**Summer Internship Report**



**STOCK MARKET ANALYSIS**

Name : Simran Chauhan

Roll No: 20CSU109

Date : 24.07.2022

Table of Contents

|  |  |  |
| --- | --- | --- |
| S.no | Topics | Page no. |
| 1. | Abstract | 5-6 |
| 2. | Introduction | 7 |
| 3. | Training description | 8-9 |
| 4. | Analysis | 10-11 |
| 5. | Bibliography | 12 |
| 6. | Project daily task | 13-15 |

***Abstract*** (*should be less than one page*)

Describe where your training was carried out (company, department, location, etc.) Summarize your training goals, activities, and accomplishments. In addition, describe the key elements you took away from your training. These should be both technical and non-technical (for example, you might describe such things as how your organization was structured, how individuals worked together in your company, how technical employees interacted with customers, what company initiatives you observed, etc.)

* In the training of big data we learnt about tools required to deal with with big data and to analyse it. We learnt about Hadoop Ecosystem and its architecture

and Configuration, Design & Role in Hadoop.

We also learnt about MapReduce real world Applications, Mapper and Reducer,Various terminologies in MR like Input Format, Output Format, , Shuffle, and Sort.

We learned how to deal with real time data using pyspark and its Components of Spark Unified stack, Spark Architecture,,Resilient Distributed Dataset (RDD)

Create Parallelized collection & external Datasets. Various operations of RDD.

We learned types of machine learning : supervised,unsupervised and  Various ML algorithms supported by MLlib, Linear regression, logistic regression, decision tree, random forest, and K-means clustering techniques.

Also the faculty was cooperetaive , I want to thank them for helping us whenever we reached out for them. We learned things important things from industrial point of view.

**1. INTRODUCTION**

Describe the company with whom your training was carried out and the background to your involvement with this company, including the location of the facility at which the work was done. Describe your position and the basic engineering areas in which you worked for your training.

* We had training in NorthCap University .We were provided with well equipped lab to perform our project . We worked amateur in the field of big data analystics.

**2 – 4. TRAINING DESCRIPTION**

Describe in detail the nature of the work or the nature of the project, which you carried out, stressing as much as possible the engineering aspects of your work. Describe the goals for the project, the methods used to approach the project, and the accomplishments. Do not disclose any proprietary information.

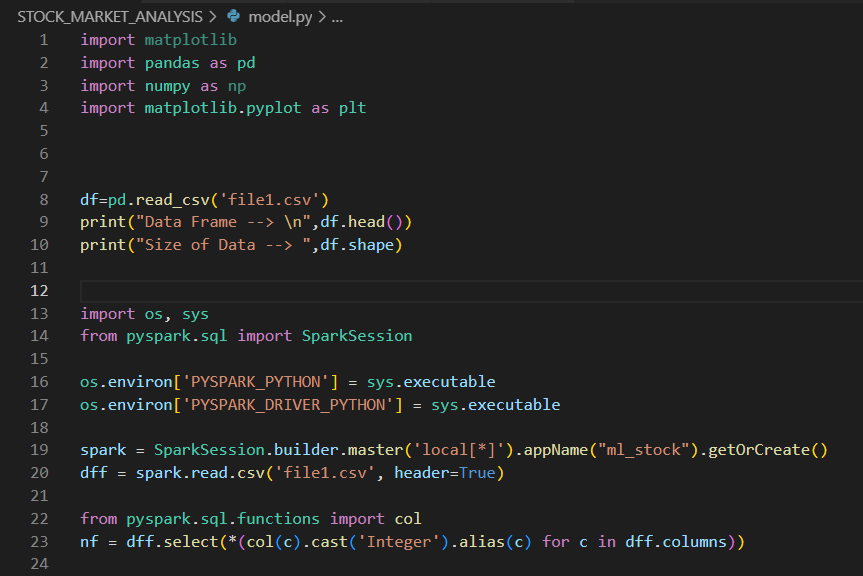
Detail the various stages of your training on a project basis. You should describe all the technical and administrative activities performed. Relevant tables and descriptive figures or pictures are encouraged, but should not be excessive. Samples of the plans or design and calculation sheets can only be included in an appendix section and must be properly referenced in the text. Materials photocopied or duplicated from the company’s previous proposals or reports should not be included.

