Contact tracing of covid 19 virus

A PROJECT REPORT

for

Artificial Intelligence

in

B.Tech (IT)

by

Bobbala Mahith Kumar Reddy (18BIT0076)

7th Sem, Year

Under the Guidance of

Dr. R.Subhashini

SITE

School of Information Technology and Engineering

NOV, 2021

DECLARATION BY THE CANDIDATE

I here by declare that the project report entitled **"CONTACT TRACING OF COVID 19 VIRUS"** submitted by me to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Artificial Intelligence (ITE2010)** is a record of bonafide project work carried out by me under the guidance of DR**. R SUBHASHINI.** I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other course.

Place : Vellore   Signature

Date : 29/nov/2021   Bobbala Mahith Kumar Reddy.

School of Information Technology & Engineering [SITE]

CERTIFICATE

This is to certify that the project report entitled **"Contact tracing of covid 19 virus"** submitted by **Bobbala Mahith Kumar Reddy (18BIT0076)** to Vellore Institute of

Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Artificial Intelligence (ITE2010)** is a record of bonafide work carried out by them under my guidance.

DR.R SUBHASHINI

GUIDE

 SITE

Contact tracing of covid 19 virus

Abstract

The increasing spread of communicable illnesses presents a huge threat to human survival. The appearance of the coronavirus COVID-19, which has properly been called a global pandemic and has resulted in so many fatalities, confusion, and massive economic losses, is currently a challenge. The World Health Organization (WHO), in collaboration with various government authorities around the world and non-governmental organisations, has suggested that efforts to contain the COVID-19 pandemic should focus primarily on social distancing, identification of infected people, tracing of possible contacts, and effective isolation of such person(s) for subsequent medical treatment. Given the current trend of infected COVID-19 patients absconding from isolation centers, the goal of this project is to create a framework for monitoring Pandemic Disease Patients' Movements and forecasting people who are more probable of getting covid infected. The methodology for this study proposes a system architecture incorporating GPS (Global Positioning System) and Assisted-GPS technologies for monitoring the geographical movements of COVID-19 patients and recording movement Trajectory Datasets under the assumption that they are assigned GPS-enabled devices such as smartphones. There is evidence that the GPS Trajectory dataset and Machine Learning techniques used in this study have the ability to function best in a real-world circumstance of monitoring a COVID-19 patient.

I. INTRODUCTION

Contact tracing is the name of the process used to identify those who come into contact with people who have tested positive for contagious diseases — such as measles, HIV, and COVID-19. During a pandemic, performing contact tracing correctly can help reduce the number of people to get infected or speed up the process of treating infected people. Doing so can help save many lives.

Contact tracing is a process used by public health ministries to help stop the spread of infectious disease, such as COVID-19, within a community.

Once a person is positive for coronavirus, it is very important to identify other people who may have been infected by the patients diagnosed. To identify infected people, the authorities follow the activity of patients diagnosed in the last 14 days. This process is called contact tracking. Depending on the country and the local authority, the search for contacts is carried out either by manual methods or by numerical methods.

II. BACKGROUND

AI technologies are now being employed to combat the COVID-19 pandemic in a variety of ways, including chatbots for virtual support, diagnostic robots to help identify the disease, serve patients in quarantine, produce treatments, and locate cures, and so on.

The digital approach of contact tracing is based on mobile devices (smartphones) that rely on tracking systems such as Bluetooth and GPS to obtain data from the user's smartphone.

Bluetooth (BLE) essentially communicates anonymous, time-shifting identifiers to other nearby devices, which then store these identifiers in a locally recorded contact history log. When it comes to tracking persons who may have become possibly infected by touching a surface or objects that an unwell person has also touched, the Bluetooth system is less effective. Another disadvantage is Bluetooth's potential inaccuracy in detecting contact events.

Location-based contact tracing, on the other hand, makes use of mobile phone tower networks or GPS, which eliminates some of the shortcomings of Bluetooth, such as the inability to locate geographic hotspots.

