VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"Jnana Sangama", Belgavi-590 018, Karnataka, India

An Internship Report
On

"Data Science and Analytics using python & R"

Along with the project

"RESTAURANT REVIEW PREDICTION ANALYSIS"

Submitted in Partial Fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By

Name: SUBHASH K V
USN: 1SJ18CS101

Under the guidance of

Internal Guide
Dr. Manjunatha Kumar BH
Professor & Head of the Department
Dept. Of CSE, SJCIT

External Guide
Mr. Lakshman annamala
Mentor & Trainer
Verzeo (CodingZen)

S J C INSTITUTE OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CHIKKABALLAPUR-562101

2021-2022

||Jai Sri Gurudev||

Sri Adichunchanagiri Shikshana Trust®

S.J.C INSTITUTE OF TECHNOLOGY, Chickballapur - 562101

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Internship work entitled **"Machine Lerning with Python" Along with the project "RESTAURANT REVIEW PREDICTION ANALYSIS"** carried out by **SUBHASH K V** bearing USN:1SJ18CS101 a bonafide student of Sri Jagadguru Chandrashekaranatha Institute of Technology in partial fulfilment for the award of **Bachelor of Engineering in Computer Science and Engineering** of **Visvesvaraya Technological University, Belgaum** during the year **2021-22**. It is certificated that all corrections / suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The Internship report has been approved as it satisfies theacademic requirements in respect of Internship work prescribed for the said Degree.

11/7/2022

Signature of Guide
Dr.Vikas Reddy S
Associate Professor
Dept. of CSE, SJCIT

12/7/22

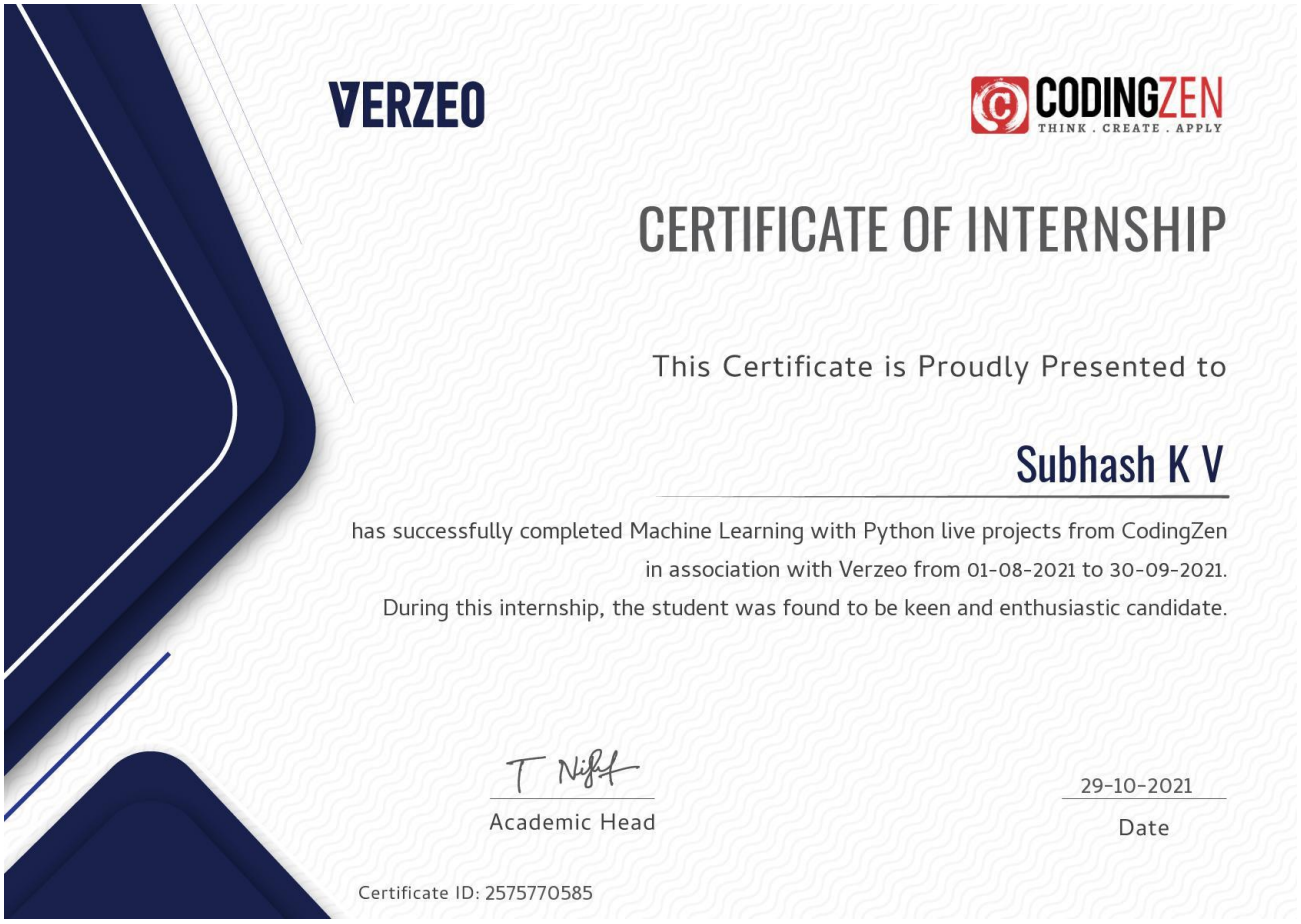
Signature of HOD
Dr. Manjunatha Kumar B H
Professor & HOD,
Dept. of CSE, SJCIT

