

COVID-19 Future Forecasting Using Supervised Machine Learning Models

Dhahabiyya Asharef

School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University Pune, India
dhahabiyya[at]gmail.com

Abstract: *AI (ML) based estimating systems have demonstrated their importance to expect in perioperative results to further develop the decision making on the future course of activities. The ML models have for some time been utilized in numerous application spaces which required the distinguishing proof and prioritization of unfavorable variables for a danger. A few expectation techniques are in effect famously used to deal with estimating issues. This study shows the capacity of ML models to gauge the quantity of impending patients impacted by Coronavirus which is by and by considered as a possible danger to humankind. Specifically, four standard gauging models, like straight relapse (LR), least outright shrinkage and choice administrator (Tether), support vector machine (SVM), and outstanding smoothing (ES) have been utilized in this review to gauge the compromising elements of Coronavirus. Three kinds of expectations are made by every one of the models, like the quantity of recently contaminated cases, the quantity of passings, and the quantity of recuperations in the following 10 days. The outcomes created by the review demonstrates it a promising system to involve these strategies for the current situation of the Coronavirus pandemic. The outcomes demonstrate that the ES performs best among every one of the pre-owned models followed by LR and Tether which performs well in anticipating the new affirmed cases, demise rate as well as recuperation rate, while SVM performs ineffectively in all the forecast situations given the accessible dataset.*

Keyword: Exponential Smoothing, Least Absolute Shrinkage And Selection Operation, Linear Regression, Support Vector Machine

1. Introduction

AI (ML) has substantiated itself as a noticeable field of study throughout the last ten years by addressing numerous extremely perplexing and complex true issues. The application regions included practically every one of this present reality spaces like medical care, independent vehicle (AV), business applications, normal language handling (NLP), canny robots, gaming, environment demonstrating, voice, and picture handling. ML calculations' learning is normally founded on experimentation technique very inverse of ordinary calculations, which adheres to the programming directions in view of choice proclamations like if-else [1].

One of the main areas of ML is gauging [2], various standard ML calculations have been utilized around here to direct the future course of activities required in numerous application regions including climate estimating, infection determining, financial exchange anticipating as well as sickness forecast. Different relapse and neural organization models have wide relevance in anticipating the states of patients in the future with a particular illness [3]. There are heaps of studies performed for the expectation of various infections utilizing AI procedures, for example, coronary conduit sickness [4], cardiovascular illness forecast [5], and bosom disease forecast [6]. Specifically, the review [7] is centered around live anticipating of Coronavirus affirmed cases and study [8] is likewise centered around the conjecture of COVID19 flare-up and early reaction. These forecast frameworks can be exceptionally useful in decision making to deal with the current situation to direct early mediations to deal with these illnesses really. This study expects to give an early figure model to the spread of novel Covid, otherwise called SARS-CoV-2, authoritatively named as Coronavirus by the World Wellbeing Association (WHO) [9].

Coronavirus is as of now an intense danger to human existence everywhere. Toward the finish of 2019, the infection was first distinguished in a city of China called Wuhan, when an enormous number of individuals created side effects like pneumonia [10]. It diversely affects the human body, including serious intense respiratory disorder and multi-organ disappointment which can at last prompt demise in an extremely brief term [11]. A huge number of individuals are impacted by this pandemic all through the world with great many passing each approaching day. Huge number of new individuals are accounted for to be positive consistently from nations across the world. The infection spreads basically through close individual to individual actual contacts, by respiratory drops, or by contacting the debased surfaces. The most difficult part of its spread is that an individual can have the infection for a long time without showing side effects. Because of the causes of its spread and the threat it poses, practically every country has imposed partial or complete lockdowns in the impacted areas and cities.

Medical experts from all across the world are working to find an effective vaccination and treatment for the illness. Because there is currently no licenced therapy to kill the virus, governments from all around the world are working on preventative measures to halt its spread.

Among all the safeguards, "being informed" on all elements of COVID-19 is seen as crucial. Numerous researchers are examining the various features of the pandemic and producing conclusions to aid mankind to contribute to this component of knowledge. The goal of this work is to construct a COVID-19 forecasting system in order to help with the present humanitarian situation. The three major elements of the disease over the next 10 days have destroyed the forecasting: 1) the number of new instances that have been confirmed. 2) the number of incidents of death 3) the total number of recoveries. This issue of estimating has been

considered as a relapse issue in this review, so the review depends on some condition-of-craftsmanship regulated ML relapse models like direct relapse (LR), least outright shrinkage and determination administrator (Tether), support vector machine (SVM), and dramatic smoothing (ES). The learning models have been prepared utilizing the Coronavirus patient details dataset given by Johns Hopkins.

