2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA) | 979-8-3503-0821-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/SITA60746.2023.10373730

The Big Data and Machine Learning in Managing Global Health Crises : The Case of the Covid-19 Pandemic

Abdessamad Essaidi
INPT
Rabat, Morocco
essaidi@inpt.ac.ma

Mostafa Bellafkih
INPT
Rabat, Morocco
bellafkih@inpt.ac.ma

El Mehdi Kandoussi
INPT
Rabat, Morocco
kandoussi@inpt.ac.ma

***Abstract*—The Covid-19 pandemic has highlighted the critical importance of Big Data analytics and Machine Learning in effectively managing global health crises. In this article, we explore how, in this era of Big Data analytics applied to Covid-19, Machine Learning techniques have played a critical role in understanding, predicting and mitigating the spread of the virus. Thanks to Machine Learning, intelligent prediction models have been developed, allowing the transmission of the virus to be monitored and future developments to be anticipated. This study aims to explore the role of Machine Learning applications and algorithms by analyzing large Covid-19 datasets from the social media platform Twitter. Finally, the insights gained from these analyzes offer valuable insights for policy makers and health professionals in their efforts to address future pandemics.**

Keywords—Big Data, Machine Learning, Big Data analytics, Covid-19, Twitter

I. INTRODUCTION

The Covid-19 pandemic, which began in 2019 and quickly spread across the world, has had an unprecedented impact on public health and the global economy. Covid-19 is characterized by symptoms such as fever, cough, dyspnea, and muscle pain, often accompanied by bilateral pneumonia [1–3]. In response to the pandemic, the WHO has approved an anti-Covid-19 vaccine [4]. This global health crisis has highlighted the importance of having advanced analytical methods to better understand and manage such crises. Big Data and Machine Learning have emerged as essential tools to address this challenge. In this era of Big Data analysis specific to Covid-19, Machine Learning has been a key driver in the management of the pandemic. Using intelligent algorithms, Machine Learning has enabled scientists and epidemiologists to understand epidemiological trends and predict the spread of the virus [5]. By analyzing health, mobility, demographic and environmental data, these prediction models have provided essential information to guide policymakers in their actions.

Another crucial aspect of Machine Learning's role in the pandemic has been the ability to anticipate future developments of the virus. Using real-time analytics, Machine Learning models were able to track infection rates [6], identify clusters of cases, and assess the effectiveness of the mitigation measures put in place. This information was crucial in making quick and informed decisions to contain the spread of the virus [7]. In this study, we focus specifically on the use of Twitter as a data source to analyze the Covid-19 pandemic. Twitter has become an important platform for communicating and sharing information in real time, making it a valuable source of data for understanding public reactions, misinformation trends, and the spread of the virus. Using Big Data collection and analysis techniques, this study examines tweets related to Covid-19 to identify key discussion topics, audience emotions, and responses to public health measures. Sentiment and theme analyzes provide additional insight into public perception of the pandemic and the effectiveness of awareness campaigns [8]. The results of this study offer information for policy makers and health professionals in their fight against future pandemics. Understanding public reactions [9], concerns and behaviors can help tailor communication strategies and build public confidence in the actions taken [10]. Additionally, by using Machine Learning to predict the spread of the virus, policymakers can make more informed and proactive decisions to mitigate the impact of future health crises [11]. Intelligent prediction models can guide the planning of medical resources [12], the application of targeted mitigation measures and the effective deployment of vaccination campaigns.

The remaining part of our study is structured as follows. Section 2 provides an overview of the existing research conducted in this field. In Section 3, we delve into the data analysis and Machine Learning techniques employed, detailing the various methods utilized to conduct our study. In Section 4, we describe in detail the ETL process to extract the data from Twitter and prepare it for analysis. Moving on to Section 5, we present the results obtained, along with discussions and comparisons with relevant prior studies. Finally, our work concludes in Section 6 with a summary and suggestions for future research directions.

II. RELATED WORKS

In recent years, many works have been carried out in the field of diagnosis and detection of COVID-19 infections using

technologies based on artificial intelligence, as well as data analysis and Machine Learning algorithms. In this section, we will present some related works, highlighting the models and methods used by the authors, as well as the results obtained. We will also highlight the differences between this work and our research proposal.

Brinati et al. [13] conducted a feasibility study utilizing Machine Learning algorithms to detect COVID-19 infection from blood exams. The authors developed two Machine Learning classifiers based on hematochemical values obtained from 280 types of data [14]. These classifiers were designed to discriminate between patients who tested negative or positive for Covid-19.