Algorithm for Contact Tracing

To create a contact tracing algorithm, we must first complete three steps:

1. Obtain the location data of various users at a certain time and location.

2. Run the data through a clustering algorithm.

3. Use the clusters to forecast who is affected.

Architecture diagram
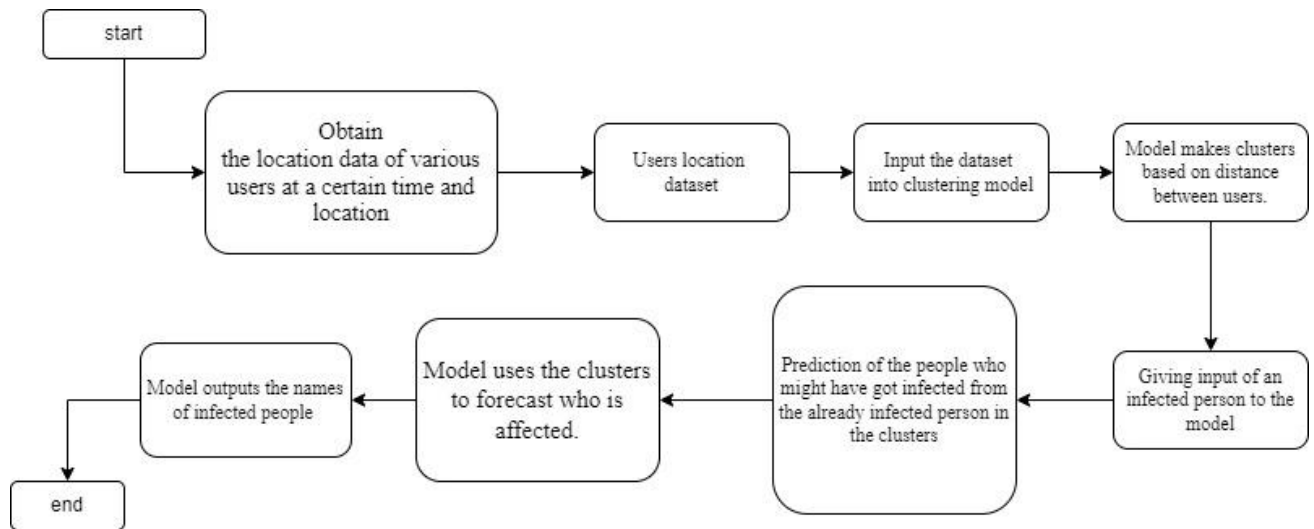
III.   Literature Survey :

There are not many articles on contact tracing of covid 19 using machine learning algorithms.
Many of the articles are based on next wave predition or cases prediction etc and are not directly related to contact tracing . the articles that are chosen here are slighly related to contact tracing and the clustering algorithms.

Soroush Ojagh[3] described the usage of low-cost proximity sensors in contact tracing. A substantial portion of their research concentrated on digital person-to-person contact tracing applications that made use of proximity sensors. The impact of SARS-CoV-2 dissemination through contacting infected surfaces in enclosed spaces is often underestimated in contact tracing applications. Their research focuses on tracing human touch within indoor environments utilising the open OGC IndoorGML standard. Their study presents a graph-based data model with a hierarchical structure that takes into account the semantics of indoor places, time, and user situations. The proposed data model's functionality is assessed for a COVID-19 contact tracing application with scalable system architecture.

[4] This research is remarkable in that it predicts the next likely location of a COVID-19 patient. Given the current pattern of infected COVID-19 patients absconding from isolation units, their study's goal is to create a framework for monitoring Pandemic Disease Patients' Movements and forecasting their future geographical locations. As a result, fifteen individuals were chosen for this study based on residency and business activity location criteria. During a three-week timeframe, the participants' movements generated 157,218 Trajectory

6

datasets. Using this dataset, Artificial Intelligence was used to mine the movement trail, Stay Points, relationships, and estimate the next probable geographical position of a COVID-19 patient .

