A Regression Model Frame with IoT to Predict COVID Zone

Subhajit Pati¹, Sourav Mallick², Snehashis Chakraborty³, Rajbinder Kaur⁴, Moloy Dhar⁵, Amrut Ranjan Jena⁶

¹CSE Dept, Guru Nanak Institute of Technology, India, subhajitpati2000@gmail.com
² CSE Dept, Guru Nanak Institute of Technology, India, souravmallicks286@gmail.com
³ CSE Dept, Guru Nanak Institute of Technology, India, snehashischakraborty1998@gmail.com
⁴ CSE Dept, Guru Nanak Institute of Technology, India, rajbinder0013@gmail.com
⁵ CSE Dept, Guru Nanak Institute of Technology, India, moloy.dhar@gnit.ac.in
⁶ CSE Dept, Guru Nanak Institute of Technology, India, amrutranjan.jena@gnit.ac.in

ABSTRACT:

In this paper, we are planning to build an application through which victims of Covid-19 can put their data. Once an effective data set is built, the application is capable to predict a COVID-19 zone based on these data by using a regression model. The parameters of the data set will be:

1. How many times a person visited outside.
2. City of the person, who visited abroad recently if yes then which country (take the ratio of victims to their population density as a parameter), if no then we will put a zero.
3. The cases in their home town by simply accessing their location (take the ratio of victims to the population density of that area).
4. Once data set is built, then train a regression model to predict the chances of a person being a victim of Covid-19 and once that information is available we can simply observe in which area the no of suspected no of people is greater through which we can tell the chance of an area being affected by this situation.
5. Although parameters of the data set are limited now but we are planning to add more so that more accurate models can be built.

Keywords: Covid-19, data set, IoT, machine learning, regression model

I. INTRODUCTION:

The world has been facing threats within the sort of pandemics periodically over the centuries. The aftermath of these pandemics has always had a huge impact on the world and has also turned the tables
over. COVID-19, the present devastating pandemic is additionally running its course currently within the world. Not only economies are crashing but the general strengths and morals of the heavily impacted nations are being compromised. Based on current statistics, the SARS-CoV-2 virus which, causes the disease COVID-19, has an R0 of two .2 which suggests one infected person will infect 2.2 others on the average. In the absence of any curative drug, the sole solution is to hamper the spread by exercising “social distancing” to dam the chain of the spread of the virus. This behavior of CoV-2 requires developing a robust mathematical basis for tracking its spread and automation of the tracking tools for online dynamic deciding. With the advent of data science and advanced analytics, we are also now able to access tons of data, collected from the various facets and phases of the virus and the disease. Multiple sources of data can be very helpful in analyzing the growth of infection with community behavior. Wrapping this data with Machine Learning (ML) and Artificial Intelligence (AI), it can be forecasted where and when, the disease is likely to spread, and notify those regions to match the required arrangements. Contact tracing is imperative to make sure that anyone who has close contact with the infected patients is traced, isolated, and monitored for symptoms.

II. Methodology:

Following steps, we have used to create this application:

1. Almost all work has been done on ANDROID STUDIO. Almost all features got in it.
2. For store Sign up details of any authenticate user during log in whether the user giving the correct email and password, we have used FIREBASE. Firebase has become a feature of Android in the updated version.
3. Firebase has two features called ‘FIREBASE AUTHENTICATION’ and ‘REAL TIMEDATA BASE’. We have connected our android project with these two features.
4. After that ‘sign up’ and ‘log in’ both options were created. All the details provided by the user are being stored in the authentication part. When new users sign up their details such as user id, email that are being stored in there. After running function and method of on click, it is now possible to go to the next activity.
   i. Some dependencies have been added in BUILD GRADEL (both app and project) of Android Studio that we have collected from internet.
   ii. After adding, find a website.
   iii. Minimizing a website in Desktop, this site automatically adjust itself as minimize screen.
   iv. Searched for a website, where we can get all the details of covid-19 i.e. common cold, flu and then copied the URL of the website by WEBSCAPPING in android studio.
      1. As a result, changes of the site in the future (if any) will be directly reflected in our app. To use GPS, Google API has been generated. To apply GPS, taken a new key from find ‘Places API’ named website.
      2. After enabling, we have got a key from there. As a result, we are able to track the GPS, nearly hospital location. So far, we have finished the work.

Apart from that, User name, how many times user gone out of house all this can be done in Firebase. All the details of user can be stored in real time database in firebase. In the place of value storing,
Ratio in and Ratio out, we have to apply DATA SCRIBBING. Also, we have some Machine Learning work left.

