Capstone Project Problem Statement

23 February 2025 10:40

Problem Statement: Credit Card Fraud Detection

Our client, a European bank, requires a robust fraud detection system for their issued credit cards. The primary objective is to identify and reject fraudulent transactions to prevent financial losses for the bank. The project involves using provided sample credit card transaction data to build and validate a fraud detection model.

Additional data for inferencing/production, if needed, can be generated using this link

Functional Requirements:

1. The training dataset is available in an S3 bucket.
2. Perform data redaction if any sensitive information is present.
3. Validate, ingest, and preprocess the data.
4. Use a feature store to maintain customer information.
5. Trigger training when code is merged into the main branch.
6. Deploy the model for both batch and online inferencing after training.
7. Deploy the model to production based on a fixed threshold of F1 score or another appropriate metric.
8. Set up data and model monitoring for the inference batch job.
9. Send email alerts when a drift is detected.

Non-Functional Requirements:

1. Use a code repository with at least a feature branch and a main branch.
2. Implement a CI/CD pipeline that triggers on code merge to the main branch.
3. Keep pipeline steps modular.
4. Ensure experiment tracking, dataset versioning, and model versioning.
5. Incorporate logging and error handling in the pipelines.

The project should leverage AWS Sagemaker features as much as possible while exploring alternative methods for implementation.