

# COVID-19 World Vaccination Progress Using Machine Learning Classification Algorithms

Nasiba M. Abdulkareem  
Duhok Polytechnic University  
Duhok, Iraq  
[nasiba.mahdi@dpu.edu.krd](mailto:nasiba.mahdi@dpu.edu.krd)

Adnan Mohsin Abdulazeez  
Duhok Polytechnic University  
Duhok, Iraq  
[adnan.mohsin@dpu.edu.krd](mailto:adnan.mohsin@dpu.edu.krd)

Diyar Qader Zeebaree  
Duhok Polytechnic University  
Duhok, Iraq  
[dqszebaree@dpu.edu.krd](mailto:dqszebaree@dpu.edu.krd)

Dathar A. Hasan  
Duhok Polytechnic University  
Duhok, Iraq  
[dathar.hasan@dpu.edu.krd](mailto:dathar.hasan@dpu.edu.krd)

<https://doi.org/10.48161/qaj.v1n2a53>

**Abstract**— In December 2019, SARS-CoV-2 caused coronavirus disease (COVID-19) distributed to all countries, infecting thousands of people and causing deaths. COVID-19 induces mild sickness in most cases, although it may render some people very ill. Therefore, vaccines are in various phases of clinical progress, and some of them being approved for national use. The current state reveals that there is a critical need for a quick and timely solution to the Covid-19 vaccine development. Non-clinical methods such as data mining and machine learning techniques may help do this. This study will focus on the COVID-19 World Vaccination Progress using Machine learning classification Algorithms. The findings of the paper show which algorithm is better for a given dataset. Weka is used to run tests on real-world data, and four output classification algorithms (Decision Tree, K-nearest neighbors, Random Tree, and Naive Bayes) are used to analyze and draw conclusions. The comparison is based on accuracy and performance period, and it was discovered that the Decision Tree outperforms other algorithms in terms of time and accuracy.

**Keywords**—COVID-19 Vaccine, Machine learning, Classification algorithm, Dataset, WEKA

## I. INTRODUCTION

Nowadays, Machine Learning (ML) algorithms are used in wide computational areas due to its effective performance. [1,2]. In December 2019, a new series of coronavirus known as SARS-Cov-2 has been detected and caused a huge number of mortality due to irregular respiratory illnesses [1,2]. Recently, this virus I known as COVID-19, it starts spreading from Wuhan, Hubie Province, China, and then became more quick spreading in all countries around the world. The virus is transmittable among people and becomes a universal pandemic. COVID-19 has shown moderate symptoms and hits people of all ages [3,4]. Although, the evidence shows that two groups of people infect this severe virus more than the others, the people who are elderly, over 60 years old and the people who are chronically ill such as Cardiovascular, Diabetes, Chronic respiratory and cancer diseases. The level of death has been increasing continuously around the world caused by COVID-19 [5], therefore many countries have been forced to do quicken actions such as: wearing masks, social distance, lockdown and curfew. The quick development of a vaccine to prevent COVID-19 has

been a global imperative. Enough vaccines against COVID-19 are forecasted to be produced in 2021 to vaccinate at least 70% of the population of every country [6]. A quantity that could blunt the pandemic years sooner than predicted if it were distributed equally, this is according to new analysis. Non-clinical methods such as data mining and machine learning techniques may help to do this. Data mining is used to derive rules from huge amounts of data, while machine learning classification algorithms are data processing techniques that simplify the development of analytical models [7,8]. Nowadays, machine learning is very important in healthcare. Machine Learning is helping to streamline administrative manners in hospitals, map and deal with infectious diseases and personalize medical cures [9-11] It can also be used to give smart alternatives to analyze huge amounts of data. By developing fast and efficient algorithms and data-driven models for real-time processing of data, Machine Learning can produce accurate results and analysis. Therefore, we must increase clinician access to additional detail so that they can make informed choices on medical diagnoses and treatment strategies while still recognizing the potential consequences and costs for all [12]. The value of machine learning in healthcare is its own ability to produce huge datasets beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights which help physicians in planning and giving care, finally leading to better outcomes, lower costs of care and increases patient satisfaction [13,14]. For this issue, we collected datasets on the success of the Covid-19 world vaccine and analyzed the information using a machine learning and data mining algorithm this is a way to derive useful information from data and an appropriate algorithm for creating a reliable predictive model from it. This research focuses on classifying machine learning algorithms and identifying the most effective algorithm with the greatest consistency and accuracy [15]. To achieve this goal, we will use a COVID-19 World Vaccination Progress dataset and the Weka tool. The following is how the rest of the job is organized: The relevant analysis is presented in section II, the approach is presented in section III, the findings of the work are discussed in section IV, and the interpretation is presented Conclusion.

