|  |  |
| --- | --- |
| CREDIT CARD FRAUD DETECTION  IBM PROJECT | Abstract  Process for creating a machine learning-based system that examines transaction data in real-time to identify credit card fraud while avoiding false positives.  IBM  APPLIED DATA SCIENCE |

Indeed, here is a thorough process for creating a machine learning-based system that examines transaction data in real-time to identify credit card fraud while avoiding false positives:

**Workflow: Credit Card Fraud Detection System**

**1. Data Collection:**

* Gather previous transaction information that includes the card details, merchant information, transaction amount, and date.
* Make sure the dataset contains both honest and dishonest transactions.

**2. Data Pre-processing:**

* Missing Values: Identify the values that need to be filled in or eliminated, as applicable.
* Identifying and dealing with outliers that can impact model performance is known as "outlier detection and treatment."
* Clean up your data by eliminating duplicates and fixing any entry mistakes.
* Scaling of the feature: Normalize or normalize the numerical characteristics.
* One-hot encoding or label encoding are both effective ways to encode categorical characteristics.

**3. Feature Engineering:**

Develop Useful Features:

* Calculate the quantity of transactions that occur for each card for a certain period of time (such as an hour or a day).
* Amount Calculate deviations from the historical transaction amounts on the card to find out whether there are any odd spending trends.
* Features Based on Time: Extract characteristics linked to time, such as weekday, hour, and holiday indications.

**4. Data Splitting:**

* Training, validation, and test sets should be divided up into three subsets from the pre-processed data (e.g., a 70-15-15 split).

**5. Model Selection:**

* Select appropriate machine learning techniques for fraud detection, such as neural networks, random forests, gradient boosting, and logistic regression.
* To perform better, take into account using ensemble techniques or deep learning models.

**6. Model Training:**

* Train the chosen model or models on the training data.
* Use suitable solutions to address class imbalance, such as over- or under-sampling, or specialised methods like Synthetic Minority Over-sampling Technique (SMOTE).

**7. Model Evaluation:**

Utilising pertinent metrics, assess the model(s) on the validation dataset:

* To gauge the general accuracy of something.
* To assess the proportion of frauds that were accurately predicted among all frauds that were forecasted.
* Remember: To calculate the proportion of frauds that were accurately predicted to all actual frauds.
* To balance recall and accuracy for the F1 score.
* To evaluate the model's capacity to differentiate across classes using ROC-AUC.

**8. Hyper-parameter Tuning:**

* To obtain the appropriate precision/recall trade-off, fine-tune the model hyper-parameters.
* Adjust model sensitivity by experimenting with various categorization criteria.

**9. Model Deployment:**

* Install the top-performing model in a system that processes credit card transactions in real time.
* Implement monitoring and warning systems for instantaneous fraud detection and reaction.

**10. Testing and Validation:**

* Before putting the system into production, carry out rigorous testing in a safe setting.
* Utilise real-time data to verify the system's performance in a practical setting.

**11. Monitoring and Maintenance:**

* Real-time performance of the system should be continuously monitored.
* Implement a model retraining procedure to accommodate changing fraud trends.

**12. Documentation and Reporting:**

* Write down every stage of the approach, including the data sources, pre-processing techniques, Engineering of features, model choice, and deployment processes.
* Inform stakeholders on a regular basis about system performance and fraud detection results.

**13. User Training and Feedback Loop:**

* In the financial institution, instruct users and pertinent staff on how to utilise the system efficiently.
* Create a feedback loop to collect suggestions from stakeholders and users for continuing improvements.

**14. Compliance and Security:**

* Verify that the system conforms with data protection laws and industry norms, such as PCI DSS.
* Put strong security procedures in place to safeguard sensitive consumer data.

**15. Project Conclusion and Evaluation:**

* Make a final assessment of the project's performance in relation to the stated goals.
* Identify areas for future improvements and lessons learned.

By using this process, you may create a real-time credit card fraud detection system that efficiently identifies fraud while reducing false positives, so promoting the safety of financial institutions and preserving client confidence