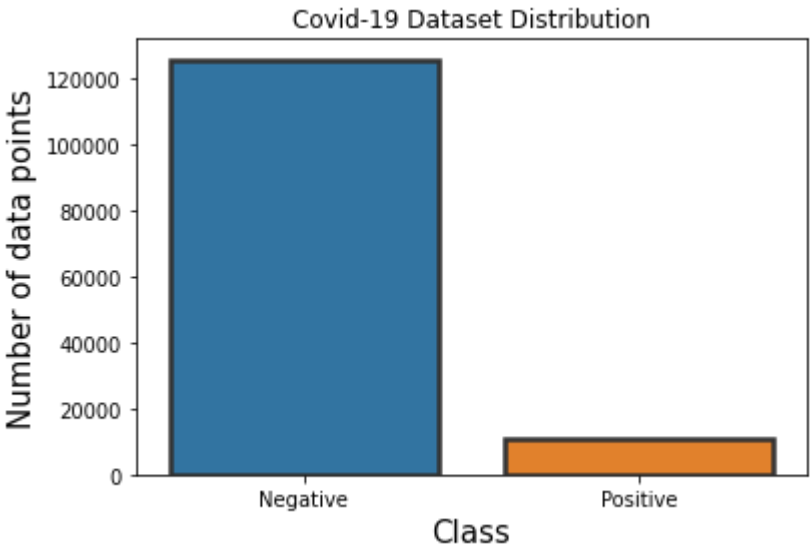
Covid-19 diagnosis prediction based on symptoms

Prediction of covid-19 based on the symptoms of patient Link

Dataset Description

- Phisiological Factors
  - Gender (Male /Female)
  - Age above 60
- Medical Symptoms
  - Cold
  - Cough
  - Fever
  - Sore Throat
  - Shortness of Breath
  - Headache
- Other Features
  - Contact with Covid-19 , or travel from abroad
- Result
  - Test Result of the patient for Covid-19



ML Models used for Prediction

- Support Vector Machine Classification (SVM)
- Lorgistic Regression (Classification)
- k-Nearest Neighbours Classification (kNN)
- Decision Trees Classifier
- Random Forest Classifier
- XG-Boost Classifier
- Gradient Boosting Classifier

DataCleanup

All the data points used in this analysis are binary type values. All String values in the dataset are appropriately transformed into 0's and 1's before training. After cleaning up the data with incomplete features, we had around 170K data points which are used for testing and training purposes.

Metrics used for Comparision

- Accuracy - Correctness of the prediction
    - f1-score-score - Harmonic mean of precision and recall
    - precision - How low is the miscassification(False positive)
    - recall - How much of positive has beeb detected as positive
    - roc_auc - Area under the ROC curve the dataset is unbalanced. so using accuracy as an error metric would not be a proper choice for this unbalanced dataset. For these kinds of problems with unbalanced datasets, we can either f1-score score or the Area under the ROC curve(AUROC).In this project, we have used AORUC as an error metric to validate the performance of our models

Experimentation

All models are modelled with two fold cross validation. WE have used the `gridsearchCV` available in the scikit-learn library to optimise the hyper parameters

Validation of Models

1. SVM

Hyper Parameters Optimisation

- Kernel Function - linear , rbg
- C - 0.1-1

Best Performed: c = 1, Kernel = Linear

Error Metric	Train	Test
Acccuracy	0.958736	0.955266

Error Metric	Train	Test
f1-score	0.693835	0.676758
Precision	0.817031	0.77853
AUROC	0.823634	0.820783