COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

GRADO UNIVERSITARIO EN BUSINESS ANALYTICS

TRABAJO FIN DE GRADO

Quantifying the Business Impact of Catenary Failures in Railway Operations

Autor: Lucía Martínez Ruiz

Director: Miguel Ángel Sanz Bobi

Madrid

Abril de 2026

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

Quantifying the Business Impact of Catenary Failures in Rail Operations

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

curso académico 2025/26 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.: Lucía Martínez Ruiz Fecha: 07/04/ 2026

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Firmado por SANZ BOBI MIGUEL ANGEL - ***6599** el día 21/04/2026 con un certificado emitido por AC FNMT Usuarios

Fdo.: Miguel Ángel Sanz Bobi Fecha: 20/04/2026

GRADO UNIVERSITARIO EN BUSINESS ANALYTICS

TRABAJO FIN DE GRADO

Quantifying the Business Impact of Catenary Failures in Railway Operations

Autor: Lucía Martínez Ruiz

Director: Miguel Ángel Sanz Bobi

Madrid

Acknowledgements

I would like to express my sincere gratitude to my project director, Miguel Ángel Sanz Bobi, for his guidance, support, and valuable feedback throughout the development of this work. His advice and continuous encouragement were essential in helping me define the direction of the project and improve its quality.

QUANTIFYING THE BUSINESS IMPACT OF CATENARY FAILURES IN RAIL OPERATIONS

Autor: Martínez Ruiz, Lucía.
Director: Sanz Bobi, Miguel Ángel.
Entidad Colaboradora: ICAI – Universidad Pontificia Comillas.

RESUMEN DEL PROYECTO

Este proyecto desarrolla un análisis de incidentes ferroviarios relacionados con la catenaria a partir de registros operativos históricos. El enfoque propuesto combina la clasificación de incidentes, la predicción del impacto, la agrupación de líneas y la generación de KPIs orientados al coste con el fin de apoyar tanto el análisis técnico como la gestión de los mismos. Los resultados muestran que los datos estructurados de incidentes contienen señales predictivas útiles, que la severidad de las disrupciones se concentra en una parte limitada de la red y que una analítica integrada puede proporcionar un apoyo significativo para la evaluación de las incidencias ferroviarias.

Palabras clave: Analítica de incidencias ferroviarias, Catenaria, Aprendizaje Automático, Mantenimiento Predictivo, Modelado de Costes.

1. Introducción

La fiabilidad ferroviaria es una cuestión crítica porque las interrupciones del servicio afectan no solo a la operativa, sino también a la experiencia del pasajero y al rendimiento económico. En las redes ferroviarias electrificadas, los incidentes relacionados con la catenaria son especialmente importantes debido a su efecto directo sobre la continuidad del tráfico y la puntualidad.

Tradicionalmente, los registros de incidentes se han utilizado con fines de reporte y documentación. Sin embargo, la creciente disponibilidad de datos operativos históricos permite avanzar hacia un enfoque más orientado a los datos. En este contexto, los métodos de aprendizaje automático pueden ayudar a identificar patrones, estimar el impacto y proporcionar indicadores que mejoren el análisis de las disrupciones. Este proyecto surge de la necesidad de transformar los registros brutos de incidentes en un marco analítico más útil para el apoyo a la toma de decisiones en el ámbito ferroviario.

2. Definición del proyecto

El objetivo principal del proyecto es desarrollar un marco analítico para el estudio de incidentes ferroviarios relacionados con la catenaria utilizando datos operativos históricos. El trabajo busca responder a tres preguntas conectadas entre sí: si un incidente está relacionado con la catenaria, cuál es la severidad de sus consecuencias operativas y cómo puede expresarse esa disrupción en términos económicos.

Para abordar estas cuestiones, el proyecto se organiza en tres tareas. La **Tarea A** se centra en la clasificación de incidentes. La **Tarea B** aborda la severidad operativa mediante la predicción de los minutos de retraso y las cancelaciones, e incluye la agrupación de líneas ferroviarias según sus perfiles de disrupción. La **Tarea C** está orientada al coste, donde la disrupción se traduce en una carga económica estimada y se agrega mediante KPIs.

3. Descripción del modelo/sistema/herramienta

El sistema parte de registros históricos de incidentes que contienen variables estructuradas, información temporal, campos relacionados con la localización y descripciones textuales. Dado que estos registros fueron creados originalmente para el reporte operativo y no para el modelado predictivo, el primer paso consiste en un preprocesado de los datos. Incluye limpieza de datos, canonización, análisis temporal y construcción de características para obtener un conjunto de datos analítico consistente.

Se combina atributos estructurados con información textual extraída de las descripciones de los incidentes mediante una representación TF-IDF [1]. Además, utiliza agrupamiento KMeans para agrupar las líneas ferroviarias según sus perfiles [2]. El sistema global se divide en tres módulos conectados: un módulo de clasificación para identificar incidentes relacionados con la catenaria, un módulo de severidad para estimar retrasos y cancelaciones y segmentar líneas, y un módulo de costes para generar estimaciones económicas y KPIs. En la Figure 1 se muestra la estructura general del modelo propuesto.

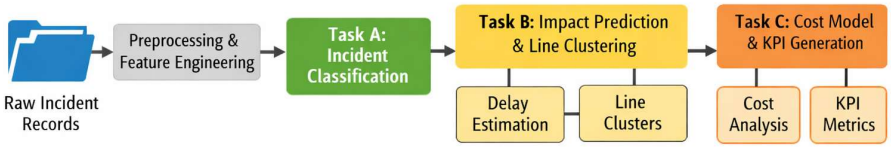

Figure 1: Descripción del Modelo

4. Resultados

Los resultados muestran que el sistema propuesto es capaz de extraer un valor analítico significativo a partir de estos datos. En la **Tarea A**, la etapa de clasificación confirma que los registros operativos estructurados contienen suficiente señal para identificar incidentes relacionados con la catenaria con un rendimiento predictivo útil. Esto respalda la idea de que los registros de incidentes ferroviarios pueden utilizarse no solo con fines de reporte, sino también como base para el apoyo analítico a la toma de decisiones.

En la **Tarea B**, los modelos son capaces de estimar el impacto en términos de retrasos y cancelaciones con una utilidad razonable para los incidentes típicos, aunque el rendimiento predictivo se debilita en la cola superior de la distribución. La etapa de agrupamiento también muestra que las líneas ferroviarias no son homogéneas: mientras que la mayoría de las líneas siguen un perfil de disrupción más regular, un subconjunto más reducido concentra actividad recurrente o resultados inusualmente severos. Este resultado es especialmente relevante porque revela que el riesgo operativo se distribuye de forma desigual a lo largo de la red.

La aportación práctica más importante surge en la **Tarea C**, donde las salidas operativas se traducen en una capa económica. El análisis muestra que los incidentes relacionados con la catenaria representan una carga anual considerable y que esta carga varía significativamente a lo largo del tiempo. Además, los resultados confirman que el impacto económico está altamente concentrado en un subconjunto limitado de líneas y

lugares, lo que refuerza el valor de combinar la predicción con la agregación basada en KPIs.

La siguiente Figure 2 ilustra esta perspectiva económica mostrando el coste anual estimado de los incidentes relacionados con la catenaria. La figura pone de manifiesto tanto la magnitud de la carga asociada a las disrupciones como su variabilidad interanual, dejando claro que el problema no solo es relevante desde el punto de vista operativo, sino también desde el económico.

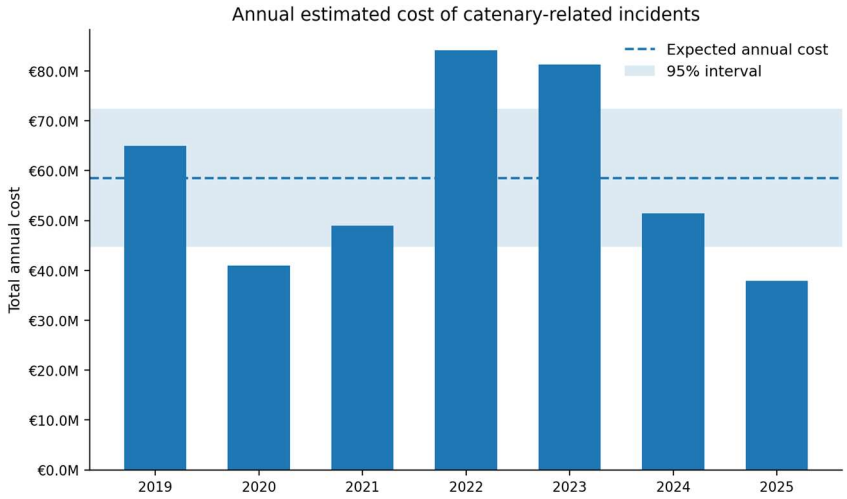

Figure 2: Estimación de costes anual de incidentes de catenaria

5. Conclusiones

Este proyecto muestra que los incidentes ferroviarios relacionados con la catenaria pueden analizarse mediante un marco integrado que combina clasificación, predicción de impacto, agrupamiento y generación de KPIs orientados al coste. La principal contribución del trabajo radica en conectar estas capas analíticas en un flujo coherente que apoye tanto la comprensión técnica como la interpretación desde la gestión.

Al mismo tiempo, los resultados confirman una limitación importante: los incidentes más severos y costosos siguen siendo los más difíciles de predecir con precisión. Por este motivo, el proyecto propuesto debe entenderse principalmente como una herramienta de apoyo a la toma de decisiones y no como un sistema de predicción totalmente autónomo. Aun así, demuestra el valor práctico de combinar aprendizaje automático con indicadores operativos y económicos en el análisis de disrupciones ferroviarias.

6. Referencias

[1] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

[2] Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

QUANTIFYING THE BUSINESS IMPACT OF CATENARY FAILURES IN RAIL OPERATIONS

Author: Martínez Ruiz, Lucía.
Supervisor: Sanz Bobi, Miguel Ángel.
Collaborating Entity: ICAI – Universidad Pontificia Comillas.

ABSTRACT

This project develops a data-driven framework for the analysis of catenary-related railway incidents using historical operational records. The proposed approach combines incident classification, impact prediction, line clustering, and cost-oriented KPI generation to support both technical analysis and managerial interpretation. The results show that structured incident data contains useful predictive signals, that disruption severity is concentrated in a limited part of the network, and that integrated analytics can provide meaningful support for railway disruption assessment.

Keywords: Railway incident analytics, Catenary, Machine Learning, Predictive Maintenance, Cost Modelling.

1. Introduction

Railway reliability is a critical issue because service disruptions affect not only operations, but also passenger experience and economic performance. In electrified railway networks, catenary-related incidents are especially important due to their direct effect on traffic continuity and punctuality. For this reason, their analysis is relevant from both a technical and a managerial perspective.

Traditionally, incident records have been used mainly for reporting and documentation. However, the increasing availability of historical operational data makes it possible to move toward a more data-driven approach. In this context, machine-learning methods can help identify patterns, estimate likely impact, and provide structured indicators that improve disruption analysis. This project is motivated by the need to transform raw incident logs into a more useful analytical framework for railway decision support.

2. Project definition

The main objective of the project is to develop an analytical framework for the study of catenary-related railway incidents using historical operational data. More specifically, the work seeks to answer three connected questions: whether an incident is catenary-related, how severe its operational consequences are likely to be, and how that disruption can be expressed in economic terms.

To address these questions, the project is organized into three tasks. **Task A** focuses on incident classification. **Task B** addresses operational severity through the prediction of delay minutes and cancellations and also includes the clustering of railway lines according to their disruption profiles. **Task C** introduces a cost-oriented layer in which disruption is translated into estimated economic burden and aggregated through KPIs. In this way, the project moves from incident data to a broader decision-support perspective.

3. Model/system/tool description

The system starts with historical incident records containing structured variables, temporal information, location-related fields, and short textual descriptions. Since these records were originally created for operational reporting rather than for predictive modelling, the first step of the framework consists of preprocessing. This includes data cleaning, canonicalization, temporal parsing, and feature engineering to obtain a consistent analytical dataset.

The framework combines structured attributes with text-based information extracted from incident descriptions using TF-IDF representation [1]. In addition, it uses KMeans clustering to group railway lines according to their disruption profiles [2]. The overall system is divided into three connected modules: a classification module for identifying catenary-related incidents, a severity module for estimating delay and cancellations and segmenting lines, and a cost module for generating economic estimates and KPIs. In the following Figure 3 it shows the framework structure.

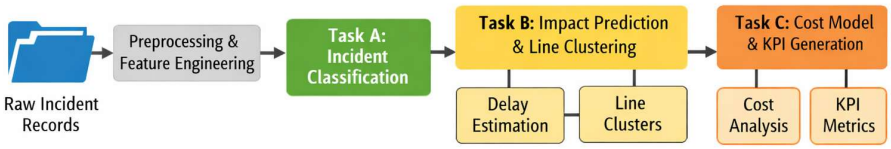

Figure 3: Model Description

This modular design makes the system both interpretable and scalable. Each module addresses a specific analytical layer, but all of them contribute to a single integrated framework capable of supporting structured railway disruption analysis.

4. Results

The results show that the proposed framework can extract meaningful analytical value from historical railway incident data. In **Task A**, the classification stage confirmed that structured operational records contain enough signal to identify catenary-related incidents with useful predictive performance. This supports the idea that railway incident logs can be used not only for reporting purposes, but also as a basis for analytical decision support.

In **Task B**, the models were able to estimate delay and cancellation impact with reasonable usefulness for typical incidents, although predictive performance became weaker in the upper tail. The clustering stage also showed that railway lines are not homogeneous: while most lines follow a more regular disruption profile, a smaller subset concentrates recurrent activity or unusually severe outcomes. This result is especially relevant because it reveals that operational risk is distributed unevenly across the network.

The strongest practical insight emerged in **Task C**, where the operational outputs were translated into an economic layer. The analysis showed that catenary-related incidents represent a substantial annual burden and that this burden varies significantly over time. In addition, the results confirmed that economic impact is highly concentrated in a

limited subset of lines and places, which reinforces the value of combining prediction with KPI-based aggregation.

The following Figure 44 illustrates this economic perspective by showing the estimated annual cost of catenary-related incidents. The figure highlights both the magnitude of the disruption burden and its interannual variability, making clear that the problem is not only operationally relevant but also economically significant.

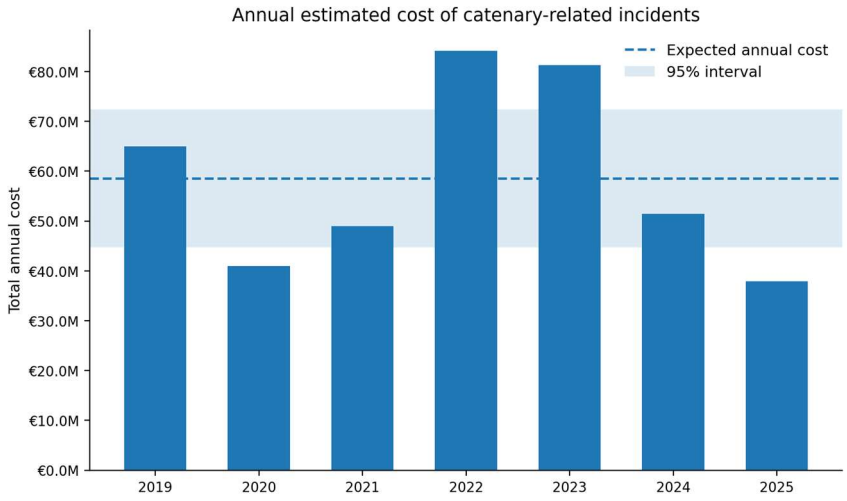

Figure 4: Annual estimated cost of catenary-related incidents

5. Conclusions

This project shows that catenary-related railway incidents can be analyzed through an integrated framework that combines classification, impact prediction, clustering, and cost-oriented KPI generation. The main contribution of the work lies in connecting these analytical layers into a coherent pipeline that supports both technical understanding and managerial interpretation.

At the same time, the results confirm an important limitation: the most severe and costly incidents remain the hardest to predict accurately. For this reason, the proposed framework should be understood primarily as a decision-support tool rather than as a fully autonomous forecasting system. Even so, it demonstrates the practical value of combining machine learning with operational and economic indicators in railway disruption analysis

6. References

[1] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

[2] Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Contents

List of Figures

List of Tables

Chapter 1. INTRODUCTION

Within Europe's mobility and decarbonization strategy, rail transportation is widely recognized as a key enabler; however, its effectiveness is constrained by service reliability and disruption risk. For passengers, reliability is experienced through service continuity and punctuality. In Great Britain, for example, the regulator reports that **84.3%** of passenger services arrived within three minutes of schedule in 2024-2025 and cancellations were **4.1%**, illustrating both the scale of operations and the persistent challenge of service performance [[4]]. Even though performance and compensation frameworks vary by country, the consequences of poor reliability are universal: immediate operational and financial costs, followed by longer-term damage to demand, trust, and modal shift.

1.1 RELIABILITY AS A SYSTEM-LEVEL CHALLENGE

Railways operate as tightly interconnected systems. A failure at a single point, such as an infrastructure fault, power interruption, or operational incident, can quickly propagate through the network, affecting timetables, rolling stock rotations, and crew schedules. As a result, many incidents are not confined to the location where they occur: they often translate into network-wide impacts on punctuality and service continuity. For this reason, infrastructure managers and operators monitor disruption at multiple levels, ranging from detailed incident logs to aggregated measures of "major incident" delay minutes and headline punctuality and cancellation indicators.

Open-data initiatives illustrate the operational importance of major incidents. In Belgium, Infrabel [1] publishes datasets that explicitly quantify the impact of "most important incidents" as those whose monthly impact **exceeds 1,000 minutes of passenger-train delay**, including the **minutes of delay**, **number of cancelled trains**, and contextual fields such as **place** and **line number**.

Complementary punctuality datasets from the same source [[5]] define "major incidents" through concrete operational thresholds. For example, a train delayed beyond a minimum duration, situations where multiple trains experience significant delays, events that trigger partial or total cancellations, and incidents with safety implications. These definitions matter because they formalize a practical distinction used in day-to-day operations: routine variability is expected, but a small subset of disruptions is treated as "major" due to its disproportionate impact. In other words, the sector already recognizes that high-impact events dominate service performance outcomes, drive recovery efforts, and concentrate a large share of the economic and reputational cost.

