International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

Implementation of Hybrid Prediction Model: An Unsupervised and Supervised Learning Perspective

Mirza Ghazanfar Baig1, Sandeep Kumar Nayak[2]
[1]Department of Computer Application,
Integral University Lucknow, UP, India
mirza.gb21@gmail.com
[2]Department of Computer Application,
Integral University Lucknow, UP, India
nayak.kr.sandeep@gmail.com

Abstract—Using raw data to make inferences is the core of data science. This might be accomplished by closely examining the complex trends and patterns in the data. Machine Learning based forecasting methods have shown useful in predicting perioperative outcomes to improve the quality of future event planning decisions. In many application sectors where machine learning models were used, it has long been necessary to identify and prioritise the negative characteristics of a threat. Wish to provide precise predictions about a certain set of data, for example, use machine learning techniques in data science. Numerous prediction algorithms are now in use to address forecasting issues. Numerous epidemiological models are being employed internationally to forecast pandemic mortality rates and the number of affected people. Making the right decisions depends on the development of reliable prediction models.

Epidemiological models have had trouble making longer-term forecasts with a higher degree of accuracy due to a lack of significant data and ambiguity. This research suggests a hybrid machine learning approach to anticipate the pandemic in contrast to Susceptible-Infected-Resistant based models, and we demonstrate its potential using COVID-19 data from India. Order to improve the identification of epidemics early, The model can also be updated using data from sources like search engine searches. Results from two well-known machine learning methods were compared to those from the improved SIR and SEIQR models.

Keywords-Data Science, Machine Learning, SIR, SEIQR, Hybrid Prediction Model, Unsupervised and Supervised Learning Perspective.

I. INTRODUCTION

Machine Learning (ML) has changed throughout the years thanks to advancements in computer technology. The idea that computers may learn without being programmed to carry out certain tasks and pattern recognition theory gave rise to it; researchers interested in artificial intelligence sought to discover if computers could learn from data. Iterative machine learning is crucial because it allows models to freely change as they are presented with fresh data. They gain knowledge from earlier calculations to generate accurate, reproducible judgements and outcomes. It's an old science, but it's gaining new vigor right now. Data scientists' management of future discoveries will depend heavily on Artificial Intelligence (AI) driven by ML. To find patterns, mine data, and apply labels across various datasets, ML employs statistical models and algorithms. Data scientists will be able to make increasingly complex and precise predictions with the aid of these models, which continuously learn from the data.

We are going to be building a Hybrid Machine Learning (HML) model using SIR Model to discuss the spread of the COVID-19 as well as certain steps taken by governments of the country and estimate the economic and social losses of the epidemic [1]. It is all in theory and practical implementation we are claiming the accuracy of the entire model or what we are saying we are just going to discuss about the inferences we can draw through the numbers through the data that we have acquired from of COVID-19 India. The objective of this exercise is to not just "flatten the curve" but to optimize social distancing to minimize the outbreak time and keep healthcare services below full capacity.

In order to find the most effective preventative measures, several research have been done to forecast the virus's spread. For instance, a hybrid model based on mobility data was developed [2], and the parameters (Susceptible, Exposed, Infected, Vaccinated, Asymptomatic, Recovered, Hospitalized and Death) in the susceptible, infected, and recovered (SIR) instances model were estimated using a particle swarm optimization technique. The findings show that, when compared to analytical approaches, the latter method has a limited margin of error and is sufficiently accurate. Results from two well-known machine learning approaches were compared with those from the improved SIR and SEIQR models [3].

We also need to take into account another compartment termed "Exposed" for disorders like COVID-19. This group comprises of people who may be infected with the virus but may not yet exhibit any symptoms (due to travel or direct or indirect

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

contact with a patient who has tested positive). In other words, they are situated between the compartments that are vulnerable and diseased. These (asymptomatic) people can still spread the illness to receptive others while not displaying any symptoms. To better record ongoing disease management efforts, additional compartments might be added, such as "Vaccinated," "Quarantined," or "Hospitalized." The same assumptions about the speeds at which persons travel between these compartments are used in the modelling process as in the prior instance. The answer enables us to estimate the number of contagious individuals at any point in the future.

II. Research Methodology

Artificial Intelligence (AI), the subset of computing science called machine learning (ML), which focuses on analyzing and comprehending data structures and patterns to enable autonomous learning, reasoning, and decision-making. Simply said, machine learning is the capability for individuals to provide massive amounts of data to computer algorithms, which analyses, recommend, and make judgements entirely based on the provided data. The algorithm can use the information to enhance its judgement going forward if any adjustments are found. The three main components of machine learning are as follows:

- The computational algorithm at the heart of decision-making.
- The decision's determining factors and characteristics.
- Basic knowledge that provides the system with the ability to learn and for which the solution is known.