## II. LITERATUR REVIEW

The healthcare system is one of the most important parts of developed countries. Nowadays the fear of COVID-19 has infected the whole world that It is associated with the rapid growth of essential data and the need to analyze the relationships and hierarchy between data leads to the need for machine learning and Data mining is in the health system that is very effective in prevention, diagnosis and treatment. COVID-19 Vaccination Progress Using Machine Learning Algorithms is the subject of these studies [16]. This paper analyzed recent studies in this area, identified and determining the most efficient algorithm with the highest accuracy and precision.

Mellado et al. [17] performed an analysis in which a Deep Neural Network (DNN) was learned to distinguish between two types of datasets: serious illness (ICU, High Care, and Mortality) and less severe illness (ICU, Low Care, and Mortality) (subject discharged from the General Ward). A simulation sample containing 70% of the data was used, with a research sample containing 30% of the data being used to verify the findings and quantify possible over-training. The population data given by the Gauteng City-Region Observatory was used to apply the DNN weights. The horizontal axis depicts the proportion of Gauteng's adult population, while the vertical axis depicts the decrease in serious disease. According to the DNN model, through vaccinating around 20% of the adult population, the incidence of serious COVID-19 disease may be decreased by over 80%. The smart algorithm's performance can be increased with further refinements such as multi-class labelling and broader datasets.

Iweendi et al. [18] suggested a Fine-Tuned RF model also with the Adaboost algorithm as a boost. To predict the urgency of the circumstance and the probable result, death, the model uses the COVID-19 patient spatial, transport, health, and demographic details. The model has an F1 Score of 0.86 and a 94% accuracy rate. The data review reveals a strong association between patient gender and mortality, as well as the fact that the bulk of patients are between the ages of 20 and 70.

Vaccine Hesitancy (VH) has been proposed by Rescue et al. [19] as a major obstacle to the next COVID-19 mass immunization program. Using area-level metrics readily accessible to policymakers, utilizing machine-learning algorithms to forecast populations at high risk of VH. Based on results from infant immunization programs for seven non-mandatory vaccinations conducted in 6408 Italian municipalities in 2016, illustrate the strategy. The region under the Receiver Operating Characteristics (ROC) curve is used to compare a collection of machine learning models. The Random Forest algorithm is shown to be the most effective in forecasting areas with a high chance of VH, improving the volatile baseline standard by 24% in terms of precision. The proportion of waste recycled, and the job rate was found to be the strongest predictors of high VH among the area-level indicators. Which will help policymakers target pro-vaccine advocacy programs at the local level.

To forecast COVID-19 vaccine candidates, Wong et al. [20] used the vaxign and the recently built-in Machine Learning based vaxign-ML reverse vaccine ology methods. Vaxign-ML expected that the Spike proteins and nonstructural proteins (nsp3 and nsp8) will induce strong

defensive antigenicity. The nsp3 gene was discovered to be more conserved in Sars-Cov-2, SARS-CoV-2, and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) than in 15 coronaviruses that infect humans and other species. The protein was also shown to include promiscuous Mhc-I and Mhc-Ii T-Cell epitopes, as well as linear B-cell epitopes, which were expected to be clustered on the protein surface. COVID-19 vaccine production may be competitive and secure if our expected future vaccine targets are realized.

