**Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **26-09-2023** |
| **Team ID** | **687** |
| **Project Name** | **AI Driven Exploration and prediction of company Registration Trends with registrar of companies (ROC)** |

# A Novel Prediction Method Based on Artificial Intelligence and Internet of Things for Detecting Coronavirus Disease (COVID-19)

**Problem Definition and Design Thinking**

**INTRODUCTION**

Mathematical models are often used by researchers to derive the nonspreading conditions of infectious diseases and predict and analyse the trends of epidemics and infected populations. Correspondingly, relevant strategies are developed accordingly. One of the currently used epidemic prediction models is the Malthusian growth model. However, this model still has a long way to go before being applied in the real world. A logistic regression , also known as the SI model to predict when the cured population without immunity is more vulnerable to reinfections. On the contrary, the classic SIR model is commonly selected for a cured population with strong immunity. This classic SIR model is widely used to describe the overall trends of epidemics owing to its ease of operation and clear and concise structure. This classic SIR model was used, for example, to analyse the 2003 SARS epidemic. The epidemic’s evolution and the overall spread patterns of the disease are described. Based on this, the SEIR model introduces the exposed, that is, class *E* population, which considers that only part of the people who are easily infected and had contact with infected people are infectious, which makes the transmission cycle of the disease longer. However, more detailed factors are not considered in the process of epidemic prediction.

**Problem Statement**

In light of the spread of novel coronaviruses, we develop one big data prediction model of novel coronavirus epidemic in the context of intelligent medical treatment, taking into account all factors of infection and death and implementing emerging technologies, such as the Internet of Things (IoT) and machine learning. Based on the different application characteristics of various machine learning algorithms in the medical field, we propose one artificial intelligence prediction model based on random forest. Considering the loose coupling between the data preparation stage and the model training stage, such as data collection and data cleaning in the early stage, we adopt the IoT platform technology to integrate the data collection, data cleaning, machine learning training model, and front- and back-end frameworks to ensure the tight coupling of each module. To validate the proposed prediction model, we perform the evaluation work. In addition, the performance of the prediction model is analyzed to ensure the information accuracy of the prediction platform.

**Key Challenges:**

At the same time, foreign scholars have conducted much research on the epidemic situation. The main research aim and findings of this paper were as follows:(1)This paper aimed to examine and compare existing machine learning algorithms in the medical field and find the best algorithm model based on AI for creating a prediction platform to prevent and control epidemics.(2)This paper aimed to combine machine learning with IoT, a platform of the IoT for the prevention and control of COVID-19, based on the traditional epidemic model.(3)This paper aimed to introduce the designed COVID-19 prediction platform in the order of data collection, data cleaning, machine learning training model, and front- and back-end frameworks.(4)The simulation results show that the predictions made by the AI designed in this paper’s random forest model are more accurate than those

**ALGORITHM AND MODEL ANALYSIS**

##### **Logistic Regression**

Logistic regression is a classification algorithm commonly used to solve binary classification problems. This algorithm has been extensively studied in the industrial and medical fields because of not only its simplicity but also its strong interpretability. The essence of logistic regression is to use maximum likelihood estimation to approximate the parameters of a given .

To date, logistic regression has been applied in many fields, in which there are many medical scenarios. Unfortunately, this algorithm cannot solve the nonlinear problems in the medical field, although it has broad prospects for health care and has been greatly studied in disease diagnosis. Typically, epidemic predictions are not linear, thus limiting the use of the logistic regression algorithm in epidemic prediction. Due to the simplicity of its form, the accuracy rate cannot be guaranteed, which makes fitting true data distribution using the logistic regression algorithm difficult.

###### **Support Vector Machines**

Before the widespread application of deep learning algorithms as machine learning algorithms, SVMs were considered the optimal method for small-sample classification problems. SVM is a nonclustering technique that can calculate distances outside of the plane. When the specific parameters of a given model are assigned by a training set, SVM classification tasks only respond to the support vectors and have no relation with the data dimensions. The computational time and storage requirements are reduced as a subset of the sample . SVM is a classifier, and the maximum intervals between different cases are its primary indicators. The hyperplane positions can be used to obtain constraints. In other words, SVM is mainly used to ensure the correct classification of two data types in the future. However, the constraint requires a maximum distance between the classification line and that point within the maximum acceptable error. Furthermore, SVM is a kernel-based technique that allows for higher-dimensional space conversion using kernel functions. The structure of SVM hyperplane solutions allows them to solve quadratic programming problems while meeting the requirements of duality and convexity. In fact, for solving the convex optimization problem, the optimal plane can be determined according to strong duality

**Objectives:**

Internet of Things (IoT) and machine learning. Based on the different application characteristics of various machine learning algorithms in the medical field, we propose one artificial intelligence prediction model based on random forest. Considering the loose coupling between the data preparation stage and the model training stage, such as data collection and data cleaning in the early stage, we adopt the IoT platform technology to integrate the data collection, data cleaning, machine learning training model, and front- and back-end frameworks to ensure the tight coupling of each module. To validate the proposed prediction model, we perform the evaluation work.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for house price prediction.

**Actions:**

The training models are created using the logistic regression algorithm, SVM, and random forest. Both the training and test sets have parameter passing settings [15, 16]. The primary goal of machine learning is to obtain a prediction model by mining the inherent patterns in historical training data and then applying the model to similar data situations [17, 18]. The general workflow diagram is presented in Figure

**Prototype**

The foundation of random forest is Bootstrap. That is, many new samples of the same size and usability are generated from a sample, and similar samples are generated again from those that have already been created. Bootstrap is also known as a self-help method as it does not use any other sample data . When the sample size is small, this method is considered useful. If the traditional method is used for verifications and segmentations, the sample size will be even smaller, resulting in a larger deviation and a nonoptimal solution. The self-service method not only fails to reduce the training sample size but also leaves.

**Iterate**

Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

**Actions:**

- Monitor the model's performance and retrain it periodically with updated data.

- Address user feedback and make necessary improvements to the web interface.

- Stay informed about advancements in machine learning and real estate pricing models for potential enhancements.

**Conclusion**

the COVID-19 prediction algorithms based on artificial intelligence were compared. Based on considerations of various characteristic constraints and prediction result accuracies, a prediction platform was established. Through simulation, it was discovered that random forest has significant advantages in epidemic prediction over logistic regression and the support vector machine. It would perform admirably when applied to the medical platform designed in this paper. Simultaneously, Singh proposed using the unmanned aerial vehicle based on blockchain to achieve contactless transmission in the COVID-19 environment. Similar application scenarios such as will be the next development direction of the platform and strived to support more application development in the epidemic environment based on prediction analysis .