Scalable Real-Time Credit Card Fraud Detection

My project focused on designing and evaluating a machine learning-based fraud detection system capable of identifying credit card fraud in real time—an essential capability for financial institutions seeking to minimize financial losses and protect customer trust.

Using a large, simulated dataset representative of real-world transactions, I engineered a solution that prioritized interpretability, scalability, and response speed. Features such as transaction amount, merchant name, location, and cardholder demographics were leveraged to build and test multiple supervised learning models. The best-performing model utilized Extreme Gradient Boosted Trees (XGBoost)—achieved strong results in identifying fraudulent transactions while maintaining a balance between low false negative rate, a critical metric in preventing financial exposure, and low false positive rate, important for maintaining customer trust and limiting inconvenience to customers.

To simulate real-time operations, I implemented a streaming pipeline using Spark, mimicking how financial institutions receive and evaluate thousands of transactions per second. The model processed incoming transactions dynamically, flagging those with high fraud probability for immediate review or intervention.

This proof of concept demonstrates not only the feasibility of deploying real-time fraud detection at scale, but also the importance of thoughtful feature selection, system efficiency, and balance between fraud prevention and customer experience. Future enhancements could include additional transaction context (e.g., card-present data, merchant distance from cardholder) and further model tuning to increase precision and reduce operational friction.