Ritonga et al. [21] used the Nave Bayes Classification algorithm on Twitter data with the keyword 'COVID-19' indexed by the keyword 'vaccine', in Indonesian tweets in the second and third weeks of January 2021. Over that time, the study reported 39% optimistic sentiment, 56% pessimistic sentiment, and 1% favourable sentiment. Since the media did not think the vaccination was effective at the time, negative opinions were created. Further research is anticipated to use a variety of algorithms to produce more accurate outcomes in opinion analysis.

Muhammad et al. [22] performed research on Data Mining models for forecasting the recovery of patients infected with 2019-nCoV. The models estimate the time it will take for Covid-19 affected patients to recover, as well as when they will be discharged from isolation centers. They also determine which patients will not recover and may die as a result of the Covid-19 pandemic. The models assist health care providers in determining the recovery and stability of freshly contaminated individuals. The models were created using the Korea Centers for Disease and Prevention (KCDC) dataset, which was used to extract instances of death and recovery reports from affected 2019-nCoV pandemic patients. Python programming and data mining techniques such as Support Vector Machine (SVM), Naive Bayes, Decision Tree, logistic regression, Random Forest (RF), and K-Nearest Neighbor (KNN) were used to create the models. The results indicate that the model developed using a Decision Tree data mining algorithm is more effective in predicting the probability of recovery of Covid-19 infected patients, with a 99.85% overall accuracy.

Brinatei et al. [23] developed two Deep Learning Classification models that use the hematochemical value from blood tests to distinguish between SARS-CoV-2 positive and negative patients, with accuracy and sensitivity varying from 82 to 86% and 92% to 95%, respectively. They've further developed a decision tree model to help doctors decide whether or not to interpret blood tests for suspected Covid-19 events (even off-line). Their findings showed that blood sample analysis and machine learning can be used to diagnose Covid-19 positive patients instead of Real-time Reverse Transcription Polymerase Chain Reaction (RRT-PCR), which is particularly useful in developing countries where RRT-PCR testing on respiratory tract specimens are scarce. They've created an open-source Web-based health reference and evaluation website.

## III. METHODS AND TECHNIQUES

The proposed Weka methods is basically consisting of four main steps: Dataset preparation, data preprocessing, Weka implementation and applying classification algorithms as shown in figure 1.

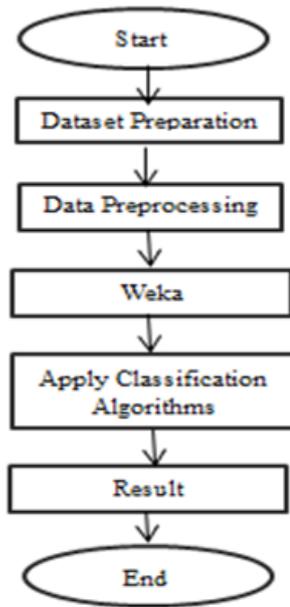


Fig. 1. Research methodology

#### A. Dataset Preparation

The ‘COVID-19 World Vaccination Progress Dataset’ is used from the Vaccine Adverse Event Reporting System (VAERS) Dataset (HHS.gov) in .CSV file format. The VAERS dataset is an open-source archive that contains the most detailed and relevant information on COVID-19 World Vaccination Progress and Events accessible everywhere in the world.

#### B. Data pre-Processing

There are 6745 instances in the dataset we used for this study, with more than 14 attributes. Prior to implementing the technique, any processing of the dataset is usually done. Data transmission has to be improved in order to meet data consistency requirements. There are two main methods of data processing: data collection, convert nominal data to numeric data

#### C. Classification

After the Data Pre-Processing steps, candidate classifiers are considered in this study: such as support vector machine (SVM), k-nearest neighbor (KNN), naïve Bayes (NB), decision tree, and random tree (RF) are used to perform classification.