Signature of Principal
Dr. G T Raju
Principal, SJCIT

External Examiners:

Name of the Examiners	Signature with Date
1. Divakar KM	25/7/22
2. Dr. Sushachan	

COMPANY CERTIFICATE



DECLARATION

I, **SUBHASH K V**, student of VIII semester B.E in Computer science & Engineering at S J C Institute of Technology, Chickballapur, hereby declare that the Internship work entitled "RESTAURANT REVIEW PREDICTION ANALYSIS" has been independently carried out by me under the supervision of **Dr.Vikas Reddy S**, Associate Professor of Department of CSE, and the coordinator **Mr. Narendra Babu C** Assistant Professor, submitted in partial fulfillment of the course requirement for the award ofdegree in **Bachelor of Engineering** in **Computer Science & Engineering** of **Visveswaraya Technological University**, **Belgavi** during the year 2021-2022. I further declare that the report has not been submitted to any other University for the award of any other degree.

PLACE: CHICKBALLAPUR
Date: 13 May 2022

STUDENT NAME: SUBHASH K V
USN : 1SJ18CS101

ABSTRACT

Restaurant Review has become the most commonly used parameter for judging a restaurant for any individual. A lot of research has been done on different restaurants and the quality of food it serves. Reviewing of a restaurant depends on factors like area situated, average cost for two people, votes, cuisines, mainly taste they serve and the type of restaurant. The main goal of this is to get insights on restaurants which people like visit and to identify the review of the restaurant.

The purpose of this analysis is to build a prediction model to predict whether a review on the restaurant is positive or negative. To do so, we will work on Restaurant Review dataset, we will load it into predicitve algorithms Multinomial Naive Bayes, SVC, XGB Regressor, Pipeline and Logistic Regression. In the end, we hope to find a "best" model for predicting the review's sentiment.

ACKNOWLEDGEMENT

With reverential pranam, we express my sincere gratitude and salutations to the feet of his holiness **Byravaikya Padmabhushana Sri Sri Sri Dr. Balagangadharanatha Maha Swamiji,** & his holiness **Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Swamiji** of Sri Adichunchanagiri Mutt for their unlimited blessings. First and foremost, we wish to express my deep sincere feelings of gratitude to our institution, **Sri Jagadguru Chandrashekaranatha Swamiji Institute of Technology.** For providing me an opportunity for completing my internship work successfully.

