

Meta-Analysis of Spread of Disease: COVID-19

Kalivaraprasad., B

Assistant Professor, Dept. of. ECE(BES-II), Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Guntur(dt)-522302. baditakali@gmail.com

Abstract—

Tropical life may spread endemic or emerging illnesses via interspecific transmission. The current outbreak of bat-borne COVID-19 emphasizes the need to protect animals. This illness has affected every nation and every age range, from babies to the elderly. Most occurrences of this condition are not lethal, however it may be life-threatening for elderly, newborn, and sick persons. The proposed work was assessed using WEKA to predict parameters and determine the optimum method based on dataset. This article analyzes the spread of CoronaVirus in key states using raw data from the Indian government's website to determine the source and how to manage the problem. Results are formulated utilizing parameters and machine learning methods. IBK, an instance-based learning method using k-parameters from the dataset, is the best classifier for the dataset since it has greatest accuracy and minimal error.

Index Terms—RT-PCR, COVID-19, Logistic, IBK, Machine learning.

I. INTRODUCTION

The world has witnessed many epidemics since many centuries and it has been observed that after every 100 years, the world faces the epidemic which takes the lives of millions of people. In this 21st century, the epidemic has emerged out in the form of COVID-19 which has become the worst pandemic. It is declared as the global health emergency by WHO (World Health Organisation) round 6 month ago. The age groups for whom this pandemic has become deadly are the children which are between the age of 0-10 years, the people which are of the age group of 60 and above and the people who are already suffering from any other disease. This pandemic has also affected adversely to the GDP of every country.

During the pandemic many people lost their lives and the major problem faced by scientists was difficulty in analyzing the behavior of COVID during the second wave in India. Through this paper we analyzed the behavior of COVID-19 on the basis of parameters which included Confirmed cases, recovered cases and death.

The research and studies carried out by epidemiologists examined that the coronavirus was first observed in bats. Bats are the ancestors of human coronavirus. This virus belongs to the family Coronaviridae, order Nidovirales, realm Riboviria and has scientific name as Orthocoronavirinae or Coronavirinae [9]. This virus has positive sense single stranded RNA genetic material and the size of the genetic material is the largest among RNA viruses i.e. approximately 26 to 32 kilobases. Coronavirus has an envelope around it which protects its genetic material during travelling between the host cells [4]. The word Corona, which originated from the Greek word "korone" which means a crown as Corona virus has the molecular structure which has spikes like projections on the surface similar to the crown which resemble the structure of AIDS virus [10]. The first most confirmed case of novel Corona virus was in Wuhan (Hubei Province), China on 17 November 2019 which emerged out rapidly. This pandemic has impacted on many countries which include the most developed and powerful nation USA. The novel Corona virus has adversely affected every country's human resources. Approximately, 743,487 people have lost their life due to this pandemic across the globe. The people which are reported confirmed are 20,405,695 world wide [1]. COVID-19 is a type of indirect contact transmission which usually occurs when an infected person sneezes and the tiny droplets of water or saliva which contains virus spread in the air. Those contaminated droplets when fall to the surface and if a person touch that contaminated surface and then touch his/her nose, mouth, eyes or if inhaled by a healthy person becomes the means of transmission of the diseases. According to the study, this virus can survive in the air for maximum three hours. COVID-19 is closely related to Severe acute respiratory syndrome i.e. SARS. The rate of transmission of COVID-19 is rapidly increasing due to the community spread.

The testing of COVID-19 is done through the samples of nasopharynx or oropharynx swab. The standard way of testing COVID-19 is through PCR (Polymerase chain reaction) in which multiple copies of single stranded RNA are created. Then RT-PCR (Reverse transcriptase PCR) occurs and many processes are done on the sample and after 48 hours the report is generated.

Machine learning is important to train the data in a form where the computer or the machine is able to recognize or learn the data to perform the operation and also help to find the result that may cause some changes and also for the better result various machine learning algorithms are designed based on the behavior of the dataset.

The paper is segregated into seven sections which include Section I consists of two figures in which Fig 1 depicts the pictorial representation of cases in the state of Haryana and the Fig 2 depicts the recent cases of the Punjab in a pictorial way that showcase the comparison between them. Section II which comprises of data acquaintance in which data clustering and a brief introduction of the data is mentioned. Section III composes of structured framework with data segregation which mainly comprises of the refining of raw data. Section IV constitutes proposed work in which detailed information of applied algorithms and techniques are on this mentioned. Section V comprises of experimental investigation in which dataset is scrutinized with help of data tables and data models. Section VI constitutes of result and discussion in which the outcomes of the applied techniques and information of selected classifier is mentioned. Section VII comprises of elicitation by analogy of data which includes a brief overview of the paper. Section VIII constitutes of conclusion in which the important points of the paper are mentioned. Section XI comprises of future work in which the possible work which can be carried out on this data is briefly described and eventually the references are mentioned.



