International e-Conference on New Horizons And Multidisciplinary Applications In Science And Technology
In Association withInternational Journal of Scientific Research in Science and Technology
Volume 9 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Covid-19 Data Analysis

Ms. Akanksha Kaple

ABSTRACT

COVID-19 outbreak was first reported in Wuhan, China and has spread to more than 50 countries. WHO declared COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020. Naturally, a rising infectious disease involves fast spreading; endangering the health of large numbers of people, and thus requires immediate actions to prevent the disease at the community level.

Therefore, Corona Tracker was born as the online platform that provides latest and reliable news development, as well as statistics and analysis on COVID-19.

This paper is done by the research team in the Corona Tracker community and aims to predict and forecast COVID-19 cases, deaths, and recoveries through predictive modeling.

The model helps to interpret patterns of public sentiment on disseminating related health information, and assess political and economic.

Keywords: SARS-COV-2, Data Analysis, Pandemic

I. INTRODUCTION

Corona virus disease 2019(COVID-19) is an infectious disease caused by severe acute respiratory syndrome corona virus (SARS-CoV-2). It was first identified in December 2019 in Wuhan, China, and has resulted in an ongoing pandemic. The first case may be traced back to 17 November 2019. As of 8th June 2020, more than 6.98 million cases have been reported across 188 countries and territories, resulting in more than 401,000 deaths. More than 3.13 million people have recovered.

The virus is primarily spread between people during contact, most often via small droplets produced by coughing, sneezing, and talking. The droplets usually fall to the ground or onto surfaces rather than travelling through air over long distances. Less commonly, people may become infected by touching a contaminated surface and then touching their face. It is most contagious during the first three days after the onset of symptoms, although spread is possible before symptoms appear, and from people who do not show symptoms.

Data Science

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to data mining machine learning and big data.

Data science is a "concept of unify statistics, data analysis, informatics, and their related methods" in order to "understand and analyze actual phenomena" with data. It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science and domain knowledge. However, data science is different from computer science and information science.

Aim

The aim of this project to design a system called Covid-19 Data Analysis that can help us to understand the cases of covid patients and to interpret patterns of public sentiment on disseminating related health information, and assess political and economic.

II. LITERATURE REVIEW

Research papers referred

1) Covid-19 Impact on Indian Economy

Authors: Mahendra Dev (Indira Gandhi Institute of Development Research) Rajeshwari Sengupta (Indira Gandhi Institute of Development Research) https://ideas.repec.org/p/ind/igiwpp/2020-013.html

2) COVID-19 in India: Potential Impact of the Lockdown and Other Longer- Term policies.

Authors: Emily Schueller(CDDEP) Eili Klein (Department of Emergency Medicine Johns Hopkins School of Medicine; CDDEP) Gary Lin (Department of Emergency Medicine Johns Hopkins School of Medicine;CDDEP) https://cddep.org/wp-content/uploads/2020/04/India-Shutdown-Modeling -Slides- Final-2.pdf

Data Sources

For the COVID-19 data have scrapped https://api.covid19india.orgwhich is a volunteer-driven, crowd sourced database for COVID-19 state & stats & patient tracing in India.

For facts and information we have referred www.wikipedia.com and www.twitter.com.

III. SYSTEM REQUIREMENTS

2.1 Hardware requirements

- OS : Windows 7 or Newer
- Processor : Minimum 1 GHz; Recommended 2 GHz or more
- Hard Drive : Minimum 32 GB; Recommended 64 GB or more
- Memory (RAM) : Minimum 1 GB; Recommended 4GB or above
- Ethernet connection (LAN) OR a wireless adapter(Wi-Fi)

2.2 Software requirements

a. Jupyter notebook

Jupyter notebook is a web-based interactive computational environment for creating jupyter notebook documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension.