I extend deep sense of sincere gratitude to **Dr. G T Raju**, **Principal, S J C Institute of Technology, Chickballapur**, for providing an opportunity to complete the Internship Work.

I extend special in-depth, heartfelt, and sincere gratitude to our HOD **Dr. Manjunatha Kumar BH, Professor and Head of the Department, Computer Science and Engineering, S J CInstitute of Technology, Chickballapur,** for his constant support and valuable guidance of the Internship Work.

I convey our sincere thanks to Internship Internal Guide **Prof. Dr.Vikas Reddy S, Assistant Professor and Head of the Department, Department of Computer Science and Engineering, S J C Institute of Technology,** for his constant support, valuable guidance and suggestions of the Internship Work.

I am thankful to Internship External Guide **Mr. Pranav jaipurkar, Knowledge Solutions India,** for providing valuable guidance and encouragement of the Internship Work.

I also feel immense pleasure to express deep and profound gratitude to our Internship Coordinator **Narendra Babu, Assistant Professor, Department of Computer Science and Engineering, S J C Institute of Technology,** for his guidance and suggestions of the Internship Work.

Finally, I would like to thank all faculty members of Department of Computer Science and Engineering, S J C Institute of Technology, Chickballapur for their support.

I also thank all those who extended their support and co-operation while bringing out this Internship Report.

Subhash k v(1SJ18CS101)

CONTENTS

LIST OF FIGURES

CHAPTER - 1

COMPANY PROFILE

1.1 History of the Organization

Witnessing the current times, we can come to this conclusion that online education is everywhere. There are numerous options for online training. We being a responsible entity realize that it has become significant for the careerists to know about the right place that will help them achieve their dreams, to feel the exhilaration of victory!

Our CEO, V.V Subrahmanyam founded Verzeo, in 2018. He aims to train students to make them industry-ready. He believes that to savour each aspirant of the country with the taste of good mentorship, it's necessary to bridge the gap between technology and education.

We have come up with a variety of courses ranging from Kids Programs, Job-Guarantee Programs, and Pro-Degree Programs packed with live projects and interactive sessions. We also provide Banking & CA training, along with technical programs. Our aim is to provide learning aids in a broad spectrum, so that students from various fields can rely on our one-stop online learning solution, Verzeo.

With more than 900 employees on board, the CEO aims to hit the company's valuation of 500 crores by the end of 2022. We, the Verzeo family, work day-in and day-out to achieve our CEO's target of shaping the future of millions of young minds.
At Verzeo, learning is not limited to any specific domain; we provide our students with immense networking opportunities with industry professionals to expand their horizons of growth and development.

1.1.1 Objectives

Their goal is to consistently deliver success to students by going the extra mile. To help their students meet their technological skills and career opportunities, they offer the right people, solutions, and services.
By leveraging leading technologies and industry best practices, they provide their students with the most efficient and effective training.

1.1.2 Operation of the Organization

The race for digital transformation is on. In this globally connected on-demand world with rapid advancements in internet technologies, businesses worldwide are under constant pressure to add innovative real-time capabilities to their applications to respond to market opportunities.

Every business worldwide is building event-driven, real-time applications - from financial services, transportation, and energy, to retail, healthcare, and Gaming companies.
Our endeavor is to make it easy to develop innovative real-time applications and efficient to operate them in production.

We have a proven record of building highly scalable, world-class consulting processes that offer tremendous business advantages to our clients in the form of huge cost-benefits, definitive results and consistent project deliveries across the globe.

We prominently strive to improve your business by delivering the full range of competencies including operational performance, developing and applying business strategies to improve financial reports, defining strategic goals and measure andmanage those goals along with measuring and managing them.

