GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

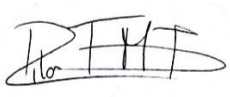
TRABAJO FIN DE GRADO

Uso de tecnologías de privatización para la compraventa de datos y de modelos de inteligencia artificial

Autor: María del Pilar Fernández-Martos Truyols

Correo electrónico del autor: 201705241@alu.comillas.edu

Director: Jaime Pereña Pinedo

Firma del alumno: 

Firma del director: 

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

Uso de tecnologías de privatización para la compraventa de datos y de modelos de inteligencia artificial

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

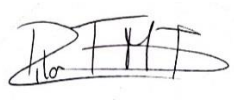
curso académico 2021/22 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: María del Pilar Fernández-Martos Truyols Fecha:14/07/2022



Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Jaime Pereña Pinedo Fecha: 14/07/2022

Agradecimientos

Me gustaría agradecer a todas aquellas personas que de una forma u otra me han guiado hasta aquí. Principalmente a mi familia, por sus ánimos continuos, y a la universidad Pontificia Comillas, por los conocimientos transmitidos a lo largo del desarrollo de mi carrera y por haberme dado la oportunidad de formarme como persona y profesional que aspira a la excelencia. Gracias a mi director de TFG, Juan de Norverto por responder a mis dudas y preguntas.

A todos los profesionales que han aportado sus opiniones, ideas y recomendaciones a mi proyecto empresarial, por dedicarme su tiempo, sus conocimientos y permitirme entrevistarles. Gracias a ellos este proyecto es más sólido y cuenta con la validación de profesionales referentes en sus sectores de los que he podido aprender.

Finalmente, me gustaría dar espacialmente las gracias a mi tutor de TFG, Jaime Pereña Pinedo, sin el cual este estudio no habría podido llevarse a cabo, por su ayuda incondicional, su paciencia, sus consejos e implicación a lo largo del desarrollo de este estudio. Por haberme dado la libertad de elegir conjuntamente el contenido de este proyecto, en el que he aprendido mucho sobre temas que me eran desconocidos hasta ese momento. Este proyecto me ha abierto un mundo, el de la tecnología, apasionante del que todavía me queda mucho por aprender. También por enseñarme metodologías de trabajo y el gusto por el trabajo bien hecho. Y por haberme facilitado los contactos de los profesionales a los que he podido entrevistar.

Muchas gracias a todos.

USO DE TECNOLOGÍA DE PRIVATIZACIÓN PARA LA COMPRAVENTA DE DATOS Y DE MODELOS DE INTELIGEMCIA ARTIFICIAL

Autor: Fernández-Martos Truyols, María del Pilar

Director: Pereña Pinedo, Jaime

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Marketplace de compraventa de datos y de modelos de IA utilizando tecnologías de privatización como la encriptación homomórfica, differential privacy, anonimización y datos sintéticos. Desarrollando una aplicación que demuestra la ejecución de modelos con datos encriptados en Azure. Además, se ha desarrollado un modelo de negocio detallado.

Palabras clave: inteligencia artificial, nube, machine learning, encriptación homomórfica, differential privacy, datos sintéticos, modelo de negocio

1. Introducción

Los datos son un activo cada vez más valioso, según un informe de BBVA *"De hecho, junto a sus empleados, muchas empresas consideran que la información es su activo más valioso y diferenciador"* (BBVA, 2018) y *"las empresas aumentarán la inversión en un 20% en análisis de datos en los próximos cinco años"* (dir&ge, 2021) y el desarrollo de herramientas que permiten utilizarlos a través de la inteligencia artificial y los modelos de machine learning. Además según el informe de IBM (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012) sobre el uso del Big Data en el mundo, la mayoría de las empresas encuestadas se encuentra todavía en fases de desarrollo y de fijar objetivos al respecto, por lo que es previsible que todavía haya mucha evolución en estos sectores.

Esta rápida evolución se debe en gran medida a la madurez de las tecnologías, a la ventaja competitiva que supone aprovechar y analizar correctamente la información de los datos. Las tecnologías han evolucionado mucho concretamente las relacionadas con la privacidad y la protección de los datos que son las que se analizan con detalle en este proyecto, como son la encriptación homomórfica, la differential privacy, la anonimización y los datos sintéticos.

Además, la Unión Europea ha unificado las legislaciones de los países sobre privacidad y protección de datos en la GDPR, por lo que los usuarios y las empresas son más conscientes de la importancia de la privacidad de los datos.

Por ello la propuesta de este proyecto es crear una empresa que facilite el uso de los datos, los modelos y las tecnologías de privacidad con el objetivo de reducir las ventajas competitivas de las grandes empresas frente a las pymes.

2. Definición del proyecto

El proyecto consta de dos partes, en primer lugar, un análisis tecnológico y en segundo lugar un análisis económico, legal y empresarial.

El análisis tecnológico hace una descripción de las tecnologías estudiadas y desarrolla un programa escrito en C# en encriptación homomórfica que se describe en el apartado 3 con más detalle.

En el grupo de las tecnologías generales, la inteligencia artificial es la capacidad de las máquinas de realizar acciones o de tener capacidades que tradicionalmente se relacionan con la inteligencia humana. Machine learning es el proceso de aprendizaje automático a través del cual las máquinas desarrollan algoritmos que se entrenan gracias a un gran volumen de datos que recibe el nombre de conjunto de entrenamiento, después se verifica el correcto funcionamiento de los algoritmos con un conjunto de datos diferente, que recibe el nombre de conjunto de test. Deep learning es una categoría del ML en la que se precisa un menor control humano para el entrenamiento de los algoritmos, siendo mucho más autónomo.

En el grupo de las tecnologías específicas de este proyecto encontramos: En primer lugar, se describe la differential privacy, que consiste en la incorporación de ruido estadístico en un conjunto de datos con el objetivo de proteger los datos sin que se afecten a los resultados derivados de la base de datos con ruido. En segundo lugar, la anonimización es el método a través del cual se elimina parte de la información de un conjunto de datos con el objetivo de mantener la privacidad y la protección de los datos de carácter personal. En tercer lugar, los datos sintéticos son datos creados de forma artificial a partir de algoritmos que pretenden simular los datos reales de una base de datos real manteniendo las proporciones, las relaciones entre los datos y los valores estadísticos de los datos originales. La encriptación homomórfica es un tipo de encriptación que a diferencia de la encriptación tradicional que solamente permite almacenar y enviar los datos encriptados, la encriptación homomórfica también permite realizar operaciones matemáticas cuando los datos están encriptados, obteniendo unos resultados que también están encriptados.

El análisis económico y legal del proyecto comienza con un análisis de otros modelos de negocio con propuestas similares como el Marketplace de Microsoft y el AWS Data Exchange analizando

sus características, los errores cometidos y comparándolas con la propuesta de este TFG. Además, se ha realizado un Bussines Model Canvas y una cuenta de pérdidas y ganancias, un análisis PEST y un DAFO. El modelo de negocio y su funcionamiento se explica con detalle en el apartado 3.

3. Descripción del modelo/sistema/herramienta

Programas de encriptación homomórfica

Primero, se hace una descripción del modelo de encriptación homomórfica y en segundo lugar se describe el funcionamiento de la empresa y sus dos líneas de negocio.

El proyecto tiene dos códigos con objetivos muy diferentes, el primero tiene como objetivo ilustrar el funcionamiento de la encriptación homomórfica, la forma en la que debe ser programado y comprender el alcance que tiene en términos de cálculo. Este primer código no busca reflejar la experiencia de un usuario real del código, sino que se comprenda cómo funciona.

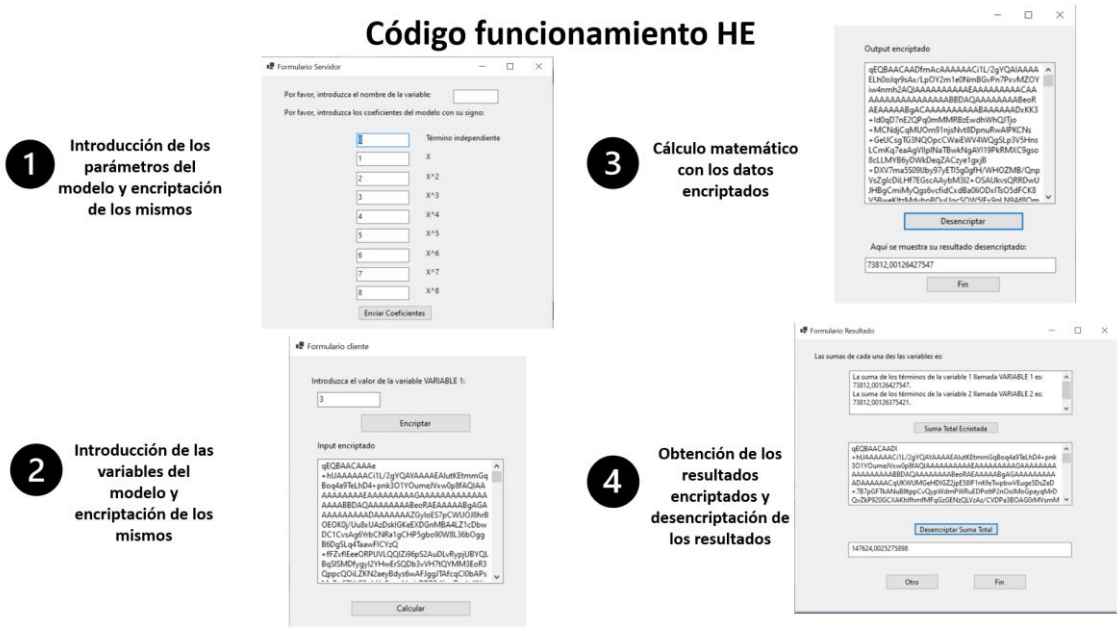


Ilustración 1: Código de funcionamiento de la HE (elaboración propia)

El objetivo del segundo código es, simplificando al máximo las operaciones que realiza, ilustrar las funciones de los tres agentes que intervienen, las funciones y las relaciones que tendrían en una experiencia real. De hecho, este modelo puede ser consultado en Azure (https://pilartfg.azurewebsites.net/).

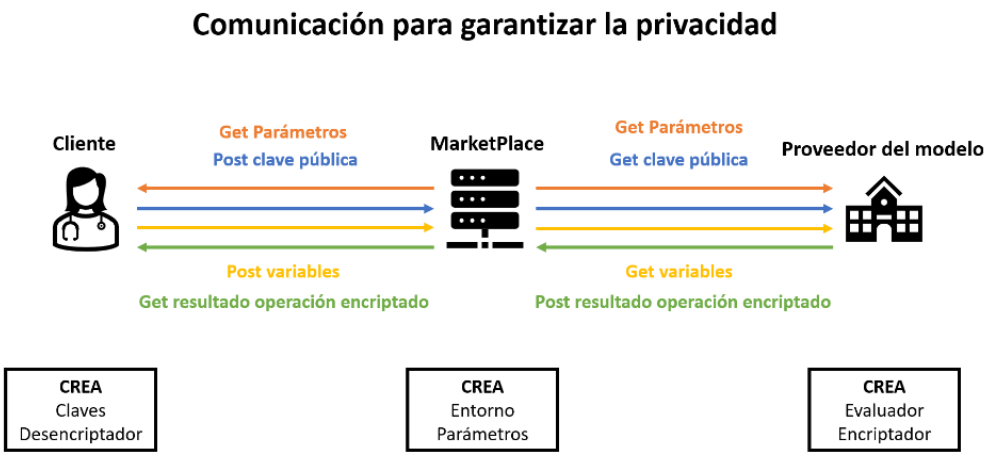


Ilustración 2: Comunicación para garantizar la privacidad (elaboración propia)

Modelo de negocio

La empresa propuesta tiene dos líneas de negocio que se complementan entre sí, la compraventa de datos y la compraventa de modelos. Ambos se han contemplado para un nicho de clientes muy específico para los inicios de la empresa, sin embargo, puede ser aplicado a cualquier sector en el futuro.

La compraventa de datos consiste en vender datos a nuestros clientes que hayan sido privatizados utilizando las tecnologías de anonimización, de differential privacy y de datos sintéticos. Este servicio sería personalizado para cada cliente buscado los datos que se ajustan a sus necesidades. Los clientes de esta línea de negocio son las empresas de márketing que únicamente en España hay unas 15.000. La razón de escoger estas empresas es que la venta de datos es cara por lo que la mayoría de las pymes no puede invertir en ello. Sin embargo, el márketing suele estar subcontratado a empresas especializadas, que podrían utilizar los datos comprados para mejorar las campañas y estrategias de varios de sus clientes distribuyendo los costes entre muchas empresas. Este modelo podría aplicarse de igual forma a la consultoría, que también suele subcontratarse. Los datos se obtendrían de fuentes públicas y privadas, y se daría al vendedor las herramientas para la privatización de los datos antes de realizar el envío, encriptado por supuesto.

Ilustración 3: Compra venta de datos (elaboración propia)

La compraventa de modelos consiste en el "alquiler-uso" de modelos que utilicen la tecnología explicada antes, la encriptación homomórfica para garantizar la privacidad. Esta línea de negocio se ha diseñado para el sector sanitario, en el que se tienen centros de investigación, universidades y laboratorios que diseñan los modelos de predicción y de ayuda en el diagnóstico. Estos modelos deberían ser accesibles para mejorar la calidad de la sanidad a nivel global. Además, los pacientes quieren mantener la protección de sus datos médicos y las instituciones creadoras de modelos no quieren renunciar a su propiedad intelectual. La propuesta de este proyecto consiste en que los modelos, al operar con encriptación homomórfica operan con los datos de los pacientes encriptados y los resultados obtenidos del modelo también están encriptados de tal forma que solamente el paciente y el médico con autorización del paciente puedan desencriptarlos. Los modelos al ser desarrollados por centros prestigiosos serían fáciles de vender y al ir a volumen de ventas podrían venderse por un precio asequible. Esto permite a los centros de investigación conseguir ingresos adicionales para continuar su labor, a la vez que dan a conocer su trabajo de forma internacional. Los pacientes pueden beneficiarse de las investigaciones a nivel global, y contribuir a que se siga investigando. Al igual que la venta de datos, esta línea de negocio puede aplicarse a otros sectores como la hostelería o la ingeniería.

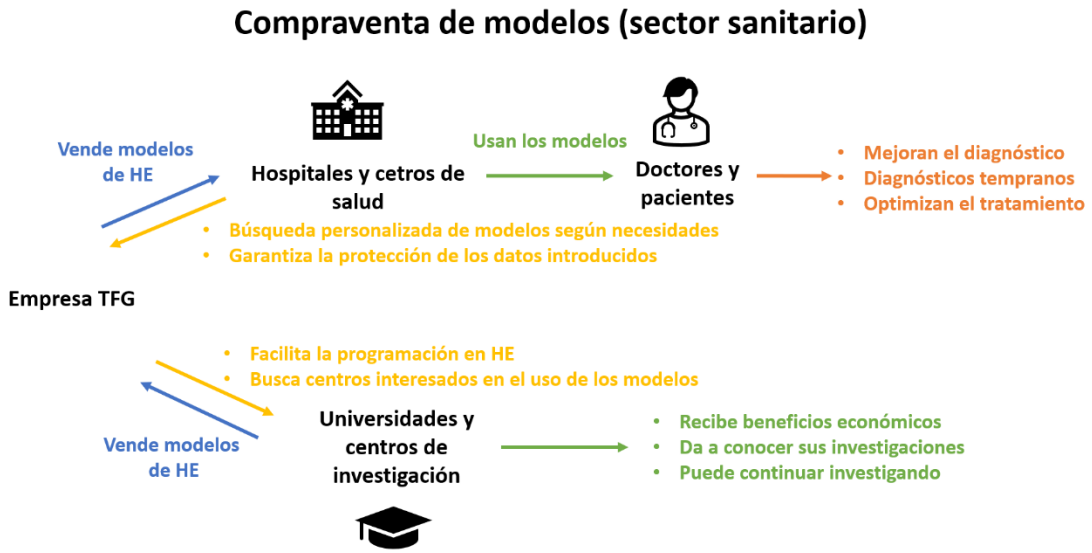


Ilustración 4: Compra venta de modelos (elaboración propia)

4. Resultados

La mayor parte de los resultados ya han sido presentados en el apartado anterior por estar muy relacionados. Aunque se mencionan brevemente resultados tecnológicos y resultados económicos.

Para obtener los códigos se han realizado comprobaciones con números decimales, negativos, con operaciones más sencillas como sumas y restas y operaciones más complejas como productos y cocientes. Además, como el segundo código se ha subido a Azure, que es la nube pública de Microsoft se han realizado comprobaciones del funcionamiento, de forma local, de forma remota en diferentes ordenadores, y de forma remota en un mismo equipo. Se ha comprobado que los resultados de operar con HE tienen una precisión de hasta la tercera cifra decimal que permite considerarlo válido para un gran número de aplicaciones.

Los resultados económicos muestran que el modelo de negocio es rentable como puede verse en las gráficas de ingresos y gastos a lo largo de tres años.

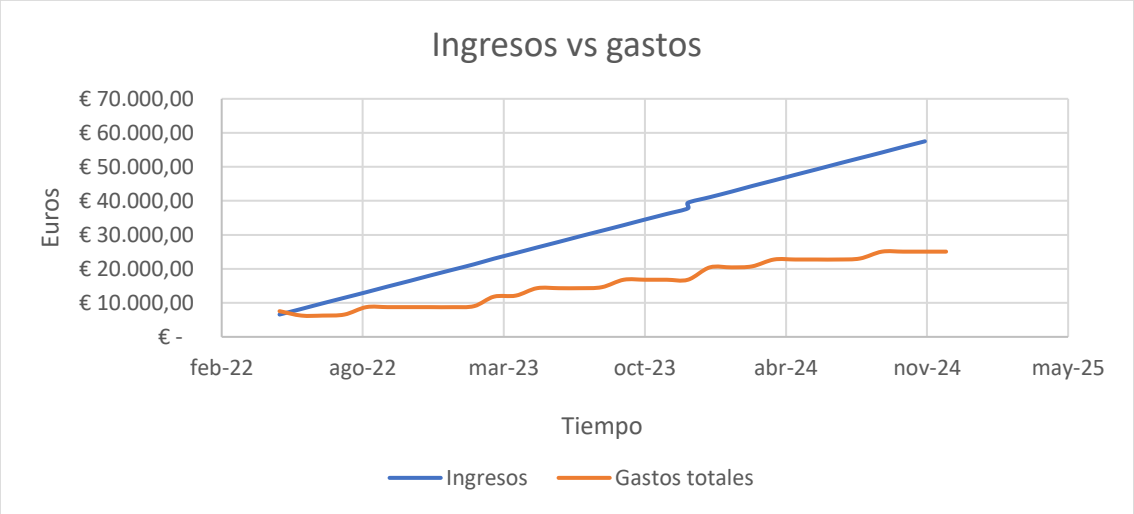


Ilustración 5. Ingresos vs gastos

5. Conclusiones

Para terminar, se concluye que existe una necesidad no satisfecha en el mercado de los datos y de los modelos. Además, las tecnologías están suficientemente maduras para poder aportar valor añadido a este proyecto. Se ha demostrado el funcionamiento y las aplicaciones reales de la encriptación homomórfica y los usos de las demás tecnologías. También, se demuestra que las líneas de negocio propuestas son aplicables a sectores muy diferentes resolviendo las necesidades particulares de cada uno de ellos. Por último, se demuestra con el análisis económico que es rentable y que puede tener éxito.

6. Referencias

BBVA. (2018). *Cómo los datos cambiarán el mundo, ¡una vez más!* Obtenido de https://www.bbva.com/es/datos-cambiaran-mundo-vez-mas/

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos.* IBM Institute for Business Value. Obtenido de https://www.fundacionseres.org/Lists/Informes/Attachments/951/IBM%20Analytics%20el%20uso%20de%20big%20data%20en%20el%20mundo%20real%20-%20Como%20las%20empresas%20mas%20innovadoras%20extraen%20valor%20de%20datos%20inciertos.pdf

dir&ge. (2021). *La inversión en análisis de datos crecerá un 20% en los próximos cinco años.* Obtenido de https://directivosygerentes.es/innovacion/inversion-analisis-datos-crecera-20-por-ciento-proximos-cinco-anos

USE OF PRIVATIZATION TECHNOLOGY FOR THE PURCHASE AND SALE OF DATA AND ARTIFICIAL INTELLIGENCE MODELS

Author: Fernández-Martos Truyols, María del Pilar.
Supervisor: Pereña Pinedo, Jaime
Collaborating Entity: ICAI– Universidad Pontificia Comillas)

ABSTRACT

Marketplace for buying and selling data and AI models using privatization technologies such as homomorphic encryption, differential privacy, anonymization, and synthetic data. Developing an application that demonstrates the execution of models with encrypted data in Azure. In addition, a detailed business model has been developed.

Keywords: artificial intelligence, cloud, machine learning, machine learning, homomorphic encryption, differential privacy, synthetic data, business model

1. Introduction

Data is an increasingly valuable asset, according to a BBVA report "*In fact, along with their employees, many companies consider information to be their most valuable and differentiating asset*" (BBVA, 2018) and "*companies will increase investment by 20% in data analysis in the next five years*" (dir&ge, 2021) and the development of tools that allow them to be used through artificial intelligence and machine learning models. Moreover, according to the IBM report (Schroeck, Shockley, Smart, Romero-Morales, & Tufano, 2012) on the use of Big Data in the world, most of the companies surveyed are still in development stages and setting goals in this regard, so it is foreseeable that there is still a lot of evolution in these sectors.

This rapid evolution is largely due to the maturity of the technologies and the competitive advantage that comes from correctly exploiting and analyzing data information. Technologies have evolved a lot, especially those related to privacy and data protection, which are the ones analyzed in detail in this project, such as homomorphic encryption, differential privacy, anonymization, and synthetic data.

In addition, the European Union has unified the laws of the countries on privacy and data protection in the GDPR, so users and companies are more aware of the importance of data privacy.

Therefore, the proposal of this project is to create a company that facilitates the use of data, models, and privacy technologies with the aim of reducing the competitive advantages of large companies over SMEs.

2. Project definition

The project consists of two parts, firstly a technological analysis and secondly an economic, legal and business analysis.

The technological analysis makes a description of the technologies studied and develops a program written in C# in homomorphic encryption which is described in section 3 in more detail.

In the group of general technologies, artificial intelligence is the ability of machines to perform actions or have capabilities that are traditionally related to human intelligence. Machine learning is the process of machine learning through which machines develop algorithms that are trained on a large volume of data called the training set, then the algorithms are checked for correct operation on a different data set, called the test set. Deep learning is a category of ML in which less human control is required for the training of the algorithms, being much more autonomous.

In the group of technologies specific to this project we find: Firstly, differential privacy is described, which consists in the incorporation of statistical noise in a dataset with the objective of protecting the data without affecting the results derived from the database with noise. Secondly, anonymization is the method by which part of the information is removed from a dataset in order to maintain privacy and protection of personal data. Thirdly, synthetic data are artificially created data based on algorithms that aim to simulate real data from a real database while maintaining the proportions, data relationships and statistical values of the original data. Homomorphic encryption is a type of encryption that, unlike traditional encryption that only allows to store and send encrypted data, homomorphic encryption also allows to perform mathematical operations when the data is encrypted, obtaining results that are also encrypted.

The economic and legal analysis of the project begins with an analysis of other business models with similar proposals such as the Microsoft Marketplace and the AWS Data Exchange analyzing their characteristics, the mistakes made and comparing them with the proposal of this TFG. In addition, a Business Model Canvas and a profit and loss account, a PEST analysis, and a SWOT. The business model and its operation are explained in detail in section 3.

3. Model/System/Tool description

Homomorphic encryption software

First, a description of the homomorphic encryption model is given and secondly, the operation of the company and its two lines of business are described.

The project has two codes with very different objectives, the first one aims to illustrate how homomorphic encryption works, how it should be programmed and to understand the scope it has in terms of computation. This first code does not seek to reflect the experience of a real user of the code, but to understand how it works.

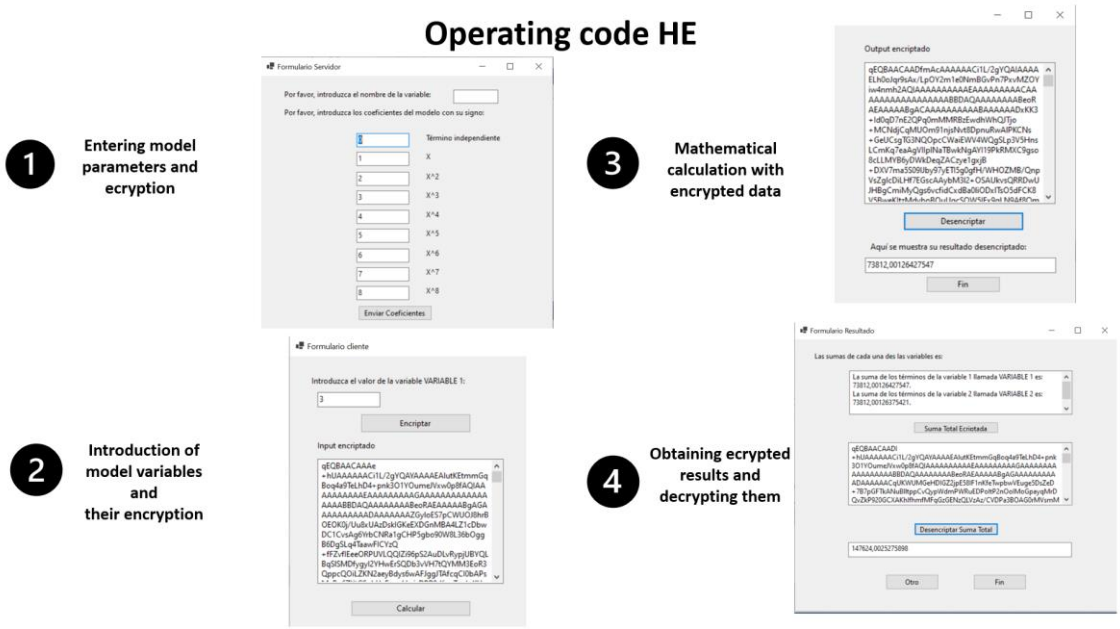


Illustration 1: HE operating code (own elaboration)

The objective of the second code is, by simplifying as much as possible the operations it performs, to illustrate the functions of the three agents involved, the roles and relationships they would have in a real experience. In fact, this model can be consulted in Azure. (https://pilartfg.azurewebsites.net/)

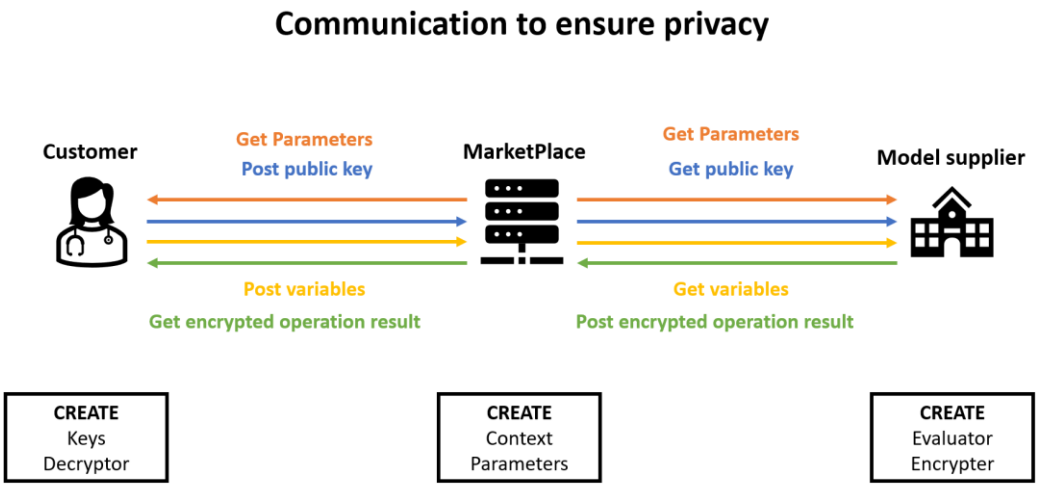


Illustration 2: Communication to ensure privacy (own elaboration)

Business model

The proposed company has two lines of business that complement each other, the purchase and sale of data and the purchase and sale of models. Both have been contemplated for a very specific customer niche for the start-up of the company, however, it can be applied to any sector in the future.

Data buying and selling consists of selling data to our customers that has been privatized using anonymization, differential privacy and synthetic data technologies. This service would be customized for each customer by finding the data that fits their needs. The clients of this line of business are marketing companies, of which there are about 15,000 in Spain alone. The reason for choosing these companies is that the sale of data is expensive, and most SMEs cannot invest in it. However, marketing is usually outsourced to specialized companies, which could use the purchased data to improve the campaigns and strategies of several of their clients by distributing the costs among many companies. This model could be applied in the same way to consulting, which is also often outsourced. The data would be obtained from public and private sources, and the vendor would be given the tools for data privatization before shipping, encrypted of course.

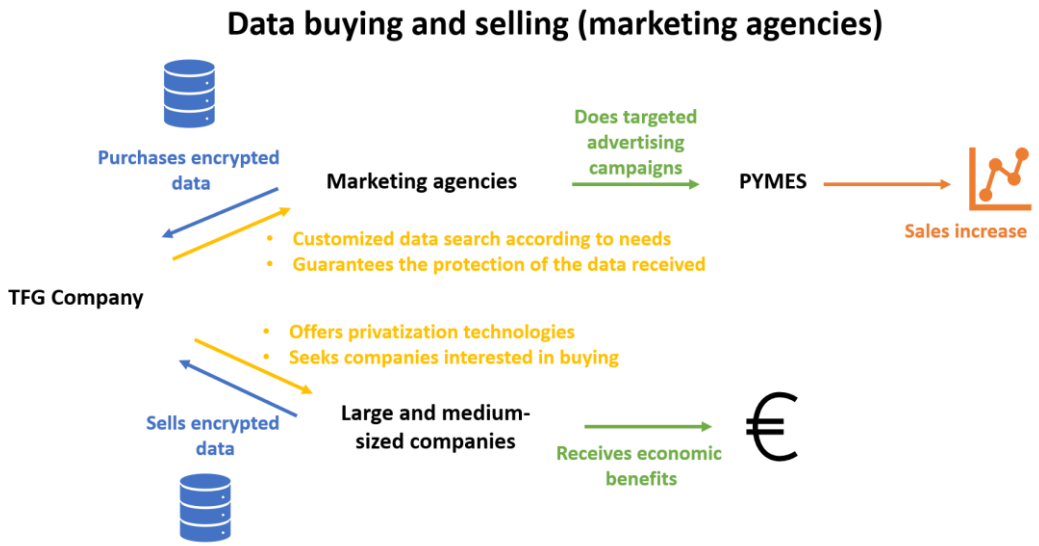


Illustration 3: Buying and selling of data (own elaboration)

The purchase and sale of models consists of the "rental-use" of models that use the technology explained above, homomorphic encryption to guarantee privacy. This line of business has been designed for the health sector, where research centers, universities, and laboratories design predictive and diagnostic models. These models should be accessible to improve the quality of healthcare globally. In addition, patients want to maintain the protection of their medical data and the institutions that create the models do not want to give up their intellectual property. The proposal of this project is that the models, operating with homomorphic encryption, operate with encrypted patient data and the results obtained from the model are also encrypted so that only the patient and the physician with the patient's authorization can decrypt them. The models being developed by prestigious centers would be easy to sell and by going to volume sales could be sold for an affordable price. This allows the research centers to earn additional income to continue their work, while making their work known internationally. Patients can benefit from research globally and contribute to further research. As with the sale of data, this line of business can be applied to other sectors such as hospitality or engineering.

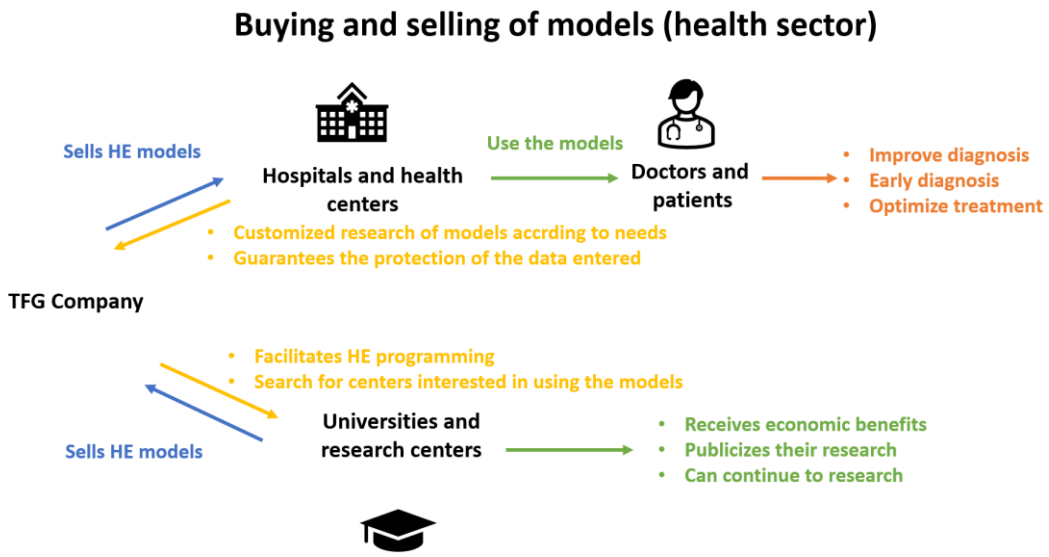


Illustration 4: Buying and selling models (own elaboration)

4. Results

Most of the results have already been presented in the previous section because they are closely related. However, technological results and economic results are briefly mentioned.

To obtain the codes, checks have been performed with decimal numbers, negatives, with simpler operations such as additions and subtractions and more complex operations such as products and quotients. In addition, as the second code has been uploaded to Azure, which is Microsoft's public cloud, performance checks have been performed locally, remotely on different computers, and remotely on the same computer. It has been proven that the results of operating with HE has an accuracy of up to the third decimal place that allows it to be considered valid for a large number of applications.

The economic results show that the business model is profitable as can be seen in the graphs of income and expenses over three years.

5. Conclusions

To conclude, there is an unmet need in the data and model market. Furthermore, the technologies are sufficiently mature to be able to add value to this project. The operation and real applications of homomorphic encryption and the uses of the other technologies have been demonstrated. It is also shown that the proposed lines of business are applicable to very different sectors, solving the particular needs of each one of them. Finally, the economic analysis demonstrates that it is profitable and can be successful.

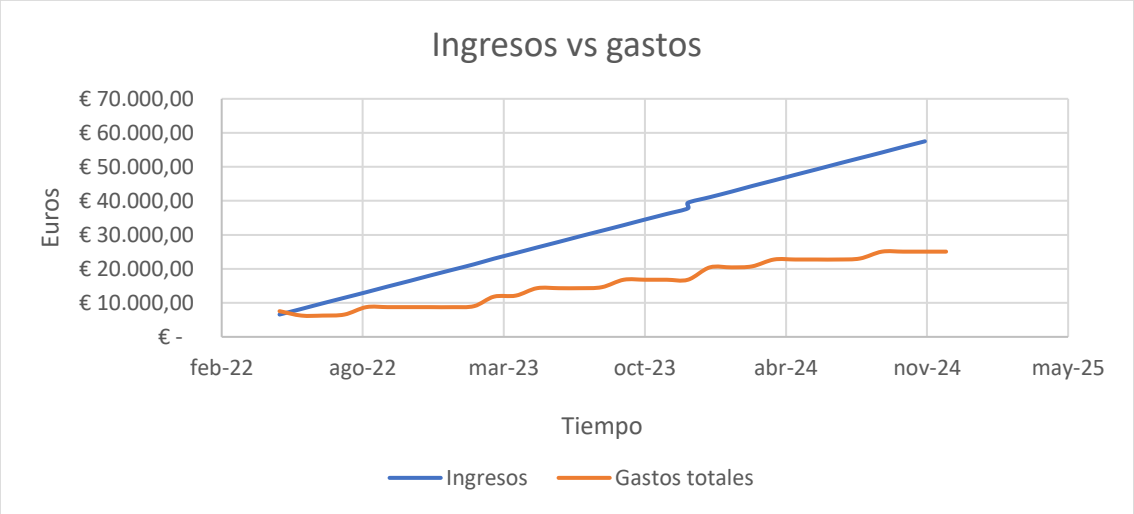


Ilustración 6: Income vs. expenses

6. References

BBVA. (2018). *Cómo los datos cambiarán el mundo, ¡una vez más!* Obtenido de https://www.bbva.com/es/datos-cambiaran-mundo-vez-mas/

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos.* IBM Institute for Business Value. Obtenido de https://www.fundacionseres.org/Lists/Informes/Attachments/951/IBM%20Analytics%20el%20uso%20de%20big%20data%20en%20el%20mundo%20real%20-%20Como%20las%20empresas%20mas%20innovadoras%20extraen%20valor%20de%20datos%20inciertos.pdf

dir&ge. (2021). *La inversión en análisis de datos crecerá un 20% en los próximos cinco años.* Obtenido de https://directivosygerentes.es/innovacion/inversion-analisis-datos-crecera-20-por-ciento-proximos-cinco-anos

GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

TRABAJO FIN DE GRADO

Uso de tecnologías de privatización para la compraventa de datos y de modelos de inteligencia artificial

Autor: María del Pilar Fernández-Martos Truyols

Correo electrónico del autor: 201705241@alu.comillas.edu

Director: Jaime Pereña Pinedo

Firma del alumno: 

Firma del director: 

ÍNDICE

ÍNDICE DE IMÁGENES

ÍNDICE DE TABLAS

Capítulo 1: Introducción y planteamiento del proyecto

En los últimos años las menciones en medios de comunicación de temas como la inteligencia artificial, el machine learning, los modelos de predicción, el Big data, la privacidad y la protección de los datos ha aumentado considerablemente. Sin embargo, la mayoría de las pequeñas y medianas empresas en España todavía no han podido implementar ninguna de estas tecnologías y probablemente pasen todavía unos años hasta que esto ocurra.