The socioeconomic impact of delays is well documented in transport economics. A 2024 study on passenger rail delay valuation [[6]] shows that passengers typically perceive a minute of arrival delay as more costly than a minute of scheduled travel time. In other words, unreliability carries a penalty beyond pure time loss, because it introduces uncertainty, stress, missed connections, and disruption to planned activities. This supports the idea that "one minute of delay" is not merely an operational metric: it is a practical proxy for customer experience and wider welfare impacts.

Finally, in some rail markets, delays and cancellations lead to large compensation outlays. For instance, UK passenger compensation under "Delay Repay" has been reported to reach record levels (over £100m in a year in some reporting), underscoring that reliability has tangible financial consequences. Although the precise mechanisms differ across countries, the broader insight generalizes **reliability problems create both operational and economic burdens**, and the ability to identify, anticipate, and mitigate high-impact incidents is strategically valuable.

1.2 ELECTRIFIED RAIL AND THE CATENARY AS A CRITICAL ASSET CLASS

In electrified railways, trains draw electricity through the interaction between the pantograph (mounted on the train) and the overhead contact line (often referred to as the catenary), as shown in Figure 5. Because this system is continuously stressed mechanically and electrically, faults can lead to sudden service stoppages, power isolation, and safety constraints that immediately affect punctuality.

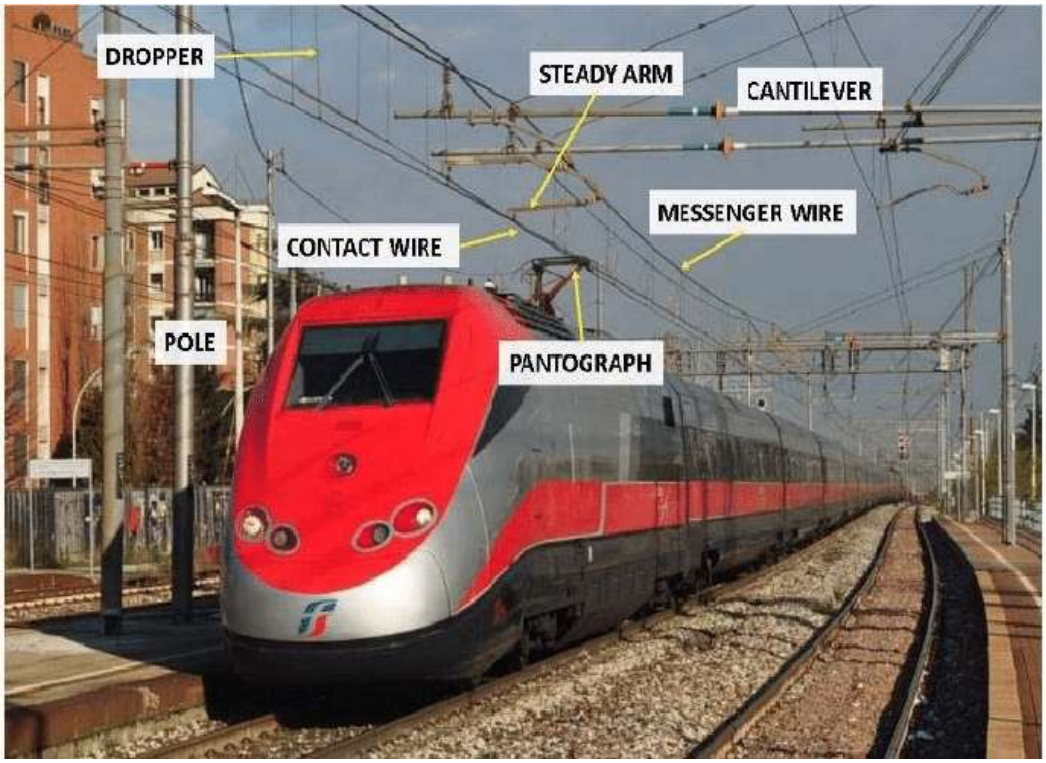

Figure 5: Electrified Railways [18]

From an engineering standpoint, failure mechanisms in overhead line systems include wear, fatigue, arcing, component looseness, geometry deviations, and contact wire rupture. Research on overhead contact wires [[8]] emphasizes that fatigue can cause **sudden catastrophic failures** due to repeated bending and tension in the pantograph-wire interaction. Recent work focusing specifically on contact wire rupture [[9]] notes that rupture

events are expected to be rare, yet they remain important because of their operational impact and the need for effective monitoring and response.

In operational terms, catenary-related disruptions tend to follow a distinctive pattern. They may require immediate power isolation, impose temporary restrictions on train movements, and trigger urgent maintenance actions on the infrastructure. These interventions frequently translate into substantial delay minutes and, in more severe cases, service cancellations. This makes catenary incidents a strong candidate for targeted analytics: if operational reports can be used to quickly flag catenary-related events and anticipate their likely impact, control rooms and maintenance teams can prioritize resources earlier, respond more consistently, and reduce both disruption time and recovery costs.

1.3 INDUSTRY DIRECTION: FROM REACTIVE MAINTENANCE TO DATA-DRIVEN PREDICTIVE MAINTENANCE

Rail infrastructure has traditionally been maintained through scheduled inspections and corrective interventions after faults. However, the sector is increasingly moving toward condition-based maintenance (CBM) and predictive maintenance, supported by sensitization, digitalization, and machine learning. European innovation programs explicitly frame predictive maintenance as a lever for improving reliability and reducing lifecycle costs. Shift2Rail [[10]] (and, more recently, the Europe's Rail Joint Undertaking ecosystem) has highlighted health monitoring systems and predictive approaches as core solutions to substitute purely preventive maintenance with condition-based methods.

A recent survey of AI-enabled predictive maintenance for railway systems [[11]] synthesizes opportunities and challenges for intelligent maintenance ecosystems, reinforcing that the field is moving quickly toward data-rich, ML-supported decision-making. Reviews of intelligent railway infrastructure monitoring [[12]] similarly emphasize advances in sensing,

feature extraction, and machine-learning methods for detection and diagnosis across railway assets.

At the subsystem level, monitoring solutions [[13]] exist for pantograph-catenary interaction and overhead line condition, including optical and sensor-based approaches that support inspection and anomaly detection. Academic work [[14]] also demonstrates ML frameworks for infrastructure event detection from vibration or other sensor signals, with the goal of earlier detection and improved maintenance response.

This context is important: **the sector is aligned on predictive maintenance**, but practical implementation is constrained by data availability, heterogeneity, and the difficulty of connecting technical signals to operational impact and business KPIs.

Chapter 2. TECHNOLOGIES DESCRIPTION

This chapter presents the main technologies and analytical techniques used in the project. Since the work combines historical incident data, machine-learning models, text-processing techniques, and cost-oriented indicators, it is necessary to describe the role of each of these components in the overall analytical framework.

2.1 ANALYTICAL ENVIRONMENT

The project is based on the analysis of historical railway incident records from 2019 to 2025, containing structured attributes, temporal information, location-related variables, and short textual descriptions. To transform these raw records into a modelling-ready dataset, several data-processing steps are required, including schema harmonization, missing-value handling, text cleaning, temporal parsing, and feature construction.

This type of analytical workflow belongs to a standard data-science environment in which raw operational data are progressively converted into structured inputs suitable for machine-learning models. In the present project, data processing is an essential technological layer because the original records were created for operational reporting rather than for predictive analytics.

2.2 MACHINE-LEARNING TECHNIQUES

- The project uses three main analytical paradigms: **classification**, **regression**, and **clustering**. Classification is applied to determine whether an incident is catenary-

related or not, which makes it possible to distinguish the events of interest from the wider pool of railway incidents.

- Regression is used to estimate the likely operational impact of incidents, measured through delay minutes and cancellations, and later to support cost estimation.
- Clustering is applied at line level to identify groups of railway lines with similar disruption profiles, providing a broader network-level view of operational behaviour.

A relevant part of the available information is contained in short free-text incident descriptions. To incorporate this information into the models, the project uses **TF-IDF** (Term Frequency-Inverse Document Frequency), a common text representation technique that transforms words into weighted numerical features according to their relative importance in the document collection [19]. Since this representation may become highly dimensional, the project also applies **Singular Value Decomposition (SVD)** to reduce dimensionality and obtain a more compact representation of the textual information [20].

For the segmentation stage, the project uses **KMeans clustering**, an unsupervised method that partitions observations into groups according to similarity in their feature values [21]. In this case, clustering is used to group railway lines according to their disruption frequency and severity patterns, helping distinguish between routine lines, recurrent high-activity lines, and rare but high-severity outliers.

2.3 COST MODELLING AND KPI GENERATION

Beyond technical prediction, the project includes an economic interpretation layer based on cost modelling and KPI construction. Cost modelling is used to translate operational disruption into an estimated economic burden, making it possible to interpret delay and cancellation patterns in managerial rather than purely technical terms. This is especially useful because operational indicators alone do not always express the business relevance of disruption clearly enough.

On top of this, the project generates **Key Performance Indicators (KPIs)** that aggregate incident behaviour by time period, location, and line cluster. These indicators support a more structured interpretation of the results, allowing the analysis to move from individual predictions to a broader decision-support perspective.

Chapter 3. STATE OF THE ART

Existing research and practice provide valuable building blocks, but several gaps remain when moving from "asset monitoring" to "actionable, end-to-end incident intelligence":

1. **Asset vs. operations focused analytics.**
   Many studies [[15]] focus on detecting physical degradation or specific failure modes of overhead line components using engineering data or sensor streams. For example, failure mode and maintenance record analyses provide structured insights into component-level failure patterns and maintenance priorities. These approaches can be highly effective when high-quality sensor or inspection data exist, but they do not always integrate with operational incident reporting, delay attribution, and cancellations, which is information that matters for service performance management.

2. **Incident impact modelling exists but is often separated from root-cause classification.**
   There is a growing body of work that models how incidents translate into delays on the network. For instance, recent research [[16]] proposes frameworks to quantify the impact of incidents on delay outcomes at network scale, incorporating exposure and rate factors. However, in day-to-day operations, one often needs both: (i) **rapid categorization** of "what kind of incident is this?" (such as catenary vs. other causes) and (ii) **impact estimation** (how many minutes or cancellations might follow). Treating these as disconnected tasks can limit usefulness for triage and prioritization.

3. **Unstructured incident descriptions are underutilized.**
   Many infrastructure organizations maintain rich free-text descriptions of incidents, sometimes in multiple languages, because operations staff record events quickly under time pressure. These descriptions contain critical semantic cues (equipment names, fault symptoms, response actions) but are inconsistent, abbreviated, and

difficult to standardize. Public datasets like Infrabel's "most important incidents" [[1]] illustrate both the opportunity (structured delay and cancellation outcomes) and the challenge (free-text description fields, multilingual context).

4. **Economic translation is often missing from technical ML pipelines.** Even when models exist for failure detection or delay prediction, they are not always connected to a transparent cost model that stakeholders can use for scenario planning (answering questions such as: "What if we reduce severe catenary incidents on high-impact lines?"). Transport economics provides methods for valuing delay impacts, and industry regimes create direct financial consequences, but operational ML models frequently stop at technical metrics [[17]].

Chapter 4. PROJECT DEFINITION

This chapter defines the project from a technical, analytical, and practical perspective. Based on the context presented in the previous chapters, the aim here is to justify why the project is worth developing, to establish its objectives, to describe the overall methodological approach, and to outline its planning and approximate economic dimension.

4.1 JUSTIFICATION

This project addresses the above practical gap by building an end-to-end, operations-oriented analytics pipeline around high-impact rail incidents, with a specific focus on catenary-related disruption and its consequences.

4.1.1 NEED FOR AN INTEGRATED ANALYTICAL FRAMEWORK

The first reason for developing this project is the lack of integrated approaches that connect different analytical layers of railway disruption. In practice, incident records can support multiple forms of analysis: identifying the type of incident, estimating its likely operational consequences, and assessing its economic implications. However, these dimensions are often treated separately. As a result, the output may remain fragmented and difficult to use for broader decision-making.

This project brings these dimensions together in a single workflow. Instead of building an isolated classifier or an isolated regression model, it proposes a multi-stage framework in which the outputs of one task help structure the next. In this way, the project reflects the reality of operational analysis more closely, since railway disruption is not only a technical issue but also an organizational and economic one.

4.1.2 OPERATIONAL RELEVANCE

A second justification is the practical relevance of the problem. Catenary-related incidents may lead to delays, cancellations, service degradation, and concentrated disruption in specific parts of the network. For infrastructure managers and operators, understanding these patterns is important not only for technical diagnosis but also for prioritization, planning, and communication.

The project is therefore justified by its ability to produce outputs that are meaningful from an operational and managerial perspective. Classification results can help distinguish relevant incidents; clustering can reveal which lines behave as recurrent or high-severity problem areas; and cost-based KPIs can translate technical disruption into business-oriented indicators.

4.1.3 FEASIBILITY OF THE PROPOSED APPROACH

A third justification lies in the feasibility of the proposed solution. The project builds on historical incident data that are already available and uses machine-learning techniques that are well-established, interpretable, and appropriate for the scale of the problem.

At the same time, the framework is applicable beyond the narrow context of a single model output. Its modular structure means that it can be extended, refined, or adapted to broader operational environments. The project is not only feasible to develop, but also relevant as a prototype of a reusable applied analytics solution

4.2 OBJETIVES

The general objective of this project is to develop a data-driven analytical framework for the study of catenary-related railway incidents that supports incident identification, impact

estimation, and cost-oriented interpretation through machine-learning techniques and KPI generation.

Objective 1. Establish a quantified baseline. Using the Infrabel dataset [[1]], incident-level and line/location aggregates of delay minutes and cancellations will be computed, normalized, and converted to euros via a transparent cost model. Parameter sets will yield a baseline of the problem's current magnitude.

Objective 2. Predict the likelihood of catenary causality. A supervised classifier will assign calibrated probabilities that incidents are catenary-related, leveraging structured fields and incident text. Performance will be assessed with temporal splits, ROC-AUC, and reliability curves.

Objective 3 Predict operational impact. Regression models suited to skewed (delay minutes) and count outcomes (cancellations) will estimate expected impact with uncertainty bands and optional severity tiers for managerial clarity.

Objective 4. Translate predictions into euros and KPIs. Predicted impacts will be monetized using a cost framework, producing decision KPIs: € per incident, € per km/train-km, and annual expected cost.

Objective 5. Prioritize preventative actions. A unified expected-loss metric (probability x severity in €) will generate a ranked risk map by line/location and simple intervention rules, with estimated effects on delays and cancellations.

Objective 6. Quantify financial value. Baseline versus mitigation scenarios will be simulated via Monte Carlo to estimate expected savings.

Objective 7. Package decision-ready deliverables. Outputs will include a concise dashboard and a brief commercial note on target customers, pricing options, and adoption roadmap.

4.3 METHODOLOGY

This project follows a structured data-driven methodology designed to transform raw railway incident records into operationally and managerially useful outputs. Rather than addressing only one isolated analytical task, the proposed approach is organized as a multi-stage framework that combines incident classification, impact prediction, and economic interpretation.

The workflow begins with the collection and preparation of historical incident records. Since the original data were generated for operational reporting rather than for machine-learning purposes, a dedicated preprocessing stage is required to harmonize the dataset and make it suitable for analysis. This stage includes schema normalization, cleaning of textual descriptions, treatment of missing and duplicated fields, temporal parsing, and the construction of canonical variables that can be used consistently throughout the project. Feature engineering is then applied to extract useful analytical signals from the available information, including structured attributes, temporal context, location-related variables, and text-based representations of incident descriptions.

Once the dataset has been prepared, the project is divided into three analytical tasks, as seen in Figure 6:

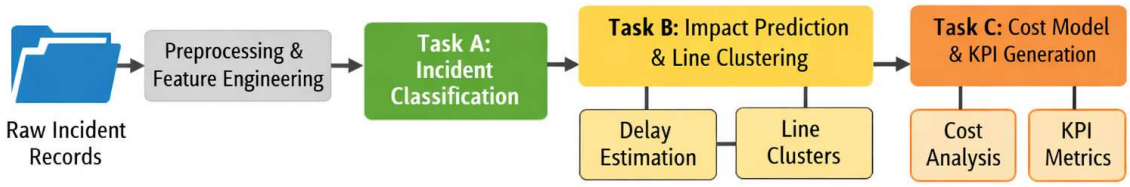

Figure 6: Model Description

- **Task A** focuses on the **classification of incidents** as catenary-related or non-catenary-related. The objective of this first stage is to determine whether structured operational records contain enough information to identify the type of disruption with

useful predictive performance. This task provides the initial interpretation layer of the framework.

- **Task B** addresses incident severity from two complementary perspectives. First, it estimates the **likely operational consequences of incidents** in terms of delay minutes and cancellations. Second, it introduces a **clustering analysis** at line level to identify groups of railway lines with similar disruption profiles. This makes it possible to move beyond event-level prediction and incorporate a broader network-level view of risk concentration.
- **Task C** extends the analysis into an **economic dimension**. In this task, the outputs of the previous ones are translated into a cost-oriented framework that supports the generation of key performance indicators. The methodology considers both a chained strategy, in which cost is derived from previously estimated operational impact, and a direct strategy, in which incident cost is predicted directly.

To ensure methodological soundness, the project uses a train/validation/test logic that prevents information leakage and preserves a realistic separation between model development and final evaluation. Model performance is assessed using metrics appropriate to each task, including classification metrics for Task A, regression metrics for Task B, and cost-oriented error indicators for Task C. In addition, the analysis includes interpretation by line, place, cluster, and time period, so that the final outputs can be understood not only at the aggregate level but also in operational context

4.4 PLANIFICATION AND ECONOMIC ESTIMATION

The project is developed through five main phases: problem definition and literature review, data preparation, model development, results analysis, and final writing. This structure made it possible to organize the work progressively, from understanding the railway disruption problem to building and evaluating the analytical models and finally documenting the results in a coherent way.

The first phase focused on defining the problem, studying the sector context, and reviewing prior work. The second phase was devoted to data preparation, including cleaning, canonicalization, temporal parsing, and feature engineering. The third phase covered the development of the three analytical tasks proposed in the project: incident classification, impact prediction with line clustering, and cost-oriented KPI modelling. The fourth phase consisted of analyzing the results and transforming them into tables, figures, and written discussion. Finally, the last phase was dedicated to writing, revising, and consolidating the final dissertation. In Table 1 the phases are summarized.