A Jupyter Notebook can be converted to a number of open standard output formats (HTML, presentation slides, LaTeX, PDF, Restructured tata, Markdown, Python) through "Download As" in the web interface, via the nbconvert library or "jupyter nbconvert" command line interface in a shell. To simplify visualization of Jupyter notebook documents on the web, the nbconvert library is provided as a service through NbViewer which can take a URL to any publicly available notebook document, convert it to HTML on the fly and display it to the user.

b. MySQL

MySQL is an open-source relational database management system (RDBMS). A relational database organizes data into one or more data tables in which data types may be related to each other: these relational help structure the data. SQL is a language programmers use to create, modify and extract data from the relational database, as well as control user access to the database.

In additional to relational database and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access and facilities testing database integrity and creation of backups.

c. Pandas

Pandas is a software library written for the Python Programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself.

d. NUMPY

NumPy is a general-purpose array-processing package. It provides a high performance multi dimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones.

A powerful N-dimensional array object, Sophisticated (broadcasting) functions, Tools for integrating C/C++ and Fortran code Useful linear algebra, Fourier transform, and random number capabilities.

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

2.3 UML Diagrams

a. Data Flow Diagram

A data flow diagram (DFD) is a way of representing a flow of data through a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow; there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart. The data flow diagram is a part of the structured-analysis modeling tool. When using UML, the activity diagram typically takes over the role of the data-flow diagram. A special form of data-flow plan is a site- oriented data-flow plan.

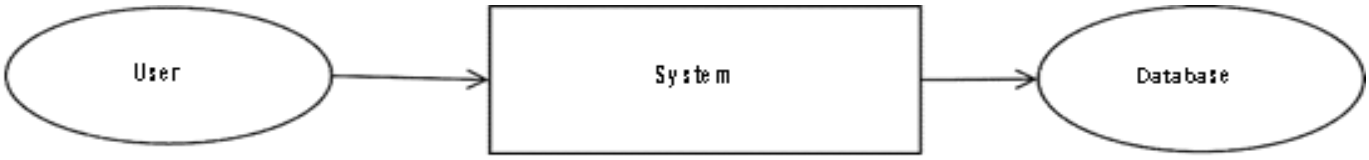

Fig. 3.5 : Level 0 Data FlowDiagram

Fig. 3.6 : Level 1 Data Flow Diagram

b. E-R Diagram

An entity-relationship (ER) model describes interrelated things in a specific domain of knowledge. A basic ER model is composed of entity types (which classify the things of interest) and specifies relationships that can exist between entities (instance of those entity types).