Esto lleva a la motivación detrás del tema escogido y a la visión con la que se ha ido moldeando este proyecto. La idea de hacer trabajo fin de grado sobre los datos y los modelos de machine learning utilizando tecnologías de privacidad, surgió con la intención de reducir las ventajas competitivas que tienen las grandes empresas frente a las pequeñas y medianas empresas, concretamente en el ámbito de la tecnología.

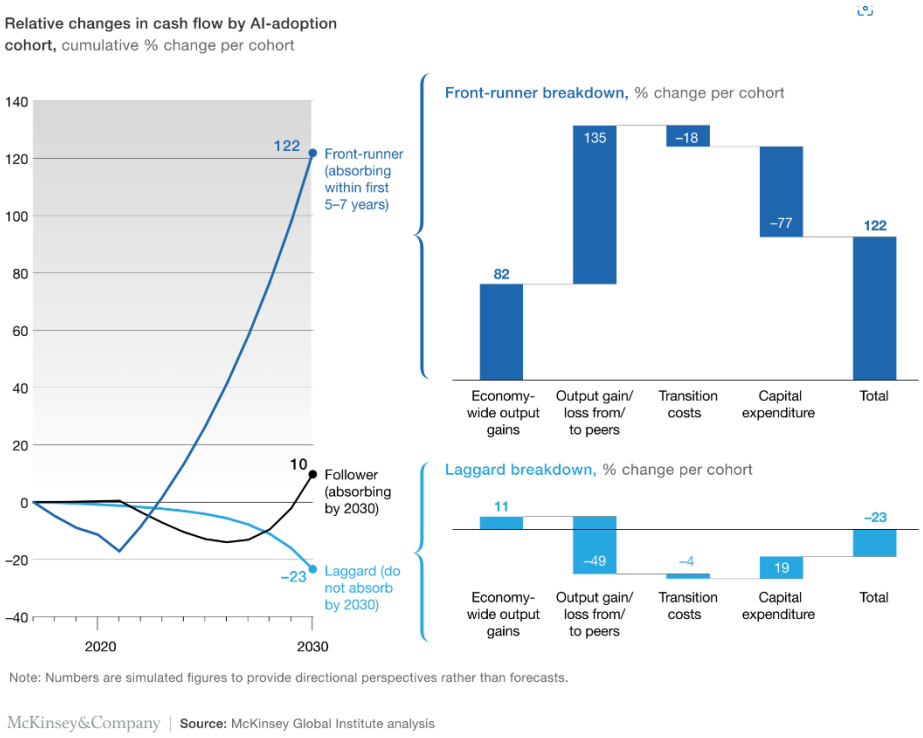


Ilustración 7: Cambios relativos en el flujo de caja por la adopción de la IA (Bughin, Seong, Manyika, Chui, & Joshi, 2018)

Había muchas formas en las que se podría haber enfocado este problema. Sin embargo, tras realizar una investigación entre las tecnologías que se estaban desarrollando dentro de

Microsoft se llegó a algunas de las que se describen en este documento y que permitían hacer frente al problema de la desventaja competitiva, especialmente en al ámbito de la tecnología, en concreto en el uso de los datos y de modelos de predicción.

El alcance de este proyecto es tratar de resolver de forma teórico-práctica el problema mencionado, para ello se realiza un enfoque teórico de las tecnologías mediante una revisión de bibliografía. El resto del proyecto es teórico práctico ya que se han realizado demostraciones tanto mediante el desarrollo de una simulación con encriptación homomórfica (se explica con detalle en el proyecto tanto la tecnología como el desarrollo de la demo) y la creación de un Business Plan en Excel a cinco años vista.

Este proyecto busca dar respuesta a algunas preguntas como, por ejemplo: ¿Existe la necesidad no resuelta de un lugar en el que se puedan comprar datos de forma personalizada?, ¿Existe la necesidad no resuelta de un lugar en el que se puedan adquirir modelos de predicción?, ¿Existe en el mercado un servicio similar al propuesto?, ¿Garantizan las soluciones del mercado las crecientes exigencias de privacidad y protección de los datos?, ¿Qué tecnologías podrían facilitar que se cumpla con la privacidad?, ¿Son accesibles y útiles las propuestas del mercado?, ¿Qué formas existen de diferenciación?, ¿Qué productos o servicios adicionales pueden ofrecerse que los clientes valoren?, ¿Es factible ejecutar esta propuesta en la nube?, ¿Se puede crear una empresa económicamente rentable que solucione este problema? ¿Contribuye este proyecto a los ODS?

Estas preguntas pueden traducirse en hipótesis a validar y por lo tanto en objetivos a perseguir en la realización de este proyecto que viene descritos de forma concreta a continuación.

Objetivos

Los objetivos principales de este Trabajo fin de grados son los siguientes:

- Estudiar y describir las diferentes tecnologías que se van a emplear en la realización de este trabajo entre las que se encuentran la encriptación homomórfica, *differential privacy*, la anonimización de datos y los datos sintéticos explicando sus ventajas y desventajas, justificando su utilización en este proyecto.
- Creación de un modelo programado en C# que utilice encriptación homomórfica en el que se demuestre su utilidad para proyectos reales, además de mostrar la comunicación

entre los diferentes agentes que participan del uso del modelo a fin de garantizar la protección de los datos utilizados.

- Demostrar la necesidad que existe en el mercado tanto de las tecnologías descritas como de la necesidad de desarrollar un Marketplace que proporcione una solución a los problemas que se mencionarán a continuación.
- Ejecutar un modelo de negocio que sea realista y viable económicamente mediante la realización de un estudio exhaustivo de los primeros años de la empresa y teniendo en cuenta su posible crecimiento a lo largo de los años, así como la expansión a diferentes sectores.

Metodología de trabajo

El proyecto y la idea original ha ido evolucionando ligeramente a medida que se avanzaba en la investigación de las tecnologías, se hablaba con expertos en la materia y se investigaban otras opciones en el mercado a fin de adaptar el proyecto a las necesidades reales de las empresas y a las capacidades y limitaciones reales de las tecnologías utilizadas. Por lo que se ha empleado una metodología de trabajo que permitiera un trabajo flexible y que pudiera adaptarse a la evolución del proyecto.

La metodología empleada para la realización de este proyecto ha consistido en:

1. Investigación de las posibles tecnologías a utilizar desde el punto de vista teórico.
2. Investigación del funcionamiento práctico de la encriptación homomórfica, las librerías disponibles, los lenguajes de programación utilizados y los ejemplos.
3. Programación de una demostración en C# de un programa que utiliza la encriptación homomórfica, tras el aprendizaje del lenguaje y de las peculiaridades de la librería.
4. Estudio de mercado y realización de una cuenta de pérdidas y ganancias en Excel.
5. Entrevistas a expertos en diferentes materias que se han tratado en este documento que ofrecían una visión externa del proyecto y han permitido ir adaptando la idea original del proyecto a la idea que finalmente se presenta.
6. En la redacción del documento se ha utilizado la metodología de trabajo ágil que ha permitido ir adaptando el proyecto a los posibles cambios que fueran surgiendo fruto de la investigación y del mayor conocimiento sobre el tema.

Recursos a emplear

Para el desarrollo de este trabajo fin de grado se hará uso de los siguientes recursos:

- Licencia educacional Microsoft Excel
- Licencia educacional Microsoft Word
- Licencia educacional Microsoft Teams para las reuniones con expertos
- Visual Studio, es un editor de código que se ha utilizado para la programación en C# de un programa de encriptación homomórfica
- Microsoft SEAL (librería de código abierto para programar utilizando encriptación homomórfica)
- Microsoft Azure, la nube de Microsoft en la que se pueden ejecutar aplicaciones y programas informáticos sin la necesidad de utilizar un editor de código.

Estructura del proyecto

La estructura del proyecto pretende facilitar la comprensión del lector para que llegue de forma natural, lógica y razonada a las soluciones aquí planteadas. Por eso existen dos grandes apartados, el primero el estudio tecnológico y el segundo el estudio económico-empresarial.

En el estudio tecnológico se comienza dando un contexto de las tecnologías, luego se hace una revisión de todas las tecnologías que en algún momento dado se van a mencionar o a utilizar en el proyecto y por último se realiza la descripción detallada del código y del funcionamiento de una de las tecnologías descritas, la encriptación homomórfica.

En el estudio económico-empresarial se comienza dando un contexto de la situación de las empresas respecto a las tecnologías mencionadas con el objetivo de dar argumentos que validen la empresa que se desarrolla más adelante, después se realiza un análisis de la competencia y de otras empresas que ofrecen servicios similares, las características que ofrecen y las formas en las que este proyecto logra diferenciarse de ellas. Después se realiza un breve estudio del contexto legal en el que se engloba el proyecto. Para continuar se realiza una descripción de los servicios propuestos y el funcionamiento de la empresa, de forma más detallada se ha realizado un Business Model Canvas, un análisis DAFO y un análisis PEST. Y una descripción de los usos y sectores en los que sería interesante implementar este modelo de negocio.

Por último, se realiza un estudio de las limitaciones tanto tecnológicas como económico-legales del proyecto, se analizan las mejoras futuras y se realiza un estudio aproximado de las emisiones de CO2 que produciría el proyecto. Finalmente, se describen las conclusiones del proyecto.

Capítulo 2: Estudio tecnológico

1. Planteamiento del problema desde el punto de vista tecnológico

Durante la última década se ha producido una gran adopción de las nuevas tecnologías como la inteligencia artificial, el machine learning o el Deep learning que se explican en este capítulo con detalle. Sin embargo, estas tecnologías tienen el principal problema de la dificultad de contar con bases de datos suficientes y de calidad que permitan entrenar modelos.

La gran barrera para la adopción masiva de modelos de IA es la falta de confianza asociada principalmente a los riesgos derivados de la seguridad y privacidad. Si dentro de una compañía ya es complicado acceder a todos los datos necesarios, cuando se necesita colaborar más allá de la propia organización los riesgos y las complicaciones se multiplican. La mayor parte de las herramientas de desarrollo de modelos parten del supuesto de que los conjuntos de datos están accesibles y se pueden utilizar sin restricciones.

Según el informe de IDC (Framingham, 2019) las empresas dan cada vez mayor importancia a los problemas de privacidad y de seguridad de sus datos. Además, la mitad de las empresas entrevistadas tiene muy presentes las consideraciones éticas del uso de estas tecnologías, los riesgos derivados, además de los problemas de privacidad ya mencionados.

Faltan herramientas que permitan compartir y proteger los datos, este problema se aborda de forma directa a lo largo de este capítulo ofreciendo las soluciones tecnológicas que hay disponibles en el mercado.

La adopción de estas nuevas tecnologías es cada vez mayor, a pesar de todo solo la mitad de las organizaciones que formaron parte de la encuesta consideran la IA una prioridad, según un informe de IDC (Framingham, 2019).

De las empresas que han implementado la inteligencia artificial cerca de un 60% ha alterado su modelo de negocio para adaptarlo a las nuevas circunstancias.

Los dispositivos tienen más sensores y más capacidad de comprender, almacenar e interpretar los datos. Un coche por ejemplo tiene múltiples sensores que permiten determinar la distancia a la que se encuentra de otros objetos, incluso detectar el tipo de objeto que hay próximo, su velocidad y actuar en consecuencia. Además de los sensores externos también cuentan con

sensores internos que permiten determinar el estado de los diferentes componentes, el aceite, la gasolina, la temperatura. Incluso los coches podrían llegar a comunicarse entre sí y con la red para mejorar la conducción.

Esto mismo ocurre a todos los niveles, en todas las empresas y en todos los dispositivos que utilizamos en nuestro día a día. Y las empresas son las encargadas de invertir y desarrollar estas tecnologías.

2. Tecnología existente (Antecedentes revisión del estado de la técnica)

En esta sección se describen en profundidad las diferentes tecnologías que en mayor o menor medida están implicadas en el desarrollo de este proyecto. Iniciando la descripción desde las más generales a las más específicas.

En líneas generales todas estas tecnologías están en pleno desarrollo y están normalizadas en el ámbito de las grandes empresas, siendo su uso cada vez más relevante en su actividad económica.

Para empezar, es necesario tener una definición sobre lo que es la privacidad y la protección de los datos. La Real Academia Española (RAE) define la privacidad como: "*Ámbito de la vida privada que se tiene derecho a proteger de cualquier intromisión*". El gran objetivo de la privacidad es prevenir que puedan sacarse conclusiones sobre individuos a partir de los datos que se publican.

El gran problema de la privacidad reside en que los datos pueden quedar expuestos o ser identificables, no por el estudio de una sola base de datos, sino por la combinación y el cruce de datos de diferentes fuentes que en muchas ocasiones son públicas y accesibles y que combinadas facilitan la identificación de individuos concretos. Algunas de las metodologías que utilizan los criminales son la reconstrucción de la base de datos original a partir de datos estadísticos o la identificación de los individuos de una población.

Asumir que por el simple hecho de eliminar de una base de datos la información que permita identificar a las personas individuales como el nombre, DNI, dirección o número de teléfono se garantiza la privacidad es erróneo.

Por este motivo las empresas tecnológicas como Microsoft, Google o IBM junto con universidades como el MIT o Harvard están constantemente desarrollando las tecnologías existentes y creando nuevas para solucionar este problema. De hecho, es necesario tener en cuenta el momento en el que se desarrolla este proyecto y que es posible que las tecnologías que aquí se presentan en un futuro cercano hayan evolucionado o incluso existan alternativas mejores.

2.1. Tecnologías generales

En esta sección se describen brevemente algunas tecnologías generales que forman parte del contexto del proyecto como son la nube, la inteligencia artificial, el *machine learning* y el *deep learning*. Se describen de forma breve porque no son las tecnologías prioritarias de este trabajo, pero se hace referencia a ellas en varias ocasiones. Además, se asume que son tecnologías relativamente conocidas por el lector.

2.1.1. La nube (Cloud)

La nube o *the cloud en* inglés, es el conjunto de servidores conectados entre sí que están situados en centros de datos en todo el mundo y que son gestionados y mantenidos por las diferentes empresas proveedoras de estos servicios como pueden ser Amazon Web Service, Microsoft Azure o Google Cloud. La nube permite acceder a los datos almacenados en estos servidores desde cualquier dispositivo que tenga conexión a internet. En ella se almacenan datos, se realizan operaciones con datos y se procesan aplicaciones.

La nube ofrece una gran flexibilidad a los usuarios y a las empresas ya que se adapta a las necesidades de cada momento. Además, tiene la ventaja de que reduce los costes respecto a la alternativa de tener servidores físicos, como los de mantenimiento.

La nube no podría funcionar sin las máquinas virtuales que son divisiones que se realizan dentro de un mismo servidor físico y que permiten que en un mimo equipo se ejecuten aplicaciones y almacenamiento de datos de usuarios que funcionan de manera totalmente independiente. Las máquinas virtuales permiten optimizar el software y el hardware de los centros de datos

permitiendo que el servicio ofrecido por los proveedores de cloud sea a precios muy competitivos.

Hay diferentes modelos de servicio de computación: SaaS, PaaS, IaaS, y FaaS.

- Software as a Service (SaaS): es el servicio más completo ya que el cloud ofrece no solo el almacenamiento de los datos, los servidores, los sistemas de operación, desarrollo o de gestión de la nube sino también la posibilidad de almacenar las aplicaciones en la nube.
- Platform as a Service (PaaS): las empresas y usuarios contratan el uso del centro de datos, sus servidores y la capacidad de almacenamiento. Además, reciben el servicio del cloud de suministrar herramientas que facilitan la gestión de los datos, así como el desarrollo de herramientas propias, el desarrollo de la infraestructura o los sistemas operativos. Los usuarios son los encargados de desarrollar sus propias aplicaciones, pero cuentan con herramientas del cloud para facilitar esa tarea.
- Infrastructure as a Service (IaaS): las empresas y los usuarios pagan al proveedor del cloud únicamente por el espacio de almacenamiento en el centro de datos físico y por el uso de los servidores y su capacidad de almacenamiento. El desarrollo y diseño de las aplicaciones es responsabilidad de las empresas y a diferencia del caso de PaaS no cuentan con herramientas de ayuda.
- Function as a Service (FaaS): es la forma más reciente de servicio de computación en la nube. Y también recibe el nombre de informática sin servidor ya que permite al usuario pagar únicamente por los servicios que utiliza en cada instante y en los momentos en los que nadie los esté usando no paga nada.

Hay diferentes tipos de nube: nube privada, nube pública, nube híbrida y multinube.

- Nube privada: un servidor que únicamente es utilizado por una empresa.
- Nube pública: un proveedor de servicios cloud que poseen y gestionan varios servidores que alquilan diferentes empresas. Esto es posible gracias a las máquinas virtuales.
- Nube híbrida: una empresa combina la nube privada con la nube pública.
- Multinube: una empresa tiene alquiladas diferentes nubes públicas a distintas empresas proveedoras de servicios cloud.

Operar en la nube supone un gran cambio en la forma tradicional en que las empresas piensan en los recursos informáticos. Con más del 90% de las empresas utilizando algún tipo de nube para alguna gestión de la empresa. La adopción y el interés por la nube pública no cesan, ya que

las organizaciones siguen una política de "la nube primero" para incorporar nuevas cargas de trabajo. La pandemia y el consiguiente aumento de los servicios digitales están convirtiendo la nube en la pieza central de las nuevas experiencias digitales. Por ejemplo, según un estudio realizado por Flexera, el 91% de las empresas estudiadas había utilizado una nube pública en el año 2019 (Right Scale. Flexera, 2019).

Las empresas son conscientes de que con tener nubes públicas no es suficiente y apuestan por modalidades de multinube y nubes híbridas. Según el estudio de Flexera, las empresas que utilizan la nube tienen entre 3 y 5 proveedores de nube diferentes entre públicas y privadas (Right Scale. Flexera, 2019). Los proveedores servicios en la nube siguen creando soluciones más allá de la nube pública a las ubicaciones privadas y locales, abordando las necesidades relacionadas con la soberanía de los datos, la portabilidad de las cargas de trabajo y la latencia de la red. A medida que las soluciones en la nube continúan madurando, los principales proveedores de nubes públicas están invirtiendo en soluciones verticales en la nube a medida para satisfacer las necesidades de la industria especializada. Sin embargo, siguen existiendo barreras para la adopción de la nube, como la seguridad de los datos y el cumplimiento de la normativa (en particular, la seguridad también se considera un resultado importante de la adopción de la nube), la gestión del gasto en la nube y los problemas de migración.

2.1.2. Inteligencia Artificial (IA)

La inteligencia artificial es la tecnología que permite que las máquinas y los ordenadores razonar, comportarse o tomar decisiones como lo haría un ser humano. Dentro de la IA existen dos enforques diferentes, en primer lugar, está en enfoque humano que aboga por replicar el pensamiento humano y en segundo lugar está el enfoque ideal que aboga por el uso de la razón. En definitiva, es un sistema de información que se inspira en un sistema biológico humano. La IA está acostumbrada a tratar con la ambigüedad y cambiar los comportamientos sin programación explícita, basándose en los datos recogidos, el análisis de uso y otras observaciones.

Esta tecnología engloba a muchas otras como el machine learning y el Deep learning que se explican con algo más de detalle a continuación, la visión por ordenador y el procesamiento del lenguaje natural (PLN) que, individualmente o en combinación, añaden inteligencia a las aplicaciones.

La IA tiene tres características principales: el aprendizaje, la percepción y el razonamiento.

El aprendizaje hace referencia a la capacidad de las máquinas de aprender, actualizarse y adaptarse por sí solas sin la necesidad de intervención humana. En la programación tradicional un humano desarrollaba el código que luego publicaba y en caso de que este tuviera que ser actualizado o mejorado se debía volver a sobrescribir la parte a modificar, este trabajo lo realizaba un desarrollador humano, por supuesto. Sin embargo, gracias a la IA se produce un cambio en la forma de concebir aplicaciones.

La percepción hace referencia a la capacidad de las máquinas de ser conscientes y de comprender el entorno que la rodea e incorporar esta información al aprendizaje antes mencionado. Algunas de estas percepciones del entorno pueden ser la comprensión de voces humanas, la lectura de textos o el reconocimiento de diversos objetos en imágenes.

Por último, el razonamiento hace referencia a la capacidad de las máquinas de tomar decisiones y de actuar en base al aprendizaje realizado y a las percepciones del momento. Algunos razonamientos incluyen la identificación de patrones, la realización de predicciones o la identificación de la mejor forma de actuar a fin de obtener un resultado concreto.

Gran parte de los fundamentos teóricos y tecnológicos de la IA se desarrollaron en los últimos 70 años, pero una confluencia de desarrollos la está impulsando: la combinación de grandes cantidades de datos, una mayor capacidad de procesamiento y algoritmos más inteligentes.

En primer lugar, la cantidad creciente de datos se debe entre otros motivos a la capacidad que se tiene de capturar y de almacenar más datos que favorece que los modelos sean cada vez más completos y precisos. Este punto se ha desarrollado en mayor profundidad anteriormente.

En segundo lugar, ahora se cuenta con una mayor capacidad de procesamiento que permite realizar muchas operaciones en relativamente poco tiempo, además esta capacidad aumentará en el futuro permitiendo hacer operaciones que hoy en día llevan mucho tiempo o que no son realizables. Hoy en día con la nube es posible que un gran número de máquinas virtuales realicen una misma operación optimizando los recursos y reduciendo el coste de las operaciones tanto de procesamiento como económicamente.

Por último, en la última década ha habido un gran desarrollo de los algoritmos, las tecnologías, las herramientas y las infraestructuras. Gracias a ello se han desarrollado mejores aplicaciones y servicios, esto también es debido al uso cada vez más generalizado del machine learning. Todo ello provoca una realimentación en la que al demandarse mejores aplicaciones se demandan mejores herramientas de ML para su desarrollo favoreciendo que evolucione y mejore.

La IA se ha convertido una tecnología esencial y las empresas y la sociedad están aprovechando los beneficios de estos avances. Los beneficios potenciales de la IA son enormes, empezando por la capacidad de dar sentido a enormes cantidades de datos y detectar (y predecir) patrones con más precisión que las personas. Los cinco principales sectores en los que la IA está preparada para tener un gran impacto son la energía, la sanidad, el sector aeroespacial, la cadena de suministro y la construcción. Es de esperar que se realicen grandes inversiones en AI en el corto medio plazo.

Sin embargo, la IA también plantea riesgos que incluyen consideraciones éticas, legales, normativas y cuestiones relacionadas con la privacidad, la ciberseguridad y la disponibilidad de talento para aprovechar al máximo esta oportunidad. El escepticismo podría ser el mayor problema de la IA, y surgen dos temas predominantes y polémicos sobre sus riesgos: la eliminación de puestos de trabajo y la posibilidad de que la IA supere la inteligencia humana en un momento dado.

2.1.3. Machine Learning (ML)

El machine learning o aprendizaje automático es una tecnología que pertenece a la inteligencia artificial explicada previamente. El ML engloba a su vez al Deep learning que se explica más adelante. La principal diferencia entre ambas tecnologías es en la forma en la que aprenden los algoritmos. En el caso de ML en el proceso de aprendizaje es necesario la intervención humana que es la encargada de clasificar, priorizar y los datos que se van a utilizar en el aprendizaje de los algoritmos. Estos algoritmos se utilizan para identificar patrones en los datos, y esos patrones se utilizan para crear un modelo de datos que pueda hacer predicciones.

La idea ha existido durante casi 60 años, pero la necesidad de dar sentido a grandes cantidades de datos está marcando la diferencia en este momento, junto con mejoras significativas en muchas disciplinas de la IA, como las redes neuronales y el aprendizaje profundo. El ML ha experimentado una innovación y un crecimiento masivo que abarca muchos sectores y aplicaciones que la mayoría de nosotros lo utilizamos a diario sin darnos cuenta para tareas que van desde la búsqueda en la web hasta los asistentes de voz. En la actualidad, las empresas están adoptando aplicaciones de ML en todas las líneas de negocio, ofreciendo resultados en diversas áreas que van desde el desarrollo farmacéutico y los servicios financieros hasta el análisis predictivo y las recomendaciones de productos. La capacidad de ML para abordar problemas del

mundo real de gran complejidad es una ventaja, aunque la escasez de competencias y las limitaciones de la calidad de los datos persisten.

Tanto el ML, junto con la IA, tienen la capacidad de revolucionar sectores enteros, y su adopción está aumentando en todos los sectores y zonas geográficas. Los principales proveedores de tecnología, como Amazon, Google y Microsoft, están adquiriendo e invirtiendo en I+D y en nuevos productos, incluyendo marcos y servicios en la nube o el aprendizaje automático como servicio (MLaaS). El principal atractivo de estas ofertas es que los clientes pueden empezar a utilizar rápidamente el ML sin tener que instalar software ni aprovisionar sus propios servidores, como cualquier otro servicio en la nube. DataDrivenInvestor afirma que Microsoft se ha convertido en el "*referente del aprendizaje automático*" (Sue, 2019) y ZDNet dice que la empresa tiene un "*derecho sustancial a la grandeza en el aprendizaje automático moderno*" gracias a sus esfuerzos pioneros en el procesamiento del habla, la visión y el lenguaje natural (Ray, 2019).

2.1.4. Deep Learning (DL)

El deep learning es una categoría dentro del machine learning y por lo tanto también está englobado por la IA. La principal diferencia con el machine learning es que el entrenamiento de los algoritmos recibe el nombre de aprendizaje supervisado, pero que reduce en gran medida la intervención humana en el proceso. Este tipo de algoritmos no precisan que los datos de entrenamiento estén etiquetados, por lo que las bases de datos que pueden utilizarse pueden ser más grandes y estar menos estructuradas. Esto tiene múltiples ventajas ya que permite el entrenamiento de modelos con menor supervisión humana, con mayor variedad de datos y mayor volumen haciendo del machine learning algo escalable.

El aprendizaje profundo se consigue gracias a la utilización de una gran colección de procesos conectados en muchas capas y exponiendo estos procesadores a un amplio conjunto de ejemplos. El diseño de esta estructura de algoritmos en capas, denominada redes neuronales artificiales, se inspira en la red neuronal biológica que utiliza el cerebro humano. Aunque las redes neuronales se concibieron por primera vez en la década de 1950, los avances en el diseño de algoritmos evolucionan rápidamente gracias a la mejora en la capacidad de almacenamiento de información rápida, la alta potencia de cálculo y la paralelización. El aprendizaje profundo es el más aceptado desde el punto de vista comercial, llegando a superar al ML tradicional.

Los casos de uso más destacados son la visión por ordenador, el reconocimiento de voz y el procesamiento del lenguaje natural (NLP). Entre la amplia gama de problemas que pueden resolverse gracias al DL se encuentran el diagnóstico médico, la predicción de la demanda, la pérdida de clientes y la predicción de fallos. Una de sus principales características es la capacidad de identificar patrones en datos no estructurados, como ya se ha mencionado. Sin embargo, como cualquier otra técnica de aprendizaje automático, el DL tiene sus limitaciones. En primer lugar, puede ser caro y complicado de configurar, y en segundo lugar la dificultad de disponer del gran volumen de los datos y de la calidad de estos para entrenar las redes neuronales puede ser una dificultada añadida que cabe suponer tendrá una gran inversión por parte de las empresas en un futuro próximo.

2.2. Differential data privacy (DP)

La *differential privacy* (DP de ahora en adelante) es una de las tecnologías clave para comprender este trabajo. Es un proyecto en el que colaboran Microsoft y *Harvard's Institute for Quantitative Social Science* cuyo objetivo es resolver la problemática del uso de datos en el desarrollo de modelos de *machine learning*. Este problema surge ante la imposibilidad de utilizar bases de datos completas para investigación por contener datos de carácter personal que para poder ser utilizados deben ser eliminados, anonimizados o falseados. Sin embargo, para ciertas investigaciones es importante contar con conjunto completo de datos a fin de obtener resultados con un mayor grado de fiabilidad.

Esta tecnología se basa en la hipótesis matemática por la cual, en un estudio estadístico, los datos de un individuo no pueden ser obtenidos a partir de los parámetros estadísticos obtenidos del conjunto de individuos. Es decir, que, a partir de la media, la mediana o la varianza de un conjunto de individuos no puede conocerse la contribución de un individuo en particular a ese dato global. De la misma forma se asume que para un conjunto de individuos suficientemente grande, la contribución individual de cada individuo podría ser omitida sin que afecte al cálculo estadístico. Esto se cumple para todos los individuos de cualquier base de datos. Este concepto recibe el nombre de ruido estadístico y también puede entenderse como añadir cierta aleatoriedad a los datos. Este ruido estadístico es suficientemente reducido para no afectar a los resultados obtenidos tras realizar diferentes análisis sobre los datos y al mismo tiempo es suficiente para evitar que los individuos sean identificables.

Aunque es algo deducible a partir del párrafo anterior, la DP funciona mejor cuanto mayor es el tamaño de la base de datos ya que el impacto que tiene cada individuo de forma independiente sobre los datos globales es menor.

En el caso de que se necesitaran obtener resultados directamente a partir de los datos reales por necesitarse resultados muy fieles o ajustados a la realidad y que no se hiciera uso de la DP se podría estar infringiendo las leyes de protección de datos.

Sin embargo, en un escenario en el que se utiliza DP, en lugar de utilizar los datos originales se utilizan unos datos aproximados. Para ello, es necesario crear una consulta a través de la cual se crean los datos aproximados a partir de los originales mediante la introducción del ruido estadístico del que se ha hablado con anterioridad, este proceso recibe el nombre de mecanismos de privacidad. Una vez se ha creado la base de datos que contienen ruido se envían para su procesamiento y estudio. De esta forma se evita romper con la privacidad de los datos y como se ha explicado antes se siguen obteniendo resultados muy precisos y prácticamente iguales a los que se habrían obtenido en el escenario en el que no se utiliza DP.

La DP ofrece un cálculo matemático medible de la privacidad de los datos de carácter individual. Comprendiendo el significado de privacidad explicado al inicio de este capítulo, se comprende que los datos privados de los individuos tienen un mayor riesgo de pérdida de privacidad cuantas más veces se utilicen, de tal manera que el riesgo que sufren de ser expuestos va aumentando. Esto significa que la privacidad no es un parámetro que pueda medirse de manera binaria en función de si los datos han sido o no expuestos.

La DP utiliza dos parámetros épsilon y delta para medir la privacidad de los datos y que estos no puedan ser recompuestos a partir de los resultados. Ambos parámetros son inversamente proporcionales. En la imagen a continuación puede comprenderse el concepto.

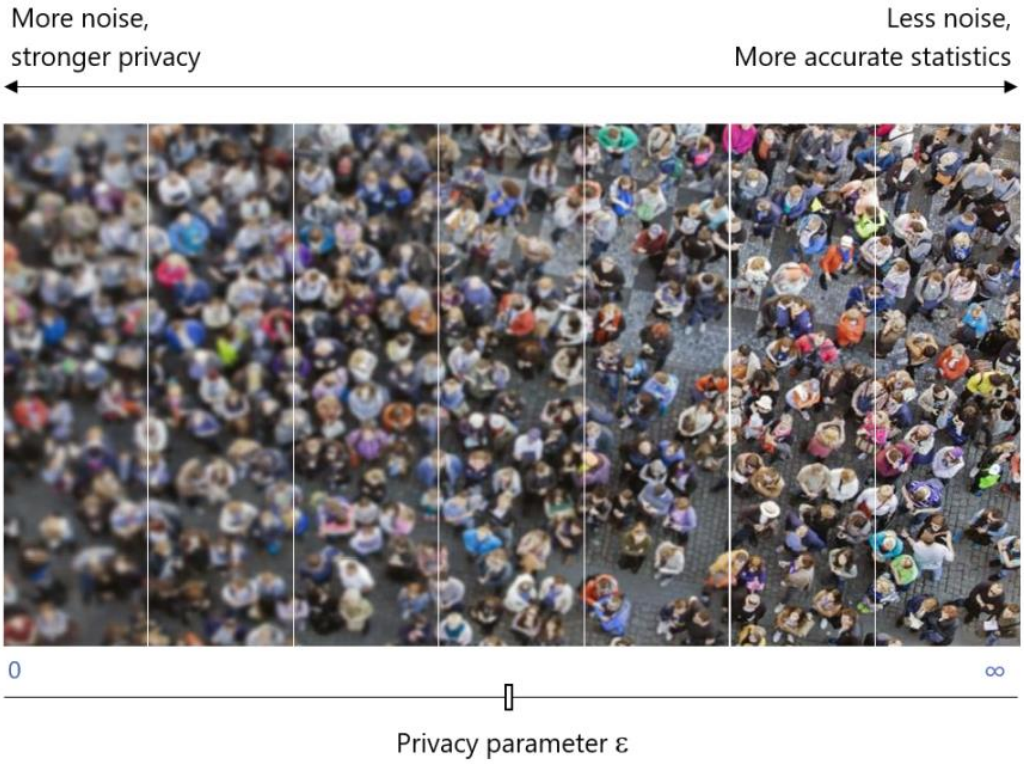


Ilustración 8: Claridad de los datos en función de épsilon (Kopp, Microsoft SmartNoise Differential Privacy Machine Learning Case Studies, 2021)

La DP es útil en escenarios en los que se realizan estudios estadísticos de los datos, en *machine learning,* la teoría de juegos o en *streaming* (Kopp, Microsoft SmartNoise Differential Privacy Machine Learning Case Studies, 2021). Aunque en el caso del entrenamiento de modelos de *machine learning* es más útil emplear otras tecnologías como el *confidential mahcine learning.*

La differential privacy impide el uso de conjuntos datos sin restricciones ofreciendo protección a nivel de grupo [OBJ][OBJ]. Sin embargo, la DP también tiene limitaciones. Una desventaja de la DP es que no funciona bien en poblaciones pequeñas y que no resulta útil en el caso de necesitar estudiar individuos concretos o a aquellos individuos que tengan características muy diferentes a las del resto, es decir que sean *outliers*. Además, es necesario tener en cuenta que al ser datos estadísticos con ruido los resultados que se obtienen son muy similares a los obtenidos con las bases de datos reales, y sumado al poco valor que tienen los datos individuales, es posible inferir, por lo tanto, a partir de los resultados estadísticos en las características de un individuo en particular, incluso si este no ha formado parte del estudio estadístico. Por ejemplo, en el caso de que el INE realice un estudio sobre el salario medio de los españoles por barrio o por calle es muy probable que a pesar de que la mayor parte de la población no haya participado en las encuestas se puedan inferir sus sueldos a partir de los resultados con bastante precisión. Es decir

que no garantiza la no inferencia estadística. Pero el objetivo de la DP es otro, en concreto ocultar a los individuos que sí que participan en las encuestas de que sus datos específicos sean descubiertos.

A pesar de todas las limitaciones explicadas en el párrafo anterior, la DO tiene muchas ventajas que ya se han explicado y tiene mayores garantías de privacidad que otras tecnologías que se mencionan en este proyecto como son la anonimización de los datos.

La DP puede combinarse con otras tecnologías que se explicarán más adelante en este documento como por ejemplo los datos sintéticos. Combinar tecnologías tiene sus ventajas ya que pueden compensar las carencias que tienen unas y otras sin comprometer la calidad de las bases de datos en exceso.

2.3. Homomorphic encryption (HE)

En primer lugar, la encriptación tradicional puede tener diferentes estructuras, las de tipo clave simétrica y las de estructura de clave pública.

Las de estructura simétrica fueron los primeros métodos de encriptación, en la que tanto el creador del mensaje encriptado como el receptor tenían la misma clave de desencriptación que coincidía con la clave de encriptación. Un ejemplo de esto es la máquina Enigma, que fue utilizada durante la segunda guerra mundial y que permitió a los alemanes comunicarse con fines militares.

