4. Answers to the following questions:

1. Based on accuracy which model is the best one?

* Support Vector Machines (SVM) with Linear kernel is the best model with 0.98 accuracy.

1. For each of the 6 other models, explain why you think it does not perform as well as the best one:

* MLP classifier: 0.9733 accuracy
  + MLP model’s accuracy do not differ much than the best model. Only one extra observation has been misclassified compared to SVM.
  + As we have not tuned hyper parameters, that could be the reason behind one misclassification.
* Linear Discriminant Analysis: 0.973333 accuracy
  + LDA model’s accuracy do not differ much than the best model. Only one extra observation has been misclassified compared to SVM.
  + Both SVM and LDA are non-parametric and high variance models. But LDA expect the probability distribution of the predictor as Gaussian. This could be the reason behind minor difference between accuracies.
* KNN Classifier: 0.973333 accuracy
  + KNN accuracy do not differ much than SVM. Only one extra observation has been misclassified. However, we used neighbors = 10.
* KNN model classifies by measuring the distance between the observations. This model expects more data to train. As we have a smaller number of observations (150), we need to use more neighbors and that could lead to overfitting.
* Distance function should be meaningful to cover all the observations. We used Euclidean distance and that could be reason behind one misclassification.
* Linear Regression: 0.966667 accuracy
  + This is a regression model which predicts a continuous value and should be rounded off to get the class. When text classes are encoded into numeric values, they become cardinal with a degree of magnitude for each class. But they are nominal values. This property can cause a problem when regression models are used for classification.
  + Linear regression model works basing on assumptions like Linear relation between feature set and label, Constant variance for all features around the regression line, and normal distribution for its residuals. Our dataset might lack few of these properties.
* Polynomial Regression (degree 2): 0.966667 accuracy
  + This is a linear regression model with Polynomial features of degree 2. All the assumptions for Linear regression are applicable to this model as well. This overfits with best accuracy if we eliminate raising the individual features to the given degree.
* Naïve Bayes Classifier: 0.960000 accuracy
  + This model assumes, the features are independent and follow gaussian distribution.
  + Our data has features that are dependent by some factor. This could be a reason for little less accuracy.
* QDA: 0.960000 accuracy
  + This model expects the features having a gaussian distribution and best works for quadratic boundaries. Our data do not have gaussian distribution and we have seen earlier, linear model works good for this data.
* Extra tree classifier: 0.95333 accuracy
  + Extremely randomized trees using boosting and a random set of features. Like SVM, this is also a non-parametric method and relies on entropy thresholds to perform splits and need more training data. Less training data is the reason behind less accuracy.
* Random forest classifier: 0.9467 accuracy
  + Random forest trees using boosting and a random set of features. Like SVM, this is also a non-parametric method and relies on entropy thresholds to perform splits and need more training data. Less training data is the reason behind less accuracy.
  + Within ensemble tree models, compared to Extra tree classifier, this one has less accuracy. Using most discriminative thresholds instead of random ones is the reason.
* Decision Tree classifier: 0.9267 accuracy
  + Like SVM, this is also a non-parametric method and relies on entropy thresholds to perform splits and need more training data. Less training data is the reason behind less accuracy.
  + Within ensemble tree models, this has lowest accuracy because, it do not use boosting or random features by default and less training data add another reason for less accuracy.
* Polynomial Regression (degree 3): 0.913333 accuracy
  + This is a linear regression model with Polynomial features of degree 3. This has less accuracy than the one with features with degree 2.
  + Polynomial features of degree 3, has a complex pattern which could be difficult for linear regression resulting into little less accuracy.