



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: VII Month of publication: July 2020

DOI: <http://doi.org/10.22214/ijraset.2020.7071>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Classification and Prediction based Enhanced J48 and REPTree Algorithms to Predict Corona Virus Pandemic

Amritpreet Kaur¹, Er. Roop Lal Sharma²

¹M.Tech (CSE), ²Assistant Professor, Department of Computer Sc. & Engg, St. Soldier Institute of Engg. & Technology, Near NIT, Jalandhar (Punjab)-144011

Abstract: Infection with Covid-19 can prompt deadly entanglements. Shockingly, there is little data about how the infection spreads and how patients are influenced. Information mining is the investigation of enormous datasets to remove covered up and beforehand obscure examples and connections. In social insurance, information mining procedures have been broadly applied in various applications including: displaying wellbeing results and anticipating persistent results, assessment of treatment adequacy, clinic positioning, and disease control. In this research work, we have comprehensively compared different data classification techniques and their prediction accuracy for Covid-19 disease. We have compared EJ48, REPTree, and User Classifier using performance measures like Accuracy, Error Rate, Consumption Time using WEKA tool. We have also compared these classifiers on various accuracy measures like TP rate, FP rate, precision, recall and f-measure by implementing on WEKA. Experimental result shows that EJ48 classifier has better classification accuracy over others for Covid-19 disease dataset

Keywords: EJ48, REPTree, Weka, Covid

I. INTRODUCTION

In 2019, world saw the flare-up of an infection called Coronavirus (Covid-19). The epic infection has a place with the coronaviruses family which is answerable for making mellow moderate colds. Covid-19 is accused for causing extreme intense respiratory sickness that lead to death by and large. As indicated by Covid-19 manifestations include: hack, fever, nose clog, breath brevity, and some of the time loose bowels. The infection started spreading quickly in China in 2019. From that point forward, the Control and Command Center of china Ministry of Health in china began recording and detailing the cases. The service site gives every day insights on new affirmed Covid-19 cases, recuperations, and passing.

II. CORONAVIRUS

AI can help speed up the medication advancement process, give knowledge into which current antivirals may give benefits, gauge contamination rates, and help screen patients quicker. Furthermore, in spite of the fact that not as of now explored, we think there are a few other fitting application territories. That said there are numerous obstructions identified with absence of restricted preparing information, the capacity to incorporate complex structures into DL models, and, maybe in particular, access to the accessible information.

III. MACHINE LEARNING

Models to foresee Classification is a broadly utilized strategy in human services. Here, we fabricate a few order the dependability and recuperation of Covid-19 contamination. We apply upgraded EJ48 and REPTree choice tree calculation. Here, we quickly portray these calculations.

- 1) *REPTree*: REPTree classifier: is a probabilistic model dependent on characterization calculation. It expect class restrictive freedom, where the conditions between class properties are overlooked. Examination has indicated that REPTree calculation have similar execution to other characterization calculations, for example, choice trees and neural systems. Furthermore, they produce exceptionally precise models and can manage enormous dataset.
- 2) *Enhanced J48*: Upgraded J48 choice Tree Algorithm: is an execution by the WEKA venture group of the notable tree enlistment calculation J48. It follows a ravenous iterative methodology in building the choice tree. The calculation parcels the dataset dependent on the best instructive trait. At every cycle, the trait with most extreme addition proportion is chosen as the parting quality. Choice tree characterization models have numerous preferences. They are anything but difficult to decipher and are known to have tantamount precision to other characterization models.

IV. RELATED WORK

- 1) *Dhara B. Mehta et al in year 2019*: The noteworthy advances in biotechnology and wellbeing science have prompted produce huge measure of information. As the coronary illness causing the serious issue since it is extremely hard to distinguish infection dependent on side effects.
- 2) *Chitra Jalota et al in year 2019*: Advanced education foundations are regularly inquisitive to think about the achievement pace of the understudies all through their investigation. Hence, they have to utilize a few strategies like physical assessment, Statistical strategies and at present winning information digging procedures for the forecast of understudy's exhibition.
- 3) *Shanjida Khan Maliha et al in year 2019*: Malignant growth is a rising lethal sickness which causes the passing of 10% among all the illnesses in Bangladesh. There are more than 100 Types of malignant growth. Foreseeing Cancer assumes an imperative job for advancing information mining applications.

