Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 6, July 2021: 5658-5671

Demand and Supply Analytical Model for Prediction to Determine the Requirement of Ventilators during COVID-19

Anil Dhasmana,

Assistant Professor, Department of Commerce,

Graphic Era Hill University, Dehradun, Uttarakhand,

Mahesh Manchanda,

Associated Professor, Department of Computer Science & Engineering,

Graphic Era Hill University, Dehradun, Uttarakhand, manchandamahesh@gmail.com

Arvind Mohan,

Professor, Department of Management,

Graphic Era (Deemed to be) University, Uttarakhand, arvindmj99@gmail.com

ABSTRACT: In the health care sector predictive analytics is always a need but right now in times of novel coronavirus 2019 (COVID-19) pandemic it became a necessity. During the early surge of the COVID-19, around the world many hospitals are facing problems to predict the exact number of required ventilators and bed capacity especially in countries like India where mostly government healthcare units are in poor shape. Various models were implemented based on the technologies available but in regions with large populations these models are not appropriate and are unable to give accurate predictions. This paper discusses a predictive analytical model which can be used to facilitate hospital's prediction power in determining the ventilator requirements for COVID-19 patients, using the provided data from hospitals. This paper aims on the implementation of a predictive algorithm based on the available data proportion of COVID-19 patients admitted in ICU and hospital for determination of required number of ventilators in advance. The implementation is carried out using a supervised machine learning technique called random forest regression algorithm. Moreover based on the model's prediction and the current supply of ventilators in hospitals, additional demand for ventilators could be prepared in advance for critical patients which may require it at any time. And the predicted increased demand of ventilators can be produced by the supplier industry, it is an example of industry intelligence.

KEYWORD: COVID-19, Demand and Supply of Ventilators in Hospital, Predictive Analytics of Ventilators, Random Forest Classification algorithm, Supervised Machine Learning Prediction Model.

INTRODUCTION

Nowadays hardly you get to see any sector of industry which doesn't use predictive analytics for quality products. Many sectors are already using these models and others are planning in one or another way. But in medical sectors

it's become a necessity seeing the current situation involving advanced technology and artificial intelligence in scientific innovations is very important. Predictive modeling is usually associated with meteorological terms and customer satisfaction hence it has many applications in the online software industry. But in the health care sector it comes with more responsibility. Given that the pandemic is claiming so many lives around the world, a second wave of cases is expected with the start of the next year itself, and at this critical moment having a predictive tool to predict resource needs such as beds or ventilators for hospitals. Based on right predictions, planning and preparation can be done responsibly and properly. It will automatically increase the faith of people in productivity of hospital staff and the hospital management. This model is also important for the hospitals to manage the upcoming surge of COVID-19 patents because if proper strategy is not considered and unfortunately not handled properly many people are literally going to lose lives due to unavailability of the resources at the required time. A major benefit is, even the private hospitals can utilize this advanced model which makes them capable of handling pandemic situations. It will provide assurance to federal governments also that in case of emergency private hospitals have sufficient resources available. They can easily keep track of availability of the ventilator to withstand any upcoming situations together and to meet expected needs of new COVID-19 patients without having to store a substantially greater number of beds or ventilators in advance.

Predictive analytics methods are for providing solutions to the problems by initially extracting information from available and existing sources of model data, and determining the patterns. And based on these experiences predicting the future results related trends and conclusions. Computer Science offers different techniques to make such predictions, like machine learning, artificial intelligence, statistical modeling, etc. predictive analytics solutions are a reliable method of forecasting, since it also focuses on what-if scenarios means, while making predictions the machine will also be competent enough to deal with some or other unexpected condition along with this risk management is another important aspect these techniques take into consideration. In market intelligence along with these fundamentals, it adapts to the needs of the industry also and looks after the needed innovations of organizations [1]. For any predictive analysis first data analytics is important, it is a process of monitoring and transforming the data and drawing useful conclusions from it. Several algorithms are used to analyze the data and understand its trend to gather the information from it which would have been missed otherwise. The information obtained can further be used to optimize and enhance the processes of a business firm. These guide the improvement areas and help the company to identify their weak areas. Data analytics comprises various techniques to analyze the data and interpret it. The diagram given below sums up all the types of data analytics which collaboratively help to make a successful prediction [2]. Figure 1 illustrate the categories of data analytic.

