# DIY Unsupervised Learning Prompts for Fraudulent Credit Card Detection

## Business Case

Credit card fraud continues to be a major issue in the financial sector, resulting in billions of dollars in losses annually. Traditional fraud detection systems often face challenges in identifying rare fraudulent transactions, as fraud cases are typically outliers in transaction data. With the surge in digital payments and the increasing volume of transactions, conventional methods struggle to keep pace. To address this, more advanced approaches are needed—ones that can detect anomalies in data without relying on labeled fraud cases. These solutions are particularly valuable in scenarios where only legitimate transaction data is available for training, enabling the detection of novel fraud patterns that have not been encountered before.

## Introduction to Unsupervised Learning

1. What is unsupervised learning, and how does it differ from supervised learning?
2. Explain the concept of clustering and its real-world applications.
3. What are anomaly detection techniques in unsupervised learning, and why are they important?

## Basics of Unsupervised Learning

1. What are the common types of unsupervised learning algorithms?
2. Explain the concepts of clustering, dimensionality reduction, and anomaly detection.
3. What are the key challenges in unsupervised learning compared to supervised learning?

## Anomaly Detection Techniques

1. What is anomaly detection, and how is it applied in real-world scenarios?
2. How does Isolation Forest work for detecting anomalies?
3. Explain the role of Local Outlier Factor in identifying outliers.
4. How does One-Class SVM work for anomaly detection?

## Evaluation of Unsupervised Models

1. How do you evaluate the performance of a clustering model?
2. What are common metrics used to evaluate anomaly detection algorithms?
3. Explain the concept of silhouette score in clustering evaluation.
4. How can you use a confusion matrix for evaluating anomaly detection results?

## Isolation Forest

1. What is the underlying principle of Isolation Forest for anomaly detection?
2. Explain how Isolation Forest isolates anomalies using decision trees.
3. What are the key hyperparameters of Isolation Forest, and how do they affect its performance?
4. What are the advantages of using Isolation Forest over other anomaly detection techniques?
5. In what scenarios would Isolation Forest struggle to detect anomalies?
6. How do you interpret the anomaly score generated by Isolation Forest?

## Local Outlier Factor (LOF)

1. What is the mathematical foundation of Local Outlier Factor for anomaly detection?
2. Explain the concept of “local density” and how it is used in LOF.
3. How does LOF compare the density of a point to its neighbors to determine outliers?
4. What are the key hyperparameters of LOF, such as n\_neighbors, and how do you tune them?
5. What are the strengths and limitations of LOF in detecting anomalies?
6. What are some strategies to scale LOF for large datasets?

## One-Class SVM

1. What is the theoretical basis of One-Class SVM in anomaly detection?
2. Explain the role of the kernel function in One-Class SVM and its impact on model performance.
3. How does One-Class SVM use a hyperplane to separate normal data from anomalies?
4. What are the strengths of One-Class SVM in handling non-linear anomaly detection problems?
5. What are the limitations of One-Class SVM, especially regarding large datasets?
6. How does One-Class SVM compare to Isolation Forest and LOF in performance and scalability?

## Comparative and Combined Use

1. How do Isolation Forest, LOF, and One-Class SVM differ in their approaches to anomaly detection?
2. What are the key scenarios where one algorithm outperforms the others?
3. How do the anomaly scores from each algorithm compare, and what insights can you gain from these differences?
4. Can Isolation Forest, LOF, and One-Class SVM handle real-time anomaly detection? If so, how?
5. How do these algorithms perform on imbalanced datasets with very few anomalies?

## Practical Applications and Case Studies

1. How is unsupervised learning applied in fraud detection?
2. Explain how anomaly detection is used in manufacturing quality control.
3. Describe a case study where clustering provided valuable business insights.
4. How can unsupervised learning be used in network security to identify threats?
5. What are some challenges in deploying unsupervised learning models in practice?

## Best Practices for Unsupervised Learning

1. How do you choose the right algorithm for your unsupervised learning task?
2. What are some tips for improving clustering or anomaly detection results?
3. Explain the importance of domain knowledge in unsupervised learning projects.
4. How do you handle the scalability challenges of unsupervised learning algorithms?
5. What are common pitfalls to avoid when implementing unsupervised learning models?