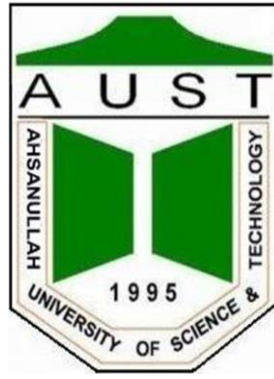# Ahsanullah University of Science & Technology
Department of Computer Science & Engineering
Semester Spring 2021

CSE 4108 Artificial Intelligence Lab

Project Name: COVID 19 DETECTION

Submitted To

Md. Siam Ansary	Tamanna Tabassum
Lecturer,	Lecturer,
CSE, AUST	CSE, AUST

Submitted By

Umma Rumman	180104055
Tahrima Akter Tamanna	180104057
Mohammad Abdur Rafi Farhab	180104068

Project Description

Our project name is Covid19 d etection. It will detect covid over/according to our dataset. For our project we choose classifier ML models. These models give accuracy of our dataset as output. It also gives confusion matric for each model. By this we can compare performances of each model.

Our Dataset

It contains 9 features with more than 300 samples. We provide a google form to random people who experienced covid19 symptoms. Then we made our dataset from the google form. We are taking 50% training data and 50% test data from the dataset.

ML Model

We have used 5 ML classifier models in our project to develop our dataset. A short brief of the models are given below :

1. **Decision Tree :** It is a tree structure classifier where decision node contains features of the dataset and split data over a decision rule. And Leaf nodes contains the outcome. Leaf node also helps to decide the class of new data. This model selects feature by Information Gain techniques. Model will split every possible condition and take the split with maximum information gain. It needs entropy to calculate information gain.

2. **Random Forest :** It is based on the ensemble learning. It is a collection of multiple random decision trees on various subsets of the given dataset. And takes majority votes to predict the test data of the dataset. In this model the number of subset features = square root of total number of features.

3. **KNN (K-Nearest Neighbour):** It is one of the simplest ML model. First it chooses k then calculates distance of test data from all training data. Then it considers majority labels of k closest neighbors. And categorized the test data into that label.

4. **SVM (Support Vector Machine):** It is a supervised learning Classifier.It chooses the extreme points that help in creating the hyperplane.The goal of SVM algorithm is to create best hyperplane boundary so that we can put the new data in the correct category in the future.It is effective in high dimensional spaces where the dimensions number is greater than number of samples.It doesn't perform well in case of large datasets.It uses a subset of training points in the decision function so it is also memory efficient.

5. **Naïve Bayes:** Naive Bayes algorithm is a supervised learning algorithm,which is based on Bayes theorem and used for solving classification problems.It includes high dimensional training dataset.It is the simple and effective algorithm that can make quick predictions.It works on the basis of the probability of an object.It perform well in case of categorial input variables compared to numerical variables.

Comparison of the Performance Metric

Model Name	accuracy	recall	precise	F-Score
Decision Tree	1	1	1	1
Random Forest	1	1	1	1
KNN(K-Nearest Neighbour)	0.8797	0.88506	0.8953	0.8901
SVM(Support Vector)	1	1	1	1
Naïve Bayes	1	1	1	1

Dicussion

After fitting our dataset in different model we can see that for KNN model ,performance metrices are less than other models . But in rest of the four models the performance metrices are same .So, we can say that our dataset works well in Decision tree,Random Forest,Naïve Bayes and SVM models where we found the precision,accuracy,recall and f-score all the metrices are 1.In contrast, for KNN model the accuracy is 0.8797 or 87.97% which is not that much good performance.

Percentage of Contribution

33.33%   33.33%   33.33%
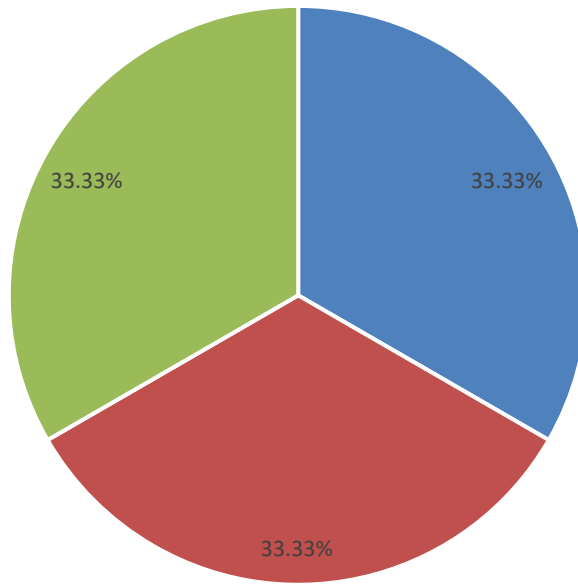
■ 180104055   ■ 180104057   ■ 180104068