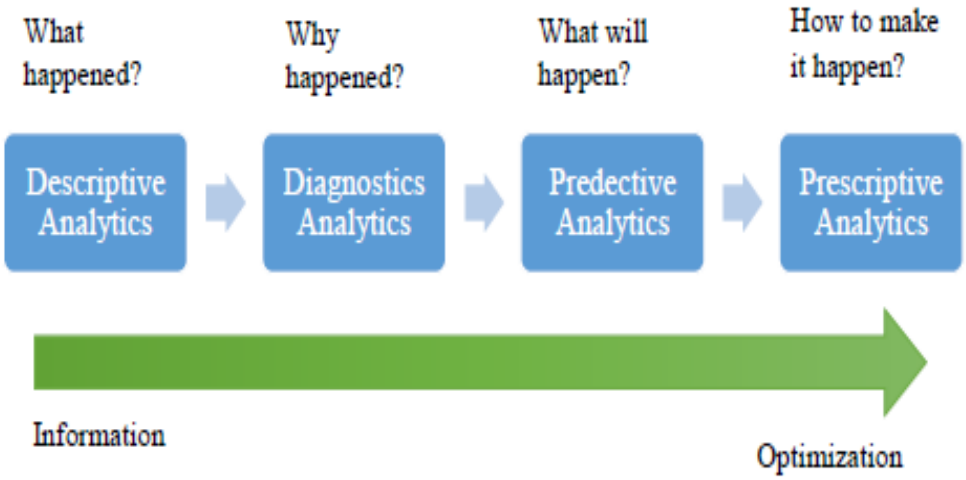

Figure 1: Categories of Data analytic

1. Applications of predictive modeling

1.1. Mechanical Ventilator:

A ventilator is a mechanical machine that provides assistance to a patient to breathe i.e. ventilate when an individual cannot breathe his or her own due to a critical respiratory illness or for some other emergency. On a ventilator the patient is connected to a hollow tube i.e. an artificial air passage that goes inside their mouth and then down into the main airway pass way or trachea. Patients remain on the ventilator until their condition improves enough to breathe on their own and no longer have they required any assistance with it [3]. Basically the machine ensures that the connected body gets sufficient oxygen and that carbon dioxide is eliminated. This becomes necessary for severely affected patients of COVID-19 because it directly harms the respiratory system of the body and hence a ventilator becomes a necessity for the patient with certain illnesses which hinders normal breathing. It all depends on what will be revealed as a common trait of those who are so severely sick with COVID-19, they need ventilation. This information will then be ready to compare with newly hospitalized patients, allowing doctors to accurately forecast whether someone will need ventilation. When does a COVID-19 patient need a ventilator how early and for how long? Once a doctor sees that a patient needs a ventilator, “it is required to be connected quickly”. The patient can be kept stable only for very short periods of time utilizing manual forms of ventilation such as using a bag and mask system with oxygen, but it's important to connect to a ventilator within half an hour in critical condition [4].

1.2. Time Series Model:

Time series model comprises a sequence of different data sets captured to implement, using time as the main input parameter. It utilizes the data of previous times to develop and get a numerical metric to estimate and predict the next weeks or months of data outcomes using that obtained matrix. A singular matrix is designed and developed over a time span with an appropriate level of accuracy which is beyond simple averages. For more understanding consider a case in which an observer is noting the calories, food weight, variety and nutrition intake of a person since a month. After an adequate time maybe a month, now an observer is able to make a chart or matrix from the observed information about the diet intakes of that person. Chosen features like calories, nutrients and food weight dataset will be used as training data. Now observers can predict the next meal of that person based on the experience obtained from the training in the time series model. Another example could be if a salon owner wishes to predict how many people are possibly going to visit the salon today? Salon owner might turn to the obscene method of averaging the gross number of customers or visitors over the past 30-60 days. However, growth is not always static or linear like this in the real world so this method could give bad predictions, but the best part of a time series model is that it can conveniently tackle such scenarios and prove a better model for exponential growths. It is capable enough to accurately align the model to the industry's trends. Moreover it also forecasts for multiple projects or at multiple regions simultaneously instead of supporting just one mode at a time [5].Now after understanding the time series model let's proceed towards the predictive algorithm for the proposed Model.

1.3. Common Predictive Algorithms

Predictive analytics algorithms can be divided into two parts that are, deep learning and machine learning. Machine learning (ML) is a subset of artificial Intelligence and it utilizes structural data that means data arranged in a tabular form for analysis. In this subpart algorithms consist of both linear and nonlinear algorithm varieties. More specifically linear algorithms train data more rapidly or immediately, while the other one nonlinear possess better optimization for the problems they face and tackle. Deep learning (DL) concept is a subset of machine learning and is more famous to deal with audio, images, text, and video like multimedia data types[5].In machine learning predictive modelling provides several different algorithms that can be implemented. Some of the most common

algorithms that are being applied to power the predictive analytics models and for this proposed model Random Forest Regression is chosen which comes under supervised machine learning type.