The dataset has been preprocessed and isolated into two subsets: preparing set (85% records) and testing set (15% records). The exhibition assessment has been done as far as significant measures including R-squared score (R2 score), Changed R-squared Score (R2 changed), mean square blunder (MSE), mean outright mistake (MAE), and root mean square blunder (RMSE). This study has a few key discoveries which are recorded beneath: • ES performs best when the time-series dataset has extremely restricted sections. • Different ML calculations appear to perform better in various class forecasts. • The majority of the ML calculations require a sufficient measure of information to foresee the future, as the size of the dataset builds the model exhibitions get to the next level. • ML model based anticipating can be exceptionally valuable for leaders to contain pandemics like Coronavirus.

1.1 Problem Definition

The current techniques utilizing most models expect a standard SEIR structure. Fraser and colleagues to assess size however roll out various improvements on the idea of the various compartments and their separate home times.

1.2 Objective

AI (ML) based gauging components have demonstrated their importance to expect in perioperative results to further develop the decision making on the future course of activities. The ML models have for quite some time been utilized in numerous application spaces which required the ID and prioritization of unfriendly factors for a danger. A few expectation strategies are by and large prominently used to deal with determining issues. This study shows the capacity of ML models to estimate the quantity of impending patients impacted by Coronavirus which is as of now considered as a likely danger to humankind. Specifically, four standard estimating models, like straight relapse (LR), least outright shrinkage and determination administrator (Tether), support vector machine (SVM), and outstanding smoothing (ES) have been utilized in this review to conjecture the undermining elements of Coronavirus.

2. Literature Survey

[1] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PloS one*, vol.13, no.3, 2018.

AI (ML) strategies have been proposed in the scholastic writing as options in contrast to measurable ones for time series determining. However, sparse proof is accessible with regards to their general exhibition as far as precision and computational prerequisites. The motivation behind this paper is to assess such execution across various determining

skylines utilizing an enormous subset of 1045 month to month time series utilized in the M3 Rivalry. In the wake of contrasting the post-test precision of famous ML strategies with that of eight conventional factual ones, we observed that the previous are ruled across both exactness measures utilized and for all determining skylines analyzed. In addition, we saw that their computational prerequisites are significantly more prominent than those of measurable strategies. The paper talks about the outcomes, clarifies why the exactness of ML models is underneath that of factual ones and proposes a few potential ways forward. The exact outcomes found in our examination stress the requirement for evenhanded and unprejudiced ways of testing the exhibition of gauging strategies that can be accomplished through sizable and open rivalries permitting significant correlations and unmistakable ends.

[2] A. Kata, "A postmodern Pandora's box: Antivaccination misinformation on the Internet," *Vaccine*, vol.28, no.7, pp.1709–1716, Feb.2010, doi: 10.1016/j.vaccine.2009.12.022.

The Web assumes a huge part in spreading against immunization data. This paper expands upon past examination by breaking down the contentions proffered on enemy of immunization sites, deciding the degree of falsehood present, and analyzing talks used to help antibody complaints. Contentions around the topics of wellbeing and viability, elective medication, common freedoms, paranoid fears, and ethical quality were found on most of sites dissected; deception was likewise predominant. The most generally proposed method of battling this misinformation is through better schooling, albeit this has demonstrated inadequate. Training doesn't consider the talks supporting antibody dismissal, like those including elective informative models of wellbeing, translations of parental obligation, and doubt of skill. Hostile to immunization protestors suggest postmodern viewpoints that reject biomedical and logical "realities" for their own translations. Favorable to inoculation advocates who center around remedying deception diminish the discussion to simply an "instructive" issue; rather, these postmodern talks should be recognized to start an exchange. With bleakness and mortality from immunization preventable sicknesses [VPDs] having arrived at record lows [1], antibodies are one of the best apparatuses for biomedical science and general well being. However amazingly, the adequacy of immunization has prompted the reappearance of hostile to inoculation opinions. Antibodies might be viewed as superfluous or perilous in light of the fact that occurrence paces of VPDs in created nations have plunged. Antibody "responses"-negative wellbeing occasions following inoculation, ascribed to the immunization-then, at that point, give off an impression of being more normal than the actual illnesses [2]. Thusly, antibodies can be viewed as casualties of their own prosperity. The media assumes an enormous part in spreading and sensationalizing antibody protests. Such complaints are essential for what has been known as the "counter inoculation development", which obviously affects immunization strategies, and individual and local area wellbeing [3].