Phase	Main activities	Estimated duration
Phase 1	Problem definition and literature review	2–3 weeks
Phase 2	Data preparation and feature engineering	3–4 weeks
Phase 3	Model development	5–6 weeks
Phase 4	Results analysis and interpretation	2–3 weeks
Phase 5	Writing and revision	3–4 weeks

Table 1: Project planning by phase

From an economic point of view, as shown in Table 2, the main cost of the project is associated with human effort, since most of the work involved data analysis, model development, evaluation, and report writing. Computational and software costs were comparatively limited because the project relied mainly on standard analytical tools and a conventional working environment. Therefore, the estimated value of the project is driven primarily by specialized analytical work rather than by expensive infrastructure or proprietary software.

Cost component	Estimated value
Human work	€4,400

Computational resources	€150
Software and tools	€100
Total estimated cost	**€4,650**

Table 2: Economic estimation of the project

Chapter 5. CONFIGURATION OF THE PLATFORM

This study is built around the principle that high-impact rail disruptions are best analyzed at the incident level, where each event can be described, quantified, and linked to operational outcomes. The project uses Infrabel's open-data dataset on the "most important incidents" in terms of punctuality impact [[1]]. Each record represents a major incident and includes a timestamp, a free-text description, and structured impact variables such as minutes of delay and the number of cancelled trains, together with contextual identifiers such as line number and location. Because the dataset focuses on high-impact events rather than routine variability, it is particularly suitable for analyzing the tail of disruption severity—the subset of incidents that dominates recovery workload and cost.

The data preparation workflow has two objectives. First, it transforms heterogeneous operational records, including multilingual fields and noisy text, into a consistent modelling table through standardization, canonicalization, and task-aware missingness handling. Second, it preserves deployment realism by enforcing a temporal train/validation/test protocol and by preventing leakage in both engineered features and weak labels. Records missing essential inputs for a given task are excluded from that modelling dataset when they cannot be safely recovered, while still being retained where appropriate for descriptive summaries.

As indicated in the Methodology, the project is structured into three connected tasks that share the same incident-level backbone:

- **Task A**: classify whether an incident is catenary-related (binary classification).
- **Task B**: predict operational impact (delay minutes and cancellations) and derive interpretable line clusters.
- **Task C**: translate operational impact into a transparent cost function and generate KPI dashboards; compare cost prediction strategies (chained vs direct).

Together, these steps ensure that the resulting dataset is multilingual-safe, reproducible, and evaluated under conditions that approximate real-world deployment.

5.1 DATA SOURCES AND SCOPE

The project relies on a single primary data source: Infrabel's open-data dataset "Most important incidents in terms of impact on trains' punctuality" [[1]]. The dataset is organized at incident level: each row corresponds to a major incident and reports both context (where/which line, description) and measured operational impact (minutes of delay and number of cancellations). The dataset is explicitly designed to capture the tail of disruption severity: it provides, for each month, the set of incidents whose combined impact exceeds 1,000 minutes of delay to passenger trains.

In practical terms, this dataset is ideal for an operations-oriented pipeline because:

- It links **unstructured incident reporting** (free-text descriptions) with **hard impact outcomes** (delay minutes and cancellations).
- It contains the identifiers required to analyze disruption concentration (by **line** and **place**).
- It targets high-impact events, which dominate recovery effort and cost.

The following Table 3 summarizes the key variables used in this study:

Columns	Type	Description	Used in
month	Categorical / date	Month in which the incident is accounted for in the "major incidents" list.	Descriptives, KPIs
incident_date	Datetime	Date/time of the incident; used to sort	A, B, C

		events and build temporal splits.	
line	Categorical	Line number / corridor where the incident occurred.	A, B, C (and clustering)
place	Categorical	Place name/location associated with the incident.	A, B, C
incident_description	Text	Free-text operational description of the incident.	A (core), B/C (features)
minutes_of_delay	Numeric	Delay minutes attributed to the incident (passenger trains).	B (target), C (cost)
number_of_cancelled_trains	Numeric	Number of cancelled passenger trains attributed to the incident.	B (target), C (cost)

Table 3: Main fields in the Infrabel "high-impact incidents" dataset [1]

The scope is intentionally constrained to ensure methodological clarity and realistic evaluation:

- **Single-source modelling**: the ML pipeline is trained and evaluated using only the incident-level dataset above; no external data sources (weather, demand, telemetry, sensor streams) are integrated in the current implementation.
- **Task-aware missingness**:
    - Task A requires a usable incident description; missing descriptions are excluded from Task A modelling.

- Task B and Task C require both delay_minutes and cancellations; incidents missing these fields are excluded from those modelling tasks.

Focus on high-impact incidents: results should be interpreted as applicable to the subset of disruptions captured by the dataset's "major incidents" definition, not necessarily to all minor disturbances.

5.2 DATA CLEANING & CANONICALIZATION

Operational incident logs are typically recorded under time pressure and later exported through reporting systems, which often introduce inconsistencies that must be resolved before modelling. Raw extracts may contain non-standard column names, duplicated fields, multilingual variants of the same concept, mixed formatting conventions, and missing values. The objective of this stage is therefore to produce a stable, multilingual-safe incident table that can be reused consistently across Tasks A to C without introducing leakage or artificial variation across exports.

5.2.1 SCHEMA NORMALIZATION, TIME PARSING, AND CANONICAL FIELDS

The pipeline first stabilizes the input schema to ensure robustness to minor export differences. Column names are normalized to a consistent snake_case format, and duplicated headers are made unique to avoid silent overwrites during ingestion. Key operational impact variables, such as delay minutes and cancellations, are explicitly coerced to numeric format, with non-parsable values mapped to missing entries.

Incident timestamps are parsed using a tolerant strategy, and a canonical chronological reference, date_naive, is created for temporal ordering and train/validation/test splitting. From this field, additional temporal variables such as year, month, and a monthly period label are derived for descriptive analysis and later KPI aggregation.

The central component of the cleaning stage is the canonicalization of multilingual or duplicated fields. In the Infrabel dataset, the same concept may appear in multiple columns due to language variants or export configuration, especially for description, place, line, and incident category. Feeding these raw variants directly into the models would fragment the feature space and risk encoding language-specific artefacts rather than the underlying incident semantics. To address this, the pipeline constructs a single canonical version for each core concept: desc_canon, place_canon, line_canon, and, where available, category_canon. Candidate columns are identified by pattern matching, the most complete variant is selected, and the remaining variants are used only to fill missing values. This produces a coherent and reproducible representation of each concept while preserving data completeness.

5.2.2 TEXT CLEANING

Incident descriptions are treated as noisy operational text and undergo light normalization aimed at removing artifacts without altering their semantic content. The pipeline trims whitespace removes placeholder strings such as literal "nan" or "none" and otherwise preserves the original wording as far as possible. For Task A, where weak labels may be derived from text keywords, an additional anti-leakage step can be applied by removing those same keywords from the input description before feature extraction. This prevents the classifier from trivially recovering the label from the original keyword itself.

Missing values are handled according to the requirements of each task. Task A requires a valid canonical description, so records missing desc_canon are excluded from the classification dataset. Tasks B and C require valid values for both delay_minutes and cancellations, and records missing either variable are excluded from impact and cost modelling. This task-aware policy avoids unjustified imputations while ensuring that each modelling dataset remains internally consistent. Records excluded from one task may still be retained for descriptive analysis when the missing fields are not required.

At the end of this stage, the raw exports are transformed into a modelling-ready table containing canonical multilingual-safe fields, a stable time reference, and consistently formatted numeric impact variables. This cleaned dataset serves as the common backbone for Tasks A to C and supports both robust modelling and transparent KPI reporting in later sections.

5.3 FEATURE ENGINEERING

Feature engineering is designed to capture three complementary sources of signal: the semantic content of incident descriptions, the structured operational context in which incidents occur, and lightweight historical indicators summarizing recent incident activity. The guiding principle is operational realism: all features should be computable at the time an incident is logged, using only information available up to that moment.

5.3.1 CORE FEATURE FAMILIES

The first feature family is based on text. The canonical incident description is transformed into numeric features using TF-IDF vectorization with unigrams and bigrams, which allows the models to capture both key terms and short operational phrases. For pipelines requiring dense inputs, the high-dimensional sparse representation is compressed using Truncated SVD, providing a lower-dimensional embedding that remains computationally efficient while preserving most of the informative variation.

The second feature family captures structured operational context. The canonical line identifier and canonical place are encoded through one-hot encoding with unknown handling, ensuring robustness to unseen values at inference time. These variables are operationally relevant because disruption patterns often vary across corridors, locations, and local infrastructure conditions.

The third feature family represents broad seasonality. A categorical season variable is derived from the incident timestamp to capture systematic calendar effects without overfitting to specific dates. This provides a lightweight proxy for recurrent seasonal influences on infrastructure performance.

The fourth feature family consists of past-only activity indicators. To summarize recent operational pressure around a line or location, the pipeline computes cumulative incident counts to the event time and rolling counts over the previous 30 days. These features are constructed strictly from earlier events, thereby preventing temporal leakage and ensuring that no future information is used in prediction.

5.3.2 DERIVED LINE CLUSTERS

In addition to the core feature families, Task B derives an interpretable segmentation of railway lines using KMeans clustering [21] on historical line-level aggregates, such as incident frequency and average impact. This unsupervised grouping makes it possible to identify line profiles with similar operational behavior and disruption patterns. The resulting line_cluster label is primarily used in Task C to structure KPI reporting and cost analysis, while the cluster centroids provide a compact summary of typical line profiles.

Feature usage then varies slightly across tasks. Task A uses incident descriptions together with contextual variables to classify catenary-related incidents; when weak labels are keyword-based, the corresponding keywords are removed from the input text to reduce label leakage. Task B combines text-derived, contextual, seasonal, and past-only activity features to predict delay minutes and cancellations, evaluating both raw-target and log1p-transformed regressors because of the heavy-tailed nature of the impact variables. Task C uses the observed or predicted operational impacts to compute incident-level cost and aggregate KPI views by time period, line, place, and line cluster. In this final task, both a chained strategy, based on Task B predictions, and a direct cost-prediction strategy are evaluated.

5.4 TRAIN/VALIDATION/TEST PROTOCOL AND LEAKAGE PREVENTION

The evaluation protocol is designed to mimic real deployment: models should be trained on historical data and evaluated on future incidents. For this reason, the study avoids random splits and instead applies chronological (temporal) splits based on the incident timestamp:

- **Train**: earliest portion of the timeline.
- **Test**: most recent portion of the timeline.

For Task A, where threshold selection is important under class imbalance, an additional validation split is created within the training window:

- **Train-inner**: earlier subset of train.
- **Validation**: later subset of train.
- **Test**: held out most recent window.

This structure ensures that no part of the test period influences either model fitting or threshold tuning.

Furthermore, rolling and cumulative counts are calculated using only past events. This ensures the features reflect what would be known at the time of the incident and prevents the model from unintentionally using future information.

In addition to these safeguards, Task A applies further controls to ensure a fair and deployment-realistic evaluation, addressing label leakage, decision-threshold selection, and class imbalance.

Task A carries a specific leakage risk because the catenary label is generated by searching for catenary-related keywords. If those same keywords remain in the text used as input features, the model can appear unrealistically accurate by relying on a simple shortcut (detecting the keyword rather than learning meaningful patterns). To avoid this, a

conservative approach is applied: when a suitable incident category field is available, labels are taken from that field; otherwise, catenary keywords are removed from the description text before TF-IDF feature extraction.

Because catenary incidents can be a minority class, a fixed decision threshold of 0.5 is not assumed to be optimal. The classification threshold is tuned on the validation split to maximize F1, and performance is then reported once on the untouched test set using the selected threshold. Finally, class imbalance is handled with SMOTE applied in a reduced SVD space, and resampling is restricted to the training data only (never validation or test).

With a consistent, multilingual-safe incident table, engineered features that reflect operational context, and a temporal evaluation protocol that prevents leakage, the next chapter details the modelling choices, hyperparameters, and evaluation outputs for Tasks A to C.

Chapter 6. SYSTEM DEVELOPMENT

This study follows a quantitative, explanatory methodology that connects operational incident records to service-performance and economic outcomes through an integrated machine-learning and cost-analytics framework. The approach is structured as an end-to-end pipeline: incident-level data are first standardized and transformed into a modelling-ready table; machine-learning models are then trained and evaluated under a temporal protocol to estimate whether an incident is catenary-related and its operational impact in terms of delay minutes and cancellations; these operational outputs are subsequently translated into incident-level costs and decision-oriented KPIs; finally, scenario and sensitivity analyses are used to quantify the potential benefits of mitigation strategies under different cost assumptions. Throughout the pipeline, strict leakage prevention and chronological validation are enforced to approximate real deployment conditions and to ensure that reported performance reflects out-of-sample behaviour.

6.1 TASK A: CATENARY INCIDENT CLASSIFICATION

The first planned task addresses the first modelling objective: identifying which high-impact incidents are likely to be catenary-related. The output of Task A is a calibrated probability score per incident, enabling flexible operational use. The task is framed as a binary classification problem, trained on incident descriptions and contextual identifiers (line and location), and evaluated under a temporal train/validation/test split.

Catenary ground truth labels are not explicitly provided as a dedicated field in all exports. Therefore, labels are generated using a keyword-based weak-labelling approach that searches for catenary-related terms (including multilingual variants). This introduces a specific leakage risk: if the same keywords are left in the input text, the classifier can achieve artificially high scores by learning a trivial shortcut. To prevent this, the implementation

removes catenary keyword patterns from the text used as input features before vectorization, forcing the model to learn contextual language and co-occurring terms rather than directly detecting the label token.

The pipeline is designed to remain interpretable while performing well on noisy operational text:

1. **Text representation:** incident descriptions are converted into TF-IDF features using unigrams and bigrams.
2. **Dimensionality reduction:** Truncated SVD compresses the high-dimensional sparse text representation into a dense embedding (30 components), improving computational efficiency.
3. **Class imbalance handling:** SMOTE is applied in the SVD space to balance the minority class, and resampling is performed strictly on the training data only.
4. **Classifier:** Logistic Regression is trained as the main model, providing a strong, interpretable baseline for text classification.

Because catenary incidents belong to the minority class, the default threshold of 0.5 may not be optimal. Therefore, the classification threshold is tuned on the validation split by selecting the value that maximizes F1-score, a metric well suited to imbalanced problems because it balances precision and recall. The selected threshold is then kept fixed for evaluation on the untouched test set. In the reported experiments, this threshold was 0.45. Performance is reported using ROC-AUC and PR-AUC as threshold-independent measures of class discrimination, together with precision, recall, and F1-score as threshold-dependent measures that capture the practical trade-offs of the final classifier.

6.2 TASK B: IMPACT PREDICTION AND LINE CLUSTERING

The second task estimates the operational impact of each high-impact incident. Two complementary targets are modelled:

- **Delay minutes:** a continuous outcome with a pronounced right-skew and extreme values.
- **Cancellations:** a non-negative count outcome, also heavy-tailed and zero-inflated in many operational settings.

The purpose of this task is twofold. First, it provides quantitative impact estimates that can support operational decision-making and subsequent cost modelling (Task C). Second, it generates an interpretable segmentation of the network by clustering lines into a small set of impact profiles (C0/C1/C2), enabling KPI reporting.

As indicated, both targets exhibit heavy-tailed distributions, where a small number of incidents account for a large share of total disruption. As a result, evaluation focuses primarily on error measures that remain meaningful under skew:

- **MAE** (mean absolute error) as the main selection criterion, due to robustness and interpretability in operational units.
- **RMSE** (root mean squared error) reported to reflect sensitivity to extreme events.
- **R^2** reported as a secondary indicator; in heavy-tailed settings it can be low or even negative on a temporal test split, which does not necessarily imply that the model is unusable but rather that a small number of extreme events dominates variance.
- A simple **prediction-interval coverage proxy** (PI80 coverage) is estimated using the distribution of training residuals, providing an additional uncertainty-oriented diagnostic.

6.2.1 IMPACT PREDICTION

This task combines a feature engineering stage with a predictive modelling strategy designed to reflect both the structure of the incident log and the statistical properties of the target variables. The input space includes a mixed feature set aimed at balancing predictive performance and operational realism. First, text information is incorporated through TF-IDF features extracted from the canonical incident description and then compressed using Truncated SVD to obtain a compact dense representation. Second, contextual information is captured through one-hot encoded line and place identifiers. Third, seasonality is represented by a categorical feature derived from the incident month. Finally, the model includes past-only activity indicators, namely cumulative and 30-day rolling incident counts for both line and place (*_cum, *_30d), computed strictly from earlier events to prevent leakage. Together, these variables allow the model to exploit semantic cues from the description while also capturing recurring disruption patterns associated with specific corridors, locations, and time periods.

On the modelling side, particular attention is given to the heavy-tailed nature of both target variables, namely delay minutes and cancellations. For this reason, Task B evaluates models both on the raw target scale and under a log1p target transformation. This is implemented through a transformed-target regression approach, in which models are trained to predict log1p(y) and predictions are later mapped back to the original scale using expm1. The final outputs are clipped to the valid non-negative domain to ensure operationally coherent predictions. This transformation reduces the influence of extreme outliers during training and often improves generalization on the temporal test set. Multiple regressors are then compared under the same feature pipeline, including models such as Ridge and Random Forest, both with and without the log1p transformation, and the final selection is based on test-oriented metrics, with MAE as the main criterion.

6.2.2 LINE CLUSTERING WITH KMEANS

In addition to point prediction, Task B produces an interpretable network segmentation by clustering lines into a small number of groups. The clustering is performed using KMeans (k equal to 3) trained on line-level aggregates computed from the training window only, such as:

- number of incidents per line,
- average delay per line,
- average cancellations per line.

The resulting clusters are mapped to ordered labels (C0, C1, C2) based on an impact-oriented ranking, and their centroids are extracted in the original scale for interpretation. This yields an operationally meaningful categorization, for example, distinguishing lines characterized by low incident frequency, high recurrence, or high severity. These clusters are subsequently used in Task C to structure KPI reporting (cost concentration by cluster).