[5][6] MST-DBSCAN study of 2877 confirmed COVID-19 cases among 12 918 tested persons in the canton of Geneva overlaid on a kernel density surface of population density. Disease clusters were calculated using a daily period beginning with the first verified case on February 26, 2020 and ending with the final accessible data on April 16, 2020. The MST-DBSCAN was calculated using a maximum spatial distance of 600 m, a minimum time distance of one day, and a maximum time distance of fourteen days (incubation period). Connectivity between COVID-19 patients is represented by edges that connect them within a given distance. To guarantee patient privacy, isolated individuals (no contacts within 1200 m) were excluded from the analysis, and random permutation geomasking was performed on precise addresses of residence before mapping. MST-DBSCAN is an abbreviation for modified space-time density-based spatial clustering of applications with a noise technique.

[8] This paper proposes a systematic research review-based methodology for numerous contact tracking applications. The framework's numerous components are related to technological working, design architecture, and application feature analysis, as well as an examination of the global acceptance of such applications. In addition, components concentrating on the security characteristics and analysis of these applications based on Data Privacy, Security Vetting, and other assaults have been added in the research framework. Many applications have yet to investigate the analytical capabilities of data supplied by contact tracing. The various use-cases identified for these applications include detecting positive case probability, identifying a containment zone in the country, locating regional hotspots, monitoring public events and gatherings, identifying sensitive routes, and allocating resources in various regions during the pandemic.

[9] Existing machine learning methods could be used to increase the speed and accuracy of determining COVID-19 clusters and contact tracing. The increased precision in finding clusters would aid in avoiding unneeded COVID-19 testing, which are inherently expensive, and would assist to slow the virus's spread. Early detection of a cluster could also restrict its radius, eliminating unnecessary lockdowns in some locations.

[1]There are numerous clustering methods, including K-means, Mean-Shift, Spectral Clustering, BIRCH, DBSCAN, and many others. These many algorithms can be classified into three types:

Density-based Clustering: Clusters are produced based on the density of the region — DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and OPTICS (Ordering Points to Identify Clustering Structure) are two examples of this type.

Hierarchal-based Clustering: Clusters are formed using a tree-type structure. Some clusters are preset and then used to generate new clusters; examples include CURE (Clustering Using Representatives) and BIRCH (Balanced Iterative Reducing Clustering, and using Hierarchies).

Partitioning-based Clustering: Clusters are generated by partitioning input data into k clusters; examples of this type include: K-means, CLARANS (Clustering Large Applications based upon Randomized Search).

[2]DBSCAN vs. k-means and hierarchical clustering.

DBSCAN may produce arbitrary formed clusters as well as clusters that are completely surrounding by (but not connected to) another cluster.

DBSCAN does not require the number of clusters for the dataset to be given, but k-means and AgglomerativeClustering do.

DBSCAN cannot cluster data sets with high density differences because the eps-min samples combination cannot be chosen appropriately for all clusters. Choosing a meaningful distance threshold eps can be difficult and requires domain expertise if the data and scale are not well understood.

Hierarchical clustering methods would divide huge groups into smaller clusters, whereas DBSCAN keeps them together in a single cluster.

K-means clustering is an unsupervised learning technique that requires the number of clusters (k) to be provided in advance, which is regarded as one of the algorithm's major shortcomings. To determine the best value of k, we must use the elbow technique, which is a heuristic process for determining the number of clusters in a dataset.

DBSCAN constructs a cluster from densely packed data points by supplying eps, which defines the radius of a neighbourhood with respect to some point. Hierarchical clustering, on the other hand, has the particular advantage of allowing any valid measure of distance to be used to build a cluster. It only makes use of the distance matrix.

IV. PROPOSED ALGORITHM

DBSCAN

It is a non-parametric density-based clustering algorithm: given a set of points in a space, it groups together points that are closely packed together (points with many nearby neighbours), identifying outliers as points that lie alone in low-density regions (whose nearest neighbours are too far away). DBSCAN is one of the most widely used clustering algorithms, as well as one of the most frequently cited in scientific literature.

