Recommended Models

Classification models would be employed since the outcome to be predicted is yes or no.

Models to be tested

Logistic Regression: It’s used for binary classification problems. It’s simple and easily interpretable. However, it assumes linear relationship between features

Naïve Bayes: It works well on relatively simple task. Naïve bayes has high-bias and low variance classifier, which has advantages over logistic regression and nearest neighbor algorithms when working with a limited amount of data. It’s a good choice for a limited CPU and memory resources.

Random Forest: It is an ensemble of multiple decision trained on different random subsets of training data. It has high performance and flexibility. It is also less prone to overfitting.

Cont’

4. Support Vector Machine: It’s mostly used when the data has exactly two classes. SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. It is vulnerable to data that has more than two classes.

5. Artificial Neural Network (ANN): ANN can learn and therefore be trained to find solutions, recognize patterns, classify data and forecast events. ANN are used for complex and large datasets.

Model evaluations

- These are techniques used to assess how well a model has learnt patterns of data.

❖ Accuracy: Ratio of correctly predicted instances to the total instances.

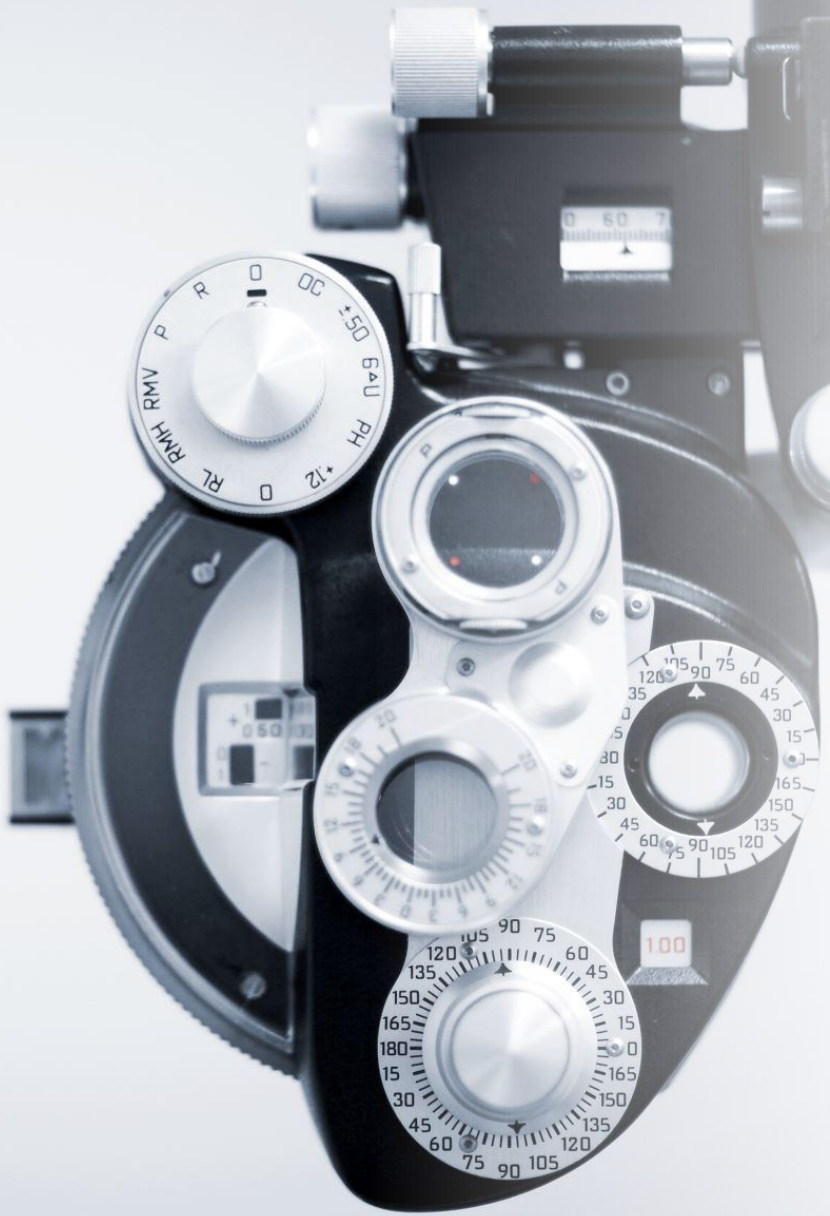
❖ Precision: Ratio of correctly predicted positive observations to the total predicted positive observations.

❖ Recall (sensitivity): Ratio of correctly predicted positive observations to the all observations in actual class.

❖ F1 Score: Harmonic mean of precision and recall.

❖ ROC-AUC: Area under the Receiver Operating Characteristic curve, which measures the ability of a classifier to distinguish between classes.

❖ Confusion Matrix: Tabular summary of actual vs predicted classes.



Hyperparameter Tuning

- It is method used to optimize model performance. Frameworks used include:

1. GridSearchCV
2. RandomizedSearchCV
3. HalvingGridSearchCV
4. HalvingRandomSearchCV