Fraud Prediction

Business Problem: Every day we do several types of transactions, but sometime people are involved in fraud, unknowingly. Is there any characteristic in the transactions that can identify if the transaction is fraudulent? Is there some way we can fit an machine learning model to recognize the fraudulent transactions before they occur?

Dataset: There are several fraud related data available, but I have taken the data from the below kaggle link for analysis and prediction.

https://www.kaggle.com/datasets/shriyashjagtap/fraudulent-e-commerce-transactions

1. **Transaction ID**: A unique identifier for each transaction.
2. **Customer ID**: A unique identifier for each customer.
3. **Transaction Amount**: The total amount of money exchanged in the transaction.
4. **Transaction Date**: The date and time when the transaction took place.
5. **Payment Method**: The method used to complete the transaction (e.g., credit card, PayPal, etc.).
6. **Product Category**: The category of the product involved in the transaction.
7. **Quantity**: The number of products involved in the transaction.
8. **Customer Age**: The age of the customer making the transaction.
9. **Customer Location**: The geographical location of the customer.
10. **Device Used**: The type of device used to make the transaction (e.g., mobile, desktop).
11. **IP Address**: The IP address of the device used for the transaction.
12. **Shipping Address**: The address where the product was shipped.
13. **Billing Address**: The address associated with the payment method.
14. **Is Fraudulent**: A binary indicator of whether the transaction is fraudulent (1 for fraudulent, 0 for legitimate).
15. **Account Age Days**: The age of the customer's account in days at the time of the transaction.
16. **Transaction Hour**: The hour of the day when the transaction occurred.

Methods: I am planning to do exploratory data analysis on the data, to identify fraudulent transactions. I will try to fit a model, that can help to identify if transactions are fraudulent.

Ethical Considerations: Transaction level data sometimes are overly critical and may contain information that can identify a person uniquely. We should always be careful and try to avoid that information being made public. While fitting the machine learning model we should be careful about what features we are using and how sensitive they are.
On the other hand, the term fraud is extremely critical, and we should be incredibly careful, before declaring a transaction as fraud, as it may give the wrong impression to the company, if the prediction is not correct.

Challenges/Issues: The nature of fraud transactions is quite less, so getting data for fraud transaction and fitting model on them is always challenging. We may have to do oversampling to handle this scenario.

References:

https://www.kaggle.com/datasets/shriyashjagtap/fraudulent-e-commerce-transactions