# Capstone Phase 1

Project Title: Credit Card Fraud Detection Analytics

Submitted by: Chinnari Abhishek

Supervisor: Sidrah Fayaz Wani (E17441)

## Week 1

### Synopsis:

Initiated the capstone project with team discussions and finalized the domain and project scope—Credit Card Fraud Detection. Outlined weekly objectives and created the initial research framework.

### Research Paper Work:

Began reviewing literature on fraud detection techniques, including supervised and unsupervised learning methods. Identified relevant journals and research on anomaly detection in finance.

### Literature Review:

Focused on classical and deep learning techniques for fraud analytics—Random Forest, Isolation Forest, and Autoencoders.

### Code Work:

Started environment setup in Jupyter Notebook. Loaded a sample dataset (Kaggle credit card dataset). Began data exploration.

### PPT Work:

Prepared a basic project introduction slide with objectives, problem statement, and roadmap.

### Conclusion:

Achieved project initialization successfully. Planned data preprocessing tasks for Week 2.

## Week 2

### Synopsis:

Focused on data preprocessing and feature understanding for credit card fraud dataset.

### Research Paper Work:

Studied the impact of imbalanced datasets on model performance and possible solutions (SMOTE, ADASYN, etc.).

### Literature Review:

Analyzed papers dealing with class imbalance and fraud probability threshold optimization.

### Code Work:

Handled missing data, normalized features, and performed PCA for dimensionality reduction.

### PPT Work:

Added data preprocessing steps and dataset summary.

### Conclusion:

Data preparation nearing completion. Next step: Model selection and training.

## Week 3

### Synopsis:

Began model selection phase and trained initial models on the dataset.

### Research Paper Work:

Reviewed performance metrics for fraud detection—precision, recall, AUC-ROC.

### Literature Review:

Focused on works comparing different classifiers like Logistic Regression, Decision Trees, and SVMs.

### Code Work:

Trained Logistic Regression and Random Forest models. Evaluated using confusion matrix.

### PPT Work:

Inserted performance charts and model architecture summary.

### Conclusion:

Initial models built. Next week to explore ensemble and anomaly detection methods.

## Week 4

### Synopsis:

Expanded model training to ensemble learning and deep learning-based techniques.

### Research Paper Work:

Reviewed boosting techniques like XGBoost and LightGBM for fraud analytics.

### Literature Review:

Investigated deep learning approaches including ANN and Autoencoder-based outlier detection.

### Code Work:

Implemented XGBoost, and built a shallow ANN for binary classification.

### PPT Work:

Updated charts comparing different model metrics.

### Conclusion:

Deep learning shows promise. Focus next week on model tuning and evaluation.

## Week 5

### Synopsis:

Dedicated this week to performance evaluation and hyperparameter tuning.

### Research Paper Work:

Explored optimization methods—Grid Search vs. Random Search.

### Literature Review:

Reviewed studies on fraud detection trade-offs: precision vs. recall.

### Code Work:

Used GridSearchCV for tuning Random Forest and XGBoost. Improved F1 score.

### PPT Work:

Added tuning approach and new performance results.

### Conclusion:

Model improvements observed. Begin unsupervised anomaly detection in Week 6.

## Week 6

### Synopsis:

Focused on unsupervised learning techniques for anomaly detection.

### Research Paper Work:

Studied Isolation Forest, One-Class SVM, and clustering-based methods.

### Literature Review:

Analyzed use cases where labels are scarce or unreliable.

### Code Work:

Implemented Isolation Forest and Local Outlier Factor (LOF). Evaluated unsupervised results.

### PPT Work:

Created side-by-side comparison of supervised vs. unsupervised methods.

### Conclusion:

Unsupervised methods show potential. Will proceed to combine models in Week 7.

## Week 7

### Synopsis:

Integrated hybrid approaches to enhance detection accuracy.

### Research Paper Work:

Explored hybrid model architectures in fraud detection.

### Literature Review:

Studied stacking and voting classifiers for ensemble models.

### Code Work:

Implemented hybrid stacking classifier combining RF, XGBoost, and ANN.

### PPT Work:

Documented the hybrid model pipeline and its advantage.

### Conclusion:

Hybrid models improved performance. Week 8 will focus on validation and real-time simulation.

## Week 8

### Synopsis:

Performed advanced model validation and real-time simulation of fraud transactions.

### Research Paper Work:

Examined real-time processing systems like Apache Kafka and Spark.

### Literature Review:

Checked literature on deploying ML models in real-world fraud systems.

### Code Work:

Created a batch simulation pipeline and tested model response times.

### PPT Work:

Included architecture diagram for real-time fraud detection system.

### Conclusion:

Real-time analysis successfully simulated. Will work on GUI and final integration.

## Week 9

### Synopsis:

Focused on building a simple GUI and preparing final deployment framework.

### Research Paper Work:

Looked into Flask and Streamlit-based ML interfaces.

### Literature Review:

Reviewed fraud dashboard tools and visualization frameworks.

### Code Work:

Built a GUI using Streamlit showing transaction status (fraud/not fraud).

### PPT Work:

Added GUI screenshots and deployment architecture.

### Conclusion:

Deployment phase underway. Final touches and documentation planned for Week 10.

## Week 10

### Synopsis:

Finalized documentation, report, and presentation materials. Performed project review.

### Research Paper Work:

Compiled references and completed report writing.

### Literature Review:

Summarized all readings and highlighted contributions.

### Code Work:

Cleaned and documented code, uploaded to GitHub.

### PPT Work:

Final capstone presentation prepared with results, graphs, and conclusions.

### Conclusion:

Project successfully completed. Fraud detection analytics model ready with documented report and prototype interface.