Anexo I. Registro del Título del Trabajo Fin de Grado (TFG)

NOMBRE DEL ALUMNO: Marta Vara Rodríguez

PROGRAMA: E6 Analytics GRUPO: 5 A FECHA: 22/07/2024

Director Asignado: Otero Peinador , Álvaro

Apellidos Nombre

Título provisional del TFG:

Application of Data Science and Predictive Models for Optimal Churn Prevention: Understanding Diverse Causes and Proposing Personalized Actions through Customer Segmentation

ADJUNTAR PROPUESTA (máximo 2 páginas: objetivo, bibliografía, metodología e índice preliminares)

- **OBJETIVO:**

In the competitive landscape of subscription-based businesses, reducing churn and boosting revenue are key priorities. This project aims to leverage predictive models not only to forecast customer cancellations but also to identify the essential strategies and resources needed for effective customer retention. By segmenting customers accurately and understanding the underlying reasons behind churn predictions, the project seeks to offer actionable insights. Through the application of advanced data science techniques, it explores the most relevant metrics and provides a comprehensive guide to interpreting and communicating results, maximizing their impact on the company's performance. Geared towards technical professionals with a background in programming and data analysis, this project aspires to empower readers to make immediate, significant contributions to reducing churn and driving revenue growth. Additionally, the study delves into the broader implications of churn and its influence on business strategy.

- **INDICE:**

SECTION 1: FUNDAMENTALS AND THEORY

1. Introduction

    1.1 Literature review

        1.1.1 Churn as a concept

        1.1.2 Chrurn Metrics

        1.1.3 Data analysis

    1.2 Motivation

    1.3 Objectives

    1.4 Methodology

    1.5 Resources

    1.6 Thesis flowchart

2. Concept of Churn

    2.1 Definition of Churn

    2.2 Impact of Churn on Companies

    2.3 Factors Contributing to Churn

3. Types of Services and Metrics

    3.1 Types of Services

        3.1.1 Saas Services

            3.1.1 Paid Consumer Products

            3.1.2 Business-to-Business (B2B) Services

            3.1.3 Freemium Business Models

        3.1.2 Non- SaaS Services

            3.1.2.1 In-App Purchase Models

            3.1.2.2 Ad-Supported Media and Apps

    3.2 Key Metrics for Combating Churn

        3.2.1 Metrics for SaaS Services

        3.2.2 Metrics for Non-SaaS Services

SECTION 2: PRACTICAL ANALYSIS AND PREDICTION

4. Data Preparation
    4.1 Description of the Transactional Database
    4.2 Data Cleaning Process
    4.2.1 Identification of Missing Data
    4.2.2 Treatment of Outliers
    4.2.3 Data Normalization
    4.3 Structuring Data for the Predictive Model
5. **Data Exploration**

    5.1 Data Exploration Techniques and Insights
    5.1.1 Descriptive Statistics
    5.1.2 Visual Exploration (Graphs and Charts)
    5.1.3 Non-Predictive Analyses (e.g., Survival Analysis, Kaplan-Meier Curves

6. Building the Predictive Model

    6.1 Selection of the Prediction Algorithm
    6.2 Model Implementation
    6.2.1 Splitting Data into Training and Test Sets
    6.2.2 Training the Model
    6.2.3 Model Validation and Tuning
    6.3 Interpretation of Model Results
    6.3.1 Significant Variables
    6.3.2 Model Performance Evaluation
    6.3.3 Analysis of Churn Predictions
    6.4 Model Interpretability
    6.4.1 SHAP Values and Interpretability

SECTION 3: ACTIONS AND POST-ANALYSIS STRATEGIES

7. Scenarios and Strategies
    7.1 Possible Scenarios Based on Predictions
    7.1.1 High Churn Probability Customers
    7.1.2 Low Churn Probability Customers
    7.1.3 Segmenting Customers by Churn Probability
    6.2 Proposals for Personalized Actions
    6.2.1 Retention Strategies for High Churn Probability Customers
    6.2.2 Loyalty Programs for Loyal Customers
    6.2.3 Reactivation Strategies for Inactive Customers
8. Implementation Planning
    7.1 Defining Priorities and Resources
    7.2 Integrating Strategies into the Business Plan
    7.3 Measuring and Monitoring Results
9. Conclusions
    8.1 Summary of Findings
    8.2 Expected Impact on the Company
    8.3 Recommendations for Future Research
10. References
11. Appendices
    11.1 Glossary of Terms
    11.2 Source Code and Scripts Used
    11.3 Additional Tables and Figures

- METHODOLOGY:

To achieve the objectives of this study, we will utilize a quantitative approach, primarily focusing on data analytics techniques. This methodology is chosen because it handles large

datasets and derives meaningful insights through statistical analysis and predictive modeling. The quantitative approach will allow us to identify patterns and trends in customer behavior, which are critical for understanding and preventing churn.

We will be using two main analytical tools in this study: Python and Graphext. Python will be employed for data exploration, cleaning, and preliminary analysis. Its extensive libraries, such as pandas, NumPy, and scikit-learn, provide robust capabilities for handling data and building predictive models. Graphext, a powerful data analytics and visualization tool, will be used throughout the model creation, iteration, and implementation phases. Graphext's intuitive interface and sophisticatics features will enable us to visualize complex data relationships, enhance the interpretability of our models, and facilitate the iterative process of refining predictive models.

The data for this study will be sourced from transactional databases, similar to those used by large companies like Amazon. This data will undergo a thorough cleaning process to ensure accuracy and reliability. The steps will include identifying and handling missing values, detecting and treating outliers, and normalizing the data to prepare it for analysis. (This paragraph will be modified when we have the specific data and sources).

Once the data is prepared, we will proceed to build a predictive model to identify customers at risk of churn. The selection of the prediction algorithm will be based on the data characteristics and the specific requirements of the study. The model will be trained using a subset of the data, and its performance will be validated through various metrics to ensure its accuracy and reliability.

After building and validating the model, we will interpret the results to identify significant variables contributing to churn. To enhance the interpretability of our model, we will employ techniques such as SHAP (SHapley Additive exPlanations) values. This will allow us to understand not only the likelihood of churn but also the underlying reasons behind it, enabling more targeted and effective intervention strategies tailored to different customer segments.

In conclusion, this methodology combines robust data analytics tools and techniques to provide a comprehensive approach to churn prevention. By leveraging Python and Graphext, we aim to deliver actionable insights that can significantly impact the business's ability to retain customers and increase revenue. Additionally, by focusing on model interpretability, we emphasize the importance of understanding the *why* behind churn predictions, which is crucial for developing personalized actions to address the unique needs of each customer segment.

Firma del estudiante:

Marta Vara

Fecha: 21/10/ 2024