1.2 Major Milestones

Over the course of the last 3 years, Verzeo has managed to make tremendous leaps in the eLearning sector and create a remarkable impact on the current Indian education dynamic.Since its inception, we have grown to specialize in 50+ departments and distribute our comprehensive courses and training programs in every part of the country.With our AI-backed platform, 150,000+ trained students.

1.3 Structure of the Organization

Our super energetic and massive team at KSI is our core strength, forming an excellent blend of IT minds with a creative bent. Their goal is to keep improving and delivering the skills that will help students have a successful career in the IT industry.

Taking advantage of our highly skilled and experienced trainers. We are primarily a student-centered organization dedicated to exceeding students' expectations in terms of meeting their needs. They successfully hosted a group of seasoned professionals.

Trainers who collaborate in order to provide their students with the knowledge they need to advance in their careers. They take pride in being a sought-after Skill development after delivering successful internships. They have successfully delivered value to our students as well as colleges over the years. They truly believe that the success of their students is their success, and they do not consider themselves to be a vendor for their program. We'd like to hear some of their stories and learn how far they've gone to ensure the success of our students, and they'll do everything they can to make that happen.

1.4Services Offered

Training / Internships form a very important part of students over all development that's why AICTE and Universities have made it mandatory for every engineer and MCA to undergo the same, we help students in achieving this goal by helping them acquire latest skills and provide them with hands on projects.

1. **Machine Learning and Internship Program.**

Learn Machine learning**,** an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed, bundled with Microsoft MTA Certification

2. **Data Science and Internship Program**

Learn Data science and how to use scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data as one of the hottest professions in the market today, bundled with Microsoft MTA Certification

3. **Java Certificate Program**

Learn Java one of the most popular programming languages used in the development of Web and Mobile applications. It is designed for flexibility, allowing developers to write code that would run on any machine, regardless of architecture or platform Bundled with Microsoft MTA Certification

4. **Cyber Security Certified Associate**

Learn the ethical way of how to do penetration testing and other testing methodologies that ensures the security of an organization’s information systems, bundled with Microsoft MTA Certification.

5. Internet of Things

Learn how to work with connected devices use sensors and raspberry PI3 and connect these devices to cloud to identify patterns and extract meaning-full information out of it, bundled with Microsoft MTA Certification

6. Business Analytics

Learn Business Analytics and how it enables companies to automate and optimize their business processes in-fact Data-driven companies treat their data as a corporate asset and leverage it for a competitive advantage as they are able to use the insights to find new patterns and relationships.

7. Digital Marketing

Learn Digital Marketing and how its used for promoting products or services online via internet, companies are gaining higher profitability and return on investment by having their Digital marketing strategies in place the program is bundled with Google Certification

CHAPTER – 2

ABOUT THE DEPARTMENT

2.1 Specific Functionalities of the Department

Our department of tech support majorly focused on manage, maintain and repair IT systems.

The Special functionalities include

- Understanding the work to be completed.
- Planning the assigned activities in more detail if needed
- Completing assigned work within the budget, timeline and quality expectations
- Informing the project manager of issues, scope changes, risk and quality concerns
- Proactively communicating status and managing expectation

2.2 Process Adopted

The department aims to first understand the user requirements. Further on, a basic structure of the product that needs to be built is drawn and understood. Eventually, the technologies that would best help in developing the product are understood. If the product has database requirements, the schema and the database design are worked upon. The department believes in “Think before you code”- the requirements and logics are first understood over a paper and then are moved to a code form. Agile processes generally promote a disciplined project management process that encourages frequent inspection and adaptation, a leadership philosophy that encourages teamwork, self-organization and accountability, a set of engineering best practices intended to allow for rapid delivery of high-quality software, and a business approach that aligns development with customer needs and company goals. Agile development refers to any development process that is aligned with the concepts of the Agile Manifesto. The Manifesto was developed by a group fourteen leading figures in the software industry, and reflects their experience of what approaches do and do not work for software development.