A typical succession to immunization alarms includes logical discussion about potential antibody hazards, which correspondence innovation sends through a manner of speaking of uncertainty; guardians fuse this with individual encounters and spread their perspectives to their gatherings [4]. These gatherings apply impressive strain on inoculation choices by making a "neighborhood immunization culture" [5]. With the noticeable quality of the Web in this day and age, Correspondence address: 110 Parkwood Cres., Hamilton, Ontario L8V 4Z7, Canada. Tel.: +1 905 387 3141. Email addresses: aniakata[at]gmail.com, kataa[at]mcmaster.ca. the perspectives, convictions, and encounters of that neighborhood culture can immediately become worldwide. Web utilization measurements show roughly 74% of Americans and 72% of Canadians are online [6]. An expected 75-80% of clients look for wellbeing data online [7]. Of these clients, 70% say the data they experience online impacts their treatment choices [8]. In 2006, 16% of clients looked online for data on inoculations or immunizations [9]. While online examination is more helpful and open than perusing clinical writing or visiting wellbeing specialists, too extraordinary a dependence on Web based data can be tricky. More than half (52%) of clients accept "practically all" or "most" data on wellbeing sites is valid [8]; yet the accessibility of mistaken and tricky data online has named the Web a "advanced Pandora's crate" [10]. The idea of the Web permits all possible feelings to spread broadly and promptly. People and gatherings gain openness online without being sifted or explored-and hostile to immunization advocates enjoy taken benefit of this reality. Against immunization messages are more normal on the Web than in different types of media, improving the probability that inoculation choices might be founded on deluding data [11]. To be sure, guardians who absolved kids from inoculation are bound to have acquired data from the Web than guardians who have their youngsters immunized; they are additionally bound to have utilized specific antivaccination sites [12]. This exhibits the significance of getting what messages are introduced on the web and why they might be acknowledged. The group of examination analyzing on the web against vaccinationism isn't huge, nor has there been a new update [11, 13-18]. Onlyone investigation [13] inspected falsehood and trickery on such locales, yet was not quantitative. Earlier examination likewise recognized the need to comprehend talks fundamental enemy of inoculation contentions [19, 20], yet didn't intricate upon them. This investigation plans to resolve these issues by responding to two primary inquiries. To begin with, what data is proffered on enemy of immunization sites, and what is its precision? Second, what talks make these immunization protests engaging? 2. Techniques 2.1. Information assortment Web look were directed on May 21, 2009 utilizing the expressions "antibody", "inoculation", and "vaccination OR inoculation" contribution to Google. com (the American form of the web search tool) and Google. ca (the Canadian rendition). Google was picked as it is the most famous web crawler, representing 73% of all Web look [21]. Results were named enemy of inoculation and included for content examination assuming they went against youth inoculation under any condition, without meeting any of the accompanying prohibition standards: (1) listserv or newsgroup pages; (2) pages exclusively containing brief notification about other

site content; (3) news results, clinical diaries or library locales; (4) video results; (5) book reviews; (6) non-English destinations; (7) locales solely about grown-up vaccination; (8) destinations only with regards to veterinary immunization and (9) inert connections. Rules (see Tables 1 and 2) were applied to the counter inoculation sites and coded as present or missing. Models were adjusted from past internet based antivaccination studies [11, 13, 14, 17, 18], as well as made by the creator. Online wellbeing data searchers analyze the initial 10 list items 97.2% of the time [22]; in this way, just the initial 10 outcomes recovered per term were inspected. Of 30 complete Google. com results, 5 of 21 inoculation locales (24%) were delegated enemy of immunization. Of 30 all out Google. ca results, 2 of 16 inoculation locales (13%) were named enemy of immunization. To store up extra sites for a more significant review, the Canadian hunts were reached out to 50 outcomes for every term. Of 150 complete outcomes, 5 of 86 inoculation destinations (6%) were named enemy of immunization (two were copies of American outcomes). Joining the American and Canadian outcomes, 8 enemy of inoculation sites were exposed to content examination. Reference section A rundowns the locales examined. The extents of supportive of and hostile to immunization destinations found per search term are delineated.

