WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

COVID-19 FUTURE FORECASTING USING MACHINE LEARNING

Prof. Aparna Patil, Sagar Shinde, Saurabh Surwase , Tejas Patil, Abhishek Chaskar

Computer deparment, Genba sopanrao moze collage of engineering,balewadi pune.411045

A B S T R A C T

Machine learning-based predictive methods (ML) have shown their importance in anticipating perioperative outcomes improve decision-making over the course of future actions. ML models have long been used in many program domains which required identification and prioritization of negative aspects of the threat. A few predictive methods are widely used dealing with forecasting problems. This study demonstrates the potential of ML models to predict the number of future patients affected by COVID-19 is currently considered a major threat to humanity in particular, the four most common predictor models, such as linear regression (LR), minimal reduction and operator selection (LASSO), vector support (SVM), and exponential smoothing (ES) was used in this study to predict the alarming properties of COVID19. Three types of prediction they are made for each model, such as the number of newly infected cases, the number of deaths, and the number of recovery for the next 10 days.

Keywords*: Supervised machine learning, Future forecast, Covid 19.*

1. INTRODUCTION

This predictive problem is considered to be a retrospective problem in this study, so the study is based on some modern supervised ML. regression models such as linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector support machine (SVM), and exponential smoothing (ES). Study models are trained using the COVID-19 patient statistics database provided by Johns Hopkins. The database was previously processed and divided into two sub-sets: training set (85% records) and test set (15% records). .In the current human problem our effort in this study is to improve the COVID-19 predictive system. The prediction was made in three significant disease variables over the next 10 days number of New confirmed cases, number of deaths, number of recovered patients

2. LITERATURE SURVEY

The purpose of this article is to introduce a new strategy for identifying high-density and mobile congestion areas, which are at risk of distributing COVID-19. The densely populated dynamic districts (so-called endangered regions) are at risk of spreading the disease in particular if they contain asymptomatic people infected with the virus and healthy people. Methods: Our system identifies endangered regions using an existing cellular network functionality — redistribution and selection of (regenerated) cells — used to store seamless coverage on machines used by the end user (UE). Frequent delivery once Cell re-selection events are a major indicator of overcrowding in the area as almost everyone is in control of UE. Results: These ratings, collected in most UEWs, allows us to identify atrium regions without compromising individual privacy and anonymity. Conclusions: Areas considered risk factors may be subject to additional employment and risk reduction.

EXISTING SYSTEM

- In our current system there are no predictions available so the government is facing many problems related to public
- Problems such as the provision of medication and oxygen in an area where more infected peoples , Deciding to set strict limits on day to day life routines

PROBLEM STATEMENT

COVID-19 Time Series Forecasting of Daily Cases, Deaths Caused and Recovered Cases Using Linear Regression

In this paper the data sets used are based on IoT-based live databases to build a modern COVID-19 breakdown model. We've included data-driven measurements as well time series analysis to predict future trends as the number of confirmed cases Well, the number of deaths caused by the virus and the number of people recovering from this novel corona virus. In order to achieve the measurements, we have used the In-Depth Learning Outline Model. Linear regression uses relationships between data points to draw a straight line This line can be used to predict future cases

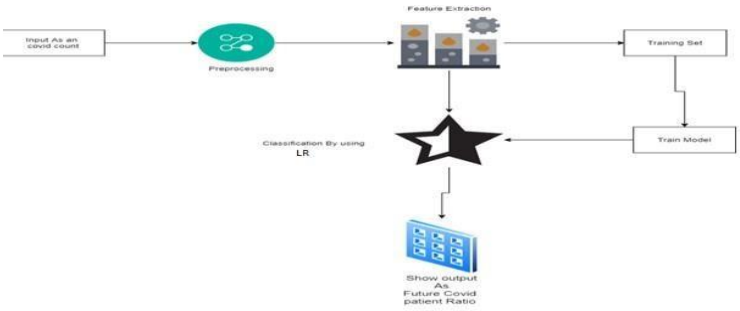
3. SYSTEM ARCHITECTURE

INPUT:

COVID-19 input data was collected to identify new daily cases of death, and recovery cases at the World Health Organization (WHO) and the World meters. Our learning region is global. This database is considered a live database so it can change every second.

PRE-PROCESSING :

Pre-data processing is the process of preparing raw data and adapting it to a machine learning model. At first and an important step in creating a machine learning model. Real-world data usually contains sounds, missing values, and perhaps in an unusable format which can be directly applied to machine learning models. Pre-data processing is a necessary task to clean the data and make it suitable for the machine learning model which also enhances the accuracy and efficiency of the machine learning model.



