

Coronavirus (COVID-19) Classification using Orange

Zark Ali Jan
zarkali8@gmail.com
Department of computer science
Riphah international University
Islamabad Lahore campus Pakistan

Muhammad Saqib
saqibaliriaz888@gmail.com
Department of computer science
Riphah international University
Islamabad Lahore campus Pakistan

Bilawal Ali Jaan
Jaanbilawal9@gmail.com
Department of computer science
Riphah international University
Islamabad Lahore campus Pakistan

Muhammad Tasawar Faraz
faraz7807@gmail.com
Department of Library and Information
Sciences Minhaj University Lahore

ABSTRACT

Sideways with the Coronavirus disease, additional crisis has established itself in the form of build horror and panic phenomena, powered by incomplete and often wrong information. There is consequently a incredible requirement to report and well know COVID-19's informational crisis and gauge public sentiment, so that suitable messaging and rule conclusions can be applied. In this paper compare different type of classification algorithm working with free available datamining and knowledge discovery programs like Orange. The Precision of classification algorithm such as Random Forest, KNN and also Naïve Bayes algorithm are contrasted utilizing each of components. COVID-19 in India [1] data-set is employed for analyzing the Classification algorithm so as to rectify the individuals without Coronavirus disease.

KEY PHRASES: classification; coronavirus, info ming, data science, machine learning, accuracy, naïve bayes, knn, orange

1. INTRODUCTION

COVID-19 virus was arisen in the end of 2019 at Wuhan region of China. COVID-19 virus displayed fever, cough, fatigue, and myalgias in human body through initial stages. The virus displays the partly alike behaviors with other virus-related pneumonia. So, the dispersion amount of the virus made the condition tough to be below control. There's a demand for automatic applications which may assist individuals in altering those big data right into Info. Now we have lots of free basis data-mining applications and tool are all offered for usage such as for instance the rapidminer [2], Waikato Environment Knowledge Analysis (WEKA) [3], Orange [4], NLTK etc. All these programs and applications furnish a variety of algorithms and methods which support in improved diagnosis of information. These programs assist in audience investigation, information conception, reversion analysis, Conclusion bushes, Prognostic analytics, Text mining.

We ran a comparison analysis amongst classification algorithm including 3 different algorithms like Random Forest, KNN and also Naïve Bayes algorithm with Orange Tool.

2.METHODOLOGY OF LITERATURE REVIEW

In Rational Records such as (sciencedirect, springerlink, pubmed, and Google Scholar) were examined between the period May 20, 2020, to May 27, 2020. Finding a Suitable title and keywords and used. After a search in detail, we have a number of articles found multiple keywords were selected, based on their significance to target studies in Coronavirus Classified,' Science on Coronavirus. A mixture of keywords, thrown in for a detailed search for each specific data. Scientific articles on practical subjects, only for humans (in human subjects), and published in English, were considered for a systematic review. Corona diseases are only beached optimistic logic RNA diseases that are recognized to hold around of the major viral genomes, awake to about 32 kbp in dimension [1–5]. Later rises in the amount of corona disease genome classifications existing following to labors study the range in the rough, the domestic Coronaviridae now holds four types (International Committee on Taxonomy of Diseases). The literature studies are mostly medical studies. The classification, segmentation studies may increase on COVID-19 in the literature.

3. CLASSIFICATION ALGORITHMS

We've researched the next Classification Algorithm within our paper:

- Random Forest
- KNN
- Naïve Bayes

Random Forest: Random forests [5] or designated forests of debate are a learning process for the classification, processing and other activities that work by creating a large number of finishing projects during a work and class outreach that is a form of classes or individual prediction of trees.

NAÏVE BAYES: is a household of uncomplicated probabilistic classifiers in line with employing Bayes' theorem with solid independence assumptions amongst your qualities [6].

KNN: k-NN is a kind of occurrence based learning, or lazy learning, where the role is only approached nearby and all computation is deferred until purpose calculation. Both for classification and regression, a suitable method can be allocate masses to the donations of the neighbors, so that the nearer neighbors contribute more to the average than the more reserved ones. For example, a common weighting scheme consists in giving each neighbor a weight of $1/d$, where d is the distance to the neighbor [7].

4. TOOLS GUIDE

4.1 ORANGE: It has always been a learning process for learning and using databases (written with Python). It is a transparent system for applications to get visual data and analysis, and it can likewise serve as a Python library. The application is archived and processed through the Bioinformatics Laboratory of the School of Computer and Information Science at the University of Ljubljana. Orange can be a very efficient software system for data mining and machine learning and statistics. The elements can be accessed by widgets and all include direct data visualization, stand-alone selection and preparation, robust investigations of understanding calculations and forecasts. After searching the data selection and the external read-out process to be deleted from the Data and Structure after the entire page has been read and you get to Explore the phylogenetic tree.

5. EXPERIMENT

5.1 DATASET: We have downloaded COVID-19 in India Dataset on Novel Corona Virus Disease 2019 in India from the Kaggle. We have total numbers of a row are 2666. In this data collection, we have deaths records, which may be suspected or confirm. Deaths records is a class label used to divide into groups (patient region wise). This data collection has India country 1st January 2020 to 31 may 2020 COVID-19 patient records. The dataset has the following attributes:

S. No

1. Date
2. Time
3. State/unionterritory
4. Confirmedindiannational
5. Confirmedforeignnational

6. Cured
7. Deaths
8. Confirmed

5.2 NAÏVE BAYES CLASSIFICATION ALGORITHM USING ORANGE:

After applying Naïve Bayes, we get 78% accuracy with 0.002% CA. We have 0.001% F1, 0.001% precision and 0.002% recall that clearly show in fig1.