There are various algorithms available in ML to implement predictive analytical analysis but for this proposed model, the algorithm will be selected according to the type of problem statement it is implementing. Next important aspect is the dataset which is going to be considered for the training and testing of the model in case of supervised machine learning. This problem statement consists of a continuous data set and for this type of problem and data, a random forest algorithm is best to implement with highest accuracy. So this accurate predictions will not only help the people, hospital or government in terms of faith and management but also the businesses. People related with ventilator manufacturing and supply would get directly benefited. As demand for ventilators will increase in surge of COVID-19 in this scenario if a predictive model can tell in advance exactly how much ventilators a hospital requires then Suppliers can increase the production based on these predictions. Moreover demand will directly affect the supply chain of companies in the right direction.

LITERATURE REVIEW

There are plenty and decent number of research papers have already been published that help to determine and analyze abundant amounts of knowledge about predictive analytics models using machine learning algorithms. one paper which utilizes the Random Forest algorithm titled as 'Random Forest Based Ensemble Classifiers for Predicting Healthcare-Associated' written by Garcia, Hernández, F. S, J. C. B., Barba, M. S., and M. N. M., Herráez. (2016). This paper implemented a model which works on Surveillance and prohibition of infections present in the hospital environment which is a tough challenge for any current healthcare systems considering both facts that the great impact these kinds of infections can have on patient mortality and the huge sanitary costs [6]. Utilization of data analysis is discussed which contributed to make these tasks easier by means of identification of risk factors and prediction of infection acquisition. This work is focused on the study of infections acquired in intensive care units by means of data mining models and random forest algorithm. The aim of the proposal was to overcome drawbacks which is common to occur while utilizing other usual strategies to implement such models. So an attempt was made by utilizing different data mining algorithms in model for predicting any device associated infections in future patients.

Another research review called 'A Predictive Model for Patient Census and Ventilator Requirements at Individual Hospitals during the COVID-19 Pandemic: A Preliminary Technical Report' written by Franklin Dexter and Richard H Epstein. They worked on an analytical model to help hospitals in estimating their census and ventilator requirements for corona patients for future periods of the virus pandemic, by using the data provided by hospitals. But the implementation was performed within an Excel such as Microsoft, Redmond, WA workbook and without the use of algorithms or any programming model. Data validity and internal consistency are checked within the workbook, and errors are identified. The outcome of the model was nearly instantaneous, producing an estimate of the census and the number of ventilators required in one, three, and seven days following the date on which the simulation is run [7]. But it is not the most efficient and productive way or model to implement in a hospital with a huge population. It won't be able to provide maximum efficiency unlike machine learning predictive analytical models such as random forest.

Research Questions

- Is there any software model available which can determine the requirement of ventilators and facilitate the ventilator management work of the hospital during COVID-19 pandemic?

- How can the government supervise the total availability count of the ventilators in states and would it help hospitals also for ventilator management?

METHODOLOGY

1. Design:

1.1. Supervised Machine Learning:

Supervised ML is a method using which a machine is trained, machine firstly, learns from the training data and now based on the experience it can make predictions independently. Basically supervised ML is a branch of artificial intelligence in which machines are trained for some purpose but since it is trained using the training data set it is called supervised. Similarly, like a function that maps an independent variable's value to a dependent variable. It deduces a function from a data set labelled as training data which is considered as a supervisor training data it helps the machines experience and learn to make accurate predictions in future. In supervised learning, an optimal scenario will be that the machine correctly determines the output for a new input i.e. predict the different class labels for unseen instances. In the most basic way, a supervised learning algorithm can be depicted with an expression y = f(x). In this statement y is the predicted output class or the dependent output that is determined by a mapping function which designates a class to an independent input value x. And the functions used to associate the input features to an accurately predicted output is generated by the machine learning model. Supervised learning can be divided into two sub divisions regression and classification. Supervised classification and regression problems can be solved using a handful of algorithms. Whichever algorithm you choose to implement it completely depends on the data set you choose and the circumstances available for that respective problem. Few of the popular classification algorithms are linear classifiers, random forest decision trees, k-nearest neighbor, and support vector machines etc. [8]. We will focus on Regression Problems since for our problem the data type is continuous in nature.

1.2. Regression Problem:

Regression is a forecast or predictive mathematically statistical process in which the model tries to find the significant correlation between independent and dependent variables. The goal of a regression algorithm is to predict continuous values or numbers such as test scores, income, sales and no. of ventilators. The equation for basic linear regression can be simply represented as: y= w [0]*x [0] + w [1]*x [1] +.....+ w[i]*x[i] +b. Where x[i] represents the features for the data set and w[i] and b represent the parameters which are generated during the training phase. For instance, with only one feature in the data the simple linear regression model can be represented using the formula y = wx + b. Where w represents the slope, x represents the single feature and b is the y-intercept. For models using two features, the panel will be used representing 2d models. Similarly, for a model having more than two or more features, will be represented using a hyperplane[6].