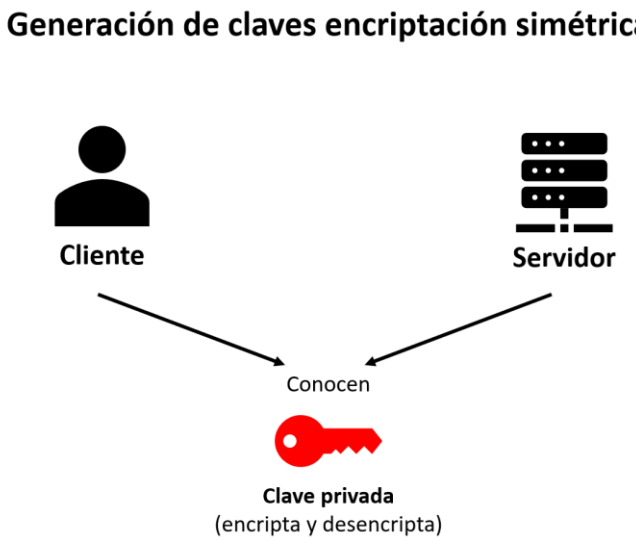


Ilustración 9: Generación de claves en encriptación simétrica (elaboración propia)

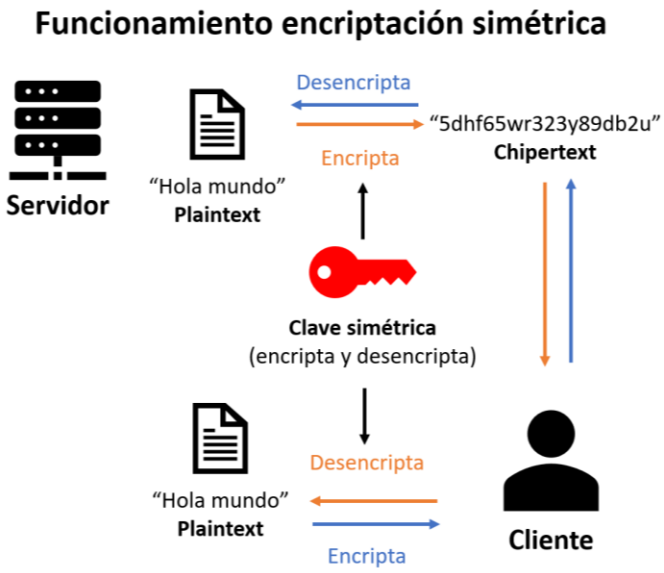


Ilustración 10: Funcionamiento de encriptación simétrica (elaboración propia)

La encriptación de estructura de clave pública, también llamada encriptación asimétrica, funciona de la siguiente manera se utiliza un algoritmo criptográfico que de forma aleatoria genera una pareja de claves, una pública y una clave privada. La clave pública, por un lado, se utiliza para encriptar la información pasando de ser un texto legible denominado *plaintext* a uno ilegible y cifrado denominado *chipertext*. La clave pública únicamente cuenta con la función de encriptar, por lo que se suele compartir con otros usuarios para que puedan enviar al destinatario la información cifrada. La clave privada se utiliza tanto para encriptar como para desencriptar la información. Esta clave debe permanecer privada y únicamente el dueño debe conocerla. Este tipo de encriptación por si sola únicamente permite la comunicación

unidireccional ya que únicamente el dueño de la clave privada puede desencriptar los mensajes cifrados. El conocimiento de la clave pública no permite de ninguna manera obtener la clave privada. Un escenario de encriptación es la comunicación con paginas SSL entre un banco (servidor) y sus clientes que buscan consultar información sobre el estado de sus cuentas.

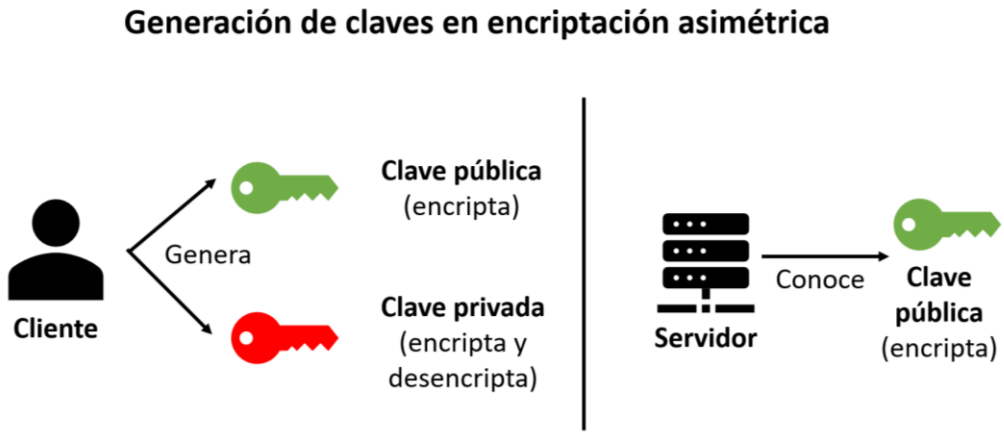


Ilustración 11: Generación de claves en encriptación antisimétrica (elaboración propia)

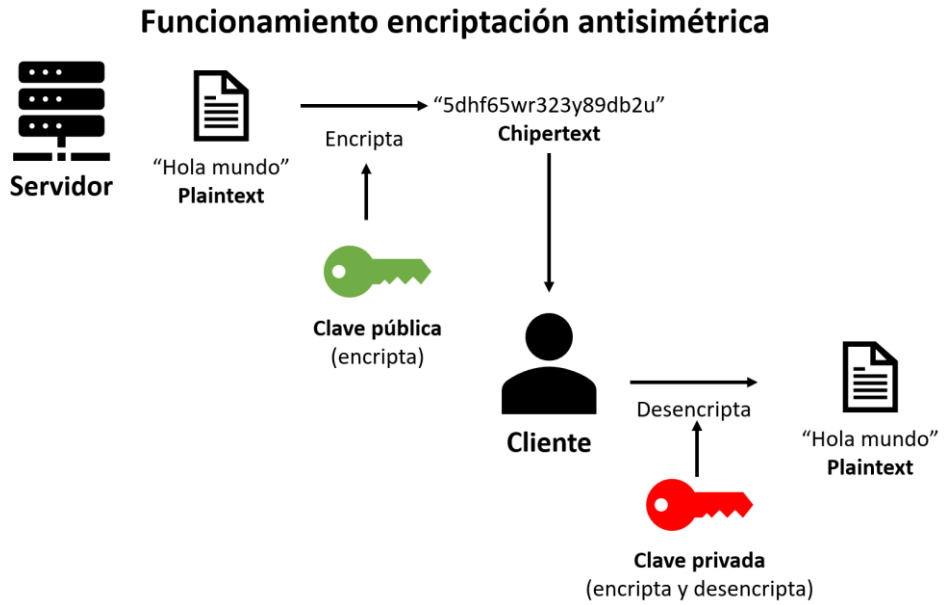


Ilustración 12: Funcionamiento de la encriptación antisimétrica (elaboración propia)

También es posible la combinación de la encriptación simétrica dentro de un contexto de encriptación asimétrico. La encriptación simétrica tiene muchas ventajas ya que permite que la comunicación y la lectura de la información pueda hacerse en ambas direcciones además de ser más eficiente y de consumir menos recursos. Su principal desventaja es que es complicado transmitir las claves simétricas entre los usuarios ya que los canales de transmisión suelen ser

inseguros. Para ello se utiliza la encriptación antisimétrica, el servidor utiliza la clave pública del cliente para encriptar la nueva clave generada para la encriptación simétrica. El cliente desencripta la clave simétrica con su clave privada antisimétrica, de esta forma tanto el servidor como el cliente conocen la nueva clave simétrica y pueden tener una comunicación bidireccional.

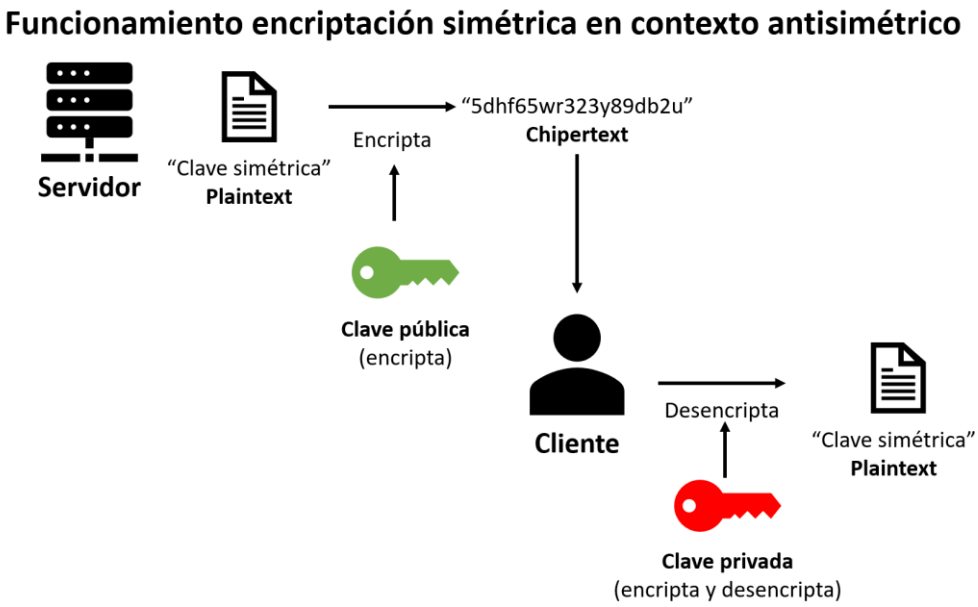


Ilustración 13: Funcionamiento de la encriptación simétrica en contexto antisimétrico (elaboración propia)

La encriptación de estructura pública permite la verificación de los usuarios que encriptan y desencriptan mensajeas, así como la validación de que la información transmitida no ha sido alterada en el proceso por agentes externos. Esto se realiza a través de firmas digitales, esta tecnología funciona gracias a un sistema certificador que valida unos certificados fiables e individuales que permiten crear las claves.

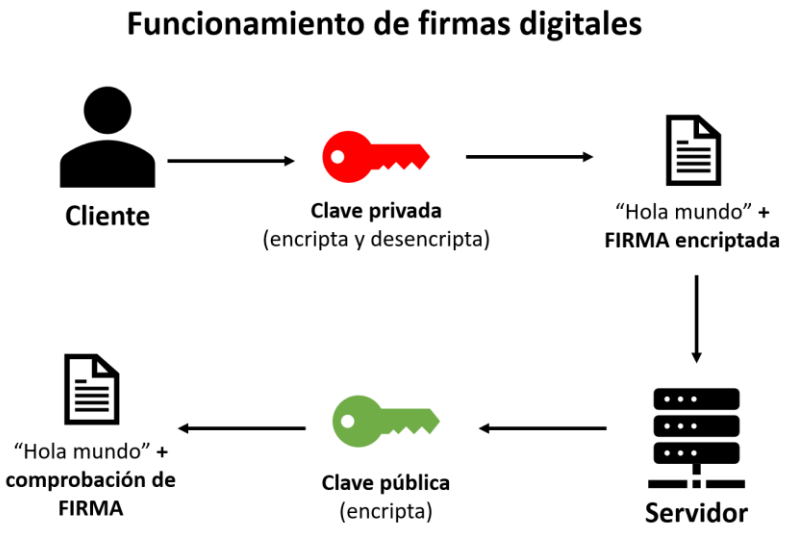


Ilustración 14: Funcionamiento de firmas digitales (elaboración propia)

La encriptación se utiliza para mantener los mensajes privados y confidenciales en sectores militares, financieros, sanitarios y otros sectores estratégicos.

Hay determinados tipos de encriptación que se consideran seguros ya que la cantidad de recursos necesarios para desencriptar un mensaje sin conocer la clave privada son demasiado elevados para que sea factible en la práctica. En caso de que se considere vulnerable se puede solucionar aumentando la longitud de las claves tanto públicas como privadas, aunque aumenta la capacidad de procesamiento necesaria para desencriptar y encriptar los *plaintext* además de hacer el proceso algo más lento.

Sin embargo, la encriptación tradicional tiene algunas desventajas. En concreto si se quieren realizar operaciones con los datos encriptados no puede realizarse de forma directa ya que es necesario desencriptarlos primero, operar con ellos y encriptar de nuevo los resultados, quedando tanto los datos como los resultados vulnerables a un ataque durante todo este proceso. Este problema se resuelve mediante la encriptación homomórfica.

El valor diferencial de la encriptación homomórfica respecto a la encriptación asimétrica tradicional es que permite la realización de operaciones matemáticas con los datos encriptados sin la necesidad de conocer las claves. No solo se realizan las operaciones con los datos encriptados, sino que los resultados obtenidos también están encriptados de tal forma que únicamente el usuario que conozca la clave de desencriptación podrá conocer el valor real, el *plaintext*, del resultado.

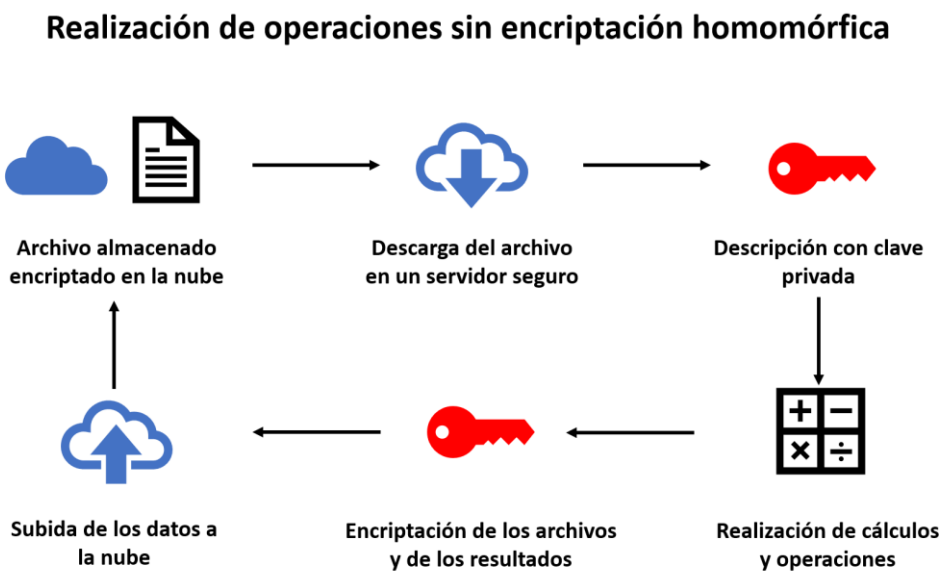


Ilustración 15: Relación de operaciones sin encriptación homomórfica (elaboración propia)

Esta ventaja cobra un mayor interés en un momento en el que la mayor parte de los datos se encuentran almacenados en la nube de forma encriptada, por lo que resulta interesante poder operar con ellos con seguridad dentro de ella, en lugar de tener que descargar los datos en un servidor seguro, donde poder desencriptarlos y operar con ellos para, a continuación, volver a subir los resultados obtenidos encriptados a la nube. Además, permite cambiar la forma de trabajar en entornos desconocidos en los que no se garantiza la protección de los datos o en el caso de que los datos no se quieran compartir, como ocurre en la mayor parte de los casos, pudiéndose cumplir con la protección de los datos en todo momento.

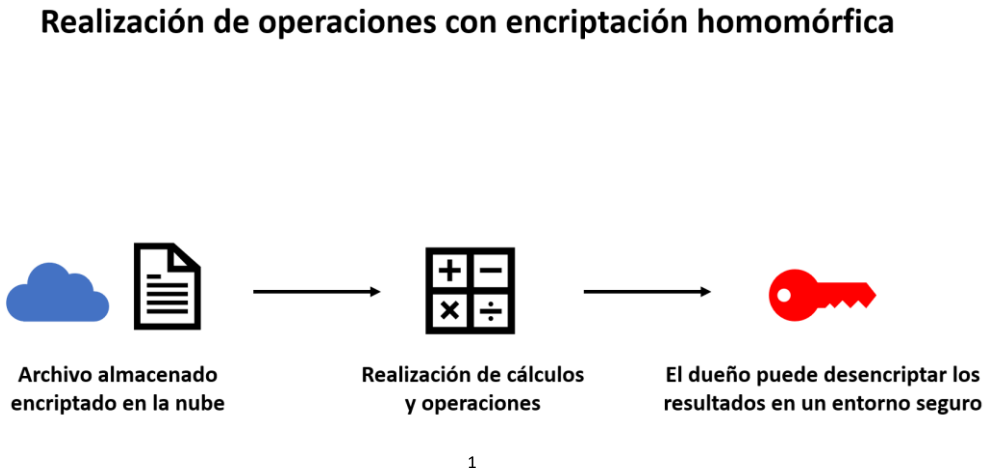


Ilustración 16: Realización de operaciones con encriptación homomórfica (elaboración propia)

La encriptación homomórfica es una tecnología que tiene diferentes esquemas o estructuras de funcionamiento como *partially homomorfic encryption*, the smomewhat homomorfic encryption, the leveled homomorphic encryption y finalmente, the fully homomorphic encryption.

La encriptación homomórfica completa (fully homomorphic encryption FHE) surgió como concepto en el año 1978 cuando fue descrita por Rivest Adleman, sin embargo, fue construida por primera vez en el año 2009 por el experto en *computer science* Craig Gentry. En los últimos años ha experimentado una cierta evolución pasando por diferentes etapas:

- Pre-FHE
- Primera generación de FHE
- Segunda generación de FHE
- Tercera generación de FHE
- Cuarta generación de FHE

En los últimos años ha experimentado una evolución especialmente en el ámbito de la normalización y estandarización de esta tecnología gracias a la asociación de "Homomorphic encryption standarization" de la que forman parte empresas como Micosoft, Samsung SDS, Intel, IBM y Google entre otras, también han participado gobiernos y organizaciones internacionales como las Naciones Unidas e instituciones académicas como Boston University, Brown, MIT o University of Columbia. Esta tecnología en la actualidad es de código abierto y

cuenta con varias librerías entre las que se encuentra HElib desarrollada por IBM y Microsoft SEAL desarrollada por Microsoft en los lenguajes de programación de C++ y C#, este último es el que se ha utilizado para la demostración del funcionamiento más adelante.

La HE tiene algunos beneficios como que permite la monetización de los datos, garantiza su privacidad y el cumplimiento de la normativa y facilita y promueve el uso de la nube. Entre los casos de uso se pueden encontrar la IA, ML y la búsqueda por coincidencia de datos en bases de datos extensas. Es decir, esta tecnología interesa a empresas pertenecientes a muchos sectores como el márketing, el análisis de riesgos, el sector financiero, los científicos de datos, las empresas que desarrollan la nube pública y analistas de seguridad.

2.4. Anonimización

La anonimización de los datos es el método a través del cual se elimina cierta información de carácter privado o personal de una base de datos con el objetivo de mantener la privacidad de los individuos que forman parte de ella. Sin embargo, no es un método que garantice la privacidad o la protección de los datos. A pesar de no ser un método de protección de los datos es necesario explicarlo para comprender por qué motivo no es un método suficiente para lograr este objetivo, además de que permite poner en valor las características de las demás tecnologías explicadas en este documento y de otras que puedan surgir en el futuro.

El motivo por el cual la anonimización no es una tecnología segura es porque en muchas ocasiones los individuos tienen una combinación única de los diferentes parámetros de la base de datos que permite por lo tanto identificarlos a pesar de haber eliminado los datos de carácter personal. En ocasiones es necesaria más de una fuente de información para lograr las identificaciones, pero una vez que se combinan es relativamente sencillo identificar a las personas individuales. Hay estudios que dicen que el 87% de los ciudadanos estadounidenses pueden ser identificados únicamente con su sexo, fecha de nacimiento y código postal (Kopp, Microsoft SmartNoise Differential Privacy Machine Learning Case Studies, 2021).

El problema del cruce de información para identificar individuos es que es posible que en el momento de la publicación de los datos sí que se pueda garantizar la privacidad, pero que al desconocer las informaciones futuras que se puedan publicar no se puede garantizar que al contar con información nueva se siga pudiendo garantizar la privacidad. Además, es necesario recordar que en la época de internet una vez se ha subido información a la web es imposible eliminarla.

Para que la anonimización sea una metodología de privacidad realmente anónima es necesario eliminar toda la información que permita identificar individuos. Hay conjuntos de datos que no permiten ese grado de abstracción, sin embargo, hay escenarios como por ejemplo el resultado de pruebas de fármacos para grandes poblaciones en las que solo se conoce la edad y el sexo de los pacientes en los que puede resultar interesante su uso.

2.5. Datos sintéticos

Esta tecnología surge para resolver el problema de la privacidad, en especial en datos personales (nombre, apellido, DNI, nacionalidad, dirección, fecha de nacimiento) y de la necesidad de utilizar los datos de forma individual y no como un conjunto de resultados obtenidos a partir de estos. Por ejemplo, como sustitutos de conjuntos de datos reales, para validar modelos matemáticos o más recientemente para el entrenamiento de modelos de *machine Learning.* Para entrenar modelos se pueden utilizar tanto datos reales, datos sintéticos como una combinación de ambos.

A pesar de ser la aplicación más reciente de los datos sintéticos, el entrenamiento de modelos de *machine Learning* ha resultado ser la aplicación más interesante de esta tecnología (Laskowski, 2018).

La definición de datos sintéticos es: “cualquier tipo de dato generado mediante simulación por ordenador”. La principal característica de esta tecnología es que los datos se crean mediante algoritmos. Estos datos creados de forma artificial mantienen las proporciones y los valores estadísticos de la base de datos original. De esta manera se obtienen resultados prácticamente iguales de utilizar las BBDD originales y una BBDD formada por completo o solo en parte por datos sintéticos.

Las dos industrias que más utilizan los datos sintéticos son aquellas en las que la naturaleza de los datos es especialmente privada o confidencial, como es el sector salud y el sector financiero. (Laskowski, 2018) . Aun así, los datos sintéticos están muy presentes en nuestra vida en otros muchos sectores, como por ejemplo los sintetizadores de sonido para crear sonidos de forma electrónica que simulan los sonidos reales de los diferentes instrumentos. Un símil de esta tecnología podrían ser el Photoshop, los efectos especiales de las películas, el diseño y dibujo a ordenador.

También se han utilizado los datos sintéticos para la mejora de los *bots* de conversación que funcionan con IA, como este de Microsoft (Reply Spur).

De forma más específica encontramos la creación de NFT para los que se diseñan unos parámetros variables que el ordenador combina de forma aleatoria a fin de conseguir todas las combinaciones posibles.

Esta tecnología se ha utilizado con éxito con datos de todo tipo. Sin embargo, el mayor problema aparece en el momento de crear datos sintéticos de imágenes, en especial de rostros humanos. Finalmente, un proyecto de Microsoft ha logrado crear una base de datos con 100.000 imágenes de caras sintéticas. Para ello se parte de una plantilla de rostro humano a la que se aplican parámetros aleatorios a distintas variables como la forma de la cara, el espesor, la forma y tamaño de los distintos rasgos faciales como los ojos, la boca, los pómulos, las orejas la frente y el pelo. Consiguiendo miles de combinaciones diferentes generadas por ordenador y que pueden ser utilizadas para entrenar modelos de machine Learning de reconocimiento facial. Una vez el modelo ha sido entrenado se utilizan imágenes reales para comprobar su correcto funcionamiento.

Los principales escenarios en los que resulta interesante utilizar datos sintéticos son:

- Cuando los datos reales tienen demasiado ruido y pueden entorpecer o dificultar la obtención de resultados.
- Cuando los datos reales son demasiado difíciles de etiquetar de forma manual para el entrenamiento de los modelos.
- Cuando en los datos reales no es posible encontrar las condiciones específicas o la calidad suficiente que se buscan para entrenar el modelo.
- Los datos sintéticos facilitan el control sobre los datos utilizados y el entrenamiento del modelo.
- Cuando los datos reales son difíciles de obtener o no existen datos suficientes.
- Cuando los datos reales son de especial protección o confidenciales y no pueden obtenerse sin poner en riesgo la información que contienen.

Los datos sintéticos pueden combinarse con otras tecnologías como la *differential privacy* para evitar uno de los problemas que presenta esta última. La desventaja de la DP que se ha mencionado anteriormente es que cada vez que se utilizan los datos estos tienen un mayor riesgo de ser identificables. Sin embargo, si al conjunto de los datos originales se le aplica un tratamiento para generar datos sintéticos combinado con el ruido estadístico introducido por la

DP, obtenemos una base de datos nueva, que no tiene ningún riesgo de exponer datos privados, aunque se utilicen muchas veces.

3. Creación de una demo

El objetivo de esta sección es la demostración mediante código del funcionamiento de la HE desde el punto de vista técnico, además de mostrar las peculiaridades que tiene a la hora de programar. También se ha hecho con el objetivo de demostrar la necesidad de un agente intermediario que se encargue de la comunicación entre las empresas, en este caso la empresa que se desarrolla en este proyecto, para garantizar la privacidad de los datos y de los modelos.

Para ello se ha programado en lenguaje C# en Visual Estudio utilizando la librería de Microsoft SEAL de código abierto que se encuentra disponible en GitHub. Hay disponibles ocho ejemplos que han sido estudiados con detenimiento a fin de comprender el funcionamiento de los diferentes tipos de HE y que han permitido escoger el que mejor se adaptaba a las necesidades específicas de este proyecto. El proceso de aprendizaje ha sido complejo por la escasa información disponible sobre esta librería, las limitaciones a la hora de programar o el funcionamiento de algunas funciones disponibles.

En los apartados siguientes se explica de forma detallada el código, su funcionamiento y los resultados que se obtienen.

3.1. Código

En este apartado se describe y explica el código desarrollado, para ello se explican únicamente las partes más relevantes y en el caso de haber partes repetitivas se explican solamente una vez a fin de facilitar la comprensión del código. En caso de querer ver el código completo se puede ver en el ANEXO 4 que puede encontrarse al final de este documento.

El proyecto tiene dos códigos con objetivos muy diferentes, el primero tiene como objetivo ilustrar el funcionamiento de la encriptación homomórfica, la forma en la que debe ser programado y comprender el alcance que tiene en términos de cálculo. Este primer código no busca reflejar la experiencia de un usuario real del código, sino que se comprenda cómo funciona.

El objetivo del segundo código tiene como objetivo ilustrar las funciones de los tres agentes que intervienen, las funciones y las relaciones que tendrían en una experiencia real, simplificando al máximo las operaciones que realiza y la encriptación homomórfica.

Código de funcionamiento de la HE

Este código consta de un *main program* en el que se encuentran todas las funciones y consta de 4 formularios, dos de ellos para el creador del modelo y dos de ellos para el cliente. En este código no se garantiza la privacidad de los datos al estar todas las claves y parámetros de encriptación almacenados en el *main*, pero el objetivo de este programa es que sea lo más ilustrativo posible desde el punto de vista de la HE. El factor privacidad se tiene en cuenta en el segundo código.

Las funciones para HE se han tomado de la librería pública de Microsoft SEAL.

- *Main program*: contiene las funciones y las llamadas a los formularios.
    - Función inicializar: en esta función se inicializa SEAL y se crean los contextos de encriptación, los parámetros de encriptación, las claves públicas y privadas, el encriptador, el evaluador y el desencriptador.

```
EncryptionParameters parms = new (SchemeType.CKKS);
            ulong polyModulusDegree = 16384;
            parms.PolyModulusDegree = polyModulusDegree;
            parms.CoeffModulus = CoeffModulus.Create(
            polyModulusDegree, new int[] { 60, 50, 50, 50, 50, 50,
60 });
            context = new SEALContext(parms);

            KeyGenerator keygen = new (context);
            SecretKey secretKey = keygen.SecretKey;
            keygen.CreatePublicKey(out  publicKey);
            keygen.CreateRelinKeys(out  relinKeys);

            encryptor = new Encryptor(context, publicKey);
            evaluator = new Evaluator(context);
            decryptor = new Decryptor(context, secretKey);
```

    - Función crear encoder: en esta función se llama a una función propia de Microsoft SEAL que crea el encoder en la modalidad CKKS de HE.

```
encoder = new CKKSEncoder(context);
```

    - Función crear Plaintext y econde x: en esta función se convierte el valor de la variable introducida por el paciente en un plaintext y a continuación se encripta almacenándolo en un string.

```
Plaintext xPlain = new Plaintext();
```

```
encoder.Encode(Convert.ToDouble(X), scale, xPlain);
Ciphertext xEncryptedTemporal=new Ciphertext();
encryptor.Encrypt(xPlain, xEncryptedTemporal);
xEncrypted[nvar] = xEncryptedTemporal;
```

- Función crear Plaintext y econde coeficientes:en esta fusión se convierten los parámetros del modelo introducidos por el cetro de investigación en un *plaintext* y se almacena en una matriz de *plaintext*, en la que cada fila se corresponde con los coeficientes de una variable. Estos coeficientes son los coeficientes de un polinomio que puede ser hasta de grado 8. Y puede tener tantas variables diferentes como quiera el creador del modelo.

```
for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                PlainCoefs[nvar,ncoef] = new Plaintext();
                encoder.Encode(Dvectorcoef[ncoef], scale,
PlainCoefs[nvar, ncoef]);

            }
```

- Función comprobar coeficientes: en esta función se comprueba que los coeficientes introducidos son iguales o distintos de 0, esto se tiene en cuenta para reducir las operaciones en el momento de calcular, obviando los términos cuyo coeficiente es 0.

```
for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                if (Dvectorcoef[ncoef] != 0)
                {
                    BCoeficientes[nvar, ncoef] = true;
                }
            }
```

- Función calcular con plaintex y chipertext: en esta función se realizan las operaciones para cada término del polinomio. Estos términos se almacenan en un string de chipertext que más tarde serán sumados. Estas operaciones se realizan con las funciones de la librería de SEAL como por ejemplo square, MultiplyPlain o MultiplyInplace. En este punto es necesario tener en cuenta las peculiaridades de la HE a la hora de programar. Primero, solo se pueden operear entre si números que estén al mismo nivel de operación. Esto da lugar a un árbol de operaciones como el que se muestra en la imagen.

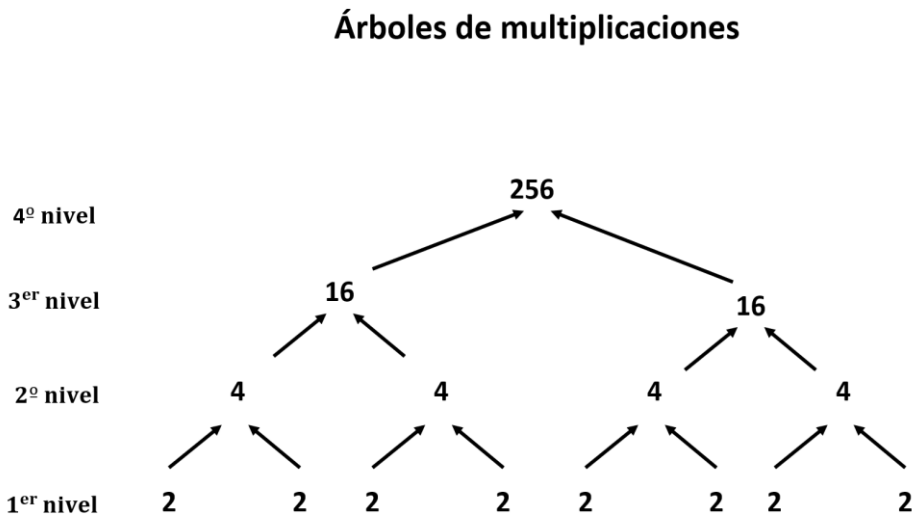


Ilustración 17: árbol de multiplicaciones (elaboración propia)

Además, es necesario reescalar (RelinearizeInplace) y relinearizar (RescaleToNextInplace) las operaciones para reducir el ruido producido por operar datos encriptados.

Primero se han realizado operaciones base que se van a utilizar para otras operaciones para optimizar el código.

```
//ENCRIPTAR 0
            Plaintext plaintext = new();
            Plaintext Plain0 = plaintext;
            encoder.Encode(0, scale, Plain0);
            Ciphertext Encripted0 = new ();
            encryptor.Encrypt(Plain0, Encripted0);

            //X^2
            Ciphertext xEncryptedauxX2 = new ();
            evaluator.Square(x1Encrypted, xEncryptedauxX2);
            evaluator.RelinearizeInplace(xEncryptedauxX2, relinKeys);
            evaluator.RescaleToNextInplace(xEncryptedauxX2);


            //X^4
            Ciphertext xEncryptedauxX4 = new ();
            evaluator.Square(xEncryptedauxX2, xEncryptedauxX4);
            evaluator.RelinearizeInplace(xEncryptedauxX4, relinKeys);
            evaluator.RescaleToNextInplace(xEncryptedauxX4);
```

```
            //X^8
            Ciphertext xEncryptedauxX8 = new ();
            evaluator.Square(xEncryptedauxX4, xEncryptedauxX8);

            evaluator.RelinearizeInplace(xEncryptedauxX8, relinKeys);
            evaluator.RelinearizeInplace(xEncryptedauxX8, relinKeys);
            evaluator.RescaleToNextInplace(xEncryptedauxX8);

            XCoefsEncriptados = new Ciphertext[numcoef];

            for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                XCoefsEncriptados[ncoef] = new();
                          }
```

Por reducir el texto repetitivo únicamente se va a presentar el código de dos factores, si se desea ver completo se puede acudir al anexo.

```
if (BCoeficientes[nvar, 7] == true)
            {
                //Multiplico X por el coeficiente de X^7
                evaluator.MultiplyPlain(x1Encrypted, PlainCoefs[nvar, 7],
XCoefsEncriptados[7]);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7], relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);


                //Multiplico Coef7*X por X^2
                evaluator.MultiplyInplace(XCoefsEncriptados[7], xEncryptedauxX2);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7], relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);


                //Multiplico Coef*x^3 *X^4
                evaluator.MultiplyInplace(XCoefsEncriptados[7], xEncryptedauxX4);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7], relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);
            }
                   if (BCoeficientes[nvar, 7] == false) { XCoefsEncriptados[7] =
      Encripted0; }

if (BCoeficientes[nvar,5] == true )
```

```
                {
                    //Multiplico X por ek coeficiente de X5
                    evaluator.MultiplyPlain(x1Encrypted, PlainCoefs[nvar,5],
XCoefsEncriptados[5]);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[5],
relinKeys);
                    evaluator.RescaleToNextInplace(XCoefsEncriptados[5]);

                    ParmsId lastParmsId = xEncryptedauxX4.ParmsId;
                    evaluator.ModSwitchToInplace(XCoefsEncriptados[5],
lastParmsId);

                    //Multiplico X^4*Coef5*X
                    evaluator.MultiplyInplace(XCoefsEncriptados[5],
xEncryptedauxX4);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[5],
relinKeys);
                    evaluator.RescaleToNextInplace(XCoefsEncriptados[5]);
                }
                      if (BCoeficientes[nvar, 5] == false ) { XCoefsEncriptados[5]
      = Encripted0; }
```

- Función sumar: en esta función se realizan las sumas encriptadas de los términos del polinomio.

```
ParmsId lastParmsId;
lastParmsId = XCoefsEncriptados[0].ParmsId;
// buscamos los parámetros del término más alto distinto de 0
 for (ncoef = 0; ncoef < numcoef; ncoef++)
{
  if (BCoeficientes[nvar, ncoef])
  {
      lastParmsId = XCoefsEncriptados[ncoef].ParmsId;
  }
}

//asignamos los parametros del termino mas alto a todos los terminos
for (ncoef = 0; ncoef < numcoef; ncoef++)
{
  evaluator.ModSwitchToInplace(XCoefsEncriptados[ncoef],lastParmsId);
  }
```

```
            lastParmsId = XCoefsEncriptados[7].ParmsId;
            evaluator.ModSwitchToInplace(XCoefsEncriptados[8], lastParmsId);
            XCoefsEncriptados[8].Scale = XCoefsEncriptados[7].Scale;


        //Estos Cipertextos auxiliares se usan porque la función de Microsoft SEAL solo admite como inputs cipertext sencillos y no acepta ni vectores ni matrices

        Ciphertext EncryptedResult = new Ciphertext();

        evaluator.Add(XCoefsEncriptados[8], XCoefsEncriptados[7], EncryptedResult);


        for (int i = 6; i >= 0; i--)
        {
            lastParmsId = XCoefsEncriptados[i].ParmsId;
            evaluator.ModSwitchToInplace(EncryptedResult, lastParmsId);
            EncryptedResult.Scale = XCoefsEncriptados[i].Scale;
            evaluator.Add(EncryptedResult, XCoefsEncriptados[i], EncryptedResult);
        }
```

- Función desencriptar resultado: en esta función se desencripta el resultado de la función suma utilizando las funciones de la librería de SEAL como Add.

```
Plaintext plainResult = new Plaintext();

decryptor.Decrypt(encryptedResult, plainResult);
result = new List<double>();
encoder.Decode(plainResult, result);
```

- Función sumar total: funciona igual que la función suma, pero para el resultado total de todas las variables.
- Función desencriptar resultado total: funciona igual que la función desencriptar resultado, pero para el resultado total de todas las variables.

NOTA: La numeración de los formularios no se corresponde con el orden en el que se ejecutan en el código, pero es el nombre que tienen en el código y para facilitar su comprensión se ha mantenido en este texto esa numeración. El orden en el que se ejecutan es: 3,2,1,4.

- Formulario 1: este formulario pide al paciente que introduzca el valor de las variables pedidas.

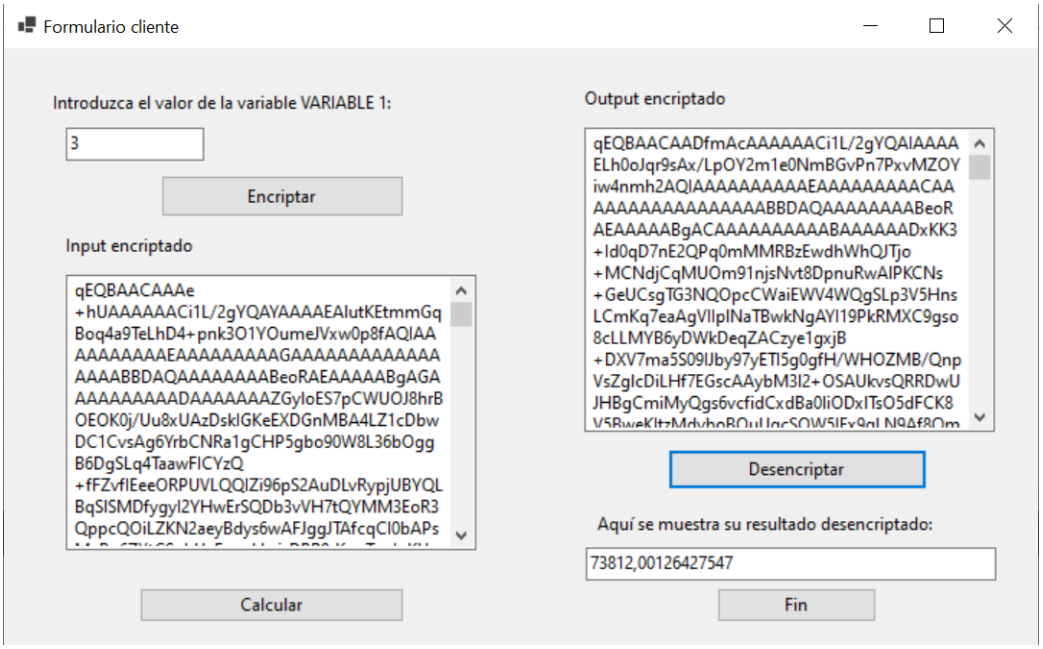


Ilustración 18:Formulario cliente (elaboración propia)

- Formulario 2: este formulario pide al creador del modelo que introduzca el nombre de las distintas variables del modelo y los coeficientes de cada variable.