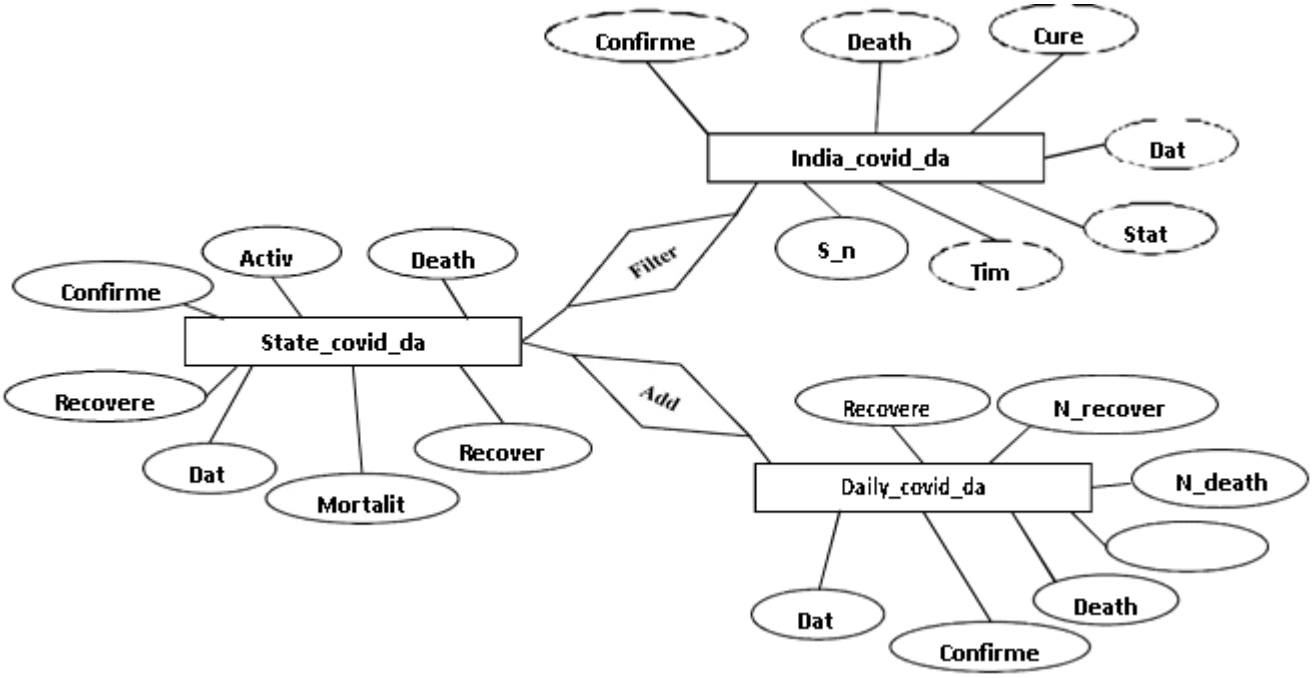

Fig. 3.7 : E-R Diagram

c. Use Case Diagram:

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved,

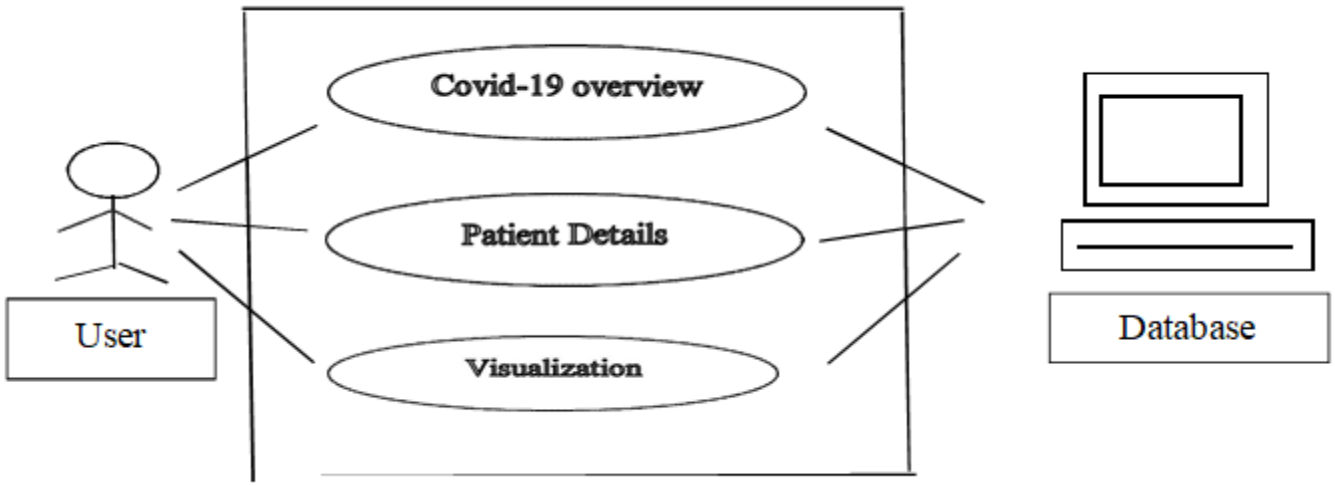

Fig 3.8 : Use Case Diagram

IV. SYSTEM IMPLEMENTATION

3.1 Data base creation

The problem of creating a covid data analysis system has two separate stages: database creation and query processing. The first stage requires the creation of database of covid patients of India. An effective storage mechanism may be implemented at this stage to improve retrieval response. In the query processing stage, gives information about the patients, number of patients with different types of cases in the existing database.

Database Creation collects the information and number of cases of patients. It stores daily cases on the basis of different cases like confirmed cases, active cases, recovered cases and deaths cases as well as store by daily cases, state wise cases and all India's cases in database. The database stored on the disk.

