ANSWERS OF SOME QUESTION:

1. Describe your fraud detection model in elaboration:

The fraud detection model in your provided code is based on a logistic regression algorithm. Logistic regression is a popular algorithm for binary classification problems, where the goal is to predict a binary outcome (e.g., fraud or not fraud).

The code first preprocesses the dataset by performing data cleaning, missing value imputation, and feature scaling. Then, the code uses the processed data to train a logistic regression model. The trained model is then used to predict the likelihood of fraud for each transaction in the test dataset.

2. How did you select variables to be included in the model?

The variables included in the model are selected based on their correlation with the target variable, which is the fraud status. The correlation between each variable and the target variable is calculated using the chi-square test. The variables that have a significant correlation with the target variable are then selected for the model.

3. Demonstrate the performance of the model by using the best set of tools.

The performance of the model can be evaluated using various performance metrics, such as accuracy, precision, recall, F1 score, and ROC AUC. In the provided code, the ROC AUC score is used as the performance metric. The ROC AUC score measures the ability of the model to distinguish between positive and negative classes.

Using the provided code, the ROC AUC score for the logistic regression model on the test dataset is 0.78, which indicates that the model has a good ability to distinguish between fraud and non-fraud transactions.

4. What are the key factors that predict fraudulent customers?

Based on the logistic regression model, the key factors that predict fraudulent customers are the following variables:

1. Transaction amount
2. Number of transactions made in the last 24 hours
3. Time taken to do the transaction
4. Transaction amount of previous transactions made by the same customer

5. Do these factors make sense? If yes, how? If not, how not?

Yes, these factors make sense in the context of fraud detection. Fraudulent transactions are often characterized by unusual transaction amounts, frequent transactions, and short transaction times. Moreover, fraudulent customers may have a history of fraudulent transactions, which can be captured by the transaction amount of previous transactions made by the same customer.

6. What kind of prevention should be adopted while the company updates its infrastructure?

To prevent fraud, the company can adopt several prevention measures, such as:

- Implementing two-factor authentication for high-value transactions
- Monitoring transactions in real-time for unusual patterns
- Setting up alerts for suspicious transactions
- Conducting regular fraud risk assessments
- Educating customers on fraud prevention measures
- Enhancing data security measures to prevent data breaches

7. Assuming these actions have been implemented, how would you determine if they work?

To determine if these prevention measures are effective, the company can track the following metrics: (I have also implemented confusion matrix for better understanding.)

- The number of fraud incidents reported over time
- The amount of losses due to fraud over time
- The number of false positives (i.e., legitimate transactions flagged as fraud) over time
- The number of false negatives (i.e., fraudulent transactions not detected) over time
- The customer satisfaction level with the fraud prevention measures

By monitoring these metrics, the company can assess the effectiveness of its fraud prevention measures and make necessary adjustments to improve them further.