6.3 TASK C: COST MODEL & KPIS

The third task translates the operational consequences of incidents into monetary estimates and decision-oriented indicators. Its purpose is twofold. First, it provides an economic interpretation of the disruption generated by each incident through a transparent cost model. Second, it aggregates those incident-level estimates into operational KPIs that can support monitoring, prioritization and scenario analysis across lines, locations and time periods. Since detailed internal accounting data for Infrabel is not publicly available, the economic framework is explicitly treated as a scenario-based model rather than as an exact accounting reconstruction. The goal is therefore not to recover the precise internal cost of each event, but to provide a reproducible methodology for estimating economic impact under clearly stated assumptions.

6.3.1 COST FUNCTION DEFINITION AND SCENARIO ASSUMPTIONS

The economic model used in this work monetizes the operational impact of each incident as:

$$C_{\text{incident}} = C_{\text{repairs}} + c_{\text{min}} \cdot \text{delay\_minutes} + c_{\text{cancel}} \cdot \text{cancellations} + C_{\text{penalties}} - C_{\text{recovered\_revenue}}$$

where c_{min} is the total cost per minute of delay (€/min), c_{cancel} is the cost per cancelled train (€/cancellation), and $C_{repairs}$, $C_{penalties}$, and $C_{recovered\_revenue}$ represent, respectively, the average fixed repair cost per incident, contractual penalties, and the revenue that may be recovered through insurance or compensation from third parties.

In the implementation, the following base values are used:

```
c_repairs            = 0.0      # €/incident
c_per_minute         = 100.0    # €/minute of delay
c_per_cancellation   = 5000.0   # €/cancelled train
c_penalties          = 0.0      # €/incident
c_recovered_revenue  = 0.0      # €/incident
```

These parameters are not intended to reproduce Infrabel's internal accounting structure exactly. Instead, they define a reasonable baseline scenario grounded in external references and in the European passenger-rights framework. The cost per minute of delay, $c_{min} =$ 100€/min, is chosen as a plausible order-of-magnitude estimate for the total burden of delay

in a high-capacity transport system. It includes not only passenger compensation, but also direct operating costs and part of the knock-on effects generated across the network.

A useful lower bound can be obtained from the rail passenger-rights framework. For example, assuming an average ticket price of 20 € and 250 passengers on board, a 60-minute delay would imply a minimum compensation of:

$$0.25 \times 20 \times 250 = 1250\ €$$

which is equivalent to approximately 21 €/minute in passenger compensation alone. Since this amount excludes staff costs, energy, rolling-stock disruption, missed connections and wider congestion effects, the total cost per minute of delay is expected to be higher. For that reason, the value of 100 €/minute is adopted as the base scenario. It should be interpreted as a bundled and conservative approximation of total delay impact rather than as a direct accounting figure.

The cost per cancellation, $c_{cancel} = 5000$€/cancelled train, is derived from a similarly simple and conservative assumption. If a cancelled train carries 250 passengers with an average fare of 20 €, the direct refund cost is:

$$20 \times 250 = 5000\ €$$

This provides a baseline estimate for the cost of cancellation. As with the delay coefficient, this value does not include all possible indirect effects, such as passenger re-routing, replacement transport, reputational damage or wider network disruption. The true cost may therefore be higher in many practical cases.

The remaining fixed terms are set to zero in the base scenario. This choice does not imply that repair costs, penalties or recovered revenues are negligible, but rather that sufficiently detailed public information is not available for the Infrabel case. Keeping these terms explicit

in the formulation preserves the extensibility of the model, since future scenario analyses may incorporate fixed repair costs for severe incidents, penalty schemes linked to service thresholds, or partial revenue recovery.

Because the main cost coefficients are based on external benchmarks and regulatory logic rather than on operator-specific financial records, the model is explicitly interpreted as scenario-based. In the software implementation, the economic parameters can be modified easily, which allows sensitivity analyses under alternative assumptions. The following reference scenarios are considered:

- **Low scenario:** 30 €/min and 3,000 €/cancellation
- **Base scenario:** 100 €/min and 5,000 €/cancellation
- **High scenario:** 150 €/min and 8,000 €/cancellation

This makes it clear that the resulting annual or monthly cost estimates are conditional on explicit assumptions. The contribution of the thesis therefore lies less in claiming a single exact euro value and more in providing a structured methodology for comparing scenarios and identifying the lines, locations and periods associated with the highest expected economic burden.

6.3.2 KPIS AND AGGREGATION FRAMEWORK

Once incident-level cost has been computed or predicted, the framework derives a set of KPIs for descriptive analysis and managerial prioritization. At incident level, the main output is the estimated total cost per event. At aggregate level, the methodology computes total cost, average cost per incident, total delay minutes, total cancellations, and cost shares by line, place and period.

To support operational monitoring, these quantities are aggregated at both monthly and quarterly levels. Monthly aggregation is useful for identifying short-term peaks, seasonal

effects and recurrent disruption periods, while quarterly aggregation provides a more stable view for strategic reporting and comparison. This dual temporal view helps capture both short-run volatility and broader trends in service impact.

In addition, the framework incorporates Pareto-style analyses to assess concentration effects. Incidents, lines or locations can be ranked by estimated cost in descending order, and cumulative shares of the total burden can then be analyzed. This makes it possible to determine, for example, whether a relatively small subset of hotspots accounts for a disproportionate share of the total disruption cost. Such concentration analyses are especially relevant in railway operations, where recurrent incidents in a limited number of corridors or places may dominate the overall economic impact.

Overall, the KPI layer transforms incident-level predictions into actionable summaries that support benchmarking, prioritization and resource allocation. Rather than focusing only on the number of incidents, the framework highlights where the economic burden is concentrated and where preventive or corrective interventions may have the greatest expected value.

6.3.3 COST PREDICTION STRATEGIES: CHAINED VERSUS DIRECT MODELLING

This task evaluates two alternative strategies for estimating incident cost. The first is a chained approach, in which the outputs predicted in Task B are used as intermediate inputs to the cost function. Under this approach, delay minutes and cancellations are first estimated through the predictive models developed in Task B, and the final cost is then obtained by applying the economic equation:

$$\hat{C}_{incident} = C_{repairs} + c_{min} \cdot \widehat{delay\_minutes} + c_{cancel} \cdot \widehat{cancellations} + C_{penalties} - C_{recovered\_revenue}$$

This approach has an important interpretability advantage, since the economic estimate remains explicitly linked to operational quantities that are meaningful in practice. It also makes scenario analysis straightforward, because alternative economic assumptions can be tested by changing the coefficients of the cost function without retraining the predictive models.

The second strategy is a direct approach, in which a regression model is trained to predict incident cost itself as the target variable. This formulation may capture nonlinear interactions between the drivers of disruption and the final monetary impact more directly. However, it is less transparent than the chained approach, because the prediction is no longer decomposed into delay and cancellation components. It is also less flexible from a scenario-analysis perspective, since any change in the cost assumptions requires redefining the target and retraining the model.

For these reasons, the chained strategy is conceptually preferred as the main methodology of the thesis, while the direct cost model is used as a benchmark. The comparison between both approaches is performed not only at incident level, but also at aggregate level, evaluating how well they reproduce total monthly and quarterly economic burden. This is important because, in practical decision-making, the ability to rank periods, lines and locations consistently may be as relevant as the predicted accuracy of individual events.

This methodological framework provides the basis for the empirical evaluation presented in the next chapter. The Results section reports the performance obtained in each task, from incident classification and impact prediction to the estimation of economic cost and aggregated KPIs. In addition to standard predictive metrics, special attention is paid to the practical usefulness of the models, particularly their ability to reproduce cost concentration patterns and to rank lines, locations and time periods according to expected operational and economic impact

Chapter 7. RESULTS ANALYSIS

This chapter presents the results obtained for the three analytical tasks developed in the project. The objective is to evaluate the performance of the proposed framework and to interpret its outputs in relation to railway disruption analysis.

7.1 TASK A RESULTS

Task A evaluates the ability of the proposed classifier to identify catenary-related incidents from operational logs under a realistic chronological split. As shown in Figure 7, the dataset is strongly imbalanced, but the prevalence of the positive class remains very stable across the different partitions. The positive class represents 7.39% of the model-selection training subset, 7.35% of the validation split, and 7.65% of the temporally held-out test split. This stability is relevant because it suggests that the chronological partition preserves the rarity of catenary incidents without introducing an artificial distribution shift. The main quantitative results for Task A are summarized in Table 4, Table 5 and Table 6.

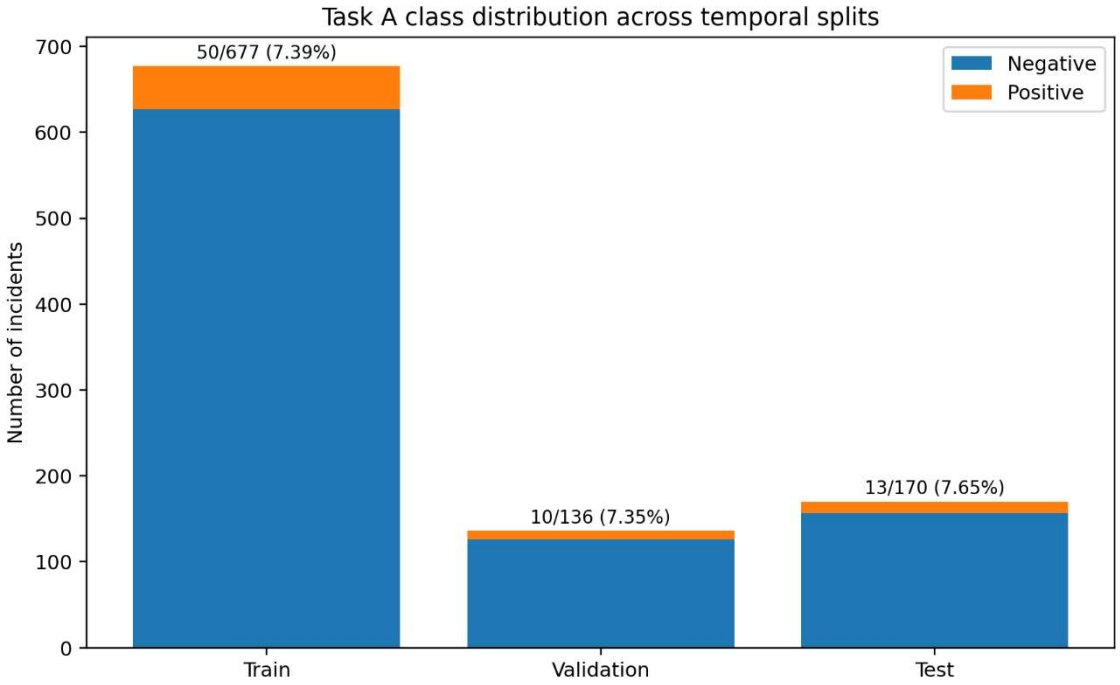

Figure 7: Class distribution across temporal splits

Split	N	Positives	Negatives	Positive rate
Train (inner)	541	40	501	0.0739
Validation	136	10	126	0.0735
Test	170	13	157	0.0765

Table 4: Sample composition by split

Split	ROC-AUC	PR-AUC	Precision	Recall	F1	Brier	Threshold
Train (inner)	0.9898	0.8307	0.7115	0.9250	0.8043	0.0337	0.4500
Validation	0.9802	0.7301	0.6923	0.9000	0.7826	0.0397	0.4500
Test	0.9706	0.7560	0.5294	0.6923	0.6000	0.0455	0.4500

Table 5: Classification performance by Split

Split	TP	FP	TN	FN
Train (inner)	37	15	486	3
Validation	9	4	122	1
Test	9	8	149	4

Table 6: Confusion-matrix counts by split

Model selection retained a logistic-regression-based pipeline with text features, contextual variables, and class rebalancing. On the validation split, the selected model achieved a ROC-AUC of 0.9802 and a PR-AUC of 0.7301, indicating a strong ability to rank catenary incidents ahead of non-catenary ones despite the strong imbalance. Since the final binary decision depends on the operating threshold, the validation threshold-selection process is shown in Figure 8, which plots precision, recall, and F1 as a function of the classification threshold. The selected operating point, τ=0.45, corresponds to the threshold that maximizes validation F1 while preserving a recall-oriented behaviour.

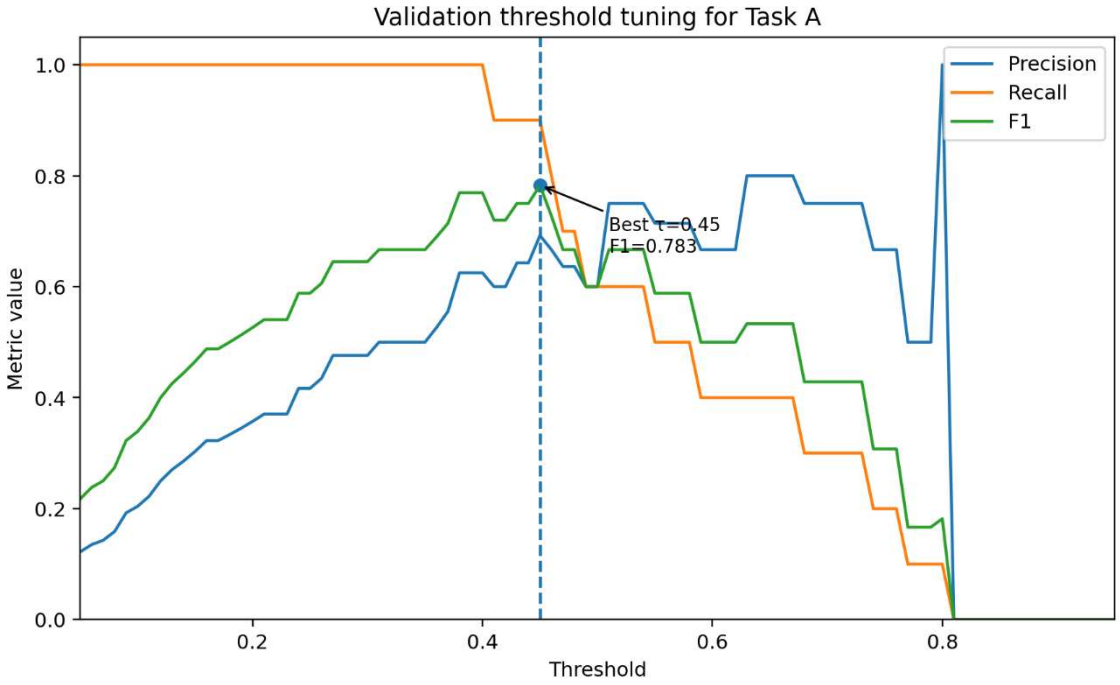

Figure 8: Validation threshold tuning

At this selected threshold, the model reached precision = 0.6923, recall = 0.9000, and F1 = 0.7826 on the validation split, with a Brier score of 0.0397. These results indicate that the classifier is able to recover most positive incidents while keeping the number of false alarms relatively contained. The ranking quality of the model is illustrated from two complementary perspectives. First, Figure 9 shows that both validation and test ROC curves remain close to the upper-left corner, confirming strong discriminative ability. Second, Figure 10 provides a more informative view under class imbalance and shows that the model maintains a favourable precision–recall trade-off in both splits.