3.2 Query processing

- Extract the information of the patients.
- Differentiate the patients according to his type of case.
- Process on the number of cases according to type of case.

3.3 Web interface

India recorded its first COVID-19 case on 30th January 2020 in kerala. The infected person was a student who had travelled to china for academic purpose. And since then cases in India is rising exponentially.

India had recorded over 500 cases till 24th March. Therefore, government declared nation-wide lockdown from 25th march to 14th April also known as lockdown 1.0 and after this government has been extending nation-wide lockdown step by step.

Situation in India,

Confirmed Cases: 9735850

Recovered: 9215581

Active: 378909

Deceased: 141360

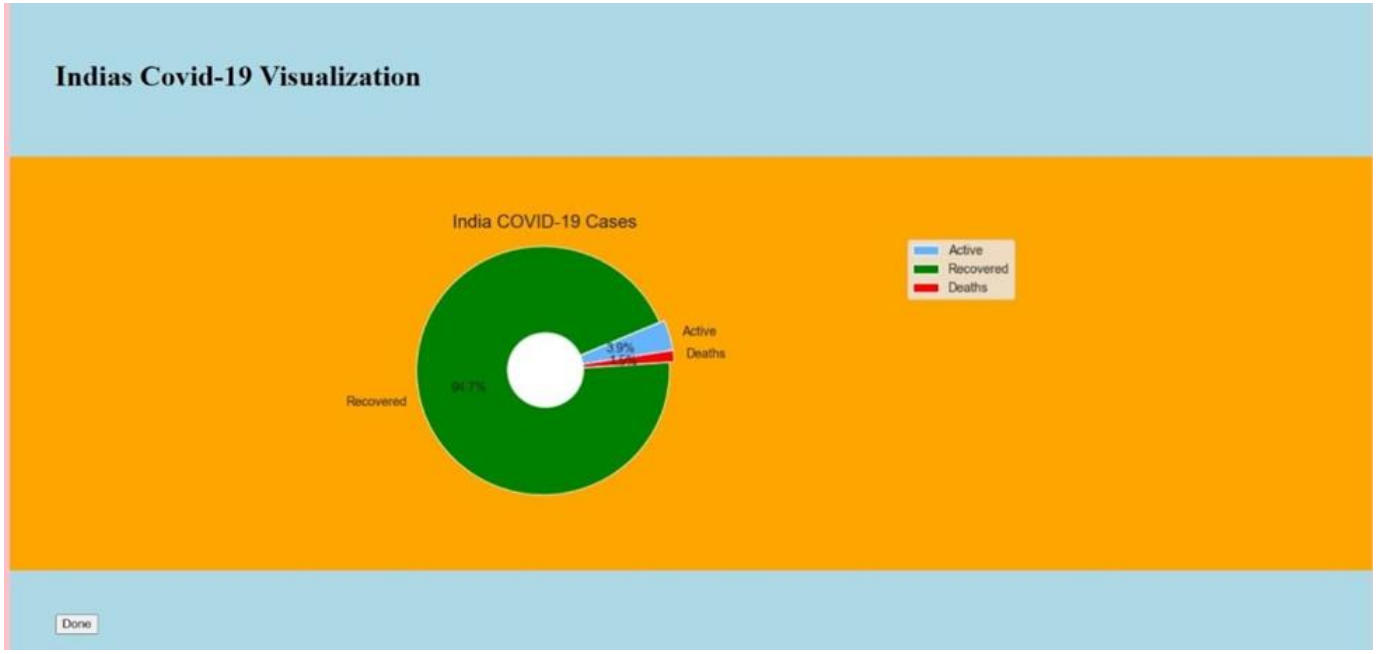

Fig 3.1 : India's Covid Ratio

India is showing good recovery rate day by day with low rate of deaths but on the other hand it is also reaching new peak of confirmed cases every day.

State wise Comparison

India consists of 28 states and 8 Union Territories with varying features such as demography, geography, location, lifestyle etc. Which can decide the spread of contagious virus? Comparing COVID-19 situation in states gives us insight of which state/area is to be focused.

