Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

FSD Mini Project Report

Topic

Text-Util

Guided By

Prof. Sanket Salvi

Group Members

PRN	NAME	ROLL NO.
1032210790	Vansh Gurnani	25
1032210573	Yuvraj singh Lamba	20
1032210606	Keshav Jha	22
1032210499	Jay Mehta	14

Table of Contents

1) Project Title

Text-Util

2) Team Members

PC14 Jay Mehta
PC20 Yuvraj Singh Lamba
PC22 Keshav Jha
PC25 Vansh Gurnani

3) Overview of various big data technologies used in the mini project

a. MongoDB:-

MongoDB is an open source NoSQL database management program. NoSQL (Not only SQL) is used as an alternative to traditional relational databases. NoSQL databases are quite useful for working with large sets of distributed data. MongoDB is a tool that can manage document-oriented information, store or retrieve information.

MongoDB is used for high-volume data storage, helping organizations store large amounts of data while still performing rapidly. Organizations also use MongoDB for its ad-hoc queries, indexing, load balancing, aggregation, server-side JavaScript execution and other features.

Why is MongoDB used?

Storage: MongoDB can store large structured and unstructured data volumes and is scalable vertically and horizontally. Indexes are used to improve search performance. Searches are also done by field, range and expression queries.

Data integration: This integrates data for applications, including for hybrid and multi-cloud applications.

Complex data structures descriptions: Document databases enable the embedding of documents to describe nested structures (a structure within a structure) and can tolerate variations in data.

Load balancing: MongoDB can be used to run over multiple servers.

The MongoDB document stores information related to a note entry in the project. Here's what each field within the document is storing:

1. "_id": This field holds a unique identifier for the note entry. It serves as a primary key for this particular document.

2. "content": This field stores the actual content of the note.

3. "category": The "category" field allows users to categorize their notes.

4. "owner": This field stores the unique identifier of the user who created the note. It helps associate the note with its respective user.

5. "bookmarked": The "bookmarked" field stores a boolean value (true or false) to indicate whether the note has been bookmarked.

6. "timestamp": This field records the current date and time when the note was created.

These fields collectively store information about the note's content, category, ownership, bookmark status, and the time it was created. This structured data enables users to effectively manage and access their notes within the project.

b. D3.js:-

In our "Notepad" website, we have harnessed the full potential of D3.js, a versatile JavaScript library, to provide users with an engaging and informative experience. D3.js allowed us to create interactive data visualizations that play a central role in helping users understand their activities on the platform. For instance, we developed dynamic charts that illustrate the relationship between the amount of time spent on the website and word count, enabling users to track their productivity over time. These visualizations are not only aesthetically appealing but also functional, making it easier for users to make data-driven decisions about their usage patterns.

Furthermore, D3.js has been instrumental in crafting categorization graphs and bar charts. Users can categorize their notes, and D3.js empowers these categories to come to life through visually appealing representations. Whether it's tracking bookmarked notes or analyzing usage patterns, D3.js enhances the user experience by providing a graphical, interactive means of understanding their data. The combination of D3.js and other data analysis tools

makes our project a comprehensive platform for users to gain insights into their online habits and maximize their productivity.

c. PySpark:-

In the context of our project, PySpark is employed for in-depth data analysis, particularly for word frequency analysis and creating word count vs. time graphs using Plotly. PySpark is a powerful data processing framework that allows you to perform complex operations on the data stored in MongoDB.

PySpark can be utilized to analyze the content field of your MongoDB documents, extracting insights such as word frequencies. This enables users to gain a better understanding of the most common words in their notes and assess their content. Additionally, PySpark can help generate word count vs. time graphs, providing users with a visual representation of how their note-taking habits evolve over time. These capabilities make PySpark an essential tool for deriving meaningful insights from the data stored in MongoDB and enhancing the analytical features of our project.

D) Node.Js

Node.js is a groundbreaking, open-source, server-side JavaScript runtime environment that has transformed the landscape of backend development. Unlike traditional server-side frameworks, Node.js operates on a non-blocking, event-driven architecture, empowering developers to build scalable, high-performance applications. Its event-driven nature enables handling multiple concurrent connections efficiently, making it particularly adept at handling real-time applications, APIs, and microservices. Powered by the V8 JavaScript engine, Node.js executes JavaScript code outside a web browser, allowing developers to leverage JavaScript both on the client and server sides of web applications. Its rich ecosystem of packages, facilitated by npm (Node Package Manager), provides access to an extensive library of modules, simplifying the development process by offering ready-made solutions for various functionalities. Node.js has emerged as a cornerstone technology in modern web development, offering speed, scalability, and flexibility to craft robust and responsive web applications.

E) React js

React.js, a powerful JavaScript library for building user interfaces, serves as the frontend framework for our "Notepad" application. Leveraging React.js enables us to create a dynamic and responsive user interface, enhancing the overall user experience.

Key Features and Implementation:

Component-based Architecture: React's component-based structure allows us to break down the UI into reusable and manageable components. For instance:

Note Component: Each note entry on the interface is represented as a React component, encapsulating its content, category, timestamp, and other relevant details.

Sidebar Component: Utilized for categorization and quick access to different types of notes.

Virtual DOM (Document Object Model): React's efficient rendering process using a virtual DOM ensures optimal performance by updating only the components that have changed. This results in a smoother and faster UI experience, especially when handling a large number of notes.

State Management: React's state management allows us to maintain the application's state seamlessly. For example:

Real-time Updates: Whenever a new note is added or edited, React efficiently updates the UI to reflect these changes without requiring a full-page reload.

Interactive Features: Implementing features like note categorization, bookmarking, and real-time search is made simpler through React's state handling.

Responsive Design: With React's flexibility, we've ensured that the "Notepad" application is responsive across various devices and screen sizes. Users can seamlessly access and manage their notes from desktops, tablets, or smartphones.

Integration with Data Analysis Tools: React.js seamlessly integrates with other technologies like D3.js and PySpark, enabling us to display data visualizations and analysis results generated from the backend.