Fig. 1 .Aggregation of recovered, confirmed and decreased cases of Harayana

II. DATAINSPECTION

In India, the first confirmed case was dated at 30, January 2020 in Kerala, a student who returned from Wuhan University of China. The main reason of increasing confirmed cases in India is community spread due to the migration of people from one place to another, lack of social distancing and lack of adapting to preventive measures. In this paper, the two states which are Haryana and Punjab have almost same population are being analysed on the basis of Confirmed cases, Death, recovered cases. The dataset of Haryana and Punjab is made by taking the statistics of each day from 1st April to 29th June from official site of COVID-19 authorized by Ministry of health and family welfare of government of India [11]. The main objective of this paper is to gather information and analyse the dataset using weka tool with the help of various classifying techniques.

The dataset of corona is made by taking the data from various sources that mainly include the official websites for corona authorized by government of India. In this dataset, the data of two states that are nearly same with respect to the population and on the basis of certain parameters, the dataset is made. The analysis of the cases of COVID-19 in each day in those states is reported.

The raw data was made and the dataset consists of 5 attributes and 180 instances. Basically, this data is made on the two states of India which is Haryana and Punjab. This pandemic situation is being analysed on these two states having nearly same population. The dataset is made by taking each and every



Fig.2. Aggregation of recovered, confirmed and decreased cases of Punjab

statistic very precisely and accurate to make the analysis more accurate. The dataset is basically divided into five columns which are the parameter of this dataset which are the attributes. The attribute of this dataset is: Date, State, confirmed cases, decreased cases, and Recovered cases.

The data of the dataset is shown through people graph in which the sun of the cases of each day is taken for every parameter in Fig.1 (a) and Fig. 1(b). In Fig.1 (a) the recent condition of COVID-

19 is shown through this diagram in which the current confirmed cases of Haryana, recovered cases of Haryana and the deaths occurred till in Haryana. In Fig.1(b) the cases in Punjab are shown through this diagram that the active cases in Punjab, recovered cases of Punjab and the deaths occurred in Punjab. This is the overview of the dataset through which analysis of this deadliest pandemic is done. For the analysis on this dataset two States are taken, the reason behind taking these two states are because these states are similar in population and also one near to each other as they have many similarities in land and weather and all the climatically measured things that's why these two states are best suited our analysis. Due to similarity in the climate condition we can analyse more deeply for this pandemic as it is beneficial that in the analysis or in comparison b/w these two states in the parameters there is no difference occur due to climate condition then the analysis will become more precise and accurate. The total population of Haryana is 2.54 crores. The rate of persons affected by Corona virus is 2 persons per 10 square km which is 0.05% of the population density of Haryana [7]. On the other hand, total population of Punjab is 2.8 crores which is approximately equal to Haryana. The rate of persons affected by Corona virus is 1 person per square km which is 0.02% of the population density of Punjab [4].

III. FRAMEWORK OF DATASET

The initial phase of dataset was raw dataset, to make it usable many techniques are applied on this dataset to refine it and make it more accurate. Initially, Class balancer is applied on the dataset, as the dataset is not balanced and after using the class balancer, another technique is applied on the dataset i.e. Ranker algorithm that gives the result to determine which attribute is most significant for the accuracy of algorithm. The selected attribute is

Confirmed case and therecovered cases which are ranked at first and second positionand show that the dataset is more accurate and after applyingthe ranker algorithm the dataset is tested with the help ofdifferent algorithm like logistic, Bayes Net, IBk and manymore.

On this dataset we perform various operation to filter thedata and find the expected output so, to find the expectedoutput we firstly use Class Balancer to balance the data usingthe unbalanced samples after that various algorithm are testedon the dataset based on the results of the algorithm whichinclude True positive rate and accuracy two algorithms areselectedwhichisLogitBoostandIBKwhichgivesusthebestresultamongallotheralgorithms.Thi datasetconsistof two states which is Haryana and Punjab and the reasonbehindselectingthesetwostatesisthepopulationofthesetwo states are nearly same and also these two states are closeto each other that makes the climatical condition very similarthat makes the comparison hurdle free and also gives the bestaccuracy.