In general, American hunts returned more enemy of inoculation results (24%) than Canadian inquiries (6%), demonstrating American guardians are bound to experience hostile to immunization locales by means of Google than are Canadian guardians. Neither web crawler returned any enemy of inoculation results for "inoculation OR inoculation"; this was normal in light of examination that observed enemy of immunization bunches try not to utilize the expression "vaccination" as they tend not to accept that antibodies present resistance [16]. Albeit earlier examinations returned more list items [11], this doesn't really mean the quantity of against immunization sites has diminished, yet rather that their pursuit rankings might have moved. By the by, the extent of destinations recovered for some, search terms is remarkable-71% of results from the Google. com "inoculation" search were delegated antivaccination. 3.2. Content and topics Fig.2 represents the level of investigated destinations with the recorded subjects present. Individual substance models are measured in Table 1. Individual plan models are measured in Table 2.3.2.1. Continuity and security "Antibodies are organic poisons that are harmful to one's health and are a significant factor in juvenile illness. " (<http://www.vaclib.org/destinations/banter/about.html>) Security topics were available on all enemy of inoculation sites examined. Each site asserted antibodies are toxic and cause idiopathic diseases. Locales focused on that immunizations contain substances toxic to people, including radiator fluid, ether, formaldehyde, mercury, and nanobacteria. Appropriate data was not expounded upon-for example, that how much possibly destructive substances in immunizations isn't to the point of creating harmful results in people, or that ether doesn't allude to the sedative yet to a synthetic compound. Sicknesses ascribed to immunizations included.

[3] B. Martin, "Texas anti-vaxxers fear mandatory COVID-19 vaccines more than the virus itself," Texas Monthly,

2020. [Online]. Available: <https://www.texasmonthly.com/news/texas-antivaxxers-fear-mandatory-coronavirus-vaccines>

The antivaccine development utilizes Facebook to advance messages on the supposed risks and outcomes of inoculating, prompting a hesitance to vaccinate against preventable transferable infections. We might want to find out about the messages these sites are sharing by means of web-based media that can impact perusers and purchasers. What messages is people in general getting on Facebook about inoculation? What content (news stories, tributes, recordings, logical investigations) is being advanced?

The World Health Organization records antibody reluctance as one of 10 dangers to worldwide wellbeing.

We proposed utilizing an online media review instrument and 3 downright inclines to catch data on sites and posts, separately. The catchphrases "antibody," "immunization truth," and "against vax" were entered in the Facebook search bar. A Facebook page was analyzed assuming it had somewhere in the range of 2500 and 150,000 preferences. Information about convictions, invitations to take action, and tributes were recorded from posts and recorded under the classifications Fantasies, Facts, and Results. Site information were entered in an online media review format.

Clients' posts reflected dread and antibody reluctance coming about because of the supposed risks of inoculation included on the site joins. Immunizations were faulted for hardships like mental imbalance, malignant growth, and barrenness. Moms shared declarations on supposed outcomes their youngsters experienced because of inoculation, which have affected different guardians to not immunize their kids. Clients denied the current measles episodes in the US to be valid, fighting back against the public authority in fights for manufacturing news.

3. Methodology

3.1 Existing System

The current strategies utilizing most models accept a standard seir structure.

Fraser and colleagues to appraise size yet roll out various improvements on the idea of the various compartments and their separate home times.

3.2 Proposed System

Here we proposed an anticipating framework for Coronavirus. The three most important aspects of the infection over the next 10 days have been completed: 1) the number of new instances that have been confirmed, 2) the number of incidents of death, 3) the total number of recoveries. This issue of estimating has been considered as a relapse issue in this review, so the review depends on some condition-of-craftsmanship regulated ML relapse models like straight relapse (LR), least outright shrinkage and choice administrator (Tether), support vector machine (SVM), and dramatic smoothing (ES). The learning models

have been prepared utilizing the Coronavirus patient details dataset given by Johns Hopkins. The dataset has been preprocessed and isolated into two subsets: preparing set (85% records) and testing set (15% records). The presentation assessment has been done as far as significant measures including R-squared score (R² score), Changed R-squared Score (R² changed), mean square mistake (MSE), mean outright blunder (MAE), and root mean square mistake (RMSE).

Advantages

- When the time-series dataset includes a small number of entries, ES works well.
- Distinct machine learning algorithms appear to be better at predicting different classes.
- Most machine learning algorithms require a large quantity of data to forecast the future; as the dataset grows larger, the model's performance improves.
- For decision-makers battling pandemics like COVID-19, ML model-based forecasting can be quite valuable.

3.3 Software Requirement Specification

3.3.1 Hardware Requirements

Because this is a project, hardware is critical. Hardware selection is equally critical to the existence and functioning of any software. The most important criteria are size and capacity.