The graph as shown in Figure 2 is representing a clear correlation between dependent variable y and independent variable x, the dots are representing the data points and the linear line is depicting the model or hypothesis implemented using these datasets. This proposed model will be implemented using regression technique to find a solution for ventilator problem statement. The considered ventilator problem statement is a regression type problem as the considered features related to this problem is of continuous data type. In case of continuous data samples, regression supervised learning technique is used for implementation. The famous types of regression algorithms are logistic regression, linear regression and polynomial regression.

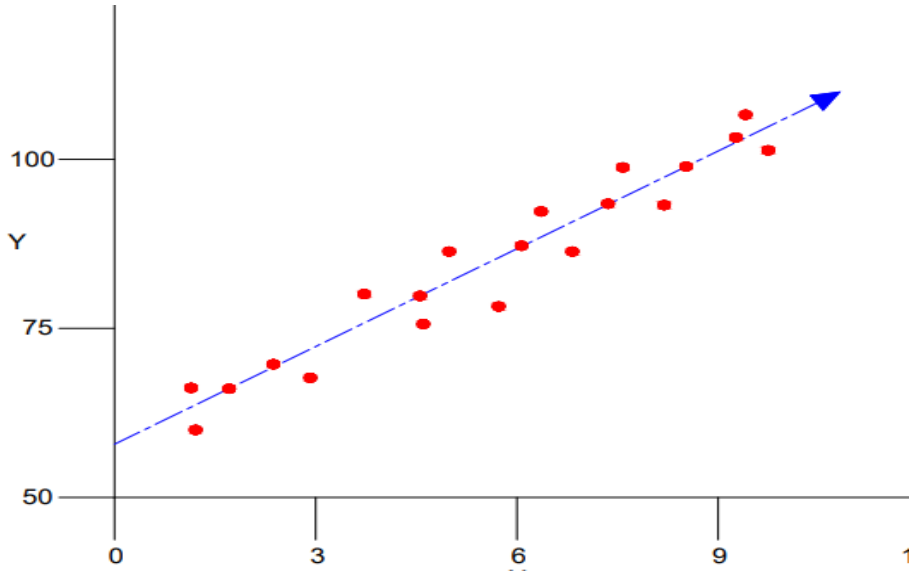

Figure 2: Visualization graph of a linear regression model representing correlation between dependent variable y and independent variable x.

1.3. Random Forest:

Random forest is probably the most popular algorithm, capable of implementing both classification and regression type problems. It can potentially and accurately classify huge volumes of data set and with maximum accuracy. Random forest name is derived from the basic fact that this algorithm is a combination of many decision trees. Each decision tree depends on the independently sampled values of a random vector. Each one can grow to the largest size possible. Every predictive analytics algorithm tries to achieve highest efficiency i.e. the lowest error possible by either using boating or bagging ensembles techniques. It includes another aspect trade off between bias and variance which is considered as a main issue in machine learning. It is basically defining the balancing between a model having high flexibility i.e. high variance that learns the training data or is trained very well but at the cost that it cannot generalize the new data now, and an inflexible model i.e. high bias that is not able to learn from the available training data. But a random forest algorithm can reduce the variance of a single decision tree which leads to the better predictions made on new data (Figure 3).