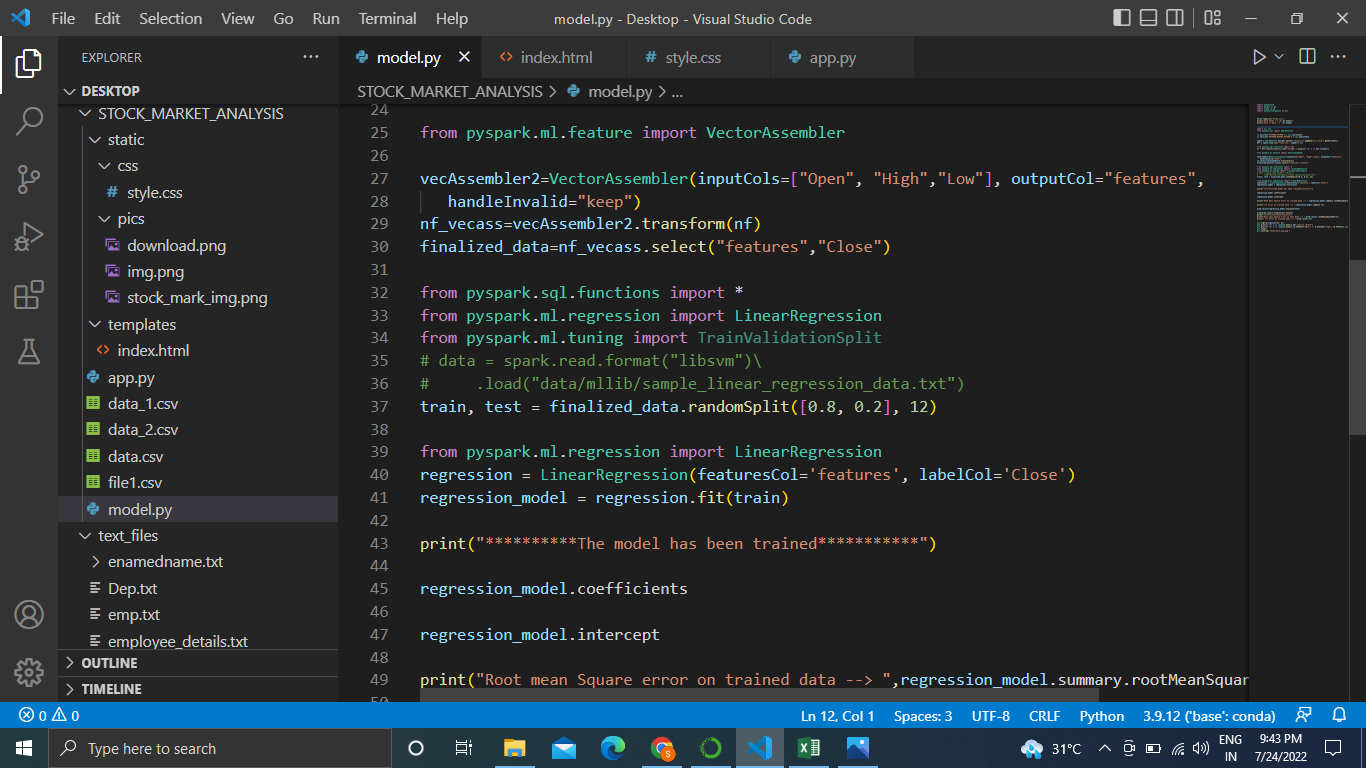
🡪 We worked on Stock Market Analysis prediction . Where we used Pyspark Mlib to evaluate over features to predict the closing stock prices. We used Flask along with HTML and CSS to deploy our model. We took our data from NIFTY 50 website of past 22 years.

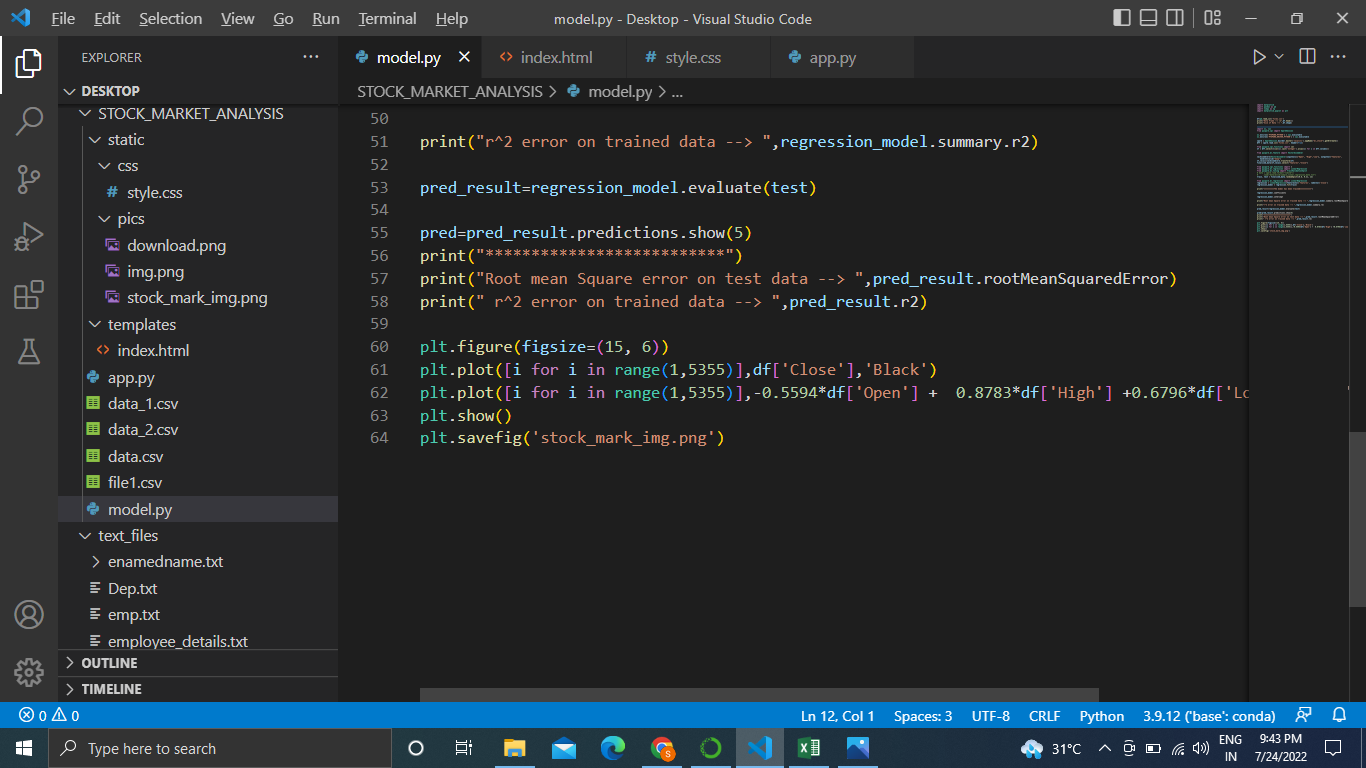
We analysed the trend of stock market .