- The K-Nearest Neighbor (KNN)

The KNN algorithm is a kind of supervised machine learning algorithm that is used to assign a class to a new data point for both discrete and continuous labels data problems. KNN keeps the training data to predict the label by computing the similarities between the input data and the training instance [24,25]. Calculating the gaps and locating the nearest neighbor are the two main steps in KNN. Any

entity casts a vote for their class, and the object with the most votes is known as a prediction. [26].

- Naïve Bayes

Naïve Bayes provides a method to predict the probability of different class based on various attributes. This algorithm is often used in text classification and when grappling with multi-class problems [27].

- Random Tree

Random tree is a classification and regression analysis algorithm that uses supervised learning. Tin Kam Ho used the random subspace method to create the algorithm. The random forest is made up of several decision trees that work together as an ensemble [28]. The Random Forest’s key steps are to take a random sample from the dataset and create a decision tree from each tree to get predication. Vote on the predication tree’s final predication and choose the one with the most votes [29,30].

- Decision Tree

For classification and regression analysis, the Decision Tree is a nonparametric supervised learning method. It worked for both categorical and continuous output variables. A decision tree has a two-stages classification process, learning and a prediction step. The model is trained using the given training data in the training process, and it is used to predict the outcome for the stated testing data in the prediction phase. [31,32].

#### D. Using Techniques with Weka Tool

WEKA 3.8.5 (Waikato Environment for Knowledge analysis) is an open-source Java platform that contains a series of machine learning algorithms that enable researchers to analyze their data for patterns, trends and the algorithms can be applied to our dataset. This study will show you how to use the Weka tool to estimation the efficiency of your machine learning model in a number of ways. To assess the idea, random train and test break to evaluate your model, and k-fold cross-validation to decide your model, use the training dataset. There are many model validation techniques available, and the Weka workbench includes four of them that this paper used Cross-Validation [33]:

- Training-Dataset: build the model using the entire training dataset, and then test it using the same data.
- Supplied-Test set: Another program is used to manually divide the dataset.
- Percentage-Split: A time you evaluate a model, divide the dataset in to training and testing partition at random.
- Cross Validation: divide your dataset into K-folds or partitions. After training the model in all partitions except one, which is normally held aside as a tests range, the average output of all k models is calculated [33,34].

After you have loaded a dataset, you can see these strategies in the Weka tool explorer. In addition, each test choice has a time limit. The aim of measurement options is to evaluate a model’s success on uncertain data. Predictive modeling aims to create a model that fits effectively in

situations where we don't know the future and have to contend with new and unknown data. In this case, we can use these types of robust statistical techniques to best predict the model's results, and Weka provides a Performance analysis when you analyze a model [35]. As a result, when testing a machine learning algorithm on a classification problem, you are faced with a vast amount of output data to digest, as classification is the most studied category of predictive modeling problem, and classification algorithms can be evaluated in a number of ways [36]. As a result, the first factor to consider in determining the success of a classification algorithm is Classification accuracy. The other one is reliability by class; actually, look at the true positive and false positive rates and make predictions about each class; the class analysis will help you determine if the problem is unequal or whether there are more than two classes [37].

#### IV. RESULT AND DISCUSSION

The previous section involves the study of each of the four techniques (Decision Tree, Random Forest, Naive Bayes-Nearest Neighbor (KNN)) introduced previously and testing each one of them using Weka Classification Tool on a set of COVID-19 Vaccine data. There are 14 attributes and 6745 instances in the whole data collection. In this paper, this study did not use SVM algorithm because the standard SVM is not suitable for the classification of large datasets, despite its strong theoretical basis and high classification accuracy. This is because the training difficulty of SVM is heavily dependent on the size of the dataset. For a large dataset, it is better to select the Decision Tree algorithm or Random tree algorithm if the dataset is consisting of more noisy data; the most suitable algorithm is the classification algorithm.