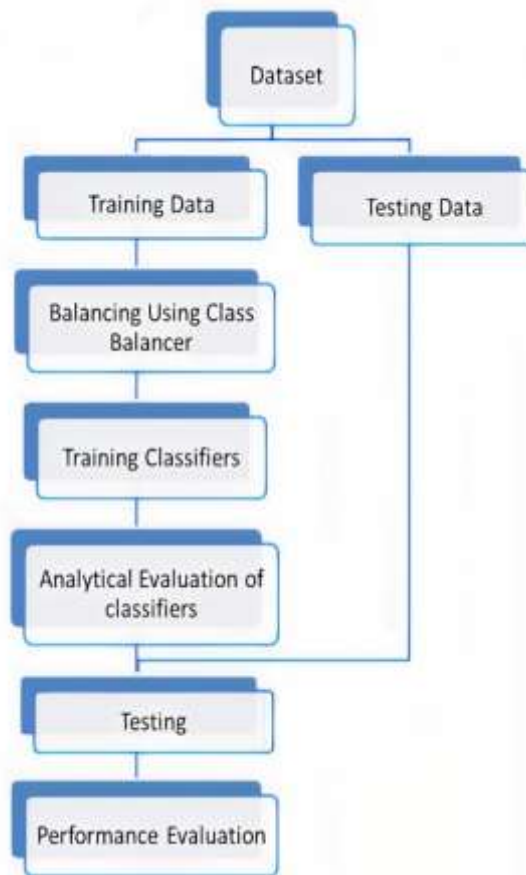


Fig.3.DataFlowDiagram

insupervisedaswellasunsupervisedlearning.Nowthedataset is load in Weka and certain balancing techniques areappliedonthedatasettoincreasetheoutputofclassifiersthedatasetisbalancedusingclassbalancer insupervisedlearning.Classbalancersynchronisetheunevenproportionsofclassesintestingsetwhi chthenincreasesthecorrectnessof the dataset. In more balances the maximum accuracy of allthe classes are explored. This techniques makes every instancepresent in the dataset balanced which helps the classificationalgorithms work more effectively and predict the accuracymoreaccurately.

IV. PROPOSEDWORK

Inthispartthedefinitedepictionofdifferentproceduresand calculations are referenced. This part contains differentsubsections in which characterization procedures are momen-tarily depicted with related flowcharts and graphical portrayal.The information is addressed with the assistance of flowchartinfig.1.ThedatasetispartitionedintotwosubgroupsinAIforexamplepreparingdatasetan dtestingdataset.Byapplying different characterizing procedures and examiningtheoutcomes,theclassifierwhichhasbestoutcomesischosen.

A. ClassBalancing

Initiallythedatasetisunbalancedanduntrained.Therearemanymachinelearningtechniquetobal ancethedataset

ClassificationAlgorithms

The dataset is tested using various classification algorithmswhich include Decision stump, Random tree, Random forest,and Logistic, Logitboost, IBk, SMO, Bayes Net, Naive Bayes,LWL and Decision table. The Decision stump takes only oneinputbecausethismodelone-leveldecisiontree.Therootofthedecisiontreeisonlyoneinternalnodewhichthenconnects to the leaves which are the terminal nodes. Randomtreeisasubgroupofrandomforestwhichclassifiesinputfeaturevectorwiththetreeandgiv esoutputaccordingtothemajorityofvotes(Randomforestisthecollectionofrandomtreepredictors. Logisticisclassifiersinwhichprobabilityofthetwopossibleclassesarepredicted.

$$y = b_0 + b_1 * x \quad (1)$$

$$p = \frac{1}{1 + e^{-y}} \quad (2)$$

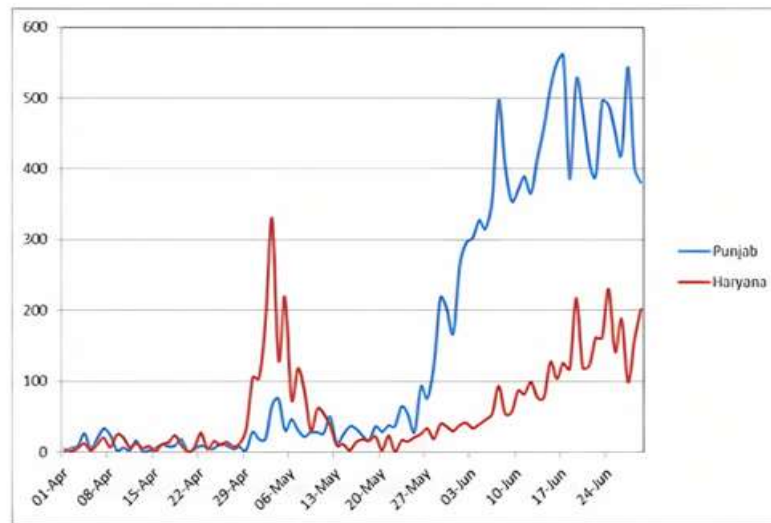


Fig.4. Confirmed Cases

Linear Regression

Logitboost is the boosting algorithm which decreases the logistic loss. IBK is a prediction algorithm which predicts the data using k nearest neighbour algorithm which measures the closest instance's distance k in the training dataset and prediction is made using these selected instances. SMO is used in the training of support vector machines which solves the quadratic problems. It is also known as sequential minimal optimization. Bayes Net is a graphical probabilistic model which computes the probability or conditional dependence with the help of directed graph. Naive Bayes is a probabilistic classifier which works on the bases of Bayes Theorem and represents the independent relation between the features. LWL (locally weighted learning) is a technique which creates the local model for each point of interest using neighbouring data. Decision table is an algorithm which performs a specification depending on the given conditions with the help of visual representation. These algorithms are implemented on the dataset and on the basis of various parameters, the best algorithm is selected.

V. EXPERIMENTAL INVESTIGATION

In this section, the dataset is analyzed with the help of graphical representation and techniques used on this dataset. The brief description of graphs of confirmed cases and re-covered cases for each state is discussed with the suitable explanation for the peak and down fall of plot lines showing graphs.