2.3 Testing

Testing was done according to the Corporate Standards. As each component was being built, Unit testing was performed in order to check if the desired functionality is obtained. Each component in turn is tested with multiple test cases to verify if it is properly working. These unit tested components are integrated with the existing built components and then integration testing is performed. Here again, multiple test cases are run to ensure the newly built component runs in co-ordination with the existing components. Unit and Integration testing are iteratively performed until the complete product is built. Once the complete product is built, it is again tested against multiple test cases and all the functionalities.

The product could be working fine in the developer's environment but might not necessarily work well in all other environments that the users could be using. Hence, the product is also tested under multiple environments (Various operating systems and devices). At every step, if a flaw is observed, the component is rebuilt to fix the bugs. This way, testing is done hierarchically and iteratively.

2.4 Structure of the Department

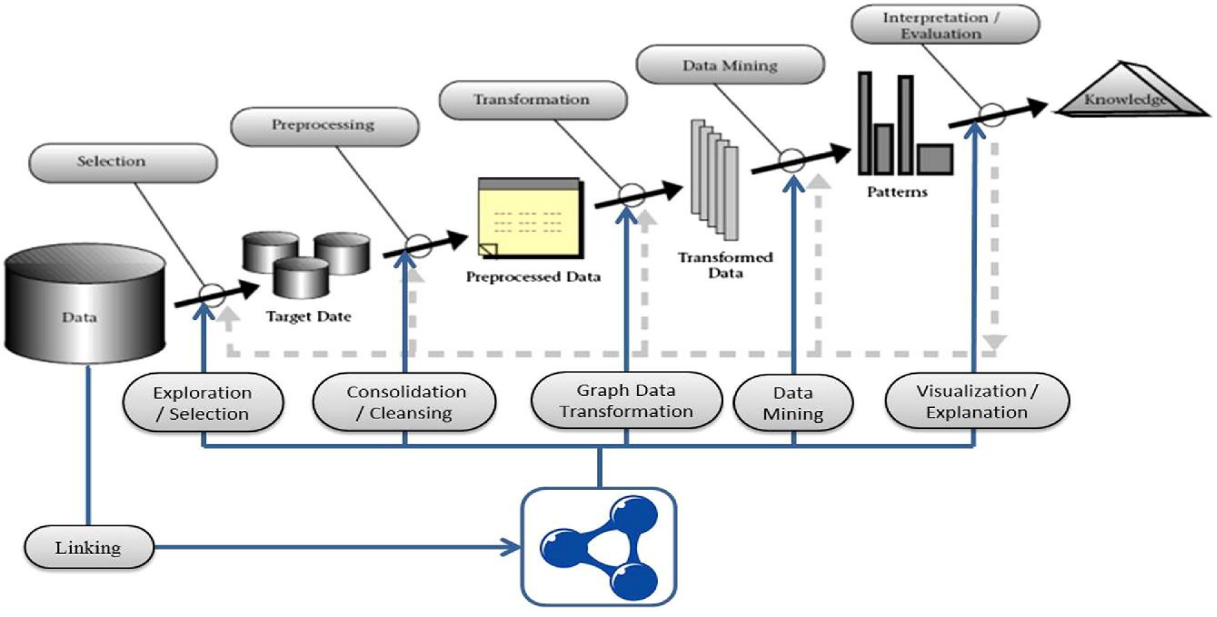


Figure 2.4.1 Structure of the Department

2.5 Roles and Responsibilities of Individuals

Since the internship was remotely conducted by the company, to ensure easy onboarding of interns, the company had individuals who took care of the smooth run of online training.

- Operation and Strategy Head- Ensured there were no difficulties for interns while onboarding. Best of mentors and doubt clarifying sessions were arranged too.
- Technical Lead- Ensured the technicalities of online training to be smooth. Best platforms were arranged for our meetings and trainings.
- Mentors- They have helped us to understand the concepts, gave us tasks to get practical take a way and clarified doubts to the best.
- Interns- Worked through the tasks given either individually or in a group

CHAPTER – 3

TASK PERFORMED

In this Internship Machine Learning with Python using MI it was a cource of making predictions using ml algorithms.