The investigation was conducted in a number of stages. Data were initially gathered through COVID-19 India's API (https://data.covid19india.org), which collects data from many organizations. The data were then pre-processed, analyzed, and checked for duplicates and missing values before being found linkages between various types of data, checked for design logic, and deployed. Figure1 offers a model serving of the study technique.

An illustration of the suggested approaches for reducing bias during the various phases of the development of ML-based applications systems: (1) Gather and prepare the necessary data. (2) Model development and model evaluation experiment. (3) Implement: Using a fresh dataset, a model may be used to forecast. Researchers are showing a lot of interest in machine learning, which is evidence that this is a very active area of research [4].

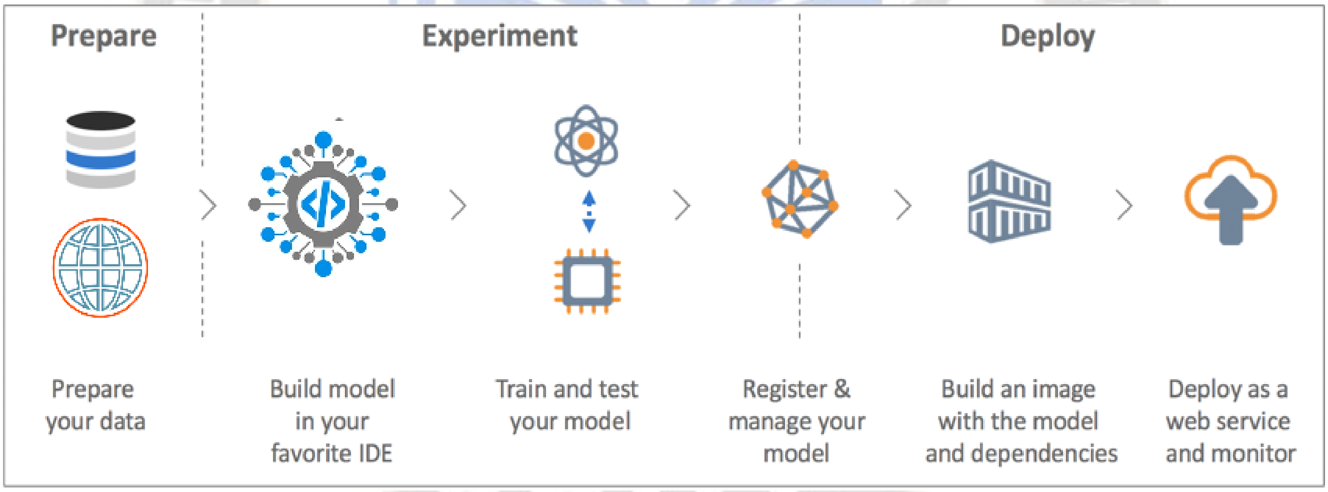


Figure 1. Model serving of research process.

III. Epidemiological Model

A. *Standard Model (SIR)*

The SIR (Susceptible, Infectious, Recovered) model and SIR modelling approaches are used in the current study to calculate the rate of disease transmission and anticipate the number of cases. SIR Classical Epidemic Models for Removing Infection from Basic States [5].

Figure 2. SIR Standard Epidemic Models

A disease's propagation among humans is depicted using the SIR model. These are the equations for the SIR model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

$$\frac{ds}{dt} = -\beta IS \qquad (1)$$

$$\frac{dI}{dt} = \beta IS - \gamma I \qquad (2)$$

$$\frac{dR}{dt} = \gamma I \qquad (3)$$

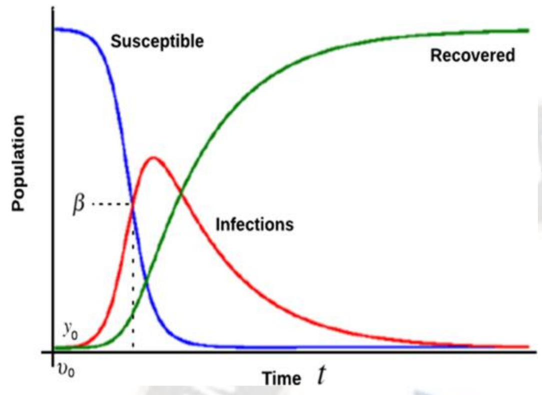


Figure 3. Infection density-based SIR model SIR Model.

The variables are the rate of transmission, the recovery rate, the infection position (y0), the initial infection velocity (υ0). The infection density (β) is adequately characterized by SIR on the presumption of a homogeneous and steady population (γ) [6]. This idea is based on the idea that everyone who is ill interacts with other people. Only susceptible persons may catch the disease, N = S + I + R states that the probability of encountering one is only equal to the percentage of sensitive persons in the population as a whole. Similar to this, a fraction of the sick people periodically eliminated from the system. The typical behavior of a SIR model is seen in Figure 3.