V. OBJECTIVES

- A. To predict COVID-19 disease. COVID-19 is one of the major causes of death in our country also. It is crucial to make people aware about the risk factor of COVID-19 diseases.
- B. To identify the COVID-19 Disease used different machine learning algorithms.
- C. To enhance the accuracy with proposed algorithm.

VI. DATASET DESCRIPTION AND PRE-PROCESSING

As referenced before, our dataset was gotten from the site <https://www.covid19india.org>. We utilized the information on Covid-19 diseases gave an account of the geographic circulation of COVID-19 instances of india

Attributes Details Of Covid-19 Dataset

Attributes	Description
Age Bracket	Age of Patient
Gender	Gender of Patient
Detected City	City name
Detected District	District name where patient positive found
Detected State	State Name where patient positive found
State code	Code of the state
Notes	Travel History of the patient and any other reason because of which he became positive
Nationality	Country Name
Type of Transmission	Where did the virus spread
Num Cases	Number of Cases
Current Status	Patient Hospitalized, Recovered or Deceased

VII. RESULT AND DISCUSSION

Table1 Major Accuracy Measures Values

Algorithms	TP RATE	FP RATE	PRECISION	RECALL	F-MEASURE
EJ48	0.921	0.139	0.939	0.921	0.93
REPTree	1	0.683	0.773	1	0.872