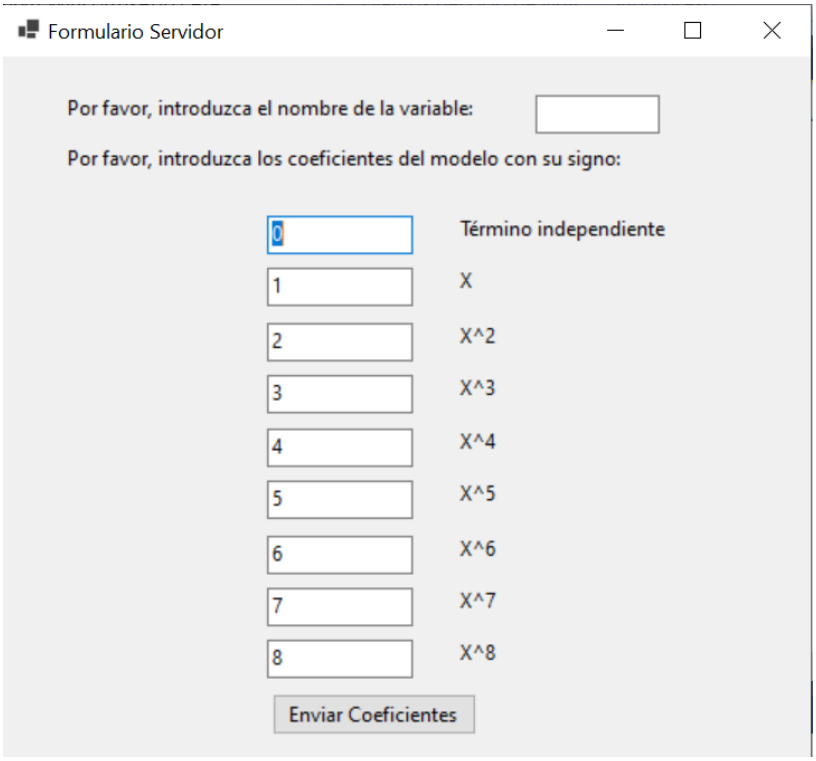


Ilustración 19: Solicitud de los parámetros (elaboración propia)

- Formulario 3: en este formulario se solicita al creador del modelo que introduzca el número de variables que quiere que tenga su modelo.

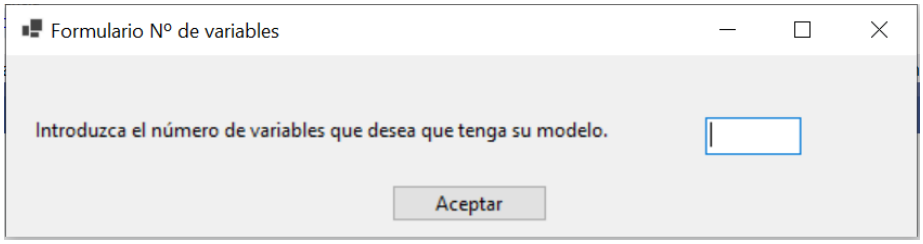


Ilustración 20: Solicitud del número de variables (elaboración propia)

- Formulario 4: en este formulario se presentan los resultados de las operaciones realizadas totales al paciente.

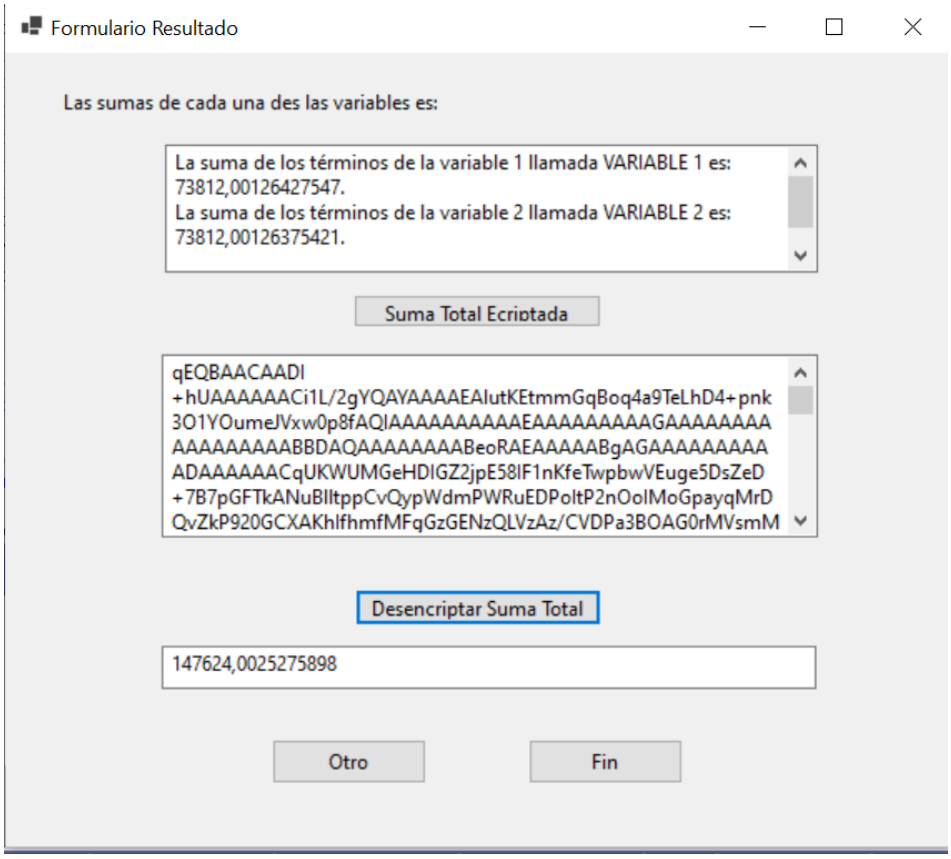


Ilustración 21: Resultados finales (elaboración propia)

Código de relaciones entre los agentes

Este código consta de tres códigos independientes que se relacionan entre sí, por lo tato se garantiza la privacidad y la protección de los datos. Este código consta de tres agentes diferenciados. Cada uno de esos programas independientes contiene sus propias funciones. Los detalles sobre el funcionamiento de la HE ya se han explicado en el código anterior por lo que no se van a repetir. La comunicación entre los agentes aparece reflejada en la siguiente imagen y se describen las funciones y el código de cada una de ellas con más detalle a continuación. Este código se ha subido a la nube de Microsoft, Azure para demostrar su implementación y que es posible ejecutarlo desde dos máquinas diferentes, como sería en la realidad la relación entre el paciente y el creador del modelo. De hecho, este modelo puede ser consultado en aquí. (https://pilartfg.azurewebsites.net/)

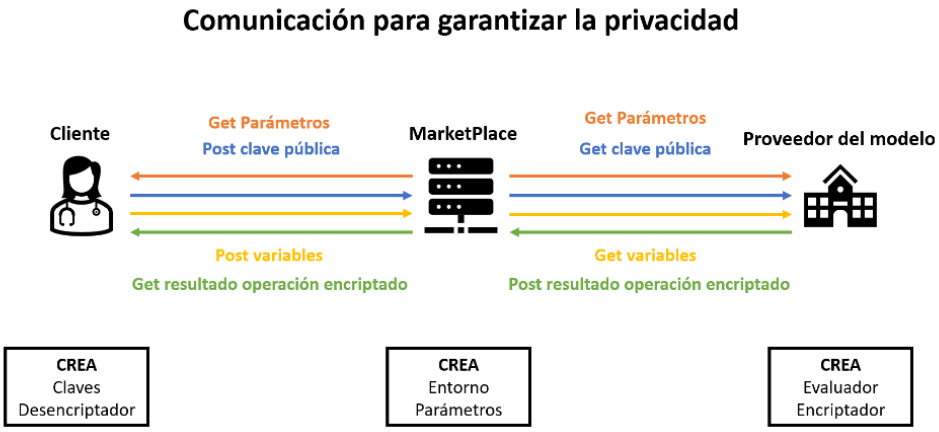


Ilustración 22: Comunicación para garantizar la privacidad (elaboración propia)

- Servidor: consta de 3 partes y representa al Marketplace
  - Values controller: gestionan los gets y los stets, es decir las llamadas del cliente y del creador de modelos.
  - Startup: crea una web a través de la cual se relacionan los agentes.
  - Program: en este programa se crean los parámetros y el contexto de SEAL.

- Creador de modelos: el proveedor del modelo queda a la espera de que un cliente haga uso del modelo y comprueba si algún cliente ha enviado los parámetros. No podrá hacer nada más hasta que no reciba los parámetros. Una vez recibe los parámetros se le pide al proveedor que introduzca el sumando que le corresponde, lo suma y envía el resultado de la suma encriptada. La imagen es del formulario que el proveedore ve en pantalla.

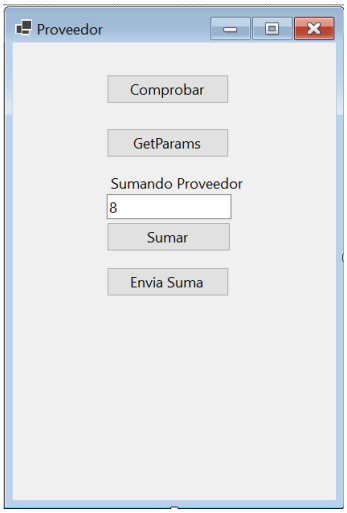


Ilustración 23: Formulario proveedor de modelos, código comunicación (elaboración propia)

- Función comprobar parámetros: comprueba si el cliente ha enviado los parámetros.

```
            using (var request = new
HttpRequestMessage(HttpMethod.Get, $"{BaseUri}/values/comprobar"))
            {
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                if (RM.Res == "true")
                {
                    bCheck = true;
                }
            }
```

- Función get params: Una vez se ha comprobado que el cliente ha enviado los parámetros se procede a conseguirlos, junto con la clave pública. Este código no se presenta ya que ha sido explicado en el código anterior, si se desea ver completo puede encontrarse en los anexos.
- Función sumar: Este código no se presenta ya que ha sido explicado en el código anterior, si se desea ver completo puede encontrarse en los anexos.
- Función enviar suma: Este código no se presenta ya que ha sido explicado en el código anterior, si se desea ver completo puede encontrarse en los anexos.

- Cliente: recibe los parámetros del servidor, crea las claves de encriptación, desencriptación y de encode. El cliente introduce el valor de una de las variables de entrada, llamada sumando y envía la clave pública y el sumando. Luego le cliente queda

a la espera del resultado final que debe enviar el creador del modelo. La imagen es del formulario que el cliente ve en pantalla.

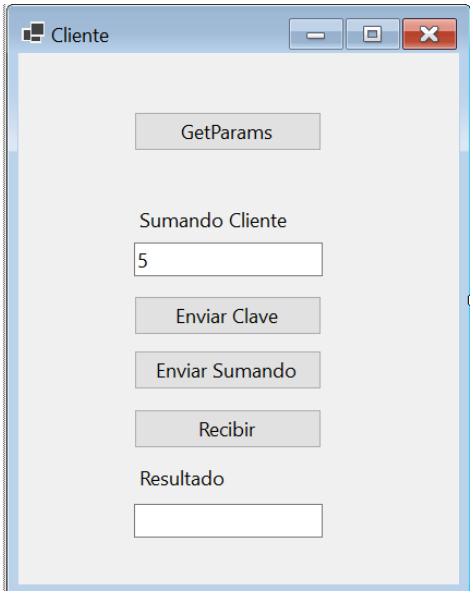


Ilustración 24: Formulario cliente, código comunicación (eleboración propia)

- Función get Params: solicita los parámetros al servidor.

```
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                _params = new EncryptionParameters();
                using (var ms = new MemoryStream(payload))
                {
                    _params.Load(ms);
                }

                _context = new SEALContext(_params);

                scale = Math.Pow(2.0, 40);
                _keyGenerator = new KeyGenerator(Program._context);
                _secretKey = Program._keyGenerator.SecretKey;
                _keyGenerator.CreatePublicKey(out
Program._publicKey);
                _encryptor = new Encryptor(_context, _publicKey);
                _encoder = new CKKSEncoder(_context);
```

```
                _decryptor = new Decryptor(_context, _secretKey);
```

- Función enviar sumando: solicita al cliente que introduzca la variable, la encripta y la envía.

```
var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                _params = new EncryptionParameters();
                using (var ms = new MemoryStream(payload))
                {
                    _params.Load(ms);
                }

                _context = new SEALContext(_params);

                scale = Math.Pow(2.0, 40);
                _keyGenerator = new KeyGenerator(Program._context);
                _secretKey = Program._keyGenerator.SecretKey;
                _keyGenerator.CreatePublicKey(out
Program._publicKey);
                _encryptor = new Encryptor(_context, _publicKey);
                _encoder = new CKKSEncoder(_context);
                _decryptor = new Decryptor(_context, _secretKey);
```

- Función enviar clave pública: envía al servidor la clave pública que permite encriptar.

```
string clave;
MemoryStream ms;
ms = new();
_publicKey.Save(ms);
clave = Convert.ToBase64String(ms.ToArray());

string sumandoRequestAsJsonStr = JsonConvert.SerializeObject(clave);
using (var request = new HttpRequestMessage(HttpMethod.Post,
$"{BaseUri}/values/clave"))
using (var content = new StringContent(sumandoRequestAsJsonStr,
Encoding.UTF8, "application/json"))
{
```

```
request.Content = content;
var response = await _client.SendAsync(request);
response.EnsureSuccessStatusCode();
            }
```

- Función recibir suma: recibe el resultado de la operación encriptado y lo desencripta usando la clave privada.

```
using (var request = new HttpRequestMessage(HttpMethod.Get,
$"{BaseUri}/values/Getsuma"))
            {
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();

                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                Ciphertext resultado;
                resultado = new Ciphertext();
                using (var mso = new MemoryStream(payload))
                {
                    resultado.Load(_context, mso);
                }

                Plaintext resultadoDesencriptado;
                resultadoDesencriptado = new();
                _decryptor.Decrypt(resultado, resultadoDesencriptado);
                List<double> dResultado = new List<double>(1);
                _encoder.Decode(resultadoDesencriptado, dResultado);

                _FormP.ImprimirResultado(dResultado[0].ToString());
```

Capítulo 3: Estudio económico-empresarial

1. Planteamiento del problema desde el punto de vista macro

Los datos han ido cobrando una mayor importancia y un mayor valor tanto para las empresas como para los usuarios. Además, la legislación al respecto será cada vez más restrictiva y limitará el uso que pueden hacer de los datos las empresas e instituciones. Por ejemplo, la GDPR (Reglamento General de Protección de Datos) dice sobre cómo deberían ser la directiva sobre datos personales:

"*tratados de manera lícita, leal y transparente en relación con el interesado («licitud, lealtad y transparencia»)"; "recogidos con fines determinados, explícitos y legítimos, y no serán tratados ulteriormente de manera incompatible con dichos fines; de acuerdo con el artículo 89, apartado 1, el tratamiento ulterior de los datos personales con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos no se considerará incompatible con los fines iniciales («limitación de la finalidad»)"* (Unión Europea, 2016).

Para las empresas los datos y el estudio de estos tiene un gran valor. En primer lugar, porque permite un conocimiento muy preciso del mercado y del público al que se dirigen y conocer bien los datos y obtener las conclusiones adecuadas de ellos puede suponer la diferencia entre una buena campaña de ventas y el crecimiento de la empresa o puede suponer grandes pérdidas tanto económicas como de clientes que luego puede ser complicado de recuperar.

Cada vez más hay empresas cuyo negocio principal se basa en la explotación económica de los datos de los usuarios como pueden ser Meta, Netflix o Google. Por lo que suponen su bien más preciado. Por ejemplo, el correo electrónico de Google, Gmail gana dinero gracias a sus usuarios a través de publicidad personalizada. A través de AdSense se subastan los anuncios a las empresas que pagan a Google por poner sus anuncios en internet para que se muestren a los usuarios con el perfil objetivo. Este perfil se conoce gracias a datos como la localización de los usuarios.

Por otro lado, los usuarios son cada vez más conscientes de los datos personales que ceden y de cómo las empresas se benefician de esta información y exigen una mayor protección y privacidad a los gobiernos que mediante legislaciones más restrictivas limitan el uso de los datos y protegen a los usuarios frente a las grandes empresas.

Uno de los grandes problemas con los datos son las colaboraciones entre empresas para campañas concretas en las que una de las dos empresas o incluso ambas deben proporcionar datos tanto de la empresa como de los clientes para poder llevar la campaña a cabo. Un ejemplo podría ser las compañías de publicidad y marketing o las consultoras, que para desempeñar su trabajo deben conocer datos específicos que tal vez la empresa preferiría mantener privada.

Sin embargo, algunos de los usos más prometedores de la inteligencia artificial aparecen cuando se combinan fuentes de datos complementarias que aportan informaciones de terceras partes. Por ejemplo, una cadena hotelera podría beneficiarse de datos de reservas de aviones de una compañía aérea para anticipar volúmenes de ocupación y personalizar ofertas dependiendo del origen de los viajeros. Existen tantos ejemplos semejantes a este como industrias o empresas hay, lo que va a generar una demanda innegable que requiere de soluciones creativas que permitan el acceso a esos datos de una forma segura.

Por otro lado, hay un número creciente de escenarios en los que el conocimiento y los datos estás repartidos por lo que la forma más eficiente de extraer conclusiones es la búsqueda de soluciones de co-innovación. Un ejemplo notable es el acuerdo entre la farmacéutica UCB y Microsoft para el desarrollo de nuevas medicinas con el uso de IA: "*UCB and Microsoft will explore how to combine diverse research data sets with four strategic objectives in mind"* (UCB, 2021).

Este proyecto pretende proponer una solución a los problemas mencionados permitiendo que las empresas puedan rentabilizar sus datos sin comprometer la privacidad ni la seguridad de estos mediante la creación de modelos matemáticos que puedan beneficiar a otras empresas o instituciones. En especial aquellas de menor tamaño y con menor capacidad de desarrollo de estos modelos. Ya sea porque el precio del desarrollo de estos modelos no se justifica con el uso que se va a hacer o porque no se disponen de suficientes datos para obtener un modelo aceptable. Además, permitirá a las entidades desarrolladoras de modelos conservar la propiedad intelectual sobre ellos y a las empresas que hacen uso de los modelos mantener la confidencialidad sobre los datos, así como los resultados obtenidos tras pasar los datos por el modelo.

2. Servicios disponibles en el mercado

En este apartado se describen con detalle otras empresas o proyectos que existes o que existieron y fracasaron en el pasado. Esto persigue dos objetivos, el primero conocer y validar la

necesidad del mercado de una empresa o institución que resuelva los problemas antes mencionados. Y por otro lado persigue el objetivo de analizar los motivos de su éxito o fracaso a fin de poder aplicarlo a este proyecto para evitar cometer los mismos errores y al mismo tiempo utilizar de inspiración aquellas cosas que sí que se han hecho correctamente y que pueden usarse como ejemplo a seguir.

Las empresas o proyectos a analizar son: el *Marketplace* de Microsoft y el *Marketplace* de Amazon.

2.1. Marketplace de Microsoft

Hace unos años Microsoft lanzón un Marketplace en el que todo el que quería podía compartir y vender sus datos, sin embargo, terminó cerrando. Tras hablar con un experto sobre el tema, Todo Singleton, empleado de Microsoft se llagaron a los motivos del fracaso.

En primer lugar, el momento en el que Microsoft sacó este proyecto no fue el adecuado. Microsoft sacó un Marketplace de datos en un momento en el que el uso de los datos todavía no estaba claro y había pocas empresas que los supieran aprovechar,

En segundo lugar, las tecnologías como la inteligencia artificial y machine learning no estaban maduras y al igual que los datos había pocas empresas que los utilizaran.

En tercer lugar, el Marketplace era un lugar descentralizado en el que todo el mundo podía subir contenido que no estaba supervisado ni controlado por nadie, por lo que era muy difícil encontrar el contenido que se ajustara a las necesidades de los clientes. También era difícil garantizar la calidad de los datos y que la estructura o la cantidad de datos fuera suficiente.

Otro aprendizaje de este proyecto fallido es la necesidad de encontrar economías de escala en este tipo de proyectos. La utilidad de un portal de estas características se incrementa de forma exponencial cuantos más datos se encuentran y más usuarios participan. Por eso es importante por un lado invertir en incorporar participantes, así como facilitar el acceso y la prueba de datos de forma gratuita. En el siguiente apartado se explica cómo Amazon ha tomado nota de esto último y ha lanzado un portal mejorando la propuesta de Microsoft.

2.2. Marketplace de Amazon

Este Marketplace a diferencia del de Microsoft sigue existiendo y recibe el nombre de AWS Data Exchange (AWS). Las primeras conclusiones que pueden extraerse al navegar por la web es que debe ser muy similar a lo que debió ser en su momento el Marketplace de Microsoft descrito en el apartado anterior. La diferencia entre los dos por tanto es probablemente los tiempos que manejó una y otra empresa, Microsoft debió ser pionera y ofreció un servicio en un mercado que no estaba suficientemente maduro o preparado y para el que no resolvía ninguna necesidad. Sin embargo, Amazon ha debido hacer una lectura más adecuada de los tiempos y de la situación del mercado.

La existencia de este servicio valida muchas de las hipótesis planteadas al inicio de este documento. Se entiende que si una empresa como Amazon ha desarrollado un intercambiador de datos es porque satisface una necesidad real de las empresas, porque el momento es el adecuado y porque hay maneras en las que ellos pueden aportar valor y tratar de resolver este problema.

La principal ventaja de AWS Data Exchange es que Amazon ofrece los servicios descritos a continuación en muchos casos de forma gratuita los primeros doce meses, lo que lleva a pensar que Amazon apenas recibe ingresos por este servicio prestado y que busca que los clientes se fidelicen y estén satisfechos con la plataforma antes de empezar a cobrarles.

El AWS Data Exchange tiene dos líneas de negocio, la primera es una librería de bases de datos y la segunda es un acceso a API´s.

La librería de bases de datos está compuesta en su mayoría de bases de datos provenientes de fuentes públicas y datos de terceros, aunque también cuenta con algunas fuentes privadas. En total AWS Data Exchange cuenta con más de 3000 fuentes de datos que pueden descargarse de forma total o parcial mediante un periodo de prueba. El valor añadido que ofrece Amazon es la recopilación en un único lugar de estas bases de datos y el buscador que permite filtrar y clasificar la información a fin de facilitar al usuario la búsqueda de una base de datos que encaje con sus necesidades. La ventaja de que sea un servicio gratuito es que los usuarios pueden probar diferentes bases de datos. Además, cuenta con una interfaz muy sencilla de utilizar. Dentro de las categorías disponibles se encuentran el sector automovilístico, la sostenibilidad y medio ambiente, el sector público, la industria de los videojuegos o las telecomunicaciones.

El acceso a API´s (application program interface), es un interfaz entre aplicaciones permite al usuario recoger información específica. Esta línea de negocio realiza las búsquedas en las bases

de datos mencionadas antes, pero en lugar de descargar las bases de datos completas permite al usuario acceder al dato específico que está buscando.

Todo esto es posible gracias a la supremacía que tiene Amazon como líder del cloud computing y que facilita el desarrollo de esta plataforma y le permite a la vez potenciar su uso entre sus clientes.

Sin embargo, el AWS Data Exchange no satisface algunas de las necesidades que en este proyecto se tratan. Por ejemplo, la necesidad de un servicio personalizado y adaptado al cliente en el que se le busquen bases de datos que cumplan con las necesidades específicas. Otra propuesta de este trabajo fin de grado es que las bases de datos no provengan únicamente de fuentes públicas, sino que los datos privados de las empresas pueden ser explotables y utilizados. Si bien es cierto, sería necesario introducir otra de las propuestas de este proyecto que es la utilización de tecnologías que garanticen o que aumenten la privacidad y la protección de los datos como la encriptación homomórfica, la differential privacy o los datos sintéticos.

La desventaja de la línea de negocio de las API´s de AWS Data Exchange es que únicamente permite consultar la información de las bases de datos, pero no permite utilizarlas como entrada de un modelo, por ejemplo. Algo que se propone en cierta forma en este trabajo fin de grado.

Respecto a la línea de negocio de los modelos que se ha propuesto en este trabajo fin de grado no hay nada ni siquiera similar dentro de la propuesta de Amazon, por lo que sería una gran innovación.

2.3. Análisis Benchmark

En este apartado se realizará un análisis Benchamark en el que se pondrá de forma esquemática una comparativa entre los tres servicios, el Marketplace de Microsoft, AWS Data Exchange y el proyecto de este trabajo fin de grado a fin de poder comparar las ventas y las desventajas de unos y de otros.

Análisis Benchmark			
	Microsoft	AWS Data Exchange	Propuesta TFG
Tiempo correcto	NO	SÍ	SÍ
Servicio personalizado	NO	NO	SÍ

Venta de BBDD públicas	SÍ	SÍ	SÍ
Venta de BBDD privadas	NO	NO	SÍ
Venta de modelos de predicción	NO	NO	SÍ
Tecnologías de privatización	NO	NO	SÍ
Grandes economías de escala	SÍ	SÍ	Mediante acción comercial
Líderes Cloud	SÍ	SÍ	NO
Acceso a API´s	NO	SÍ	Se podría simplificar el acceso a modelos para ofrecer esta característica

Tabla 1: Análisis Benchmark (elaboración propia)

3. Legislación sobre protección de datos

Una de las mayores limitaciones a la hora de trabajar con los datos en la legislación. Es necesario tener en cuenta que la tecnología siempre va un paso por delante de la ley por lo que en ocasiones hay tecnologías que todavía no se encuentran contempladas.

Este es un proyecto de tecnología y no se tienen los conocimientos legales suficientes para analizar esta sección en profundidad, aunque sin duda sería una investigación muy relevante para el trabajo en su conjunto ya que una empresa debe funcionar dentro de un contexto legal. Aun así, a continuación, hay unas breves explicaciones que permiten comprender de forma superficial la situación legal actual.

Unión Europea y España

La legislación europea sobre protección de datos viene regulada en la GDPR que es el Reglamento General de Protección de Datos. En ella se busca unificar el criterio y las legislaciones de los estados que componen la Unión Europea a fin de reforzar la protección sobre

los ciudadanos. También regula la exportación de los datos fuera de la Unión. Las empresas que no cumplan con el reglamento se verán sometidas a grades sanciones económicas. Este reglamento es ejecutivo desde el año 2018 y la previsión es que con el tiempo se vuelva más restrictivo.

El objetivo de la ley es doble, por un lado, pretende devolver progresivamente el control de sus datos personales a los ciudadanos europeos y por otro lado busca facilitar el cumplimiento da la ley al armonizar la legislación de todos los países. Esta normativa aplica a todas aquellas empresas que procesen datos de ciudadanos de la UE.

Los datos personales según el GDPR incluyen las direcciones IP, los identificadores en internet, datos del nivel económico, cultural o relacionados con la salud. Y gracias a la ley la seguridad y protección de estos datos pasa a ser prioridad de las empresas, por lo que tendrán un protocolo de seguimiento de los datos desde su obtención, su uso y su destrucción.

A fin de garantizar la protección de estos datos las empresas deberán estudiar las diferentes alternativas existentes entre las que se encuentran algunas de las tecnologías propuestas al inicio de este documento como son la encriptación, la anonimización o la DP entre otras.

En España antes de que entrara en vigor la GDPR existía la LOPD, Ley Orgánica de Protección de Datos de 1999. Además, la Constitución española considera en su artículo 18.4 que la protección de los datos personales de las personas físicas es un derecho fundamental: "*la ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos*" por lo que España siempre ha sido una nación pionera en este aspecto (Jefatura del Estado, 2018).

Estados Unidos

Debido a que la legislación en materia de protección de datos de los Estados Unidos es menos restrictiva que la de la Unión Europea y la zona inicial de operación de la empresa sería la UE se considera que si el proyecto cumple para la legislación europea cumple para la estadounidense.

Implicaciones para este proyecto

Dado que este proyecto comenzaría en la UE deberá como todas las demás empresas cumplir con la legislación vigente en materia de protección de datos. Dado el desconocimiento de esta autora en derecho se contrataría a un experto en leyes. De todas maneras, para la realización de este trabajo se ha consultado con dos abogados diferentes a los que se les ha explicado con

detalle el proyecto y no han visto que pudieran existir problemas siempre y cuando se pida a los usuarios que firmen el consentimiento del uso de los datos de carácter personal.

4. Planteamiento de soluciones

Tras el estudio de las tecnologías disponibles, los problemas con los datos, la evolución de la mentalidad de las empresas e instituciones ante el análisis de datos y la legislación vigente en Europa y Estados Unidos surge la segunda innovación de este trabajo de fin de grado que es la creación de una empresa que solucione estos problemas respondiendo a las necesidades del mercado. Para ello se ha elaborado un Marketplace, el cual se explica con detalle en los siguientes apartados. Entre ellos se incluye una descripción del funcionamiento y de los servicios prestados, el modelo de negocio a través de un business canvas model, una cuenta de pérdidas y ganancias de los primeros cinco años expuesta en Excel, un análisis DAFO y un análisis PEST.

4.1. Marketplace de datos encriptados y modelos

En esta sección es necesario aclarar que esta empresa no pretende ser un Marketplace tradicional en el que todo el mundo pueda subir sus ofertas y todo el mundo pueda comparar de manera indiscriminada y libre. No se persigue este objetivo por la sencilla razón de que tal y como se ha analizado en el análisis dónde se han estudiado otras empresas que en el pasado o bien en el presente han tratado de resolver el mismo problema, se ha llegado a la conclusión de que este tipo de modelos de negocio terminan fracasando debido a la excesiva libertad que dificulta que los clientes encuentren datos y modelos que se ajusten a sus necesidades reales. Por lo que el modelo de negocio está basado en el de una empresa tradicional que cuenta con clientes a los que tiene que conectar entre sí a fin de satisfacer las necesidades individuales y específicas de cada uno de ellos dándoles asistencia en el proceso.

4.2. Servicios

En esta sección se describe con detalle la propuesta de valor del modelo de negocio y los servicios prestados. La empresa tiene principalmente dos líneas de negocio que se complementan entre sí, ya que el objetivo es solucionar el problema de los datos en todas las etapas de su uso. La primera, el uso de los datos en el entrenamiento de modelos y en segundo

lugar el tratamiento de los datos privados que deben ser introducidos en modelos para la obtención de resultados.

4.2.1. Uso de modelos con HE

La primera línea de negocio es la puesta en contacto de empresas que buscan dar salida a sus modelos con empresas que no cuentan con los recursos suficientes para crearse sus propios modelos y que esperan encontrarlos en otras empresas.

Este modelo de negocio es especialmente interesante en sectores dónde las necesidades de todos los compradores son las mismas y dónde hay pocos oferentes. Este sector es el de la medicina. Existen relativamente pocos centros de investigación, en parte porque las investigaciones son muy caras y hay escasez de recursos para ello, además de una escasez de datos con los que trabajar. Además, este tipo de centros cuentan con una gran confianza en el sector y tienen buena reputación. Por otro lado, hay muchos hospitales de todo el mundo que podrían beneficiarse de este conocimiento y de estas investigaciones. Sin embargo, existe el problema de introducir datos de pacientes en modelos de predicción o de ayuda al diagnóstico puede comprometer la privacidad de los pacientes o que a pesar de contribuir a la curación y tratamiento del paciente este puede mostrarse reticente a que sus datos se utilicen o compartan.

El funcionamiento de este modelo de negocio sería el siguiente: nuestra empresa se pone en contacto con creadores de modelos (universidades, centros de investigación, empresas especializadas, farmacéuticas, organismos públicos) mediante visitadores del sector, ferias especializadas y conferencias. Simultáneamente se contacta con empresas o instituciones que pudieran estar interesadas en hacer uso de ciertos modelos específicos o de modelos que sean muy similares a los disponibles. De forma personalizada se trata de vender los modelos que ofrecen los vendedores a los compradores de tal manera que estos se adapten a sus necesidades específicas. Además, se ofrece como servicio el facilitar toda la tecnología relacionada con la encriptación homomórfica explicada con detalle al inicio de este documento y que permite garantizar la privacidad y seguridad tanto del vendedor como del comprador. En la imagen a continuación se muestra lo explicado.

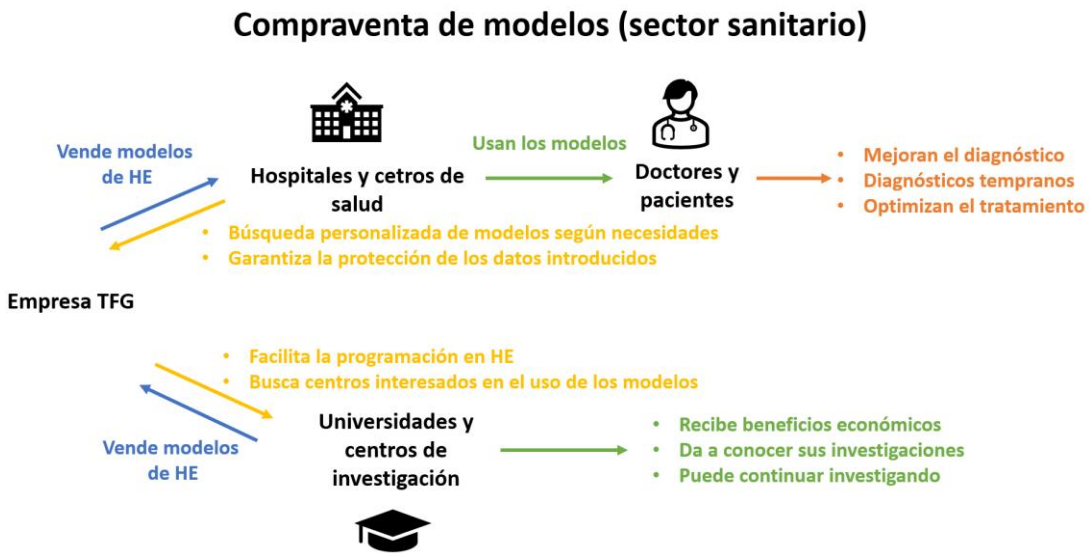


Ilustración 25: Compraventa de modelos (elaboración propia)

Los aspectos técnicos de esta parte se han abordado en el capítulo de descripción de las tecnologías y puede verse un ejemplo concreto de código descrito con detalle en ese capítulo o completo en los anexos.

4.2.2.Compraventa de datos

La segunda línea de negocio es la puesta en contacto de empresas que quieren comercializar con sus bases de datos con otras empresas que buscan bases de datos externas para diferentes fines. Esta compraventa se haría vendiendo no los datos reales, sino vendiendo bases de datos alteradas por las diferentes tecnologías explicadas al principio de este documento a fin de mantener la privacidad y confidencialidad de los datos. Estas tecnologías son: la *differential privacy* y los datos sintéticos principalmente.

Tal y como se ha explicado previamente el objetivo de esta línea de negocio no es crear un lugar dónde todo el mundo pueda comprar y vender los datos poniéndose en contacto entre sí. Sino que el objetivo es crear un servicio personalizado de compraventa en el que se atiendan a las necesidades individuales tanto de compra como de venta de los diferentes clientes de forma personalizada.

Debido a los elevados precios de cierto tipo de datos este modelo de negocio puede resultar inaccesible para las pymes. Por lo que la intención sería trabajar con empresas de sectores específicos que ofrecen servicios que suelen estar subcontratados por las pequeñas y medianas

empresas y que pueden por tanto distribuir los costes de compra de estas bases de datos entre diferentes clientes. Un ejemplo de estas empresas pueden ser las agencias de publicidad y de márketing que suelen estar enfocadas a sectores específicos y que es un servicio que muchas empresas tienden a externalizar. Otro ejemplo puede ser las empresas de recursos humanos que también es un servicio que se subcontrata. Por último, también podría ser interesante para pequeñas y grandes consultoras que también se subcontratan para situaciones concretas.

Al ser una transacción personalizada se busca que la empresa que vende los datos sea de confianza y que los datos que se venden sean de calidad y que realmente satisfagan las necesidades del comprador.

El funcionamiento de este modelo de negocio sería el siguiente: una empresa contacta con nosotros y nos cuenta las necesidades que tiene, el tipo de base de datos que necesita, la cantidad de datos, la estructura de los datos, el uso que va a hacer de ellos… Y nuestra empresa se encargaría de buscar entre los clientes que tenemos empresas que estén dispuestas a vender datos (no los originales) de esas características. Una vez se ha decidido quién va a ser el proveedor de datos se fija el tratamiento que van a tener los datos brutos para su posterior comercialización y se le facilita al vendedor el acceso a las tecnologías antes mencionadas. Posteriormente se envían los datos tratados al comprador. En la imagen a continuación se muestra lo explicado.