Operating System supported by

- 1) Windows 7
- 2) Windows XP
- 3) Windows 8

Processor – Pentium IV or higher

RAM--256 MB

3.3.2 Software Requirements

At the conclusion of the analytical work, the software requirements specification is created. The development of the application requires the usage of the following technologies and languages:

- 1) Python
- 2) Mysql
- 3) Mysql client
- 4) WampServer 2.4

Technologies and Languages utilised to Develop-Python

3.4 Purpose

The point of this study is the future determining of COVID19 spread zeroing in on the quantity of new certain cases, the quantity of passings, and the quantity of recuperations. The dataset utilized in the review has been acquired from the GitHub vault gave by the Middle to Frameworks Science and Designing, Johns Hopkins College [12]. The archive was essentially made accessible for the visual dashboard of 2019 Novel Covid by the college and was upheld by the ESRI Living Map book Group. Dataset records are contained in the organizer on the GitHub archive named (csse_covid_19_time_series). The envelope contains day by day time series synopsis tables, including the quantity of affirmed cases, passings, and recuperations. All

information are from the every day case report and the update recurrence of information is one day.

3.5 Scope

A huge number of individuals are impacted by this pandemic all through the world with great many passings each approaching day. Large number of new individuals are accounted for to be positive consistently from nations across the world. The infection spreads principally through close individual to individual actual contacts, by respiratory drops, or by contacting the sullied surfaces. The most difficult part of its spread is that an individual can have the infection for a long time without showing side effects. The reasons for its spread and thinking about its risk, practically every one of the nations have pronounced either incomplete or severe lockdowns all through the impacted areas and urban communities. Clinical scientists all through the globe are presently involved to find a fitting immunization and prescriptions for the illness. Since there is no supported medicine till now for killing the infection so the states of all nations are zeroing in on the precautionary measures which can stop the spread. Out of all insurances, "be educated" pretty much every one of the parts of Coronavirus is thought of as critical. To add to this part of data, various scientists are concentrating on the various components of the pandemic and produce the outcomes to help humankind.

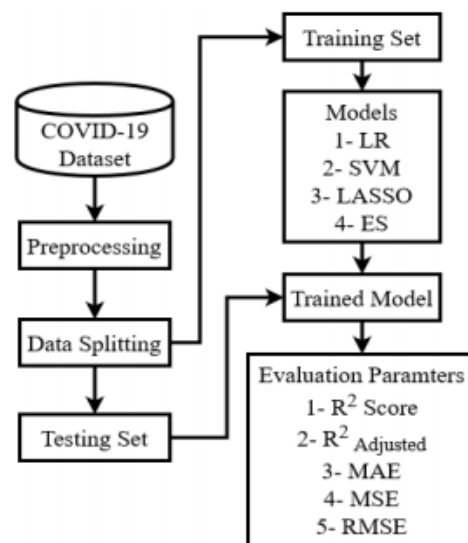
Some Facebook messages empower winning fantasies about the security and outcomes of immunizations and logical add to guardians' antibody aversion. Profoundly concerning is the question web-based media can possibly give occasion to feel qualms about the connection between medical services suppliers and people in general. A grip of normal misinterpretations can assist with supporting medical care supplier practice. Many sicknesses have been nearly, or totally, destroyed because of inoculation. Inoculation against infection forestalls 2-3 million passings each year universally and could forestall considerably more with worldwide immunization upgrades [1]. Vaccination has incomprehensibly diminished mortality because of preventable transferable sicknesses. For instance, before the presentation of the measles antibody, 300, 000-400, 000 Canadians were contaminated each year, for certain recuperations and numerous passings [2]. Since the end of measles in 1998 because of immunizations, there have been not many cases in Canada [2]. Additionally, when the polio antibody was presented in Canada during the 1950s, cases decreased significantly, and the current danger to the Canadian populace is incredibly low [3].

The World Health Organization (WHO) has announced antibody reluctance as one of the best 10 dangers to worldwide wellbeing [4]. Online media has helped fuel the development of the antivaccine development, with Facebook being recognized as a vital disseminator of deception encompassing the mission [5-7]. Facebook is the biggest online media stage, with multiple billion dynamic month to month clients [8]. There have been not kidding endeavors to decrease how much deception spread on the web-based media website by bringing down the positioning of Gatherings and Pages making bogus cases [7]. Online media managers have been encouraged to eliminate these Pages

and Gatherings through and through; be that as it may, counterarguments refer to an infringement of common liberties to get to uncensored data [7]. This paper uncovered the messages of the antivaccine development on the web and how people see inoculation. We meant to uncover the legends and facts that clients of Facebook Pages notice and participate in. Medical services customers and medical services suppliers might wind up on far edges of the discussion. Absence of vaccination puts general society in danger and diminishes general wellbeing endeavors to control measles and polio and forestall episodes of flu (influenza) alongside other transferable illnesses. The change in power among specialists and patients because of simple admittance to data online has prompted the scrutinizing of medical services suppliers and expanded shared decision making [6, 9].