V. EXPERIMENTS RESULTS

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 241 entries, 0 to 240
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   id         241 non-null    object
 1   timestamp  241 non-null    datetime64[ns]
 2   latitude   241 non-null    float64
 3   longitude  241 non-null    float64
dtypes: datetime64[ns](1), float64(2), object(1)
memory usage: 7.7+ KB
```

Output of df.info() call on Jupyter Notebook

The attributes that are used in the dataset are name, timestamp,latitude and longitude.

9
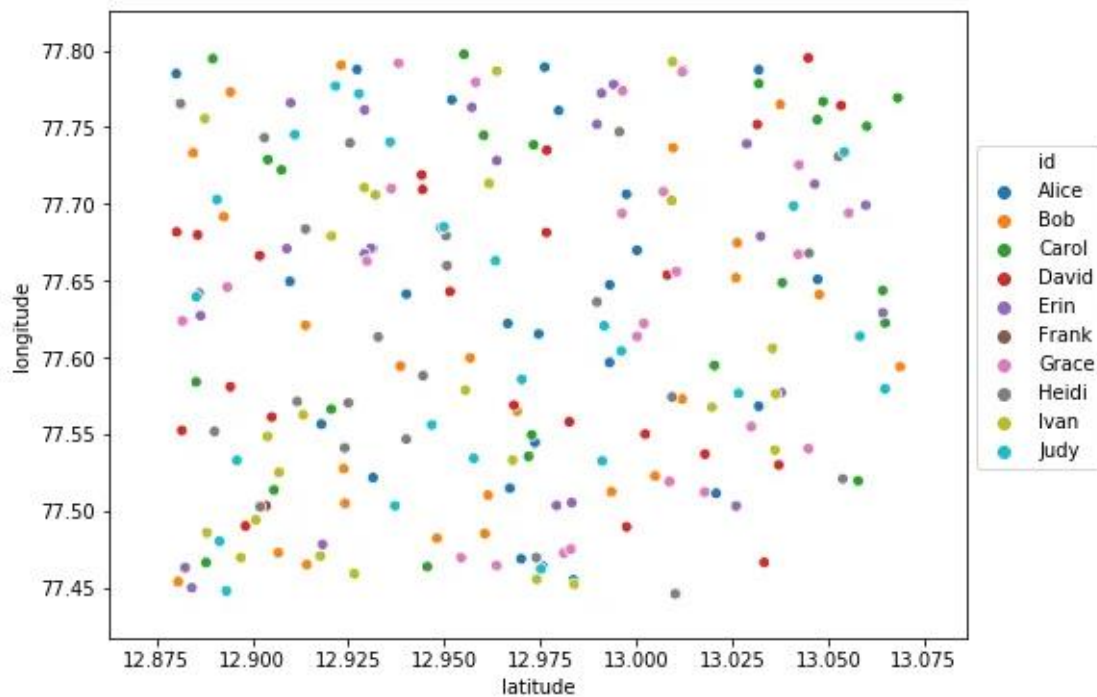
analyzing the dataset using the scatter plot showing the ids with their latitudes and longitudes on the x-axis and the Y-axis respectively

