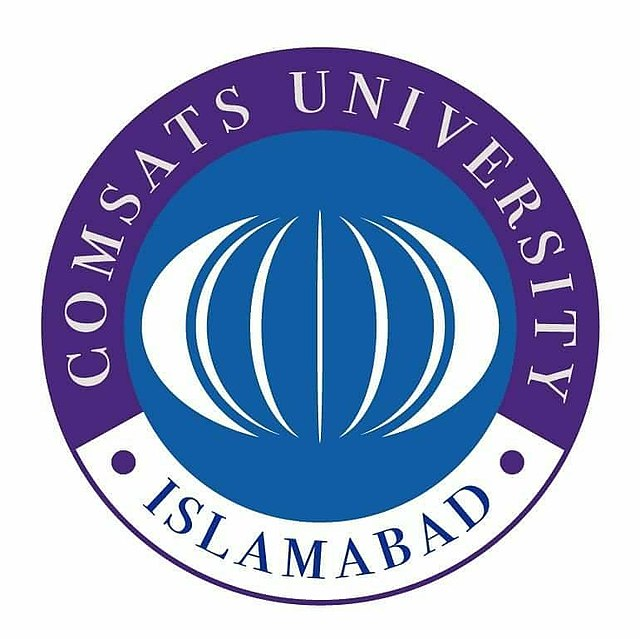
****

**Project Deliverable**

**Group Members:**

- Cybil Fatima(SP23-BAI-013)

-Esha Alvi(SP23-BAI-015)

**Project: Credit Card Fraud Detection Using Machine Learning Models**

---

**Objective:**

The goal of this project is to develop an efficient machine learning model to detect credit card fraud using a dataset of credit card transactions. Multiple machine learning models, including Naive Bayes, Logistic Regression, Random Forest, Xgboost classifier and Support Vector Machine (SVM), are implemented and their performances enhanced using hyperparameter optimization techniques such as Grid Search and Randomized Search.

---

**Methodology:**

**1. Data Preprocessing:**

* **-Data Extraction:** The dataset, provided as a ZIP file, was extracted and loaded into a pandas DataFrame for further analysis.

- **Data Cleaning:**Missing values and duplicate entries were removed to ensure clean data.

- **Outlier Handling:** Extreme values in the 'Amount' feature were capped using the Interquartile Range (IQR) method to minimize their impact on the model.

**2. Feature Scaling:**

- The 'Amount' and 'Time' features were standardized to bring all features to a similar scale, improving the model's performance, especially for models sensitive to feature scaling.

**3. Data Visualization:**

- A correlation heatmap was generated to identify relationships between features.

- Visualizations, such as boxplots, were used to explore the distribution of the 'Amount' feature for fraudulent and non-fraudulent transactions.

**4. Model Training:**

**- Naive Bayes:**

- A simple probabilistic classifier that assumes feature independence. It served as a baseline model for fraud detection, providing a benchmark for comparing more complex models.

**- Logistic Regression:**

- A linear model that predicts the probability of fraud based on input features. It’s useful when the relationship between features and the target is linear and provides clear interpretations of how individual features influence fraud prediction.

**- Random Forest:**

- An ensemble of decision trees that can capture non-linear patterns and feature interactions in the data. This model’s complexity made it effective in modeling fraud due to its ability to handle intricate relationships between features.

-**XGBoost Classifier:**

- An advanced ensemble technique leveraging gradient boosting, applied with hyperparameter optimization. It demonstrated high accuracy and efficiency in handling imbalanced datasets.

**- Support Vector Machine (SVM):**

- A powerful algorithm that maximizes the margin between classes in a transformed feature space. SVM with SMOTE balancing and hyperparameter tuning using Randomized Search was applied, resulting in competitive performance metrics.

**5. Handling Imbalanced Data:**

- Given the imbalanced nature of the dataset (fraudulent transactions being rare), oversampling of the minority class (fraud) was performed using SMOTE to create a balanced training set.

- Additional techniques like custom oversampling with StratifiedKFold Cross-Validation were utilized to ensure robust training on balanced subsets.

**6. Model Evaluation:**

- Each model was evaluated on the test set using accuracy, confusion matrix, and classification report (including precision, recall, and F1-score) to assess performance in detecting fraud.

- Custom cross-validation was implemented for SVM with SMOTE to ensure fair evaluation and provide detailed metrics for each fold.

**7. Hyperparameter Optimization:**

- Grid Search and Randomized Search were used to fine-tune the hyperparameters of the models, optimizing their performance further.

- The SVM model, using Randomized Search, identified optimal values for the regularization parameter (C) and kernel type, significantly improving its fraud detection capability.

- The XGBoost Classifier also benefitted significantly from hyperparameter tuning, with the best parameters identified for enhanced detection capabilities.

---

**Expected Outcomes:**

- The project delivers a robust fraud detection system capable of identifying fraudulent credit card transactions with high accuracy.

- Model performance is improved by addressing class imbalance, optimizing hyperparameters, and selecting the best model based on evaluation metrics.

---

**Impact and Benefits:**

- **Real-world Application**:This fraud detection system can be integrated into payment systems to identify and prevent fraudulent activities, reducing financial losses.

- **Scalability:** The methodology can be applied to large-scale transaction datasets to detect fraud on a global scale, offering immense potential for use in financial institutions and online transaction platforms.

---

**Conclusion:**

This approach aims to build an accurate and reliable system for credit card fraud detection. By safeguarding financial transactions, the system enhances trust in digital payment systems while providing a scalable solution to combat fraud effectively.

Additionally, the integration of advanced techniques such as SMOTE for handling class imbalance, SVM with Randomized Search for hyperparameter tuning, and XGBoost Classifier with optimized settings highlights the project’s innovative approach to tackling this critical problem. The system's evaluation metrics underscore its potential for real-world deployment, ensuring accuracy and reliability in fraud detection.