AI training - hackathon

Use case: Credit Card Fraud Detection

Team: Samjog, Vadi, Krishna, Vamsee, Srinivasan

Dataset: Credit Card Fraud Prediction (kaggle.com)

This dataset offers a variety of attributes valuable for comprehensive analysis. It contains 555,719 instances and 22 attributes, a mix of categorical and numerical data types. Importantly, the dataset is complete with no null values. Here's a breakdown of the attributes:

- Trans_date_trans_time: Timestamp of the transaction (date and time).
- Cc_num:Unique customer identification number.
- MerchantThe merchant involved in the transaction.
- Category Transaction type (e.g., personal, childcare).
- Amt:Transaction amount.
- First: Cardholder's first name.
- Last: Cardholder's last name.
- Gender: Cardholder's gender.
- Street: Cardholder's street address.
- City: Cardholder's city of residence.

Dataset: Credit Card Fraud Prediction (kaggle.com)

- State: Cardholder's state of residence.
- Zip: Cardholder's zip code.
- Lat:Latitude of cardholder's location.
- Long: Longitude of cardholder's location.
- City_pop:Population of the cardholder's city.
- Job:Cardholder's job title.
- Dob: Cardholder's date of birth.
- Trans_num: Unique transaction identifier.
- Unix_time: Transaction timestamp (Unix format).
- Merch_lat:Merchant's location (latitude).
- Merch_long: Merchant's location (longitude).
- Is_fraud:Fraudulent transaction indicator (1 = fraud, 0 = legitimate). This is the target variable for classification purposes.

Classification using Auto ML

- ✓ Create Data
- ✓ Create a compute cluster
- ✓ Create a job – automated ML job and submit
- ✓ It takes approximately 15 minutes to complete the job
- ✓ Once the job is completed, you can view the child jobs, metrics, algorithm chosen etc.
- ✓ Deploy the model for consumption

Question: How do we select the algorithm that we prefer in Auto ML?

Classification using code

- ✓ Import libraries
- ✓ Read data
- ✓ Perform EDA
- ✓ Train the model
- ✓ Predict the results
- ✓ Print the accuracy