We had Four components in our project as follow

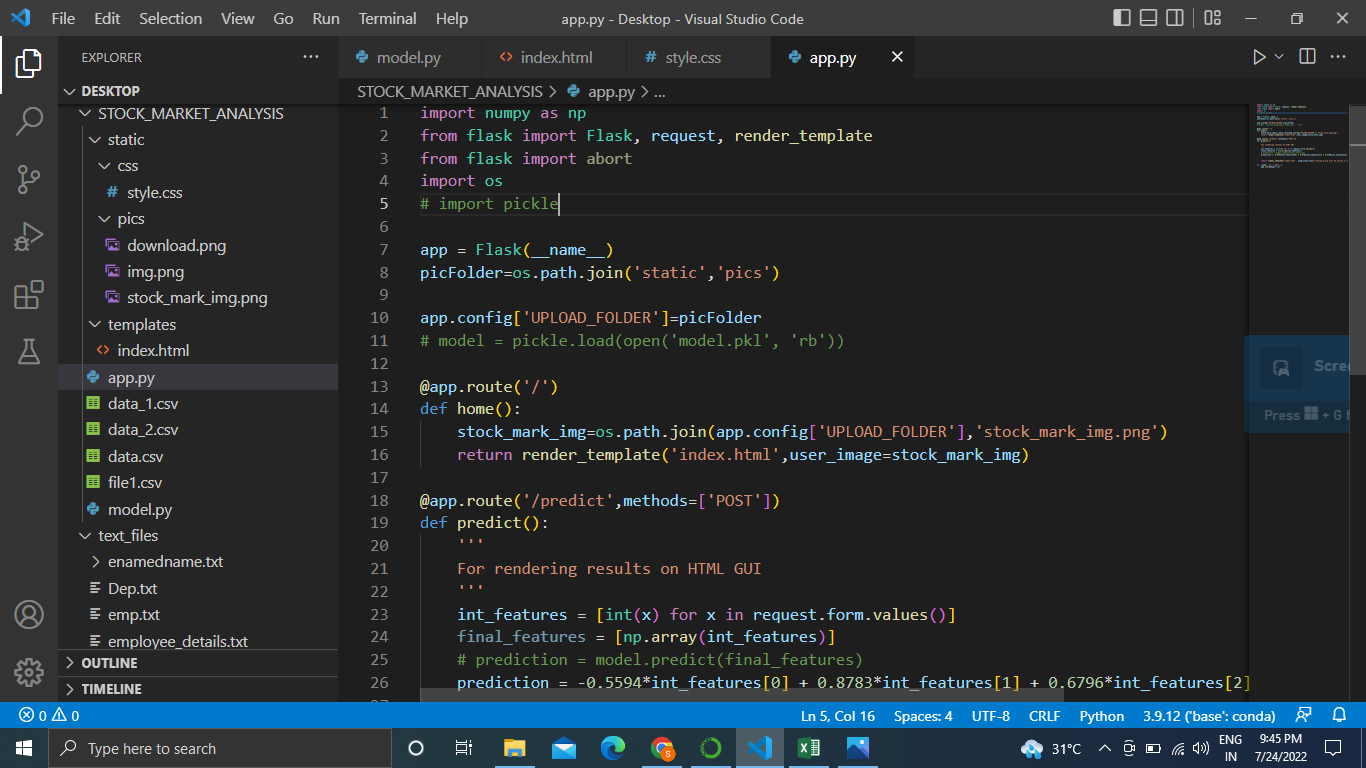
1. Model.py

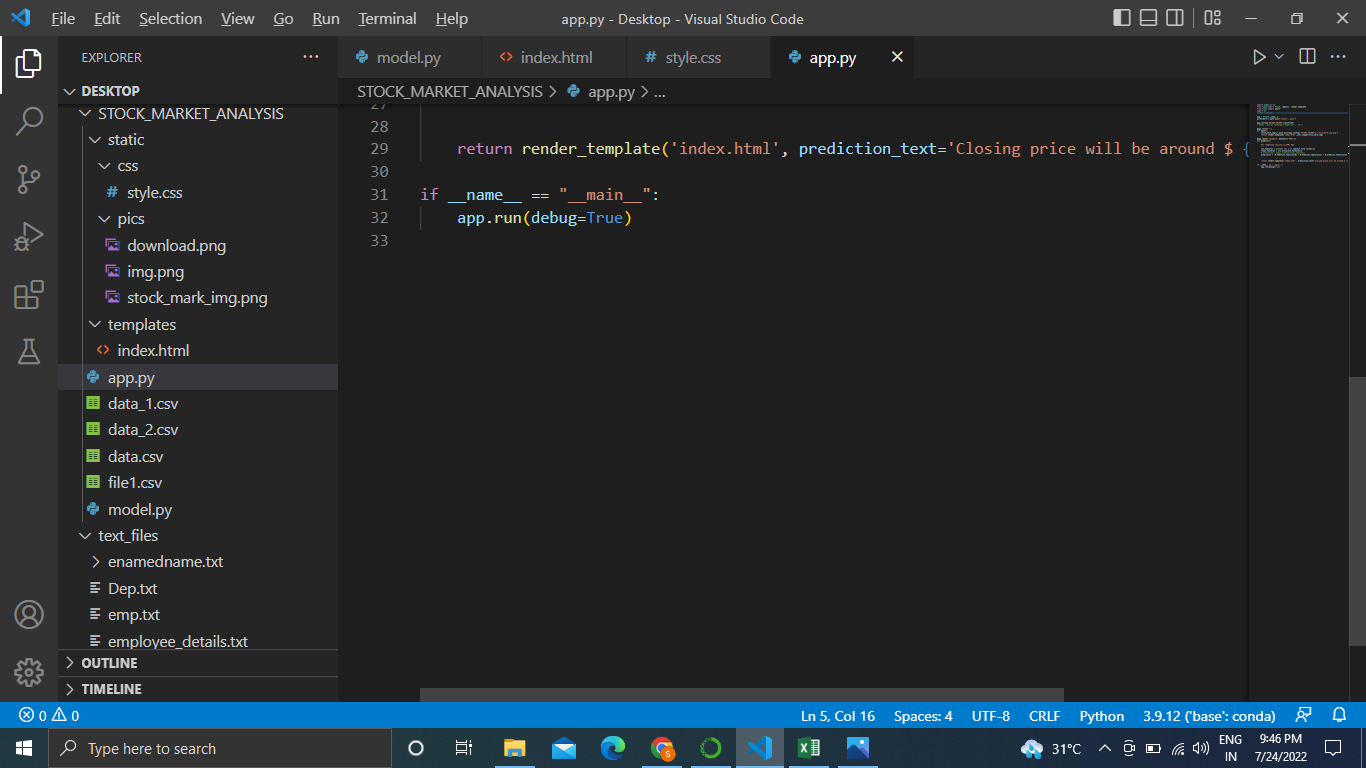




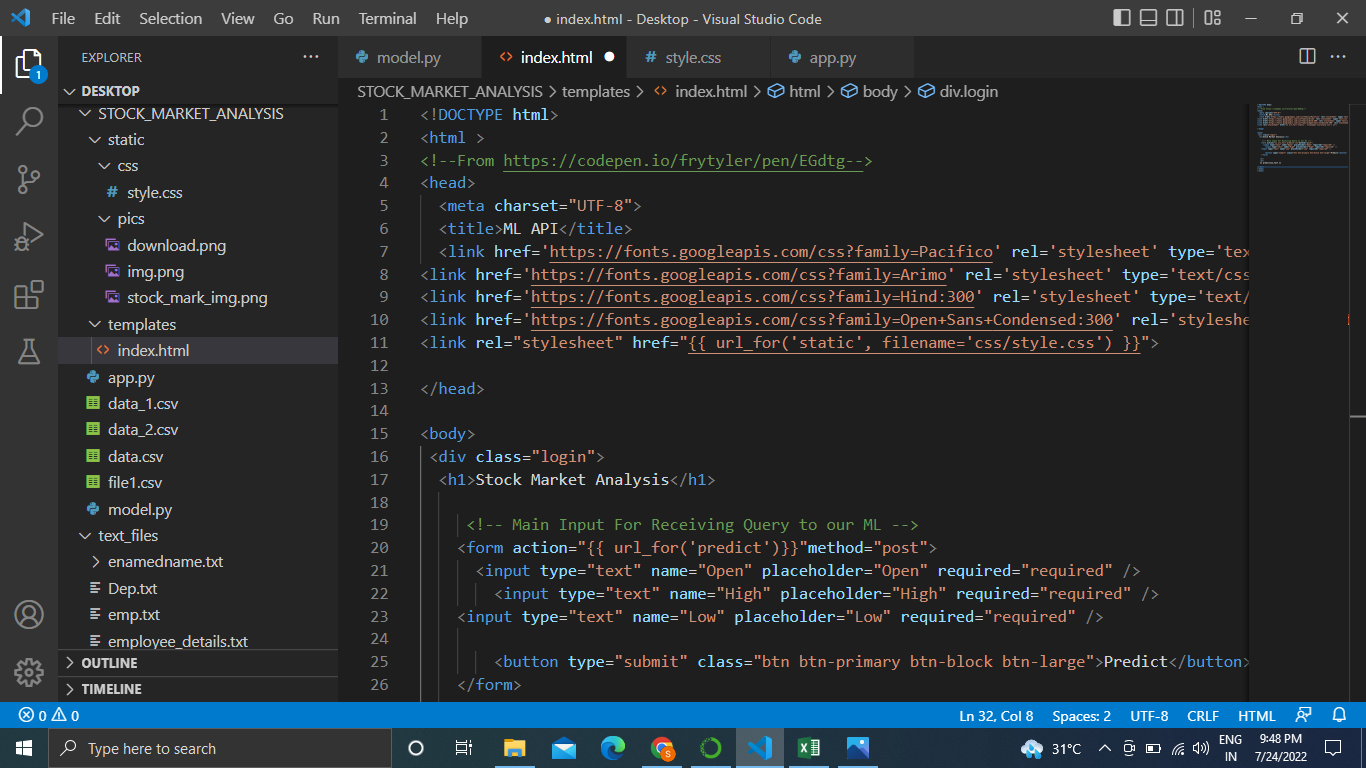


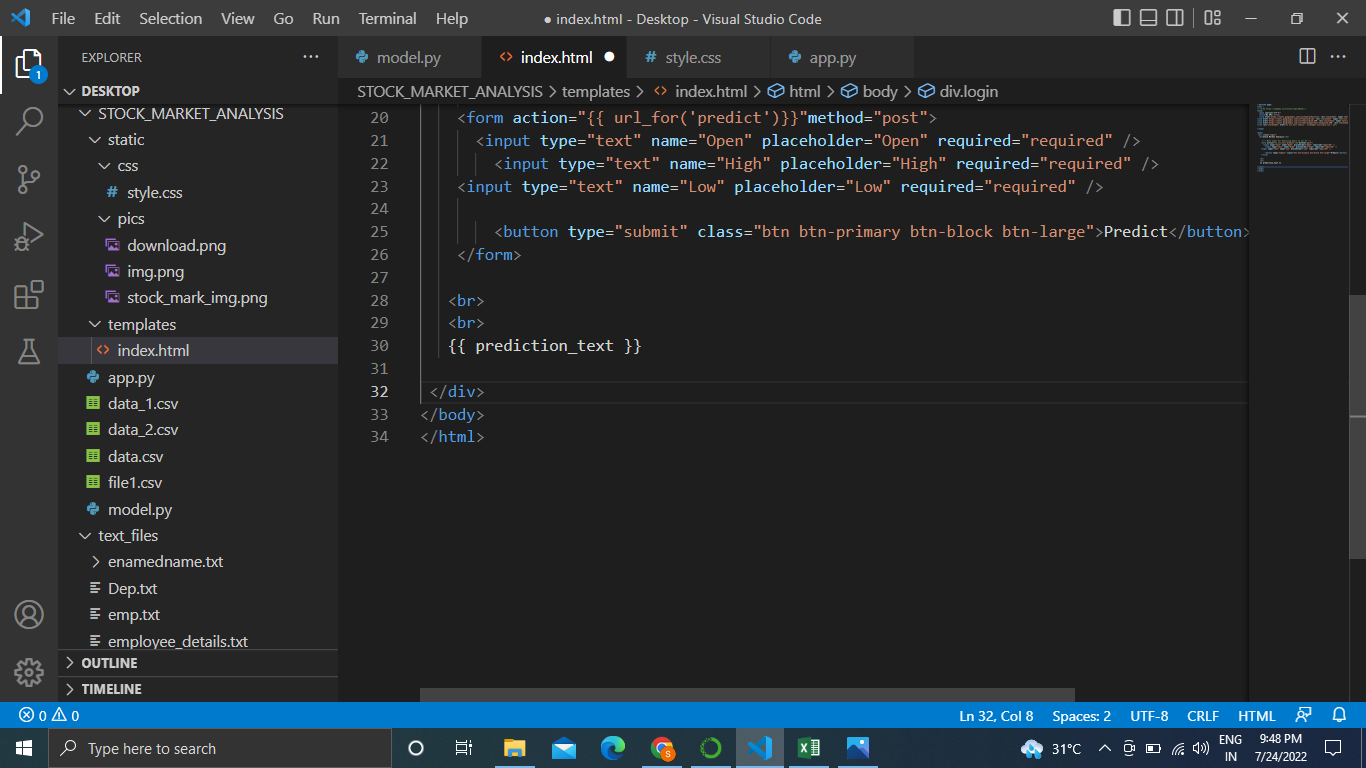
1. app.py



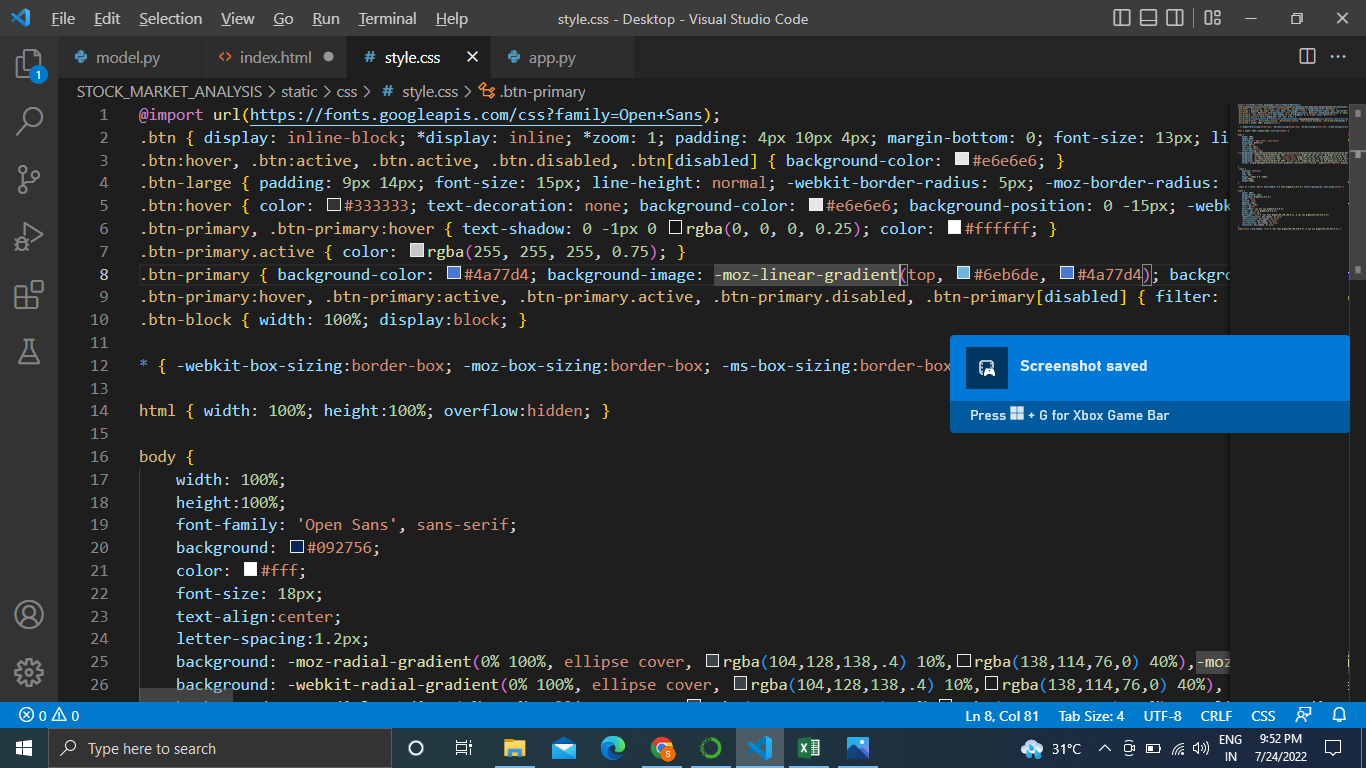


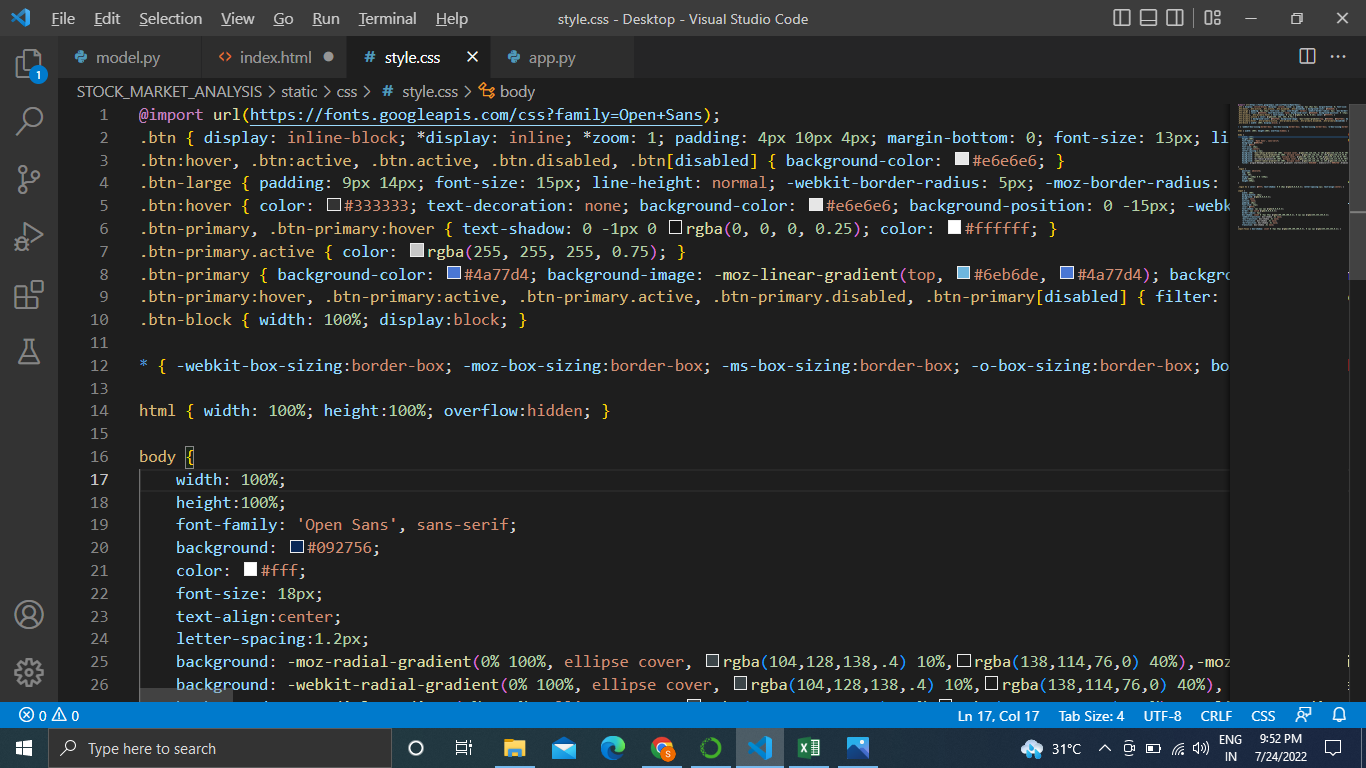
3. index.html

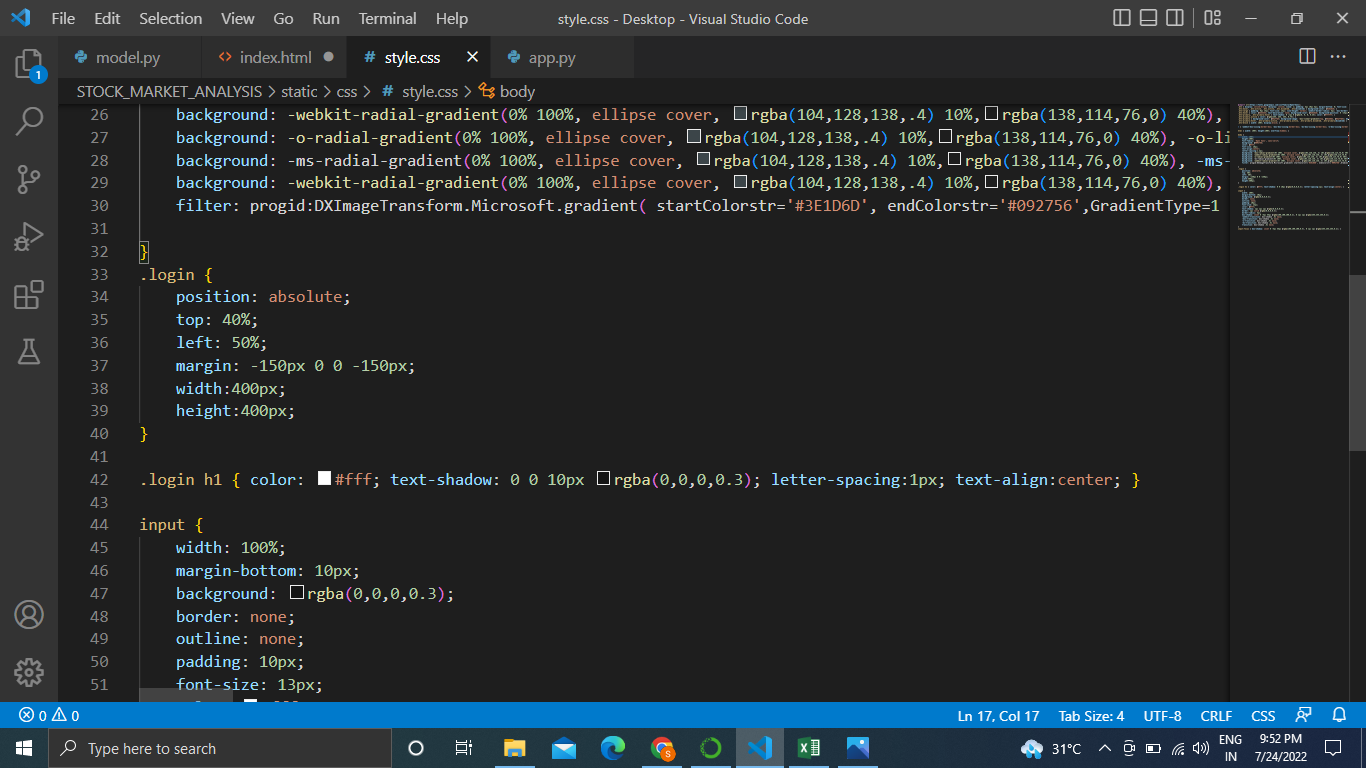




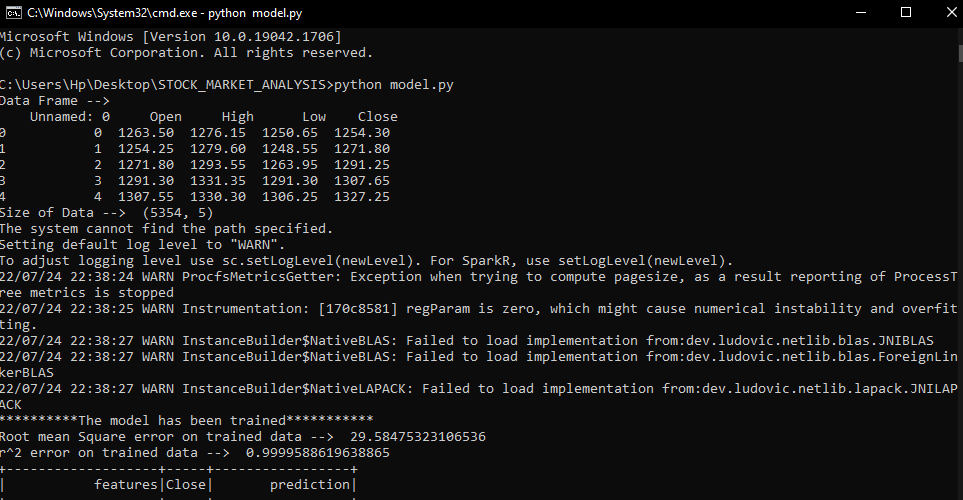
4.Style.css

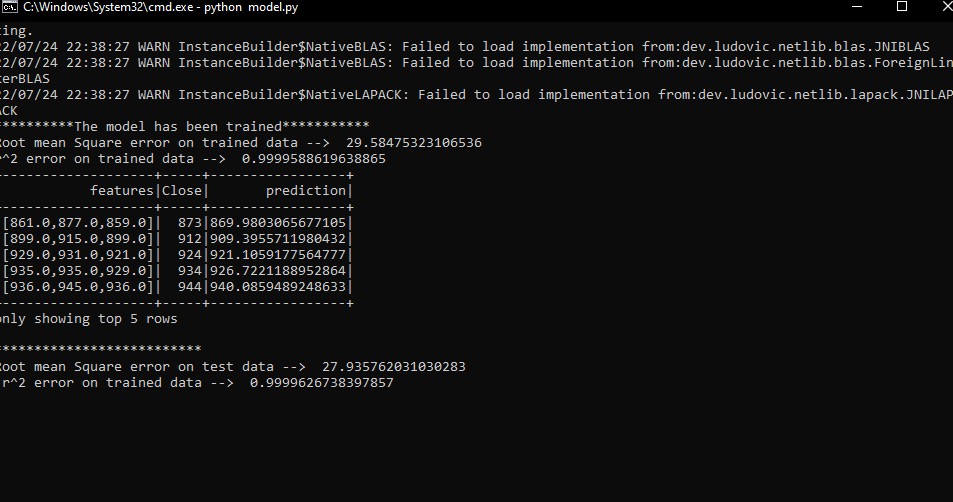


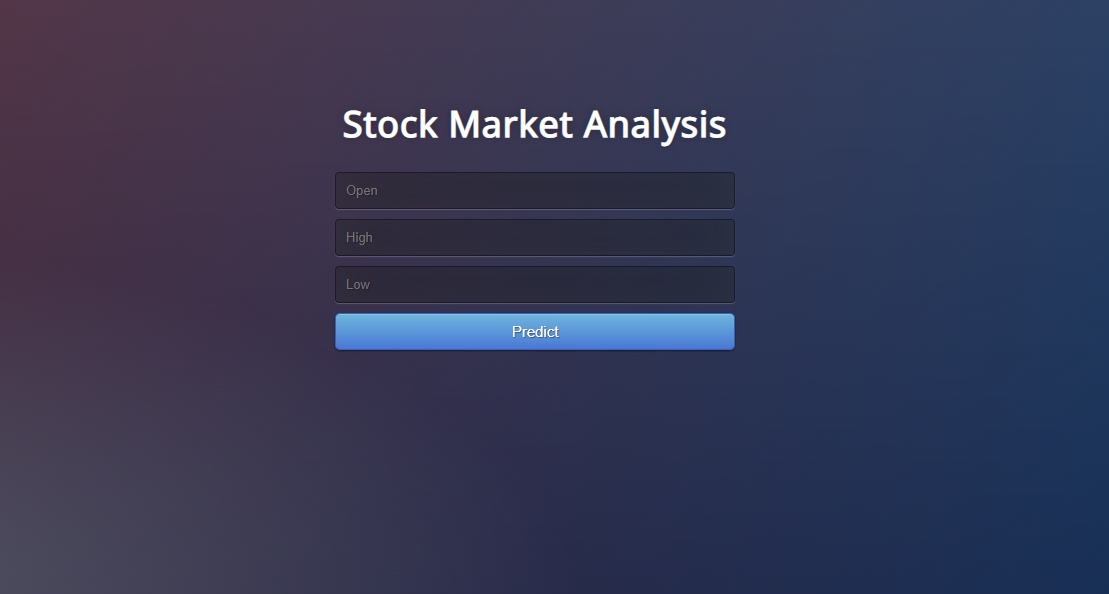


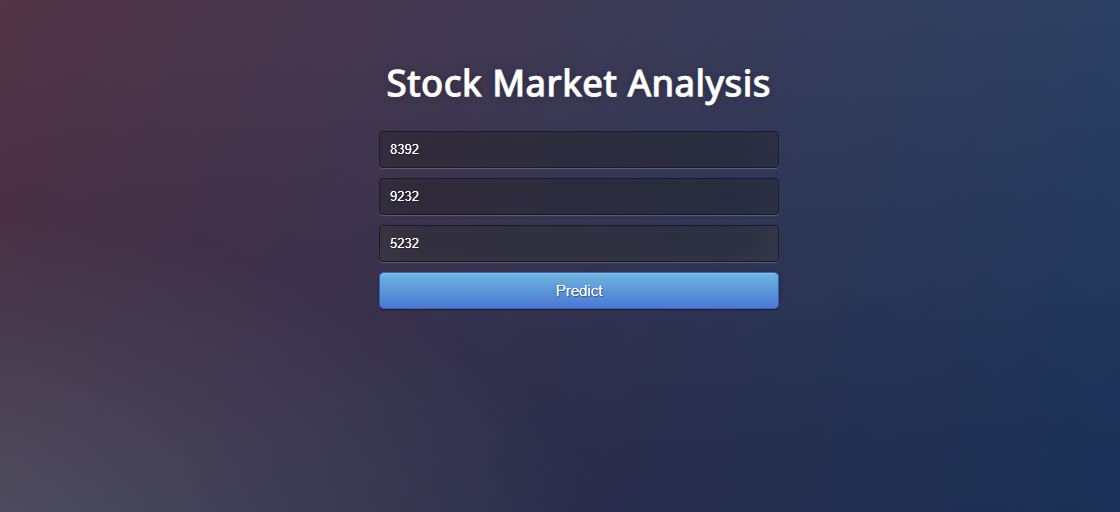


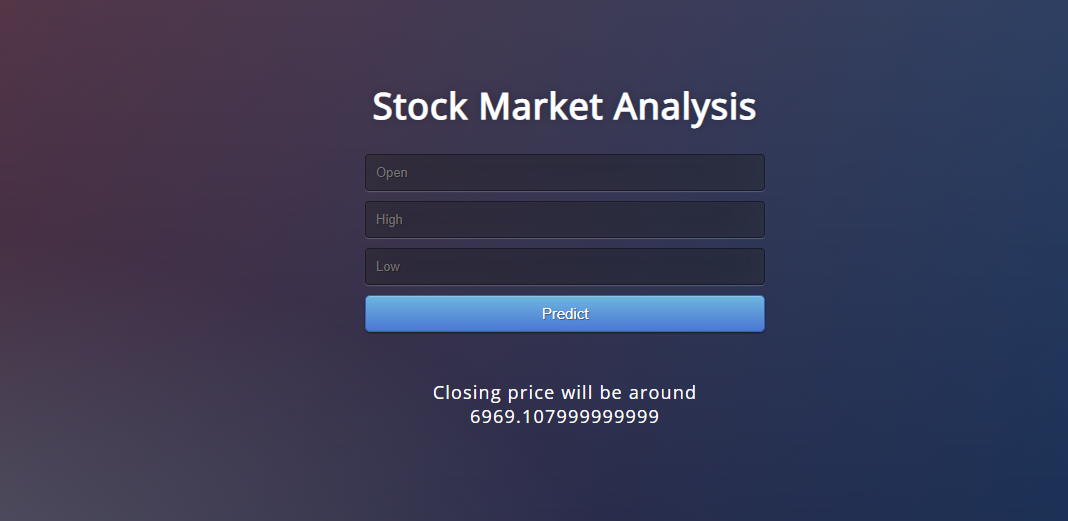
OUTPUT 🡪











* In the training of big data we learnt about tools required to deal with with big data and to analyse it. We learnt about Hadoop Ecosystem and its architecture