B. Improved Model (SEIQDR)

SEIQR model is an extension of the SIR, simulates interactions between individuals under various settings, The following are included: Susceptibility (S), Exposure (E), Infection (I), Quarantine (Q), Death (D), and Recovery (R). The SIR model's parameters S, I, and R, are all equal, whereas E displays the percentage of infected individuals who do not display any symptoms [7].

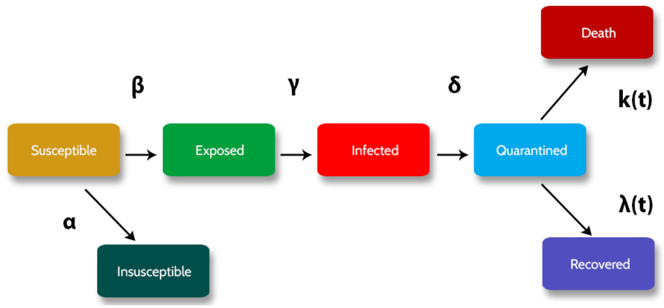


Figure 4. Improved SIR Model with Exposed and Quarantine

SEIQR Compartmental Model the SEIQR population percentage is provided via a compartmental model with four differential equations (Susceptible, Exposed, Infected, Quarantine and Recovered). Equations of the SEIQR model are:

$$\frac{dS(t)}{dt} = -\beta\frac{S(t)I(t)}{N} - \alpha S(t) \qquad (1)$$

$$\frac{dE(t)}{dt} = \beta\frac{S(t)I(t)}{N} - \gamma E(t) \qquad (2)$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t) \qquad (3)$$

$$\frac{dQ(t)}{dt} = \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t) \qquad (4)$$

$$\frac{dR(t)}{dt} = \lambda(t)Q(t) \qquad (5)$$

$$\frac{dD(t)}{dt} = \kappa(t)Q(t) \qquad (6)$$

where α represents the protection rate; β is the infection rate and illustrates the inverse of the average latent time; γ shows the rate of recovery (removal); δ represents the inverse of the average quarantine time; λ0, λ1 and λ2 are coefficients used in the time-dependent cure rate; κ0, κ1 and κ2 are coefficients used in the time-dependent mortality rate and S(t), P(t), E(t), I(t), Q(t), R(t), D(t) refer to the susceptible, insusceptible, exposed (infected but not yet infectious, in a latent period), infectious (with infectious capacity and not yet quarantined), quarantined (confirmed and infected), recovered, and closed cases.

IV. HYBRID MODEL

Currently, a variety of epidemiological models are being used to forecast individual infection rates and pandemic fatality rates. However, due to a lack of pertinent, trustworthy, and important data, very few models are able to produce time-bound precise judgements for the same. The SEIR model's Hybrid Model variant, which is utilized to characterize COVID-19 transmission throughout the nation, is introduced in this section. Figure 4 shows the flowchart of the SIR model. The hybridized SIR model includes supervised and unsupervised learning perspectives. SIR and Regression models are the two most often utilized machine learning algorithms for predicting disease infection globally.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

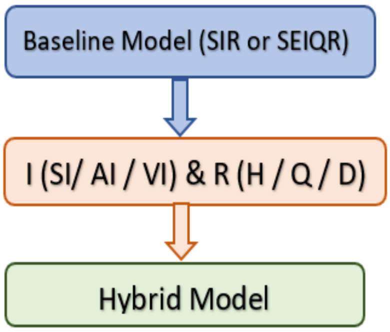


Figure. 5. Illustration of Task Procedure.

The COVID-19 case forecasts across the nation were done by using ML (Supervised and Unsupervised) techniques are mentioned below. Algorithms are taught by data in the subject of machine learning, and learning is accomplished through unsupervised and supervised learning techniques. This section discusses a few techniques that were used to forecast verified COVID-19 cases as well as fatalities [8].

Modified SIR model includes the following elements: Deaths (D), Exposed (E), Infected (I), Vaccinated (V), Asymptomatic (A), Quarantine (Q), Hospital (H), and Recovered (R) (D). S stands for "susceptible individuals," or those who are able to catch the illness; E stands for "exposed individuals," or those who have been exposed but are not yet contagious; I stands for "infective individuals," or those who can spread the illness; V stands for "individuals," or those who can spread the illness; A stands for "asymptomatic individuals," or those who can spread the illness while exhibiting no symptoms; Q stands for "quarantine," or those who can recover from the illness.