In 2020, Soares et al. [15] proposed a novel artificial intelligence-based method for identifying Covid-19 cases using simple blood exams. They constructed an artificial intelligence classification framework based on a reduced dataset, aiming to achieve a classifier with high specificity and high negative predictive values, while maintaining reasonable sensitivity.

Banerjee et al. [16] proposed the integration of artificial intelligence and Machine Learning for predicting Covid-19 infection from blood samples. By employing full blood counts and employing shallow learning, random forest, and artificial neural network models, they achieved high accuracy in predicting Covid-19 patients, especially among populations in regular wards. Similarly, in 2020, Moraes Batista et al. [17] explored the application of Machine Learning algorithms to diagnose and predict Covid-19 in emergency patients. They used the same dataset as the authors of and trained their model with five Machine Learning algorithms, including neural networks, gradient boosting trees, random forests, support vector machines, and logistic regression.

In 2021, AlJame et al. [18] introduced an ensemble learning model for COVID-19 diagnosis. To handle null values in the dataset, they utilized a K-Nearest Neighbors algorithm and employed an isolation forest method to eliminate outlier data. The proposed model was trained and evaluated using publicly available data from [19]. Remarkably, the ensemble model achieved an outstanding overall accuracy of 99%. Alves et al. [20] also presented a Machine Learning model for diagnosing COVID-19 based on blood tests. The authors tested various Machine Learning models on a public dataset obtained from [19].

III. Data analysis and Machine Learning techniques

Big Data analytics is gaining popularity in health research and could provide predictive models for public health systems. New technologies such as Big Data play a crucial role in providing solutions to health problems. Nowadays, health data is growing tremendously every day, which requires effective, relevant and timely solutions to reduce the mortality rate. The use of Big Data analytics and Machine Learning techniques can play a vital role in tackling health challenges. Identifying the best performing prediction algorithm can help improve public health decision making and develop accurate predictive models to anticipate health care trends and needs. These results open new perspectives for the use of data science in health, contributing to the improvement of health systems and the reduction of mortality rates.

A. Data analysis

The entire process of knowledge discovery in databases can be made clearer by incorporating several key steps, including preprocessing, data mining and evaluation [21]. As illustrated in Fig. 1., these essential operators enable the construction of a comprehensive data analytics system, starting with data gathering and proceeding to information extraction and knowledge presentation to the user. Notably, research articles and technical reports often concentrate on data mining, with more emphasis compared to other operators. The subsequent sections of this study will delve into the major parts depicted in Fig. 1. The knowledge discovery process in databases is a multi-step process that involves extracting useful knowledge from large datasets, providing insights into the significance and implementation of each step. The process involves the following steps:

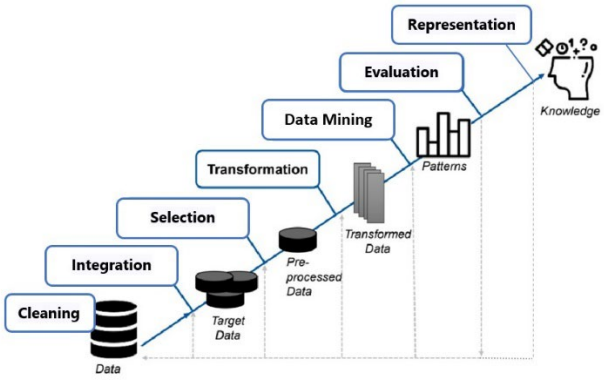

Fig. 1. The knowledge discovery process in databases

- Data cleaning

This step involves removing noise and inconsistencies from the data. organizations face the challenge of transforming this data into clean and usable forms to address complex problems. Therefore, there is a pressing need for data analysis focused on cleaning and refining raw data to make it suitable for further analysis [22]. This can include removing duplicate records, correcting errors, and dealing with missing data.

- Data integration

The second step of the knowledge discovery process in databases is data integration. In this step, data from multiple sources is combined into a single dataset. This can be a challenging task, as the data may come from different formats, structures, or systems. Therefore, it is important to ensure that the data is in a consistent format and that any conflicts between different datasets are resolved.

Data integration also involve dealing with conflicts between different datasets. For example, different datasets may use different codes or formats for the same variable. Therefore, it is important to resolve these conflicts to ensure that the resulting dataset is consistent and accurate.

Overall, data integration is a critical step in the knowledge discovery process in databases. By combining data from multiple sources into a single dataset, it is possible to gain a more comprehensive understanding of the data and extract useful knowledge from it. However, this step requires careful attention to detail and a thorough understanding of the data being integrated.