Healthcare delivery requires the support of new technologies like Artificial Intelligence (AI), Internet of Things (IoT), Big Data and Machine Learning to fight and look ahead against the new diseases [1]. We aim to review the role of AI as a decisive technology to analyze, prepare us for prevention and fight with COVID-19 (Coronavirus) and other pandemics [2].

Rahmatizadeh, Valizadeh-Haghi has proposed an AI-based management model for the critical care of COVID-19 patients. Vaishya, Javaid has compared non-AI-based treatment systems with AI-based systems and recommended that an AI-based system reduces the time and processes in the treatment of a patient and can also help with the control of disease [3]. Additionally, the AI-based system can identify the level of infection in a location which can assist with contact tracing [4]. Additionally, the AI-based system can identify the level of infection in a location which can assist with contact tracing.

Additionally, an LSTM and RNN model was used to predict the epidemic of COVID-19 [5]. The AI applications' interventions for the COVID-19 pandemic include the early "detection and diagnosis of the infection", prevention of disease, monitoring the treatment, contact tracing of the individuals, drugs and vaccines development, reducing the workload of healthcare workers and projections of cases and mortality [6] [7].

![Diagram](image)

**Figure 1**

### III. RESULT ANALYSIS:

This app is used for keeping only one thing that people should get the most out of it during this global pandemic. The objectives of this app are:

1) *To predict the chance of a person being affected by the corona virus so that we should get aware of it and should take all the precautions that one should take in order to save them and the whole world from the virus.*
2) *Help to gather adequate knowledge about common cold, flu and covid virus as all of them have some common symptoms.*
3) *With the help of this app, it will find all the nearby hospital, isolation centers and doctors within a range of 2 Km radius and also the contact details of the doctors, that will be provided to us in case if any person want to contact.*
4) Get all the updates regarding this pandemic.

TRAINING THE CLASSIFIER MODEL:

```python
# [39]:     AGE  Outings  CITY  RatioIn  RatioOut  Outcome
0   2.969732    2   10   -0.916291  0.207014   H
1   2.944439    3   11  -2.120264 -1.609438   L
2   3.044522    9   12  -1.427118 -1.968113   L
3   3.737670   10   13  -0.916291  -2.302585   H
4   3.891820    6   14  -0.356675  -1.021651   H

# [41]: dataset.shape
(349, 6)

Performing train test split

```python
# [155]: from sklearn.model_selection import train_test_split
from sklearn.utils import shuffle
from sklearn import preprocessing

X_dataset = shuffle(dataset)
X = dataset.drop(['Outcome'], axis=1)
y = dataset['Outcome']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,
                                                  random_state=42)
# X_train=preprocessing.scale(X_train)
# X_test=preprocessing.scale(X_test)

0.11 Model building (Logistic Regression)

```python
# [156]: from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
# fitting model
model.fit(X_train, y_train)
# predictions
y_pred = model.predict(X_test)
```
0.12 Accuracy checking

```python
from sklearn.metrics import confusion_matrix, accuracy_score
print(confusion_matrix(y_test, y_pred))
print(f'accuracy is {accuracy_score(y_test, y_pred)*100}%')
```

```
[[36  9]
 [ 6 37]]
accuracy is 82.95454545454545
```

0.13 Performing PCA for better accuracy

```python
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
scaler.fit(X)
scaled_data=scaler.transform(X)
from sklearn.decomposition import PCA
pca=PCA(n_components=4)
pca.fit(scaled_data)
x_pca=pca.transform(scaled_data)
X=dataset.drop(['Outcome'], axis=1)
y=dataset['Outcome']
X_train, X_test, y_train, y_test = train_test_split(x_pca, y, test_size=0.25, ...
random_state=42)
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
#fitting model
model.fit(X_train,y_train)
#predictions
y_pred=model.predict(X_test)
from sklearn.metrics import confusion_matrix, accuracy_score
print(confusion_matrix(y_test, y_pred))
print(f'accuracy is {accuracy_score(y_test, y_pred)*100}%')
```
In the result analysis part, we have applied four different Machine Learning techniques such as: (i) Logistic Regression (LR), (ii) Principal Component Analysis (PCA), (iii) Random Forest (RF) and (iv) K-nearest neighbors’ algorithm (KNN). The result obtained by these techniques is expressed in the form of

\[
\begin{bmatrix}
35 & 10 \\
7 & 36
\end{bmatrix}
\]