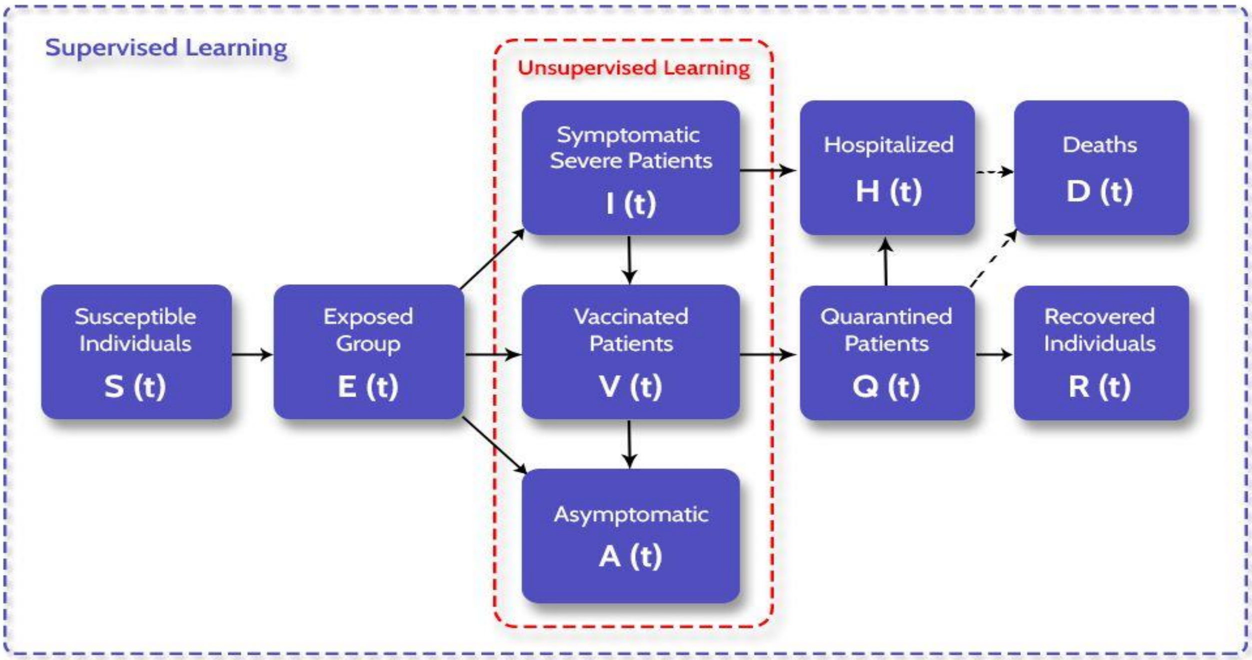


Figure 6. Illustration of Task Procedure

The approach appears to be suitable for data that is growing in a pandemic-related environment. By changing their model parameters, the SIR and SEIQR models' projected accuracy was greatly improved. Although there are several methods for predicting patterns in this epidemic, no single algorithm would be the most effective choice in every situation [9].

V. MACHINE LEARNING

Data science is all about drawing conclusions from unprocessed data. This may be achieved by studying the intricate patterns and trends in the data at a very detailed level. Machine learning is useful in this situation. Compartmental models and machine learning are frequently integrated to replace the fixed parameters of the former with time-varying parameters that may be learned from data. Machine learning, on the other hand, excels in extracting patterns from data. A lot of models would need to be built from scratch, which is unfortunate, especially when learning complex patterns like the spread of a pandemic.

There are six phases in the data science life cycle, and it begins with defining-

- Business Understanding: It is the procedure for discovering, analyzing, and outlining the needs connected to a certain company objective.
- Data Understanding: Prior to being placed in a data warehouse or another sort of storage system, it is this

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

process that has come to be known as data gathering, filtration, and cleansing [10].

- Data Preparation: The process of data exploration includes feature engineering, missing value imputation, outlier detection, and variable development in the data.
- Modelling: Making a detailed depiction of the connections between various sorts of information that will be kept in a database is known as modelling.
- Evaluation: To examine and assess each component of the provided data using analytical or logical reasoning.
- Deployment: In data science, the use of a new dataset to apply a model for prediction is referred to as "deployment."

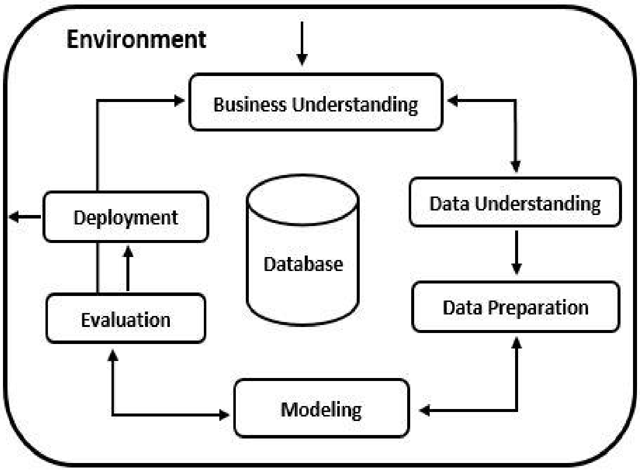


Figure 7. Cyclic Process of Machine Learning Implementation.

Making machines learn and behave like people without the need of explicit programming is the main goal of the field of machine learning research. Training data is used to train machine learning algorithms. They can precisely forecast the future and decide based on the past when fresh data is collected [11]. Machine learning divided into two main groups.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

Technique	Algorithm	Parameter	Sustainability/ Attribute	Data Limitations
Regression	Linear Regression	fit_intercept, normalize, copy_X, n_jobs	Coefficients, Rank, Singular, Intercept	Continuous
	Polynomial Regression	degree, interaction_only, include_bias, order	Powers, Input Features, Output Features	Continuous
	SVR (Support Vector Regression)	kernel, degree, gamma, coef0, tol, epsilon, shrinking, cache_size, verbose, max_iter	Support Vectors, Coefficients, Fitted Status, Intercept	Continuous
Classification	Logistic Regression	penalty, default, tolerance, fit_intercept, intercept_scaling, class_weight, random_state, solver, max_iterations, multi_class, verbose, warm_start, n_jobs, l1_ratio	Labels Classes, Coefficient, Intercept, Iterations,	Categorical
	KNN (K- Nearest Neighbors)	n_neighbors, weights, algorithm, leaf_size, power, metric, metric_params, n_jobs	Labels Classes, Effective Metric/ Metric Params, outputs_2d	Categorical
	Naive Bayes Classification	Prior Probabilities , var_smoothing (variances for calculation stability)	Labels Classe, Class Count/Prior, Epsilon, Sigma, Theta, Probability	Categorical
	SVM (Support Vector Machines)	kernel, degree, gamma, coefficients, shrinking, probability, tolerance , cache_size, class_weight, verbose, max_iterations, decision_function_shape, break_ties, random_state,	Labels Classe, Support Vectors, Coefficients, Probability, Intercept	Categorical
Regression & Classification	Decision Tree	criterion, splitter, max_depth, min_samples_split, min_samples_leaf, min_weight_fraction_leaf, max_features, random_state, max_leaf_nodes, min_impurity_decrease, min_impurity_split, presort, ccp_alpha	Feature Importances, Max Features, n_outputs, Tree	Continuous & Categorical
	Random Forest (Ensemble Learning Method)	n_estimators, criterion, max_depth, min_samples_split, min_samples_leaf, min_weight_fraction_leaf, max_features, max_leaf_nodes, min_impurity_decrease, min_impurity_split, bootstrap, oob_score, n_jobs, random_state, verbose, warm_start, ccp_alpha, max_samples	Base Estimator, Feature Importances, n_outputs, oob_score, oob_prediction	Continuous & Categorical
Clustering	K-Means Clustering	n_clusters, initialization, max_iterations , tolerance, precompute_distances, verbose, random_state, copy_x, n_jobs, algorithm	Cluster Centers, Labels , Inertia, Iterations , random_state	Continuous
	Hierarchical Clustering	n_clusters, affinity, memory, connectivity, compute_full_tree, linkage, distance_threshold	clusters, labels, leaves, connected components, children	Continuous
	Fuzzy C-Means Clustering	n_clusters, max_iter, memory, error, random_state	Cluster Centers, Labels , Inertia, Iterations , random_state	
	Hidden Markov Models	n_components, covariance_type, min_covar, startprob_prior, transmat_prior, means_prior, means_weight, covars_prior, covars_weight, algorithm, random_state, n_iter, tolerance, verbose, params, init_params	Features, Monitor, startprob, transmat, means, covars	Continuous
Dimensionality Reduction	SVD (Singular-Value Decomposition)	n_components, algorithm, n_iterations , random_state, tolerance	components, Iterations, Random State, Tolerance,	Continuous
	PCA (Principal Component Analysis)	n_components, copy=True, whiten, svd_solver, tolerance, iterated_power, random_state	components, explained_variance, singular_values, mean	Continuous
	Random Projection (Gaussian/Sparse)	n_components, eps, random_state, density	components, density	Continuous
Association Rule Mining	Apriori	min_support, min_confidence, min_lift, max_length	Support, Confidence, Lift, Length of Relation	Categorical
	FP-Growth	patterns, confidence_threshold, transactions, support_threshold,	Patterns, Confidence, Transactions, Support	Categorical

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

A. Supervised Learning

It has a model that can forecast outcomes using labelled data. When a dataset has been tagged, the intended answer is already known. Using historical data, the machine can make precise predictions. Classification and regression are the two additional categories under which supervised learning may be subdivided.

Classification is used when an output variable contains two or more classes and is categorical. For instance, yes or no, true or untrue, male or female, etc. Regression is used when the output variable has a real or continuous value. In this case, a change in one variable affects the other since there is a connection between the two or more variables. Consider salary that is determined by prior employment experience or weight that is determined by height, etc.

B. Unsupervised Learning

Unsupervised learning uses unlabeled data to allow the computer to learn on its own. The system replies by searching for patterns in the unlabeled data [12]. Based on patterns, similarities, and other factors, the computer divides the provided collection of patterns into groups. Clustering and association are the two subtypes of unsupervised learning.

Clustering is the process of putting things into groups that are different from one another yet comparable to one another. For instance, identifying the consumers who bought related items. Through association, a form of rule-based machine learning, one may determine the possibility that items in a collection will show up together. Identifying, for instance, the items that were bought concurrently.

The following list summarizes the key distinctions between supervised and unsupervised learning:

- Machine learning models are developed for supervised learning tasks utilizing labelled training data. In contrast, training data for an unsupervised machine learning job does not have labels or categories attached to it.
- Unsupervised learning enables you to find hidden patterns inside a dataset without the need for human input, whereas supervised learning models assist predict outcomes for future data sets.

In machine learning, models must be trained to emulate human behavior in order to simulate complex psychological aspects or functions including inference, reasoning, and decision-making. It is a subfield of AI that employs methods with a statically-based foundation to aid machines in improving over time. In order to model the scenario in any type of mathematical framework, it may obtain knowledge using statistical approaches. For navigation, recognition, prediction, or description, machine learning is necessary [13]. Based on how closely a text or message resembles emails with the same tag, spam filters often determine if it is spam. This is an everyday use of machine learning. Before utilizing machine learning, you must, however, properly comprehend the demands of the organization.

Let's say the agent is in the environment and has a specified mission to do based on the following:

Agent=A	Environment=E
Task= T	Output=O
Error= e	Performance=P
Max Error=e+	High Performance Output=Po+
Min Error=e-	Low Performance Output=Po-

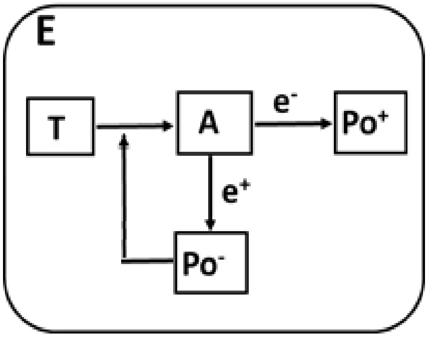


Figure 8. Machines Learning

Therefore, the task's performance is directly correlated with how well the particular environment has improved, which might be growing or decreasing.

Therefore, the task's performance is directly correlated with how well the particular environment has improved, which might be growing or decreasing.

$$\boxed{P \propto E}$$

Figure 8 indicates that depending on the particular environment, when a job is entered into the agent, the performance of error handling will either grow or decrease.

In this study, machine learning classifiers and clustering for pandemic forecasting are introduced. Both supervised and unsupervised learning perspectives are the foundation of the suggested paradigm. Three phases make up the operation. First, we classified our observations on the environment's performance (Random Forest). Second, we enhanced performance by enhancing the environment ($P \propto E$) [14]. To increase classification performance, we integrated the SIR to baseline models and SVM algorithm. Based on classification accuracy using SVM as a measure of infection characteristics, it is demonstrated that the suggested hybrid models, when combined with SIR and SEIR, generated highly pleasing outcomes.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6652
***Article Received**: 01 March 2023 **Revised**: 12 April 2023 **Accepted**: 28 April 2023*

VI. COMPUTATIONAL RESULTS

In simple terms, machine learning technology helps analysis and automate large chunks of data and make predictions in real-time without involving people. The ML algorithms are trained using the supervised learning paradigm on marked data sets, which requires them to generate a ground-truth output (continuous or discrete) for each input. Unsupervised learning, in contrast, does not provide a ground-truth result, and the algorithms frequently search the data for instances. In order to improve the cumulative reward and make it more suitable for upcoming decision-making tasks, reinforcement learning aims to raise the cumulative reward. Clumps analysis and dimensionality reduction are characteristics of unsupervised learning, whereas regression and classification are characteristics of supervised learning. In order to properly forecast if a patient has COVID-19 on their lab results, The program will keep learning from these encounters [15].

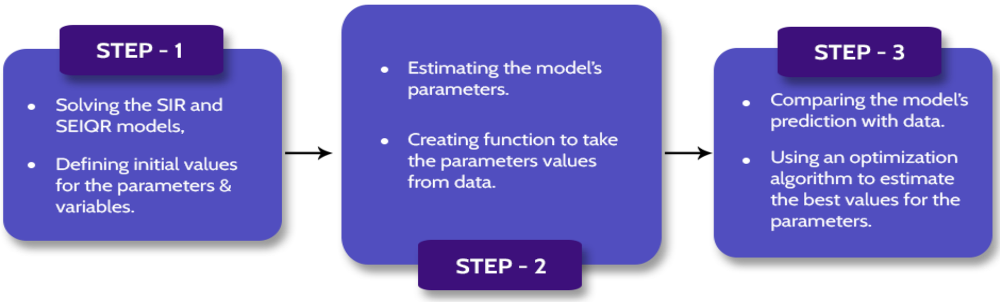
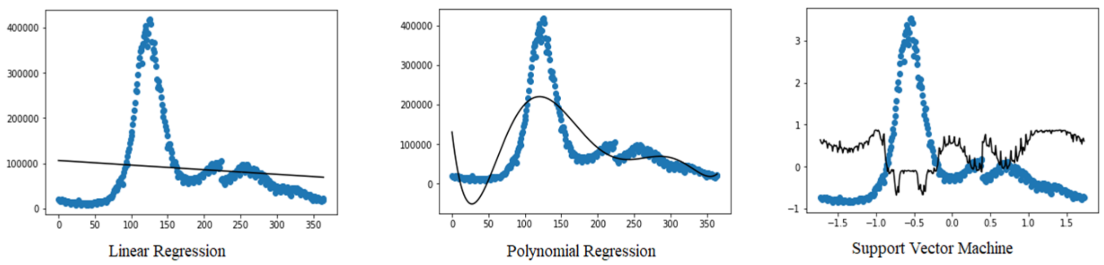
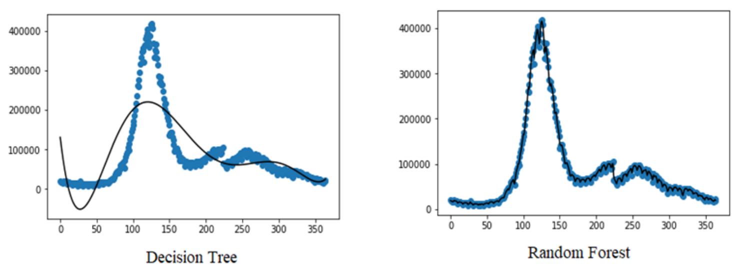


Figure 9. Workflows of Machine Learning Methods.

Based on cutting-edge AI methods to automatically identify COVID-19 using medical images, we categories the approaches and points of view that address the detection of (COVID-19) occurrences using machine learning. The algorithms employed in edge computing and for their cognition of machine learning activities analysis the imaging processes that can motivate machine learning techniques and are capable of looking for an analysis of complicated image and text data [16].

To help with future decision-making, machine learning (ML) based forecasting algorithms have shown useful in predicting perioperative outcomes. Identifying and prioritizing the harmful aspects of a threat has long been important in many application domains where ML models have been deployed. Numerous prediction algorithms are now in use to address forecasting issues [17]. This study shows how machine learning (ML) models may be used to predict how many COVID-19 patients will be impacted in the future, which is now considered to be a threat to humanity. Standard forecasting models, in particular Classification and Regression, to help with future decision-making, machine learning (ML) based forecasting algorithms have shown useful in predicting perioperative outcomes. Identifying and prioritizing the harmful aspects of a threat has long been important in many application domains where ML models have been deployed. Numerous prediction algorithms are now in use to address forecasting issues [18].





SN.	Algorithm	Accuracy
	(A) Classification	(Categorical Data)
1-	Logistic Regression	00.00%
2-	K-Nearest Neighbors	00.00%
3-	Naive Bayes Classifier	00.00%
4-	Decision Tree	89.04%
5-	Random Forest	99.91%
	(B) Regression	(Continuous Data)
6-	Linear Regression	10.32%
7-	Polynomial Regression	60.71%
8-	Support Vector Machine	58.34%

Table 1. Supervise Learning Algorithm Accuracy Percentage.

Figure10. Machine Learning Types and Approaches.

However, the majority of solutions focus on finding the factors that might be able to explain the historical data rather than predicting the number of newly infected people. There is no assurance that the parameters will be able to accurately forecast behavior in the future, despite the fact that these approaches aid in our understanding of the dynamics of the illness [19].

VII. Discussion & Conclusion

In this streamlined model, the birth rates is not taken into consideration. In order to improve the early identification of local epidemics, there is also an opportunity to update the model utilizing data sources like search engine searches. Data on outbreaks on cruise ships offer a controlled study to more precisely analysis cases that go unreported. The model also disregards the varying proportion of infected individuals who will require medical assistance [20]. This proportion varies according on how well-protected vulnerable populations, such as the elderly or those with weakened immune systems, one of the negative effects of protracted social isolation is economic upheaval, which sees businesses fail, unemployment rise, resources become limited, and help is needed to care for the most vulnerable. Repeated outbreak cycles that extend for several years are common when the virus mutates and a significant number of people are exposed. Instead, then merely "flattening the curve," the objective of this exercise is to increase social distance in order to shorten epidemic times and save underutilized healthcare resources.

- There was a large mismatch between the real data and the typical SIR model. However, configuring the SIR model's parameters greatly enhanced the forecast [21].
- Parameter tuning significantly improved the prediction accuracy of the SIR and SEIQR models.
- The modified SEIQR model outperformed the SIR model in terms of performance [22].
- For India, the SEIQR model with optimization makes better forecasts.
- The optimal values of the parameters were determined by the SEIQR model algorithm and the SIR model approach.

