

A Big Data Analytics Capstone Project in Finance (Banking, Financial services, Insurance)

A.SHALINI
24MBMA02
MBA (GENERAL)

USE CASES:

1. Predicting customer churn in Banking
2. ATM cash demand forecasting
3. Credit card fraud detection
4. Customer segmentation for banking
5. Loan default prediction

The financial sector generates vast amounts of data daily, yet struggles to extract actionable insights. This comprehensive big data analytics portfolio addresses five critical business imperatives. Through five strategic use cases, we demonstrate how machine learning transforms raw data into measurable business value across banking and financial services.



Driving Revenue Through Customer Intelligence

Customer Segmentation for Banking

Challenge: Generic marketing approaches fail to engage diverse customer bases effectively.

Solution: Unsupervised Clustering (K-Means / DBSCAN) on transaction + demographic data.

Impact: Enables targeted offers and personalized marketing → expected 8–10% growth in digital transactions..

Result: Created 4–5 customer groups (e.g., “High-Value Loyal,” “Credit-Heavy Risk,” “Digital Engager”).

Predicting Customer Churn in Banking

Challenge: Acquiring new customers costs five times more than retaining existing ones.

Solution: Binary classification models predict churn probability by analysing activity levels, account balances, and service interactions.

Impact: Helps focus retention efforts on high-risk customers → higher retention ROI.

Result: Model gives each customer a churn probability score.

ATM Cash Demand Forecasting

Challenge: Inefficient cash management leads to ATMs running out of cash, customer frustration, or excessive idle capital and increased security risks.

Solution: Time-series forecasting models (e.g., Prophet) predict daily/hourly cash withdrawals based on historical data, seasonal patterns, and local events.

Impact: Achieves a 10–15% reduction in ATM operational expenses, minimises cash-in-transit frequency, and enhances customer satisfaction through consistent availability.

Result: Accurate 7-day cash demand forecast for each ATM.

Minimising Financial Loss Through Advanced Fraud Detection

Credit Card Fraud Detection

Challenge: Fraudulent transactions erode customer trust and generate millions in losses annually

Solution: Imbalanced classification techniques handle datasets where fraud represents just 0.2% of cases

Impact: Reduces chargeback losses and increases customer trust.

Result: Real-time model that predicts fraud before payment is processed.

Loan Fraud Prediction

Challenge: Fraudulent loan applications lead to substantial bad debt write-offs

Solution: Predictive models evaluate applicant demographics, credit history, and employment verification

Impact: 90–93% detection rate for fraudulent applications

Result: Model accuracy of 90–93% in identifying fraudulent applications.



Key Outcomes & Strategic Impact

Digital Transaction Growth

Increase in digital transactions through personalised customer segmentation.

Loan Fraud Detection Accuracy

High precision in identifying fraudulent loan applications at the point of submission.

Enhanced Customer Lifetime Value

Improved retention ROI and customer lifetime value through proactive churn prediction.

Operational Cost Reduction

Savings in ATM operational expenses through optimised cash demand forecasting.

Real Time Fraud Prevention

Instantaneous detection and prevention of credit card fraud, reducing chargeback costs.

Transaction Growth

Increase in digital transactions via personalised segmentation

Cost Savings

Monthly reduction in ATM operational expenses through accurate forecasting

Fraud Reduction

Conclusion & Future Outlook

Big Data Analytics helps financial institutions achieve:

Risk Reduction – Accurate and fast fraud detection

Revenue Growth – Better segmentation and retention models

Cost Savings – Smarter forecasting and cash management

Technical Skills Demonstrated:

Data Handling: Large, imbalanced, and time-series dat

Modeling: Clustering, Classification, Forecasting

Tools Used: Python, Scikit-learn, Prophet, Databricks, Cloud platforms

Big data analytics empowers financial institutions to achieve **risk reduction** through intelligent fraud detection, **revenue growth** via sophisticated customer strategies, and **operational excellence** through predictive resource management.

Technical capabilities demonstrated: Large-scale data processing, imbalanced classification, unsupervised clustering, time-series forecasting with Prophet, and end-to-end model deployment using Python, Scikit-learn, and cloud platforms.