FEATURE EXTRACTION :

In real-world machine learning problems, there are often too many features (features) based on what the last prediction was made. The higher the number of features, the harder it becomes it visualizes a training set and works on it. Sometimes, many of these factors are related or unwanted. This is where Feature Release begins.

Algorithm used in proposed system is as follows:

- SVM Algorithm
- Linear Regression
- Classification Techniques

4. PROPOSED ALGORITHM/METHODOLOGY

Algorithm used in proposed system are as follows:

1. **LASSO**: LASSO is a regression model belongs to the linear regression technique which uses shrinkage . Shrinkage in this context refers to the shrinking of extreme values of a data sample towards central values.

   The shrinkage process thus makes LASSO better and more stable and also reduces the error . LASSO is considered as a more suitable model for multicollinearity scenarios. Since the model performs L1 regularization and the penalty added in this case is equal to the magnitude of coefficients.

   So LASSO makes the regression simpler in terms of the number of features it is using. It uses a regularization method for automatically penalizing the extra features.

   That is, the features that cannot help the regression results enough can be set to a very small value potentially zero.

   An ordinary multivariate regression uses all the features available to it and will assign each one a coefficient of regression.

   In contrast, the LASSO regression attempts to add them one at a time and if the new feature does not improve the fit enough to out-way the penalty term by including that feature then it could not be added meaning as zero.

   Thus the power of regularization by applying the penalty term for the extra features is that it can automatically do the selection for us.

Thus the models are made sparse with few coefficients in this case of regularization since the process eliminates the coefficients when their values are equal to zero. That means LASSO regression works on an objective to minimize the following:

$$\mathbf{Xn\ i=1\ (yi - X\ j\ xij\beta j)\ 2 + \lambda\ X\ p\ j=1\ |\beta j\ |}$$

It sets the coefficient, which can be interpreted as min(sum of square residuals + λ |slope|), where, λ |slope| is penalty term.

1) **SVM Algorithm**: SVM is a classification based method or algorithm. There are some cases where we can use it for regression. However, there are rare cases of use in unsupervised learning as well. SVM in clustering is under research for the unsupervised learning aspect. Here, we use unlabeled data for SVM.

1) **Linear Regression**: Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

5. OBJECTIVE AND SCOPE

OBJECTIVE :

The forecasting is done for the three important variables of the disease for the coming 10 days:

1) The number 0f new confirmed cases.

2) The number of death cases.

3) The number of recoveries.

SCOPE :

The forecasting is done for the three important variables of the disease for the coming 10 days:

- The number 0f new confirmed cases.
- The number of death cases.
- The number of recoveries.

6. CONCLUSION

So in this study we have use a ML based prediction model that has been proposed for predicting the risk of covid 19 outbreak this system analyses dataset contain day to day actual data and makes prediction for coming days using ML algorithms

The result proves that ES performs best in the current forcasting domain given the nature and size of dataset LR and LASSO also perform well to some extend to predict death rate and confirm cases

REFERENCES

[1] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Computer Science, vol. 83, pp. 1064– 1069, 2016.

[2] A. Maqbool, "Coronavirus: why has the virus hit African Americans so hard?," BBC News, 2020. [Online]. Available:

[3] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus covid-19," Plos one, vol. 15, no. 3, p. e0231236, 2020.

[4] S. Almasy, H. Yan, and M. Holcombe, "Coronavirus pandemic hitting some African-American communities extremely hard," CNN Health, 2020. [Online]. Available: https://www.cnn.com/2020/04/06/health/uscoronavirus-updatesmonday/index.html. [Accessed: 13-Apr-2020].

[5] G. Grasselli, A. Pesenti, and M. Cecconi, "Critical care utilization for the covid-19 outbreak in lombardy, italy: early experience and forecast during an emergency response," Jama, 2020.

[6] WHO. Naming the coronavirus disease (covid-19) and the virus that causes it. [Online].

[7] Available: https://www.who.int/emergencies/diseases/novelcoronavirus-

[8] 2019/technicalguidance/naming-the-coronavirus-disease- (covid-2019)-and-the-virus-that-causesit

[9] World Health Organization, "Coronavirus disease (COVID-19) advice for the public: myth busters," World Health Organization, 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novelcoronavirus-2019/advice-forpublic/myth-busters. [Accessed: 13-Apr-2020].

[10] H. Yan, N. Chen, and D. Naresh, "What's spreading faster than coronavirus in the US? Racist assaults and ignorant attacks against Asians," CNN.com, 2020