- Data selection

The third step of the knowledge discovery process in databases is data selection. In this step, a subset of the data is selected for analysis. This is an important step because it allows the analyst to focus on the most relevant data and avoid analyzing irrelevant or redundant data.

Data selection can involve filtering out irrelevant data or selecting specific variables of interest. For example, if the analysis is focused on a specific population or time period, it may be necessary to filter out data that does not meet these criteria. Similarly, if the analysis is focused on a specific variable, such as sales or customer satisfaction, it may be necessary to select only the data that is relevant to that variable.

One important consideration in data selection is bias. If the data selected for analysis is not representative of the entire dataset, the results of the analysis may be biased. Therefore, it is important to carefully consider the criteria used for data selection and to ensure that the selected data is representative of the entire dataset.

- Data transformation

This step involves transforming the data into a format that is suitable for analysis. This can involve converting categorical variables into numerical variables, normalizing the data, or creating new variables based on existing ones.

The fourth step of the knowledge discovery process in databases is data transformation. In this step, the selected data is transformed into a format that is suitable for analysis. This is an important step because the raw data may not be in a format that is directly usable for analysis.

Data transformation involve a variety of techniques, depending on the nature of the data and the analysis being performed. One common technique is converting categorical variables into numerical variables. Categorical variables are variables that take on a limited number of values, such as gender or product type. Converting these variables into numerical variables can make them easier to analyze using statistical techniques.

Data transformation also involve creating new variables based on existing ones. For example, if the analysis is focused on customer behavior, it may be useful to create a new variable that combines information about the customer's age, income, and education level. This new variable may provide more insight into the customer's behavior than any of the individual variables alone.

- Data mining

Data mining methods extend beyond specific approaches tailored to particular data problems. Instead, various technologies, including Machine Learning techniques, have been employed for many years to analyze data. Early data analysis primarily relied on statistical methods to gain insights into the situation at hand, such as understanding public opinions on social media platforms [26]. Alongside traditional data analysis, problem-specific data mining methods were utilized to extract meaningful patterns and knowledge from the collected data. Embracing this diverse set of approaches allows for efficient exploration, analysis, and comprehension of information within datasets, enabling informed decision-making and effective resolution of complex problems.

This step involves applying statistical and machine learning techniques to the data to extract patterns and relationships. This can involve clustering, classification, regression, or association rule mining.

- Pattern evaluation:

Evaluation is a fundamental part of the data mining process. Its main role is to measure the results obtained from the methods used. It also plays a key role in the choice and optimization of data mining algorithms, such as the selection of a genetic algorithm to solve clustering problems. The evaluation thus makes it possible to determine the efficiency, the precision and the relevance of the results produced by the algorithm, thus offering an essential means of evaluating the quality of the performance of the model or the approach employed. In summary, evaluation plays a vital role in the validation and continuous improvement of data mining methods, thus ensuring informed and reliable decision-making in various fields of application.

This step involves evaluating the patterns and relationships discovered in the previous step to determine their usefulness and relevance.

- Knowledge representation

The final step of the knowledge discovery process in databases is data presentation. In this step, the knowledge discovered through the previous steps is represented in a format that is understandable and usable by humans. This is an important step because the insights gained from the data are only valuable if they can be effectively communicated and acted upon.

Data presentation can involve a variety of techniques, depending on the nature of the data and the audience for the presentation. One common technique is creating visualizations, such as charts or graphs, that summarize the key findings of the analysis. Visualizations can be useful for highlighting patterns or relationships in the data that may not be immediately apparent from the raw data.

Another common technique in data presentation is creating reports or summaries that provide a more detailed overview of the analysis. Reports can be useful for providing context and background information, as well as for summarizing the key findings and recommendations.

Data presentation can also involve creating interactive dashboards or other forms of output that allow users to explore the data and gain insights on their own. This can be particularly

useful for stakeholders who may not have a background in data analysis but still need to make decisions based on the insights gained from the data.

- Knowledge utilization

This step involves using the knowledge discovered to make decisions or take actions. This can involve using the knowledge to improve business processes, develop new products, or make policy decisions.

B. Machine Learning techniques

In this study, our goal is to conduct an in-depth and comparative analysis of various Machine Learning algorithms with the aim of identifying the best prediction algorithm for Covid-19 data. To evaluate the performance of the algorithms, we use different measures such as accuracy, precision, sensitivity and specificity. We used a Covid-19 specific dataset to conduct this study. The algorithms were chosen for their ability to process complex data and provide reliable predictive results.