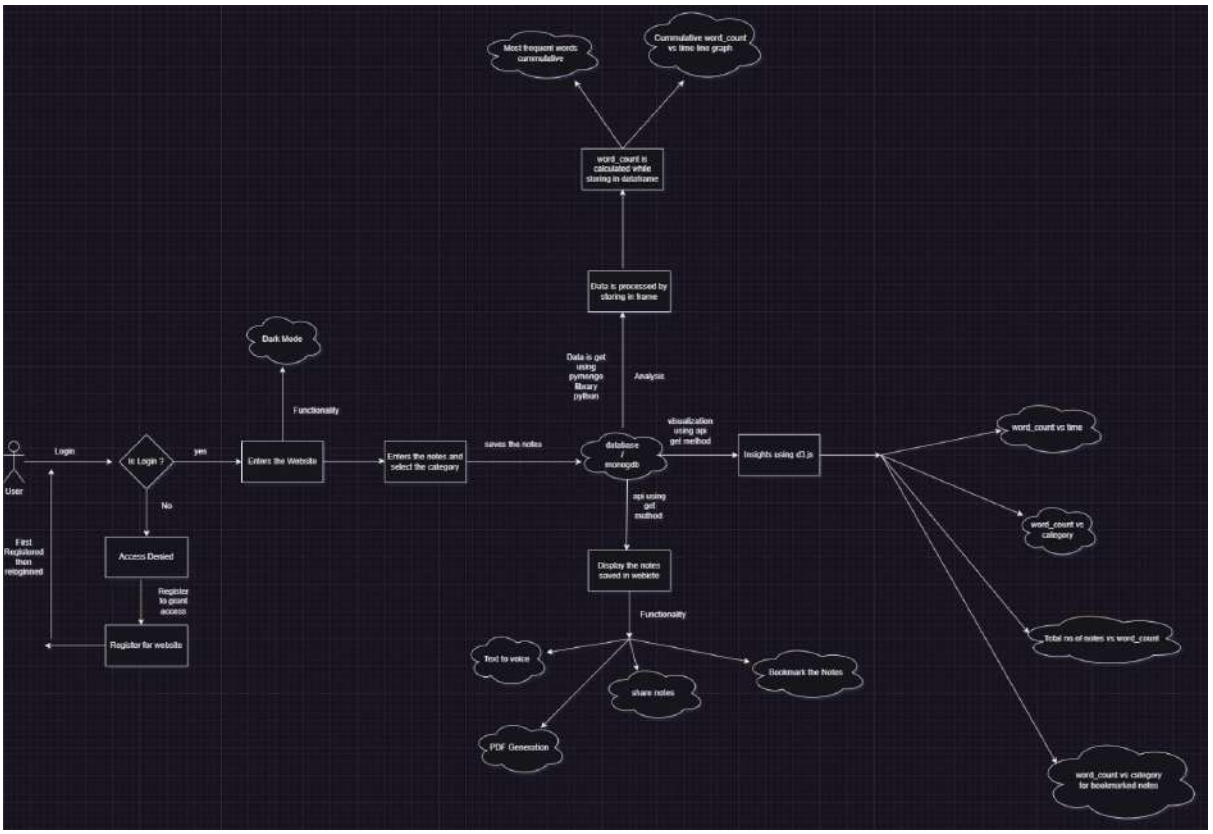
F) Express Js

Express.js stands as a minimalist and flexible web application framework built on top of Node.js, simplifying the process of building robust and scalable web applications. Renowned for its unobtrusive yet powerful features, Express.js offers a myriad of functionalities for routing, middleware, and templating, making it an ideal choice for both small-scale and large-scale projects. Its intuitive routing system allows developers to define and handle various HTTP request methods on different URL endpoints, ensuring clean and organized code structure. Express.js seamlessly integrates with numerous middleware, providing additional functionalities such as logging, authentication, and error handling, enhancing the security and performance of applications. Furthermore, its support for various templating engines enables the creation of dynamic views and content rendering, empowering developers to craft interactive and engaging user interfaces. Express.js' flexibility, extensive documentation, and active community support make it a go-to framework for building modern, feature-rich web applications with speed and efficiency.

G) Bootstrap

Bootstrap remains a widely acclaimed front-end framework that revolutionizes the creation of responsive, mobile-first web applications and websites. Offering a comprehensive suite of pre-built CSS and JavaScript components, Bootstrap streamlines the design process, enabling developers to craft visually appealing, consistent, and functional interfaces with minimal effort. Its grid system facilitates the creation of responsive layouts that adapt seamlessly across various devices and screen sizes, ensuring a consistent user experience. Bootstrap's extensive collection of customizable components, including navigation bars, buttons, forms, and modals, empowers developers to create polished and modern interfaces while adhering to industry best practices. Moreover, its documentation and active community support make it an accessible tool for both novice and experienced developers, cementing its status as a go-to framework for building attractive and responsive web applications efficiently.

4) Workflow diagram



5) (i) Future Scope:-

Collaborative Editing: Implement real-time collaboration features that allow multiple users to edit the same document simultaneously. This could be useful for team projects or brainstorming sessions.

Voice Recognition and Dictation: Integrate voice recognition technology to allow users to dictate their notes instead of typing them manually. This can be especially useful for users on the go or those with accessibility needs.

Privacy and Security: Incorporating robust encryption and privacy features to ensure the safety of sensitive information stored in the notepad. For example, using a pin to access notes regarding bank details.

Customizable Templates: Provide users with the ability to create and save custom templates for different types of notes (e.g., meeting minutes, to-do lists, project plans).

Machine Learning for Predictive Text: Implement a predictive text feature that suggests the next word or phrase based on the user's writing patterns and context.

(ii) Conclusion:-

In conclusion, "Notepad" is a robust and versatile project that combines the power of technologies like D3.js, PySpark, and MongoDB to offer users a comprehensive solution for note-taking, data analysis, and time management. With its current feature set and the potential for future expansion, our peoject provides a user-friendly and data-driven approach to note-taking, fostering productivity and enhancing the overall note-taking experience.

