Capstone Project Data Science: Fraud Detection Analysis

DATASET

Project Objectives and Scope

1. **How do you define success for your fraud detection model?**
2. **What are the key challenges expected during the implementation of this model?**

Data Analysis

1. **How do you handle multicollinearity in your dataset?**
2. **What role does feature importance play in your analysis?**

Data Preprocessing

1. **What techniques are used to handle categorical data?**
2. **Why is it important to split the dataset into training and testing sets?**

Model Training

1. **How does Gaussian Naive Bayes differ from other variants of Naive Bayes?**
2. **What strategies are used to optimize model parameters?**

Model Evaluation

1. **How do you ensure the reliability of your evaluation metrics?**
2. **What is the impact of false positives and false negatives in fraud detection?**

Results and Interpretation

1. **How do you validate the results of your model?**
2. **What are the implications of your findings for stakeholders?**

Model Improvement

1. **How do you incorporate domain knowledge into your model?**
2. **What advanced techniques, such as ensemble methods, could be considered for improving performance?**

Practical Implementation

1. **How do you ensure the scalability of your fraud detection model?**
2. **What steps are taken to ensure data privacy and security?**

Technical Implementation

1. **How do you manage dependencies and version control for your implementation?**
2. **What are the benefits of using pipelines in machine learning workflows?**