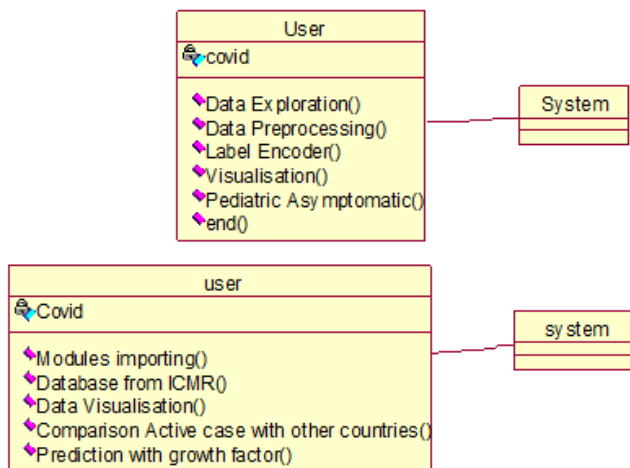
As the vast majority of the world anticipates an immunization to stop the Coronavirus (otherwise called the 2019 novel Covid) pandemic, "devotees" of antivaccine Facebook Pages appear to fear the antibody more than the actual infection [10]. In the midst of the Coronavirus pandemic, online media destinations, for example, Facebook can't handle the wellbeing deception that is spread on its Pages [11]. Antivaccine Pages have been giving paranoid fears, security concerns, and elective wellbeing medicine that grip the consideration of "unsure" people riding the web for data on immunizations [11]. The WHO is battling to stop the spread of deception online by teaming up with web-based media monsters to figure out how to control bogus cases [12]. A few instances of such cases incorporate that Coronavirus is a bioweapon supported by the Bill and Melinda Doors Establishment or that it can just be restored by polishing off natively constructed mixtures (some incorporate drinking blanch) [13]. Our point was to find out about the messages that can impact perusers and buyers that these sites are sharing by means of online media.

3.6 Overall Description

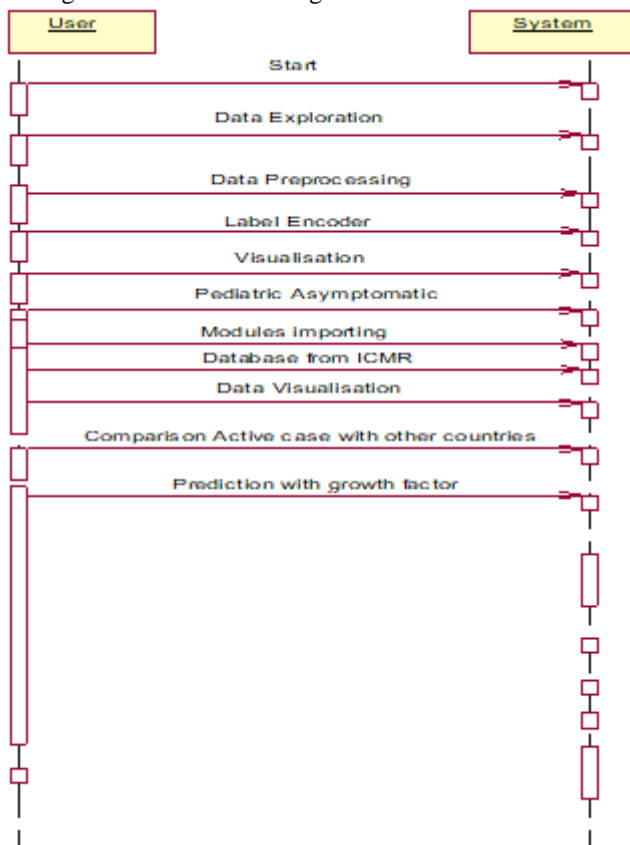


3.7 Design

In this there are three classes: data owner, data user and cloud server. Cloud activates owner and user. Data owner uploads files and data user view and download files.