Table 1 shows that the Decision Tree classifier's percentage of correctly classified instances (6744) is higher than the other mentioned classifiers, and it is roughly comparable to the Random Tree classifier (6487). As a result, the Decision Tree classifier has a lower percentage of incorrectly classified instances (1) than the other classifiers, with the Nave Bayes classifier getting the largest percentage of incorrect classification instances (2041).

TABLE 1: CORRECTLY AND IN CORRECTLY- CLASSIFIED INSTANCES.

Classifiers	Correctly Classified Instances	In Correctly- Classified Instances	Kappa Statistic	Accuracy
Naïve Bayes	4704	2041	0.6856	69.7405%
KNN	5481	1264	0.7934	81.260%
Decision Tree	6744	1	0.9998	99.9852%
Random Tree	6487	258	0.9565	96.1749%

In table 2, Mean Absolute Error (MAE) or sometimes it is referred as L1 and Root Mean Square Error (RMSE) are used to measure the classification accuracy for each algorithm. MAE refers to the absolute error between predicted and observed values, while RMSE is a quadratic measure of the error between predicted and observed values. The main difference between two metrics is that RMSE provides a higher weight to large magnitude errors due to averaging the square of errors. As it is illustrated, Naïve Bayes the classifier has the largest value of MAE (0.008), and the Decision Tree has less, MAE value (0). Naïve Bayes

the classifier has a large percentage of relative absolute error (36.3292).

TABLE 2: MSE, RMSE CALCULATION

Classifiers	MAE	RMSE	RAE
Naïve Bayes classifier	0.008	0.076	36.3292
K-nearest neighbors	0.0049	0.0676	22.2405
Decision Tree	0	0.0014	0.0167
Random Tree	0.0013	0.0261	5.8538

While from Table 3 shows that the Decision Tree is more likely to be correct, at highest True Positive rate (1.000), it is the Naïve Bayes that has the lowest overall True Positive (0.697). The Precision values of the Nave Bayes classifier and the Random Tree are also strong (0.856 and 0.819 respectively). The Random Tree has the highest number of false positives (0.014).

TABLE 3: TP, FP, PRECISION AND RECALL CALCULATION

Classifiers	TP Rate	FP Rate	Precision	Recall
Naïve Bayes classifier	0.697	0.002	0.856	0.697
K-nearest neighbors	0.813	0.005	0.754	0.813
Decision Tree	1.000	0.000	0.000	1.000
Random Tree	0.962	0.014	0.819	0.962

Table 4 clearly demonstrates that the Decision Tree has the maximum MCC, and the K-nearest neighbors has a low MCC value. Decision Tree Covers highest percentage of ROC (1.000).

TABLE 4: F-MEASUREMENT, MCC AND ROC AREA CALCULATION

Classifiers	F- measure	MCC	Roc Area
Naïve Bayes classifier	0.689	0.705	0.958
K-nearest neighbors	0.541	0.702	0.905
Decision Tree	1.000	1.000	1.000
Random Tree	0.725	0.821	0.981

Fig.2 shows that, among the data mining techniques Decision tree, naïve Bayes, Random Tree, and K-nearest neighbor (KNN)), the model developed using The Decision Tree data mining algorithms are more efficient and have a 99.9% overall accuracy.

