Task 2: Description of the Fraud Detection Model

I used the Random Forest Classifier for Fraud Detection because:

1. Random Forest Classifier is an ensemble method for classification task, which predict through majority vote hence increase accuracy and decrease error rate.
2. It can handle numerical and categorical variables more efficiently which makes the model Robust( to Overfitting).

Task 3: Selecting the Variables for the model

For selecting the features, I used domain knowledge that can impact fraudulent behaviour such as ‘amount’: Transaction amount, ‘type’: Type of Transaction, ‘oldbalance and newbalance’: Balance before and after the Transaction.

Task 4: Demonstration of Model’s Performance and Tools:

I used sklearn matrices such as Accuracy, Precision, Recall, F1-Score and Confusion Matrix to evaluate the model’s performance.

I used the following tools which are required:

1. Pandas and Numpy: For Data Preprocessing and Data Manipulation
2. Matplotlib and Seaborn: For Data Visualisation
3. Sklearn and Scipy: For feature Transformation, model’s training and testing, model’s evaluation etc.

Task 5: Key Factors that Predict Fraudulent Customers:

1. Transaction Type: In the provided Dataset, the Types of Transactions are ‘CASH-IN’, ‘CASH-‘OUT’ , ’DEBIT’ , ‘PAYMENT’ and ‘TRANSFER’ among them ‘CASH-OUT’ and ‘TRANSFER’ are more likely to Fraud.
2. Transaction Amount: Amounts that are closest to the threshold set by organisation are more likely to Fraud.
3. Old and New Balance: Immediate Drop or Immediate Hike in Transaction Balance is indicators of Fraudulent Activities.

Task 6: Do these factors make sense:

Yes, These factors make sense,

1. Some Type of Transactions are red flags usually are those involved in high liquidating such as high amount withdrawal or transfer.
2. Large Transaction Amount is indeed a red flag because most fraudsters are maximise their gains in a single transaction by transferring or withdrawing in lump –sum, which is abnormal in comparing to the regular transaction amounts.
3. Unusual balance record also is a red flag because customers involved or likely to be involved in fraud have unbalanced records of Transactions.

Task 7: kind of Prevention should be adopted while the company update its infrastructure:

1. Regular Auditing
2. Real-Time Monitoring

Task 8: Assuming these actions have been implemented, how would I determine if they work?

Assuming these actions have been implemented, then I determine that it works effectively by following some approaches that are:

1. A/B Testing: Implementing changes in a controlled manner and comparing the outcomes with a control group.
2. Collect feedback from the monitoring system, customers, and fraud detection teams