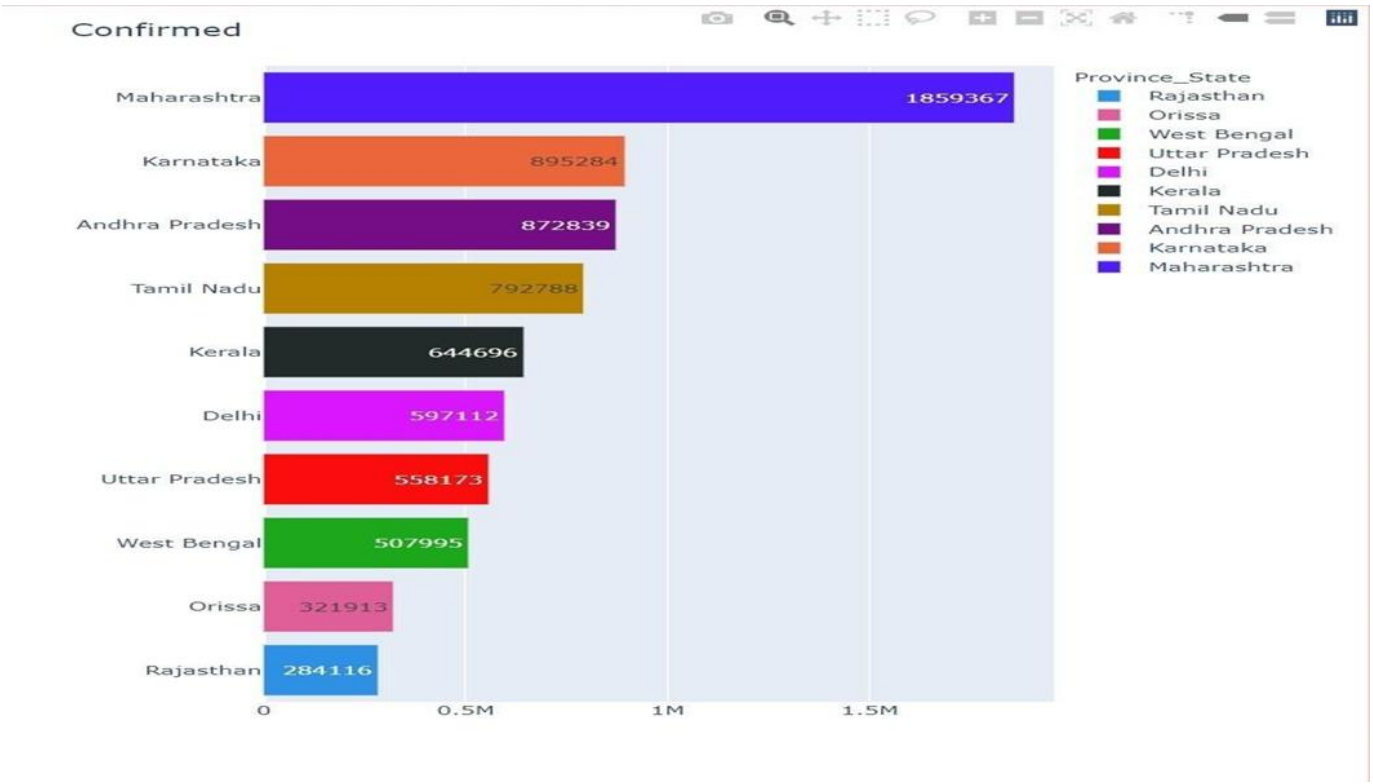

Fig 3.2 : State Wise Covid Ratio

As shown above, Maharashtra, Karnataka, Andra Pradesh, Tamil Nadu, Kerala, Delhi, Uttar Pradesh, West Bengal, Orissa, Rajasthan are top 10 states in confirmed cases in India. When the cases started to be detected in India since that time only Maharashtra has been leading this tally.

V. TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses specific testing requirements.

4.1 Types of testing

a. Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application . It is done after the completion of

an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

b. Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

c. Acceptance testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input: identified classes of valid input must be accepted.