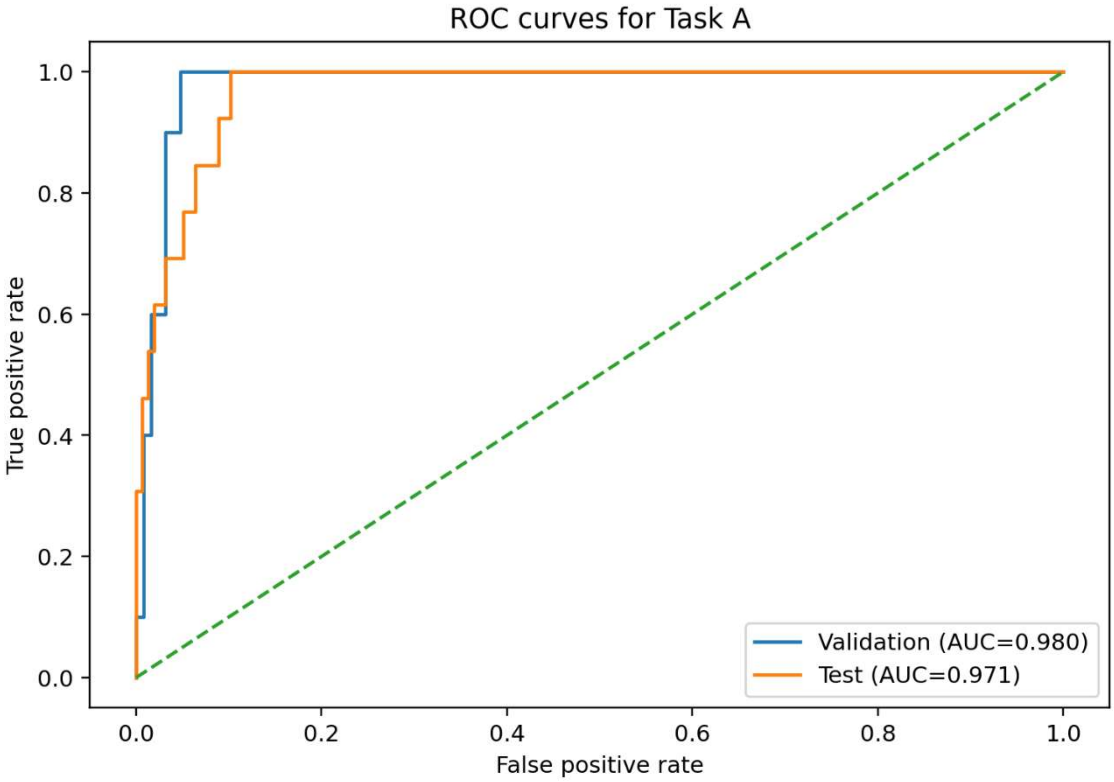

Figure 9: ROC Curve

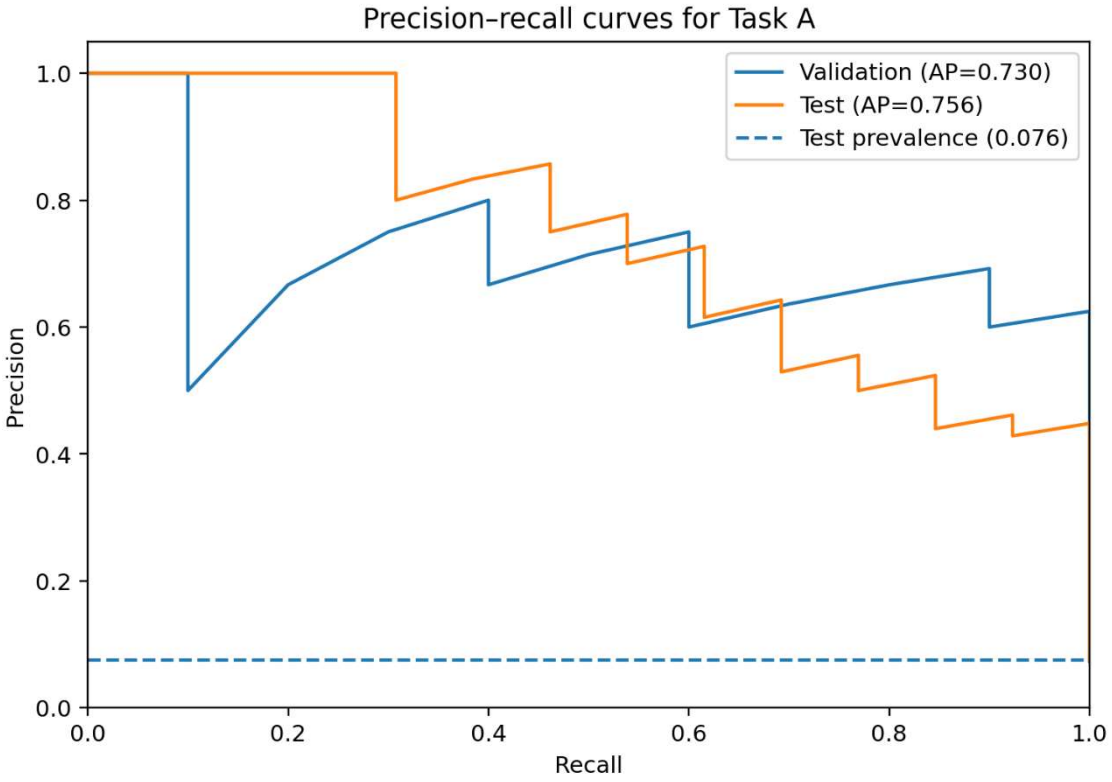

Figure 10: Precision–recall curve

When applied to the test set, the classifier preserves a high-ranking capability. As reported in Table 5, test performance reaches ROC-AUC = 0.9706, PR-AUC = 0.7560, and Brier = 0.0455. The small decrease in ROC-AUC relative to validation is consistent with moderate temporal generalization errors, but overall, the results remain strong. In fact, the slightly higher PR-AUC on the test split suggests that the model continues to separate positive from negative cases effectively in ranking terms, even when evaluated on future incidents.

However, when the fixed threshold selected on validation is applied to the test set, binary classification performance becomes more modest. At τ=0.45, the model obtains precision = 0.5294, recall = 0.6923, and F1 = 0.6000. This means that the classifier still captures most positive incidents, but with a larger number of false positives than in validation. This behaviour can be examined in Figure 11, which shows the distribution of predicted

probabilities in the test horizon. Positive incidents tend to receive clearly higher scores than negative ones, confirming that the model has learned a meaningful separation between both classes. Nevertheless, the overlap between both distributions in the intermediate region explains why some false positives and false negatives remain around the chosen threshold.

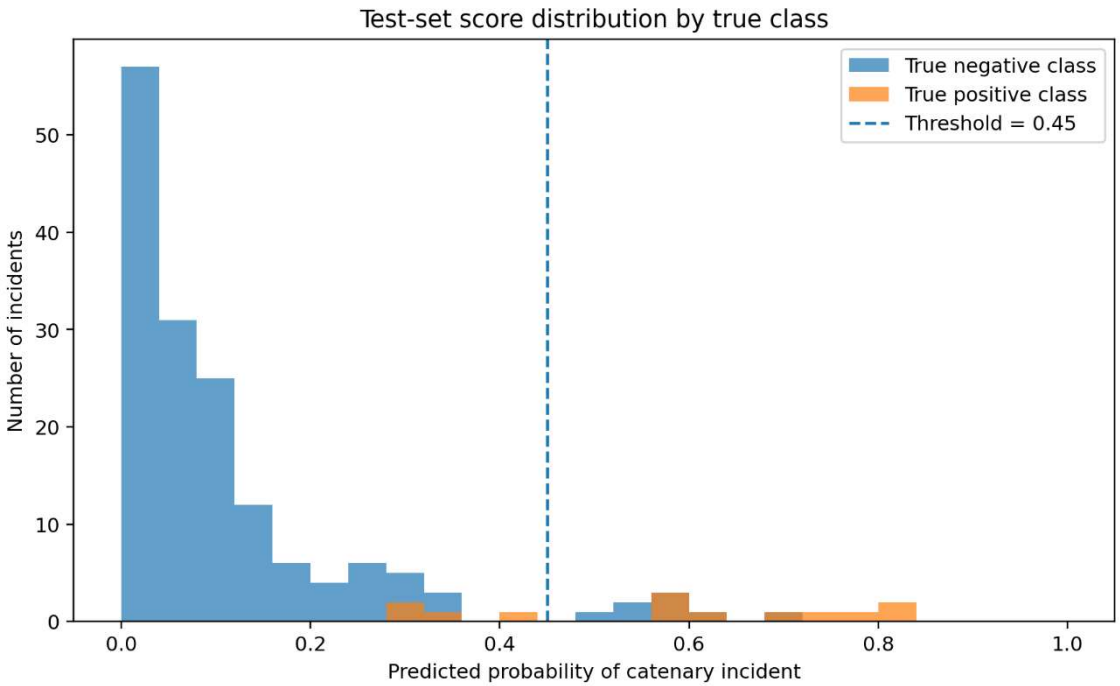

Figure 11: Test score distribution by true class

The practical consequences of this operating point are shown in Figure 12. At the selected threshold, the model produces 9 true positives, 8 false positives, 149 true negatives, and 4 false negatives on the test set. This confirms that the classifier is not a perfect hard filter, but it is still useful as a screening and prioritization tool. It identifies most catenary incidents while keeping the false-alarm volume at a manageable level.

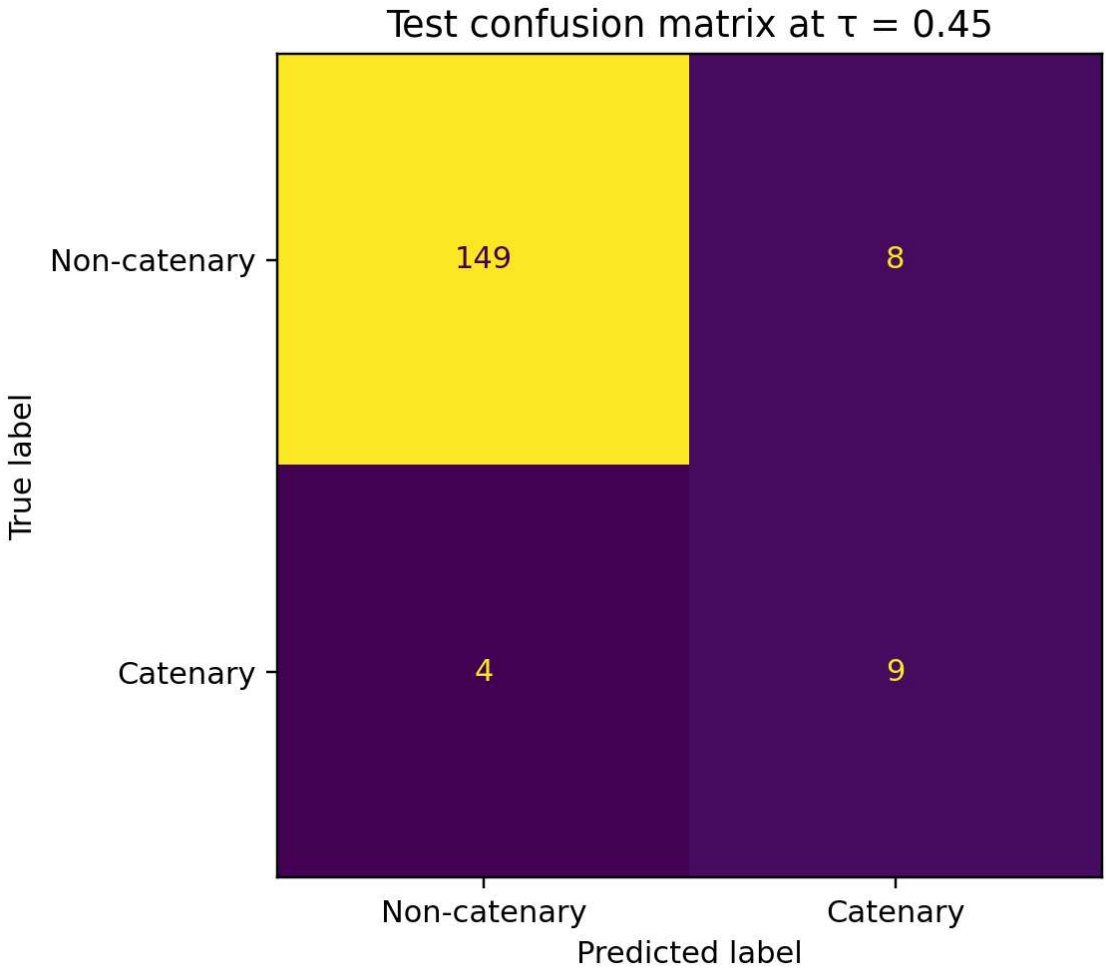

Figure 12: Test confusion matrix

Overall, the results of this task indicate that the proposed classifier is best interpreted as a probabilistic ranking model rather than as a fully reliable binary decision system. Its main strength lies in assigning systematically higher scores to catenary-related incidents than to non-catenary ones, which makes it valuable for downstream prioritization and further review. The selected threshold offers a reasonable balance between recall and precision, but it can also be adjusted: lower thresholds would favor sensitivity, whereas higher thresholds would reduce false alarms.

7.2 TASK B RESULTS

Task B had two complementary objectives: (i) to predict operational impact in terms of delay minutes and cancellations, and (ii) to identify groups of railway lines with similar disruption profiles. The results show that the incident process is highly concentrated in a limited set of lines and places, that predictive performance is acceptable for typical events but weak for extreme disruptions, and that line-level clustering provides a useful operational segmentation of the network.

Before evaluating the regression models, it is useful to examine how incidents are distributed across the network. Incidents are not evenly spread across lines: the ten most affected lines account for 54.4% of all 847 incidents, with lines 50A (70 incidents), 25 (64), 161 (59), and 36 (54) standing out clearly as shown in Figure 13. A similar concentration appears at station or place level, although it is less pronounced: the ten most affected places explain 25.4% of incidents, led by BRUSSEL-ZUID (40 incidents), BRUSSEL-NOORD (25), and GENT-SINT-PIETERS (25) as shown in Figure 14. This concentration justifies the inclusion of cumulative and recent frequency features at both line and place level, because historical exposure is clearly not homogeneous across the network.

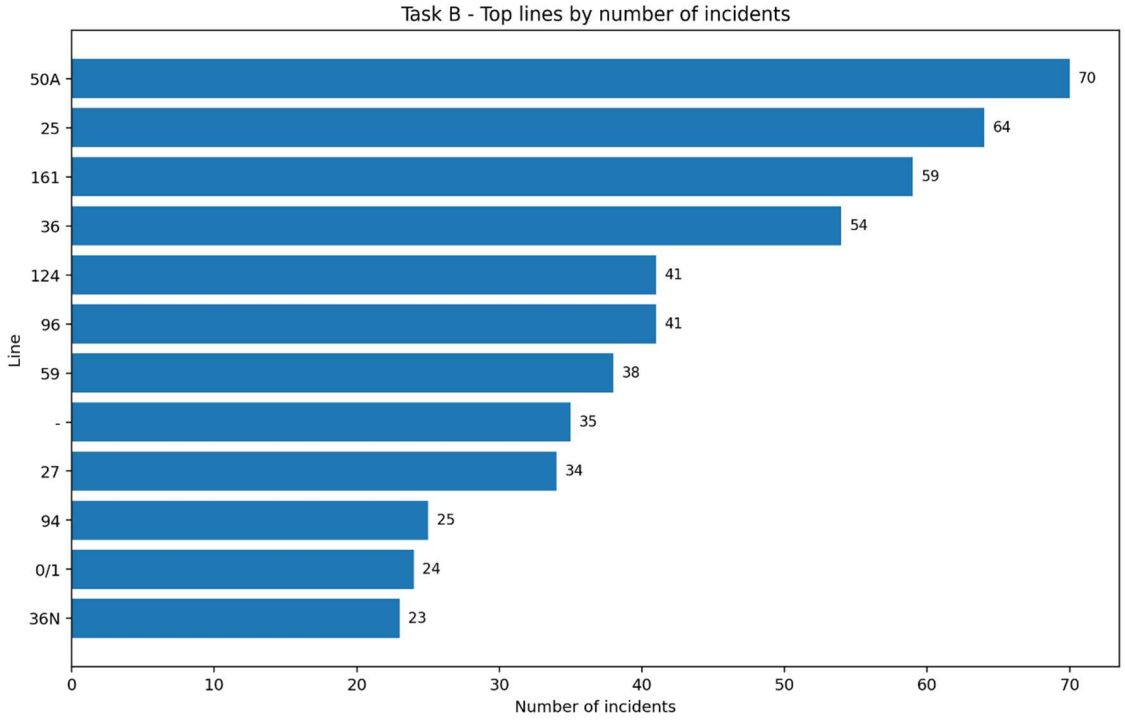

Figure 13: Top lines by number of incidents

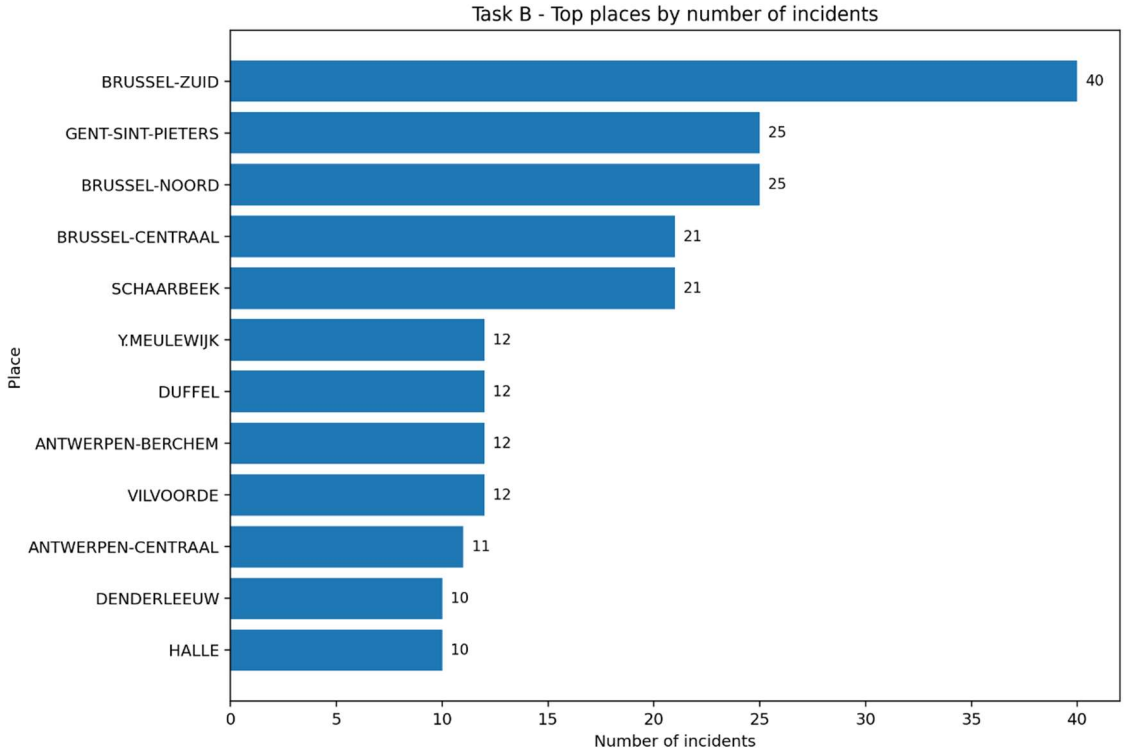

Figure 14: Top places by number of incidents

Model comparison on the test set indicates that the best-performing specification in terms of absolute error is the random-forest model with a log1p target transformation for both regression targets. For delay prediction, this model achieves a test MAE of 877.6 minutes and a test RMSE of 1508.0 minutes as indicated in Figure 15. For cancellation prediction, the same specification obtains a test MAE of 34.8 cancellations and a test RMSE of 66.9 as shown in Figure 16. However, the comparison also reveals that performance remains modest in explanatory terms. For delays, ridge_log1p slightly improves RMSE (1481.8) and is the only specification with a positive test R^2 (0.025), whereas rf_log1p is favoured only when MAE is prioritized. For cancellations, all models yield low or negative test R^2 values, which suggests that cancellations are much noisier and harder to explain with the available structured features. In both targets, the reported PI80 test coverage is consistently above the nominal 80% level (roughly 0.90-0.95), which suggests that the prediction intervals are conservative and contain the observed values.

