PoC Anomaly Detection in Pensions

Reflection

Bachelor's degree in Computer Science
field AI

Aynur Guliyeva

Academic year 2025-2026

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

1. Introduction

This reflection document provides an overview and critical evaluation of the internship project carried out during my internship period. The document is structured in two main parts.

First, a **substantive reflection** is given on the internship project itself, focusing on the objectives, the work that was carried out, the results for the organisation, and the current and future state of the system that was developed.

Second, a **personal reflection** is presented, describing what the internship has meant for me on a personal and professional level, which competencies were developed, which challenges were encountered, and how these challenges were addressed.

2. Substantive Reflection on the Internship Project

Building on the context and objectives described above, the following sections reflect on the execution and outcomes of the internship project. Section 2.2 describes the concrete work that was carried out and the system components that were delivered, followed by an evaluation of the value for the organisation and the current project status.

2.1 Context and Objectives of the Project

The internship project was carried out at Pension Architects, a company specialised in pension administration and actuarial software. The organisation manages large volumes of pension-related data across multiple pension plans, each with its own regulatory and calculation logic.

The main objective of the internship project was to **design and implement an AI driven anomaly detection framework** to automatically detect irregularities in pension contribution and reserve data. The focus was on four major plan groups:

- PF1
- PF2
- PF3 120/130/140
- PF4

The project aimed to reduce manual control effort, increase consistency in anomaly detection, and introduce a scalable solution that could improve over time through human feedback.

2.2 What Was Accomplished

During the internship, a complete end-to-end anomaly detection system was designed and implemented. Concretely, the following components were realised:

- **Data cleaning and feature engineering pipelines** for each plan type, converting raw CSV inputs into model-ready datasets.

- **Synthetic anomaly injection mechanisms**, introducing realistic anomalies to enable supervised learning despite limited historical anomaly data.
- **Supervised machine learning models** (Random Forest–based) trained to detect anomalies while controlling false positives.
- **Automated scoring pipelines** using AWS SageMaker, triggered by new data uploads to Amazon S3.
- **Human-in-the-loop retraining pipelines**, allowing domain experts to label anomalies and improve model performance over time.
- **AWS automation** using S3 event notifications, Lambda functions, and SageMaker Pipelines.
- **Monitoring and transparency** via Slack notifications for scoring, retraining, and error events.

This resulted in a production-ready framework that can be reused across different pension plans with minimal configuration changes.

2.3 Value for the Organisation and Users

For Pension Architects, the project delivers several concrete benefits:

- **Early detection of calculation errors** in pension contributions and reserves.
- **Reduction of manual review workload**, allowing experts to focus on truly relevant cases.
- **Consistent anomaly detection** across large datasets and multiple pension plans.
- **Scalability**, as new plans can be added following the same architectural pattern.
- **Continuous improvement**, as the models learn from human feedback.

For end users (internal analysts and domain experts), the system provides clear anomaly flags, confidence scores, and traceable results, improving both efficiency and trust in the system.

2.4 Project Status and Remaining Work

The project is functionally complete and operational. The core pipelines for scoring and retraining are fully automated and already in use for testing and validation purposes.

Possible future improvements include:

- Further refinement of feature engineering as new domain insights emerge.
- Expansion to additional pension plans.
- Integration of more advanced explainability techniques for anomaly justification.
- Long-term monitoring of model drift and periodic retraining strategies.

2.5 Advice for the Organisation

Based on the project experience, the following recommendations can be made:

- Continue investing in **human-in-the-loop feedback**, as this significantly improves model quality.
- Maintain clear **data governance and naming conventions** for S3 inputs to avoid automation failures.
- Gradually build a **shared anomaly knowledge base** from labeled feedback.
- Keep documentation up to date as pipelines evolve.

3. Personal Reflection

In addition to the technical and professional competencies developed during the internship, the experience also contributed significantly to my personal growth. The following section reflects on how increased responsibility and exposure to real-world constraints influenced my way of working and thinking.

3.1 Personal Learning Experience

This internship was highly valuable for my personal and professional development. It allowed me to work on a complex, real-world problem where technical decisions had direct business and regulatory impact.

I gained a much deeper understanding of how AI systems are designed **beyond experimentation**, focusing on reliability, automation, and maintainability.

3.2 Competencies Developed

During the internship, I developed several competencies:

- **Technical competencies** ○ Advanced Python programming ○ Machine learning model design and evaluation ○ Feature engineering for structured financial data ○ Cloud architecture using AWS (S3, Lambda, SageMaker)
- **Analytical competencies** ○ Translating business problems into data-driven solutions ○ Evaluating trade-offs between recall and false positives
    - Designing systems that generalise rather than overfit
- **Professional competencies** ○ Communicating complex technical concepts to non-technical stakeholders
    - Working autonomously on long-term projects ○ Structuring documentation and pipelines for reuse

3.3 Personal Growth

I grew significantly in terms of **independence and responsibility**. Rather than working on isolated tasks, I was responsible for the full lifecycle of the solution, from design to deployment.

I also improved my ability to **handle uncertainty**, as not all requirements were known upfront and many decisions had to be validated iteratively.

3.4 Challenges and How They Were Addressed

The main challenges encountered were not purely technical:

- **Complex domain logic**
  Pension rules are complex and differ per plan. This was addressed through close collaboration with domain experts and iterative validation.
- **Data quality issues**
  Real-world datasets contained inconsistencies and missing values. Robust preprocessing and defensive programming were applied.
- **Automation reliability**
  Cloud automation introduced new failure modes (missing files, incorrect triggers). These were mitigated through validation checks, logging, and monitoring.

Each challenge contributed to a deeper understanding of building reliable systems in production environments.

4. Conclusion

The internship project resulted in a robust and scalable AI-driven anomaly detection framework that delivers tangible value to Pension Architects by improving the efficiency and consistency of data validation processes. Beyond the technical outcome, this internship was a highly formative experience for me personally.

Working on a real-world system with direct business and regulatory impact challenged me to think beyond model performance alone and to focus on reliability, maintainability, and stakeholder trust. Being responsible for the full lifecycle of the solution significantly increased my confidence as a future AI professional.

This internship confirmed my interest in applied artificial intelligence and data-driven systems in complex domains and has strongly influenced my ambitions for my further studies and career.