6) References:-

https://d3js.org/ https://www.mongodb.com/

https://spark.apache.org/docs/latest/api/python/index.htm

l https://colab.google/ https://plotly.com/

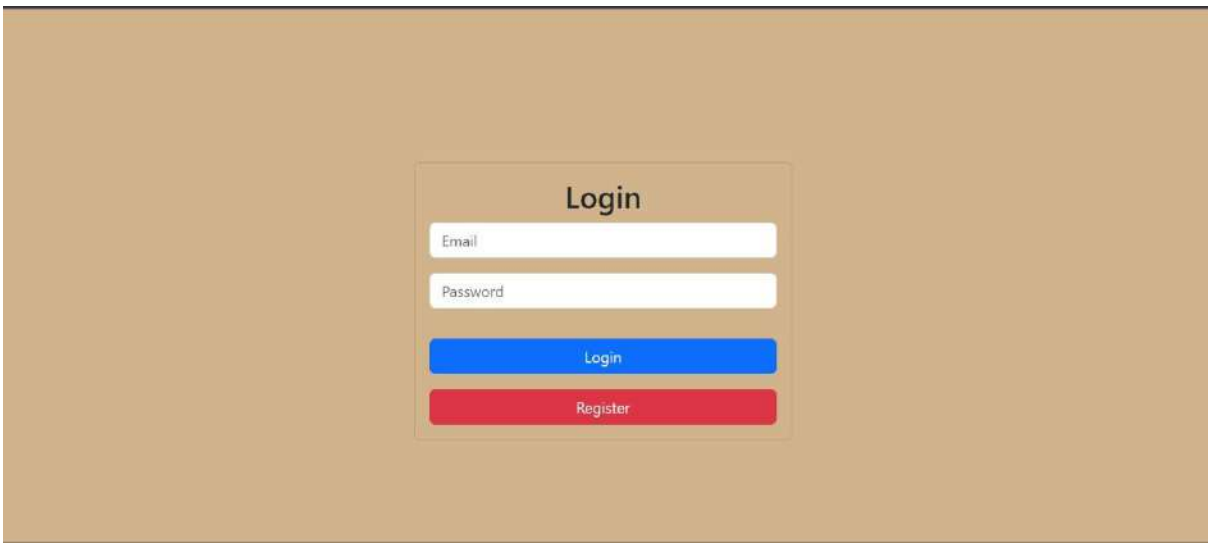
7) Source Code :-

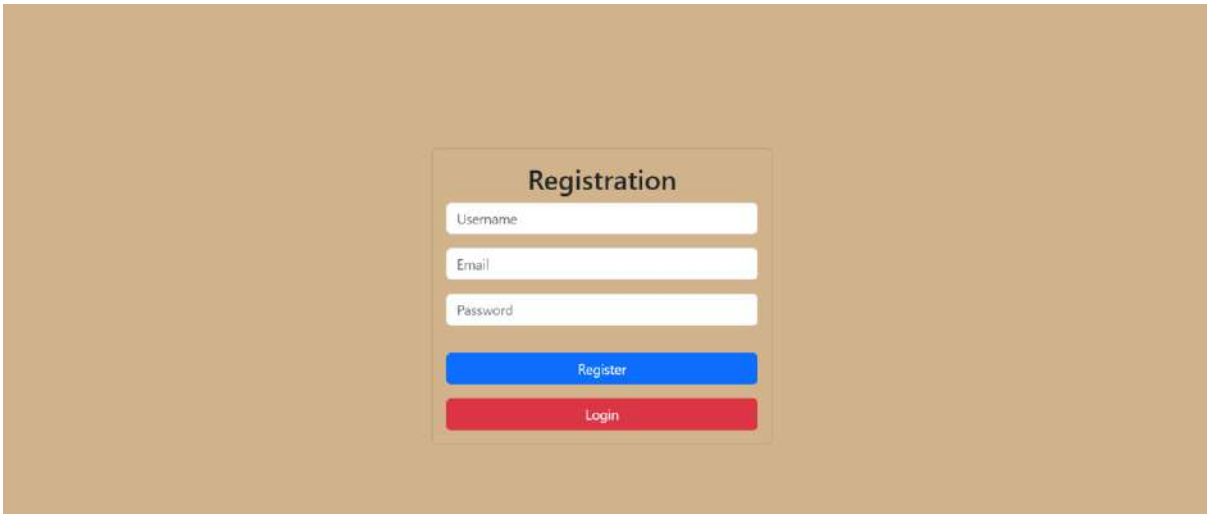
https://github.com/vanshgurnani/Text-Util

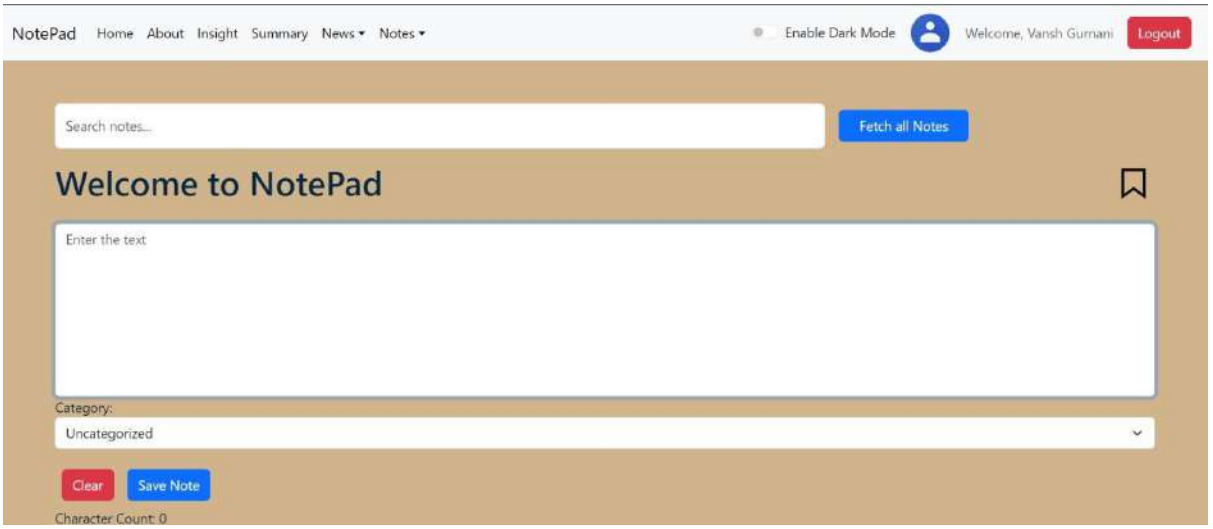
https://github.com/vanshgurnani/flask-test