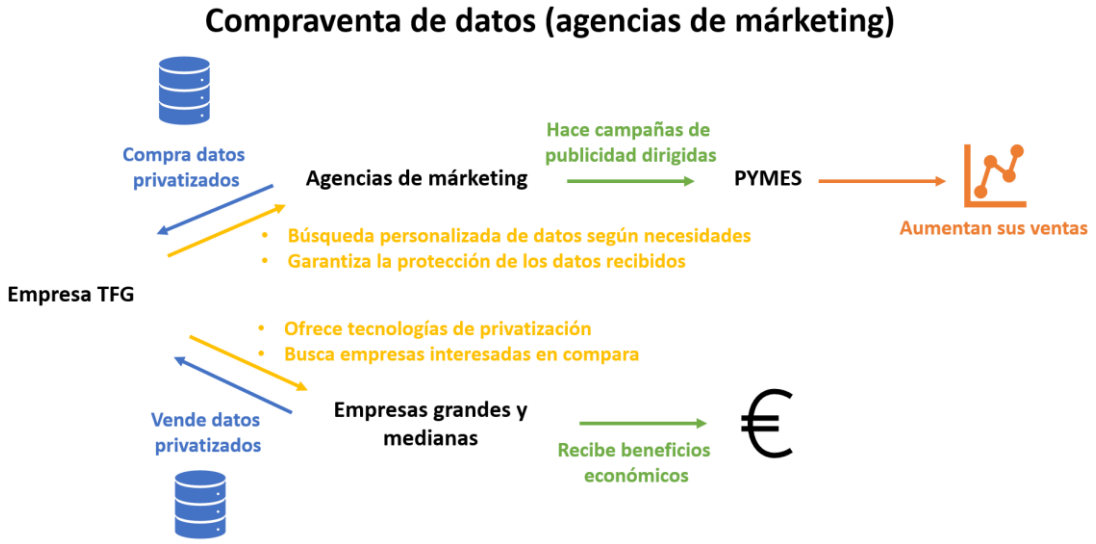


Ilustración 26: Compra venta de datos (elaboración propia)

4.2.3.Matching de datos y modelos

Una parte vital de la empresa tal y como se ha explicado en los párrafos anteriores y en el apartado en el que se estudian otros modelos de negocio similares, es el trato personalizado a los vendedores y compradores. Esto es el gran valor añadido de este proyecto y que facilita que se terminen llevado a cabo esos intercambios de datos y de modelos. Evitando que se convierta en un espacio en el que se venden muchos datos que realmente no se ajustan a las necesidades de nadie o que no cumplen con los estándares de calidad necesarios. Este valor añadido también es una barrera de entrada para posibles competidores que pretendan hacer un modelo de negocio similar.

5. Modelo de negocio y viabilidad económica

En esta sección se desarrolla en profundidad el modelo de negocio del proyecto contando con secciones como el canvas model, una cuenta de pérdidas y ganancias en Excel, un análisis DAFO y un análisis PEST.

5.1. Canvas Model

Para el estudio del modelo de negocio y del funcionamiento del Marketplace se ha utilizado un business model Canvas que es una herramienta que forma parte de la metodología de Lean *Start-Up* y que permite de forma sencilla y completa comprender las distintas dimensiones del negocio, así como su funcionamiento. Se ha escogido esta metodología por las ventajas que ofrece ya que facilita la comprensión y favorece el análisis de todo el negocio en su conjunto. La estructura tradicional suele ser en forma de tabla, pero para facilitar la lectura de este documento se ha dividido en 9 partes.

1. Segmento de clientes

En esta parte se busca satisfacer necesidades existentes en el mercado o crear nuevas necesidades a un consumidor específico. Para ello es necesario comprender bien al cliente para poder ofrecerles lo que necesitan creando un perfil específico.

Dentro del *Marketplace* hay dos líneas de negocio claramente diferenciadas pero complementarias en cierta forma, la primera es el alquiler de modelos y la segunda es la venta de datos. Para cada línea de negocio hay un vendedor y un comprador.

Alquiler de modelos

Para el alquiler de modelos el perfil del cliente que suministra los modelos es una empresa grande con recursos suficientes tanto por disponibilidad de datos para entrenar el modelo como por disponibilidad de recursos y tecnología para poder entrenar esos modelos. Estas empresas están interesadas en obtener una rentabilidad económica de los modelos que ya tienen. También podría ser una buena forma de rentabilizar los datos que no puede vender de manera indirecta, mediante la creación de modelos de Machine Learning que se entrenen con esos datos y poniendo en alquiler ese modelo. La ventaja es que al ser un alquiler de modelos las empresas no ceden su propiedad intelectual a otras empresas.

El perfil del cliente que hace uso de los modelos es una pequeña o mediana empresa que por su tamaño no tiene los datos suficientes para poder entrenar un modelo propio o no dispone de los recursos técnicos para poder crear ese modelo. Estas empresas, sin embargo, podrían beneficiarse de modelos realizados por empresas e instituciones de renombre.

El perfil cliente también puede ser organismos públicos o empresas de países en vías de desarrollo que se benefician de los modelos de las grandes empresas.

Para este modelo de negocio se ha considerado que el sector de la medicina puede ser un buen *earlyadopter*. Este escenario se explicará más adelante.

Compraventa de datos

La compraventa de datos puede ser bidireccional y pueden ocurrir diferentes escenarios que se van a describir a continuación.

Los vendedores de datos de nuevo podrían ser las grandes empresas que buscan obtener una rentabilidad económica de sus bases de datos sin comprometer la privacidad de los datos gracias a las tecnologías mencionadas al principio de este documento.

También en un momento dado podría ser interesante que las pequeñas empresas vendan sus datos, sin embargo, es difícil en este caso que se vendan a un buen precio al ser una cantidad pequeña de datos y de zonas geográficas y sectores muy específicos.

Los compradores de datos pueden ser otras empresas que necesiten datos específicos ya sea para ampliar sus propias bases de datos o para crear bases de datos nuevas con las que poder trabajar y que tienen los recursos para crear modelos o para poder sacar provecho de los datos adquiridos.

Los *earlyadopters* en este caso serían las pequeñas y medianas empresas de márketing que a fin de ofrecer un mejor servicio a sus clientes estarían interesadas en conocer mejor los sectores de sus clientes y estaría muy interesadas en acceder a beses de datos.

2. Propuesta de valor o ventaja competitiva

En esta parte se explica el valor innovador de la empresa respecto a otras del mercado que ya satisfacen esa necesidad o que tienen propuestas de valor similares. El objetivo es conocer el fator diferenciador y comprender lo que valora el cliente de la empresa. También se explica la necesidad que se satisface o que se ha creado y los mecanismos utilizados para ello, incluye una breve descripción del producto o servicio.

La propuesta de valor del *Marketplace* tiene dos partes, el alquiler de modelos y la venta de datos.

En primer lugar, el alquiler de modelos ofrece la posibilidad a las empresas de monetizar los modelos que ya tiene desarrollados y a la vez conservar la propiedad intelectual ya que los modelos no saldrían de sus servidores en la nube. Como se ha mencionado anteriormente también les permite monetizar de forma indirecta los datos que no pueden vender de forma directa mediante la creación de un modelo que haya sido entrenado con esos datos y que pueden poner en alquiler en la plataforma.

A los compradores les permite acceder a modelos desarrollados por empresas que disponen de muchos datos para entrenar los modelos por lo que es una manera de reducir la ventaja competitiva de las grandes empresas frente a las PYMES en un mondo en el que cobran cada vez más importancia el uso de los datos, así como su capacidad para procesarlos. Además, el acceso a estos modelos será más económico que desarrollar desde cero un modelo similar al ofrecido en el *Marketplace*.

La tecnología utilizada en esta parte es la encriptación homomórfica que permite que los valores de entrada del modelo estén encriptados, que las operaciones con esos datos se realicen de forma encriptada y que los resultados obtenidos estén también encriptados por lo que se

solucionan los problemas de privacidad del usuario evitando que la empresa suministradora del modelo pueda utilizar los datos introducidos para mejorar su modelo.

Además, la encriptación homomórfica resuelve el problema de trabajar con datos encriptados en la nube ahorrando capacidad de procesamiento, tiempo y energía.

En el caso específico de los *earlyadopters* que son universidades y centros de investigación en el ámbito de la medicina que ponen al servicio de la sanidad pública los resultados de sus investigaciones a través de modelos de predicción. A cambio los hospitales públicos pagarían un precio por el uso puntual del modelo. Esto es beneficioso para ambas partes ya que los centros de investigación consiguen fuentes de financiación adicionales que les permiten seguir invirtiendo en investigación y por lo tanto en el desarrollo de más modelos. La parte pública se beneficia al acceder a investigaciones de instituciones de renombre a un precio reducido que permite mejorar la calidad de los diagnósticos y ayudar a realizar diagnósticos tempranos que a largo plazo son beneficiosos para el paciente ya que le permite mantener su calidad de vida y localizar las enfermedades de forma temprana o a determinar el tratamiento óptimo reduciendo el tiempo de tratamiento y de recuperación. Esto también supone ventajas para los hospitales públicos ya que al ser las detecciones más tempranas se reducen los costes de los tratamientos y se reducen las listas de espera de los hospitales a largo plazo.

En segundo lugar, la propuesta de valor de la compraventa de datos es agrupar tanto la oferta como la demanda de bases de datos con el valor añadido que tienen el uso de las tecnologías que se han explicado al inicio de este documento. En concreto se utilizarían la *differential privacy*, la anonimización, los datos agregados y los datos sintéticos. La propuesta de valor para los vendedores de datos es la facilidad para encontrar compradores gracias a la agrupación de la demanda. Al poder vender los datos a múltiples compradores la rentabilidad aumenta.

Gracias a las tecnologías descritas se le da un valor añadido a este servicio que permite garantizar la privacidad y la seguridad gracias a las tecnologías utilizadas permitiendo a las empresas cumplir con la legalidad al compartir los datos.

Por otro lado, puede ocurrir que una empresa no disponga de los datos necesarios o suficientes para desarrollar los modelos que necesitan y sea necesario recurrir a bases de datos externas. El valor que ofrece el *Marketplace* es al tener agrupada la oferta es más sencillo conseguir lo que se necesita con exactitud e incluso obtener datos más diversificados por venir de diferentes zonas geográficas o de diferentes sectores. La compra de bases de datos puede resultar

interesante no solo para las empresas que deseen desarrollar modelos sino para cualquier empresa que quiera conocer mejor su sector o sus clientes actuales o potenciales.

En España la mayoría de las empresas son pequeñas y medianas empresas por lo que suelen subcontratar el márketing. Por este motivo se ha decidido que los *earlyadopters* de este modelo de negocio sean las agencias de márketing ya que les permite ofrecer un mejor servicio a sus clientes al poder acceder a nuevas bases de datos de diferentes sectores, además al tener a muchas PYMES como clientes les permite rentabilizar los datos que compren al repartir el coste entre un gran número de beneficiarios. Este escenario se describirá con mayor detalle más adelante.

3. Canales

Los canales son los medios utilizados para dar a conocer la propuesta de valor a los clientes objetivo y la forma en la que van a poder acceder al producto o servicio ofrecido. Se valorarán también los diferentes medios y la forma más rentable de distribuir el producto o servicio.

El *Marketplace* sería una página web con dos partes diferenciadas, una destinada a los modelos y otra destinada a la venta de datos. Todas las transacciones económicas se realizarían a través de la web. También los datos introducidos en los modelos necesitarían pasar a través de la web para poder utilizar las diferentes tecnologías. Toda la información de la web estará almacenada en la nube permitiendo que la empresa crezca con flexibilidad al poder contratar más espacio en caso de ser necesario.

Respecto a los canales de comunicación con los clientes se utilizaría el correo electrónico, y las consultas telefónicas para resolver dudas o problemas que puedan surgir.

Para captar clientes se utilizarían diferentes técnicas en función de los diferentes perfiles. La captación de grandes empresas tanto para que pongan sus modelos en alquiler como para que vendan sus datos se realizará a través de personal especializado en el sector que visite a las empresas y les venda la propuesta de valor. Para el escenario específico de la medicina se buscaría firmar contratos bilaterales entre el *Marketplace* y las universidades y entre el *Marketplace* y los hospitales públicos.

Sin embargo, para las pequeñas y medianas empresas se utilizarían canales como revistas y publicaciones específicas de cada sector, publicidad centrada en empresas y la participación en congresos y ferias relacionados con la tecnología y de los sectores específicos ya que permiten acceder a un gran número de clientes potenciales. En algunas circunstancias podría ser

interesante tener personal dedicado a convencerlas de utilizar los servicios ofrecidos. También se intentaría utilizar el boca-oído.

4. Relación con los clientes

En esta sección se describen las relaciones entre la empresa y los distintos segmentos de clientes y los costes asociados de estas relaciones. También está relacionado con el apartado de canales ya que a través de ellos se va a dar esta comunicación.

Las grandes empresas esperarán un servicio técnico y de asesoramiento o tal vez ofertas personalizadas de modelos que les puedan ser de utilidad.

Las empresas desarrolladoras de modelos esperan una red de clientes de calidad que paguen bien por los modelos que desarrollan. Y por lo tanto debemos conseguir esos clientes.

Las empresas pequeñas y medianas esperarán modelos relativamente asequibles o que merezcan la inversión y que satisfagan sus necesidades individuales a un precio razonable.

5. Flujo de ingresos

Los flujos de ingresos describen las diferentes formas en la que se monetiza la propuesta de valor y cuáles de ellas son las principales vías de ingresos. También se justifica el precio del producto o servicio sabiendo lo que el cliente está dispuesto a pagar por ello. Por último, se define el método de pago.

La propuesta de valor se monetiza gracias al cobro de comisiones por las transacciones económicas que se dan en el *Marketplace*, tanto en el negocio de la venta de datos como en el negocio del alquiler de modelos. El cobro de comisiones se justifica debido a los costes fijos de personal y mantenimiento de la web entre otros costes, el resto de los costes aparecen de forma detallada en la parte de estructura de costes del *bussines model canvas* y en la descripción del modelo de Excel de pérdidas y ganancias.

Además de las comisiones, se cobrará un porcentaje adicional para reinvertirlo de forma exclusiva en la captación de nuevos clientes y en la publicidad. Esto se justifica ya que es beneficioso tanto para el *Marketplace* como para las empresas que forman parte de él ya que al haber más usuarios habrá más bases de datos a la venta y también más compradores de datos. En la parte de alquiler de modelos también se verían beneficiados de esta medida debido a que habría más oferta de modelos y mayor demanda.

Por último, se podrían obtener unos ingresos adicionales por ser una empresa que sea *partner* tecnológico de Microsoft y ofrecer nuestros servicios a través de su portal. Estos ingresos pueden llagar a suponer hasta un porcentaje 10% del volumen de ventas, pero respecto al resultado neto podría suponer hasta un 33%.

6. Recursos clave

En esta sección se analizan los recursos necesarios para llevar a cabo el negocio permitiendo localizar los costes y las inversiones que son necesarias para más adelante poder estructurar el plan de negocios. Estos recursos clave deban estar optimizados y debe estudiarse su utilidad y si son realmente imprescindibles para el funcionamiento de la empresa.

Los recursos clave sin los cuales no se podría llevar a cabo el funcionamiento correcto del *Marketplace* son:

En primer lugar, la web que es el lugar en el que se realizan todas las transacciones económicas y que por tanto es imprescindible. En segundo lugar, los servidores en la nube son necesarios para almacenar la información de la empresa, así como para operar ya que la nube ofrece flexibilidad para poder adaptar los recursos a las necesidades cambiantes que se tengan. Los clientes que suministran modelos o que veden datos tendrán almacenados sus recursos en sus propios servidores a los que se llamará desde la web. En ningún momento se almacenarán esos recursos en los servidores del *Marketplace*. El último recurso clave y sin el cual no se podría ni comenzar a operar son los clientes ya que son a la vez los que aportan el valor al *Marketplace* por lo que un recurso clave son sus modelos y sus datos y por lo tanto un recurso clave es los empleados que se encargan de la captación de clientes.

7. Actividades clave

En este apartado se describe el funcionamiento diario de la empresa explicando la forma en la que se aporta valor al cliente y la forma a través de la cual esto ocurre. Es decir, que se explica en detalle el funcionamiento desde que el cliente se pone en contacto con la empresa hasta que recibe el producto o servicio.

La empresa se pone en contacto con grandes empresas a través de agentes de vetas que explican y asesoran durante todo el proceso. Estos agentes se encargan de buscar y convencer a estas empresas y organizaciones de los beneficios y las ventajas que pueden tener por ofrecer sus modelos en el *Marketplace*. Por lo que el trato con los grandes clientes es muy personalizado.

Las actividades clave se han ido describiendo en los apartados anteriores con más detalle.

8. Aliados clave

Este apartado describe los proveedores necesarios para llevar a cabo la actividad económica y los recursos que suministran. También se describen los colaboradores que pueden facilitar algunas partes del proceso y la forma en la que ayudan.

Los principales aliados en la venta de modelos en especial al principio son los hospitales y centros de investigación que disponen de modelos matemáticos que puedan tener un uso comercial y que precisen de especial privacidad debido a la naturaleza de los datos. Otro aliado importante en el sector salud va a ser la sanidad pública que podría utilizar los modelos por un precio asequible y respetando en todo momento la privacidad del paciente.

Los principales aliados en la compraventa de datos son las grandes empresas y las agencias de márketing.

Por último, un aliado que no es imprescindible pero que puede suponer ingresos adicionales es ser *partner* de Microsoft lo que podría aportar ingresos adicionales sin suponer un aumento excesivo de los costes.

9. Estructura de costes

En esta sección se describen los costes derivados de la actividad económica definiendo las prioridades y aquellos que son imprescindibles para el modelo de negocio, así como los precios de llevar a cabo las actividades clave y los costes de los recursos clave.

Los principales costes del *Marketplace* son de personal y de infraestructura. En primer lugar, los costes de personal engloban a agentes de ventas especializados por sector. Otros costes de personal son debidos a perfiles técnicos que se encargan de la logística y del funcionamiento de los diferentes modelos. Los costes de infraestructura son los derivados del funcionamiento del *Marketplace* como es la página web y la nube. Además de manera puntual se requerirán servicios de abogados subcontratados para redactar los contratos firmados por las empresas sobre los modelos y los datos comercializados. También se subcontratará a una agencia de márketing especializada que será la responsable de vender el servicio a empresas más pequeñas que puedan estar interesadas en utilizar los servicios ofrecidos.

Además de los costes de personal y de infraestructura hay algunos costes derivados de la actividad como puede ser la necesidad de un espacio de trabajo para los empleados en un momento dado. En el modelo de negocio en Excel se han contemplado dos posibilidades, la primera que solo exista la opción de teletrabajo y la segunda en la que se tiene una oficina o un espacio de *coworking* en el que poder trabajar de forma presencial.

Los costes se ven reflejados de forma concreta en el modelo de negocio de Excel que se presenta a continuación.

5.2. Modelo de negocio en Excel

Para el estudio del modelo de negocio del Marketplace se ha realizado un modelo en Excel con valores estimados de ingresos y de gastos tratando en todo momento de que sea lo más realista posible e incluso calculando con ciertos márgenes a fin de demostrar la viabilidad económica de la empresa. Para las estimaciones de gastos y de impuestos se han adjuntado en el propio Excel los enlaces a los datos que los justifican.

NOTA: El año 2022 comienza en el mes de mayo.

La estructura del Excel tiene las siguientes hojas:

- Cuenta de pérdidas y ganancias supuesto presencialidad: En este supuesto se considera que todos los empleados trabajan de forma presencial, se ha realizado un cálculo que estima en qué momento pasa a ser más económico el alquiler de una oficina o el alquiler de espacios de coworking, a partir de cierto número de empleados resulta más económico alquilar una oficina que un coworking con todos los gastos adicionales que una oficina implica como son los del montaje de la oficina, el renting de los pcs, la electricidad, el servicio de limpieza, mantenimiento y reparaciones de las instalaciones, el mobiliario, la conexión a internet, los suministros de la cocina (café, leche y azúcar) o las fotocopias.

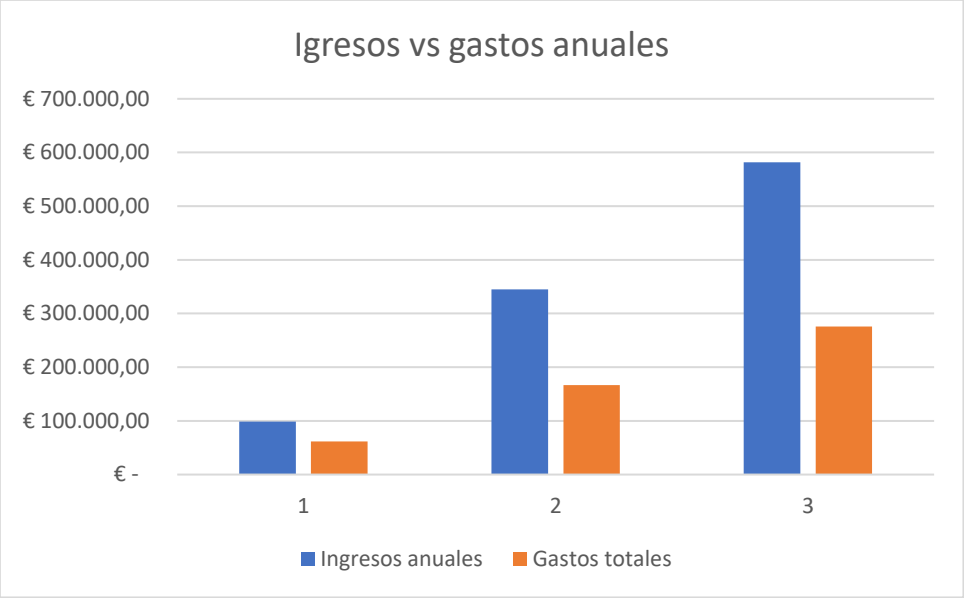


Ilustración 27: Gráfico de los ingresos vs los gastos anuales modelo presencial (elaboración propia)

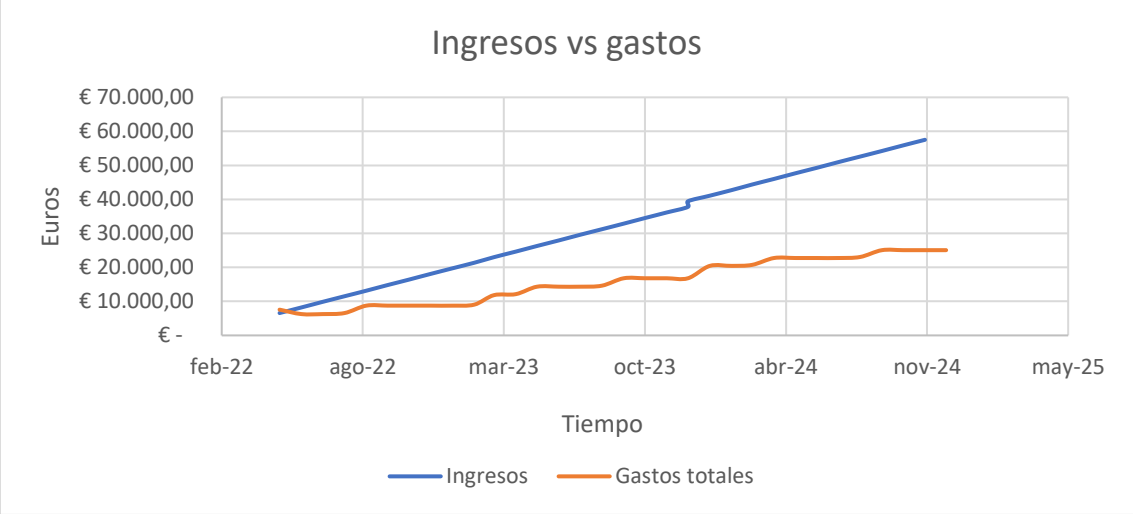


Ilustración 28: Gráfico de los ingresos vs los gastos modelo presencial (elaboración propia)

- Cuenta de pérdidas y ganancias supuesto teletrabajo: En este supuesto se considera que la mayoría de los empleados teletrabajan, de todas formas, se da la opción a los empleados de acudir presencialmente a un coworking. Para el cálculo se ha considerado que únicamente la mitad de los empleados trabaja presencialmente.

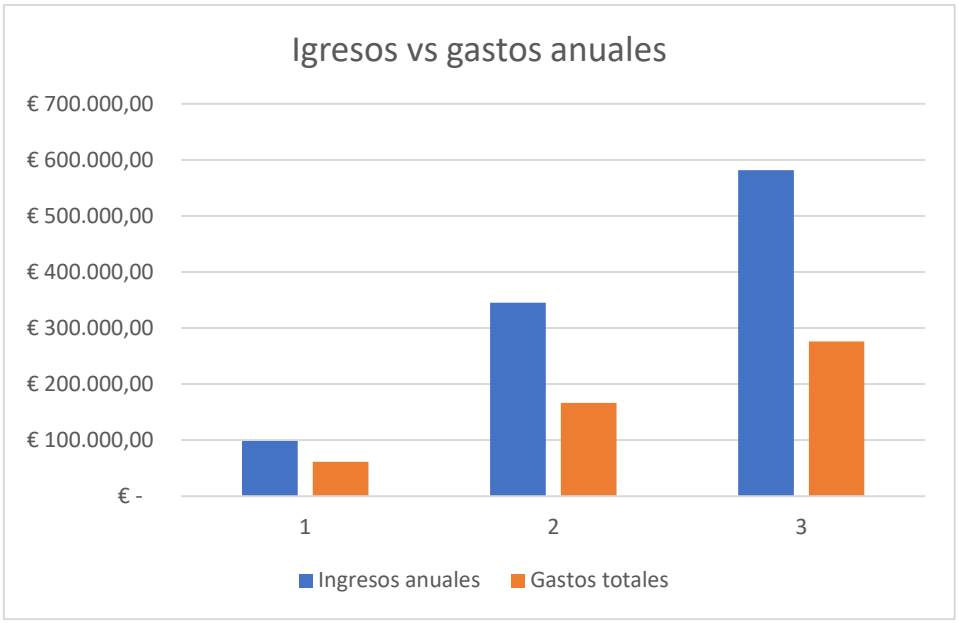


Ilustración 29:Gráfico de los ingresos vs los gastos anuales modelo semipresencial (elaboración propia)

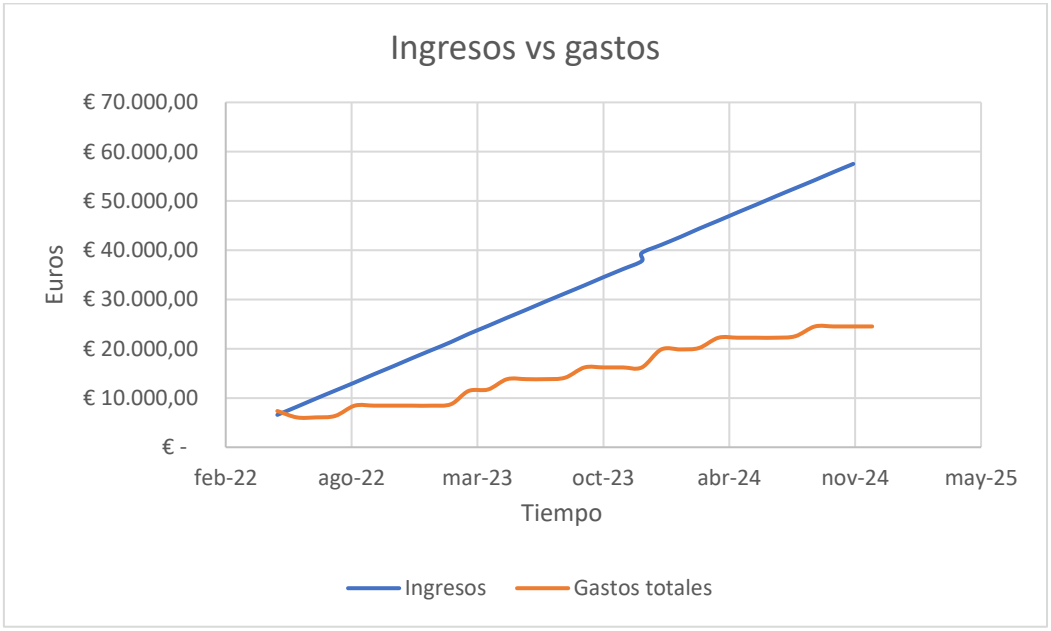


Ilustración 30: Gráfico de los ingresos vs los gastos modelo semipresencial (elaboración propia)

- Personal: en esta pestaña se calcula la evolución del número de empleados de cada categoría en función del crecimiento y las necesidades de la empresa. También se analizan sus sueldos, los costes para la empresa y los impuestos que deben pagarse a la SS.

  Entre los empleados se tienen agentes de ventas para cada sector, informáticos encargados del desarrollo de la web y de las aplicaciones para cliente. Más adelante se contratarían los servicios de un abogado de forma fija, al inicio se subcontratarían. En

el ámbito de las operaciones se contaría con personal administrativo y financiero y de márketing, aunque al igual que el abogado se contratarían de forma fija una vez la empresa esté establecida.

En la siguiente tabla se refleja el número de empleados de cada tipo junto con sus sueldos cada año.

Nº DE EMPLEADOS	Bruto mensual	Bruto Anual	AÑO 2022	AÑO 2023	AÑO 2024
Agente de ventas sectror	€ 1.500,00	€ 18.000,00	2	4	6
Informático	€ 2.158,33	€ 25.900,00	1	2	2
Abogado	€ 2.950,00	€ 35.400,00	0	0	0
Mánager	€ 2.708,33	€ 32.500,00	0	0	0
Administrativo	€ 1.500,00	€ 18.000,00	0	0	0
Márketing	€ 2.666,67	€ 32.000,00	0	0	1

Tabla 2: Número de empleados (elaboración propia)

- Seguridad social: en esta pestaña se realiza el cálculo desglosado de la seguridad social, que supone un 35%, si se desea ver con detalle pueden consultarse los anexos.
- Ingresos por venta de modelos: en esta pestaña se realiza una estimación de los ingresos por la venta de modelos. Se recuerda que el sector al que va dirigido esta línea de negocio es el sanitario. Para esta estimación se han buscado los datos relativos al número de muertes al año en España por causas cardiovasculares y por tumores de diferentes tipos. A partir de las muertes se ha estimado el número total de pacientes que hay en España de ambas categorías. A continuación, se ha hecho una estimación del porcentaje de esos pacientes que utilizarían nuestros servicios al mes. Y a partir del porcentaje se ha obtenido el número de pacientes mensuales. Se ha considerado que el servicio tiene un coste para el paciente de 10 euros de los cuales 7 van para la empresa creadora de modelos y 3 van de comisión por el servicio. De los 7 euros que recibe la empresa creadora de modelos 4 son directamente para ellos y 3 invierten el márketing y en los visitadores encargados de dar a conocer el modelo. En las tablas a continuación pueden verse las enfermedades más comunes en España con el número de muertes y de pacientes. Además de cómo se ha realizado el cálculo de los ingresos obtenidos por la venta de modelos mensual.

ENFERMEDADES CON MÁS FALLECIDOS EN ESPAÑA	Nº DE MUERTES
Enfermedades isquémicas del corazón	29.654,00
Enfermedades cerebrovasculares	25.817,00
Cáncer de bronquios y pulmón	21.893,00
Demencia	20.822,00
Insuficiencia cardíaca	19.358,00
Alzheimer	15.571,00
Enfermedad hipertensiva	14.271,00
Enf. Crónica de las vías respiratorias	12.734,00
Diabetes melitus	11.297,00
Cáncer de colon	11.131,00
Neumonía	8.768,00
Cáncer de páncreas	7.427,00
Insuficiencia renal	7.351,00

Tabla 3: Número de muertes en España por enfermedad (INE)

Para la estimación se han utilizado datos simplificados:

ENFERMEDADES CON MÁS FALLECIDOS EN ESPAÑA	Nº DE MUERTES	Nº DE PACIENTES EN ESPAÑA
Enfermedades cardiovasculares	116.215,00	3.290.000,00
Tumores	112.741,00	3.290.000,00

Tabla 4: Estimación del número de enfermos en España

	Nº DE MUERTES	Nº DE PACIENTES EN ESPAÑA									AÑO 2022
TOTAL PACIENTES MODELO	228.956,00	6.580.000,00	1.316,00	1.645,00	1.974,00	2.303,00	2.632,00	2.961,00	3.290,00	3.619,00	19.740,00
COSTE DE LOS MODELOS PARA EL PACIENTE	€ 10,00		€ 236,75	€ -	€ -	€ -	€ -	€ -	€ -	€ -	€ 236,75
INGRESOS TOTALES MEDICINA(MI COMISIÓN)	€ 3,00	30%	€ 71,03	€ -	€ -	€ -	€ -	€ -	€ -	€ -	€ 71,03
INGRESOS PARA LA CREADORA DE MODELOS			€ 165,73	€ -	€ -	€ -	€ -	€ -	€ -	€ -	€ 165,73
COSTE DE LOS MODELOS PARA EL PACIENTE MODELO SIMPLIFICADO	€ 10,00		€ 13.160,00	€ 16.450,00	€ 19.740,00	€ 23.030,00	€ 26.320,00	€ 29.610,00	€ 32.900,00	€ 36.190,00	€ 197.400,00
INGRESOS TOTALES MEDICINA(MI COMISIÓN) MODELO SIMPLIFICADO	€ 3,00		€ 3.948,00	€ 4.935,00	€ 5.922,00	€ 6.909,00	€ 7.896,00	€ 8.883,00	€ 9.870,00	€ 10.857,00	€ 59.220,00
INGRESOS PARA LA CREADORA DE MODELOS MODELO SIMPLIFICADO			€ 9.212,00	€ 11.515,00	€ 13.818,00	€ 16.121,00	€ 18.424,00	€ 20.727,00	€ 23.030,00	€ 25.333,00	€ 138.180,00

Tabla 5: Ingresos por venta de modelos del primer año (elaboración propia)

	AÑO 2022	AÑO 2023	AÑO 2024
TOTAL PACIENTES MODELO	19.740,00	69.090,00	116.466,00
COSTE DE LOS MODELOS PARA EL PACIENTE MODELO SIMPLIFICADO	197.400,00	690.900,00	1.164.660,00
INGRESOS TOTALES MEDICINA(MI COMISIÓN) MODELO SIMPLIFICADO	59.220,00	207.270,00	349.398,00
INGRESOS PARA LA CREADORA DE MODELOS MODELO SIMPLIFICADO	138.180,00	483.630,00	815.262,00

Tabla 6: Ingresos por venta de modelos (elaboración propia)

- Ingresos por venta de datos: en esta pestaña se realiza una estimación de los ingresos por la venta de datos. Se recuerda que el sector ala que va dirigido esta línea de negocio es el de las agencias de márketing. Aunque es necesario recordar que es extrapolable a otros sectores como la ingeniería o la consultoría. Para este cálculo se ha buscado el dato del número de agencias de márketing que hay en España y se ha hecho un cálculo aproximado del presupuesto anual y mensual que podrían tener para comprar modelos en función del tamaño de la empresa.

15000	TOTAL AGENCIAS	**AGENCIAS DE MÁRKETING (número)**
5%	750	Agencias grandes
25%	3750	Agencias medianas
70%	10500	Agencias pequeñas

Tabla 7: Número de agencias de márketing según tamaño (elaboración propia)

AGENCIAS DE MÁRKETING (dinero)	PRESUPUESTO MENSUAL	PRESUPEUESTO ANUAL
Agencias grandes	250	3000
Agencias medianas	125	1500
Agencias pequeñas	62,5	750

Tabla 8: Presupuesto de las agencias de márketing según su tamaño (elaboración propia)

A continuación, se ha realizado una estimación en porcentaje del número de agencias de cada tipo que son nuestros clientes.

Y con el presupuesto que se ha estimado se calculan los ingresos mensuales por la venta de modelos.

AGENCIAS DE MÁRKETING	may-22	jun-22	jul-22	ago-22	sep-22	oct-22	nov-22	dic-22	**AÑO 2022**
Agencias grandes	0,200%	0,250%	0,300%	0,350%	0,400%	0,450%	0,500%	0,550%	0,550%
Agencias medianas	0,200%	0,250%	0,300%	0,350%	0,400%	0,450%	0,500%	0,550%	0,550%
Agencias pequeñas	0,200%	0,250%	0,300%	0,350%	0,400%	0,450%	0,500%	0,550%	0,550%

Tabla 9: Porcentaje de agencias de márketing alcanzadas (elaboración propia)

AGENCIAS DE MÁRKETING	may-22	jun-22	jul-22	ago-22	sep-22	oct-22	nov-22	dic-22	**AÑO 2022**
Agencias grandes	375	468,75	562,5	656,25	750	843,75	937,5	1031,25	5625
Agencias medianas	937,5	1171,875	1406,25	1640,625	1875	2109,375	2343,75	2578,125	14062,5
Agencias pequeñas	1312,5	1640,625	1968,75	2296,875	2625	2953,125	3281,25	3609,375	19687,5
	2625	3281,25	3937,5	4593,75	5250	5906,25	6562,5	7218,75	39375

Tabla 10: Número de agencias de márketing alcanzadas por mes (elaboración propia)

INGRESOS POR SECTORES	may-22	jun-22	jul-22	ago-22	sep-22	oct-22	nov-22	dic-22	**AÑO 2022**
Agencias de márketing	€ 2.625,00	€ 3.281,25	€ 3.937,50	€ 4.593,75	€ 5.250,00	€ 5.906,25	€ 6.562,50	€ 7.218,75	€ 39.375,00

Tabla 11: Ingresos del primer año por venta de datos (elaboración propia)

INGRESOS ANUALES	**AÑO 2022**	**AÑO 2023**	**AÑO 2024**
Agencias de márketing	€ 39.375,00	€ 137.812,50	€ 232.312,50

Tabla 6: Ingresos anuales por veta de modelos (elaboración propia)

- Gastos: en esta pestaña se analizan los gastos de la empresa. El objetivo de esta pestaña es calcular el momento en el que los costes de una oficina son menores que los de un coworking. Si se desea ver con más detalle puede consultarse el Excel completo en el anexo.