A. Analysis of dataset

The dataset is made from the daily analysis of COVID-19 cases in Haryana and Punjab. The dataset is based on five attributes that are date, State, confirmed, decreased and recovered. The graphical representation of the dataset is shown through the Fig. 3(a). According to the dataset, the ratio of confirmed cases in Haryana and Punjab was almost same from 26th April to 28th April and a rapid increment can be seen in Haryana from 29th April to 5th May. The peak of confirmed cases of COVID-19 was due to the migration of people from other states to Haryana and many people including 1 police officer was found positive which led to the rapid increase in the confirmed cases in Haryana. Then, there was the down fall in the confirmed cases in Haryana as the ratio of confirmed cases decrease with respect to recovered cases from 4th June to 29th June. The increase of confirmed cases in Punjab started from 26th May to 27th June. The peak in Punjab was from 7th June to 27th

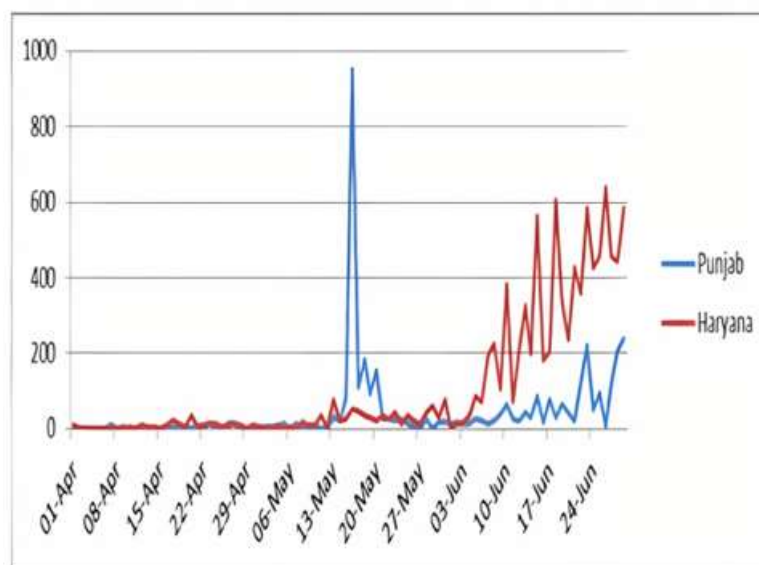


Fig. 5. Recovered Cases

June. It was due to the migrants that returned from the pilgrimage which led to the ratio of recovered cases was very as compared to confirmed cases which can be analyzed from Fig.3(b).

It has been observed from the fig.3(a) and fig.3(b) that the rate of increase of confirmed cases is 157 persons per day and rate of increase of recovered cases is 105 persons per day in Haryana. The rate of increment in confirmed cases is 60 persons per day and increment in recovered cases is 50 persons per day in Punjab.

B. Experimental Setting

This dataset is trained by using various algorithms that are used for finding the best suitable algorithm for this dataset to classify the data with the help of WEKA tool. The main objective of this paper is to find the accuracy obtained by the algorithms on this dataset and using the various filters. This dataset is made more concise and appropriate for making the prediction more efficient. This dataset is gone through both training and testing phase without splitting it into several parts.

The dataset is trained and tested with the help of the best tool for applying machine learning algorithms on the data which is weka tool. The dataset is loaded on the tool and supervised instance-based class balancer is applied to balance the dataset so that the further algorithm gives us the correct and accurate results. For analysing the algorithm, some data is considered on the basis of selected parameters and the best algorithms are selected for the dataset. The parameters used for selection of the best machine learning algorithms are MCC, ROC, TP-Rate, FP-Rate, Accuracy, Error and F-measure.

VI. RESULT AND DISCUSSION

This section describes the output, findings and the related discussions which are gathered after applying various machine learning algorithms on the dataset. This consists the overall result by applying the algorithm in which it is briefly discussed about the element associated with each and every parameter and the main outcome from these parameters and find out the correlation of the elements used for the classification with the dataset and what they represent is examined. The main factors on which the classification is done is also shown in this part. There are many algorithms applied on the dataset to find the best among all the

algorithm and various techniques is also applied on the dataset to pick the best algorithm associated with the data.