Invalid Input: identified classes of invalid input must be rejected.

Functions: identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems / Procedures :

Interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

d. Tests Results

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

e. White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

f. Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box

.you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

4.2 Unit Testing

Unit testing is usually conducted as part of a combined code and unit test phase of the software life cycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

4.3 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications,

e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

4.4 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

VI. ADVANTAGES AND LIMITATIONS

5.1 Advantages

- To understand the ratio of cases.
- Helpful to understand the ratio of cases in terms of different parameters
- Helps to recognize COVID balance.
- User friendly application.
- Secure database.
- Easy maintenance.
- Data can be updated easily.
- Database records can be maintained computer.
- Easy data retrieval.
- System reduces manual workload.

- Less error prove and accuracy is maintained.
- Redundancy of data is avoided.
- Provides report of data.
- Fast processing and handy.

5.2 Limitations

The ratio of these data varies on daily basis, therefore we are unable to predict exact prediction about ratio.
Due to the uncertainty to predict exact ratio, we can't say when the cases will increase or decrease.

VII. CONCLUSION

Covid-19 cases are increasing daily, and it is very important to analyze these data. In this study, Covid-19 patients' data were analyzed to determine the relationship between different variables. Results show that there are dependence in age group and current status and in age group and gender only in gender, and current status variables are independent.
Our estimation show that the COVID-19 epidemic trend in India will rise from Feb 28, 2020 and will peak during May 15 to June 15.
Under such circumstances, two types of hypotheses can be considered for the adequacy of the current measures and as to whether the epidemic peak will occur during June 15-31, 2020.
Assuming that the current measures are inefficient and inadequate, the number of estimated cases will continue to increase by the end of Oct.
The current control measures are effective and sufficient and the number of cases estimated after the epidemic peak will go down.
At the time of the outbreak of COVID-19, the best and most urgent steps must be taken to overcome the corona virus epidemic. The fight against corona virus infection should be an emergency.

Future Scope

To learn more about other attributes such as patient gender, ethnicity and age and how it causes the fatality rate. A dash board of interactive charts to provide an overall summery.

VIII. REFERENCES

[1] . D. Hui et al, "The continuing 2019-nCoV epidemic threat of novel corona virus to global health - The latest 2019 novel corona virus outbreak in Wuhan, China, "International Journal of Infectious Diseases, vol. 91, pp. 264-266,2020.
[2] . World Health Organization (WHO), "Coronavirus disease 2019 (COVID-19) Situation Report - 35," WHO, 2020.
[3] . World Health Organization (WHO), "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCov)," WHO,2020.
[4] . World Health Organization. (2020). COVID-19 Strategy Update. Geneva: WHO. Retrieved 11 May 2020, from http://www.int/publicationsdetail/covid-19- strategy-update-14-april-2020.

[5] . WHO, "Advice for Public," WHO Int., 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novelcoronavirus2019/advice- for-public [Accessed 27 February2020].

[6] . World Health Organization, "Report of the WHO-China Joint Mission on Corona virus Disease 2019 (COVID-19)," World Health Organization, 2020.

[7] . "Corona virus Disease 2019 (COVID-19)," Centers for Disease Control and Prevention, 2020.[Online]. Available: https://www.cdc.gov/coronavirus/2019- ncov/about/symptoms.html [Accessed 27 February 2020].

[8] . World Health Organization, "2019 Novel Corona virus (2019-nCoV): Strategic Preparedness and Response Plan," World Health Organization, Geneva,2020.

[9] . World Health Organization, "Corona virus disease 2019 (COVID-19) Situation Report- 13," World Health Organization,2020.

[10] . Corona Tracker Community, "Corona Tracker," Corona Tracker, 2020. [Online]. Available: https://www.coronatracker.com[Accessed 28 February2020].

[11] . World Health Organization. (2020). COVID-19: Operational Planning Guidelines And COVID-19 Partners Platform to support country preparedness and response. Retrieved 28th April 2020, from https://openwho.org/courses/UNCT-COVID19- preparedness-response- EN.