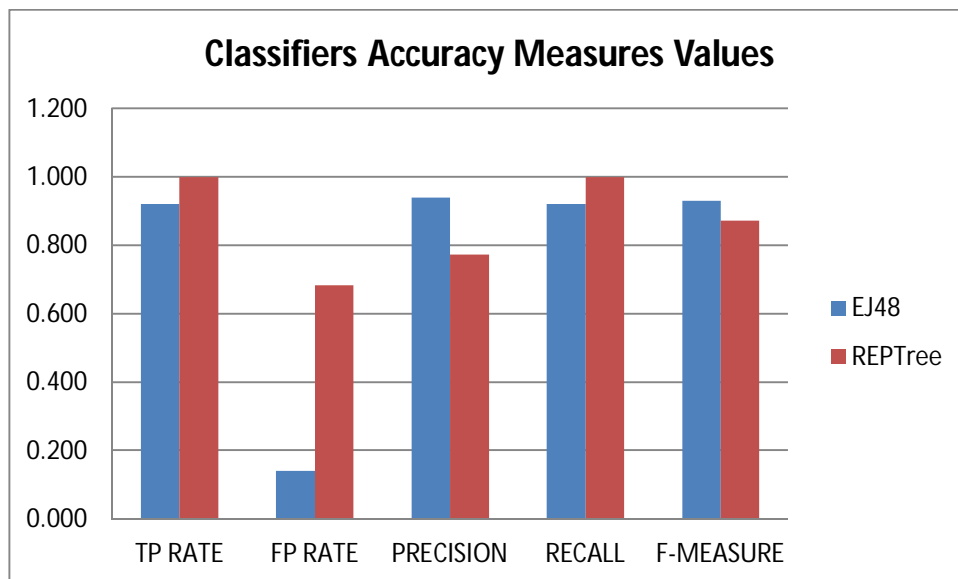


Fig 1 Classifier Accuracy Measures Values

Table 2 Classification Error Rate Values of Algorithms

Algorithm	Error Rate
EJ48	54.82
REPTree	83.94
User Classifier	99.99

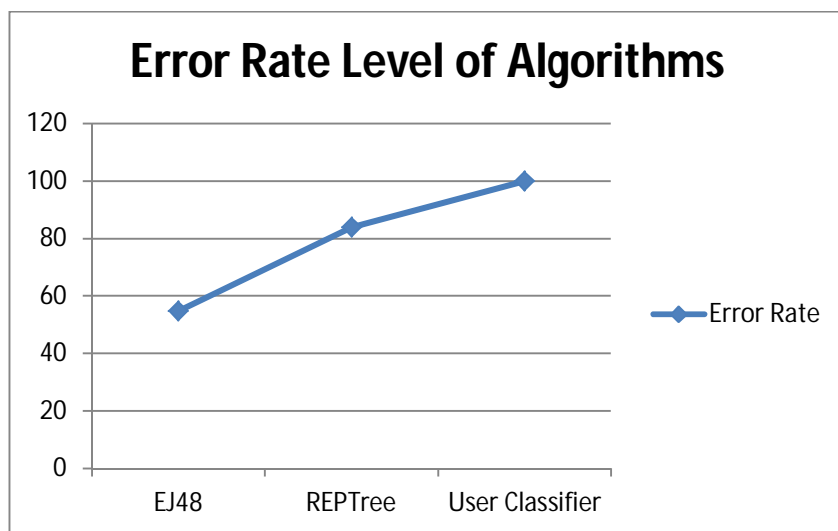


Fig 2 represents Error Rate values for the chosen classifiers. The K values show the better results for EJ48 over other classifiers

Table 3 Classification Accuracy Values of Algorithms

Algorithm	Accuracy
EJ48	89.810
REPTree	79.46
User Classifier	69.46

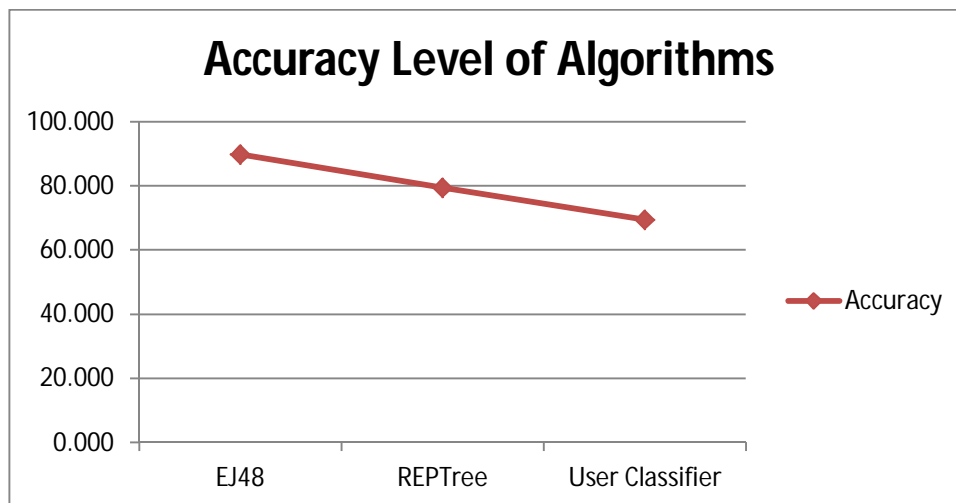


Fig 3 shows the accuracy levels for all classifiers. It shows that J48 has better accuracy whereas the REPTree shows poor accuracy levels.

Table 4 Accuracy and Error Rate values of EJ48, REPTree and User Classifier

Algorithms	Accuracy	Error Rate
EJ48	89.810	54.82
REPTree	79.46	83.94
User Classifier	69.46	99.99