A. Analysis of dataset

The dataset is classified with help of certain classifiers such as Naive Bayes, Random tree, Logistic, Bayes Net, Logitboost, IBk, Random forest, SMO, Decision stump, LWL and Decision table. The performance of the classifiers is closely examined through various parameters which include TP-Rate, MCC, AUC, FP-Rate, F-measure, Accuracy and Error. The graphical representation of the analysis of the parameters is shown through Fig.4 which has further subdivisions to represent every parameter. The accuracy is represented through Fig.4 (a) which shows that accuracy of IBk is greatest among other classifiers which is constructed using the data in the Table.1. The slope of FP-Rates shows that the fraction of people which are confirmed for the test of COVID-19. The graph is shown through Fig.4 (b). The Fig.4(c) represents the TP-Rate and the slope represents the fraction of people which include the recovered cases from COVID-19 and deaths of the people. The graphical representation of ROC curve is shown through Fig.4 (d). The ROC also known as AUC which represents the area under curve which is the curve between the True positive rates i.e. TP-Rate and False positive rate i.e. FP-Rate. The value of ROC is highest for IBk which is reported as 1 among other classifying algorithms. The IBK has the maximum accuracy which is 100.00% and minimum errors which is 0.00%

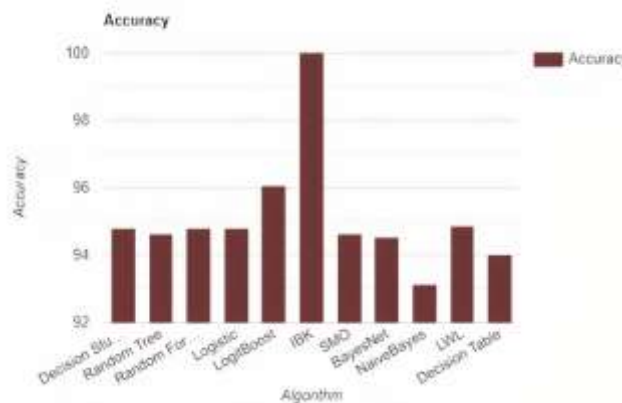


Fig.6. Accuracy

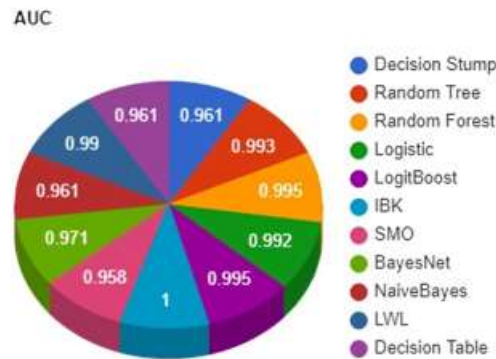


Fig.7. AUC (AreaUnderCurve)

which makes this classifying algorithm best among all other classifying algorithms for this dataset. The graphs shown above give the overview about the comparison done between the algorithm mentioned are to pelevan algorithm from the many algorithm applied on the dataset. It has been noted that the comparison between the algorithm is very precise as the algorithms having very minutedifferences and again the dataset is refined and again all these selected algorithms are applied on the dataset to find the best among these algorithms.

B. Analytical estimation of applied algorithm

In this section, the analysis of the algorithms is presenting comprises of parameters; from these parameters the best algorithm is selected. This section consists of different matrices which show the comparison of all the algorithms applied on the dataset.

The different matrices consist of the parameters which are Error, Accuracy, MCC, F-Measure, FP-Rate, TP-Rate, AUC on the basis of these parameters the algorithms are

TABLE I PERFORMANCE EVALUATION OF DIFFERENT ALGORITHMS BASED ON PARAMETERS

Classifier	DecisionStu mp	RandomTre e	RandomFore st	Logistic	LogitBoos t	IBK	SMO	BayesNet	NaiveBaye s	LWL	DecisionT able
Accuracy(%)	94.80	94.63	94.80	94.80	96.06	100	94.63	94.55	93.16	94.88	94.80
Error(%)	5.20	5.37	5.20	5.20	3.94	0.00	5.37	5.45	6.84	5.12	5.20
TPRate	0.948	0.946	0.948	0.948	0.961	1.00	0.946	0.945	0.932	0.949	0.948
FPRate	0.026	0.112	0.026	0.064	0.061	0.00	0.03	0.032	0.062	0.025	0.026
F-Measure	0.949	0.945	0.949	0.948	0.96	1.00	0.947	0.946	0.932	0.95	0.949
MCC	0.891	0.879	0.891	0.882	0.909	1.00	0.887	0.885	0.850	0.893	0.891
AUC	0.961	0.993	0.995	0.992	0.995	1.00	0.958	0.971	0.961	0.99	0.961

TABLE II
PERFORMANCE COMPARISON AMONG THE TOP FIVE ALGORITHMS

Classifier	Accura(%) cy	Error(%)	TPRate	FPRate	F- Measure	MCC	AUC
BayesNet	94.55	5.45	0.945	0.032	0.946	0.885	0.971
SMO	94.63	5.37	0.946	0.03	0.947	0.887	0.958
Logistic	94.80	5.20	0.948	0.064	0.948	0.882	0.992
LogitBoo st	99.06	3.94	0.961	0.061	0.960	0.909	0.995
IBK	100.0	0.00	1.000	0.000	1.000	1.000	1.000