and Configuration, Design & Role in Hadoop.

We also learnt about MapReduce real world Applications, Mapper and Reducer,Various terminologies in MR like Input Format, Output Format, , Shuffle, and Sort.

We learned how to deal with real time data using pyspark and its Components of Spark Unified stack, Spark Architecture,,Resilient Distributed Dataset (RDD)

Create Parallelized collection & external Datasets. Various operations of RDD.

We learned types of machine learning : supervised,unsupervised and  Various ML algorithms supported by MLlib, Linear regression, logistic regression, decision tree, random forest, and K-means clustering techniques.

**5. ANALYSIS**

Describe the usefulness of your training work in advancing your understanding of the engineering profession. What did you learn from this experience? Include both the technical and non-technical elements (for example, you might describe such things as how your organization was structured, how individuals worked together in your company, how technical employees interacted with customers, what company initiatives you observed, etc.)

Analyze your performance during the training. What are your primary strengths? Give examples of what you did well. What are the areas that you would like to improve? Give examples of what you would do differently. Ask your supervisor for input on this aspect of the report.

🡪 I had a great great experience in learning new things,new tools , new ways of analyzing and fetching data.

I learned Hadoop Ecosystem and its architecture , how different tools work together. I learned how we can utilize different machines,devices as one cluster to store our enormous data.

My primary strengths was in machine learning during the training.

**6. CONCLUSION**

Summarize the usefulness of your summer training in advancing your understanding of the engineering profession, both technically and organizationally. Summarize what you learned from the experience. Discuss the benefits acquired from the training experience and the ways in which it enriched your knowledge. Any deficiencies in your education and suggestions for improvements of the training program should also be mentioned.

* As more and more data is generated and collected, data analysis requires scalable, flexible, and high performing tools