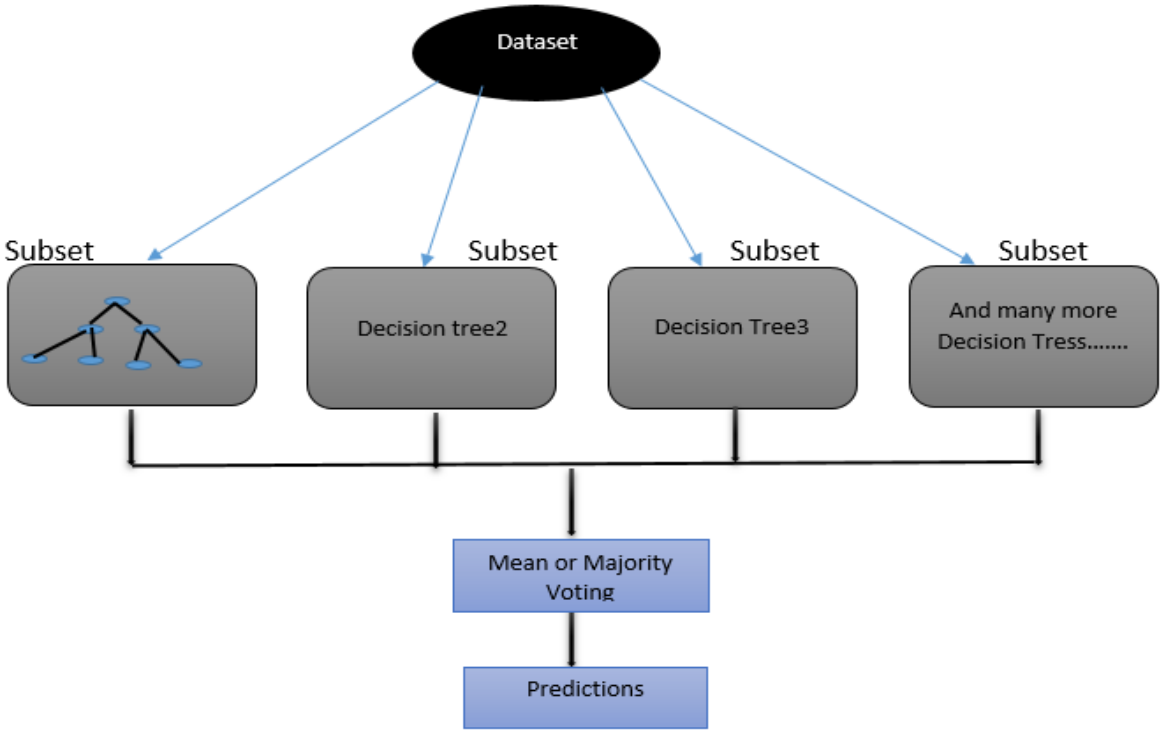

Figure 3: Functionality of random forest algorithm. Steps showing how the dataset is divided to build many decision trees called 'forest', to obtain a collaborative prediction from each decision tree.

The principle of a random forest algorithm is that while individual trees may be weak learners, together with each other they can comprehend a single strong learner. The popularity of the random forest model is explained by its various advantages, some of them are its accuracy and highest efficiency rate when running on large databases which is not common with other algorithms. It is resistant to over fitting unlike decision trees. Multiple decision trees together reduces the bias and variance value of a smaller set or single decision tree. It is capable of handling thousands of input variables without any variable deletion. It properly handles and provides effective methods for determining the missing data. Maintains accuracy even when a large proportion of the dataset is missing or unavailable [5]. Now after knowing the techniques and selecting the algorithm for our model next thing is working on data samples for it the first thing is to do is:

1.4. Features selection:

1.4.1. Principal Component Analysis (PCA):

Instances based on the features i.e. the column names for our model which we will consider. Then Feature scaling, it is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step. After collecting large datasets which are increasingly extensive in many dimensions and disciplines. In order to interpret correctly in a perfect order and form such datasets, after sampling specific methods are needed to significantly decrease their dimensionality and volume in a proper interpretable way, such that highest possible original information in the data is protected and preserved. Principal component analysis is oldest and most widely utilized. In simple sense reducing the size of a dataset but keeping the 'variability' preserved i.e. statistical information as much as possible

[9]. PCA is a process of computing the principal components of the model and using them to perform a change in model on the basis of the data, sometimes using only the first few principal components and ignoring the rest. PCA is used in experimental data analysis and for making predictive models. Main purpose is dimensionality reduction by projecting each data point only on the first few principal components.

2. *Sample:*

For perfect data samples, selection of right features is important so that the proposed model can understand the relevance of each feature with the problem statement. Including unnecessary or irrelevant features can cause situations like over fitting. Table 1 illustrate the data set used by random forest algorithms. The selected features (columns) for the sample dataset and proposed model are:

- Feature which represents no. of COVID-19 patient's admission in hospital on a daily basis.
- Count of patients with severe of COVID-19 symptoms
- Count of patients with past health or medical history like respiratory issues etc.
- Duration of mechanical ventilator occupied by each patient.
- Availability Ventilators Present in the Hospital.
- Average age group of COVID patients.
- Count of patients who are already on ICU beds.

Table 1: The data set used by random forest algorithms include features like new admissions, available ventilators, Patients with Severe condition, age group etc.[10]**.**

S.No	Beds	Bed_occupancy	Bed_change_%	ICU_beds	Ventilators	V_Availability	population_change	Medical_History	Severe_Symptoms	age 1 to 64	age over 65
0	1305	75.5	-0.48	7.3	0.02569	334	-0.27	46	48.7	60.1	27
1	1227		2.85	10.6	0.01912	517	0.1	40.2	43.4	72.6	13.9
2	805		-3.04	8.3	0.02742		0.09	36.6	42.5	68.2	14.2
3	800	62.1	-0.86	38.7	0.02993	233	0.47	46	48.2	65.5	21.5
4	737	73.8	-0.9	21.8	0.02792	106	0.72	42.8	45.1	66.7	19.2
5	702	65.5	-0.07	13.8	26.433	105	-0.24	40.4	44.3	67.1	18.6