5.3. Análisis DAFO

Él análisis DAFO cuyas siglas significan debilidades, amenazas, fortalezas y oportunidades permite realizar un análisis de la estructura interna de la empresa, así como la situación externa a fin de ver las ventajas competitivas y hacia dónde debe evolucionar la empresa.

No se ha desarrollado en mayor profundidad este gráfico ya que a lo largo de todo el documento se han explicado con detenimiento las diferentes partes de este DAFO, y con este gráfico simplemente se pretende visualizarlo de forma general.

Análisis DAFO

DEBILIDADES	AMENAZAS
• Empresa pequeña y con pocos recursos iniciales. • Elevado coste de captación de clientes. • Necesidad de creación de economías de escala	• Modelo de negocio que necesita gran cantidad de clientes para ser rentable económicamente. • La desconfianza de las empresas en el tratamiento de sus datos. • Poco interés por parte de las empresas en la venta de uno de sus principales activos. • Dificultad de establecer barreras de entrada
FORTALEZAS	**OPORTUNIDADES**

• No hay apenas competencia en el mercado. • Conocimiento profundo de las tecnologías a implementar. • Doble línea de negocio que se complementa entre sí.	• Aumento del uso de los datos. • Aumento en la confianza y el uso de modelos de ML y de AI. • Aumento de la inversión por parte de las empresas en tecnología y análisis de datos. • La legislación en materia de protección de datos va a ser cada vez más rígida. • Gran desarrollo reciente de tecnologías que permiten tratar con los datos de forma privada. • Demanda creciente de fuentes de datos de terceros • Posibilidad de crear un estándar de facto al ser la primera empresa ofreciendo el servicio

Tabla 12: Análisis DAFO (elaboración propia)

5.4. Análisis PEST

El análisis PEST cuyas siglas significan político, económico, social y tecnológico. Este análisis permite comprender en profundidad el contexto externo en el que se sitúa la empresa y realizar un estudio estratégico de la situación. En este caso se ha estudiado el análisis legal dentro del político por estar muy relacionados entre sí al ser las instituciones políticas las que toman decisiones sobre la legislación.

No se ha desarrollado en mayor profundidad este gráfico ya que a lo largo de todo el documento se han explicado con detenimiento las diferentes partes de este PEST, y con este gráfico simplemente se pretende visualizarlo de forma general.

Análisis PEST

POLÍTICO-LEGAL	ECONÓMICO
• Mayor protección por los datos personales. • Legislación en ley de protección de datos más restrictiva. • La legislación va con retraso respecto al desarrollo de las tecnologías. • Diferencias entre las legislaciones en materia de protección de datos europea y norteamericana.	• Situación económica de desaceleración de la economía y de elevada inflación. • Los datos ya se consideran el activo más preciado de las empresas. • Incremento de las inversiones de las empresas e instituciones públicas en análisis de datos al llegar a ser una ventaja competitiva.
SOCIOCULTURAL	**TECNOLÓGICO**
• Madurez del sector de los datos y de los modelos. • Cambio de mentalidad en la percepción de las personas en general sobre sus datos de carácter personal y el uso que hacen las empresas de estos. • Cambio de mentalidad en las empresas tradicionales que empiezan a invertir en nuevas tecnologías y que en ocasiones van muy retrasadas en materia de datos y seguridad.	• Gran evolución en los últimos años de tecnologías existentes que permiten que su uso pase de ser teórico a aplicable a la vida real. Madurez de las tecnologías. • Grandes inversiones por parte de empresas tecnológicas en el desarrollo de las tecnologías existentes y en la investigación de nuevas tecnologías. • Sector de los datos en auge con modelos de negocio que todavía no existen en lo que va a haber una gran innovación.

Tabla 13: Análisis PEST (elaboración propia)

6. Usos y aplicaciones (igual en lugar de usos y aplicaciones puedo poner escenarios específicos y explicarlos) (mover al 5)

Durante todo el trabajo se han ido mencionando los sectores que podrían beneficiarse de la creación de esta empresa y del funcionamiento del Marketplace en algunos casos concretos. En esta sección se incide de forma concreta en los casos de uso que se han considerado más interesantes o en los que resultaría más sencillo llevarlo a la práctica. Entre estos sectores se encuentra el sector de la salud y el sector del márketing.

6.1. Salud e investigación

Una de las motivaciones de este proyecto era contribuir al sector sanitario, que como se ha podido comprobar durante la pandemia es un sector clave y para el que es vital contar con los datos suficientes y con los modelos de predicción necesarios para facilitar el trabajo del personal sanitario.

Por ello la parte del modelo de negocio enfocada en el uso de los modelos está dirigida al sector sanitario. En este sector hay agentes diferenciados como los hospitales, los centros de salud, las universidades, los laboratorios, las farmacéuticas y los centros de investigación. Estos agentes tienen funciones diferentes y contribuyen unas a otras en su labor. El objetivo de este proyecto es hacer más accesible los modelos generados garantizando la independencia entre las diferentes instituciones a fin de garantizar la protección absoluta de los datos del paciente.

Como se ha mencionado en múltiples ocasiones este modelo de negocio resulta beneficioso para ambas partes ya que permite acceder a modelos punteros, innovadores y asequibles económicamente. A la vez a la empresa generadora de modelos le permite obtener ingresos adicionales que podría utilizar en la mejora de esa investigación o bien en otras investigaciones que no estén relacionadas. Además, cuentan con la ventaja de que la empresa que aquí se crea ofrecería servicios adicionales como el de los visitadores que se encarguen de vender a los médicos, hospitales y centros de salud los modelos creados facilitando que se usen de forma masiva.

6.2. Campañas de márketing y ventas

Como ya se ha mencionado la principal motivación de este proyecto era ayudar a las Pymes reduciendo la ventaja competitiva de las grandes empresas en el ámbito tecnológico, concretamente en el uso de los datos.

Por ello la parte del modelo de negocio enfocada en el uso de los datos está dirigida a las empresas de márketing y de publicidad. Las Pymes suelen subcontratar este tipo de servicios para campañas puntuales o porque es más económico que tener este servicio integrado dentro de la propia empresa. En España puede haber entre 15,000 y 20,000 agencias de marketing y publicidad (ReasonWhy, 2015) de diversos tamaños por lo que es un volumen de potenciales clientes a tener en cuenta.

En general las empresas de márketing suelen estar especializadas en clientes de sectores similares. También puede ocurrir que una misma base de datos pueda servir para dos empresas que no tengan nada que ver la una con la otra por lo que en caso de que no estén especializadas también les pueda resultar interesante el uso de los datos que se les propone.

El coste de comprar bases de datos es elevado, pero al poder distribuir el coste de adquisición de los datos entre muchos clientes pasa de ser un capricho innecesario para las pymes a ser algo que de manera indirecta les permite beneficiarse de las nuevas tecnologías. De manera indirecta porque se haría mediante las campañas de publicidad creadas por las agencias de márketing que serían mucho más especializadas y que tendrían un mayor índice de captación de clientes que antes de utilizar bases de datos adicionales.

6.3. Otros sectores

Además de los casos de uso mencionados se considera que este modelo de negocio puede ser exportado a prácticamente todos los sectores entre los que se destacan: el financiero, la consultoría y la hostelería y turismo.

El sector financiero es uno de los que más datos utiliza y en el que más modelos de predicción se utilizan para comprender o explicar la situación económica y hacer previsiones de mercado.

Por eso se considera que en un futuro podría ser interesante estudiar la adaptación del proyecto a las necesidades específicas de este sector.

La consultoría es un sector en el que las consultoras trabajan para diferentes clientes que buscan soluciones puntuales a problemas específicos, para resolver estos problemas es necesario que los clientes compartan información privada de la empresa con la consultora. Es cierto que para ello se firman estrictos acuerdos de confidencialidad. Sin embargo, para algunos datos específicos que el cliente prefiera proteger especialmente podría utilizarse la HE para proteger los datos y a la vez permitir que la consultora trabaje y obtenga resultados sin la necesidad de conocer los datos específicos de las empresas clientes.

La hostelería y el turismo es un sector en el que los datos de los viajeros están distribuidos entre diferentes empresas como las aerolíneas, los hoteles, las webs de comparadores, agencias de viajes, empresas de alquiler de vehículos, blogs de turismo y agencias de entretenimiento. Sin embargo, podría lograrse una fusión y un tráfico de modelos de predicción y/o de datos utilizando las tecnologías mencionadas que permitan a todas las empresas obtener ventajas de un mayor conocimiento del comportamiento de los clientes, a fin de ofrecer mejores servicios o servicios más personalizados.

Capítulo 4: Conclusiones

Limitaciones

En este apartado se describen de forma detallada las limitaciones del proyecto realizado, tanto las limitaciones tecnológicas como las limitaciones económico-empresariales.

Las limitaciones tecnológicas se han ido mencionando a lo largo del proyecto, pero en este apartado se agrupan todas juntas para poder tener una visión de conjunto.

La primera limitación está relacionada con las operaciones realizables, la encriptación homomórfica tiene muchas ventajas, sin embargo, solo puede realizar las operaciones polinómicas, sumar, restar, multiplicar y dividir. Esto limita en cierta medida los modelos que pueden hacer uso de ella. Es de esperar que tal vez en el futuro aumenten el tipo de operaciones que se pueden realizar.

La segunda limitación también es de la encriptación homomórfica, y está relacionada como ya se ha explicado con los niveles de operaciones. A más operaciones realizadas sobre un valor encriptado mayor ruido y menor precisión en los resultados obtenidos. Una forma de reducir este problema es utilizando los cambios de escala. Además, para poder operar dos valores es necesario que estos estén en el mismo nivel de operaciones, es decir que se hayan operado el mismo número de veces. Esto puede ser una dificultad a la hora de programar ya que debe tenerse en cuenta al operar valores entre sí.

La tercera limitación tecnológica es la limitación del volumen de datos que pueden tratarse con las diferentes tecnologías de privatización. Está claro que si se opera con los datos en bruto la capacidad de procesamiento es menor que si se les aplican transformaciones. A mayor transformación de los datos mayor capacidad de computación implicada, por ejemplo, en la encriptación homomórfica se produce una gran transformación de los datos al pasar del *plaintex* al *cipertext.*

La cuarta limitación tecnológica está relacionada con la tipología de datos tratables con la encriptación homomórfica ya que únicamente puede utilizarse con números decimales. Aunque es posible que en un futuro se aumenten los tipos de datos que procesa.

Las limitaciones económicas también se han ido mencionando a lo largo del trabajo.

La primera limitación económica es que para que el Marketplace sea rentable y funcione correctamente es necesario contar con un volumen muy elevado de empresas que trabajen con

nosotros y de grandes empresas que den credibilidad a los servicios prestados. Además, el coste de adquisición de un cliente nuevo es elevado.

La segunda limitación empresarial es que es un modelo de negocio que actualmente apenas tiene competencia directa, pero que las grandes empresas tecnológicas como Amazon, Google o Microsoft podrían replicar con relativa facilidad y en relativamente poco tiempo. Esta sería una competencia difícil de hacer frente. Hay que tener en cuenta que hay ventajas de que este proyecto haya sido pionero en muchas de sus propuestas ya que lleva una ventaja grande respecto a posibles competidores que puedan surgir. Además, se ofrecen servicios muy personalizados que las grandes empresas no pueden ofrecer.

La tercera limitación económica es la necesidad de generar confianza, este sistema sólo puede funcionar si los clientes confían en que sus datos estarán seguros y en que los modelos funcionarán de forma correcta. Para lo primero se pueden conseguir certificaciones y validaciones técnicas, así como incluir cláusulas contractuales, pero para lo segundo siempre se estará expuesto ya que el Marketplace no podrá ni verificar ni comprobar la calidad de los datos y modelos aportados.

Otras áreas de investigación y mejoras futuras

Para el desarrollo de este proyecto se ha realizado un estudio integral del problema y de la solución considerando los aspectos tecnológicos, económicos, legales y logísticos. A pesar de todo por falta de tiempo y conocimientos algunos temas no se han estudiado con la profundidad necesaria.

Como por ejemplo el aspecto legal de este proyecto que daría por sí solo para un trabajo de fin de grado de derecho y en el que no se ha hecho un mayor énfasis por motivos obvios entre los que se encuentra que el proyecto es de ingeniería y por lo tanto no es motivo de este trabajo hacer un mayor énfasis en esta parte. Es de suponer que la GDPR se vuelva más restrictiva con los años por lo que podría llegar a afectar al modelo de negocio que aquí se presenta, por lo que también podría resultar interesante el estudio futuro de este aspecto.

Como se ha mencionado en varias ocasiones a lo largo del trabajo, las tecnologías que se mencionan son tecnologías vivas que se encuentran en constante evolución y mejora y cabe esperar que en los próximos años hayan cambiado bastante y que muchas cosas mencionadas en este proyecto estén obsoletas. Por este motivo sería interesante realizar una revisión de las

tecnologías existentes y de las que puedan surgir en un futuro cercano para evaluar el impacto en este proyecto.

Otro aspecto que sería interesante revisar en el futuro es la adopción de las pequeñas y medianas empresas españolas del análisis de los datos, la forma en la que se ha llevado a cabo esta transición, en cuanto tiempo y con que tecnologías. También sería interesante evaluar las consecuencias sufridas por las empresas que no han sabido adaptarse a los nuevos tiempos y a las nuevas tecnologías.

Un aspecto interesante que habría sido interesante validar mediante entrevistas es la necesidad que tienen las empresas de diferentes sectores en mayor profundidad conociendo mejor sus necesidades reales y sus dificultades a fin de adaptar todavía más el proyecto. Como finalmente se ha dedicado gran parte de la carga de trabajo a la parte tecnológica por ser más relevante para un proyecto de ingeniería, estas entrevistas no se han llevado a cabo, aunque sería interesante realizarlas.

Por último, es necesario tener en cuenta que en los momentos en los que se termina de escribir este proyecto la economía mundial se encuentra en una situación excepcional, con inflación, crisis energética y de suministros especialmente tecnológicos. Esta variable no se ha tenido en cuenta en el desarrollo económico del trabajo y sin duda sería interesante realizar una revisión en profundidad cuando se tengan más datos de la situación económica y financiera.

A pesar de todos los temas que podrían ser interesantes revisar en el futuro se considera que el trabajo realizado ha cumplido con los objetivos marcados al inicio del proyecto y que ha sabido adaptarse a los conocimientos que se iban adquiriendo para que el proyecto siguiera teniendo sentido. También se han priorizado partes más técnicas como el desarrollo del código de HE que en un principio iba a ser más sencillo de lo que finalmente ha sido, sobre otras partes más ilustrativas pero que aportaban un menor valor al proyecto.

Conclusiones

Al principio de este trabajo se comenzó mencionando las diferencias entre la adopción de nuevas tecnologías relacionadas con los datos entre las grandes empresas y las pymes. Y las ventajas competitivas que daba a unas sobre las otras. El objetivo de este proyecto era contribuir en la medida de lo posible a reducir esta brecha. Creo que el proyecto ha cumplido con los objetivos que se plantearon en primer lugar, que se ha cumplido con la metodología propuesta y que se han utilizado las herramientas propuestas.

En apartados anteriores ya se ha explicado las limitaciones del proyecto, las investigaciones futuras y otras áreas de investigación. Por lo que se procede con la validación de las hipótesis iniciales.

También se planearon unas preguntas que debían ser respondidas. A lo largo del proyecto se ha ido respondiendo de forma detallada a casi todas ellas. Para muchas de ellas se ha logrado una respuesta satisfactoria y completa y otras tienen una respuesta que evolucionará con el tiempo. Otras de las preguntas como las relacionadas con la ley se han tratado de forma superficial ya que no se considera que sea materia de este proyecto analizarlas en más profundidad. A continuación, se realiza una recopilación de las conclusiones obtenidas tras la labor de investigación y desarrollo.

Primero, mediante un análisis Benchmark se ha realizado un estudio de otros Marketplace que hay en el mercado y se ha observado que los datos son un modelo de negocio interesante y que las grandes empresas tecnológicas están interesadas en entrar en el mercado. Sin embargo, no hay ningún modelo de negocio que ofrezca modelos de machine learning.

Segundo, se confirma la existencia de tecnologías como la encriptación homomórfica, la *differential privacy* o los datos sintéticos cuyo objetivo es dar mayor protección a los datos y a la vez permitir que sigan pudiéndose utilizar y que puedan obtenerse resultados fieles a la realidad. Estas tecnologías son la gran propuesta de este proyecto para diferenciarse de otros de la competencia ya que aportan un valor diferencial.

Tercero, se concluye que la nube es la mejor opción en cuanto a dónde almacenar y ejecutar el proyecto por muchos motivos como el ahorro económico comparado a usar servidores locales, mayor adaptabilidad al crecimiento y necesidades de cada momento, cuidado del medio ambiente reduciendo de forma considerable las emisiones de CO_2 a la atmósfera y gran disponibilidad de los archivos almacenados y de los programas que deban ejecutarse a través de cualquier dispositivo con conexión a internet.

Cuarto, se ha demostrado que el modelo de negocio es económicamente viable y que es posible crear esta empresa. Analizando el servicio al cliente, las relaciones entre las empresas a las que se prestaría servicio, la estructura interna de la empresa, los empleados y recursos necesarios para llevar a cabo el proyecto y la forma en la que se captarían nuevos clientes.

Quinto, este proyecto contribuye a la sociedad concretamente al sector sanitario ya que es el principal beneficiario, con este proyecto se facilitan las relaciones entre los diferentes agentes del sector a fin de impulsar la aceptación de los modelos de predicción desarrollados por grandes

centros de investigación que contribuyen ayudando a los médicos con el diagnóstico de los pacientes, ayuda en la toma de decisiones sobre el mejor tratamiento dando una segunda opinión al doctor o ayudando en la gestión hospitalaria a fin de agilizar las esperas y ofrecer un servicio más eficiente y de mejor calidad. Además, se contribuye al desarrollo de las pymes a través de las empresas de márketing que subcontratan ya que las campañas de publicidad y visibilización serán más dirigidas y con una mayor captación de clientes que favorece el aumento de las ventas.

Sexto, con este proyecto se contribuye a la consecución de algunos de los ODS, como, la salud y bienestar, el crecimiento económico, la reducción de las desigualdades, la acción por el clima, la paz, la justicia y las instituciones sólidas y por último las alianzas para lograr objetivos.

Se concluye por lo tanto que el desempeño del trabajo ha sido satisfactorio, que ha cumplido con los objetivos marcados y se ha sabido adaptar a los cambios que pudieran ocurrir al evolucionar la investigación sobre la tecnología, el mercado, la competencia o sobre las dificultades técnicas o logísticas. Además, se han respondido a las preguntas que se plantearon en la introducción y se han llevado a cabo demostraciones de tipo práctico tanto tecnológicas como económicas.

ANEXO 1: OBJETIVOS DE DESARROLLO SOSTENIBLE

Los objetivos y metas de desarrollo sostenible (ODS) son el conjunto de objetivos globales que desde el año 2015 se marcaron los países con el objetivo de resolver algunos de los grandes problemas de nuestra época. Estos objetivos tienen como fecha de cumplimiento el año 2030. Son un total de 17 y están relacionados con problemas como: la pobreza, el hambre, la salud, la educación, la igualdad, el agua, la energía, el crecimiento económico, la industria, la sostenibilidad, el clima, la paz y las alianzas.

Los Objetivos de Desarrollo Sostenible (ODS) con los que se alinea el proyecto son los siguientes: objetivo de salud y bienestar, objetivo de crecimiento económico, objetivo de reducción de las desigualdades, objetivo de paz, justicia e instituciones sólidas y objetivo de alianzas para lograr objetivos:

Objetivo 3: Salud y bienestar

El objetivo de salud y bienestar tiene como objetivos reducir la mortalidad materna e infantil, terminar con las epidemias de SIDA, tuberculosis, malaria y enfermedades relacionadas con el agua. También busca reducir el consumo de sustancias adictivas como el tabaco, el alcohol o los estupefacientes y reducir las muertes provocadas por la contaminación del suelo, aire o el agua. Reducir los fallecimientos por accidentes de tráfico. Que el acceso a la salud sea universal y de calidad, con acceso a medicamentos, vacunas y tratamientos tanto de prevención como de curación.

El proyecto de Marketplace tiene como objetivo centrarse en las empresas e instituciones dedicadas a la salud, en especial en los primeros años del desarrollo del proyecto por ser de los sectores en los que los datos son de especial sensibilidad y que requieren una protección especial.

Además, este sector puede verse muy beneficiado por la creación de este proyecto al permitir de manera accesible que el máximo número posible de instituciones pueda beneficiarse de las nuevas investigaciones llevadas a cabo por todo el mundo.

Este proyecto permitiría que hospitales de todo el mundo puedan hacer uso de programas y modelos de predicción y de Machine Learning desarrollados por universidades y centros de

investigación asegurando la confidencialidad. Y tal y como se ha explicado anteriormente los desarrolladores de modelos conserven la propiedad intelectual y puedan obtener un beneficio económico del modelo favoreciendo investigaciones futuras al tener nuevas fuentes de financiación. Por otro lado, los hospitales que usarán esos modelos en el diagnóstico de enfermedades o en modelos de eficiencia sobre gestión hospitalaria podrán introducir los datos encriptados y recibir los resultados encriptados sin comprometer en ningún momento los datos.

Objetivo 8: Crecimiento económico

El objetivo 8 de crecimiento económico tiene como metas el crecimiento económico de los países en vías de desarrollo, la inversión en tecnología e innovación. Además, promueve el desarrollo de pequeñas y medianas empresas, que era uno de los principales objetivos de este proyecto. Reducir la dependencia entre el crecimiento económico y el uso y explotación de los recursos naturales que dañan el medio ambiente. También busca reducir el desempleo especialmente entre los jóvenes y que las remuneraciones y las condiciones de trabajo sean dignas e iguales para hombres y mujeres. Por último, también propone proteger los derechos laborales de los empleados, así como los ambientes laborales seguros.

Este objetivo está relacionado con el proyecto en la medida en la que las pequeñas y medianas empresas se podrán beneficiar de los modelos desarrollados por instituciones de mayor tamaño un precio reducido. Permitiendo a estas empresas ser más competitivas y que tengan mayores oportunidades de crecimiento económico. Por lo que supondrá un crecimiento económico a nivel individual de las empresas y también un crecimiento económico a nivel global de la economía crenado un tejido empresarial más fuerte y competente frente a las grandes empresas.

Objetivo 10: Reducción de las desigualdades

El objetivo de reducción de las desigualdades mediante el cumplimiento de una serie de metas entre las que se encuentran: Incremento de los ingresos de las personas más pobres, trabajar en la inclusión social, económica y política de todas las personas y que no sean discriminadas por ningún motivo. También busca la obtención de la igualdad de oportunidades mediante políticas fiscales y legales. Por último, busca que los estados más desarrollados inviertan en los países en vías de desarrollo a fin de promover su crecimiento.

Mediante el desarrollo de este proyecto se reducirán las desigualdades entre empresas e instituciones ya que hará más accesible la tecnología de Machine Learning y el procesamiento de los datos, que en este momento queda limitado a las grandes empresas. Por lo que las pequeñas y medianas empresas, que en España suponen la mayor parte del tejido empresarial, sean más competentes y tengan más oportunidades frente a las grandes empresas.

Además, al mantenerse la privacidad total sobre los datos por el uso de la encriptación homomórfica los usuarios también verán reducida su vulnerabilidad frente a las grandes empresas reduciendo esta relación desigual entre empresas y usuarios.

Objetivo 13: Acción por el clima

El objetivo de acción por el clima tiene como metas el establecimiento de políticas que busquen mejorar la educación y conciencia de las personas hacia el cambio climático a fin de reducir los efectos provocados por la acción humana. Especialmente en países subdesarrollados o en vías de desarrollo.

Este objetivo se cumple en el proyecto gracias a varios factores. En primer lugar, al reducirse la capacidad de computación al no tener que encriptar y desencriptar los datos constantemente se produce un ahorro importante en capacidad de computación y por lo tanto un ahorro en el consumo energético contribuyendo al cuidado del medio ambiente. En segundo lugar, el proyecto contará con un estudio sobre el impacto de C02 buscando siempre que sea el menor posible. Por último, al estar el proyecto en la nube se evitará sobredimensionar el sistema comprando equipos físicos innecesarios ya que la nube permite adaptar el almacenamiento y la capacidad de procesamiento a las necesidades de cada momento, haciendo que el proyecto esté muy optimizado, contribuyendo así al cuidado del medio ambiente.

Objetivo 16: Paz, justicia e instituciones sólidas

El objetivo de paz, justicia e instituciones sólidas busca la reducción de la violencia y de la mortalidad debido a ella. Además, busca poner fin a todas las formas de violencia como pueden ser el maltrato, la tortura, la explotación, los robos, la corrupción y la delincuencia organizada. Para ello busca crear instituciones sólidas, eficaces y transparentes que tomen decisiones que persigan este objetivo, además promover el acceso a la información y proteger las libertades fundamentales. Promover la seguridad jurídica y la no discriminación.

Gracias a este proyecto se contribuiría a mejorar la privacidad de los usuarios y de las empresas además de protegerlos frente a los ciberataques que en los últimos años han aumentado los casos de ransomware (secuestro de datos) tanto a empresas como a particulares, en concreto en el caso de las empresas aumentó de 100 millones de casos en 2018 a 200 millones en 2021. Esto es posible gracias a la encriptación homomórfica ya que evita tener que desencriptar los datos para tener que operar con los datos por lo que hay menos posibilidades de que se intercepten. Por lo tanto, este proyecto contribuye a este objetivo de varias formas. En primer lugar, al contribuir a la protección de datos facilita a las empresas cumplir con la ley contribuyendo al objetivo de justicia. En segundo lugar, al reducir el riesgo de ciberataque contribuye al desarrollo de instituciones sólidas.

Objetivo 17: Alianzas para lograr objetivos

El objetivo de alianzas para lograr objetivos tiene como metas en el ámbito financiero la inversión de los países desarrollados en países menos desarrollados mediante la movilización de recursos adicionales vía impuestos. También busca promover que estos países tengan una deuda pública reducida y que en caso de ser necesario hagan frente a las deudas de los países con deudas muy elevadas. Las metas en el ámbito tecnológico buscan que los países se apoyen y ayuden entre si compartiendo e intercambiando conocimientos, innovación, ciencia y tecnologías. Favoreciendo de manera especial aquellas tecnologías que sean de carácter sostenible. En el ámbito del comercio internacional favorecer un mercado justo, equitativo y que favorezca las exportaciones de los países en especial de los más desfavorecidos. Por último, buscar la estabilidad a nivel macroeconómico sin eliminar la legitimidad de los países para legislar y tomar medidas económicas y legales para reducir la pobreza y las desigualdades en sus fronteras. Todo ello mediante el intercambio de conocimientos entre las diferentes instituciones y empresas mediante alianzas. Y la parte más relacionada con este proyecto la del valor de que los países tengan datos de calidad, fiables, suficientes que permitan tener un control sobre el estado de su población a fin de poder medir las mejoras y localizar los problemas y poder tomar medidas específicas.

Este es claramente uno de los objetivos del proyecto ya que depende de la colaboración entre instituciones y empresas. Esta colaboración sopone una relación de simbiosis ya que todos los participantes ganan algo con el desarrollo de este Marketplace en especial pequeñas y medianas instituciones con un acceso limitado al desarrollo de modelos matemáticos.

ANEXO 2: ENTREVISTAS A EXPERTOS

Para el desarrollo de este proyecto se han realizado entrevistas a diferentes expertos a fin de validar la idea, obtener ideas de mejora y aprender de profesionales que conocen mejor las tecnologías empleadas. En este anexo se pretende resumir muy brevemente las conversaciones con las distintas personas entrevistadas y su aportación a este proyecto.

Pablo Montoliu, Chief Information and innovation officer en Aon Seguros, en esta entrevista se validó el modelo de negocio ya que se le presentó la idea y se le explicó el funcionamiento de la empresa y de los servicios ofrecidos a fin de comprobar si se pudiera aplicar al sector de las aseguradoras. Pablo encontró el modelo muy interesante especialmente para las empresas internacionales que deben someterse a legislaciones diferentes en función de dónde operen. Además, consideró que las aseguradoras se podrían beneficiar de trabajar con un mayor volumen de datos a fin de reducir el riesgo implícito de las empresas de seguros.

David Rodriguez, director desarollo de negocio en Turing Challenge, en esta entrevista se centró en validar el modelo de negocio y ver cómo debía hacerse la cuenta de pérdidas y ganancias, la estructura y las características de una empresa tecnológica. David ha fundado Turin empresa en la que trabaja, que se dedica a las nuevas tecnologías como la IA o el 5G.

Ester de Nicolás,strategic Missions and technologies director en Microsoft, ha estado implicada en el proyecto de differential privacy, además a lo largo de su vida profesional ha colaborado con muchas empresas y sugirió buenas ideas de partnerships que se podían hacer o de empresas a las que les podría interesar este proyecto. Ester además validó la idea del proyecto y de las líneas de negocio del proyecto, apoyando de forma especial el uso de la differential privacy.

Todd Singleton, Cloud Data Executive en Microsoft, aún no trabajaba en Microsoft cuando existía el Marketplace de datos de Microsoft, pero tenía muchos conocimientos sobre el tema por el desarrollo de informes al respecto en los que ha colaborado. Por eso el objetivo de esta entrevista era comprender como era el funcionamiento del proyecto, los servicios que ofrecía y lo más importante los motivos por los que terminó fracasando. Esta entrevista fue especialmente importante para validar las ideas que se tenían de este TFG y de las formas en las que se podía mejorar la propuesta de Microsoft y de evitar cometer los mismos errores.

Andrea Ortega, abogada, en esta entrevista se trataron los aspectos relacionados con las formas legales en las que se podía proteger esta idea de negocio frente al plagio de las grandes

empresas. Lamentablemente no existe ningún mecanismo que proteja las ideas por lo que este modelo de negocio podría ser copiado por las grandes empresas tecnológicas. Para evitarlo Andrea propuso que se buscaran valores diferenciales y barreras de entrada como por ejemplo las tecnologías propuestas y los servicios personalizados que se ofrecen. También se trataron los problemas relacionados con la protección de los datos y se validó la idea de que este proyecto podría abrir nuevas opciones a la explotación de los datos gracias a las tecnologías de privatización cumpliendo la ley.

Ignacio Pereña, general counsel and secretay of the board of director en Redexis, abogado. En esta entrevista al igual que en la de Andrea Ortega se ha analizado el problema legal de este proyecto, de forma algo superficial y explicada de forma sencilla al no tener esta autora los conocimientos necesarios de derecho. El objetivo era conocer los requisitos que se debían cumplir y comprender que la ley no contempla de forma específica estas tecnologías, por lo que es necesario aplicar la ley general y el proyecto puede llevarse a cabo siempre y cuando se cumpla con la legislación europea y se firmen los acuerdos correspondientes entre los clientes sobre el uso y el tratamiento de los datos.

Kim Laine, principal research manager en Microsoft, es el principal desarrollador de la encriptación homomórfica y está dedicado a la investigación y en el desarrollo de ejemplos y se las librerías disponibles en GhitHub, además se dedica a la investigación de nuevas aplicaciones de esta tecnología y en hacerla más eficiente y con mayor capacidad de cálculo. En esta entrevista se analizaron y resolvieron dudas respecto a la encriptación homomórfica, concretamente dudas técnicas, en la que explicó la importancia de respetar los árboles de operaciones y el escalado de los resultados tras cada operación, además dio algún ejemplo sobre usos de la encriptación homomórfica como para la verificación de si una cuenta concreta ha sufrido un secuestro de datos o no.

Jaime Pereña, director growth innovation and strategy en Microsoft, ha sido el tutor de este trabajo fin de grado y ha colaborado aportando su conocimiento sobre AI, ML y ofreciendo perspectiva y experiencia para orientar el proyecto.

ANEXO 3: ESTUDIO MEDIOAMBIENTAL

Este trabajo fin de grado está comprometido no solo con las empresas, los usuarios y su privacidad, sino que también está comprometido con el cuidado del medio ambiente, esto se describe en el anexo de los ODS. Por este motivo la idea es que toda la empresa esté almacenada en la nube y que los programas servidor, cliente y administrador se ejecuten en la nube. Tal y como se ha explicado en la revisión de las tecnologías en el apartado específico de la nube, trasladar los servidores locales a la nube supone un ahorro en coste económico, pero también un ahorro considerable en las emisiones de C02 a la atmósfera. Para demostrarlo se ha realizado una comparativa aproximada del CO2 enviado a la atmósfera si se utilizaran servidores locales y si se utilizan centros de datos públicos como Azure.

La estimación se ha hecho de la siguiente manera:

Primero se han obtenido los datos numéricos sobre los kilos de dióxido de carbono producidos por cada usuario tanto en la nube como en servidores locales, se han obtenido también los datos sobre las emisiones de CO2 por cada GB ejecutado, se han mirado la electricidad consumida por 6 servidores ejecutando el modelo de negocio de venta de datos y 6 servidores ejecutando el modelo de negocio de venta de modelos al año. También se ha buscado el precio del kW/h en España actualmente y en base al porcentaje de energías renovables y no renovables se obtiene los kg de C02 por cada kW/h.

KgCO2e per user (Public cloud)	6,00
KgCO2e per user (Avg on premises)	30,00
CO2 emissions per GB Public cloud (Avg)	7,90
CO2 emissions per GB On Premises (Avg)	1,40
Electricity(6+6 servers+5GB) - year	2.132,80
Precio por kw/h	0,33
Total Kw/h - Year	**6.531,31**
Kg CO2 Per Kw/h	0,19
Azure Kg Co2 per year per VM	**40,65**
KgCO2 per 10 VMs	**406,52**

Tabla 14: Datos utilizados en el cálculo y las fuentes consultadas (elaboración propia)

Input Web
Calculation

Parameter
Output

Tabla 15: Leyenda (elaboración propia)

Para el cálculo se han tomado unos valores de usuario sencillos para comprender la magnitud del ahorro, a lo largo de 5 años. La descripción de las operaciones realizadas se encuentra tras la tabla.

	Y1	Y2	Y3	Y4	Y5
# Users	5,0	10,0	15,0	20,0	25,0
KgC02 BackOffice on Premises	150,0	300,0	450,0	600,0	750,0
KgCO2 BackOffice Public Cloud	30,0	60,0	90,0	120,0	150,0
Storage on Premises	5,0	10,0	15,0	20,0	25,0
kgCO2Storage on Premises (#GB)	7,0	14,0	21,0	28,0	35,0
kgCO2Storage Public Cloud	39,5	79,0	118,5	158,0	197,5
KgCO2 Servers On Premise	1.254,0	1.254,0	1.254,0	1.254,0	1.254,0
kgCO2 Servers Azure	406,5	406,5	406,5	406,5	406,5
Total KgCO2 on premises	1.411,0	1.568,0	1.725,0	1.882,0	2.039,0
Total KgCO2 on Azure	476,02	545,52	615,02	684,52	754,02
Savings	934,99	1.022,49	1.109,99	1.197,49	1.284,99
Total savings					5.549,97

Tabla 16: Cálculo medioambiental (elaboración propia)

Para el primer bloque se ha multiplicado el número de usuarios por los KgCO2 emitidos por usuario en servidores locales, repitiéndose el proceso para la nube pública.
Para el segundo bloque se ha multiplicado el almacenamiento en servidores locales por las emisiones de CO2 por gigabyte en servidores locales, siguiéndose la misma metodología para sacar los KgCO2 de almacenamiento en la nube pública. A continuación, se ha multiplicado los KW/h totales en un año por los KgCO2 por KW/h, a fin de extraer los KgCO2 en servidores locales.

Para el último bloque, se han sumado los respectivos apartados anteriores según el servidor que se utiliza, siendo una suma entre los KgCO2 por usuarios, los KgCO2 por almacenaje y los KgCO2 de servidores, a fin de extraer una cifra total. Finalmente, los ahorros son el resultado de la resta del total de KgCO2 en Azure (nube pública) al total de KgC02 en servidores locales.

Estos cálculos son repetidos para un periodo de 5 años mediante los usuarios y almacenaje establecidos en las tablas, suponiendo un mantenimiento de los datos de la primera tabla en el tiempo, obteniendo de esta forma un ahorro acumulado durante ese periodo de 5.549,97 Kg de CO2.

ANEXO 4: CÓDIGO COMPLETO

En este anexo se presenta el código completo ya que se considera que constituye una parte imprescindible del proyecto y que refleja gran parte del trabajo realizado. En el documento en si se mencionan las partes más relevantes y se eliminan las partes repetitivas a fin de facilitar la lectura y comprensión de este.