Then create a model for contact tracing using the DBSCAN model. Using this model we will generate clusters, which will help identify infections by filtering the data in the clusters.

let's generate clusters using our model.
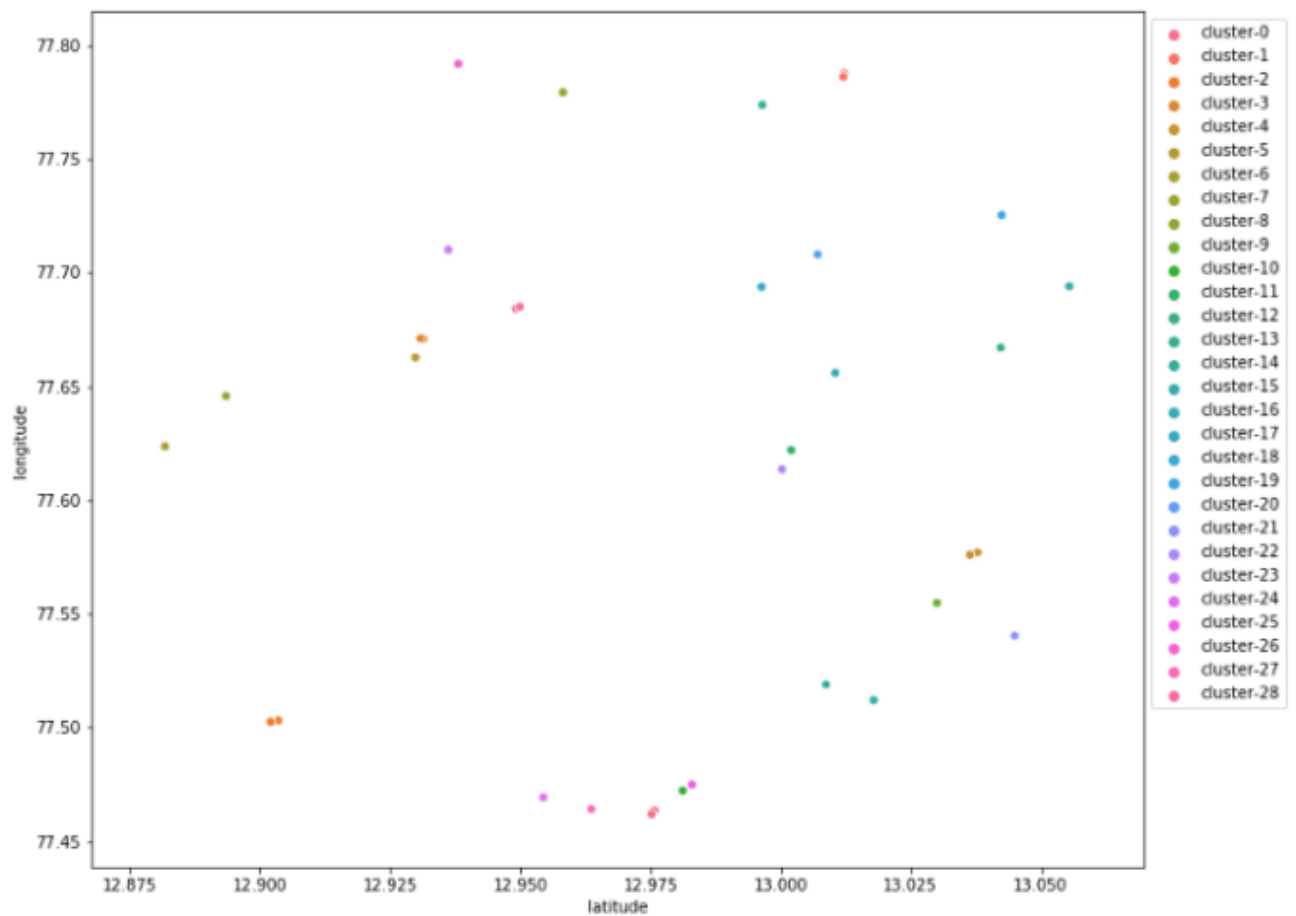
Output obtained by scatter plotting all the clusters except the noise on Jupyter Notebook

To find persons who may have been infected by the patient, we'll simply run the method get_infected_names and pass a name from the dataset as an argument.

The screenshots below show the results of potentially infected people's names.