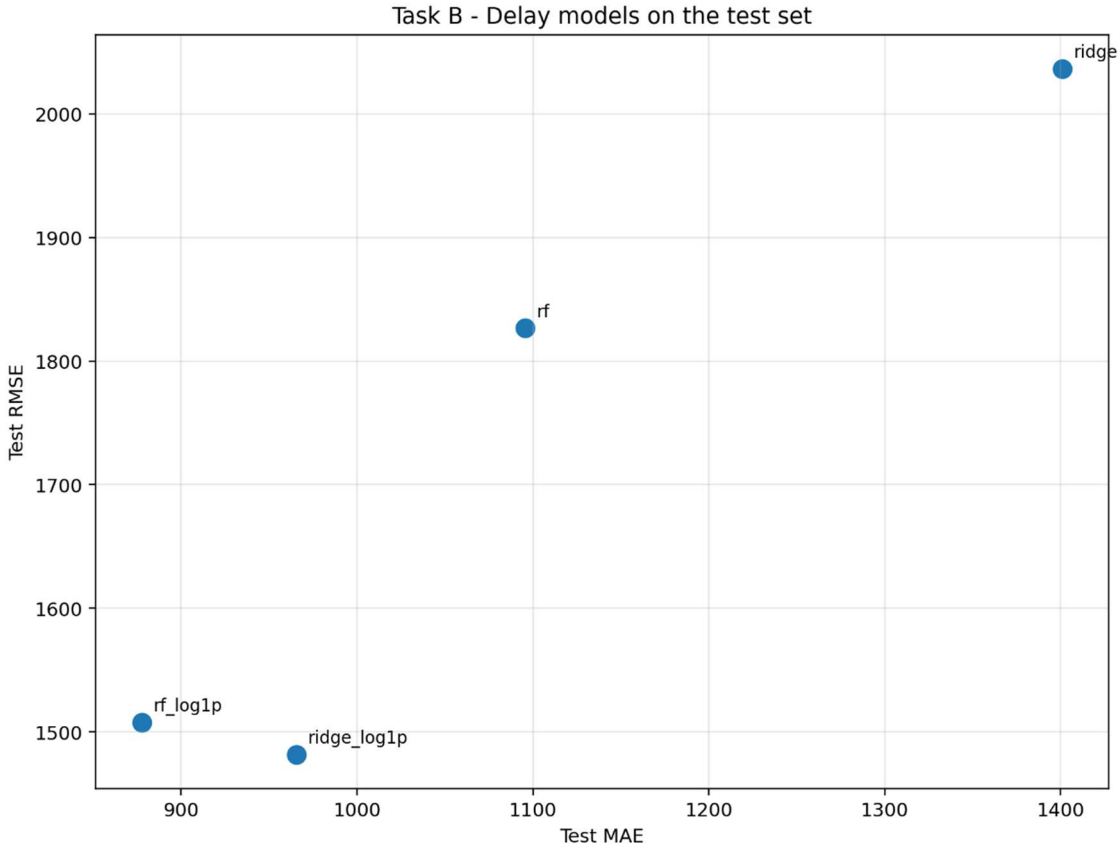

Figure 15: Delay models on test set

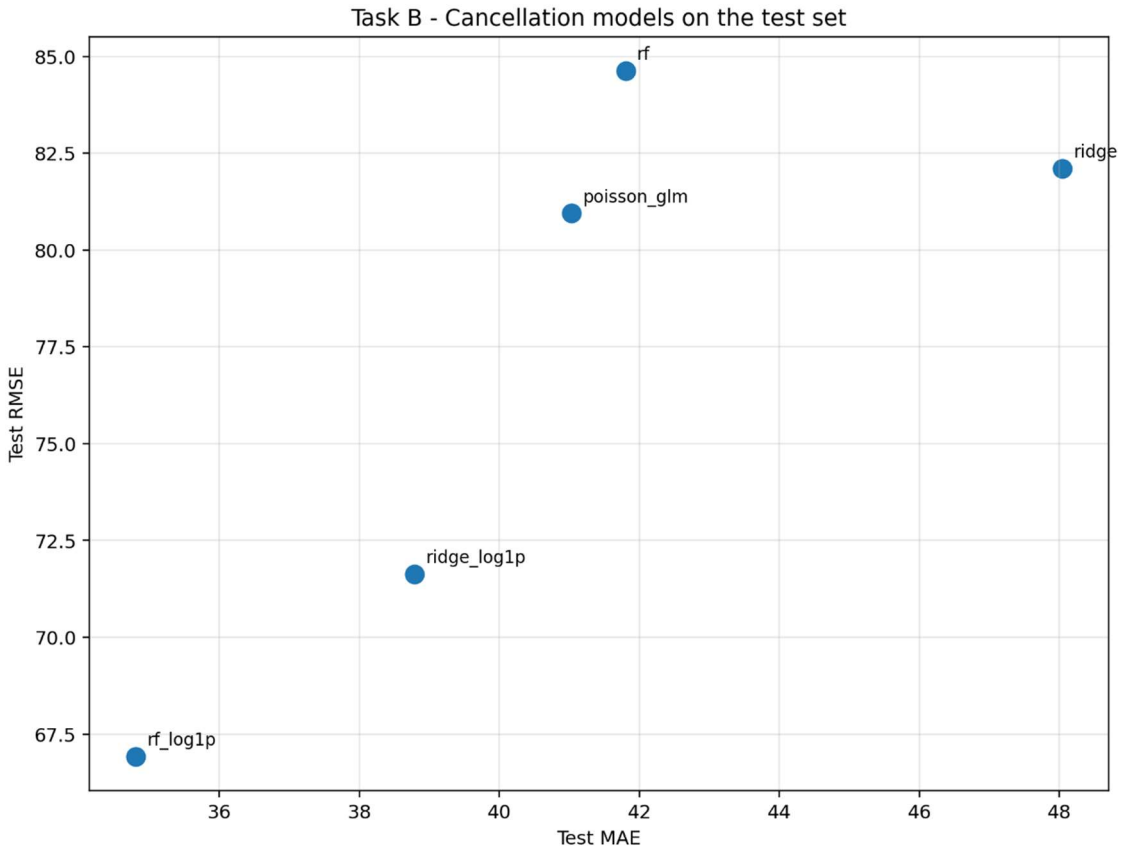

Figure 16: Cancellation models on the test set

The fitted-versus-observed plots confirm that the selected models capture the central mass of the distribution better than the tails. For delays, the cloud of observations is concentrated around roughly 1000-2500 minutes, where predictions are comparatively stable, but the model compresses many high-impact events into a much narrower predicted range, as seen in Figure 17. This effect becomes very clear for severe disruptions above the 90th percentile: the MAE rises from 549.3 minutes in the lower 90% of the test set to 3832.6 minutes in the top decile, with a strong average underprediction of the most disruptive cases. The same pattern appears in cancellations, as in Figure 18. For the lower 90% of cases, the MAE is 21.3 cancellations, but for the top decile it increases to 156.1, again with systematic underestimation of extreme events. Therefore, regression models are useful to approximate

routine operational impact, but they should not be interpreted as reliable estimators of the most severe, low-frequency disruptions.

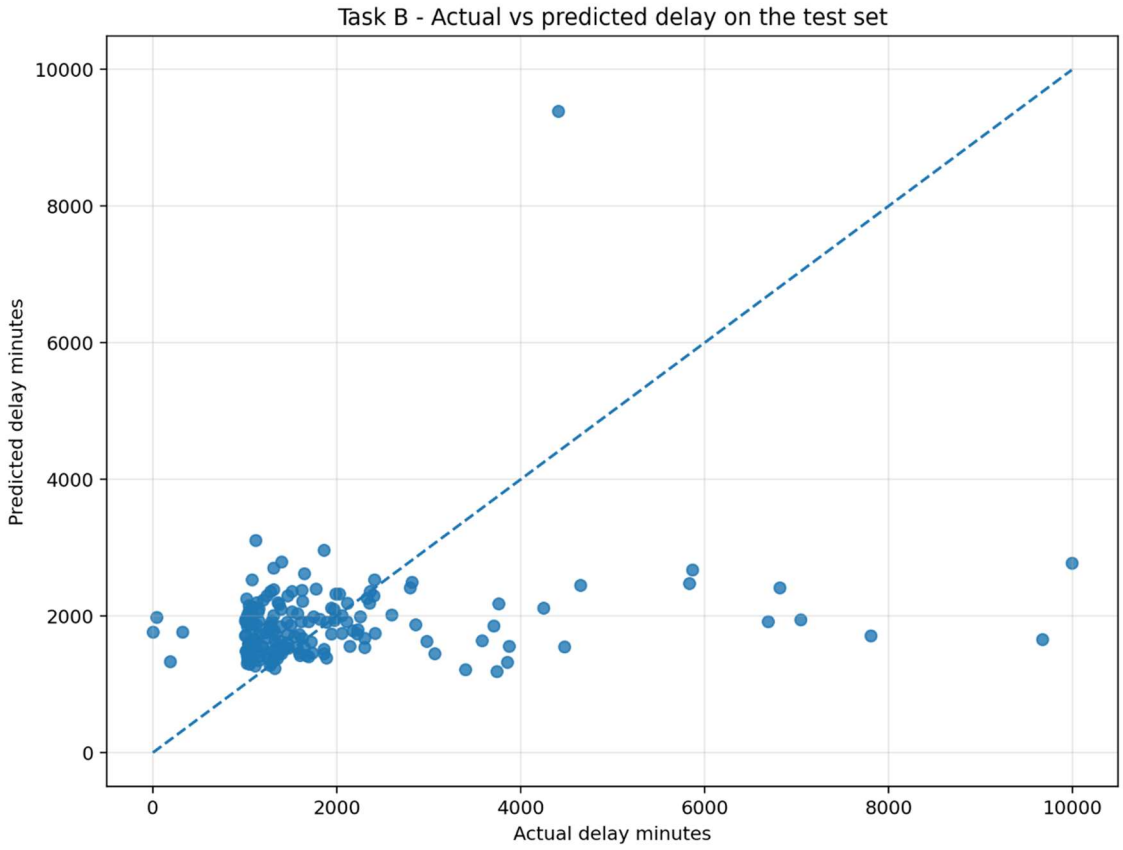

Figure 17: Actual vs predicted delay on the test set

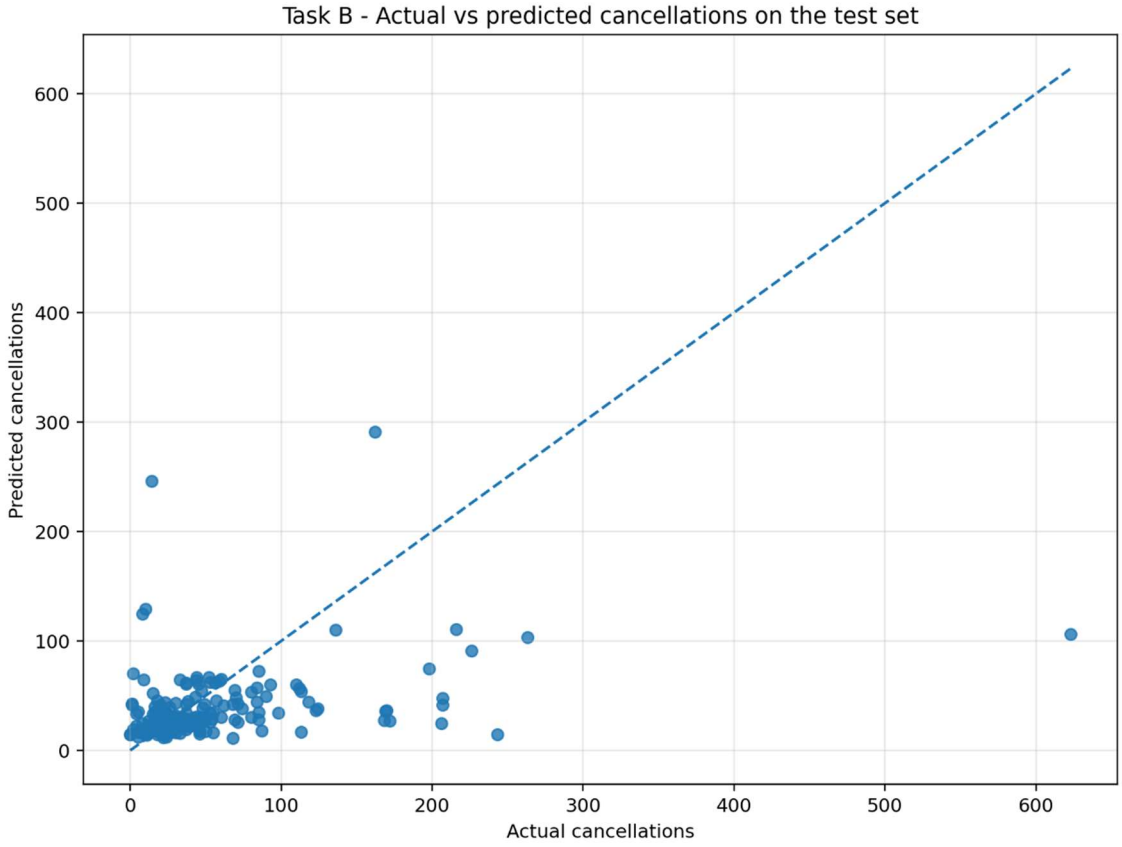

Figure 18: Actual vs predicted cancellations on the test set

The second part of Task B is the clustered lines using their incident frequency and average operational impact. The resulting three-cluster solution shown in Figure 19 is operationally interpretable. Cluster C0 contains 66 out of 77 lines (85.7%) and represents the broad baseline profile of the network: low exposure and moderate consequences, with an average of 4.85 incidents per line, 1672.2 delay minutes, and 39.0 cancellations. Cluster C1 contains only 8 lines (10.4%) but concentrates 48.7% of all incidents, meaning that it captures the lines with recurrent disruption activity but still moderate average severity. Its centroid corresponds to 41.25 incidents per line, 1946.9 delay minutes, and 45.7 cancellations. Finally, Cluster C2 contains only 3 lines (3.9%) and represents a small outlier group characterized by very high severity: 6366.9 mean delay minutes and 313.7 mean cancellations on average, despite a much lower incident frequency than C1. In practical

terms, C1 is the cluster of operationally busy lines that deserve continuous monitoring, while C2 isolates rare but highly disruptive profiles that are especially relevant for risk management and anomaly-oriented analysis.

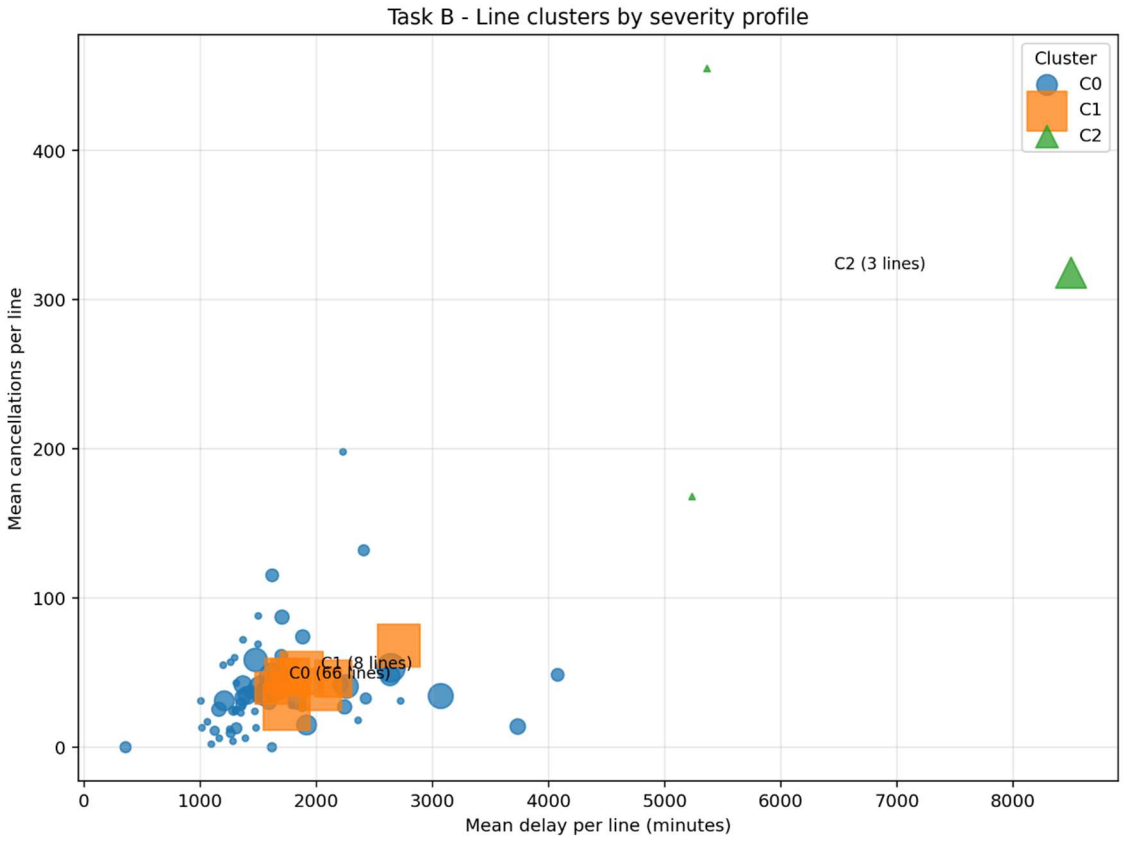

Figure 19: Line clusters

This clustering structure is also consistent with the prediction results. When the test set is segmented by line cluster, the largest errors appear in C2, where the MAE reaches 1924.8 minutes for delay and 163.6 cancellations, far above the corresponding values for C0 and C1. In other words, the same lines that appear as severity outliers in the unsupervised analysis are also the hardest to model accurately in the supervised regression task. In Table 7 a summary of the line cluster characteristics is provided. This is an important result for the project: line clustering is not only descriptive but also helps identify where standard

regression models are likely to fail and where specialized anomaly-detection logic will add most value.

Clúster	Lines	Avg. incidents/line	Avg. delay	Avg. cancellations	Interpretation
C0	66	4.85	1672.2	39.0	Baseline profile, low exposure and moderate impact
C1	8	41.25	1946.9	45.7	Recurrent high-activity lines
C2	3	9.00	6366.9	313.7	Rare but very severe outlier lines

Table 7: Line cluster table summary

Overall, this task provides three main findings. First, operational incidents are concentrated in a relatively small subset of lines and major nodes, which supports the use of historical frequency features. Second, the selected regression models are adequate for estimating the impact of ordinary disruptions, but they systematically underpredict extreme cases. Third, the three-cluster solution offers a compact and meaningful segmentation of the network, separating routine lines (C0), recurrent high-activity lines (C1), and rare high-severity outliers (C2).

7.3 TASK C RESULTS

Task C translates incident-level disruption into an economic layer and then aggregates the results into KPIs that can support monitoring and decision-making. The results show that annual incident-related cost is both high and volatile, that economic impact is concentrated in a relatively small subset of lines and places, and that a direct cost model performs slightly better overall than the chained approach based on Task B predictions. At the same time, both approaches remain substantially weaker on the most extreme incidents, which is consistent with the tail behaviour already seen for delay and cancellation prediction.

