**Assignment # 1**

**Machine Learning and Data Analytics Case Studies (ML 467T)**

1. Discuss the challenges faced in high-dimensional datasets. Compare and contrast Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in terms of objectives, methodology, and applications.
2. Differentiate between K-Means clustering and Hierarchical clustering. Mention use cases where each algorithm is suitable. List and explain any two applications of clustering in real-world business scenarios.
3. Explain the Naïve Bayes algorithm with derivation from Bayes’ theorem. Briefly explain how Naïve Bayes algorithm works in text classification. What are the strengths and limitations of Naïve Bayes? Suggest techniques to handle correlated features and zero probability problems.
4. Explain the concept of Support Vector Machines (SVMs). Illustrate with a diagram how an optimal hyperplane is constructed. What is the kernel trick in SVM? Explain different kernel functions (linear, polynomial, RBF) and their applications.
5. How does a decision tree method handle categorical vs continuous variables? Describe how splitting criteria such as Information Gain and Gini Index are calculated.