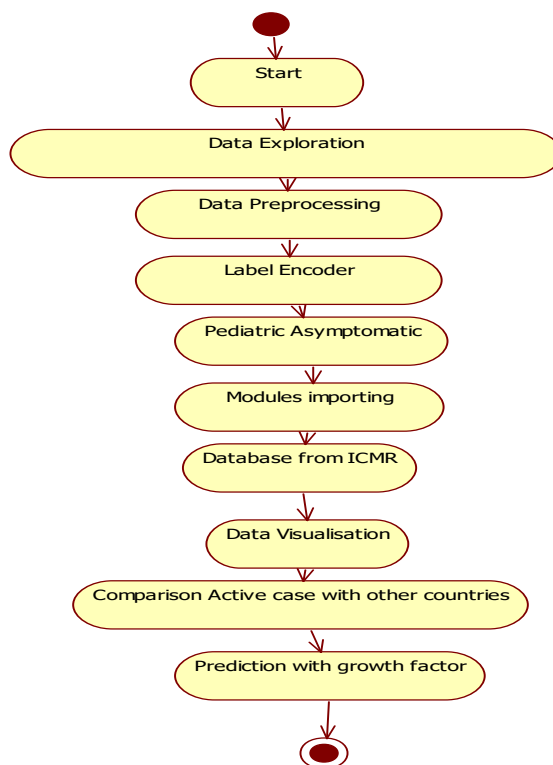


Interaction diagrams are shown below the diagrams, and they show how operations are carried out. They capture the interaction between things in a collaborative setting. They are time-focused, and they use the vertical axis of the diagram to indicate time, indicating when messages are delivered and when they are received. The diagrams are time focused and depict the interaction order; they not only demonstrate interaction but also some control over the sequences. There's also a time line to illustrate which half is sending and which is receiving.



The activity diagram is a crucial diagram in UML for describing the system's dynamic characteristics. An activity diagram is a flowchart that depicts the movement of information from one action to the next. The action can be defined as a system operation. From one action to the next, the control flow is depicted. An activity diagram is a flowchart that depicts the movement of information from one action to the next. The activity diagram has a start point, which is at the top of the diagram and serves as the

diagram's starting point, as well as an end point, which is at the bottom of the diagram and serves as the diagram's ending point. All of the activities are linked to the beginning and ending points. The activity diagram is essentially a flowchart that depicts the fields' activity flow. The action can be defined as a system operation.



4. Algorithms

A method or set of rules to be followed in computations or other problem-solving activities, " according to the definition of the word algorithm. As a consequence, an algorithm is a series of rules/instructions that outline how a work should be carried out step-by-step in order to get the desired outcomes.

The different algorithm used here are:

4.1 Linear Regression

The most fundamental type of regression analysis is linear regression. It presupposes that the dependent variable and the predictor have a linear relationship (s). We strive to find the best fit line in regression to explain the connection between the predictors and the predictive/dependent variable.

Because of its straightforward representation, linear regression is a popular model.

The representation is a linear equation that combines a collection of input values (x), with the solution being the projected output for that set of input values (y). As a result, both the input (x) and output (y) values are numeric.

Each input value or column is assigned one scale factor, referred to as a coefficient and denoted by the capital Greek letter Beta in the linear equation (B). One more coefficient is

added, which gives the line an extra degree of freedom (for example, going up and down on a two-dimensional plot) and is known as the intercept or bias coefficient.

For instance, in a basic regression situation with only one x and one y , the model would be $Y = 0 + 1x$.

When there are several inputs (x) in higher dimensions, the line is termed a plane or a hyper-plane. As a result, the representation takes the form of an equation with precise values for the coefficients.

4.2 LASSO Regression

Lasso regression technique is a variation of the commonly used linear regression algorithm. It improves normal linear regression by significantly altering its cost function, resulting in models that are less overfit. Although lasso regression and ridge regression are extremely similar, there are several fundamental distinctions between the two that you must grasp if you wish to utilise them successfully. This article will teach you all you need to know about lasso regression, including the distinctions between lasso and ridge and how to use it in your own machine learning projects.

The dataset and problem statement must be considered before selecting a model. Understanding the dataset and how characteristics interact with one another is critical. Less significant characteristics of your dataset are penalised using Lasso regression, which makes their coefficients zero and so eliminates them. As a result, you have the benefit of feature selection and model construction that is straightforward. As a result, lasso regression can be utilised if the dataset has a high dimensionality and strong correlation.

4.3 Logistic Regression

The supervised learning classification method logistic regression is used to predict the likelihood of a target variable. Because the nature of the goal or dependent variable is dichotomous, there are only two classifications. In basic terms, the dependent variable is binary in nature, with data represented as 1 (representing success/yes) or 0 (representing failure/no). A logistic regression model predicts $P(Y=1)$ as a function of X mathematically. It's one of the most basic machine learning algorithms, and it may be used to solve a variety of classification issues including spam detection, diabetes prediction, cancer diagnosis, and so on.