At annual level, the estimated cost of catenary-related incidents over the seven observed years is €58.54 million per year on average, with a 95% interval from €44.70 million to €72.37 million, as seen in Figure 20. However, this hides strong interannual variation. The maximum annual cost is observed in 2022, reaching €84.14 million, whereas the minimum appears in 2025, with €37.90 million. In other words, the range between the highest and lowest annual values exceeds €46.24 million, which confirms that incident cost should be treated as a highly variable operational and financial burden rather than as a stable background expense.

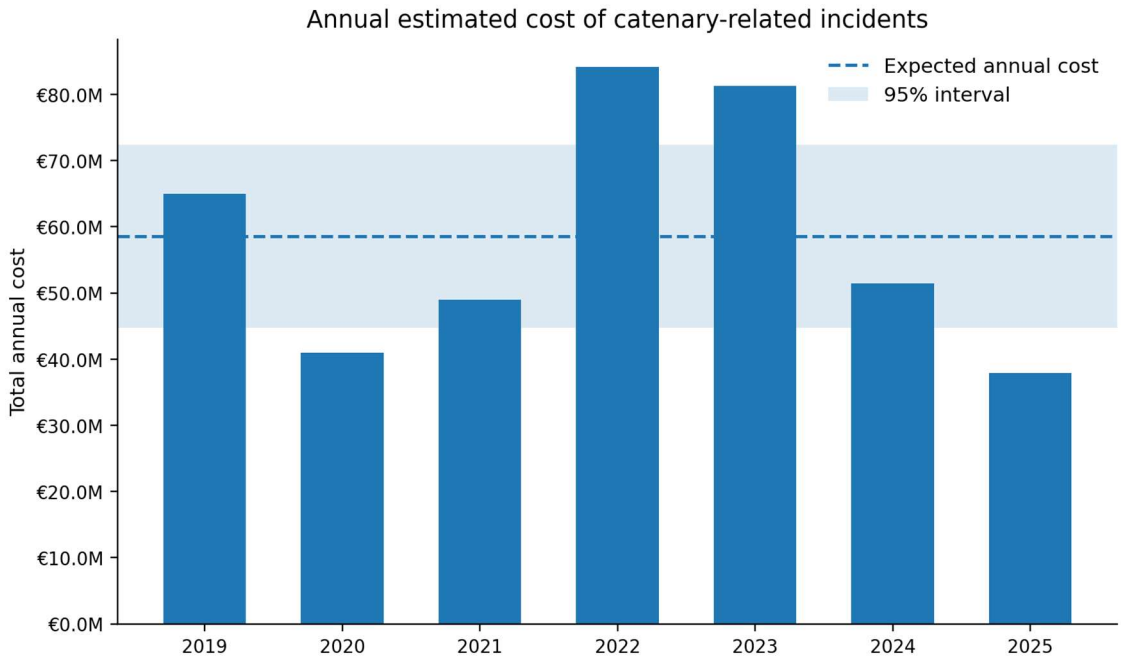

Figure 20: Annual estimated cost of catenary incidents

This cost is also clearly concentrated across the network. At line level, the top 5 lines account for 44.0% of total cost, and the top 10 already explain 63.0%, as in Figure 21. The largest contributor is the group of incidents with unknown or unspecified line, which alone represents 18% of total cost. Among identified lines, line 25 stands out with €37.90 million,

followed by 161 (€26.17 million), 50A (€23.58 million) and 36 (€21.86 million). At place level the concentration is less extreme but still substantial: the top 5 places explain 31.0% of total cost and the top 10 explain 40.0%, shown in Figure 22. Excluding the unspecified category, the most expensive nodes are BRUSSEL-NOORD (€19.20 million), BRUSSEL-ZUID (€12.98 million) and BRUSSEL-CENTRAAL (€11.16 million). This reinforces the conclusion from Task B that the network is not homogeneous: a limited set of corridors and hubs absorbs a disproportionate share of disruption cost.

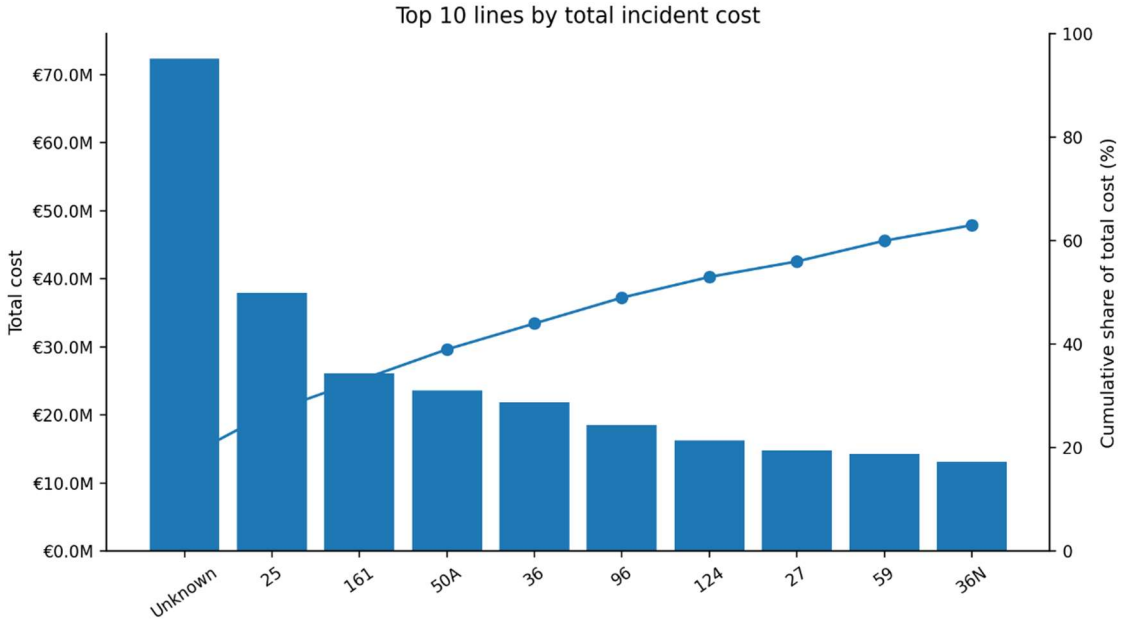

Figure 21: Top 10 lines by total incident cost

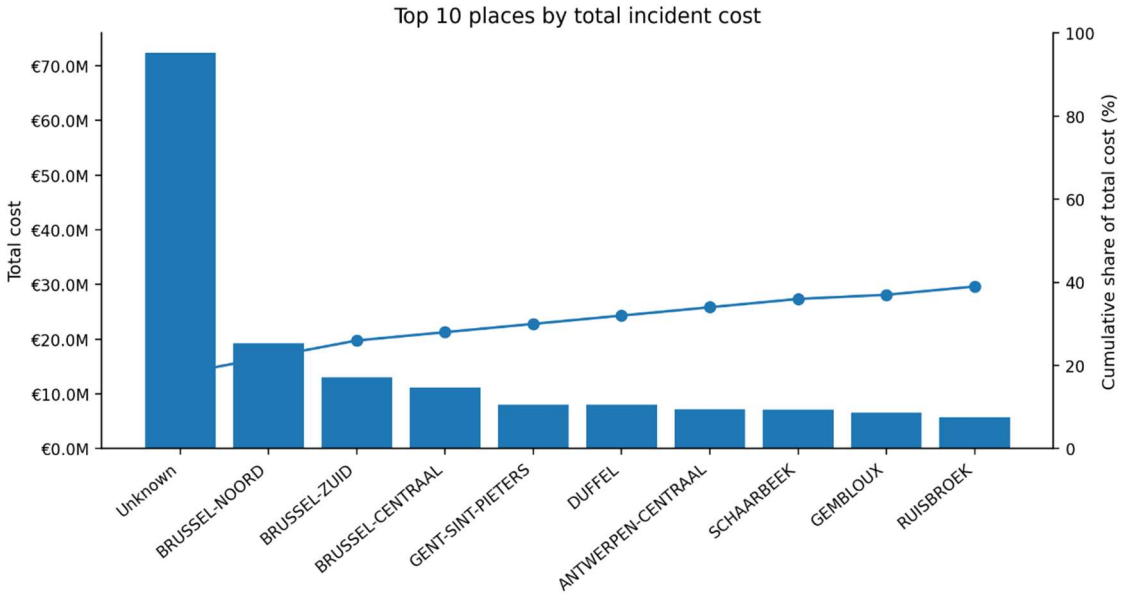

Figure 22: Top 10 places by total incident cost

For incident-level cost prediction, two alternatives were evaluated on the same test set of 170 incidents: a chained approach, in which Task B predictions for delay and cancellations are converted into cost, and a direct model, which predicts incident cost directly. The direct model performs slightly better overall, as indicated in Figure 23. Its test MAE is €236,242, compared with €242,872 for the chained model; its RMSE is €402,598, versus €426,400; and its correlation with actual cost rises to 0.372, compared with 0.282 for the chained approach. This suggests that some cost information is lost when prediction is decomposed into intermediate targets, and that direct estimation is better able to preserve the structure of the final economic variable.

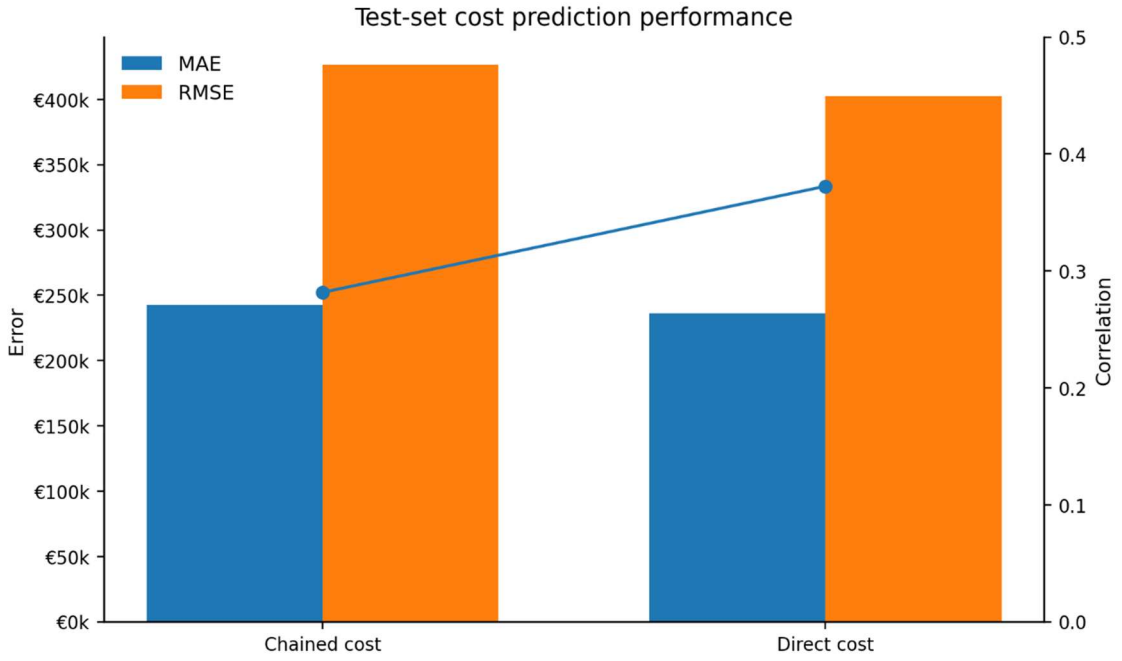

Figure 23: Test- set cost prediction performance

The fitted-versus-observed plots confirm that both approaches approximate the central mass of incident cost better than the tails, as seen in Figure 24 and Figure 25. The direct model produces a slightly tighter cloud around the identity line for mid-range incidents, which explains its better aggregate metrics. Even so, both methods struggle with very severe events. For the lower 90% of the test set, the chained model attains a lower MAE (€153,386) than the direct model (€171,331). However, in the top cost decile, the direct model is clearly better, reducing MAE from €1,048,246 to €820,433. Both models also show strong negative bias in that upper tail, with average underprediction of about €913,930 for the chained method and €774,035 for the direct model. Therefore, the direct approach is preferable for overall cost forecasting, but neither model should be considered fully reliable for rare, very high-impact incidents.

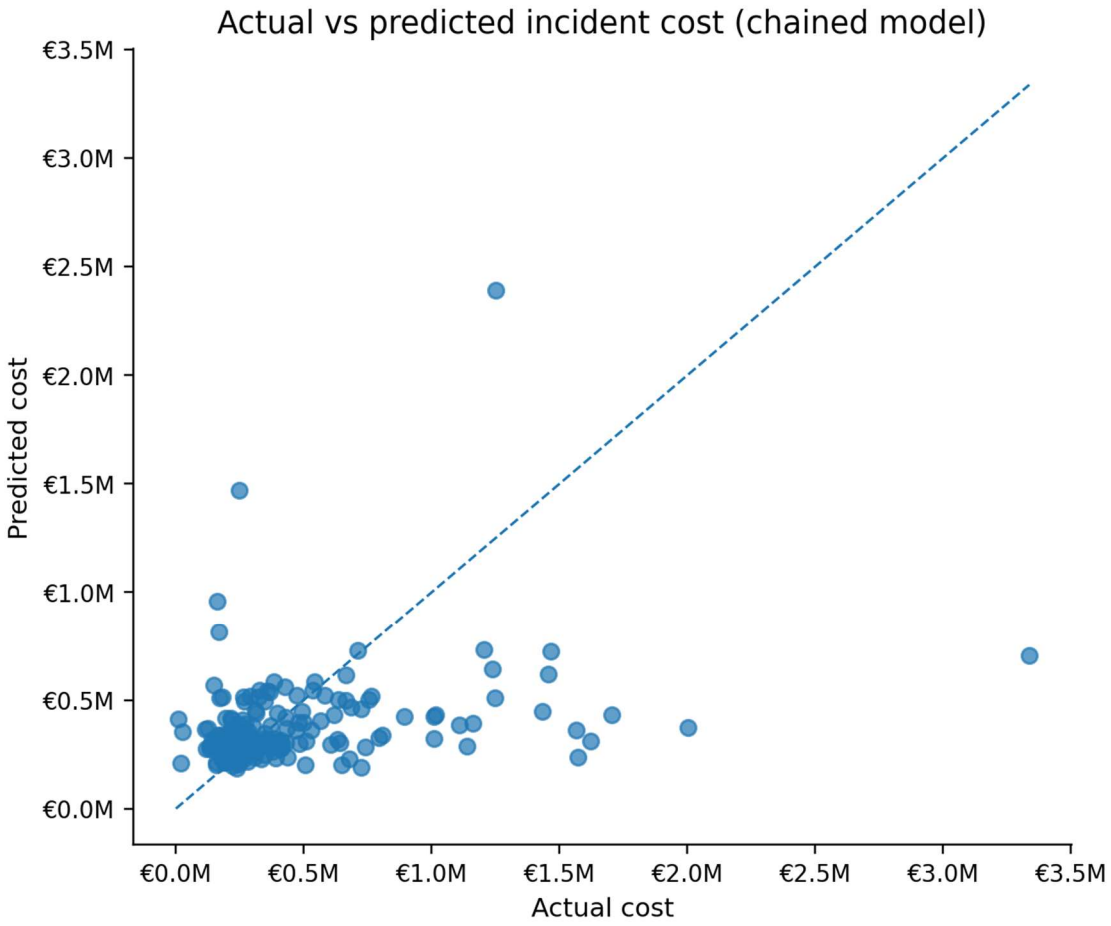

Figure 24: Actual vs predicted chained model

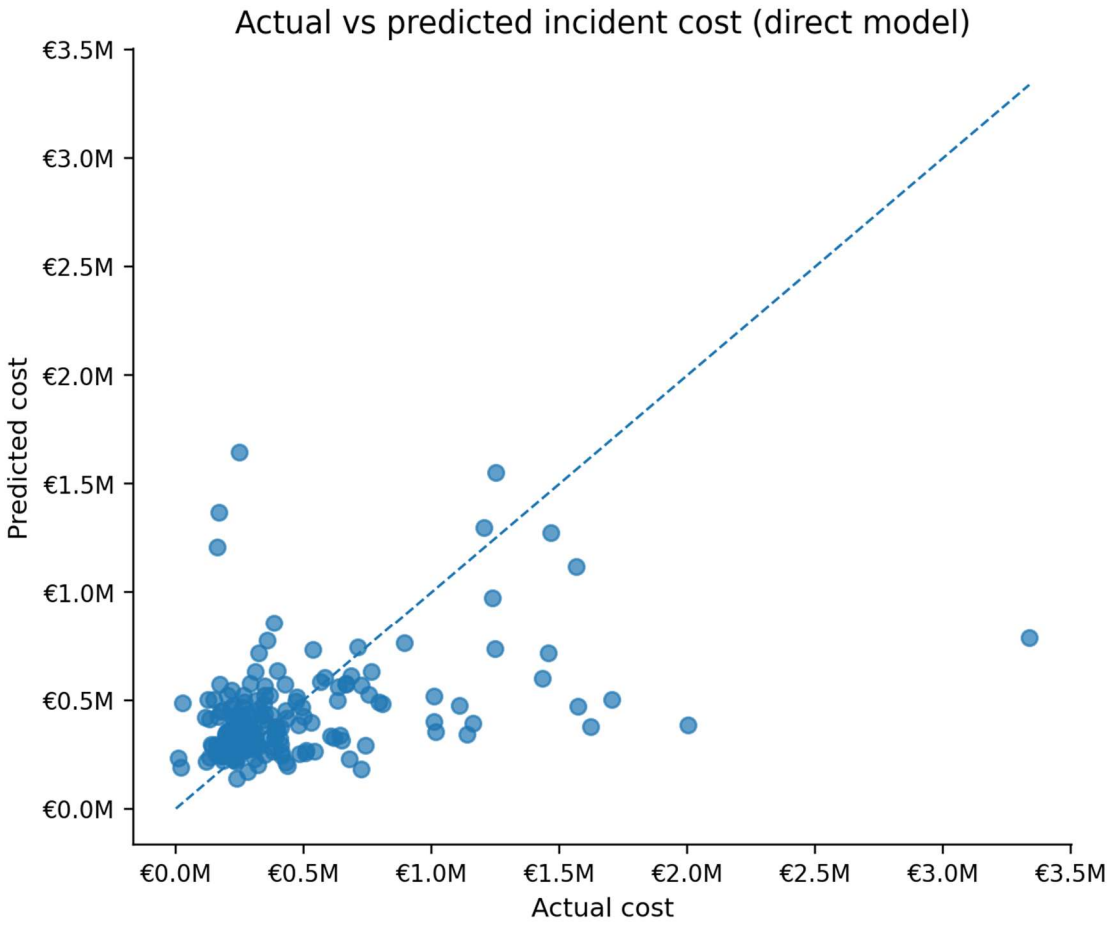

Figure 25: Actual vs predicted direct model

Temporal aggregation adds an important operational perspective. The most expensive quarter in the series is 2022Q1, with a total estimated cost of €15.60 million generated by only 3 incidents, which implies an average cost per incident of €5.20 million. This illustrates a recurring pattern in the KPI tables: total economic burden is not driven only by incident frequency, but also by the presence of a small number of unusually severe events. When cost is decomposed by line cluster, the same structure identified in Task B reappears very clearly, as seen in Figure 26. Clusters C0 and C1 account for almost the same share of incidents (47.2% and 47.3%, respectively), but C1 absorbs a larger share of total cost (42.3% versus 38.3%). Most importantly, C2 represents only 4.4% of incidents yet it explains 18.7% of

total cost, with an average quarterly cost per incident above €1.15 million. This makes C2 the clearest high-severity economic outlier in the network. In addition, Figure 26 suggests that the cost contribution of C2 is not evenly distributed over time, but tends to concentrate in specific quarters, particularly Q1 and Q4. This pattern may indicate a seasonal effect, potentially associated with harsher weather or more adverse operating conditions during those periods, although this interpretation should be regarded as exploratory because no explicit meteorological variables were included in the model.