					5						
6	663	70.1	-0.26	11.6	0.03301	135	0.22	40.8	43.4	65.6	19
7	662		0.04	6.9	0.02666	123	-0.09	39	42.4	68.4	16.8
8	656	73.2	-2.65	15.5	36.2368	43	-1.49	39.7	47.1	66.2	19
9	598	75.6	-1.23	11.6	0.01418	123	0.21	39.6	43.1	62.2	19.7
10	582	67.8	0.08	9.2	10.995	111	0.07	38.8	42.3	69.6	15.1
11	576	81.8	-0.73	15.9		377	0.5	40.2	42.7	64.3	18.6
12	557	71.1	-1.03	9.7		30	-1.13	39.7	46.9	64.8	19.8
13			0	7.1			0.87	43.5	45	72.2	16.3
14	469	70.4	-1.68	14.6		29	0.21	39.4	46.1	64.2	19.5
15	466	70.7	-2.63	24.8		237	1.9	38.7	39.9	69.2	14.3
16	453	82	-0.82	11		208	0.77	41.4	43.4	66.7	18.4
17	450	69.5	-0.28	6.4		103	0.04	42.8	46.2	66	19.1
18	434		6.55	3.6		145	0.43	36.5	38.4	71.7	10.6
19	421	61.6	-0.18	6		81	-0.46	43.5	45.6	65.4	20.4
20			0.88	9.1	5.2136	3	1.23	37.9	39.5	65.5	15.5

					2						
21	360	80.7	-1.76	8	14.873	16	0.77	38.4	40	65.4	16.8
22	339	66.8	-0.01	4.2	0.01369	112	-0.29	40.2	44.4	64.9	21.5
23	332	65.4	0	6.4		419	0.22	41.5	43.6	64.8	18.8
24	328		-10.56	6.1		16	0.17	40.9	44.3	62.4	21.2
25	318	78.9	-1.01	12.5	0.00879	200	-0.13	44.4	46.5	63.5	23
26	306		-1.29	9.1		3.5	0.69	35.9	37.1	65.5	14.4
27	302	93.3	-0.58			417	1.64	29.3	30.6	60.4	11.7
28	297	75.3	0.08	9.7		93	0.09	41.5	43.9	65.9	19.4

3. Instrument:

3.1. Python

Python is a high-level, general-purpose and interpreted programming language. Most widely used as a backend development language in the software industry. This language has a constructive and object-oriented approach which aims to help programmers write clear, fast logical code for small and large-scale projects. Widely used to implement the machine learning algorithms and implementation of predictive analytical models [11]. Python provides a number of libraries that can be used to implement different requirements of data science. Some of them are pandas, a data analysis tool, numpy which facilitates mathematical operations on arrays data structure and their vectorization.

3.2. Anaconda Distribution Edition

Anaconda is an open-source tool available for free for the distribution of the Python and R programming languages mainly for scientific computing works like data science, ML applications, predictive analytics, large-scale data processing etc. It aims to simplify the package management activities and deployment of the software. The reason for installing Anaconda is that it comes with a lot of preinstalled packages and spyder is one of them. Spyder is an integrated development environment (IDE) and a powerful tool for python based machine learning i.e. spyder also known as scientific python development environment. Some of its basic features give you an upper edge over the conventional IDE's. It includes editing, debugging, testing, and introspection features[11].

4. Data Collection:

After getting the correct dataset in proper format and with required dimensionality. After selecting technique, algorithm and data set, the next task is to design the model. The block diagram as shown in Figure 4 is representing the algorithm according to which these methodologies are going to be implemented. That means this block diagram is representing the algorithm processing. The steps it includes are firstly, scaling and then normalization of the obtained data set. Then the next step according to the random forest algorithm is divide the final dataset into two parts one as training data and another is testing data. Third step is model training. The training data set is used for the purpose of training the model, it gives input to the user and then output for that respective input this step is executed again and again for each data set. Fourth step is model testing in which a testing data set is used to test the trained model. Fifth and final step involves the predicted result obtained from the testing data set is then compared with the actual result and based on the outcomes the mean square error for the model is estimated. This value of mean square error represents the predictive power of the model and random forest algorithm.

