Predicting Covid-19 using machine learning

Author: Ana Osorio-Alvarado

Faculty Mentor: Dr. Christoforos Christoforou

ST. JOHN'S UNIVERSITY 150 1870–2020 Faithful to the Mission

Neuro-Intelligence Innovation Lab

Introduction

The global outbreak of COVID-19 has led to a health crisis, currently affecting many countries worldwide. Its rapid spread and variants have led to millions of deaths. Therefore, to reduce the spread of COVID-19, there is a need for effective methods to screen a patient's symptoms. This would result in the prevention of new cases and the reduction in the number of global deaths.

Study Objective

Utilize machine learning approaches to predict positive COVID-19 cases, based on patient's:

- Symptoms
- medical conditions
- exposure
- precautionary measures

Key Findings

Machine learning algorithms can successfully predict COVID-19 cases based on commonly measured symptoms.

Future work:

Model prediction uncertainty using probabilistic modeling.

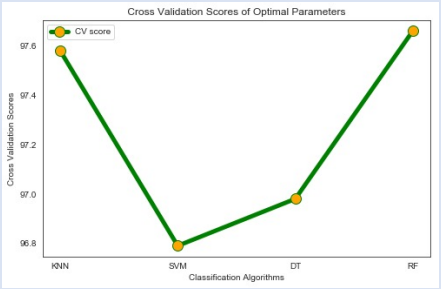
Methods

Classification models considered: K Nearest Neighbor (KNN); Support Vector Machines (SVM); Decision Trees (DT); Random Forest (RF)

Features/ Independent Variables: Medical symptoms; history of medical conditions; indicator of possible exposure; precautionary measures.

Model Selection: We employed cross-validation, for selecting the optimal parameters of each model.

Model	Cross Validation scores	Optimal Parameters
RF	97.66	max_depth = 10, n_estimators = 500
KNN	97.05	k = 2
DT	96.98	criterion=entropy,max_depth= 8.99,splitter= best
SVM	96.79	C=20, degree=3 gamma = auto,kernel = poly

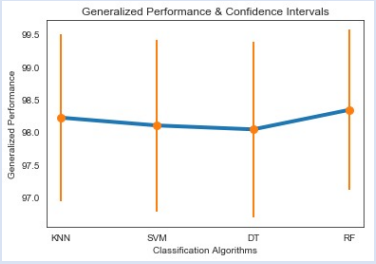


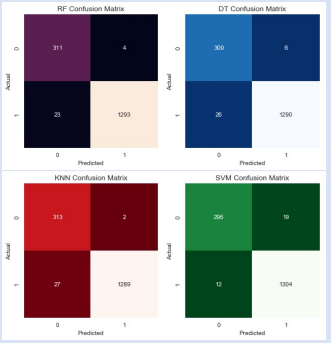
Model Evaluation: Evaluated the model's generalization performance in terms of accuracy and area under the ROC curve on an independent test set.

Results

Generalization performance, using optimal parameters of each model.

Model	Generalized Performance	Confidence Interval	ROC AUC
RF	98.34	(97.73, 98.96)	98.49
KNN	98.22	(97.58, 98.86)	98.66
SVM	98.10	(97.44, 98.76)	96.53
DT	98.04	(97.36, 98.71)	98.06





Confusion matrices