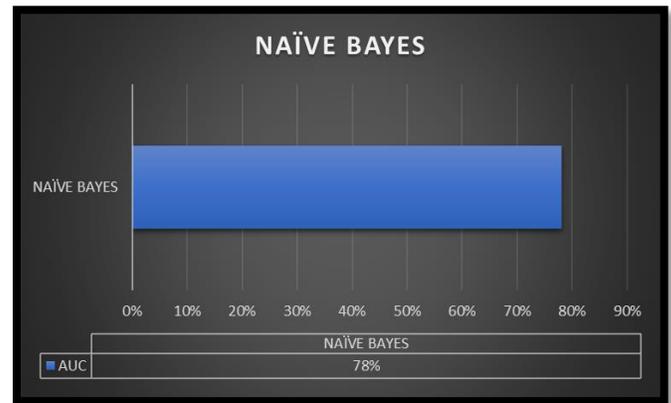


Figure 1: Naïve Bayes

5.3 RANDOM FOREST CLASSIFICATION ALGORITHM USING ORANGE:

After applying Random Forest, we get 86% accuracy with 65% CA. We have 63% F1, 62% precision and 65% recall that clearly show in fig2.

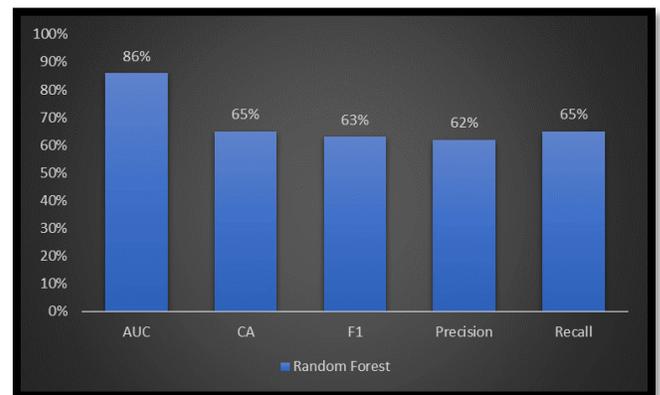


Figure 2: Random Forest

5.4 KNN CLASSIFICATION ALGORITHM USING ORANGE:

After applying KNN, we get 53% accuracy with 37% CA. We have 25% F1, 19% precision and 37% recall that clearly show in fig 3.

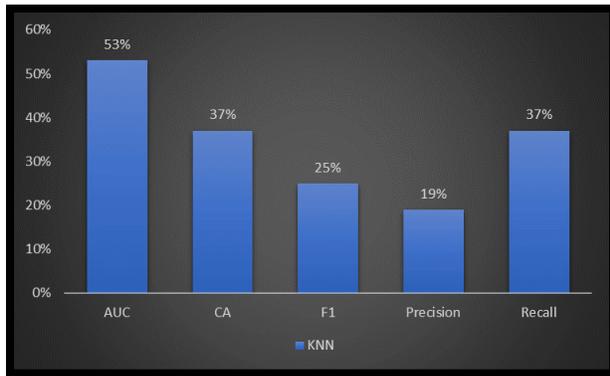


Figure 3: KNN

CONCLUSIONS:

In this paper COVID-19, patient data collection used and get data from the COVID-19 in India Dataset on Novel Corona Virus Disease 2019 in India Kaggle Data Set. It has 2666 rows with 9 attributes and 1 class variable. There was a concert of accuracy classified models result attached, which is as clear as the tracks in Figure 4.

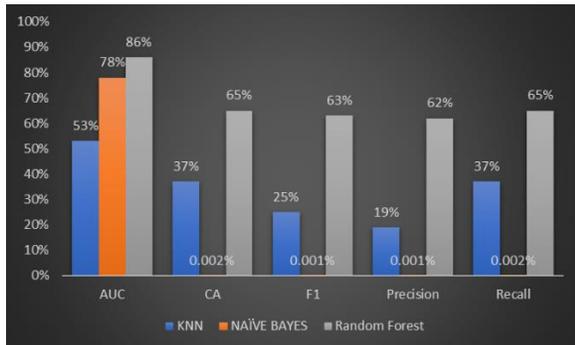


Figure 4: Overall Result

We clearly see that highest accuracy get from random forest and lowest accuracy get from KNN.

REFERENCES:

- [1] <https://www.kaggle.com/sudalairajkumar/covid19-in-india>
- [2] <https://rapidminer.com/>
- [3] <https://www.cs.waikato.ac.nz/ml/weka/>
- [4] <https://orange.biolab.si/>
- [5] https://en.wikipedia.org/wiki/random_forest
- [6] https://en.wikipedia.org/wiki/naive_bayes_classifier
- [7] https://en.wikipedia.org/wiki/k-nearest_neighbors_algorithm
- [8] randhawa, g.s., soltysiak, m.p.m., el roz, h., de souza, c.p.e., hill, k.a. and kari, l. (2020). Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: covid-19 case study. Plos one, 15(4), p.e0232391.
- [9] barstugan, m., ozkaya, u. And ozturk, s. (n.d.). Coronavirus (covid-19) classification using ct images by machine learning methods. [online] available at: <https://arxiv.org/ftp/arxiv/papers/2003/2003.09424.pdf> [accessed 5 jun. 2020].