- Random Forest

Random Forest (RF) is a popular algorithm in Machine Learning, belonging to the category of supervised learning, and it is used to solve both classification and regression problems. It is based on the principle of set learning, which consists of combining several classifiers to improve the overall performance of the model and solve complex problems more efficiently [23].

The essence of the Random Forest is its ability to create multiple decision trees, each of which is built on a different subset of the original data. Then it makes individual predictions on each tree and, using a majority voting technique, it aggregates those predictions to determine the final output. This approach of combining the results of multiple decision trees helps to strengthen the predictive accuracy of the algorithm, the Fig. 2. [28] explains the working of the Random Forest algorithm.

Due to its generalizability and robustness, the Random Forest algorithm is widely adopted in various classification and regression applications. Its performance generally improves with the number of trees used in the forest, meaning that the more trees there are, the better the algorithm is at solving complex problems and providing accurate predictions.

The Random Forest algorithm has several hyperparameters that must be adjusted by the user. For example, it is important to determine the number of randomly drawn observations for each tree and whether to carry out these random draws with or without replacement. Similarly, the number of randomly drawn variables for each division, the division rule, and the minimum number of samples that a node must contain are also parameters to consider in order to optimize the performance of the model.

The Random Forest is a powerful Machine Learning algorithm, capable of dealing with classification and regression problems. Thanks to its ensemble-based nature and its majority voting approach, it offers high predictive accuracy and can be used successfully in a variety of fields, for example the analysis of Covid-19 data. However, it is essential to tune its hyperparameters well to obtain the best possible performance in a given context.

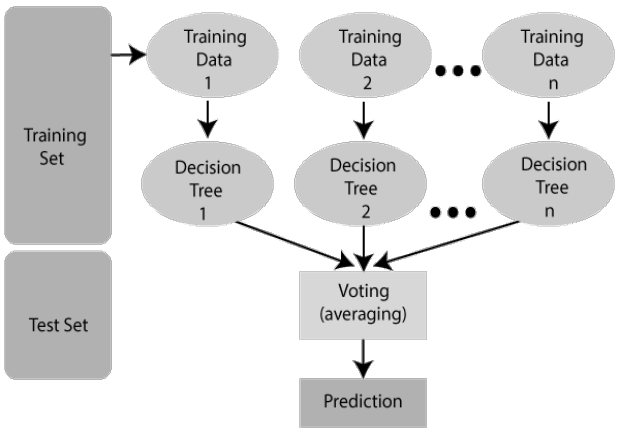

Fig. 2. The diagram of functioning of the Random Forest algorithm

- K-Nearest Neighbor

The K-nearest neighbor (K-NN) is one of the simplest algorithms in Machine Learning, using the supervised learning technique. Its principle is based on the similarity between the new data and the data already available, thus making it possible to classify the new data in the closest category among those existing [24]. The operation of the K-NN algorithm is to remember all available data and classify a new data point based on its similarity to existing data. Thus, when new data appears, it can easily be classified into a relevant category using the K-NN algorithm, as illustrated in Fig. 3 [29].

K-NN can be used for both regression and classification, but its use is mainly dedicated to classification problems. It is particularly effective when the data to be classified has specific groupings or structures. Besides its simplicity and ease of implementation, the K-NN algorithm however requires a careful choice of the parameter K, which represents the number of nearest neighbors to be considered during classification. A too small K may lead to a classification that is too sensitive to variations, while a too large K risks giving a too global classification, thus neglecting the local details of the data.

The K-NN algorithm offers a simple and efficient approach to Machine Learning, ideal for data classification. However, it is essential to carefully select the value of the parameter K to obtain optimal results and to take into account the particularities of the data to be processed.

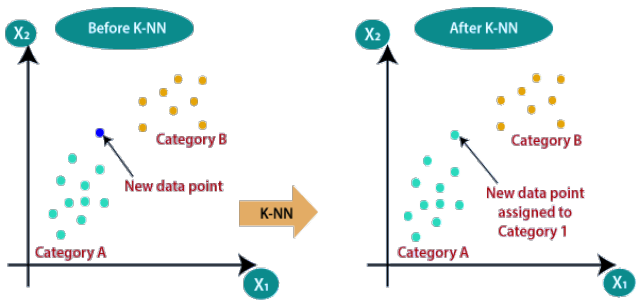

Fig. 3. K-Nearest Neighbor

- Support Vector Machine

Support Vector Machine (SVM) is one of the most popular Supervised Learning algorithms, utilized for both Classification and Regression problems, with a primary focus on Classification tasks in Machine Learning [25]. The main objective of the SVM algorithm is to create the best line or decision boundary, known as a hyperplane, that effectively separates the data points in n-dimensional space into distinct classes. This hyperplane enables easy classification of new data points in the correct categories in the future.

SVM achieves this by selecting the extreme points or vectors, referred to as support vectors, to construct the hyperplane. Consequently, the algorithm is named Support Vector Machine. The Fig. 4 [30]. below illustrates how two different categories are classified using the decision boundary or hyperplane.

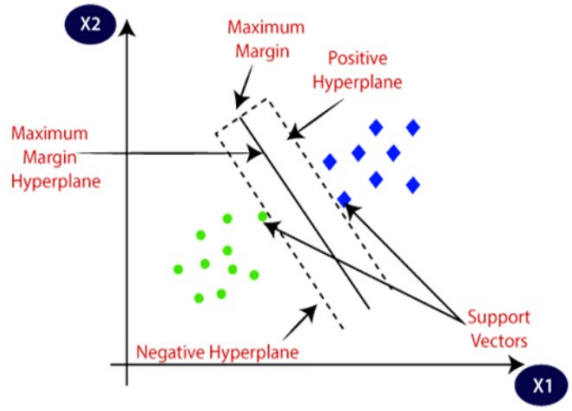

Fig. 4. Support vector machine

IV. ETL PROCESS TO EXTRACT THE DATA

In our approach to managing the information, data analysis plays a crucial role. Social media data, especially from Twitter, is rich in potential information but requires a rigorous process of extraction, transformation, and loading (ETL) to make it usable. The integration of the cleaned data into our analysis system is of paramount importance in our approach. Data collected from Twitter can often be messy and contain duplicate information, making it difficult to interpret. Therefore, we resort to data cleaning methods to eliminate disturbances and redundancies, in order to keep only the relevant elements for our analysis. In addition, we implement data transformation techniques to adapt them to the requirements of our Machine Learning algorithms.

These algorithms play a vital role in allowing us to dissect data from Twitter. They help us identify trends, detect patterns and spot anomalies, all of which are valuable factors in understanding the dynamics of the spread of the Covid-19 pandemic [27]. Finally, the success of our ETL process is instrumental in ensuring that the data collected from Twitter is adequately prepared for our analysis, thus consolidating the quality and relevance of our findings and conclusions relating to the management of the pandemic Covid-19.

As mentioned earlier, the data collected from Twitter is rich in COVID-19 information but requires a rigorous ETL process to make it usable. Our ETL process involves several steps, including data extraction, data cleaning, and data transformation.

In the data extraction step, we use Twitter's API to collect tweets related to Covid-19. We filter the tweets based on relevant keywords and hashtags to ensure that we only collect tweets that are relevant to our analysis. Once we have collected the tweets, we move on to the data cleaning step.

In the data cleaning step, we use various techniques to remove noise and redundancy from the data. This includes removing duplicate tweets, removing retweets, and removing tweets that contain irrelevant information. We also use natural language processing techniques to identify and remove tweets that contain offensive or inappropriate language.

Once the data has been cleaned, we move on to the data transformation step. In this step, we use various techniques to transform the data into a format that is compatible with our Machine Learning algorithms. This includes converting the text data into numerical data, creating features from the text data, and normalizing the data to ensure that all features are on the same scale.

Overall, our ETL process is crucial to ensuring the accuracy and reliability of our analysis. By using a rigorous ETL process, we are able to extract relevant information from Twitter and prepare it for analysis using advanced Machine Learning algorithms. This allows us to make accurate predictions about Covid-19 epidemiological trends and provide valuable insights to public health policymakers and healthcare professionals.

V. RESULTS AND DISCUSSION

During the course of this experiment, we employed Machine Learning strategies to forecast the sentiment of the collected data. Subsequently, we compared the outcomes of these models to determine the optimal classification model (RF, K-NN, and SVM) for prediction. Moreover, our analysis focused on hashtag keywords such as "COVID-19 pandemic," "coronavirus," "Health," "Health Departments," and "Health Services."

A. Dataset

Our dataset includes historical tweets regarding the COVID-19 pandemic from the Twitter platform. We acquired this data using Twitter's scraping API in conjunction with Python's Scweet library, which allowed us to extract a substantial volume of big data, as shown in Table 1.

This dataset exclusively includes tweets directly related to the COVID-19 pandemic, all from the Twitter platform. This data is of utmost importance in our research efforts, where we aim to improve our understanding of the shock of the pandemic and provide information for strategic decision-making in public health.

We employed the Twitter API to search for tweets associated with significant keywords, including "COVID-19 pandemic," "coronavirus," "Health," "Health Departments," and "Health Services." The gathered tweets were subsequently stored in CSV files, generating a substantial and well-organized dataset, representing a form of Big Data ready for future analysis.

TABLE I. DATASET

id	full_text	favorite _count	user_ id	user_ followers
123	Al Duhail Football Club players participate in an awareness campaign about Coronavirus (Covid-19).	23	31695 1125	288
124	Ensuring Continuity of Essential Health Services for GBV Survivors During the COVID-19 Crisis.	51	31695 111	529
125	Where was Schumer and Pelosi's hearings on #coronavirus in Jan?? ðŸ‘‡ #COVID19Pandemic	56	31695 11456	14360680
126	Wise words from our director of Sports Medicine, Kelsey Logan. Ways to Keep Kids Exercising During COVID-19 https://t.co/5IZQSPeE8U @ACTION4HF @CincyKidsHeart",,,NaN,,2,4,11925296131021 70112,413	71	33226 04940	14586
127	Countryâ€™s first drive-through testing facility for covid-19 set up in Clifton, Karachi by Sindh government and health departments.	23	22562 57457	79
128	NIDA Director outlines potential risks to people who smoke and use drugs during COVID-19 pandemic | National Institutes of Health (NIH) https://t.co/XAxKI9VDUe,,, NaN,,0,2, 785274118442586113,897	25	22562 57458	345
129	health minister, wife test positive for #coronavirus	36	22562 57459	212

B. Accuracy Tabel

In the classification task, the performance of these algorithms is assessed using metrics such as accuracy and precision, which provide valuable insights into their effectiveness in making correct predictions.

Accuracy is a fundamental metric that measures the proportion of correct predictions made by an algorithm over the total number of predictions. It gives an overall indication of how well the algorithm classifies the data, and on the other hand, is a metric that measures the proportion of true positive predictions (correctly predicted positive cases) over the total number of positive predictions made by the algorithm.

In addition to accuracy and precision, the time of execution is another critical factor to consider. The execution time reflects the efficiency and scalability of the algorithm, especially when dealing with large datasets. A fast-performing algorithm is advantageous, as it allows for real-time or near-real-time predictions, making it more suitable for time-sensitive applications. Hence, when evaluating classification algorithms, a comprehensive analysis of accuracy, precision, and time of execution is essential to select the most appropriate model for a particular task as shown in Table 2. A high-performing algorithm that achieves both accurate predictions and fast execution times is highly desirable for successful and efficient data classification.

TABLE II. ACCURACY TABLE

Classifier	Accuracy	Precision	Speeds
Random Forest	0.9834	0.9819	0.023
Support vector machine	0.7156	0.6267	0.067
K-Nearest Neighbor	0.7239	0.8134	0.032

C. Comparison Between Algorithms

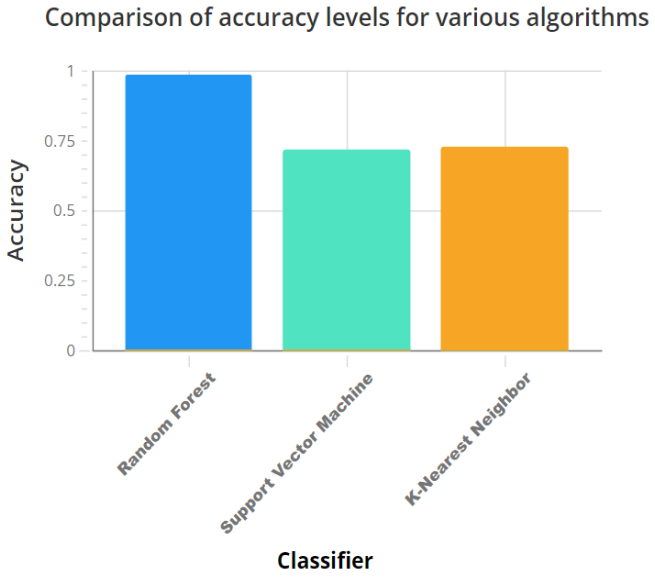

Fig. 5. Comparison between Algorithms

Following the conducted analyses and comparisons, the results clearly demonstrate the superior performance of the RF algorithm in comparison to the other algorithms investigated. The RF algorithm excelled in delivering more accurate predictions, enabling enhanced anticipation of Covid-19 epidemiological trends as shown in Figure 6.

These findings hold significant importance for public health policymakers and healthcare professionals, offering valuable insights to better comprehend and address the Covid-19 pandemic. By identifying the most effective prediction algorithm, this study opens avenues for refining pandemic prevention and management strategies and facilitating more effective decision-making. The RF algorithm's superior performance in predicting Covid-19 epidemiological trends can help policymakers and healthcare professionals to anticipate the spread of the virus and take proactive measures to prevent its transmission. This can include implementing targeted

interventions in high-risk areas, allocating resources to areas with the highest predicted case counts, and adjusting public health messaging to better address the needs of specific populations.

Moreover, the findings of this study have implications beyond the Covid-19 pandemic. The use of advanced Machine Learning algorithms to analyze epidemiological data can be applied to other infectious diseases, such as influenza, tuberculosis, and HIV/AIDS. By identifying the most effective prediction algorithms for these diseases, public health policymakers and healthcare professionals can make more informed decisions and take proactive measures to prevent their spread.

The superior performance of the RF algorithm can be attributed to its ability to handle complex and non-linear relationships between variables. This is particularly relevant in the context of the Covid-19 pandemic, where multiple factors can influence the spread of the virus, including population density, age, comorbidities, and social distancing measures. The RF algorithm's ability to handle missing data and outliers also contributes to its superior performance, as missing data is a common issue in epidemiological studies.

Overall, the findings of this study demonstrate the potential of Machine Learning to improve public health outcomes. By leveraging the power of advanced algorithms, policymakers and healthcare professionals can make more accurate predictions, identify high-risk areas, and take proactive measures to prevent the spread of infectious diseases. This study highlights the importance of continued investment in Machine Learning research and development to improve public health outcomes and mitigate the impact of future pandemics.

VI. Conclusion

This study aimed to investigate and analyze the outcomes achieved by employing various Machine Learning algorithms in the medical domain to predict COVID-19. We presented a prediction algorithm designed to foresee the to fight coronavirus at an early stage. The dataset comprised input parameters extracted from Twitter using hashtag keywords, and the models were trained and validated accordingly. The study involved constructing learning models, including RF, SVM, and K-NN, to perform the prediction diagnosis. The models' performance was assessed based on the accuracy of their predictions. The research results demonstrated that the RF model outperformed the K-NN and SVM models in terms of prediction accuracy.

Furthermore, comparisons were conducted based on factors like the execution time and feature set selection to enhance the effectiveness of this research. These findings shed light on the potential of Machine Learning algorithms in the medical field and may contribute to early detection and improved predictions for COVID-19 cases.

Machine Learning assumes a pivotal role in navigating the Covid-19 pandemic within the realm of Big Data analytics. Its capacity to expedite medical research, monitor the virus's dissemination, and facilitate informed decision-making presents a promising avenue for more effective handling of global health emergencies in the future. Nonetheless, deploying Machine Learning should be guided by ethical and data protection principles to ensure its positive influence on public health and society at large.

References

[1] J. S. M. Peiris, K. Y. Yuen, A. D. M. E. Osterhaus, and K. St¨ohr, “"e severe acute respiratory syndrome,” New England Journal of Medicine, vol. 349, no. 25, pp. 2431–2441, 2003.

[2] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, “Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges,” International Journal of Antimicrobial Agents, vol. 55, no. 3, Article ID 105924, 2020.

[3] W. J. Wiersinga, A. Rhodes, A. C. Cheng, S. J. Peacock, and H. C. Prescott, “Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19),” Journal of the American Medical Association, vol. 324, no. 8, pp. 782–793, 2020.

[4] Tchagna Kouanou, Aurelle, et al. "An overview of supervised machine learning methods and data analysis for COVID-19 detection." Journal of Healthcare Engineering 2021 (2021).

[5] Gaur, Loveleen, Gurinder Singh, and Vernika Agarwal. "Leveraging artificial intelligence tools to combat the COVID-19 crisis." Futuristic Trends in Network and Communication Technologies: Third International Conference, FTNCT 2020, Taganrog, Russia, October 14–16, 2020, Revised Selected Papers, Part I 3. Springer Singapore, 2021.

[6] Baldominos, Alejandro, et al. "A scalable machine learning online service for big data real-time analysis." 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD). IEEE, 2014.

[7] Hua, Jinling, and Rajib Shaw. "Corona virus (Covid-19)“infodemic” and emerging issues through a data lens: The case of china." International journal of environmental research and public health 17.7 (2020): 2309.

[8] Samuel, Jim, et al. "Covid-19 public sentiment insights and machine learning for tweets classification." Information 11.6 (2020): 314.

[9] Porat, Talya, et al. "Public health and risk communication during COVID-19—enhancing psychological needs to promote sustainable behavior change." Frontiers in public health (2020): 637.

[10] Wong, Catherine Mei Ling, and Olivia Jensen. "The paradox of trust: perceived risk and public compliance during the COVID-19 pandemic in Singapore." Journal of Risk Research 23.7-8 (2020): 1021-1030.

[11] Sheng, Jie, et al. "COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions." British Journal of Management 32.4 (2021): 1164-1183.

[12] Lv, Zhihan, et al. "Digital twins in unmanned aerial vehicles for rapid medical resource delivery in epidemics." IEEE Transactions on Intelligent Transportation Systems 23.12 (2021): 25106-25114.

[13] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, “Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study,” Journal of Medicine, vol. 44, no. 8, p. 135, 2020.

[14] https://zenodo.org/record/3886927#.YIluB5AzbMV Bood Routing DataSet San Raffaele Hospital (Milan, Italy).

[15] V. A. Soares, F. S. Fogliatto, M. H. P. Rigatto, M. J. Anzanello, M. Idiart, and M. Stevenson, “A novel specific artificial intelligence-based method to identify covid-19 cases using simple blood exams,” 2020.

[16] A. Banerjee, S. Ray, B. Vorselaars et al., “Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population,” International Immunopharmacology, vol. 86, Article ID 106705, 2020.

[17] A. F. de Moraes Batista, J. L. Miraglia, T. H. R. Donato, and A. D. P. Chiavegatto Filho, “Covid-19 diagnosis prediction in emergency care patients: a machine learning approach,” 2020.

[18] M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed, “Ensemble learning model for diagnosing COVID-19 from routine blood tests,” Informatics in Medicine Unlocked, vol. 21, Article ID 100449, 2020.

[19] E Data4u, “Diagnosis of COVID-19 and its clinical spectrum,” retrieves from https://www.kaggle.com/einsteindata4u/ covid19, 2020.

[20] M. A. Alves, G. Z. Castro, B. Oliveira et al., “Explaining machine learning based diagnosis of COVID-19 from routine blood tests with

decision trees and criteria graphs," Computers in Biology and Medicine, vol. 132, Article ID 104335, 2021.

[21] Molina-Coronado, Borja, et al. "Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process." IEEE Transactions on Network and Service Management 17.4 (2020): 2451-2479.

[22] Haoxiang, Wang, and S. Smys. "Big data analysis and perturbation using data mining algorithm." Journal of Soft Computing Paradigm (JSCP) 3.01 (2021): 19-28.

[23] Alzubi, Jafar, Anand Nayyar, and Akshi Kumar. "Machine learning from theory to algorithms: an overview." Journal of physics: conference series. Vol. 1142. IOP Publishing, 2018.

[24] Bansal, Malti, Apoorva Goyal, and Apoorva Choudhary. "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning." Decision Analytics Journal 3 (2022): 100071.

[25] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9.1 (2020): 381-386.

[26] Essaidi, Abdessamad, and Mostafa Bellafkih. "A New Big Data Architecture for Analysis: The Challenges on Social Media." International Journal of Advanced Computer Science and Applications 14.3 (2023).

[27] Essaidi, Abdessamad, Dounia Zaidouni, and Mostafa Bellafkih. "COVID-19: Analysis and Measurement of the Influence of the Tweets to Help the Public Health Sector to Fight Coronavirus." Advances on Smart and Soft Computing: Proceedings of ICACIn 2021. Springer Singapore, 2022.

[28] Deshpande, Nilkanth Mukund, Shilpa Gite, and Rajanikanth Aluvalu. "A review of microscopic analysis of blood cells for disease detection with AI perspective." PeerJ Computer Science 7 (2021): e460.

[29] Mohsen, Saeed, Ahmed Elkaseer, and Steffen G. Scholz. "Human activity recognition using K-nearest neighbor machine learning algorithm." Proceedings of the International Conference on Sustainable Design and Manufacturing. Singapore: Springer Singapore, 2021.

[30] Anwar, Muhammad Zohaib, Zeeshan Kaleem, and Abbas Jamalipour. "Machine learning inspired sound-based amateur drone detection for public safety applications." IEEE Transactions on Vehicular Technology 68.3 (2019): 2526-2534.