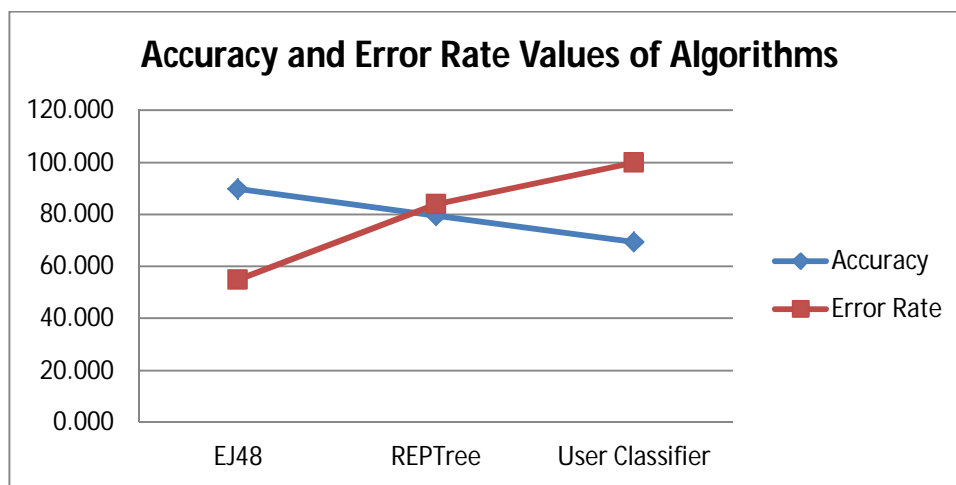


Fig 4 Analysis of Accuracy and Error Rate values of EJ48, REPTree and User Classifier

VIII. RESULTS AND DISCUSSION

Execution estimates factors appears, that classifier having K esteem more noteworthy than 0 indicating superior to possibility of understanding. On account of EJ48 classifier, in which estimation of k is coming 0.921, 0.139, 0.939, 0.921, 0.93 gets an opportunity of understanding. Table 5.2 unmistakably shows that EJ48 classifier has a more noteworthy estimation of $K=0.921$ when contrasted with different classifiers utilized in this examination on Covid-19 dataset. The Error Rate esteems appeared in table 5.3, speaks to that EJ48 accomplishes least MAE values 54.82 though REPTree and User Classifier displays most extreme MAE esteem 83.94 and 99.99. In that sense it very well may be inferred that expectation result created by EJ48 is closer to reality estimations of current status referenced in dataset. Exactness appeared in Table 5.4, EJ48 with the most noteworthy 89.81 in examination of different calculations considered for this trial., in any case, REPTree and User Classifier gives helpless outcomes for Accuracy measure. The TP Rate, Precision, Recall, F-Measure esteems for EJ48 is superior to different classifiers.

IX. CONCLUSION AND FUTURE WORK

In this exploration work, three classifiers EJ48, REPTree and User Classifier have been utilized for experimentation on WEKA information mining instrument. These classifiers have been contrasted all together with decide the arrangement precision dependent on execution estimates like accurately characterized occurrence, all out an ideal opportunity to assemble model, Error Rate and Accuracy esteems. In this exploration work, the exactness estimates like accuracy, review and FMeasures have additionally been considered so as to decide the precision level of every classifier all the more decisively. After experimentation it has been distinguished that EJ48 characterization calculation is demonstrating higher precision and expectation rate among all picked classifiers and existing writing deal with Covid-19 illness dataset. In future, better expectation models can be created with a more prominent precision by presenting classifier exactness execution changes and consideration of more test information for a specific infection.

REFERENCES

- [1] Dhara B. Mehta, Nirali C. Varnagar, "Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.
- [2] Chitra Jalota, Rashmi Agrawal, "Analysis of Educational Data Mining using Classification", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019..
- [3] Esther Gyimah, Delali Kwasi Dake, "Using Decision Tree Classification Algorithm to Predict Learner Typologies for Project-Based Learning", IEEE, 2019.
- [4] Shanjida Khan Maliha, comRomana Rahman Ema, comSimanta Kumar Ghosh, "Cancer Disease Prediction Using Naive Bayes,K- Nearest Neighbor and J48 algorithm", 10th ICCCNT 2019 July 6-8, 2019, IIT - Kanpur Kanpur, India.
- [5] Azwa Abdul Aziz, Andrew Starkey , Elissa Nadia Madi, " Predicting Supervise Machine Learning Performances for sentiment analysis using contextual-based approaches", IEEE, 2019.
- [6] Muhammad Hammad, Haris Anwar, "Sentiment Analysis of Sindhi Tweets Dataset using Supervised Machine Learning Techniques", IEEE, 2019.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)