Assignment
Supervised Learning

1. Which technique(s) from the following would you use… [7 Marks]
   - Logistic Regression
   - Support Vector Machines
   - Bayesian classifier
   - K-Nearest Neighbour
   - Principal Components Analysis
   - Decision Trees

   …if you…

   (a) Care only about accuracy and not about interpretation?
   (b) Care about interpretation more than accuracy?
   (c) Scoring throughput is high?
   (d) Want to build a robust and deterministic model?
   (e) Each class has potentially multiple sub-classes?
   (f) Want to visualize data that has no class labels?
   (g) Don't want to build a model at all

2. Five types of problems: [5 Marks]
   1. Classification Problem
   2. Regression Problem
   3. Retrieval Problem
   4. Recommendation Problem
   5. Reasoning Problem

   Identify and explain that below domains belong to which class of problems (any 2 areas)

   i. news.google.com
   ii. youtube.com
   iii. linkedin.com
   iv. ola.com / uber.com
   v. 99acres.com / magicbricks.com

3. List one difference and one similarity between the following: [5 Marks]
   (a) K-nearest Neighbour and K-Means
   (b) PCA and MDS
   (c) Classification and Regression

(d) Bag-of-words and market Basket Analysis
(e) Naïve Bayes and Decision Trees

4. In this problem we will explore what dataset and within those datasets what features you want to create for solving the following problems. [Give at least four features for each problem] **[4 Marks + 4 Marks]**

a) You are a new Fintech start up in town. You build credit models based on people's SMS data (financial related SMS's only). What kind of features will you engineer that will help you predict credit score of your customers?

b) Let's say WhatsApp hires you to build the following model. WhatsApp wants to find the most "popular" people in its customer base for viral marketing. What features on each person you will you use to come up with a popularity score of a person on WhatsApp. (Note: A person might be part of multiple groups)