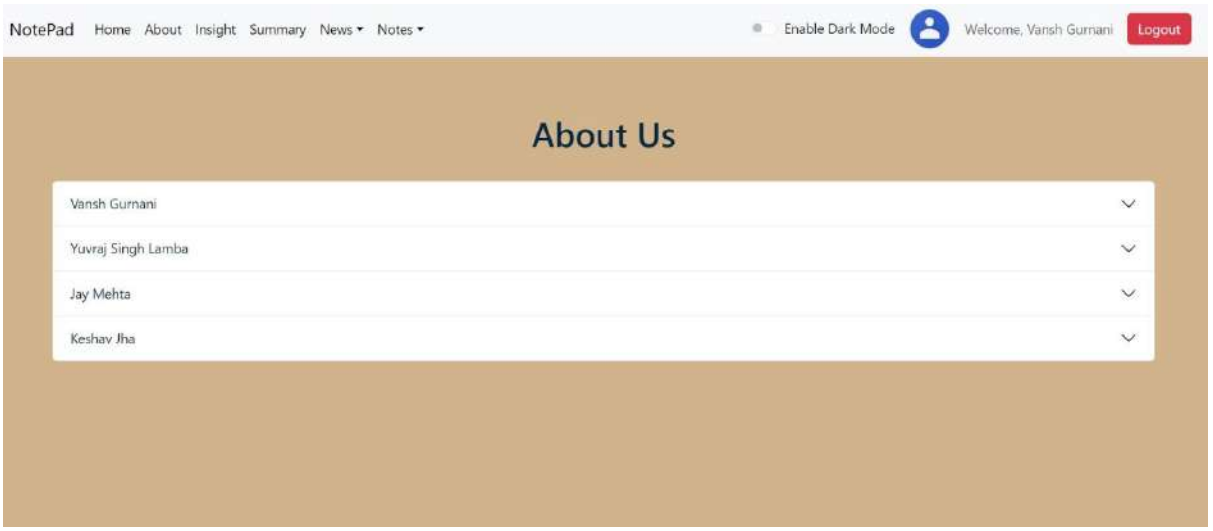
Screenshots of the Project:-

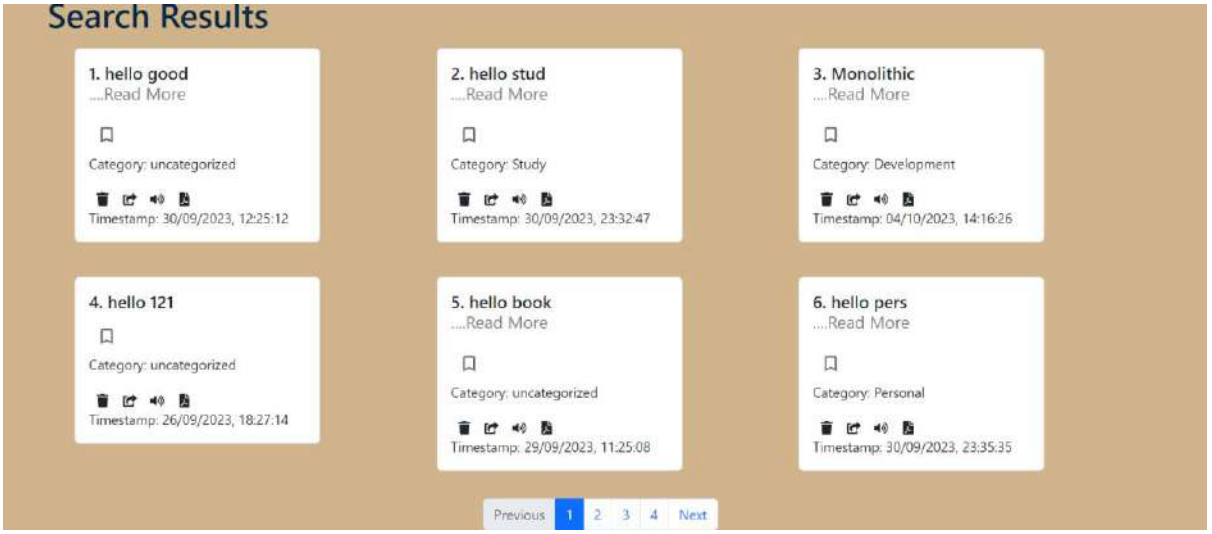
Light Mode :-

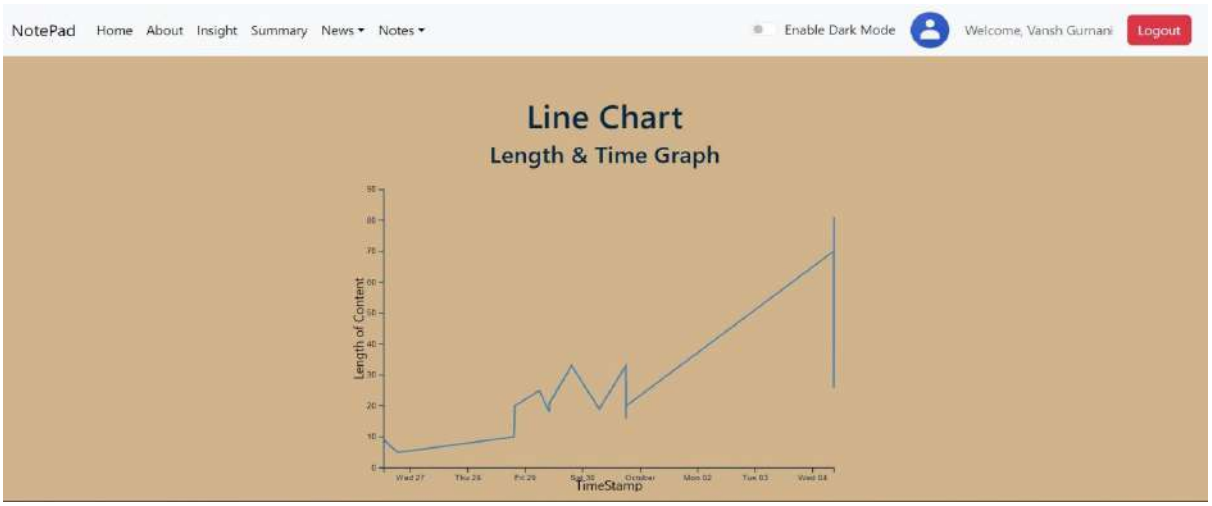





NotePad
Home
About
Insight
Summary
News
Notes
Enable Dark Mode
Welcome, Vansh Gurnani
Logout
Search notes...
Fetch all Notes
Welcome to NotePad
Enter the text
Category:
Uncategorized
Clear
Save Note
Character Count: 0

