Architecture Design

Fraud Transaction Detection

Revision Number: 0.1

Author - Deepraj Arya

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	14 JAN 2024	Deepraj Arya	Architecture Design

Reviews:

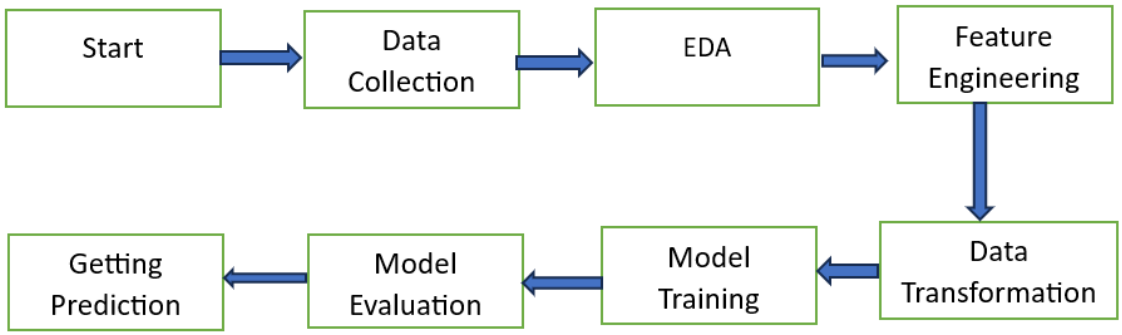
VERSION	DATE	REVIEWER	COMMENTS
0.1	14 JAN 2024		

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY	APPROVED BY	COMMENTS

Architecture Design

Architecture



- src: Contains the source code for the project.
  - Fraud_TX: Main module for the fraud detection system.
    - components: Components for data ingestion, data transformation, and model training.
    - pipelines: Pipelines for data processing, model training and prediction.
    - utils.py: Utility functions.
- app.py: Flask application for web interface to get prediction
- notebooks: Contains Jupyter notebooks (e.g., for exploratory data analysis) for understanding the data to implement CI/CD pipelines.

1. Data Ingestion and Storage:

1.1 **Data Sources:** Data may come from various sources like databases, e-commerce platforms, or external sources. Ingest data from these sources. This is project of iNeuron so they provided the data for this project.

1.2 **Data Storage:** Store the data in a database or data warehouse. Consider using technologies like PostgreSQL, MySQL, or cloud-based solutions like Amazon Redshift or Google BigQuery for scalability.

2. Data Preprocessing:

Implementing ETL (Extract, Transform, Load) processes to clean and preprocess the data. This includes handling missing values, outliers, and transforming the data into a usable format.

- ### Exploratory Data Analysis (EDA):

  Exploratory data analysis is one the most important stage of project. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, find interesting relations among the variables. Tools (python libraries) like Pandas, Matplotlib, or Seaborn are used to perform EDA.

- ### Feature Engineering:

  Feature engineering is a crucial aspect of building an effective fraud transaction detection system. In the context of fraud detection, feature engineering involves selecting, transforming, and creating relevant features (variables) from raw data to improve the performance of machine learning models. These features help the model distinguish between legitimate and fraudulent transactions.

3. Data Transformation :

Explaining the data transformation phase involves detailing how raw data is processed and modified to make it suitable for consumption by downstream systems or applications. Following are data transformation steps used in this model.

- Simple Imputer
- Standard Scalar
- Robust Scalar
- One Hot Encoder

4. Model Building :

- **Model Training –**

  Explaining the model building phase involves detailing how machine learning models are developed and trained to achieve specific objectives. In this project for fraud detection, logistic regression and decision tree algorithm is used for classification model. Model training stage is initiated using training_pipeline.py file in one-shot.

- **Model Validation –**

  The model validation stage is a critical step in the machine learning lifecycle, and its importance cannot be overstated. This stage is designed to assess the performance and generalizability of a machine learning model. Moving with model validation stage, Cross validation and accuracy_score is used in this project.

5. Prediction :

The preprocessed and transformed features are input into the trained fraud detection model. The model applies its learned patterns and algorithms to generate a prediction or a class score indicating the likelihood that the given transaction is fraudulent where 0 is a valid transaction and 1 is fraud transaction.