Training Program

The internship is a platform where the trainees are assigned with the specific task. In the initial days of the internship, I was trained on the following:

- Python Programming
- Machine Learning Algorithms

DATA SET

This section describes, in brief, the data that has been used for the research. Data from restourant was used in this project, the major amount of data was extracted from public website Kaggle (Kaggle.com), data regarding the review and linked was obtained from a leading Restaurant in India. Data from restaurant sources was integrated together to form a staging data-set. For predicting the review is either positive or negative which uses for the people to say that the which restaurant is best in class and it also uses for restaurant to improve there levelk of standarsds in their quality items either it may be the quality food , private space, surrounding of the place, etc.

Below table shows the different types of reviews present in the data-set.

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1
5	Now I am getting angry and I want my damn pho.	0
6	Honesity it didn't taste THAT fresh.)	0
7	The potatoes were like rubber and you could te...	0
8	The fries were great too.	1
9	A great touch.	1
10	Service was very prompt.	1
11	Would not go back.	0
12	The cashier had no care what so ever on what I...	0
13	I tried the Cape Cod ravoli chicken with cranb...	1
14	I was disgusted because I was pretty sure that...	0
15	I was shocked because no signs indicate cash o...	0
16	Highly recommended.	1

DATASET EXTRACTION AND TRANSFORMATION

Data related to the Restaurant review was collected in .csv format, the data related to review was extracted using data extraction tool provided by (Mozenda (n.d.)) in .csv files. Data being from public portal had multiple records which got mixing and irrelevant values; data cleaning was performed in Microsoft Excel by collecting all the records to a record having unwanted and missing values. Once the data-set was added to google colab Unwanted columns were left over there, and extracted only wanted (correctly organized) and then divided them into two parts that review and tlinked.1 and then the cleaned data was transformed to be suitable for the model. The original data-set had only the review as a representation of language, to have a consistent metrics for the language score that is either 0 or 1. Similarly, by undertaking the training and testing data we created a prediction model using SVC machine learning algorithm.

Algorithms

- Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.

- SVC

  The Linear Support Vector Classifier (SVC) method applies a linear kernel function to perform classification and it performs well with a large number of samples. If we compare it with the SVC model, the Linear SVC has additional parameters such as penalty normalization which applies 'L1' or 'L2' and loss function.

- Pipeline

- pipeline is a means of automating the machine learning workflow by enabling data to be transformed and correlated into a model that can then be analyzed to achieve outputs. This type of ML pipeline makes the process of inputting data into the ML model fully automated

- Another type of ML pipeline is the art of splitting up your machine learning workflows into independent, reusable, modular parts that can then be pipelined together to create models. This type of ML pipeline makes building models more efficient and simplified, cutting out redundant work.

- This goes hand-in-hand with the recent push for microservices architectures, branching off the main idea that by splitting your application into basic and siloed parts you can build more powerful software over time. Operating systems like Linux and Unix are also founded on this principle. Basic functions like ‘grep’ and ‘cat’ can create impressive functions when they are pipelined together.

In my two months Internship I have undergone through three phases:

- Training Phase
- Designing and Development Phase
- Testing and Maintenance Phase

As the final task, a main project was developed using machine learning models to predict the chance of a student to be admitted to a master’s program. This will assist students to know in advance if they have a chance to get accepted. This project predicts the admission of a student based on different features including university rating, student’s undergraduate GPA, GRE score, research experience and etc. This predicts that how much chances are there that the student will get admission in his selected university or not. In this project I have used multiple algorithms including linear regression, artificial neural network (ANN), random forest regressor, decision tree regressor. In the end I have deployed this model on a Web Based GUI to check student’s admission chances and these models are working fine.