For other nations, the Prophet algorithm performed better. The improved versions of the SIR outperformed the Prophet algorithm, the logistic function, and the original SIR model, whereas the Prophet algorithm was surpassed by the logistic function for instances in all three nations.

When the baseline models were coupled with RF and assessed by being categorized into classes, the recommended hybrid models performed exceptionally well, showing that they may be employed as a management technique for prediction and classification issues. Additionally, none of the predictions made in this paper took social isolation and quarantine into account, even though this is an important area for future research. Evaluation of additional epidemiological models might be interesting in addition to the upgraded SIR and SEIQR models that were examined in this paper. Additionally, since these factors may enhance prediction accuracy, it would be advantageous to incorporate mathematical models with other variables such as changes in laws, societal norms, and constraints. According to these results and studies on related viruses, COVID-19 may be vulnerable to environmental variables. In light of the arguments above, it appears that each country has a different growth rate.

Acknowledgment

This work is acknowledged under Integral University Manuscript No IU/R&D/2023 – MCN0001790.

References

[1] Jin, Weiqiu, et al. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning." Computers in Biology and Medicine 146 (2022): 105560.
[2] Azar, Ahmad Taher, and Aboul Ella Hassanien, eds. Modeling, control and drug development for COVID-19 outbreak prevention. Springer, 2022.
[3] Rahimi, Iman, et al. "Analysis and prediction of COVID-19 Using SIR, SEIQR, and machine learning models: Australia, Italy, and UK Cases." Information 12.3 (2021): 109.
[4] Kamiri, Jackson, and Geoffrey Mariga. "Research methods in machine learning: A content analysis." International Journal of Computer and Information Technology (2279-0764) 10.2 (2021).
[5] Bayat, Sayeh, et al. "Identifying preclinical Alzheimer disease from driving patterns: a machine learning approach." Alzheimer's & Dementia 17 (2021): e057316.
[6] Rinaldi, Gianmario, et al. "Epidemiological model based periodic intervention policies for COVID-19 mitigation in the United Kingdom." Scientific Reports 12.1 (2022): 1-17.
[7] Huang, Guangqiu. "Artificial infectious disease optimization: A SEIQR epidemic dynamic model-based function optimization algorithm." Swarm and evolutionary computation 27 (2016): 31-67.
[8] Aràndiga, Francesc, et al. "A spatial-temporal model for the evolution of the COVID-19 pandemic in Spain including mobility." Mathematics 8.10 (2020): 1677.
[9] Fields, R., et al. "Age-stratified transmission model of COVID-19 in Ontario with human mobility during pandemic's first wave." Heliyon 7.9 (2021): e07905.
[10] Varanasi, Venkata Ramani, and Shaik Razia. "Intrusion detection using machine learning and deep learning." Int J Recent Technol Eng (2019).
[11] Kardani, Navid, et al. "Modelling of municipal solid waste gasification using an optimised ensemble soft computing model." Fuel 289 (2021): 119903.

[12] Varanasi, Venkata Ramani, and Shaik Razia. "Intrusion detection using machine learning and deep learning." Int J Recent Technol Eng (2019).
[13] Dutta, Krittika, et al. "A data mining based target regression-oriented approach to modelling of health insurance claims." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.
[14] Baig, Mirza Ghazanfar, and Sandeep Kumar Nayak. "Critical analysis on data science and big data avenues." International Journal of Scientific and Technology Research 8.11 (2019).
[15] Shadabfar, Mahdi, et al. "Time-variant reliability-based prediction of COVID-19 spread using extended SEIVR model and Monte Carlo sampling." Results in Physics 26 (2021): 104364.
[16] Pinter, Gergo, et al. "COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach." Mathematics 8.6 (2020): 890.
[17] Gudigar, Anjan, et al. "Role of artificial intelligence in COVID-19 detection." Sensors 21.23 (2021): 8045.
[18] Jamshidi, Mohammad, et al. "Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment." Ieee Access 8 (2020): 109581-109595.
[19] Baig, Mirza Ghazanfar, and Sandeep Kumar Nayak. "Machine Intelligence in COVID-19 Prediction: An Adaptable Approach For Automation." International Journal of Mechanical Engineering Vol. 6 No. 2 : 2021
[20] Vega, Roberto, Leonardo Flores, and Russell Greiner. "SIMLR: Machine Learning inside the SIR model for COVID-19 Forecasting." Forecasting 4.1 (2022): 72-94.
[21] Khaloofi, Huda, et al. "Performance evaluation of machine learning approaches for covid-19 forecasting by infectious disease modeling." 2021 International Conference of Women in Data Science at Taif University (WiDSTaif). IEEE, 2021.
[22] Agarwal, Praveen, et al., eds. Analysis of infectious disease problems (Covid-19) and their global impact. Singapore: Springer, 2021.