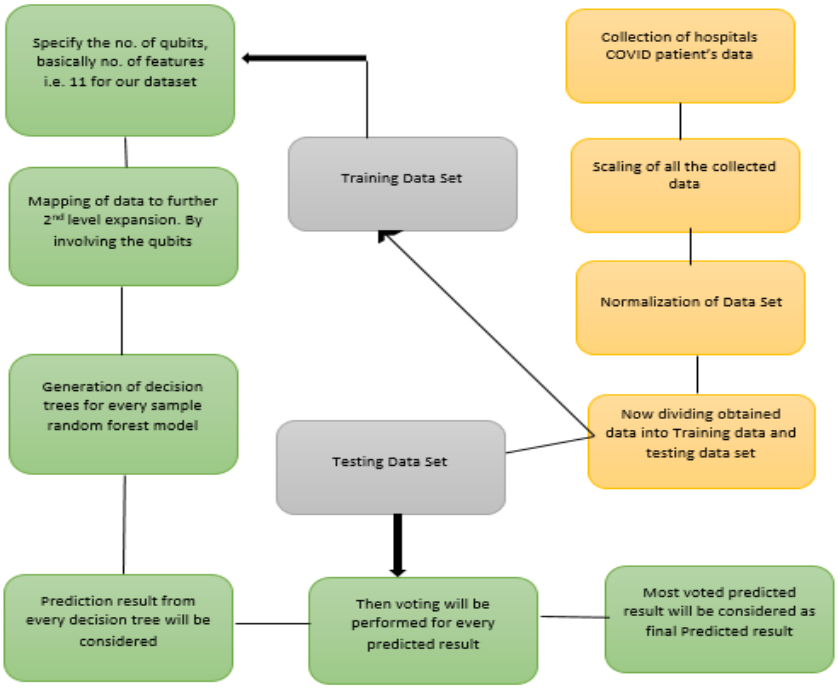

Figure 4: System design for determining the number of required ventilators in a hospitals based on random forest algorithm.

5. Data Analysis:

5.1. Model Training:

Final thing after getting sampled and modified dataset we to train our model used on the training dataset. We know that the random forest model is generated from many decision trees. Rather than just traditionally averaging the outcome predictions from trees which called as a "forest", this model uses two key concepts that gives it the name random, first is randomly performing sampling of training data points while generating decision trees and second is it can be random subsets of features when splitting the nodes i.e. random sampling of training observations, when

training, each tree in a random forest learns from a random sample of the data points. The samples are drawn with replacement, known as bootstrapping, which means that some samples will be used multiple times in a single tree. The idea is that by training each tree on different samples, although each tree might have high variance with respect to a particular set of the training data. Overall, at the final stage the entire forest will have lower variance but not at the cost of high bias.

5.2. Model testing:

Model testing is performed using the, K-fold cross validation technique breaks the training data into equal-sized k "folds." It traverses through each fold, one by one and for instance treats that fold as hold up data, and training a model on all the other left K-1 fold, based on it evaluating the model's performance on that one considered holdout fold. This results in generation of K different models, each one with an accuracy score out of the sample model on a different taken hold out set. The model's cross-validation score is obtained by averaging these K models' scores [11]. Since it provides a lower variance estimation of the model's true out of sample score hence cross validation is very useful in comparison of a single train test decimal tree split method. Training every individual learner on different jack up subsets of the data and then taking average of the predictions is called as bagging or bootstrap aggregating. Bagging is a ML ensemble Meta algorithm technique used to improve the accuracy and stability of machine learning algorithms utilized by classification and regression techniques [12]. Random forest algorithms use this technique because it reduces the variance and helps to avoid overfitting problems.

RESULTS AND DISCUSSION

The purpose of this paper was to develop a predictive analytical model for determining availability and requirement of ventilators in hospitals based on available data of COVID-19 patients and hospitals to accomplish these tasks like selecting the best type of machine learning technique for implementing an efficient algorithm. After collection of proper data from the hospital and performing sampling the proposed system adopted a regression technique and implemented the random forest algorithm for getting the most accurate prediction and an efficient model. This model could be implemented using a decision tree which is also a good machine learning algorithm but it could arise problems like over fitting which occurs due to unnecessary numbers of feature selection. Decision tree algorithms are prone to over fitting problems to avoid issues such as over fitting and other difficulties. Random forest algorithm is considered.

The results proved that multi regressions that are random forest showed better behavior than single regression i.e. better than a decision tree, especially behavior like bagging and boosting technique when random forest was utilized as base regression. The proposed model study has revealed that the random forest is a powerful way to handle the imbalanced data problems and can be an alternative approach to the usual procedures for large amounts of data. The model inputs for each currently hospitalized patient with COVID-19 are the duration of hospitalization, whether the patient is currently receiving or has previously received mechanical ventilation, and the duration of the current ventilation episode, if applicable. The model effectively takes these inputs and gives a model with mean square error as 10.64 as outcome. Based on these obtained predictions from this proposed model now the hospitals can manage their ventilator requirements. This model can also detect the number of ventilators to be manufactured just by comparing the predicted value of the model with the number of available ventilators of the hospital. So that vendors can manage the production and supply accordingly and this is where a predictive analytics model impacts the people and market effectively.