CHAPTER – 4

REFLECTION NOTES

4.1 Experience

As per our experience during the internship, Verzeo India follows a good work culture and it has friendly employees, starting from the staff level to the management level. The trainers are well versed in their fields and they treat everyone equally. There is no distinguishing between fresher graduates and corporates and everyone is respected equally. There is a lot of teamwork followed in every task, be it hard or easy and there is a very calm and friendly atmosphere maintained at all times. There is a lot of scope for self-improvement due to the great communication and support that can be found. Interns have been treated and taught well and all our doubts and concerns regarding the training or the companies have been properly answered. All in all, Knowledge Solutions India was a great place for a fresher to start career and also for a corporate to boost his/her career. It has been a great experience to be an intern in such a reputed organization.

4.2 Technical Outcomes

4.2.1 System Requirements and Specification

HARDWARE REQUIREMENTS:

- Processor : x86 or x64
- Hard Disk : 216 GB or more.
- Ram : 512 MB (minimum), 1 GB(recommended)

SOFTWARE REQUIREMENTS:

- Operating System : Windows or Linux
- Development Environment : Anaconda Navigator (Jupiter Notebook or Spyder)

4.3 System Analysis and Design

4.3.1 Existing System

Used multiple machine learning models to create a system that would help the restaurant owner to get review that is either positive or negative by predicting using the given review. The secondly it helps the customers to get the best hotel near by his location by seeing the review. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables.

Review system was developed by (Waters and Miikkulainen (2013)) to support the visitors to get the best hotel . which are Categorical variables and for machine learning to model work we should input numerical values to perform. hence use Label Encoding on these 2 Features that encode Yes/No as 0/1**.** After Encoding split the Dataset to X and Y variables and again split to Train and Test sets of 70% and 30%. Apply Standardisation on Dataset as we have different scale ranges for different Features. Hence after applying Standard scaling it will bring all the values to a common range which is easy for model to compute and makes computation fast.. Logistic regression and SVC were used to create the model, both models performed equally well and the final system was developed using Logistic regression due to its simplicity. The time required by the admission committee to review the Restaurant was reduced by 74% but human intervention was required to make the final decision on status.

4.3.2 Disadvantages of the Existing System

- Limitation of this system only relied on the restaurant makes the restaurant to go down and it takes so much time as if they change their behaviour, quality, surroundings etc.
- The existing system lagged the factor of the research work in the related field.
- This model achieved only 80% accuracy.
- To improve the accuracy we need to use more number of training data and also we need to use high performing algorithms

13

4.3.3 Proposed System

The principal objective of the research is to help the restourants to get there level of standards in a way that positive or negative reviews and also helps to who are aspiring to go visit the restaurant. The Restaurant review Prediction system will help them to evaluate the chances of success in improving the customers needs. It will help them in saving a huge amount of time and money spent in the knowing each and every customers decisions. Also, it will help them to limit the number of customers liking the restaurant and what the customers are expecting from the restaurant and it also helps the customers by suggesting them the best Restaurants where they have high chances of their needs.

4.3.4 Advantages of the Proposed System

- Information about the prediction analysis is clear to enter all the required information to predict the review is either positive or negative.
- The user interface code will interact with the Linear Regression, KNN, SVC to provide the users with the required result.
- User reviews may redirect consumers to more qualitative restaurants which leads lower quality restaurants to close or to improve quality in response to changes in consumer demand.

4.1 System Architecture

4.3.1 Data Flow Diagram

The machine learning models are trained with the given dataset. The machine learning models used in this project are linear regression, linear simple vector classifier(svc), random forest regressor, decision tree regressor. Once the models are trained, the model are entered to predict the chances of getting positive or negative review.