4.3.1 Logistic Regression Types

In general, logistic regression refers to binary logistic regression with binary target variables, but it may also predict two additional types of target variables. Logistic regression may be classified into the following types based on the number of categories.

4.3.1.1 Binomial or binary

In this method of categorization, a dependent variable can only be one of two types: 1 or 0. These variables might, for example, indicate success or failure, yes or no, victory or loss, and so on.

4.3.1.2 Multinomial

The dependent variable might have three or more alternative unordered categories or types with no quantitative significance in this sort of categorization. These variables may, for example, represent "Type A," "Type B," or "Type C."

4.3.1.3 Ordinal

In this sort of categorization, the dependent variable can have three or more potential ordered categories or quantitatively significant types. For example, these variables may indicate "bad" or "good," "very good," or "excellent," with scores ranging from 0 to 3.

Algorithms are required for Machine Learning models to function. A specific ML model may perform better with one or the other approach depending on the set of circumstances. As a result, Machine Learning engineers and fans should be informed of the many types of algorithms that may be employed in various settings so that they can choose the best one for the job. In Machine Learning, there is no such thing as a one-size-fits-all solution, and experimenting with multiple algorithms might yield the required results.

You must, for example, be familiar with Linear Regression. This approach, however, cannot be used with categorically dependent variables. This is where Logistic Regression is useful. Logistic Regression is a supervised method of learning used in Machine Learning to predict the likelihood of a dependent or target variable. You may anticipate and construct correlations between dependent and one or more independent variables using Logistic Regression. Predictive analytics for binary classification often use logistic regression equations and models. You may also use them to classify several classes.

4.4 LGBM Regression

A large number of Data Enthusiasts compete in a variety of online Machine Learning competitions. Everyone has their own distinct method for determining the appropriate model and accurately predicting the result of a particular problem statement. Feature Engineering is an important aspect of the Machine Learning process that takes up the majority of the time. Modeling, on the other hand, plays a significant element when there isn't a lot of preprocessing or particular limits on the characteristics. There are a variety of ensemble approaches that can assist in the creation of powerful, resilient models that can provide extremely precise forecasts. But what precisely is the buzz about the word 'Ensemble'? Let's have a look at what the word Ensemble entails.

5. Conclusion and Future Work

The precariousness of the COVID-19 pandemic can ignite a massive global crisis. Some researchers and government agencies throughout the world have apprehensions that the pandemic can affect a large proportion of the world population [26], [27]. In this study, an ML-based prediction system has been proposed for predicting the risk of COVID19 outbreak globally. The system analyses dataset containing the day-wise actual past data and makes predictions for upcoming days using machine learning

algorithms. The results of the study prove that ES performs best in the current forecasting domain given the nature and size of the dataset. LR and LASSO also perform well for forecasting to some extent to predict death rate and confirm cases. According to the results of these two models, the death rates will increase in upcoming days, and recoveries rate will be slowed down. SVM produces poor results in all scenarios because of the ups and downs in the dataset values. It was very difficult to put an accurate hyperplane between the given values of the dataset. Overall we conclude that model predictions according to the current scenario are correct.

References

- [1] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PloS one*, vol.13, no.3, 2018.
- [2] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, "Machine learning strategies for time series forecasting," in *European business intelligence summer school*. Springer, 2012, pp.62–77.
- [3] F. E. Harrell Jr, K. L. Lee, D. B. Matchar, and T. A. Reichert, "Regression models for prognostic prediction: advantages, problems, and suggested solutions." *Cancer treatment reports*, vol.69, no.10, pp.1071–1077, 1985.
- [4] P. Lapuerta, S. P. Azen, and L. LaBree, "Use of neural networks in predicting the risk of coronary artery disease," *Computers and Biomedical Research*, vol.28, no.1, pp.38–52, 1995.
- [5] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel, "Cardiovascular disease risk profiles," *American heart journal*, vol.121, no.1, pp.293–298, 1991.
- [6] 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.
- [7] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus covid-19," *Plos one*, vol.15, no.3, p. e0231236, 2020.
- [8] 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.
- [9] WHO. Naming the coronavirus disease (covid-19) and the virus that causes it. [Online]. Available: [https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novelcoronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
- [10] C. P. E. R. E. Novel et al., "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19) in china," *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi*, vol.41, no.2, p.145, 2020.