Capstone Project Data Science: Fraud Detection Analysis

DATASET

Project Objectives and Scope

1. **What is the primary goal of your fraud detection model?**
2. **Why are sensitivity and precision important for this project?**

Data Analysis

3. **What is the class distribution of fraud vs. non-fraud transactions in your dataset?**
4. **Does the 'Time' feature help in predicting fraud? How?**

Data Preprocessing

5. **Why should the 'Amount' feature be standardized?**
6. **Which features are dropped during preprocessing and why?**

Model Training

7. **How does Gaussian Naive Bayes handle continuous features?**
8. **What are the steps in training the Naive Bayes model?**

Model Evaluation

9. **How are sensitivity and precision calculated?**
10. **What metrics do you use to evaluate model performance?**

Results and Interpretation

11. **What are the key findings from your model's predictions?**
12. **How do different threshold values affect model performance?**

Model Improvement

13. **What are the limitations of Naive Bayes for fraud detection?**

14. **What other algorithms could improve performance?**

Practical Implementation

15. **How can your model be integrated into a real-time fraud detection system?**
16. **What are the ethical implications of deploying your fraud detection model?**

Technical Implementation

17. **What are the steps to implement Naive Bayes in Python?**
18. **How can cross-validation improve your model?**