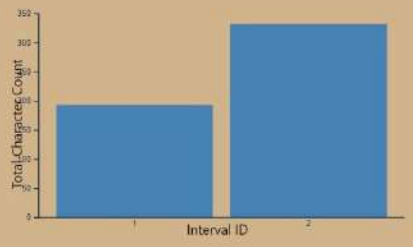

NotePad
Home
About
Insight
Summary
News
Notes
Enable Dark Mode
Welcome, Vansh Gurnani
Logout
About Us
Vansh Gurnani
Yuvraj Singh Lamba
Jay Mehta
Keshav Jha


Search Results
1. hello good
....Read More
Category: uncategorized
Timestamp: 30/09/2023, 12:25:12
2. hello stud
....Read More
Category: Study
Timestamp: 30/09/2023, 23:32:47
3. Monolithic
....Read More
Category: Development
Timestamp: 04/10/2023, 14:16:26
4. hello 121
Category: uncategorized
Timestamp: 26/09/2023, 18:27:14
5. hello book
....Read More
Category: uncategorized
Timestamp: 29/09/2023, 11:25:08
6. hello pers
....Read More
Category: Personal
Timestamp: 30/09/2023, 23:35:35
Previous
1
2
3
4
Next


NotePad
Home
About
Insight
Summary
News
Notes
Enable Dark Mode
Welcome, Vansh Gumani
Logout
Line Chart
Length & Time Graph
Length of Content
TimeStamp