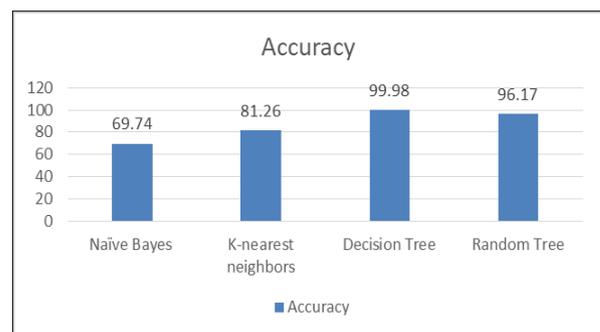


Fig. 2. Classification accuracy

#### CONCLUSION

COVID-19 pandemic impacts millions of lives all around the world as a major public health concern. To win the fight against the COVID-19 pandemic, we'll need an efficient vaccine that can be distributed equally and broadly. Therefore, we retrieved datasets on covid-19 world vaccine progress and have processed this information using four algorithms (Decision Tree, K-nearest neighbors, Random Tree, and Naïve Bayes) classification algorithms found in the WEKA data mining tool. Data mining algorithm

performance assessment is important since it helps users to pick the suitable algorithm for their classification/prediction activities. The final results after applied the dataset on Weka software and calculate the accuracy (The accuracy can be defined as the percentage of correctly classified instances divided by instances) for each classification algorithm shows that the best algorithm based on our data is the Decision Tree classifier with an accuracy of 99.9% and the total time taken to build the model is at 0.08 seconds.