to provide insights in a timely fashion. However, organizations are facing a growing big data ecosystem where new tools emerge and become outdated very quickly. Therefore, it can be very difficult to keep pace and choose the right tools.
* It a first step to help you solve this challenge. With a broad set of managed services to collect, process, and analyze big data, it makes it easier to build, deploy, and scale big data applications. This enables you to focus on business problems instead of updating and managing these tools.
* This training provided us many solutions to address your big data analytic requirements. Most big data architecture solutions use multiple tools to build a complete solution. This approach helps meet stringent business requirements in the most cost-optimized, performant, and resilient way possible. The result is a flexible big data architecture that is able to scale along with your business.

**BIBLIOGRAPHY**

You must give credit to the ideas or words of another person. You do this by making a bibliography, which is a separate page at the end of your report that lists all the resources you used. A bibliography is a list of every source which you have used in preparing your project. The IEEE format is recommended.

PROJECT DAILY WORK

|  |  |  |
| --- | --- | --- |
| S.no | Topic | Date |
| 1. | Data Analysis with Pyspark | 18.07.2022 |
| 2. | Flask Code | 19.07.2022 |
| 3. | Html file | 20.07.2022 |
| 4. | Css file | 21.07.2022 |
| 5. | Presention | 22.07.2022 |
| 6. | Report Writing | 23.07.2022 |

# 