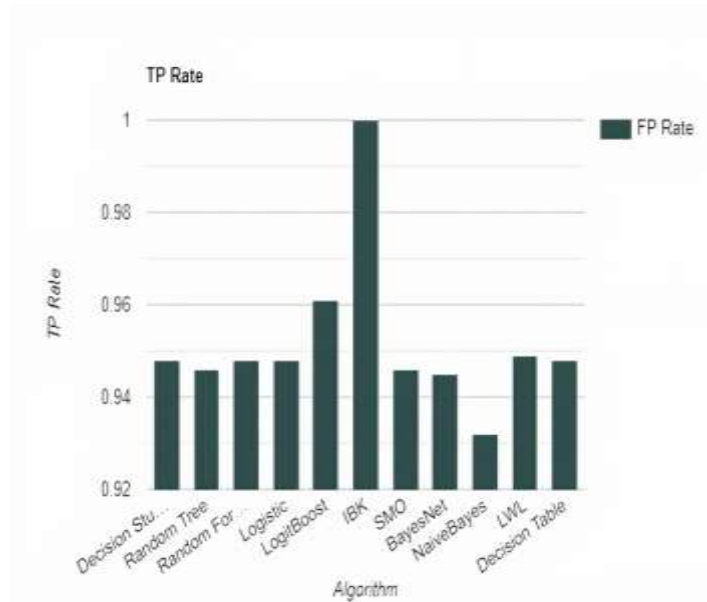


Fig.8.TP Rate

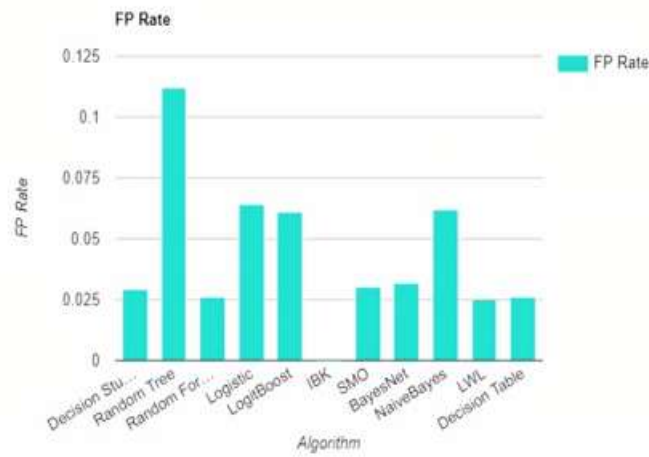


Fig.9.FP Rate

Research paper

© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

compared which includes Decision Stump, Random Tree, Random Forest, Logistic, Logitboost, IBK, SMO, Bayes Net, Naive Bayes, LWL, and Decision Table. The Table I illustrates the values of all the applied algorithms which are reclassified on the parameters such as TP-Rate, FP-rate, Accuracy, Error, AUC (Area under Curve) and MCC. The least accuracy is determined for Naive Bayes which is 94.80%. The second least is Bayes Net which reaches to 94.55%. Three classifying algorithms are analyzed and found IBk which is 100.00%, Logitboost which is 96.06% and LWL which is 94.88%. The algorithm such as Decision Stump, Random Forest, Decision Table and Logistic gives the same values for accuracy

i.e. 94.80%. In Table II, the Algorithms which gain the top maximum values of all the parameters are illustrated. The algorithm which found best on the basis of these parameters is IBK, which attain 100% accuracy on this COVID-19 dataset. The optimized algorithm which is IBk also gives the highest values for TP-Rate, F-Measure, MCC and AUC which is

1.00 and least values of Error and FP-Rate which is 0.00. The accuracy by this algorithm is observed when the dataset is balanced and only contains the relevant attributes for the analysis of COVID-19 by this algorithm. Hence, IBk is found to be the best algorithm for the dataset.

VII. ELICITATION BY ANALOGY OF DATA

The graph shown above shows the analytical view of parameters used for the dataset. The Fig. 4(a) shows the accuracy which means correlation of distinct algorithm applied on the dataset. It depicts that the ratio of prediction by the algorithm or the comparative analysis of correct prediction with the total prediction. Accuracy and Error show the overview of the confusion matrix by taking in consideration of framework used in confusion matrix as these two elements are showing the ratio among those frameworks. From the above-mentioned figure, it can be concluded that the IBk algorithm used for the comparison of the elements which can be interpreted as the correct prediction for all the elements. The Fig. 4(b) shows the AUC of the dataset

that means the area under curve made between the true positive rate and the false positive rate which indicates the accumulative estimation of performance among all the feasible classification approach. The AUC of IBk represents that the perfection of classifier to resemble all the elements of confusion matrix. In Fig. 4(c) shows the plot of FP-Rate which is false positive rate means the instances which are incorrectly classified due to certain circumstances in the plot. It is clearly visible that the algorithm is able to classify them correctly. In Fig. 4(d) shows the diagram of TP-Rate which is true positive rate mean sensitivity indicates the ratio among number of true positive cases by the total number of cases which means all the cases are analysed correctly causes no error in the prediction by the algorithm.