```
In [17]: get_infected_names('Frank')
Out[17]: ['Grace', 'Carol']
```

```
In [19]: get_infected_names('Erin')
Out[19]: ['Ivan']
```

```
In [20]: get_infected_names('Heidi')
Out[20]: ['David', 'Judy']
```

Website design and screenshots:

Alice

Choose File | dataset.json

Predict

Predict

infected persons are ['Judy']

Web app code:

```python
import os
from flask import Flask, request, render_template
import pandas as pd
from sklearn.cluster import DBSCAN

app = Flask(__name__)


@app.route('/')
def home():


    return render_template('index.html')

@app.route('/predict',methods=['GET','POST'])
def predict():
  if request.method == 'POST':
```

```python
        variable=request.form['infectedperson']
        file = request.files['jsonfile'] # fet input
        filename = file.filename
        print("@@ Input posted = ", filename)

        file_path = os.path.join('static', filename)
        file.save(file_path)
    def get_infected_names(input_name):
        df = pd.read_json(file_path)
        epsilon = 0.0018288 # a radial distance of 6 feet in kilometers
        model = DBSCAN(eps=epsilon, min_samples=2,
metric='haversine').fit(df[['latitude', 'longitude']])
        df['cluster'] = model.labels_.tolist()

        input_name_clusters = []
        for i in range(len(df)):
            if df['id'][i] == input_name:
                if df['cluster'][i] in input_name_clusters:
                    pass
                else:
                    input_name_clusters.append(df['cluster'][i])

        infected_names = []
        for cluster in input_name_clusters:
            if cluster != -1:
                ids_in_cluster = df.loc[df['cluster'] == cluster, 'id']
                for i in range(len(ids_in_cluster)):
                    member_id = ids_in_cluster.iloc[i]
                    if (member_id not in infected_names) and (member_id != input_name):
                        infected_names.append(member_id)
                    else:
                        pass
        return infected_names
    output=get_infected_names(variable)

    return render_template('index.html', prediction_text='infected persons are
{}'.format(output))


if __name__ == "__main__":
    app.run(port=5000,debug=True)
```

VI. USE TO SOCIETY AND FUTURE ASPECTS

- Once a person is positive for coronavirus, it is very important to identify other people who may have been infected by the patients diagnosed.
- This should be definetely implemented in all areas as this is highly effective and shows high efficiency in giving out accurate results.
- It can protect its citizens from this lethal virus to the greatest extent possible. In essence, this application connects Indian health services to their people during this volatile time.
- It identifies the people who might be infected way ahead and those people can be alarmed and made quarantined.
- This not only saves that person life but also stops the spread of the virus drastically by many folds.
- This is not very costly to implement but the results that it demonstrates is very highly appreciable when it comes to its efficiency.

VII. CONCLUSION AND FUTURE WORK

Based on the preceding sections, we may conclude that a clustering algorithm such as DBSCAN can cluster data points without prior knowledge of datasets. The same clustering study can be performed using hierarchical and centroid-based clustering, which can find the best number of clusters even in the presence of noise and outlier points. DBSCAN can detect non-linearly separable clusters in a dataset that cannot be successfully grouped using k-means or Gaussian Mixture EM.

Contact tracing is one of the ways we can use technology to save people's lives and get them the care they need as soon as feasible. There is evidence that the GPS Trajectory dataset and Machine Learning techniques used in this study have the ability to function best in a real-world circumstance of monitoring a COVID-19 patient.

VIII.   REFERENCES

[1] "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" Ester, M., H. P. Kriegel, J. Sander, and X. Xu, In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, AAAI Press, pp. 226–231. 1996

[2] "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). In ACM Transactions on Database Systems (TODS), 42(3), 19.

[3] A Person-to-Person and Person-to-Place COVID-19 Contact Tracing System Based on OGC IndoorGML Soroush Ojagh,Sara Saeedi

[4] A framework for monitoring movements of pandemic disease patients based on GPS trajectory datasets Authors: Paulinus O. Ugwoke, Francis S. Bakpo, Collins N. Udanor, Matthew C. Okoronkwo

[5] Geospatial digital monitoring of COVID-19 cases at high spatiotemporal resolution David De Ridder, José Sandoval, Nicolas Vuilleumier, Silvia Stringhini

[6] Keesara S Jonas A Schulman K Covid-19 and health care's digital revolution.

[7] Kuo F-Y Wen T-H Sabel CE Characterizing diffusion dynamics of disease clustering: a modified space-time DBSCAN (MST-DBSCAN) algorithm.

[8] Technological and analytical review of contact tracing apps for COVID-19 management Rajan Gupta ,Gaurav PandeyO,Poonam Chaudhary &Saibal K Pal

[9] Application of Machine Learning Algorithms for Effective Determination of COVID-19 Clusters Vazeerudeen Abdul HameedEmail authorSelvakumar SamuelKesava Pillai Rajadorai