**Capstone Project Summary: Credit Card Fraud Detection**

**1. Project & Model Development Summary**

The goal of this project was to develop a robust machine learning model that can detect fraudulent credit card transactions while minimizing false positives and maximizing recall. The dataset contained 259,335 records simulated from 2019–2020, with a significant class imbalance between legitimate and fraudulent transactions.

To address this, a comprehensive modeling pipeline was followed:

* **Preprocessing & Scaling**: Handled missing values and normalized features using StandardScaler.
* **Imbalance Handling**: Applied both **SMOTE** and **ADASYN** to generate synthetic fraud samples.
* **Model Selection**: Compared **KNN**, **XGBoost**, and **Logistic Regression** using cross-validation and multiple metrics.
* **Hyperparameter Tuning**: Used **StratifiedKFold** and GridSearchCV to tune XGBoost.
* **Evaluation Metrics**: Used confusion matrix, F1-score, TPR, FPR, and ROC AUC for evaluation.

**2. Root-Cause Analysis (Hypothesis Framework)**

Based on exploratory data analysis and domain reasoning, I generated several hypotheses that potentially indicate fraudulent behaviour:

1. **H1**: Fraudulent transactions deviate significantly from typical spending patterns.
2. **H2**: They often occur at unusual times (e.g., late night or holidays).
3. **H3**: They involve unfamiliar merchant categories or locations.
4. **H4**: Fraud tends to be clustered (multiple frauds in a short time).
5. **H5**: Transaction amount and frequency differ sharply from a user’s usual behaviour.

These hypotheses informed feature engineering and helped in validating model behaviour during evaluation.

**3. Model Performance Summary**

**Final Metrics (Best XGBoost Model):**

| **Metric** | **Value** |
| --- | --- |
| **TPR (Recall)** | 90.0% |
| **FPR** | 1.27% |
| **Accuracy** | 99.0% |
| **ROC AUC** | 0.993 |
| **F1-Score (fraud class)** | 0.50 |
| **Best Parameters** | max\_depth=6, n\_estimators=200, learning\_rate=0.1 |

**Class-wise Scores:**

Class 0 (Legitimate): Precision = 1.00, Recall = 0.99

Class 1 (Fraud): Precision = 0.35, Recall = 0.90

**Cross-Validation F1 Scores:**

| **Model** | **F1 Score** |
| --- | --- |
| **KNN** (SMOTE) | **0.9916** ✅ Best |
| XGBoost (raw) | 0.9891 |
| Logistic Reg. | 0.8326 |

**4. Business Impact**

**Fraud Detection Benefits**

* **90% of fraud cases were successfully identified**, significantly reducing the risk of monetary loss.
* A **low 1.27% false positive rate** ensures that only a small fraction of genuine users experience false alerts, protecting the customer experience.

**Cost Saving Potential**

Assuming an average fraud cost of ₹5,000 per missed transaction:

* Prevented frauds: **1,351 (TP)** × ₹5,000 = **₹6,755,000** saved

Missed frauds:

* **150 false negatives** still pose risk → Model can be retrained regularly for further improvement

**Real-World Deployability**

* The model is scalable and deployable with minimal latency using **XGBoost**.
* KNN, while slightly better in F1-score, may be slower in production due to memory and distance computations.

**Conclusion**

After extensive model evaluation, **KNN and XGBoost emerged as top-performing models**. XGBoost was selected for deployment due to its speed, scalability, and solid performance (F1 = 0.989, ROC AUC = 0.993).  
The system demonstrates excellent **fraud-catching capability** with minimal customer disruption, and offers a **direct financial benefit** by reducing fraudulent losses. The solution is effective, interpretable, and suitable for real-time fraud detection systems.

**Business Impact**

The chosen model (KNN) has significant **real-world implications**:

* **Fraud Detection Accuracy**: With a 90% recall, the system detects 9 out of 10 fraudulent transactions, reducing financial loss and regulatory risk.
* **Customer Satisfaction**: With a false positive rate of only 1.27%, very few genuine transactions are wrongly flagged—preserving customer trust and experience.
* **Efficiency**: The ROC AUC of 0.993 demonstrates the model’s strong overall classification performance, allowing financial institutions to act with high confidence.

**🧩 Final Takeaway**

“By combining data balancing techniques like SMOTE/ADASYN with powerful algorithms such as KNN and XGBoost, I developed a robust fraud detection system capable of high recall and minimal false alarms. The model's impact translates directly into fraud loss prevention and improved customer satisfaction—making it a valuable asset for any financial service provider.”