Capstone Project 2

Fraud Detection in Credit Card Transactions using Logistic Regression, Support Vector Machine (SVM), Random Forest, Artificial Neural Networks (ANN)

Name: CHEVALA AKSHAY RAJ

Batch: DST 20923

Start Date: SEPTEMBER 15 2023

Problem Statement:

The primary goal of this project is to develop an effective fraud detection system for credit card transactions using machine learning techniques. The dataset taken from Kagle and contains information about various transactions, including features such as transaction amount, time, and a binary class indicating whether the transaction is fraudulent or not. The objective is to build and evaluate different machine learning models that can accurately identify fraudulent transactions, thereby enhancing the security of credit card transactions.

Techniques/Approaches Used:

1. **Logistic Regression:** A widely used linear classification algorithm suitable for binary classification tasks.
2. **Support Vector Machine (SVM):** Utilized for its ability to create a hyperplane that effectively separates classes in the feature space.
3. **Random Forest:** An ensemble learning method combining multiple decision trees to improve accuracy and robustness.
4. **Artificial Neural Network (ANN):** A deep learning model with multiple layers, capable of learning complex patterns in the data.

Why These Techniques:

- **Diversity of Models:** Different models provide diverse approaches to address the fraud detection problem, allowing for a thorough evaluation of their performance.
- **Interpretability and Complexity:** Logistic Regression provides simplicity and interpretability, while SVM and Random Forest handle non-linear relationships. ANN, as a deep learning model, captures complex patterns.

Steps Followed:

1. **Data Exploration:** Explored the dataset's dimensions, characteristics, and distribution of classes. Investigated the statistics of the 'Amount' feature and checked for missing values.
2. **Data Processing:**
    - Dropped rows and columns with missing values.

- Scaled the 'Amount' feature using StandardScaler.
- Eliminated duplicate rows in the dataset.

3. **Model Training:** Trained Logistic Regression, SVM, Random Forest, and ANN models using the processed data.
4. **Model Evaluation:**
    - Evaluated models using accuracy, confusion matrix, and classification reports.
    - Plotted precision-recall and ROC curves for logistic regression.
5. **Visualization:**
    - Visualized support vectors and the hyperplane in the SVM model.
    - Plotted feature importance for Random Forest.

Results:

- **Logistic Regression:** Accuracy: 99.90%
- **SVM:** Accuracy: 99.93%
- **Random Forest:** Accuracy: 99.95%
- **ANN:** Accuracy: 99.94%

Qualities:

- High accuracy across all models, indicating their effectiveness in fraud detection.
- Support Vector Machine demonstrated high precision and recall.
- Random Forest exhibited strong feature importance, providing insights into influential features.

Suggestions:

- **Model Deployment:** Consider deploying the selected model in a real-time credit card transaction monitoring system.
- **Continuous Monitoring:** Regularly update and retrain models to adapt to evolving fraud patterns.
- **Explainability:** Provide explanations for model predictions, especially in cases where interpretability is crucial.
- **Feature Engineering:** Explore additional feature engineering techniques to enhance model performance.

Project Conclusion:

This project successfully addressed the challenge of fraud detection in credit card transactions using a variety of machine learning models.