PROGRAMA DESCRIPTIVO DEL FUNCIONAMIENTO DE LA ENCRIPTACIÓN HOMOMÓRFICA

PROGRAM

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.IO;
using System.Numerics;

namespace Formulario_TFG
{
    static class Program
    {
        //HAY QUE NOMBRAR BIEN A MI FORMULARIO 3 y 4
        static FormSIniz MiFormFormSIniz;
        static FormCliente MiFormC;
        static FormServidor MiFormS;
        static FormResTot MiFormRes;

        static PublicKey publicKey;
        static RelinKeys relinKeys;

        static Encryptor encryptor;
        static Evaluator evaluator;
        static Decryptor decryptor;

        static SEALContext context;
        static double scale;
        static CKKSEncoder encoder;
```

```
        static Plaintext[,] PlainCoefs;
        static Plaintext Coef7;
        static bool[,] BCoeficientes;

        static Ciphertext[] XCoefsEncriptados;
        static public Ciphertext[] ArrayEncryptedResult;


        static public bool Bmore;

        //Número de filas (Variables) que tiene la matriz
        static public int numvars;
        //Número de columnas (Coeficientes) qeu tiene la matriz
        static public int numcoef;
        //FILA VARIABLES (para recorrer la matriz)
        static public int nvar;
        //COLUMNA COEFICIENTES DE CADA VARIABLE (para recorrer la
matriz)
        static public int ncoef;

        //Vector de texto que contiene los nombres de as distintas
variables, estos nombres se asignan por el servidor, el vector tiene
la dimensión qeu le da el servidor
        static public string[] NombreVar;

        //Array de ciphertext que contiene los valores de las
variables
        static public Ciphertext[] xEncrypted;

        //Vector de sumandos desencriptados
        // static public double[] SumandosDes;

        //No se si me convence lo de tener el result como variable de
todo el programa, debería ser una variable local del formulario

        static public List<double> result;
        static public List<double> SumandosDes;
        static public Ciphertext ChiperResultado_total;

        [STAThread]
```

```
        static void Main()
        {
            numcoef = 9;
            //PARA EL FORMULARIO, SE PONEN POR DEFECTO
            Application.EnableVisualStyles();
            Application.SetCompatibleTextRenderingDefault(false);

            //Falta nombrar bien a mi formulario 3 y 4
            MiFormFormSIniz = new FormSIniz();
            MiFormC = new FormCliente();
            MiFormS = new FormServidor();


            //Declarada en la clase y asignado el valor (2^50 por
semejanza con los params de encriptacion)
            scale = Math.Pow(2.0, 50);
            context = Inicializar();
            CrearEncoder();

            Bmore = true;

            while (Bmore) {
                //Falta nombrar bien a mi formulario 3 y 4
                MiFormFormSIniz= new FormSIniz();
                MiFormC = new FormCliente();
                MiFormS = new FormServidor();


                //Formulario 3
                Application.Run(MiFormFormSIniz);
                NombreVar = new string[numvars];
                numcoef = 9;

                //VARIABLES PARA COMPLETAR EL FORMULARIO SERVIDOR
                //NOTA, EN REALIDAD NO NECESITO ALMACENAR ESTE VALOR
EN LA MATRIZ, CON TENER UN VECTOR QUE VOY CREANDO CADA LOOP SOLO CON
LA VARIABLE CORRESPODIENTE BASTA
                //Creación del array de cipertextos que contienen el
valor de las distintas variables
                xEncrypted = new Ciphertext[numvars];
```

```
                //Creación del array de Plains que contienen el valor
de los coeficientes de las distintas variables
                //NOTA, EN REALIDAD NO NECESITO ALMACENAR ESTE VALOR
EN LA MATRIZ, CON TENER UN VECTOR QUE VOY CREANDO CADA LOOP SOLO CON
LA VARIABLE CORRESPODIENTE BASTA
                PlainCoefs = new Plaintext[numvars, numcoef];
                //Creación del array de Bools que contienen el valor 1
si el coeficiente es distito de 0 y un 0 si el coeficiente es igual a
0
                //NOTA, EN REALIDAD NO NECESITO ALMACENAR ESTE VALOR
EN LA MATRIZ, CON TENER UN VECTOR QUE VOY CREANDO CADA LOOP SOLO CON
LA VARIABLE CORRESPODIENTE BASTA
                BCoeficientes = new bool[numvars, numcoef];

                //NOTA: Antes de que esto estuviera aquí, lo que
pasaba es que como en el loop del formulario se hacian todos los
cálculos no pasaba nada porque funcionaba


                //Se desplaza por las distintas filas (variables) de
la matriz
                for (nvar = 0; nvar < numvars; nvar++)
                {
                    //Formulario 2
                    MiFormS = new FormServidor();
                    Application.Run(MiFormS);
                }

                // VARIABLES PARA COMPLETAR EL FORMULARIO CLIENTE
                XCoefsEncriptados = new Ciphertext[numcoef];
                SumandosDes = new();
                ArrayEncryptedResult = new Ciphertext[numvars];
                //Se desplaza por las distintas filas (variables) de
la matriz
                for (nvar = 0; nvar < numvars; nvar++)
                {
                    //Formulario 1
                    MiFormC = new FormCliente();
                    Application.Run(MiFormC);
                }
```

```
                MiFormRes = new FormResTot();
                //Formulario 4
                if (numvars>1)
                {
                    Application.Run(MiFormRes);
                }

            }
        }

        #region Funciones Form 1 (Servidor)
        //DESCRIPCIÓN: Sirve para declarar los parámetros de
encriptación y para crear las claves de encriptación tanto públicas
como privadas
        //ESTADO: TERMINADO
        //REFERENCIAS: Desde el formulario 1, Cliente desde el botón
de encriptar
        public static SEALContext Inicializar()
        {
            EncryptionParameters parms = new (SchemeType.CKKS);
            ulong polyModulusDegree = 16384;
            parms.PolyModulusDegree = polyModulusDegree;
            parms.CoeffModulus = CoeffModulus.Create(
            polyModulusDegree, new int[] { 60, 50, 50, 50, 50, 50, 60
});
            context = new SEALContext(parms);

            KeyGenerator keygen = new (context);
            SecretKey secretKey = keygen.SecretKey;
            keygen.CreatePublicKey(out  publicKey);
            keygen.CreateRelinKeys(out  relinKeys);

            encryptor = new Encryptor(context, publicKey);
            evaluator = new Evaluator(context);
            decryptor = new Decryptor(context, secretKey);

            //DUDA ENTONCES NO HACE FALTA QUE DEVUELVA EL CONTEXT? -
JPP: NO ES NECESARIO
            return context;
```

```
        }


        //DESCRIPCIÓN: Crea el encoder y la lista en la que almacenar
las X
        //ESTADO: Hay que cambiar el formato lista o bien encontar una
forma de de uso de la lista
        //REFERENCIAS: TODAVÍA NO TIENE REFERENCIAS
        //JPP: yo incluira el unico paso mas arriba y eliminaria esta
funcion
        public static void CrearEncoder()
        {

            encoder = new CKKSEncoder(context);

        }

        //DESCRIPCIÓN: Convierte a plaintext y encodifica la variable
X que ha introducido el cliente
        //ESTADO: Terminada
        //REFERENCIAS: Se llama desde el formulario1, el del cliente
desde el botón de encriptar
        public static Ciphertext CrearPlainTextyEncodeX(string X, int
nvar)
        {
            Plaintext xPlain = new Plaintext();
            encoder.Encode(Convert.ToDouble(X), scale, xPlain);
            //REVISAR
            Ciphertext xEncryptedTemporal=new Ciphertext();
            encryptor.Encrypt(xPlain, xEncryptedTemporal);
            xEncrypted[nvar] = xEncryptedTemporal;

            return xEncrypted[nvar];
        }

        #endregion

        #region Funciones Form 2 (Cliente)
        //DESCRIPCIÓNConvierte a plaintext y  NO encodifica los
coeficientes introducidos por el servidor
```

```
        //ESTADO:TERMINADO
        //REFERENCIAS:Se llama desde el formulario2, el del servidor
desde el botón de enviar
        public static void CrearPlainTextyEncodeCoef(double[]
Dvectorcoef)
        {
            //// OPCIÓN CON VARIABLE LOCAL QUE SOLO USO DENTRO DE CADA
LOOP
            ////Da error por estar declarado a nivel global como
matriz
            ////Habría que declaralo a nivel de función yo creo,
aunque hay que ver como se devuelve a otras funcioens
            //PlainCoefs = new Plaintext[numcoef];
            //for (ncoef = 0; ncoef < numcoef; ncoef++)
            //{
            //    encoder.Encode(Dvectorcoef[ncoef], scale,
PlainCoefs[ncoef]);
            //}

            Coef7 = new Plaintext();
            encoder.Encode(Dvectorcoef[7], scale, Coef7);
            //Este Cipertexto auxiliar se usa porque la función de
Microsoft SEAL solo admite como inputs cipertext sencillos y no acepta
ni vectores ni matrices
            //Plaintext CoefAux = new Plaintext();
            for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                PlainCoefs[nvar,ncoef] = new Plaintext();
                encoder.Encode(Dvectorcoef[ncoef], scale,
PlainCoefs[nvar, ncoef]);

            }
        }

        //DESCRIPCIÓN: Comprueba si los coeficientes introducidos por
el servidor son iguales a 0 o diferentes a fin de reducir el
procesamieto consmido durante la fase de encriptación
        //ESTADO:TERMINADO
        //REFERENCIAS:Se llama desde el formulario2, el del servidor
desde el botón de enviar
```

```
        public static void ComprobarCoeficientesArray(double[] Dvectorcoef)
        {
            //OPCIÓN CON VARIABLE LOCAL QUE SOLO USO DENTRO DE CADA LOOP
            //BCoeficientes = new bool[numcoef];
            //for (ncoef = 0; ncoef < numcoef; ncoef++)
            //{
            //    if (double.Parse(SCoeficientes[ncoef]) != 0)
            //    {
            //        BCoeficientes[ncoef] = true;
            //    }
            //}

            for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                if (Dvectorcoef[ncoef] != 0)
                {
                    BCoeficientes[nvar, ncoef] = true;
                }
            }

        }
        #endregion

        #region Funciones Form 1 (Servidor)

        //DESCRIPCIÓN: Se realizan los cálulos de cada uno de los términos del polinomio utilizando vectores tanto para el bool como para almacear los plaintext
        //ESTADO: TERMINADO, Igual falta el return para poder pasarlo a la función suma, IGUAL SE PUEDE CREAR UN ARRAY DE CHIPHERTEXT PARA QEU QUEDE MÁS LIMPIO
        //FALTA RELLENAR LOS IF SI EL NÚMERO ES IGUAL A 0, PARA QUE NO LO CALCULE TODO
        //REFERENCIAS:FORM 1
        public static Ciphertext CalcularConPlainTextYCiphertextArray(Ciphertext x1Encrypted)
        {
```

```
//ENCRIPTAR 0
Plaintext plaintext = new();
Plaintext Plain0 = plaintext;
encoder.Encode(0, scale, Plain0);
Ciphertext Encripted0 = new ();
encryptor.Encrypt(Plain0, Encripted0);

//X^2
Ciphertext xEncryptedauxX2 = new ();
evaluator.Square(x1Encrypted, xEncryptedauxX2);
evaluator.RelinearizeInplace(xEncryptedauxX2, relinKeys);
evaluator.RescaleToNextInplace(xEncryptedauxX2);


//X^4
Ciphertext xEncryptedauxX4 = new ();
evaluator.Square(xEncryptedauxX2, xEncryptedauxX4);
evaluator.RelinearizeInplace(xEncryptedauxX4, relinKeys);
evaluator.RescaleToNextInplace(xEncryptedauxX4);

//X^8
Ciphertext xEncryptedauxX8 = new ();
evaluator.Square(xEncryptedauxX4, xEncryptedauxX8);

evaluator.RelinearizeInplace(xEncryptedauxX8, relinKeys);
evaluator.RelinearizeInplace(xEncryptedauxX8, relinKeys);
evaluator.RescaleToNextInplace(xEncryptedauxX8);

XCoefsEncriptados = new Ciphertext[numcoef];

for (ncoef = 0; ncoef < numcoef; ncoef++)
{
    XCoefsEncriptados[ncoef] = new();
}

//CASO ESPECIÍFICO X^8
if (BCoeficientes[nvar,8] == true )
{

//Multiplico por el coeficeite
```

```
            ParmsId lastParmsId = xEncryptedauxX8.ParmsId;
            evaluator.ModSwitchToInplace(PlainCoefs[nvar,8],
lastParmsId);
            xEncryptedauxX8.Scale = PlainCoefs[nvar,8].Scale;
            evaluator.MultiplyPlain(xEncryptedauxX8,
PlainCoefs[nvar,8], XCoefsEncriptados[8]);
            evaluator.RelinearizeInplace(XCoefsEncriptados[8],
relinKeys);
            evaluator.RescaleToNextInplace(XCoefsEncriptados[8]);

            }
            if (BCoeficientes[nvar,8] == false) { XCoefsEncriptados[8]
= Encripted0; }

            if (BCoeficientes[nvar, 7] == true)
            {
                //Multiplico X por el coeficiente de X^7
                evaluator.MultiplyPlain(x1Encrypted, PlainCoefs[nvar,
7], XCoefsEncriptados[7]);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);


                //Multiplico Coef7*X por X^2
                evaluator.MultiplyInplace(XCoefsEncriptados[7],
xEncryptedauxX2);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);


                //Multiplico Coef*x^3 *X^4
                evaluator.MultiplyInplace(XCoefsEncriptados[7],
xEncryptedauxX4);
                evaluator.RelinearizeInplace(XCoefsEncriptados[7],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[7]);
            }
```

```
            if (BCoeficientes[nvar, 7] == false) {
XCoefsEncriptados[7] = Encripted0; }

            if (BCoeficientes[nvar,6] == true )
                {

                ParmsId lastParmsId = xEncryptedauxX2.ParmsId;
                evaluator.ModSwitchToInplace(PlainCoefs[nvar,6],
lastParmsId);

                //Multiplico X^2 por el coeficiente de X^6
                evaluator.MultiplyPlain(xEncryptedauxX2,
PlainCoefs[nvar,6], XCoefsEncriptados[6]);
                evaluator.RelinearizeInplace(XCoefsEncriptados[6],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[6]);

                //Multiplico X^4* Coef6*X^2
                evaluator.MultiplyInplace(XCoefsEncriptados[6],
xEncryptedauxX4);
                evaluator.RelinearizeInplace(XCoefsEncriptados[6],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[6]);
            }
                if (BCoeficientes[nvar,6] == false) {
XCoefsEncriptados[6] = Encripted0; }

                if (BCoeficientes[nvar,5] == true )
                {
                    //Multiplico X por ek coeficiente de X5
                    evaluator.MultiplyPlain(x1Encrypted,
PlainCoefs[nvar,5], XCoefsEncriptados[5]);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[5],
relinKeys);

evaluator.RescaleToNextInplace(XCoefsEncriptados[5]);

                    ParmsId lastParmsId = xEncryptedauxX4.ParmsId;
                    evaluator.ModSwitchToInplace(XCoefsEncriptados[5],
lastParmsId);
```

```
                    //Multiplico X^4*Coef5*X
                    evaluator.MultiplyInplace(XCoefsEncriptados[5],
xEncryptedauxX4);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[5],
relinKeys);

evaluator.RescaleToNextInplace(XCoefsEncriptados[5]);
                }
                if (BCoeficientes[nvar, 5] == false ) {
XCoefsEncriptados[5] = Encripted0; }

                if (BCoeficientes[nvar, 4] == true )
                {
                    //Multiplico X^4 por ek coeficiente de X^4, DUDA:
No se si puedo operar esto porque no se si están al mismo nivel
                    ParmsId lastParmsId = xEncryptedauxX4.ParmsId;
                    evaluator.ModSwitchToInplace(PlainCoefs[nvar, 4],
lastParmsId);

                    evaluator.MultiplyPlain(xEncryptedauxX4,
PlainCoefs[nvar, 4], XCoefsEncriptados[4]);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[4],
relinKeys);

evaluator.RescaleToNextInplace(XCoefsEncriptados[4]);
            }
            if (BCoeficientes[nvar, 4] == false ) {
XCoefsEncriptados[4] = Encripted0; }

                if (BCoeficientes[nvar, 3] == true )
                {
                //Multiplico X por ek coeficiente de X^3
                evaluator.MultiplyPlain(x1Encrypted, PlainCoefs[nvar,
3], XCoefsEncriptados[3]);
                evaluator.RelinearizeInplace(XCoefsEncriptados[3],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[3]);

                //Multiplico X^2*Coef3*X
```

```
                evaluator.MultiplyInplace(XCoefsEncriptados[3],
xEncryptedauxX2);
                evaluator.RelinearizeInplace(XCoefsEncriptados[3],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[3]);
                }
                if (BCoeficientes[nvar, 3] == false ) {
XCoefsEncriptados[3] = Encripted0; }

                if (BCoeficientes[nvar, 2] == true )
                {
                    ParmsId lastParmsId = xEncryptedauxX2.ParmsId;
                    evaluator.ModSwitchToInplace(PlainCoefs[nvar, 2],
lastParmsId);

                    //Multiplico X^2 por el coeficiente de X^2
                    evaluator.MultiplyPlain(xEncryptedauxX2,
PlainCoefs[nvar, 2], XCoefsEncriptados[2]);
                    evaluator.RelinearizeInplace(XCoefsEncriptados[2],
relinKeys);

evaluator.RescaleToNextInplace(XCoefsEncriptados[2]);
            }
            if (BCoeficientes[nvar, 2] == false ) {
XCoefsEncriptados[2] = Encripted0; }

                if (BCoeficientes[nvar, 1] == true)
                {
                //Multiplico X por el coeficiente de X
                evaluator.MultiplyPlain(x1Encrypted, PlainCoefs[nvar,
1], XCoefsEncriptados[1]);
                evaluator.RelinearizeInplace(XCoefsEncriptados[1],
relinKeys);
                evaluator.RescaleToNextInplace(XCoefsEncriptados[1]);
            }
            if (BCoeficientes[nvar, 1] == false ) {
XCoefsEncriptados[1] = Encripted0; }

                if (BCoeficientes[nvar, 0] == true )
                {
```

```
                    encryptor.Encrypt(PlainCoefs[nvar, 0],
XCoefsEncriptados[0]);
                }
                if (BCoeficientes[nvar, 0] == false ) {
XCoefsEncriptados[0] = Encripted0; }


            return Sumar();
        }

        //DESCRIPCIÓN: Se suman todos los términos calculados
anteriormente en formato encriptado
        //REFERENCIAS: FORM 1
        public static Ciphertext Sumar()
        {
            ParmsId lastParmsId;
            lastParmsId = XCoefsEncriptados[0].ParmsId;
            // buscamos los parametros del termino mas alto distinto
de 0
            for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                if (BCoeficientes[nvar, ncoef])
                {
                    lastParmsId = XCoefsEncriptados[ncoef].ParmsId;
                }
            }

            //asignamos los parametros del termino mas alto a todos
los terminos
            for (ncoef = 0; ncoef < numcoef; ncoef++)
            {
                evaluator.ModSwitchToInplace(XCoefsEncriptados[ncoef],
lastParmsId);
            }

            // SUMAMOS LOS TÉRMINOS DEL POLINOMIO

            lastParmsId = XCoefsEncriptados[7].ParmsId;
            evaluator.ModSwitchToInplace(XCoefsEncriptados[8],
lastParmsId);
```

```
            XCoefsEncriptados[8].Scale = XCoefsEncriptados[7].Scale;

            //Estos Cipertextos auxiliares se usan porque la función
de Microsoft SEAL solo admite como inputs cipertext sencillos y no
acepta ni vectores ni matrices

            Ciphertext EncryptedResult = new Ciphertext();

            evaluator.Add(XCoefsEncriptados[8], XCoefsEncriptados[7],
EncryptedResult);


            for (int i = 6; i >= 0; i--)
            {
                lastParmsId = XCoefsEncriptados[i].ParmsId;
                evaluator.ModSwitchToInplace(EncryptedResult,
lastParmsId);
                EncryptedResult.Scale = XCoefsEncriptados[i].Scale;
                //DUDA AQUÍ NO SE QUEJA DE QUE NO FUNCIONA, DEBERÍA
PONER LOS AUXILIARES, igual no llega a entrar en el loop?
                evaluator.Add(EncryptedResult, XCoefsEncriptados[i],
EncryptedResult);
            }

            return EncryptedResult;
        }
        //public static List<double> DesencriptarResultado(Ciphertext
encryptedResult)
        public static List<double> DesencriptarResultado(Ciphertext
encryptedResult)
        {
            Plaintext plainResult = new Plaintext();

            decryptor.Decrypt(encryptedResult, plainResult);
            //List<double> result = new List<double>();
            //encoder.Decode(plainResult, result);
            result = new List<double>();
            encoder.Decode(plainResult, result);

            return result;
```

```
        }

        #endregion

        #region Form 4 (Cliente resultado)

        //DESCRIPCIÓN: Se realizan las sumas de todas las variables
encriptadas pero haciendo la mini trampa de desencriptarlas antes y
volver a encriptarlas,
        //ESTADO:
        //REFERENCIAS:
        //PROBAR A SUMAR SIN DESENCRIPTAR
        public static Ciphertext SumarTotal()

        {
            Plaintext[] PlainResultencripted = new Plaintext[numvars];
            Ciphertext[] CiphResEncrypted = new Ciphertext[numvars];

            for (nvar=0; nvar<numvars; nvar++)
            {

                Plaintext PlainEncryptedTemporal = new Plaintext();
                Double ResAux = result[nvar];


                encoder.Encode(ResAux, scale, PlainEncryptedTemporal);
                PlainResultencripted[nvar] = PlainEncryptedTemporal;

                Ciphertext CiphEncryptedTemporal = new Ciphertext();

                encryptor.Encrypt(PlainEncryptedTemporal,
CiphEncryptedTemporal);
                CiphResEncrypted[nvar] = CiphEncryptedTemporal;

            }

            ChiperResultado_total = new Ciphertext();
            for (nvar = 1; nvar < numvars; nvar++)
            {
```

```
                //Estos Cipertextos auxiliares se usan porque la
función de Microsoft SEAL Add solo admite como inputs cipertext
sencillos y no acepta ni vectores ni matrices

                Ciphertext Auxiliar1 = CiphResEncrypted[nvar - 1];
                Ciphertext Auxiliar2 = CiphResEncrypted[nvar];

                evaluator.Add(Auxiliar1, Auxiliar2,
ChiperResultado_total);
            }

            return ChiperResultado_total;
        }


        public static List<double>
DesencriptarResultadoTotal(Ciphertext ChiperResultado_total)
        {

            Plaintext plainresultado_total = new();

            decryptor.Decrypt(ChiperResultado_total,
plainresultado_total);
            List<double> resultado_total = new List<double>();
            encoder.Decode(plainresultado_total, resultado_total);

            return resultado_total;
        }

        #endregion
    }


}
```

FORM 1:

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
```

```
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.IO;

namespace Formulario_TFG
{
    public partial class FormCliente : Form
    {
        static Ciphertext x1Encrypted;

        public FormCliente()
        {
            InitializeComponent();
        }
        private void Form1_Load(object sender, EventArgs e)
        {
            label2.Text = $"Introduzca el valor de la variable
{Program.NombreVar[Program.nvar]}: ";
        }

        #region Botones
        private void buttonEncriptar_Click(object sender, EventArgs e)
        {
            Cursor.Current =Cursors.WaitCursor;
            x1Encrypted =
Program.CrearPlainTextyEncodeX(textBoxX.Text, Program.nvar);

            string Sx1Encrypted;
            using (var ms = new MemoryStream())
            {
                x1Encrypted.Save(ms);
                Sx1Encrypted = Convert.ToBase64String(ms.ToArray());
            }
```

```
            textBoxXencriptado.Text = Sx1Encrypted.Substring(1,
32000);
            Cursor.Current = Cursors.Default;
        }
        private void buttonCalcular_Click(object sender, EventArgs e)
        {
            Cursor.Current = Cursors.WaitCursor;
            string SResEncrypted;
            Program.ArrayEncryptedResult[Program.nvar] =
Program.CalcularConPlainTextYCiphertextArray(x1Encrypted);

            using (var ms = new MemoryStream())
            {
                Program.ArrayEncryptedResult[Program.nvar].Save(ms);
                SResEncrypted = Convert.ToBase64String(ms.ToArray());
            }

            textBoxOutputEncriptado.Text = SResEncrypted.Substring(1,
32000);
            Cursor.Current = Cursors.Default;
        }
        private void buttonDesencriptar_Click(object sender, EventArgs
e)
        {
            Cursor.Current = Cursors.WaitCursor;
            //FUCIÓN DESENCRIPTAR RESULTADO
            List<double> ResDesencriptado;

            ResDesencriptado =
Program.DesencriptarResultado(Program.ArrayEncryptedResult[Program.nva
r]);

            Program.SumandosDes.Add(ResDesencriptado[0]);
            textBoxResDesencipt.Text = ResDesencriptado[0].ToString();
            Cursor.Current = Cursors.Default;
        }
        #endregion

        private void Botro_Click(object sender, EventArgs e)
        {
```

```
            Program.Bmore = true;
            FormCliente.ActiveForm.Close();
        }

        private void bFin_Click(object sender, EventArgs e)
        {
            Program.Bmore = false;
            FormCliente.ActiveForm.Close();
        }
    }
}
```

FORM 2:

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.IO;

namespace Formulario_TFG
{
    public partial class FormServidor : Form
    {
        public FormServidor()
        {
            InitializeComponent();
        }

        private void buttonEnvCoef_Click(object sender, EventArgs e)
        {
            //FORMA CON VECTORES PARA QUE PUEDA SER VARIABLE
            //a gracia está en que sea un número que me da el usuario
```

```
            //No se si esto lo debería meter en una función en el main
mejor
            try
            {
                Program.NombreVar[Program.nvar]
=textBoxNombreVar.Text;
                int n = 9;

                double[] Dvectorcoef = new double[n];


                Dvectorcoef[0] = Convert.ToDouble(txBTermInd.Text);
                Dvectorcoef[1] = Convert.ToDouble(txBCoefX.Text);
                Dvectorcoef[2] = Convert.ToDouble(txBCoefX2.Text);
                Dvectorcoef[3] = Convert.ToDouble(txBCoefX3.Text);
                Dvectorcoef[4] = Convert.ToDouble(txBCoefX4.Text);
                Dvectorcoef[5] = Convert.ToDouble(txBCoefX5.Text);
                Dvectorcoef[6] = Convert.ToDouble(txBCoefX6.Text);
                Dvectorcoef[7] = Convert.ToDouble(txBCoefX7.Text);
                Dvectorcoef[8] = Convert.ToDouble(txBCoefX8.Text);

                //Esta función determina si el coeficiente introducido
es 0 o distinto de 0, se utilizará para reducir procesamiento de
cálculos a la hora de encriptar
                Program.ComprobarCoeficientesArray(Dvectorcoef);

                //Esta función convierte los strings en plaintext,
pero me da un error en el función que está explicado dentro de la
propia función
                Program.CrearPlainTextyEncodeCoef(Dvectorcoef);
                this.Close();

            }
            catch
            {
                MessageBox.Show($"Error, algún valor itroducido como
coeficiente no es válido, por favor, introduzca un número.");
            }
```

```
        }
    }
}
```

FORM 3:

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.IO;

//SE supone que debería estar en
//namespace Formulario_TFG
namespace Formulario_TFG
{
    public partial class FormSIniz : System.Windows.Forms.Form
    {
        public FormSIniz()
        {
            InitializeComponent();
        }

        private void buttonEnviarNVars_Click(object sender, EventArgs
e)
        {
            int res;
            bool EsTexto= int.TryParse(textBoxNVars.Text,out res);

            if (EsTexto == true)
            {
                Program.numvars = int.Parse(textBoxNVars.Text);
            }
            else {
```

```
                MessageBox.Show("El valor introducido no es un número
entero, por favor introduzca un valor válido.");
            }
            this.Close();

        }
    }
}
```

FORM 4:

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.IO;

namespace Formulario_TFG
{
    public partial class FormResTot : System.Windows.Forms.Form
    {
        public FormResTot()
        {
            InitializeComponent();
            string auxiliar="";

            for (Program.nvar = 0; Program.nvar < Program.numvars;
Program.nvar++)
            {
                auxiliar = auxiliar + $"La suma de los términos de la
variable " + Convert.ToString(Program.nvar + 1) + $" llamada
```

```
{Program.NombreVar[Program.nvar]}" + $" es:
{Program.SumandosDes[Program.nvar]}." + Environment.NewLine;
            }
            textBoxCadaVarDes.Text = auxiliar;
        }

        private void buttonSumaTotalEncriptada_Click(object sender,
EventArgs e)
        {
            //REVISAR ESTE BOTÓN PARA VER SI FUNCIONA
            Cursor.Current = Cursors.WaitCursor;
            string SSumTotEncript;
            Ciphertext SumTotEncript;
            SumTotEncript = Program.SumarTotal();

            using (var ms = new MemoryStream())
            {
                SumTotEncript.Save(ms);
                SSumTotEncript = Convert.ToBase64String(ms.ToArray());
            }

            textBoxSumaTotalEncriptado.Text =
SSumTotEncript.Substring(1, 32000);
            Cursor.Current = Cursors.Default;
        }

        private void buttonSumaTotalDesencriptar_Click(object sender,
EventArgs e)
        {
            Cursor.Current = Cursors.WaitCursor;
            //FUCIÓN DESENCRIPTAR RESULTADO
            List<double> ResDesencriptado;
            //REVISAR
            ResDesencriptado =
Program.DesencriptarResultadoTotal(Program.ChiperResultado_total);

            Program.SumandosDes.Add(ResDesencriptado[0]);
            textBoxSumaTotalDesencriptado.Text =
ResDesencriptado[0].ToString();
            Cursor.Current = Cursors.Default;
```

```
            //textBoxSumaTotalDesencriptado.Text =
Program.DesencriptarResultadoTotal(ChiperResultado_total);
        }

        private void buttonOtro_Click(object sender, EventArgs e)
        {
            Program.Bmore = true;
            FormCliente.ActiveForm.Close();
        }

        private void buttonFin_Click(object sender, EventArgs e)
        {
            Program.Bmore = false;
            FormCliente.ActiveForm.Close();
        }
    }
}
```

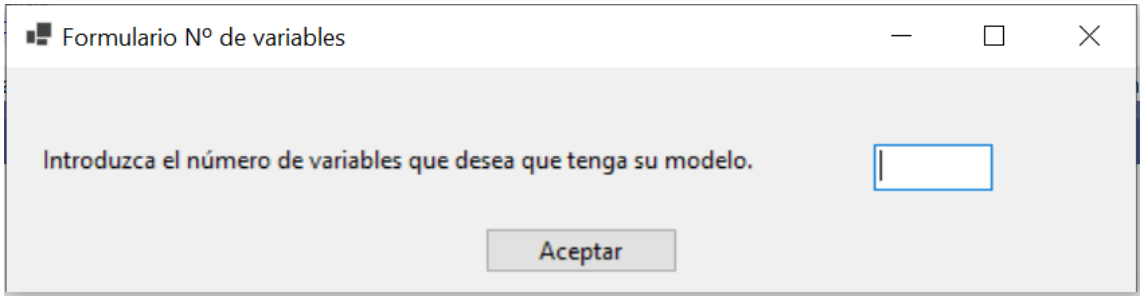


Ilustración 31:Formulario número de variables (elaboración propia)

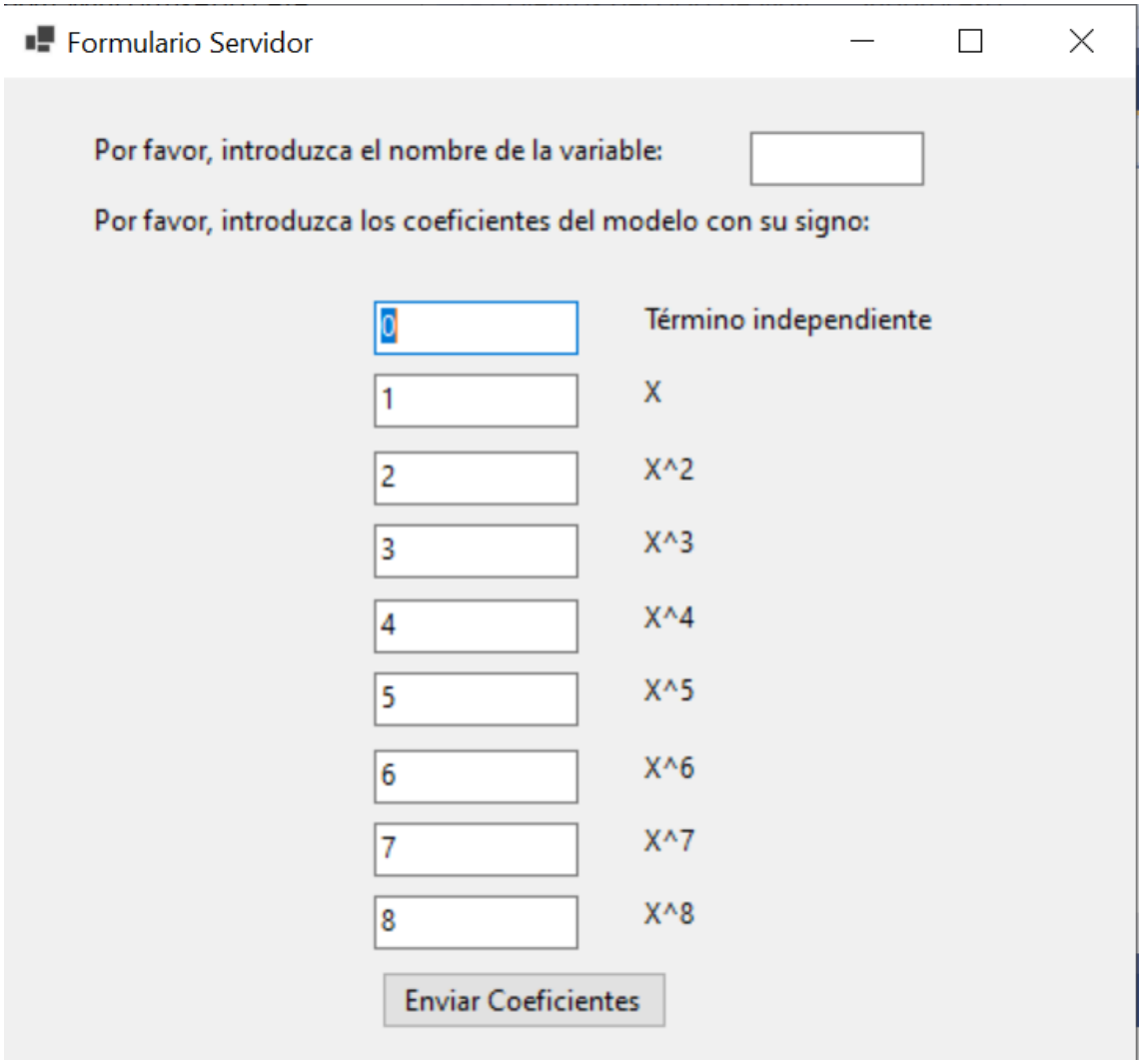


Ilustración 32: Formulario obtención de los parámetros de las variables del modelo (elaboración propia)

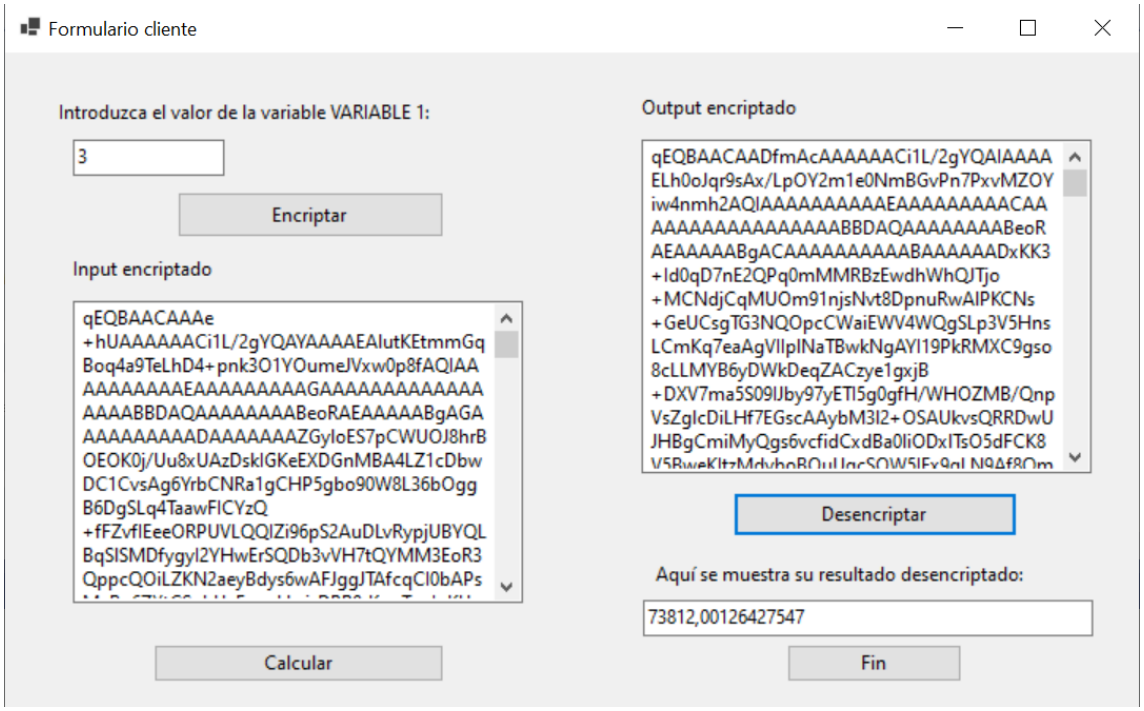


Ilustración 33:introducción de las variables del modelo y obtención de resultados (elaboración propia)

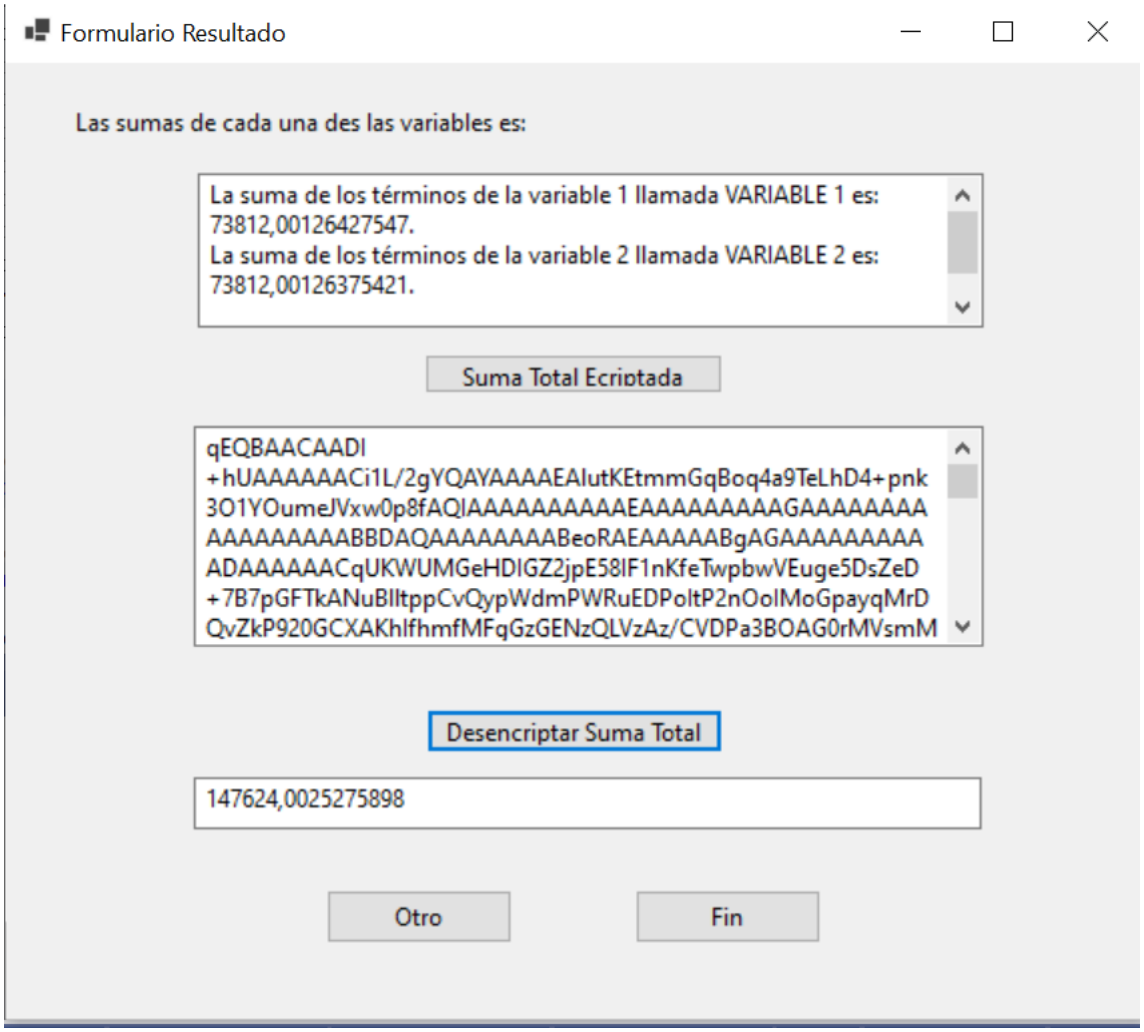


Ilustración 34: Obtención del resultado total (elaboración propia)

PROGRAMA DE COMUNICACIÓN

SERVIDOR:

```
using Microsoft.AspNetCore.Hosting;
using Microsoft.Extensions.Configuration;
using Microsoft.Extensions.Hosting;
using Microsoft.Extensions.Logging;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Threading.Tasks;
using Microsoft.Research.SEAL;

namespace WebApplication5
```

```
{
    public class Program
    {

        public static SEALContext _sealContext;
        public static double scale;
        public static EncryptionParameters _params;
        public static PublicKey _publickey;

        public static Ciphertext sumandoEncriptado;
        public static Ciphertext sumandoClienteEncriptado;
        public static Ciphertext suma;

        public static bool bIni;
        public static bool bSum;
        public static bool bPK;

        public static void Main(string[] args)
        {
            Program._params = new EncryptionParameters(SchemeType.CKKS);
            ulong polyModulusDegree = 8192;
            Program._params.PolyModulusDegree = polyModulusDegree;
            Program._params.CoeffModulus = CoeffModulus.Create(polyModulusDegree,
new int[] { 60, 40, 40, 60 });
            Program.scale = Math.Pow(2.0, 40);

            Program._sealContext = new SEALContext(Program._params);
            bIni = true;
            bSum = false;
            bPK = false;

            CreateHostBuilder(args).Build().Run();
        }

        public static IHostBuilder CreateHostBuilder(string[] args) =>
            Host.CreateDefaultBuilder(args)
                .ConfigureWebHostDefaults(webBuilder =>
                {
                    webBuilder.UseStartup<Startup>();
                });
    }
```

}

VALUES CONTROLLER SERVIDOR:

```
using Microsoft.AspNetCore.Mvc;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Threading.Tasks;
using Microsoft.Research.SEAL;
using System.IO;

// For more information on enabling Web API for empty projects, visit
https://go.microsoft.com/fwlink/?LinkID=397860

namespace WebApplication5
{
    [Route("api/[controller]")]
    [ApiController]
    public class ValuesController : ControllerBase
    {
        public class ResModel
        {
            public string Res { get; set; }
            public string S1 { get; set; }
            public string S2 { get; set; }

        }

        public ValuesController()
        {

        }

        [HttpPost]
        [Route("sumando")]
        public void GuardarSumando([FromBody] string request)
        {
            var payload = Convert.FromBase64String(request);
```

```
            Program.sumandoClienteEncriptado = new Ciphertext();
            using (var ms = new MemoryStream(payload))
            {

Program.sumandoClienteEncriptado.Load(Program._sealContext, ms);
            }
            Program.bSum = true;
        }

        [HttpPost]
        [Route("guardarsuma")]
        public void GuardarSuma([FromBody] string request)
        {
            var payload = Convert.FromBase64String(request);

            Program.suma = new Ciphertext();
            using (var ms = new MemoryStream(payload))
            {
                Program.suma.Load(Program._sealContext, ms);
            }
        }

        [HttpPost]
        [Route("clave")]
        public void GuardarClave([FromBody] string request)
        {
            var payload = Convert.FromBase64String(request);

            Program._publickey = new PublicKey();
            using (var ms = new MemoryStream(payload))
            {
                Program._publickey.Load(Program._sealContext, ms);
            }
            Program.bPK = true;
        }


        [HttpGet]
        [Route("Getsuma")]
```

```
public ActionResult<ResModel> GetSumando()
{
    ResModel RM;
    RM = new();

    MemoryStream mso = new();
    Program.suma.Save(mso);
    RM.Res = Convert.ToBase64String(mso.ToArray());

    return RM;
}

[HttpGet]
[Route("comprobar")]
public ActionResult<ResModel> Comprobar()
{
    ResModel RM;
    RM = new();

    if (Program.bIni && Program.bSum && Program.bPK)
    {
        RM.Res = "true";
    }
    else
    {
        RM.Res = "false";
    }

    return RM;
}


[HttpGet]
[Route("params")]
public ActionResult<ResModel> GetParams()
{
    ResModel RM;
    RM = new();

    MemoryStream ms = new();
```

```
            Program._params.Save(ms);
            RM.Res= Convert.ToBase64String(ms.ToArray());
            return RM;
        }

        [HttpGet]
        [Route("paramsP")]
        public ActionResult<ResModel> GetParamsP()
        {
            ResModel RM;
            RM = new();

            MemoryStream ms = new();
            Program._params.Save(ms);
            RM.Res = Convert.ToBase64String(ms.ToArray());

            MemoryStream ms2 = new();
            Program._publickey.Save(ms2);
            RM.S1 = Convert.ToBase64String(ms2.ToArray());

            MemoryStream ms3 = new();
            Program.sumandoClienteEncriptado.Save(ms3);
            RM.S2 = Convert.ToBase64String(ms3.ToArray());

            return RM;
        }

    }
}
```

FORM PROVEEDOR:

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.Net.Http;
```

```
using Newtonsoft.Json;
using System.Text;
using System.IO;

namespace FormProveedor
{
    public class ResModel
    {
        public string Res { get; set; }
        public string S1 { get; set; }
        public string S2 { get; set; }

    }

    class Program
    {
        /// <summary>
        ///  The main entry point for the application.
        /// </summary>
        ///

        private static Encryptor _encryptor;
        private static Decryptor _decryptor;
        private static SEALContext _context;

        private static KeyGenerator _keyGenerator;
        private static CKKSEncoder _encoder;
        private static SecretKey _secretKey;
        private static PublicKey _publicKey;
        private static double scale;
        private static EncryptionParameters _params;

        private static HttpClient _client;
        private static string BaseUri;

        public static FormCli _FormP;

        [STAThread]
        static void Main()
        {
```

```
            Application.SetHighDpiMode(HighDpiMode.SystemAware);
            Application.EnableVisualStyles();
            Application.SetCompatibleTextRenderingDefault(false);

            _client = new();
            BaseUri = "http://localhost:63231/api";
            _FormP = new();
            Application.Run(_FormP);
        }

        static public async void GetParams()
        {
            using (var request = new
HttpRequestMessage(HttpMethod.Get, $"{BaseUri}/values/params"))
            {
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                _params = new EncryptionParameters();
                using (var ms = new MemoryStream(payload))
                {
                    _params.Load(ms);
                }

                _context = new SEALContext(_params);

                scale = Math.Pow(2.0, 40);
                _keyGenerator = new KeyGenerator(Program._context);
                _secretKey = Program._keyGenerator.SecretKey;
                _keyGenerator.CreatePublicKey(out Program._publicKey);
                _encryptor = new Encryptor(_context, _publicKey);
                _encoder = new CKKSEncoder(_context);
                _decryptor = new Decryptor(_context, _secretKey);

            }
        }
```

```
        static public async void EnviarSumando(string sumando)
        {

            Plaintext plaintext;
            plaintext = new Plaintext();
            _encoder.Encode(Convert.ToDouble(sumando), scale,
plaintext);
            Ciphertext ciphertextSumando;
            ciphertextSumando = new Ciphertext();

            _encryptor.Encrypt(plaintext, ciphertextSumando);

            string sumandoencriptado;
            MemoryStream ms;
            ms = new();
            ciphertextSumando.Save(ms);
            sumandoencriptado = Convert.ToBase64String(ms.ToArray());

            string sumandoRequestAsJsonStr =
JsonConvert.SerializeObject(sumandoencriptado);
            using (var request = new
HttpRequestMessage(HttpMethod.Post, $"{BaseUri}/values/sumando"))
            using (var content = new
StringContent(sumandoRequestAsJsonStr, Encoding.UTF8,
"application/json"))
            {
                request.Content = content;
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
            }

        }

        static public async void EnviarClave()
        {
            string clave;
            MemoryStream ms;
            ms = new();
            _publicKey.Save(ms);
```

```
            clave = Convert.ToBase64String(ms.ToArray());

            string sumandoRequestAsJsonStr =
JsonConvert.SerializeObject(clave);
            using (var request = new
HttpRequestMessage(HttpMethod.Post, $"{BaseUri}/values/clave"))
            using (var content = new
StringContent(sumandoRequestAsJsonStr, Encoding.UTF8,
"application/json"))
            {
                request.Content = content;
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
            }

        }

        static public async void RecibirSuma()
        {
            using (var request = new
HttpRequestMessage(HttpMethod.Get, $"{BaseUri}/values/Getsuma"))
            {
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();

                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                Ciphertext resultado;
                resultado = new Ciphertext();
                using (var mso = new MemoryStream(payload))
                {
                    resultado.Load(_context, mso);
                }

                Plaintext resultadoDesencriptado;
                resultadoDesencriptado = new();
                _decryptor.Decrypt(resultado, resultadoDesencriptado);
```

```
                List<double> dResultado = new List<double>(1);
                _encoder.Decode(resultadoDesencriptado, dResultado);

                _FormP.ImprimirResultado(dResultado[0].ToString());

            }

        }


    }


}
```

FORM CLIENTE:

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.Research.SEAL;
using System.Net.Http;
using Newtonsoft.Json;
using System.Text;
using System.IO;

namespace WinFormsApp1
{

    public class ResModel
    {
        public string Res { get; set; }
        public string S1 { get; set; }
        public string S2 { get; set; }

    }
    class Program
    {
        /// <summary>
```

```
        ///  The main entry point for the application.
        /// </summary>
        ///

        private static Encryptor _encryptor;
        private static SEALContext _context;
        private static CKKSEncoder _encoder;
        private static PublicKey _publicKey;
        private static double scale;
        private static EncryptionParameters _params;
        public static Evaluator _evaluator;

        private static HttpClient _client;
        private static string BaseUri;

        public static Ciphertext sumandoCliente;
        public static bool bCheck;
        public static FormPro _FormC;
        public static Ciphertext suma;

        [STAThread]
        static void Main()
        {
            Application.SetHighDpiMode(HighDpiMode.SystemAware);
            Application.EnableVisualStyles();
            Application.SetCompatibleTextRenderingDefault(false);

            _client = new();
            BaseUri = "http://localhost:63231/api";
            _FormC = new();
            Application.Run(_FormC);
        }

        static public async void GetParamsP()
        {
            using (var request = new
HttpRequestMessage(HttpMethod.Get, $"{BaseUri}/values/paramsP"))
            {
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
```

```
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                var payload = Convert.FromBase64String(RM.Res);
                _params = new EncryptionParameters();
                using (var ms = new MemoryStream(payload))
                {
                    _params.Load(ms);
                }

                _context = new SEALContext(_params);
                scale = Math.Pow(2.0, 40);

                var payload2 = Convert.FromBase64String(RM.S1);
                _publicKey = new PublicKey();
                using (var ms2 = new MemoryStream(payload2))
                {
                    _publicKey.Load(_context, ms2);
                }

                _encryptor = new Encryptor(_context, _publicKey);
                _encoder = new CKKSEncoder(_context);
                _evaluator = new Evaluator(_context);

                var payload3 = Convert.FromBase64String(RM.S2);
                sumandoCliente = new Ciphertext();
                using (var ms3 = new MemoryStream(payload3))
                {
                    sumandoCliente.Load(_context, ms3);
                }
            }
        }

        static public async void Comprobar()
        {
            using (var request = new
HttpRequestMessage(HttpMethod.Get, $"{BaseUri}/values/comprobar"))
            {
                var response = await _client.SendAsync(request);
```

```
                response.EnsureSuccessStatusCode();
                ResModel RM = new();
                RM = JsonConvert.DeserializeObject<ResModel>(await
response.Content.ReadAsStringAsync());

                if (RM.Res == "true")
                {
                    bCheck = true;
                }
            }
        }

        static public async void Sumar(string sumando)
        {
            Plaintext plaintext;
            plaintext = new Plaintext();
            _encoder.Encode(Convert.ToDouble(sumando), scale,
plaintext);
            Ciphertext ciphertextSumando;
            ciphertextSumando = new Ciphertext();
            _encryptor.Encrypt(plaintext, ciphertextSumando);

            suma = new();
            _evaluator.Add(ciphertextSumando, sumandoCliente, suma);

        }

        static public async void EnviarSuma()
        {

            string sumaencriptada;
            MemoryStream ms;
            ms = new();
            suma.Save(ms);
            sumaencriptada = Convert.ToBase64String(ms.ToArray());

            string sumaRequestAsJsonStr =
JsonConvert.SerializeObject(sumaencriptada);
            using (var request = new
HttpRequestMessage(HttpMethod.Post, $"{BaseUri}/values/guardarsuma"))
```

```
            using (var content = new
StringContent(sumaRequestAsJsonStr, Encoding.UTF8,
"application/json"))
            {
                request.Content = content;
                var response = await _client.SendAsync(request);
                response.EnsureSuccessStatusCode();
            }

        }

    }
}
```

ANEXO 5: MODELO ECONÓMICO COMPLETO

En este anexo se presenta de forma completa el análisis económico realizado en Excel que al igual que el código constituye una parte imprescindible del proyecto y que refleja gran parte del trabajo realizado. En el documento en si se mencionan las partes más relevantes y se eliminan las partes repetitivas a fin de facilitar la lectura y comprensión de este.

La cuenta de pérdidas y ganancias se ha realizado en Excel y podrá descargarse a través de este enlace: Modelo de negocio TFG

Bibliografía

En esta sección se describen las fuentes consultadas y utilizadas para la redacción de este trabajo.

Álvarez, G. (2020). What Differential Privacy Is and Why Google and Apple Are Using It with Your Data. *Think Big. Telefónica Tech. Cybersecurity*. Obtenido de https://business.blogthinkbig.com/differential-privacy-google-apple-using-it-with-your-data/

Andrews, G. (2021). What Is Synthetic Data? *Nvidia*. Obtenido de https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/

Apple Differential Privacy Team. (2017). Learning with Privacy at Scale. *Apple. Machine Learning Research*. Obtenido de https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Apple Differential Privacy Team. (s.f.). *Learning with Privacy at Scale.* Obtenido de https://docs-assets.developer.apple.com/ml-research/papers/learning-with-privacy-at-scale.pdf

AWS. (s.f.). ¿Qué es la informática en la nube? Obtenido de https://aws.amazon.com/es/what-is-cloud-computing/

AWS. (s.f.). AWS Data Exchange. Obtenido de https://aws.amazon.com/es/data-exchange/

AWS. (s.f.). Informática en la nube con AWS. Obtenido de https://aws.amazon.com/es/what-is-aws/

Azure Microsoft. (s.f.). ¿Qué es la nube? Obtenido de https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-the-cloud/

BBVA. (2018). *Cómo los datos cambiarán el mundo, ¡una vez más!* Obtenido de https://www.bbva.com/es/datos-cambiaran-mundo-vez-mas/

Brubaker, M., & Prince, S. (2021). Tutorial #12: Differential Privacy I: Introduction. *BorealisAI*. Obtenido de https://www.borealisai.com/en/blog/tutorial-12-differential-privacy-i-introduction/

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). Notes from the AI frontier: Modeling the impact of AI on the world economy. *McKinsey & Company*. Obtenido de

https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-AI-frontier-modeling-the-impact-of-ai-on-the-world-economy

Cloudflare. (s.f.). ¿Qué es la nube? | Definición de nube. Obtenido de https://www.cloudflare.com/es-es/learning/cloud/what-is-the-cloud/

Cordeiro, B. (2021). How to anonymize data with Presidio. Microsoft's SDK for Data Protection and Anonymization. *Medium. Bravo Lab.* . Obtenido de https://medium.com/bravo-lab/how-to-anonymize-data-with-presidio-44ec40aeb611

Data anonymization. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Data_anonymization

DataCouncil. (s.f.). Obtenido de https://www.datacouncil.ai/

Devaux, E., & Statice. (2020). Data Protection Techniques Needed to Guarantee Privacy. *KDnuggets*. Obtenido de https://www.kdnuggets.com/2020/10/data-protection-techniques-guarantee-privacy.html

Differential privacy. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Differential_privacy#:~:text=Differential%20privacy%20(DP)%20is%20a,about%20individuals%20in%20the%20dataset

Dilmegani, C. (2022). Differential Privacy: How it works, benefits & use cases. *AI Multiple*. Obtenido de https://research.aimultiple.com/differential-privacy/

Dilmegani, C. (2022). What is Synthetic Data? What are its Use Cases & Benefits? *AI Multiple*. Obtenido de https://research.aimultiple.com/synthetic-data/

dir&ge. (2021). *La inversión en análisis de datos crecerá un 20% en los próximos cinco años.* Obtenido de https://directivosygerentes.es/innovacion/inversion-analisis-datos-crecera-20-por-ciento-proximos-cinco-anos

Disponibilidad. (s.f.). *Wikipedia*. Obtenido de https://es.wikipedia.org/wiki/Disponibilidad

Dong, Y., & Scoullos, E. (2022). Generating Synthetic Data with Transformers: A Solution for Enterprise Data Challenges. *Nvidia. Developer*. Obtenido de https://developer.nvidia.com/blog/generating-synthetic-data-with-transformers-a-solution-for-enterprise-data-challenges/

Douglas, W. (2022). Synthetic data for AI. *MIT Technology Review*. Obtenido de https://www.technologyreview.com/2022/02/23/1044965/ai-synthetic-data-2/

Dowli, N., Gilad-Bachrach, R., Laine, K., Naehrig, M., & Wernsing, J. (2016). *CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy.* Microsoft. Obtenido de https://www.microsoft.com/en-us/research/publication/cryptonets-applying-neural-networks-to-encrypted-data-with-high-throughput-and-accuracy/

Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3-4). Obtenido de https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf

Encryption. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Encryption

Enigma machine. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Enigma_machine

Fiabilidad de sistemas. (s.f.). *Wikipedia*. Obtenido de https://es.wikipedia.org/wiki/Fiabilidad_de_sistemas

Framingham, M. (2019). IDC Survey Finds Artificial Intelligence to be a Priority for Organizations But Few Have Implemented an Enterprise-Wide Strategy. *businesswire*. Obtenido de https://www.businesswire.com/news/home/20190708005039/en/IDC-Survey-Finds-Artificial-Intelligence-Priority-Organizations

Frankenfield, J. (2021). Data anonymization. *Investopedia*. Obtenido de https://www.investopedia.com/terms/d/data-anonymization.asp

Gad, A. (2020). Breaking Privacy in Federated Learning. *KDnuggets*. Obtenido de https://www.kdnuggets.com/2020/08/breaking-privacy-federated-learning.html

Georgian. (2018). A Brief Introduction to Differential Privacy. *Georgian Impact Blog*. Obtenido de https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b

Gimenez, M. (2020). Amazon Web Services (AWS): ¿qué es y qué ofrece? *hiberus blog*. Obtenido de https://www.hiberus.com/crecemos-contigo/amazon-web-services-aws-que-es-y-que-ofrece/

GitHub. (s.f.). Obtenido de https://github.com/microsoft/SEAL-Demo

Google. (2021). ¿Qué es la nube y cómo funciona? *uCloudStore*. Obtenido de https://www.ucloudstore.com/blog/que-es-la-nube-y-como-funciona/

Google. (s.f.). HOW GOOGLE ANONYMIZES DATA. Obtenido de https://policies.google.com/technologies/anonymization?hl=en-US

Greenberg, A. (2017). How One of Apple's Key Privacy Safeguards Falls Short. *Wired*. Obtenido de https://www.wired.com/story/apple-differential-privacy-shortcomings/

Guevara, M. (2021). How we're helping developers with differential privacy. *Google Developers*. Obtenido de https://developers.googleblog.com/2021/01/how-were-helping-developers-with-differential-privacy.html

Harvard University. (s.f.). Differential Privacy. *Harvard University Privacy Tools Project*. Obtenido de https://privacytools.seas.harvard.edu/differential-privacy

Historiadelaempresa.com. (s.f.). ¿Qué son los datos agregados? (Más 6 ejemplos). Obtenido de https://historiadelaempresa.com/datos-agregados

Homomorphic encryption. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Homomorphic_encryption

Homomorphic Encryption Standardization. (2019). HomomorphicEncryption.org Standards Meeting. Obtenido de http://homomorphicencryption.org/aug-17-2019-homomorphicencryption-org-standards-meeting/

IBM. (s.f.). Fully Homomorphic Ecryption. How to achieve data privacy by design. Obtenido de https://research.ibm.com/labs/uk/fhe.html?_ga=2.228307933.1971778217.1603398947-1997620092.1602871268

IBM Security. (2020). *IBM Security Homomorphic Ecryption Services.* Obtenido de https://www.ibm.com/downloads/cas/KQ27PWBO

IBM. (s.f.). Unluck value of sensitive data without decryption. Obtenido de https://research.ibm.com/labs/uk/fhe.html?_ga=2.228307933.1971778217.1603398947-1997620092.1602871268

IMMUTA. (s.f.). *How to Enhance Privacy in Data Science.* Obtenido de https://www.kdnuggets.com/2019/08/immuta-ebook-privacy-data-science.html

Información agregada. (s.f.). *hmong*. Obtenido de https://hmong.es/wiki/Aggregate_data

Jefatura del Estado. (2018). Legislación consolidada. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. *BOE*. Obtenido de https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

Jones, T. (2019). Privacy and Data Science: Protecting Sensitive Data in the Age of Aalytics. *IQT*. Obtenido de https://medium.com/in-q-tel/privacy-and-data-science-protecting-sensitive-data-in-the-age-of-analytics-16791e6cd56e

Jones, T. (2019). Privacy and Data Science: Protecting Sensitive Data in the Age of Analytics. *Medium. IQT*. Obtenido de https://medium.com/in-q-tel/privacy-and-data-science-protecting-sensitive-data-in-the-age-of-analytics-16791e6cd56e

Kahan, J. (2020). How differential privacy enhaces Microsoft´s privacy and security tools: SmartNoise Early Adopter Acceleration Program Launched. *Microsoft*. Obtenido de https://blogs.microsoft.com/on-the-issues/2020/12/10/differential-privacy-smartnoise-early-adopter-acceleration-program/

Kahan, J. (2020). New differential privacy platform co-developed with Harvard´s OpenDP unlocks data while safeguarding privacy. *Microsoft*. Obtenido de https://blogs.microsoft.com/on-the-issues/2020/06/24/differential-privacy-harvard-opendp/

Klucar, J. (s.f.). Privacy Techniques for Data Science. *Data Council*. Obtenido de https://www.datacouncil.ai/talks/privacy-techniques-for-data-science

Kopp, A. (2021). Create privacy-preserving synthetic data for machine learning with SmartNoise. *Microsoft. Open Source Blog*. Obtenido de https://cloudblogs.microsoft.com/opensource/2021/02/18/create-privacy-preserving-synthetic-data-for-machine-learning-with-smartnoise/

Kopp, A. (2021). *Microsoft SmartNoise Differential Privacy Machine Learning Case Studies.* Microsoft. Obtenido de https://azure.microsoft.com/mediahandler/files/resourcefiles/microsoft-smartnoisedifferential-privacy-machine-learning-case-studies/SmartNoise%20Whitepaper%20Final%203.8.21.pdf

Laskowski, N. (2018). Definition synthetic data. *TechTarget. Search CIO*. Obtenido de https://www.techtarget.com/searchcio/definition/synthetic-

data#:~:text=Synthetic%20data%20is%20information%20that's,to%20train%20machine%20learning%20models

Lauter, K. (2018). Second homomorphic encryption standardization workshop delivers the goods. *Microsoft*. Obtenido de https://www.microsoft.com/en-us/research/blog/second-homomorphic-encryption-standardization-workshop-delivers-goods/

Lauter, K. (2018). Tales from the Crypt(ography) Lab with Dr. Kristin Lauter. *Microsoft*. Obtenido de https://www.microsoft.com/en-us/research/podcast/tales-from-the-cryptography-lab-with-dr-kristin-lauter/

Lozano, E. (2018). Datos individuales vs. Datos agregados. *Vocacion Estadistica*. Obtenido de http://vocacionxestadistica.blogspot.com/2018/04/datos-individuales-vs-datos-agregados.html

Lucini, F. (2021). The Real Deal About Synthetic Data. *MIT Sloan*. Obtenido de https://sloanreview.mit.edu/article/the-real-deal-about-synthetic-data/

Melo, S. (2018). Qué es y cómo funciona la nube. *DataScope*. Obtenido de https://datascope.io/es/blog/que-es-y-como-funciona-la-nube/

Microsoft. (2022). Aggregate Functions (Transact-SQL). Obtenido de https://docs.microsoft.com/en-us/sql/t-sql/functions/aggregate-functions-transact-sql?view=sql-server-ver15

Microsoft. (2022). Aggregations and Aggregation Designs. Obtenido de https://docs.microsoft.com/en-us/analysis-services/multidimensional-models-olap-logical-cube-objects/aggregations-and-aggregation-designs?view=asallproducts-allversions

Microsoft. (2022). Anonymization of Visual Studio subscriber information. Obtenido de https://docs.microsoft.com/en-us/visualstudio/subscriptions/anonymization

Microsoft. (2022). Automatic aggregations. Obtenido de https://docs.microsoft.com/en-us/power-bi/enterprise/aggregations-auto

Microsoft. (2022). Cloud Discovery data anonymization. Obtenido de https://docs.microsoft.com/en-us/defender-cloud-apps/cloud-discovery-anonymizer

Microsoft. (2022). Use Aggregate Functions. Obtenido de https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/use-aggregate-functions?view=asallproducts-allversions

Microsoft. (2022). User-defined aggregations. Obtenido de https://docs.microsoft.com/en-us/power-bi/transform-model/aggregations-advanced

Microsoft. (2022). What is differential privacy in machine learning (preview)? Obtenido de https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy

Microsoft. (s.f.). AI Lab Project: Synthetic Data Generator. Obtenido de https://www.microsoft.com/en-us/ai/ai-lab-synthetic-data-showcase

Microsoft. (s.f.). AI Lab projects. Differential privacy. Obtenido de https://www.microsoft.com/en-us/ai/ai-lab-differential-privacy

Microsoft. (s.f.). Explore differential privacy. Obtenido de https://docs.microsoft.com/en-us/learn/modules/explore-differential-privacy/

Microsoft. (s.f.). Homomorphic Encryption. Obtenido de https://www.microsoft.com/en-us/research/project/homomorphic-encryption/

Microsoft. (s.f.). Introducción a los aspectos básicos de Azure. Obtenido de https://docs.microsoft.com/es-es/learn/modules/intro-to-azure-fundamentals/introduction

Microsoft. (s.f.). Microsoft SEAL. Obtenido de https://www.microsoft.com/en-us/research/project/microsoft-seal/#overview

Microsoft Research. (2021). Homomorphic Encryption with Microsoft SEAL. Obtenido de https://www.youtube.com/watch?v=GUOBxFKM5a8

MIT News. (2020). The real promise of synthetic data. Obtenido de https://news.mit.edu/2020/real-promise-synthetic-data-1016

Near, J., Darais, D., & Boeckl, K. (2020). Differential Privacy for Privacy-Preserving Data Analysis: An Introduction to our Blog Series. *NIST. Cybersecurity Insights*. Obtenido de https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our

Nelson, D. (2022). What Is Synthetic Data? *UNITE.AI*. Obtenido de https://www.unite.ai/what-is-synthetic-data/

Nguyen, A. (2019). Understanding Differential Privacy. For Intuitions behind a Theory to a Private AI Application. *Towards Data Science*. Obtenido de https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

Nvidia. (2021). NVIDIA Announces Omniverse Replicator Synthetic-Data-Generation Engine for Training AIs. Obtenido de https://nvidianews.nvidia.com/news/nvidia-announces-omniverse-replicator-synthetic-data-generation-engine-for-training-ais

Oliveros, B. (s.f.). ¿Qué es la nube en informática? ¿Qué es la nube o cloud computing? *Stratus Media Solutions*. Obtenido de https://www.stratusmedia.io/blog/desarrollo-informatico/que-es-la-nube-o-cloud-computing/

PowerData. (s.f.). Definición y principales conceptos relacionados con Data Aggregation. Obtenido de https://www.powerdata.es/concepto-data-aggregation#:~:text=Data%20Aggregation%20es%20un%20tipo,as%C3%AD%20como%20para%20realizar%20an%C3%A1lisis

PowerData. (s.f.). GDPR: Lo que debes saber sobre el reglamento general de protección de datos. Obtenido de https://www.powerdata.es/gdpr-proteccion-datos

Public-key cryptography. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Public-key_cryptography

RAE. (s.f.). Privacidad. Obtenido de https://dle.rae.es/privacidad

Ramaswamy, L. (2021). Common Miscoceptions About Differential Privacy. *KDnuggets*. Obtenido de https://www.kdnuggets.com/2021/11/common-misconceptions-differential-privacy.html

Ray, T. (2019). Enterprise AI and machine learning: Comparing the companies and applications. *ZD Net*. Obtenido de https://www.zdnet.com/article/guide-to-enterprise-ai-and-machine-learning-companies-and-applications/

ReasonWhy. (2015). ¿Cuántas agencias de publicidad hay en España? Obtenido de https://www.reasonwhy.es/actualidad/sector/cuantas-agencias-de-publicidad-hay-en-espana-2015-10-13#:~:text=No%20hay%20forma%20de%20saber%20cu%C3%A1ntas%20agencias%20d

e,hay%20estudios%20actualizados%20sobre%20el%20sector%20publicitario%20espa%C3%B1ol

Record.evolution. (2020). *Anonymization Techniques and Best Practices: A Quick Guide*. Obtenido de https://www.record-evolution.de/en/blog/data-anonymization-techniques-and-best-practices-a-quick-guide/

Reglamento General de Protección de Datos. (s.f.). *Wikipedia*. Obtenido de https://es.wikipedia.org/wiki/Reglamento_General_de_Protecci%C3%B3n_de_Datos

Reply Spur. (s.f.). Microsoft. Client Story: Synthetic Data Improves Microsoft's Next-Gen Conversational AI Engine. Obtenido de https://www.thespurgroup.com/synthetic-data-improves-microsoft-conversational-ai

Right Scale. Flexera. (2019). *State of the Cloud Report.* Obtenido de https://resources.flexera.com/web/media/documents/rightscale-2019-state-of-the-cloud-report-from-flexera.pdf

Ring learning with errors. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Ring_learning_with_errors

RMIT Online. (2020). *Data Science and Privacy. How much is it okay to know?* Obtenido de https://online.rmit.edu.au/blog/data-science-and-privacy-how-much-it-okay-know

Sabharwal, N. (2019). Privacera and Microsoft Azure, Part 4: Data Anonymization and Pseudonymization in Databricks File System. *Medium. Privacera*. Obtenido de https://blog.privacera.com/azure-and-anonymization-databricks-file-system-1d3d91f7ba7e

Satori. (s.f.). Guide: Data Masking. Data Anonymization: Use Cases and 6 Common Techniques. Obtenido de https://satoricyber.com/data-masking/data-anonymization-use-cases-and-6-common-techniques/

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos.* IBM Institute for Business Value. Obtenido de https://www.fundacionseres.org/Lists/Informes/Attachments/951/IBM%20Analytics%20el%20uso%20de%20big%20data%20en%20el%20mundo%20real%20-%20Como%20las%20empresas%20mas%20innovadoras%20extraen%20valor%20de%20datos%20inciertos.pdf

Sue, A. (2019). 6 reasons Microsoft has become the go-to for machine learning. *Medium. DataDrivenInvestor*. Obtenido de https://medium.com/datadriveninvestor/6-reasons-microsoft-has-become-the-go-to-for-machine-learning-e642864ef5f5

Symmetric-key algorithm. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Symmetric-key_algorithm

Synthetic data. (s.f.). *Wikipedia*. Obtenido de https://en.wikipedia.org/wiki/Synthetic_data

Team CFI. (2021). Data Anonymization. Obtenido de https://corporatefinanceinstitute.com/resources/knowledge/other/data-anonymization/

The Synthetic Data Vault. (s.f.). Obtenido de https://sdv.dev/

Tyagi, N. (2021). What is Differential Privacy and How does it Work? *analyticSteps*. Obtenido de https://www.analyticssteps.com/blogs/what-differential-privacy-and-how-does-it-work

UCB. (2021). UCB and Microsoft Expand Collaboration to Accelerate Drug Discovery and Development. Obtenido de https://www.ucb.com/stories-media/Press-Releases/article/UCB-and-Microsoft-Expand-Collaboration-to-Accelerate-Drug-Discovery-and-Development

Unión Europea. (2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. *BOE*. Obtenido de https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807

Walsh, N. (2020). Microsoft Sustainability Calculator helps enterprises analyze the carbon emissions of their IT infrastructure. *Azure Microsoft*. Obtenido de https://azure.microsoft.com/en-us/blog/microsoft-sustainability-calculator-helps-enterprises-analyze-the-carbon-emissions-of-their-it-infrastructure/

Wood, E. (s.f.). *Synthetic Data with Digital Humans.* Microsoft. Obtenido de https://www.microsoft.com/en-us/research/uploads/prod/2019/09/2019-10-01-Synthetic-Data-with-Digital-Humans.pdf

Wood, E., Baltrusaitis, T., Hewitt, C., Dziadzio, S., Johnson, M., Estellers, V., . . . Shotton, J. (2021). *Fake It Till You Make It: Face analysis in the wild using synthetic data alone.* Microsoft.

Obtenido de https://www.microsoft.com/en-us/research/publication/fake-it-till-you-make-it-face-analysis-in-the-wild-using-synthetic-data-alone/

Zewe, A. (2022). When it comes to AI, can we ditch the datasets? *MIT News*. Obtenido de https://news.mit.edu/2022/synthetic-datasets-ai-image-classification-0315