## REFERENCES

- [1] Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review. *Machine Learning*, 62(03).
- [2] Jahwar, A. F., & Abdulazeez, A. M. (2020). Meta-Heuristic Algorithms For K-Means Clustering: A Review. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 12002-12020.
- [3] Adans-Dester, C. P., Bamberg, S., Bertacchi, F. P., Caulfield, B., Chappie, K., Demarchi, D., ... & Bonato, P. (2020). Can mHealth technology help mitigate the effects of the COVID-19 pandemic?. *IEEE Open Journal of Engineering in Medicine and Biology*, 1, 243-248.
- [4] Swayamsiddha, S., & Mohanty, C. (2020). Application of cognitive Internet of Medical Things for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*.
- [5] Shinan-Altman, S., Levkovich, I., & Tavori, G. (2020). Healthcare utilization among breast cancer patients during the COVID-19 outbreak. *Palliative & Supportive Care*, 18(4), 385-391.
- [6] Eisenstadt, M., Ramachandran, M., Chowdhury, N., Third, A., & Domingue, J. (2020). COVID-19 antibody test/vaccination certification: there's an app for that. *IEEE Open Journal of Engineering in Medicine and Biology*, 1, 148-155.
- [7] Sear, R. F., Velasquez, N., Leahy, R., Restrepo, N. J., El Oud, S., Gabriel, N., ... & Johnson, N. F. (2020). Quantifying covid-19 content in the online health opinion war using machine learning. *Ieee Access*, 8, 91886-91893.
- [8] Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(4), 140-147.
- [9] Abdulkareem, N. M., & Abdulazeez, A. M. (2021). Machine Learning Classification Based on Radom Forest Algorithm: A Review. *International Journal of Science and Business*, 5(2), 128-142.
- [10] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Machine learning and region growing for breast cancer segmentation. In *2019 International Conference on Advanced Science and Engineering (ICOASE)* (pp. 88-93). IEEE.
- [11] Omar, N., Abdulazeez, A. M., Sengur, A., & Al-Ali, S. G. S. (2020). Fused faster RCNNs for efficient detection of the license plates. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 974-982.
- [12] Muhammad, G., Alhamid, M. F., & Long, X. (2019). Computing and processing on the edge: Smart pathology detection for connected healthcare. *IEEE Network*, 33(6), 44-49.
- [13] Begli, M., Derakhshan, F., & Karimipour, H. (2019, August). A layered intrusion detection system for critical infrastructure using machine learning. In *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)* (pp. 120-124). IEEE.
- [14] Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10-19.
- [15] Nithya, B., & Ilango, V. (2017, June). Predictive analytics in health care using machine learning tools and techniques. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 492-499). IEEE.
- [16] Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access*, 8, 203097-203116.
- [17] Mellado, B., Wu, J., Kong, J. D., Bragazzi, N. L., Asgary, A., Kawonga, M., ... & Orbinski, J. (2021). Leveraging Artificial Intelligence and Big Data to optimize COVID-19 clinical public health and vaccination roll-out strategies in Africa. Available at SSRN 3787748.
- [18] Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., ... & Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health*, 8, 357.
- [19] Carrieri, V., Lagravinese, R., & Resce, G. (2021). Predicting vaccine hesitancy from area-level indicators: A machine learning approach. *medRxiv*.
- [20] Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in immunology*, 11, 1581. [5] Ritonga, M., Al Ihsan, M. A., Anjar, A., & Rambe, F. H. (2021, February). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012045). IOP Publishing.
- [21] Ritonga, M., Al Ihsan, M. A., Anjar, A., & Rambe, F. H. (2021, February). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1088, No. 1, p. 012045). IOP Publishing.
- [22] Muhammad, L. J., Islam, M. M., Usman, S. S., & Ayon, S. I. (2020). Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. *SN Computer Science*, 1(4), 1-7.
- [23] Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *Journal of medical systems*, 44(8), 1-12.
- [24] Tang, H., Xu, Y., Lin, A., Heidari, A. A., Wang, M., Chen, H., ... & Li, C. (2020). Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers. *IEEE Access*, 8, 35546-35562.
- [25] Gallego, A. J., Calvo-Zaragoza, J., & Rico-Juan, J. R. (2020). Insights Into Efficient k-Nearest Neighbor Classification With Convolutional Neural Codes. *IEEE Access*, 8, 99312-99326.
- [26] Salim, N. O., & Abdulazeez, A. M. (2021). Human Diseases Detection Based On Machine Learning Algorithms: A Review. *International Journal of Science and Business*, 5(2), 102-113.
- [27] Mahboob, T., Irfan, S., & Karamat, A. (2016, December). A machine learning approach for student assessment in E-learning using Quinlan's C4. 5, Naive Bayes and Random Forest algorithms. In *2016 19th International Multi-Topic Conference (INMIC)* (pp. 1-8). IEEE.
- [28] Choudhury, S., & Bhowal, A. (2015, May). Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection. In *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)* (pp. 89-95). IEEE.
- [29] Khorshid, S. F., & Abdulazeez, A. M. (2021). BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 18(4), 1927-1951.
- [30] Abdulqadir, H. R., & Abdulazeez, A. M. (2021). Reinforcement Learning and Modeling Techniques: A Review. *International Journal of Science and Business*, 5(3), 174-189.
- [31] Das, H., Naik, B., & Behera, H. S. (2020). An experimental analysis of machine learning classification algorithms on biomedical data. In *Proceedings of the 2nd International Conference on Communication, Devices and Computing* (pp. 525-539). Springer, Singapore.
- [32] Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- [33] Mohammed, D. Y., & Karabatak, M. (2018, March). Terrorist attacks in Turkey: An evaluate of terrorist acts that occurred in 2016. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)* (pp. 1-3). IEEE.
- [34] Kumar, N., & Khatri, S. (2017, February). Implementing WEKA for medical data classification and early disease prediction. In *2017 3rd international conference on computational intelligence & communication technology (CICT)* (pp. 1-6). IEEE.
- [35] Kiranmai, S. A., & Laxmi, A. J. (2018). Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy. *Protection and Control of Modern Power Systems*, 3(1), 1-12.

- [36] Shafiq, M., Yu, X., Laghari, A. A., Yao, L., Karn, N. K., & Abdessamia, F. (2016, October). Network traffic classification techniques and comparative analysis using machine learning algorithms. In 2016 2nd IEEE International Conference on Computer and Communications (ICCC) (pp. 2451-2455). IEEE.
- [37] Daku, H., Zavorsky, P., & Malik, Y. (2018, August). Behavioral-based classification and identification of ransomware variants using machine learning. In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 1560-1564).