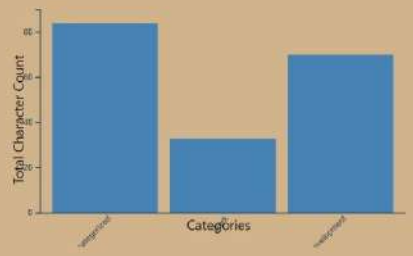
Bar Graph

Total Character Count for Each 10-Note Interval



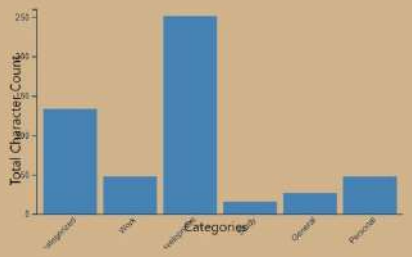
Bookmark Graph

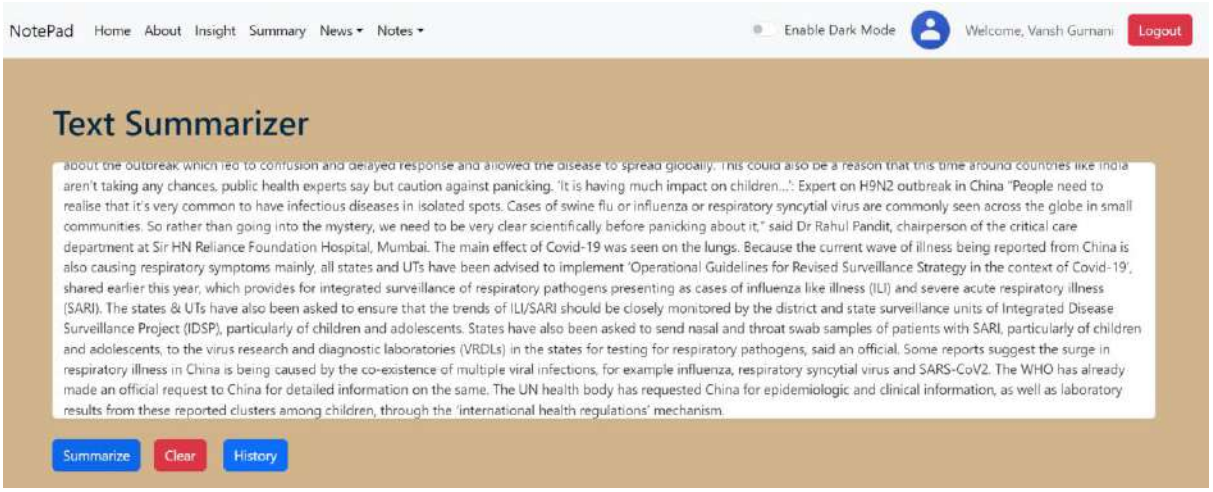
Category Graph for Bookmarked Notes

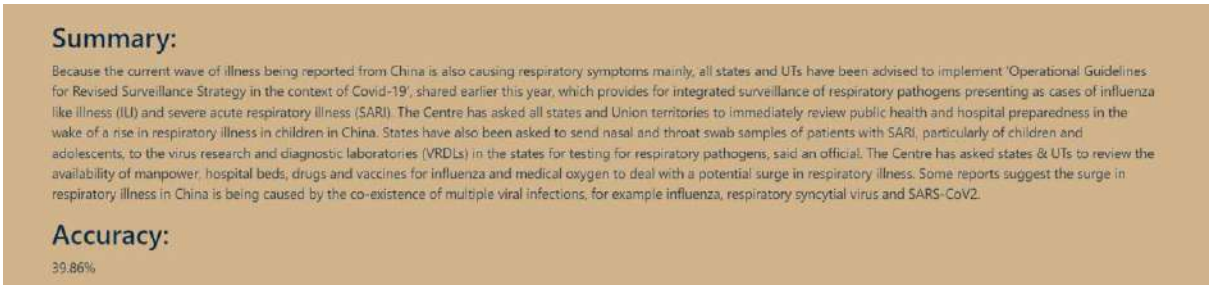


Bar Chart

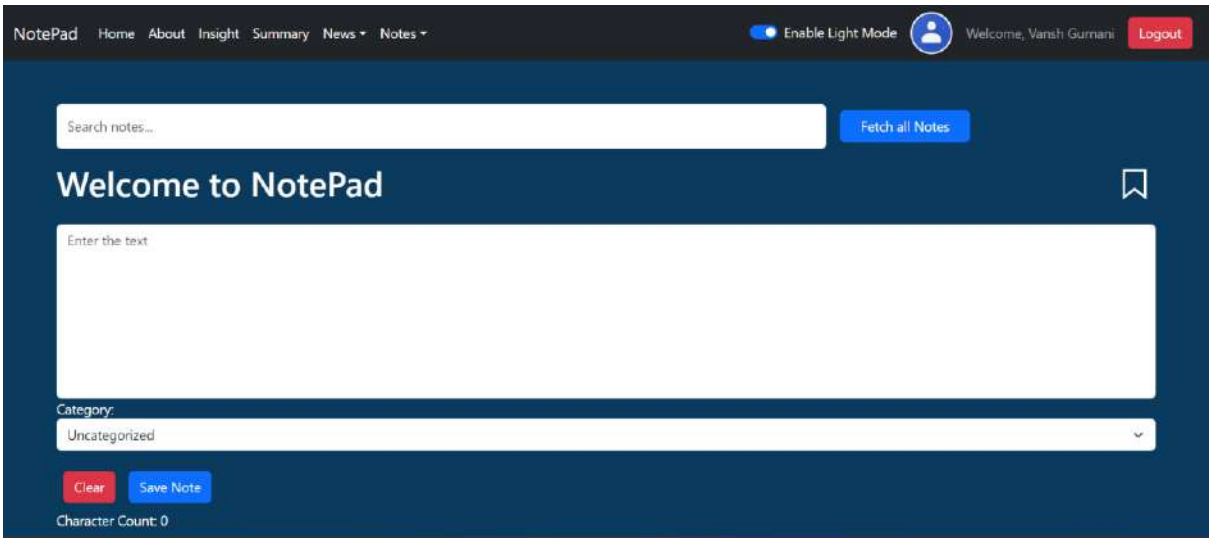
Category Graph

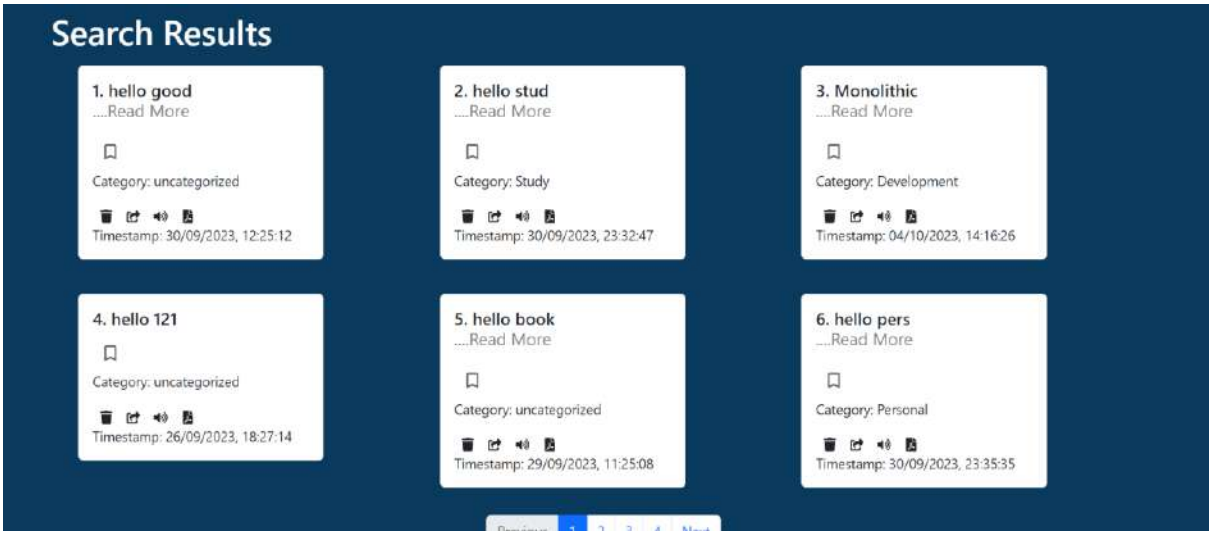


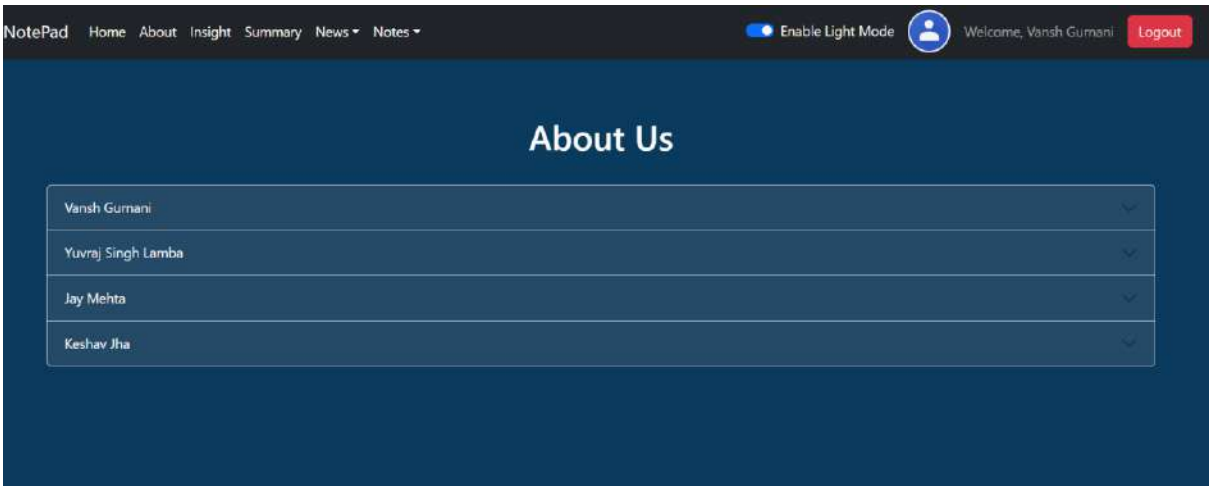

NotePad
Home
About
Insight
Summary
News
Notes
Enable Dark Mode
Welcome, Vansh Gurnani
Logout
Text Summarizer
about the outbreak which led to confusion and delayed response and allowed the disease to spread globally. This could also be a reason that this time around countries like India aren't taking any chances, public health experts say but caution against panicking. 'It is having much impact on children...': Expert on H9N2 outbreak in China "People need to realise that it's very common to have infectious diseases in isolated spots. Cases of swine flu or influenza or respiratory syncytial virus are commonly seen across the globe in small communities. So rather than going into the mystery, we need to be very clear scientifically before panicking about it," said Dr Rahul Pandit, chairperson of the critical care department at Sir HN Reliance Foundation Hospital, Mumbai. The main effect of Covid-19 was seen on the lungs. Because the current wave of illness being reported from China is also causing respiratory symptoms mainly, all states and UTs have been advised to implement 'Operational Guidelines for Revised Surveillance Strategy in the context of Covid-19', shared earlier this year, which provides for integrated surveillance of respiratory pathogens presenting as cases of influenza like illness (ILI) and severe acute respiratory illness (SARI). The states & UTs have also been asked to ensure that the trends of ILI/SARI should be closely monitored by the district and state surveillance units of Integrated Disease Surveillance Project (IDSP), particularly of children and adolescents. States have also been asked to send nasal and throat swab samples of patients with SARI, particularly of children and adolescents, to the virus research and diagnostic laboratories (VRDLs) in the states for testing for respiratory pathogens, said an official. Some reports suggest the surge in respiratory illness in China is being caused by the co-existence of multiple viral infections, for example influenza, respiratory syncytial virus and SARS-CoV2. The WHO has already made an official request to China for detailed information on the same. The UN health body has requested China for epidemiologic and clinical information, as well as laboratory results from these reported clusters among children, through the 'international health regulations' mechanism.
Summarize
Clear
History

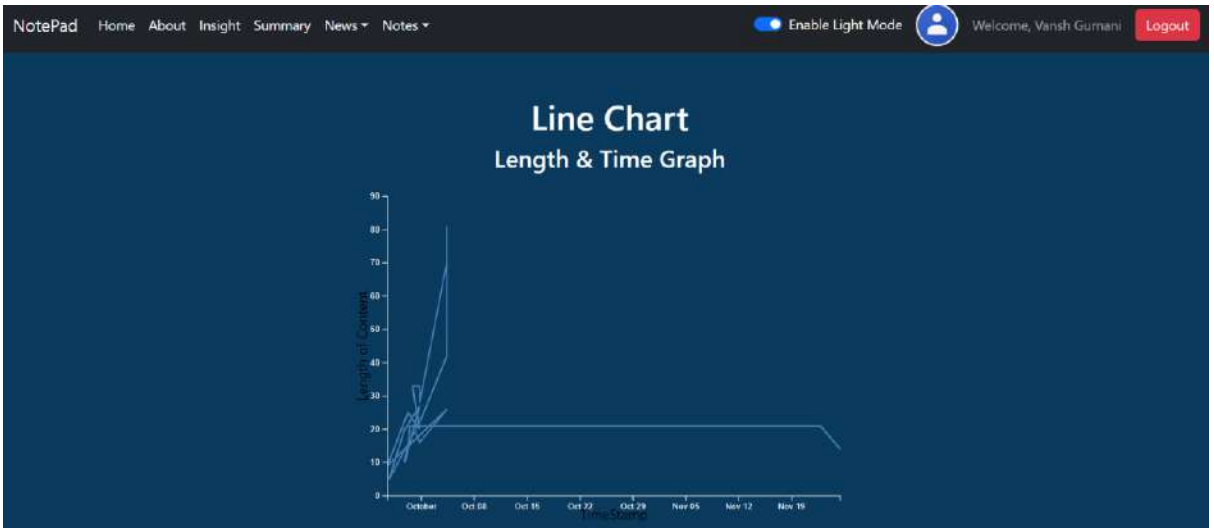

Summary:
Because the current wave of illness being reported from China is also causing respiratory symptoms mainly, all states and UTs have been advised to implement 'Operational Guidelines for Revised Surveillance Strategy in the context of Covid-19', shared earlier this year, which provides for integrated surveillance of respiratory pathogens presenting as cases of influenza like illness (ILI) and severe acute respiratory illness (SARI). The Centre has asked all states and Union territories to immediately review public health and hospital preparedness in the wake of a rise in respiratory illness in children in China. States have also been asked to send nasal and throat swab samples of patients with SARI, particularly of children and adolescents, to the virus research and diagnostic laboratories (VRDLs) in the states for testing for respiratory pathogens, said an official. The Centre has asked states & UTs to review the availability of manpower, hospital beds, drugs and vaccines for influenza and medical oxygen to deal with a potential surge in respiratory illness. Some reports suggest the surge in respiratory illness in China is being caused by the co-existence of multiple viral infections, for example influenza, respiratory syncytial virus and SARS-CoV2.
Accuracy:
39.86%

Dark Mode :-


NotePad
Home
About
Insight
Summary
News
Notes
Enable Light Mode
Welcome, Vansh Gurnani
Logout
Search notes...
Fetch all Notes
Welcome to NotePad
Enter the text
Category:
Uncategorized
Clear
Save Note
Character Count: 0

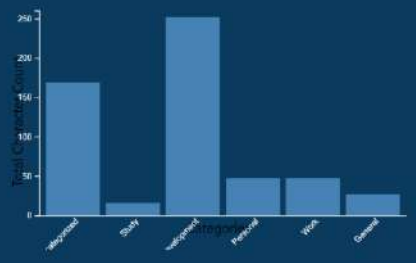

Search Results
1. hello good
....Read More
Category: uncategorized
Timestamp: 30/09/2023, 12:25:12
2. hello stud
....Read More
Category: Study
Timestamp: 30/09/2023, 23:32:47
3. Monolithic
....Read More
Category: Development
Timestamp: 04/10/2023, 14:16:26
4. hello 121
Category: uncategorized
Timestamp: 26/09/2023, 18:27:14
5. hello book
....Read More
Category: uncategorized
Timestamp: 29/09/2023, 11:25:08
6. hello pers
....Read More
Category: Personal
Timestamp: 30/09/2023, 23:35:35


NotePad
Home
About
Insight
Summary
News
Notes
Enable Light Mode
Welcome, Vansh Gumani
Logout
About Us
Vansh Gumani
Yuvraj Singh Lamba
Jay Mehta
Keshav Jha


NotePad
Home
About
Insight
Summary
News
Notes
Enable Light Mode
Welcome, Vansh Gumani
Logout
Line Chart
Length & Time Graph
Length of Content
Time Stamp
October
Oct 08
Oct 15
Oct 22
Oct 29
Nov 05
Nov 12
Nov 19

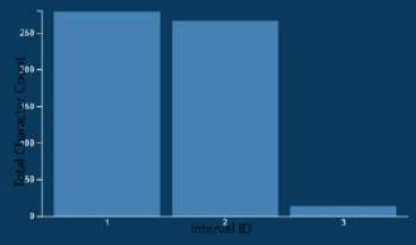
Bar Chart

Category Graph



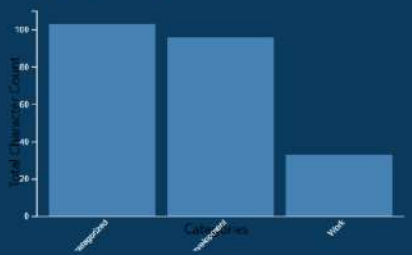
Bar Graph

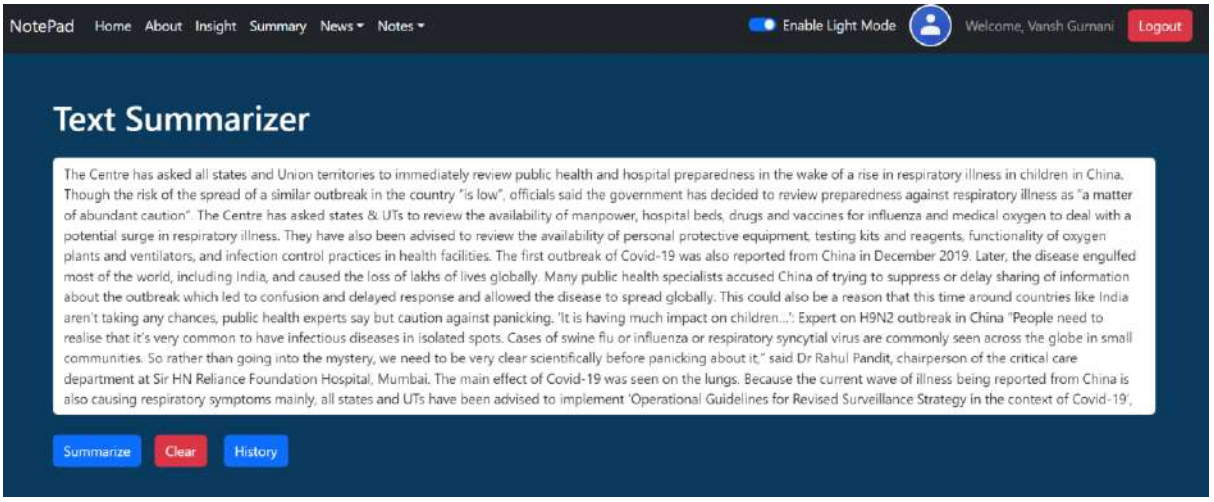
Total Character Count for Each 10-Note Interval

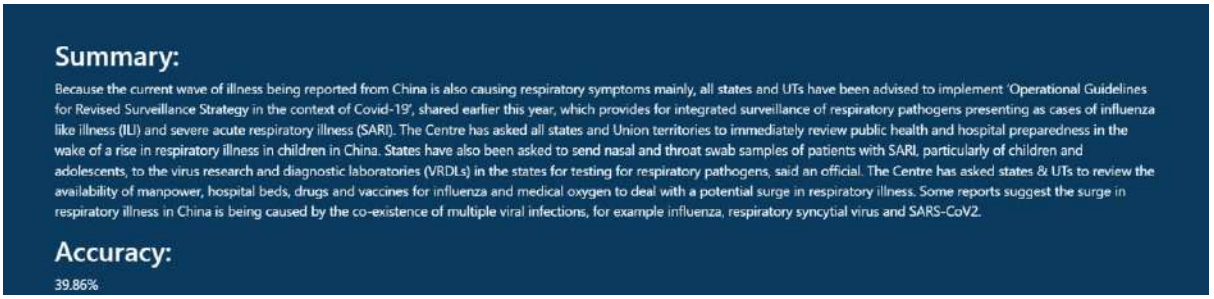


Bookmark Graph

Category Graph for Bookmarked Notes




NotePad
Home
About
Insight
Summary
News
Notes
Enable Light Mode
Welcome, Vansh Gurnani
Logout
Text Summarizer
The Centre has asked all states and Union territories to immediately review public health and hospital preparedness in the wake of a rise in respiratory illness in children in China. Though the risk of the spread of a similar outbreak in the country "is low", officials said the government has decided to review preparedness against respiratory illness as "a matter of abundant caution". The Centre has asked states & UTs to review the availability of manpower, hospital beds, drugs and vaccines for influenza and medical oxygen to deal with a potential surge in respiratory illness. They have also been advised to review the availability of personal protective equipment, testing kits and reagents, functionality of oxygen plants and ventilators, and infection control practices in health facilities. The first outbreak of Covid-19 was also reported from China in December 2019. Later, the disease engulfed most of the world, including India, and caused the loss of lakhs of lives globally. Many public health specialists accused China of trying to suppress or delay sharing of information about the outbreak which led to confusion and delayed response and allowed the disease to spread globally. This could also be a reason that this time around countries like India aren't taking any chances, public health experts say but caution against panicking. 'It is having much impact on children...': Expert on H9N2 outbreak in China "People need to realise that it's very common to have infectious diseases in isolated spots. Cases of swine flu or influenza or respiratory syncytial virus are commonly seen across the globe in small communities. So rather than going into the mystery, we need to be very clear scientifically before panicking about it," said Dr Rahul Pandit, chairperson of the critical care department at Sir HN Reliance Foundation Hospital, Mumbai. The main effect of Covid-19 was seen on the lungs. Because the current wave of illness being reported from China is also causing respiratory symptoms mainly, all states and UTs have been advised to implement 'Operational Guidelines for Revised Surveillance Strategy in the context of Covid-19',
Summarize
Clear
History


Summary:
Because the current wave of illness being reported from China is also causing respiratory symptoms mainly, all states and UTs have been advised to implement 'Operational Guidelines for Revised Surveillance Strategy in the context of Covid-19', shared earlier this year, which provides for integrated surveillance of respiratory pathogens presenting as cases of influenza like illness (ILI) and severe acute respiratory illness (SARI). The Centre has asked all states and Union territories to immediately review public health and hospital preparedness in the wake of a rise in respiratory illness in children in China. States have also been asked to send nasal and throat swab samples of patients with SARI, particularly of children and adolescents, to the virus research and diagnostic laboratories (VRDLs) in the states for testing for respiratory pathogens, said an official. The Centre has asked states & UTs to review the availability of manpower, hospital beds, drugs and vaccines for influenza and medical oxygen to deal with a potential surge in respiratory illness. Some reports suggest the surge in respiratory illness in China is being caused by the co-existence of multiple viral infections, for example influenza, respiratory syncytial virus and SARS-CoV2.
Accuracy:
39.86%