Further discussion on topics like what can we do to minimize the variance and increase the accuracy rate of the model. Moreover, data analytics can be prioritized for obtaining better data samples. Good training set is basically the good our dataset will be, the better training dataset would be which would result in a brilliant prediction model. In coming to conclusions based on what the analysis already shown in this paper the techniques and processes of

predictive analytics is vast to understand and implement. Predictive analytics is boon for business intelligence and profit, it just not only makes predictions but also gives scope to innovation in industry at so many levels.

CONCLUSION

Aim for this paper was to implement a model for determining the number of required ventilators in a regression problem using the random forest algorithm. While implementing this proposed predictive analytics model a thorough study of the different aspects related to problem statement, technologies and algorithms is done to get a proper model. Shortage of ventilators in hospitals can put lives at risk, so health authorities around the world are doing everything they can to get more advanced healthcare technology in the medical department. Healthcare authorities around the world are trying their best to find ways to fill their hospitals with more equipment and ventilators to handle the increasing numbers of COVID-19 patients so to fulfill this requirement the model proposed in this paper can be considered.

To achieve this model firstly, an analysis is done on data analytics and its importance for the proposed model. How data analytics is used with the machine learning techniques for implementation of predictive analytical models. Predictive analytics use previously experienced data to predict what pattern will occur in the upcoming future and exactly what actions can affect those outcomes directly or indirectly. Advanced organizations use various predictive analytics to make right decisions that help their business to grow or in case of the medical industry to save lives. Predictive analytics facilitates businesses to optimize efficiency and performance of business. Implementing this proposed system into the business model means companies can effectively reduce costs by estimating right demands and based on products or ventilators supply in the market. This proposed system is also useful in identifying many other efficient ways of doing business just by storing large amounts of relevant data. Companies can also use predictive analytics to make better business decisions by analyzing customer trends and satisfaction, which can lead to new and better products and services in the market which ultimately increase the brand value of the company or Industry.

Supervised learning is the simplest subcategory of machine learning and it is widely implemented and used in industry right now. Supervised learning is commonly used and has proven to be an excellent tool in many fields. Determining numbers of ventilators for a hospital required during a pandemic is a regression type problem and can be tackled using regression techniques of supervised machine learning and assuming how continuous and large the hospital data can be, accuracy is a main requirement here, performance and cost are also factors which vote for choosing a random forest algorithm to implement this regression problem. This model has a lot of scope in hospitals for now as well for the future as according to expert studies this pandemic is going to affect the nations even next year so to cope up with every possible situation hospitals must be prepared. Moreover future work can be done to improve different other broad aspects of this proposed model to implement it in the medical industry of any scale.

REFERENCES

[1] B. Nithya and V. Ilango, “Predictive analytics in health care using machine learning tools and techniques,” 2017, doi: 10.1109/ICCONS.2017.8250771.

[2] P. Rod Koch, CMA, “Business Intelligence to Predictive Analytics,” [Online]. Available: https://sfmagazine.com/wp-content/uploads/sfarchive/2015/01/TECH-PRACTICES-From-Business-Intelligence-to-Predictive-Analytics.pdf.

[3] Clevelandclinic, “mechanical-ventilation.” https://my.clevelandclinic.org/health/articles/15368-mechanical-ventilation#:~:text=a mechanical ventilator%3F-,A mechanical ventilator is a machine that helps a

patient,their main airway or trachea.

[4] kaggle, “Ventilators Shortage.” https://www.kaggle.com/mpwolke/ventilators-shortage/comments.

[5] SRIRAM PARTHASARATHY, “Top 5 Predictive Analytics Models and Algorithms.” https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/.

[6] M. N. M. García, J. C. B. Herráez, M. S. Barba, and F. S. Hernández, “Random forest based ensemble classifiers for predicting healthcare-associated infections in intensive care units,” 2016, doi: 10.1007/978-3-319-40162-1_33.

[7] R. H. Epstein and F. Dexter, “A Predictive Model for Patient Census and Ventilator Requirements at Individual Hospitals During the Coronavirus Disease 2019 (COVID-19) Pandemic: A Preliminary Technical Report,” *Cureus*, 2020, doi: 10.7759/cureus.8501.

[8] Machinelearningmastery, “supervised-and-unsupervised-machine-learning-algorithms/.” https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[9] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016, doi: 10.1098/rsta.2015.0202.

[10] kaggle, “hospital dataset.” https://www.kaggle.com/treich/country-metadata.

[11] James LeDoux, “K-Fold Cross Validation,” *James LeDoux*. https://jamesrledoux.com/code/k_fold_cross_validation#:~:text=K-fold cross validation works,on the one holdout fold.

[12] Machinelearningmastery, “bagging-and-random-forest-ensemble-algorithms-for-machine-learning.” https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/.