VIII. CONCLUSION

The extremely unstable dataset of COVID-19 is balanced using techniques of machine learning shown through this paper. This dataset is balanced with the help of supervised instance-based class balancer. Further, the ranker algorithm is applied on this dataset to find the most important attributes of this dataset which is confirmed cases and then recovered cases. Then the accuracy on this dataset is measured by applying various algorithms and found that IBk is the best algorithm that suits with this data. Overall performance is shown as a result and noted down that IBk gives perfect Accuracy which tells the prediction ratio of correct prediction with the total no. of instances which shows that the algorithm is perfectly able to evaluate the elements of confusion matrix gives us 100% result, TP-Rate means the classification done by the algorithm in a correct class with respect to the all the elements, among all the algorithm is has been obtained that IBk algorithm is able to classify correctly all the features from the class which is 1 out of 1, FP-Rate represents the capability of the algorithm shows the percentage of incorrectly classified components of the dataset with the total components in the dataset which is 0 that means the algorithm IBk is able to classify correctly all the element from class, MCC means it is a correlation coefficient which is based on the all the grids of the confusion matrix it ranges from -1 to +1 there -1 denotes the poor algorithm for this dataset and +1 shows the algorithm for this dataset and F-measure is the measured by taking the harmonic mean of precision and recall shows the stability of Dataset with respect to the algorithm and the area under curve is a plot draw using TP-Rate and FP-

Rate which shows that the capability of algorithm who is able to classify the true positive and negative class correctly in which the value 1 represent that the algorithm is perfectly matched on this dataset for the classification. The results appeared after accomplishment of the techniques describes that IBk is giving the best result or performance in comparison to all other algorithms of machine learning.

IX. FUTURE WORK

Different machine learning and deep learning techniques can be carried out on this dataset to increase precision. Various Machine learning techniques like neural network, enabling, clustering and many more can be applied to do prediction and find better results in future. Further work can be carried out on the dataset using various predictions algorithms. Numerous distinct techniques which involve Time series, Time average, Hybrid algorithm and Perceptron can be executed on this dataset to find the prediction for the peak and due dates of novel corona virus from the two states which are Haryana and Punjab. The similar work can be applied for other states of India or other countries in this world to find the end of this COVID-19 pandemic from the universe.

REFERENCES

- [1] <https://covid19.who.int/region/searo/country/in>
- [2] Zhao, J. et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin. Infect. Dis. ciaa344, March 2016 <https://doi.org/10.1093/cid/ciaa344>.
- [3] Qin, C. et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. Clin. Infect. Dis. Clinical Infectious Diseases, Vol. 71, Issue 15, pp. 762–768, August 2017. <https://doi.org/10.1093/cid/ciaa248>.
- [4] Kuo L, Masters PS. The small envelope protein E is not essential for murine coronavirus replication. J Virol; Vol. 77(8): pp. 4597–608 (2003).
- [5] <https://www.covid19india.org/state/PB>

Research paper

© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

- [6] Zhao S, Lin Q, Ran J, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019: A data-driven analysis in the early phase of the outbreak. *Int J Infect Dis* 2019. S1201-9712(20)30053-9. 10.1016/j.ijid.2019.01.050M. *International Journal of Infectious Diseases*, Vol. 94, May 2019, pp.72-73.
- [7] Huang C, Wang Y, Li X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*; Vol.395:pp.497–506(2015).
- [8] <https://www.covid19india.org/state/HR>
- [9] Wu, J.T., Leung, K., Bushman, M. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 26, pp.506–510(2013). (<https://doi.org/10.1038/s41591-020-0822-7>).
- [10] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7095448>
- [11] Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*; Vol.367:pp.1260-1263(2019).
- [12] <https://www.covid19india.org>
- [13] Al-Baadani AM, Elzein FE, Alhemyadi SA, Khan OA, Albenmoussa AH, Idrees MM. Characteristics and outcome of viral pneumonia caused by influenza and Middle East respiratory syndrome-coronavirus infections: a 4-year experience from a tertiary care center. *Ann Thorac Med* 2019; Vol.14:pp.179-185.
- [14] Peiris JSM, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*; Vol.361:pp.1319–1325(2003).
- [15] Tao Y, et al. Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. *J. Virol*; Vol.91:pp.01953–16(2017).
- [16] World Health Organization best practices for the naming of new human infectious diseases (World Health Organization, 2015); <http-s://apps.who.int/iris/handle/10665/163636>.
- [17] <https://apps.who.int/iris/bitstream/handle/10665/331780/nCoVsitrep11Apr2015-eng.pdf>.
- [18] Wang M, Cao R, Zhang L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*; Vol.30:pp.269-271(2014).