accuracy is 80.681818181817

0.14 Model building (Random Forest)

```python
[177]: from sklearn.ensemble import RandomForestClassifier
model=RandomForestClassifier()
model.fit(X_train,y_train)
y_pred=model.predict(X_test)
from sklearn.metrics import confusion_matrix,accuracy_score
print(confusion_matrix(y_test,y_pred))
print(f" accuracy is \{accuracy_score(y_test,y_pred)*100\}%")
```

\[
\begin{bmatrix}
31 & 14 \\
13 & 30
\end{bmatrix}
\]

accuracy is 69.318181818183

0.15 Model building (KNN)

```python
[187]: from sklearn.neighbors import KNeighborsClassifier
model=KNeighborsClassifier(n_neighbors=13)
model.fit(X_train,y_train)
y_pred=model.predict(X_test)
from sklearn.metrics import confusion_matrix,accuracy_score
print(confusion_matrix(y_test,y_pred))
print(f" accuracy is \{accuracy_score(y_test,y_pred)*100\}%")
```

\[
\begin{bmatrix}
30 & 15 \\
17 & 26
\end{bmatrix}
\]

accuracy is 63.63636363636363
accuracy. The accuracy of LR, PCA, RF, and KNN is given by 82.955\%, 80.681\%, 69.318\% and 63.636\% respectively. LR has given a better accuracy as compared to other machine learning techniques.

The sample for research work is given in Table-1.

| SL.NO | NAME   | AGE | TIMES OF OUTING | CITY        | COVID RATIO IN | COVID RATIO OUT |
|-------|--------|-----|----------------|-------------|----------------|----------------|
| 1     | RAKESH | 45  | 6              | BANGALORE   | 0.75           | 0.25           |
| 2     | SUVENDU| 36  | 3              | HYDERABAD   | 0.65           | 0.66           |
| 3     | ASISH  | 54  | 4              | CHENNAI     | 0.55           | 0.38           |
| 4     | AMIT   | 45  | 6              | BANGALORE   | 0.75           | 0.25           |
| 5     | YOGESH | 49  | 9              | AGRA        | 0.8            | 0.3            |
| 6     | ANKITA | 45  | 4              | AJMER       | 0.65           | 0.56           |
| 7     | POULAMI| 18  | 6              | NADIA       | 0.5            | 0.6            |
| 8     | TAPOS  | 32  | 7              | KOLKATA     | 0.65           | 0.23           |
| 9     | MANALI | 27  | 15             | DELHI       | 0.5            | 0.6            |
| 10    | SNEHA  | 44  | 0              | KANPORINDORE| 0.65           | 0.98           |

The graphical representation of the analysis is described as follows:

![Figure 2](image)

Figure 2: Plotted the correlation of each continuous variable in the form of heatmap so that we can analyze how each variable is correlated with each other in order to remove those features having less correlation.
Figure 3: Plotted a count plot graph in order to check whether our dataset is balanced or imbalanced.

Figure 4: Pair plot for multivariable analysis between the features and it is done only on continuous variables.
From figure 5 to 8, we have used dist plot in order to see what type of distribution the continuous variables follow i.e. (SND, Gaussian Distribution). If the distribution is skewed, using Log Normalization techniques to convert them into SND.
In figure 8 to 10, we have used Box Plot in order to check whether there are some outliers present in our dataset to take some measures.

IV. CONCLUSION

In this pandemic period, people should very much careful about their health. Therefore, it is required to identify the COVID zones, so that they can move carefully in their surroundings. In this context, we have proposed a regression model to predict the COVID zone as per the travelling history of different persons of that locality. We have compared four machine learning techniques to test proposed model. However, the LR technique gives better accuracy of 82.955% as compared to other three techniques.

References

[1] Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental, 2(1), 35.

[2] Pannu, A. (2015). Artificial intelligence and its application in different areas. Artificial Intelligence, 4(10), 79-84.

[3] Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews.

[4] Jelodar, H., Wang, Y., Orji, R., & Huang, H. (2020). Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. arXiv preprintarXiv:2004.11695.

[5] Guescini, M., Sisti, D., Rocchi, M. B., Stocchi, L., & Stocchi, V. (2008). A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics, 9, 326, doi:10.1186/1471-2105-9-326.
[6] Peeri, N. C., Shrestha, N., Rahman, M. S., Zaki, R., Tan, Z., Bibi, S., et al. (2020). The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? Int J Epidemiol, doi:10.1093/ije/dyaa033.

[7] Gandon, S., Day, T., Metcalf, C. J. E., & Grenfell, B. T. (2016). Forecasting Epidemiological and Evolutionary Dynamics of Infectious Diseases. Trends Ecol Evol, 31(10), 776-788, doi:10.1016/j.tree.2016.07.010.