Sentiment Analysis on tweets about remote work

Problem:
Credit card fraud by definition is the fraudulent use of a credit card through the theft of the cardholder's personal detail. With the rise of e-commerce, the problem of fraudulent use of credit cards has become more acute.

Given previous transaction datas, can we predict whether or not a new transaction is a fraudulent one ?

Losses to fraud incurred by payment card issuers worldwide reached $27.85 billion in 2018 and are projected to rise to $40.63 bullion in 10 years. Answering the above problem can reduce these damages faced by merchants and credit card issuers. Our goal is to correctly classify the minority class of fraudulent transactions.

Data:
Industrial dataset

Approach:

1. Preprocessing:
    - One hot encoding for categorical data
    - Remove redundant columns
    - Remove duplicate entries
    - Impute missing values
2. Feature Engineering:
    - Create new features: length of last 4 digits, merchant that contains '.com'
3. EDA:
    - Check data distribution
    - Measure Spread
    - Datetime trends
    - Correlation matrix
    - Hypothesis Test
4. Training**:**
    - One hot encoding for categorical variables
    - 3 models: Logistic Regression, Random Forest, XG Boost
5. Evaluation:
    - ROC AUC
    - Confusion Matrix