Research paper

© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 1, 2019

- [19] Kujawski SA, Wong KK, Collins JP, et al. First 12 patients with coronavirus disease 2019 (COVID-19) in the United States. medRxiv, March 12, 2020. (<https://www.medrxiv.org/content/10.1101/2020.03.09.20032896v1>).
- [20] Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev*; Vol. 22: pp. 370–85 (2009).
- [21] Peiris, J. S. et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* Vol. 361, pp. 1767–1772 (2003).
- [22] Chu, K. H. et al. Acute renal impairment in coronavirus-associated severe acute respiratory syndrome. *Kidney Int.* Vol. 67, pp. 698–705 (2005).
- [23] Hamming, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* Vol. 203, pp. 631–637 (2004).
- [24] Imai, Y., Kuba, K. & Penninger, J. M. The discovery of angiotensin-converting enzyme 2 and its role in acute lung injury in mice. *Exp. Physiol.* Vol. 93, pp. 543–548 (2008).
- [25] Kuba, K., Imai, Y. & Penninger, J. M. Angiotensin-converting enzyme 2 in lung diseases. *Curr. Opin. Pharmacol.* Vol. 6, pp. 271–276 (2006).
- [26] DiGiovanni C, Conley J, Chiu D, Zaborski J. Factors influencing compliance with quarantine in Toronto during the 2003 SARS outbreak. *Biosecur Bioterror*; Vol. 2: pp. 265–272 (2004).
- [27] Chan KS, Lai ST, Chu CM, et al. Treatment of severe acute respiratory syndrome with lopinavir/ritonavir: a multicentre retrospective matched cohort study. *Hong Kong Med J.* Vol. 9: pp. 399–406 (2003).
- [28] Coleman CM, Sisk JM, Mingo RM, Nelson EA, White JM, Frieman MB. Abelson kinase inhibitors are potent inhibitors of severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus fusion. *J Virol.* Vol. 90: pp. 8924–8933 (2016).
- [29] Soo YO, Cheng Y, Wong R, et al. Retrospective comparison of convalescent plasma with continuing high-dose methylprednisolone treatment in SARS patients. *Clin Microbiol Infect.* Vol. 10: pp. 676–678 (2004).

- [30] Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Trans-mission dynamics and control of severe acute respiratory syndrome. *Science*; Vol.300:pp.1966–70.10.1126/science.1086616(2003).
- [31] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*; Vol.395:pp.689–97.10.1016/S0140-6736(20)30260-9(2017).
- [32] Donnelly, C. A. et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* Vol.361, pp.1761–1766(2003).
- [33] Guy, J.S., Breslin, J.J., Breuhaus, B., Vivrette, S. & Smith, L.G. Characterization of a coronavirus isolated from a diarrheic foal. *J. Clin. Microbiol.* Vol.38, pp.4523–4526(2000).
- [34] De Vries, A. A. F., M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot. The genome organization of the nidovirales: similarities and differences between arteri-, toro-, and coronaviruses. *Semin. Virol.* Vol.8: pp.33-47(1997).
- [35] Fischer, F., C. F. Stegen, C. A. Koetzner, and P. S. Masters. Analysis of a recombinant mouse hepatitis virus expressing a foreign gene reveals a novel aspect of coronavirus transcription. *J. Virol.* Vol.71:pp.5148-5160(1997).
- [36] Lau, E.H. et al. A comparative epidemiological analysis of SARS in Hong Kong, Beijing and Taiwan. *BMC Infect Dis.* Vol.10, article.5(2010)
- [37] Narayanan, K., and S. Makino. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J. Virol.* Vol.75:pp.9059-9067(2001)
- [38] Glowacki I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol*; Vol.85(9):pp.4122–34(2011).
- [39] Oshiro LS, Schieble JH, Lennette EH. Electron microscopic studies of coronavirus. *J Gen Virol*; Vol.12:pp.161-168(1971).

[40] LiW, ShiZ, YuM, RenW, SmithC, EpsteinJH, WangH, CramerG, HuZ, ZhangH, et al: Bats are natural reservoirs of SARS-like coronaviruses. *Science*, Vol.310: pp. 676-679. 10.1126/science.1118391(2005).

1. Punjab and Haryana are currently reporting higher COVID-19 cases compared to other states in India. In comparison to other states, they have a higher daily positivity rate and a rapidly increasing number of cases. The reason behind the higher numbers could be attributed to factors such as people disregarding COVID-19 protocols, a more infectious variant of the virus, and a high population density. These two states are currently more prone to COVID-19 cases, and it is crucial that the citizens take necessary precautions to prevent the spread of the virus. The state governments have implemented measures such as restrictions and night curfew, but it is essential for people to also play their part by wearing masks, practicing social distancing, and getting vaccinated.

2. The number of cases in India has been fluctuating since the beginning of the pandemic, and the same is true for the states of Punjab and Haryana. It is crucial to understand that the COVID-19 data is complex, and the values can be influenced by several factors such as the pace of vaccination, compliance with COVID-19 protocols, and the emergence of new variants. It is challenging to accurately predict the number of cases as the situation is constantly changing.

Additionally, the dataset for COVID-19 cases in India is not consistent for each state, and there can be disparities in the data collection and reporting processes. It is important to understand that the COVID-19 data is complex, and the values can be influenced by several factors.

In conclusion, the values for COVID-19 cases in Punjab and Haryana, as well as in other states in India, are very fluctuating. It is challenging to accurately predict the number of cases as the situation is constantly changing. The complexity of the data, disparities in the data collection and reporting processes, and the influence of various factors make it difficult to examine each state of India accurately.