



DOCUMENTATION

INTERNSHIP REFLECTION REPORT

Bachelor in Applied Computer Science.

Academiejaar 2024-2025

Campus: Geel

Emmanuel Akpandara

Learning Points

- Improved machine learning skills, particularly with LLMs like PatentBERT.
- Understood patent data processing and classification challenges.
- Gained experience with ML tools such as **MLflow**, **Databricks**, and fine-tuning pre-trained models.
- Enhanced problem-solving and critical thinking through research and experimentation.

Key Responsibilities and Tasks

- Analyzed the limitations of the existing **Bag of Words (BoW)** classification system.
- Implemented and fine-tuned **PatentBERT** for patent classification.
- Addressed challenges like class imbalance, overfitting, and large text processing.
- Experimented with **data augmentation** with contextual **embeddings** to improve model performance.
- Used **MLflow** for tracking experiments and model performance.
- Presented progress and results to mentors, integrating feedback for improvements.

Challenges Faced and Solutions

- **Understanding Existing Code:** Took time to understand the legacy BoW system before developing new models.
- **Class Imbalance:** Implemented weighted sampling and fine-tuning strategies to balance model predictions.
- **Overfitting:** Gradual unfreezing of BERT layers and learning rate adjustments helped prevent overfitting.
- **Large Text Handling:** Explored chunking strategies and considered alternative models for long sequences.
- **Experimentation:** Tested various hyperparameters and architectures to refine model performance.

Skills Developed

- **Technical Skills:**
 - Fine-tuning pre-trained models (**PatentBERT**, **BatteryBERT**).
 - Experiment tracking with **MLflow** and data handling with **Databricks**.
 - Implementing advanced machine learning techniques and data augmentation.
- **Analytical Skills:**
 - Evaluating model performance with metrics like accuracy, F1-score, and PR AUC.
 - Critical thinking in identifying and solving model limitations.
- **Soft Skills:**
 - Improved presentation and communication skills through regular updates.
 - Time management and problem-solving in a professional setting.

Personal Growth

- Gained confidence in applying theoretical ML knowledge to real-world problems.
- Learned the importance of **iterative experimentation** and continuous improvement.
- Developed a structured approach to tackling complex problems.
- Understood the value of integrating feedback from mentors for project success.

Impact of the Internship

- Gained insight into the R&D challenges in the battery industry.
- Strengthened technical expertise and clarity on pursuing a career in **machine learning** and **data engineering**.

Conclusion

This internship at Umicore Belgium was a transformative experience, enhancing both my technical and professional skills. The challenges I faced and overcame allowed me to grow as a machine learning practitioner. The opportunity to contribute to real-world projects has piqued my interest in considering a career in AI-driven solutions and data engineering.