Title: Detecting Fraudulent Automobile Claims for Atlannta Insurance

Role: Data Scientist

Objective: To develop a robust system for detecting fraudulent automobile insurance claims using machine learning techniques.

Team: Worked in a team of two data scientists

Methodology:

1. Data Preparation and Feature Engineering:
   * Performed data cleaning, feature engineering, duplicate detection, and correction of incoherent values in the insurance claims dataset.
2. Handling Unbalanced Data:
   * Addressed the issue of unbalanced data using a semi-supervised methodology to ensure the models were trained effectively on both fraudulent and non-fraudulent cases.
3. Supervised Models for Fraud Detection:
   * Created supervised machine learning models, including AdaBoost, XGBoost, and Support Vector Machine (SVM), to detect fraudulent claims while minimizing false positive cases.
4. Outlier-based Fraud Detection:
   * Utilized unsupervised learning techniques, such as Isolation Forest and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), to detect potential fraud as outliers in the dataset.
5. Technologies:
   * Implemented the project using Python, scikit-learn, CUDA, PySpark, Power BI, and XGBoost.

Results:

The team successfully developed a comprehensive fraud detection system for Atlannta Insurance's automobile claims. By utilizing a combination of supervised and unsupervised machine learning techniques, the system effectively identified fraudulent cases while minimizing false positives. This allowed Atlannta Insurance to streamline their claims processing and mitigate the financial impact of fraudulent claims.