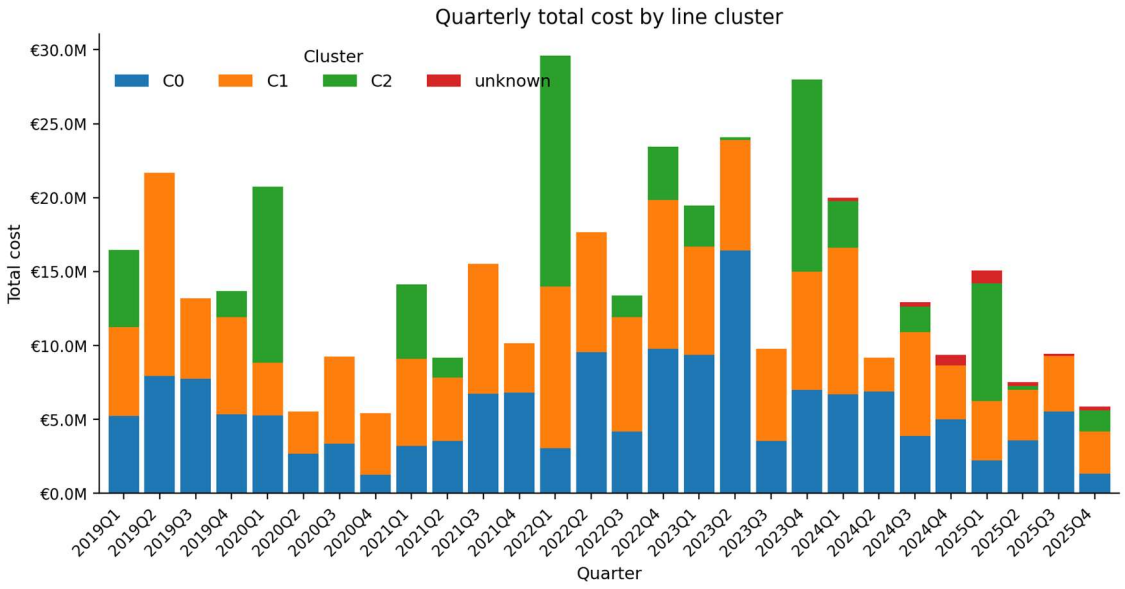

Figure 26: Quarterly total cost by line cluster

The same effect appears in the model error by cluster. On the test set, the average actual incident cost in C2 is about €1,135,180, far above the values for C0 and C1, and both models are much less accurate there. The chained approach reaches a cluster-specific MAE of €946,737 in C2, while the direct model reduces it to €751,855, although this remains far above the error levels observed in C0 and C1. This is an important result for the business analysis: the lines identified as severity outliers are also the ones that dominate cost risk and are hardest to forecast accurately.

Overall, this task delivers three main conclusions. First, catenary-related incidents represent a material annual economic burden, with an expected cost of roughly €58.5 million per year and substantial volatility. Second, cost is strongly concentrated on a limited subset of lines, places, and especially in the high-severity C2 cluster. Third, a direct cost model is preferable to the chained alternative in aggregate terms and especially for the most expensive incidents, although both approaches still underpredict the extreme tail. These findings strengthen the case for combining predictive modelling with cluster-aware monitoring, since the largest business risk is concentrated precisely where standard regression is least reliable.

Chapter 8. CONCLUSIONS AND NEXT STEPS

This project has shown that catenary-related railway incidents can be analyzed through an integrated data-driven framework that combines classification, impact prediction, and economic interpretation. Rather than treating these tasks separately, the study has demonstrated the value of linking them into a coherent analytical pipeline that supports both technical understanding and managerial decision-making.

The first main conclusion is that structured incident records contain sufficient information to identify catenary-related events with useful predictive performance. This confirms that operational logs, when properly cleaned and standardized, can serve not only as administrative documentation but also as a meaningful source of analytical insight.

The second conclusion is that the operational impact of incidents can be approximated with reasonable usefulness for typical cases, although prediction becomes less reliable for the most severe events. In particular, the results of Task B show that delay and cancellation patterns are not uniformly distributed across the network. Some lines behave as routine low-exposure elements, whereas others concentrate on recurrent disruption activity or unusually severe outcomes. This finding highlights the importance of moving beyond incident-level analysis and incorporating network segmentation into the interpretation of disruption risk.

The third conclusion is that translating operational disruption into an economic layer adds substantial practical value. The cost modelling and KPI analysis developed in Task C show that catenary-related incidents represent a significant and highly concentrated economic burden. Even if these monetary values should be interpreted as model-based estimates rather than exact accounting losses, they provide a much clearer view of where disruption risk is most materially concentrated.

At the same time, the study also makes clear that the most severe and costly events remain the hardest to model. It shows that the main challenge in railway disruption analytics is not

the prediction of ordinary cases, but the treatment of rare, high-impact incidents that dominate operational and economic risk. For that reason, the framework proposed should be understood primarily as a structured decision-support system, not as a fully autonomous forecasting tool.

8.1 MAIN CONTRIBUTIONS OF THE PROJECT

First, it develops an **end-to-end analytical workflow** for catenary-related incident analysis, covering data preparation, incident classification, impact prediction, clustering, and cost-oriented KPI construction. This gives the study coherence and practical relevance beyond isolated modelling exercises.

Second, it shows that **structured operational incident data can support multiple analytical objectives simultaneously**. The same dataset is used to determine incident type, estimate operational severity, and approximate economic burden, demonstrating the broader analytical potential of railway disruption records.

Third, it introduces a **network-level perspective** through clustering, showing that disruption severity and cost are concentrated in a limited subset of lines and operational profiles. This adds strategic value to the analysis by helping identify where standard models are less reliable and where monitoring efforts should be prioritized.

Fourth, it provides a **business-oriented interpretation layer** by translating operational disruption into cost-based indicators. This is especially important because it connects technical railway analytics with managerial concerns such as prioritization, reporting, and resource allocation.

8.2 NEXT STEPS

The most important next step is to improve the explanatory richness of the data. Future work should incorporate additional variables such as weather conditions, maintenance history, asset age, infrastructure interventions, traffic density, or inspection information. These data sources could help explain part of the uncertainty that remains unobserved in the current framework.

A second step is to strengthen the modelling of rare high-impact events. Since the results consistently show weaker performance in the upper tail, future research should focus more directly on extreme-event modelling, anomaly detection, and hybrid approaches that combine statistical learning with expert-informed rules.

A third step is to move the framework closer to operational use. This could involve the development of a dashboard or monitoring interface capable of presenting classification outputs, impact estimates, cost concentration, and cluster-level KPIs in an interpretable way for decision-makers.

In summary, this project shows that catenary-related railway incidents can be analyzed through an integrated framework that combines classification, impact prediction, and cost-oriented KPI generation. The results confirm that structured incident data contain useful operational signals, while also showing that disruption severity and economic burden are highly concentrated in a limited part of the network. At the same time, the study highlights that the most severe events remain the hardest to predict, which means the proposed framework is best understood as a decision-support tool rather than a fully autonomous forecasting system.

Chapter 9. BIBLIOGRAPHY

[1] Infrabel. (n.d.). *Most important incidents in terms of impact on trains punctuality* [Dataset]. Infrabel Open Data. Retrieved October 21, 2025, from https://infrabel.opendatasoft.com/explore/dataset/belangrijkste-incidenten/

[2] Universidad Pontificia Comillas. (n.d.). *Cómo plantear un TFG* [PDF]. SIFO (LMS). Retrieved October 21, 2025, from https://sifo.comillas.edu/…/1%20C%C3%B3mo%20plantear%20un%20TFG.pdf

[3] Universidad Pontificia Comillas. (n.d.). *Para hacer una parte del plan de negocio como TFG* [PDF]. SIFO (LMS). Retrieved October 21, 2025, from https://sifo.comillas.edu/…/3%20Para%20hacer%20una%20parte%20del%20Plan%20de%20negocio%20como%20TFG.pdf

[4] Office of Rail and Road. (n.d.). *Train service performance*. *Annual assessment of Network Rail 2024 to 2025*. Retrieved February 19, 2026, from https://www.orr.gov.uk/annual-assessment-network-rail-2024-2025/train-service-performance

[5] Infrabel. (n.d.). Oorzaken vertraging per maand (Causes of delay per month) [Data set]. Infrabel Open Data. Retrieved February 19, 2026, from https://opendata.infrabel.be/explore/dataset/oorzaken-vertraging-per-maand/information/

[6] Rossa, K., Smith, A. S. J., Batley, R. P., & Hudson, P. (2024). The valuation of delays in passenger rail using journey satisfaction data. *Transportation Research Part D: Transport and Environment, 129*, 104088. https://doi.org/10.1016/j.trd.2024.104088

[7] Topham, G. (2024, April 17). *Compensation payouts to UK rail passengers for delays hit £100m a year*. *The Guardian*. https://www.theguardian.com/business/2024/apr/17/record-compensation-payouts-passengers-uk-train-delays

[8] Sunar, O., & Fletcher, D. (2023). A new small sample test configuration for fatigue life estimation of overhead contact wires. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 237*(4), 438–444. https://doi.org/10.1177/09544097221116531

[9] Collina, A., Lo Conte, A., & Bucca, G. (2025). Feasibility analysis of monitoring contact wire rupture in high-speed catenary systems. *Vibration, 8*(2), 22. https://doi.org/10.3390/vibration8020022

[10] Shift2Rail Joint Undertaking. (2019). Shift2Rail catalogue of solutions (Catalogue of Solutions—Web version) [Report]. https://rail-research.europa.eu/wp-content/uploads/2019/10/Catalogue-of-Solutions-Web.pdf

[11] Bris-Peñalver, F. J., Verdecia-Peña, R., & Alonso, J. I. (2026). *A survey of AI-enabled predictive maintenance for railway infrastructure: Models, data sources, and research challenges*. Sensors, *26*(3), 906. https://doi.org/10.3390/s26030906

[12] Zhang, Y., Dai, P., Sysyn, M., Hu, Y., Kou, L., Song, H., & Shi, J. (2026). *From point clouds to predictive maintenance: A review of intelligent railway infrastructure monitoring*. Sensors, 26(4), 1131. https://doi.org/10.3390/s26041131

[13] HBK. (n.d.). *Measurements in high-voltage environments: Overhead line and pantograph maintenance*. Retrieved February 19, 2026, from https://www.hbm.com/10479/pantograph-monitoring-using-optical-technology/

[14] Amin, M. A., Najeh, T., Sridharan, N. V., Ghoul, A., & Karim, R. (2026). Enhancing railway infrastructure monitoring with AI: A machine learning approach for event detection. Transportation Engineering, 23, 100414. https://doi.org/10.1016/j.treng.2025.100414

[15] Du, J., Li, J.-P., Chaplin, G., Gardiner, A., & Hu, F.-Y. (2023). Failure mode analysis and maintenance of railway overhead line rigid stainless steel droppers and multi-strand copper jumpers. IET Electrical Systems in Transportation, 2023, Article 8858919. https://doi.org/10.1049/2023/8858919

[16] Mukunzi, G., & Palmqvist, C.-W. (2024). The impact of railway incidents on train delays: A case of the Swedish Railway Network. Journal of Rail Transport Planning & Management, 30, 100445. https://doi.org/10.1016/j.jrtpm.2024.100445

[17] Rossa, K., Smith, A. S. J., Batley, R. P., & Hudson, P. (2024). The valuation of delays in passenger rail using journey satisfaction data. *Transportation Research Part D: Transport and Environment, 129*, 104088. https://doi.org/10.1016/j.trd.2024.104088

[18] Lupi, C., Felli, F., Ciro, E., Paris, C., & Vendittozzi, C. (2021). Railway overhead contact wire monitoring system by means of FBG sensors. Fracture and Structural Integrity, 15(57), 246–258. https://doi.org/10.3221/IGF-ESIS.57.18

[19] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

[20] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information

Science, 41(6), 391–407. https://asistdl.onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9

[21] Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

APPENDIX A. CONTRIBUTION TO THE SDGS

This project contributes to the Sustainable Development Goals (SDGs) through its focus on railway reliability, data-driven infrastructure analysis, and cost-oriented decision support. Although the work is developed as an analytical and predictive framework rather than as a physical intervention on the railway network, its objectives and expected applications are aligned with several SDGs related to innovation, sustainable transport, efficiency, and resilience.

The most direct connection can be established with **SDG 9: Industry, Innovation and Infrastructure**. One of the main aims of this goal is to promote resilient infrastructure and encourage innovation in industrial processes. In this context, the project contributes by proposing a data-driven framework for analyzing catenary-related railway incidents using machine-learning techniques, clustering, and KPI generation. The work supports a more intelligent and structured understanding of infrastructure behaviour, moving beyond descriptive reporting toward analytical decision support.

A second important link is with **SDG 11: Sustainable Cities and Communities**. Railway transport is a key component of sustainable mobility systems, especially in urban and interurban environments. Improving the reliability of electrified railway services contributes indirectly to safer, more efficient, and more dependable transport networks. By helping identify catenary-related incident patterns, estimate disruption severity, and highlight where operational risk is concentrated, the project can support decisions that improve service continuity and reduce the negative effects of disruption on passengers and mobility systems.

The project also shows a relevant connection with **SDG 12: Responsible Consumption and Production**. More efficient management of infrastructure incidents can help reduce unnecessary operational inefficiencies, optimize maintenance priorities, and improve the use of available resources. Although the project does not directly control maintenance

interventions, its analytical outputs can support better prioritization and more informed allocation of effort.

Another indirect contribution can be associated with **SDG 13: Climate Action**. Railways are generally considered one of the more sustainable forms of mass transport, particularly in electrified networks. Supporting the reliability and continuity of railway services may help reinforce the attractiveness and efficiency of rail transport relative to less sustainable alternatives.

A further connection may be made with **SDG 8: Decent Work and Economic Growth**. Railway disruptions generate economic costs, affect productivity, and may reduce operational efficiency. By translating incident behaviour into cost-oriented indicators and KPIs, the project provides a framework that can support more informed operational and managerial decisions. This may help organizations reduce disruption burden, improve planning, and allocate resources more effectively. While this is not the primary purpose of the SDG, the project contributes to it through its economic and organizational perspective.

Overall, the strongest alignment of the project is with SDG 9, followed by SDG 11, while SDG 12, SDG 13, and SDG 8 represent secondary but still relevant connections. The work does not claim to solve these goals directly. Rather, its contribution lies in providing an analytical tool that supports more reliable, efficient, and sustainable railway infrastructure management.

APPENDIX B. DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS IN THE BACHELOR'S THESIS

WARNING: The University considers ChatGPT and other similar tools to be very useful in academic life. However, their use always remains under the student's responsibility, since the answers provided may not be accurate. In this regard, their use in the preparation of the Bachelor's Thesis to generate code is **not permitted**, because these tools are not reliable for that task. Even if the code works, there are no guarantees that it is methodologically correct, and it is highly likely that it is not.

I, **Lucía Martínez Ruiz**, a student of **Double Degree in Telecommunications Engineering and Business Analytics and Master´s in Telecommunications Engineering** at **Comillas Pontifical University**, upon submitting my Bachelor's Thesis entitled **"Quantifying the Business Impact of Catenary Failures in Railway Operations"**, hereby declare that I have used ChatGPT or other similar Generative Artificial Intelligence (GAI) tools only in the context of the activities described below **[the student must keep only those activities in which ChatGPT or similar tools have been used and delete the rest. If none have been used, delete all of them and write: "I have not used any."]**:

1. **Research idea brainstorming:** Used to generate and outline possible research areas.
2. **References:** Used together with other tools, such as Science, to identify preliminary references, which I then reviewed and validated.
3. **Code interpretation:** Used to carry out preliminary data analysis.
4. **Multidisciplinary studies:** Used to understand perspectives from other fields on multidisciplinary topics.

5. **Style and language improvement:** Used to improve the linguistic and stylistic quality of the text.
6. **Preliminary generator of flowcharts and content:** Used to draft initial diagrams.
7. **Summarizer and explainer of complex books:** Used to summarize and understand complex literature.
8. **Reviewer:** Used to receive suggestions on how to improve and refine the thesis with different levels of rigor.
9. **Translator:** Used to translate texts from one language into another.

I affirm that all the information and content presented in this thesis are the result of my own individual research and effort, except where otherwise indicated and properly acknowledged (I have included the appropriate references in the thesis and have explicitly stated how ChatGPT or other similar tools have been used). I am aware of the academic and ethical implications of submitting non-original work and accept the consequences of any violation of this declaration.

Date: 07/04/2026

Signature: