As the project lead for a team of two Data Scientists, I worked on a comprehensive fraud detection system for Atlanta Insurance's automobile claims. The objective was to develop a robust system that could effectively detect fraudulent claims using machine learning techniques.

Our team utilized a meticulous methodology that began with data preparation and feature engineering, which involved data cleaning, duplicate detection, correction of incoherent values, and feature engineering in the insurance claims dataset.

We also addressed the issue of unbalanced data by using a semi-supervised methodology to ensure the models were trained effectively on both fraudulent and non-fraudulent cases. Our team created supervised machine learning models, including AdaBoost, XGBoost, and Support Vector Machine (SVM), to detect fraudulent claims while minimizing false positive cases. Additionally, we 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.

To implement the project, we used Python, scikit-learn, CUDA, PySpark, Power BI, and XGBoost technologies. Our team's approach successfully developed a comprehensive fraud detection system that effectively identified fraudulent cases while minimizing false positives. This allowed Atlanta Insurance to streamline its claims processing and mitigate the financial impact of fraudulent claims.

As the project lead, I was responsible for ensuring that the project adhered to best practices, timelines, and quality standards. My contribution to the project included driving the team's methodology, supervising feature engineering and model development, as well as ensuring that the project was completed within schedule and budget. Working as a team of two data scientists, we were able to deliver a project that met the company’s needs and exceeded its expectations.