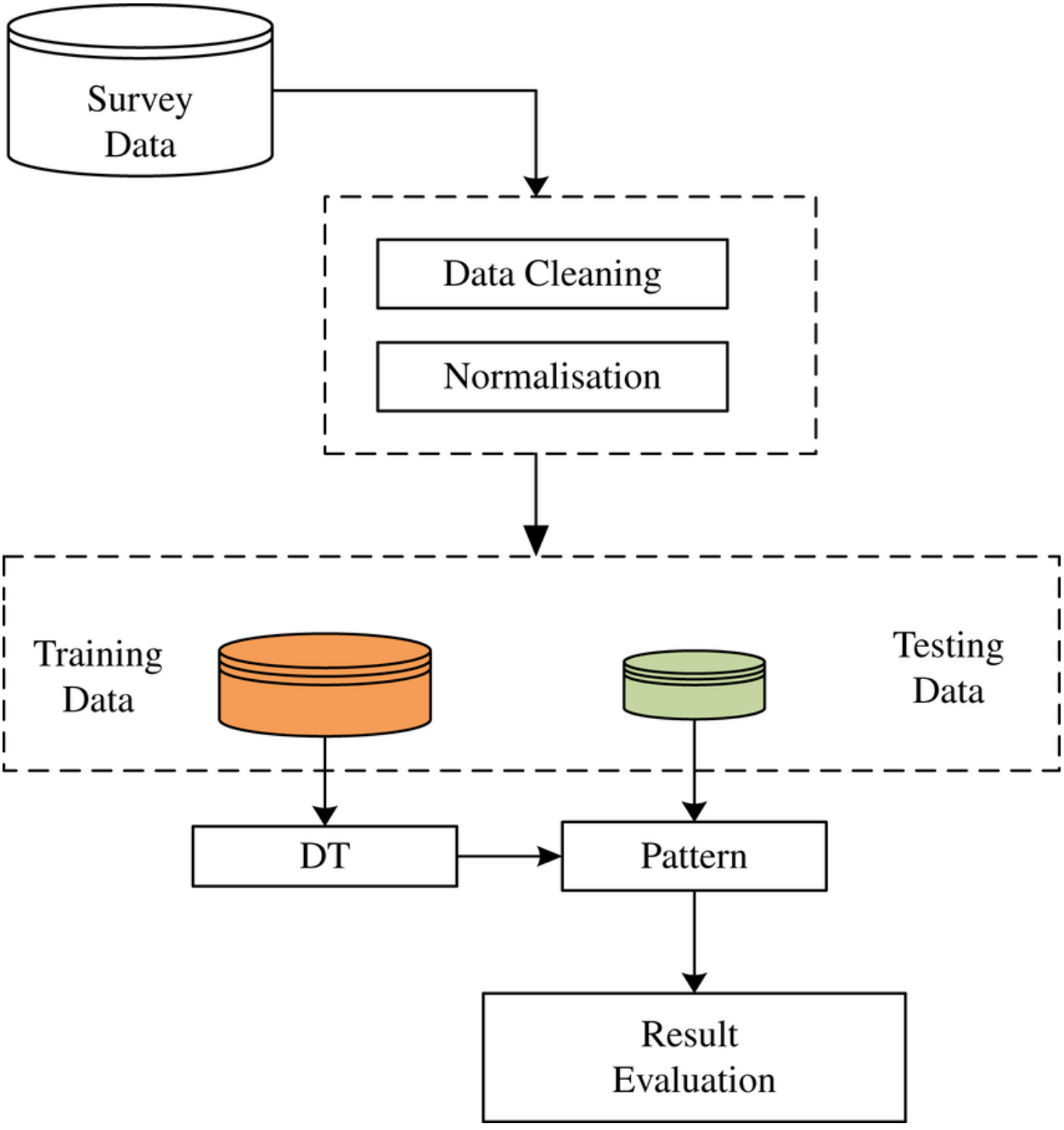


Figure 4.4.1.1 Dataflow diagram of Restaurant Review Prediction

13

Implementation

4.2.1Modules

1. Exploratory Data Analysis in Machine Learning
2. Data Visualization
3. Training and Testing
4. Train and Evaluate Linear Support Vector Classifier
5. Train and Evaluate pipeline in machine learning

MODULES DESCRIPTION

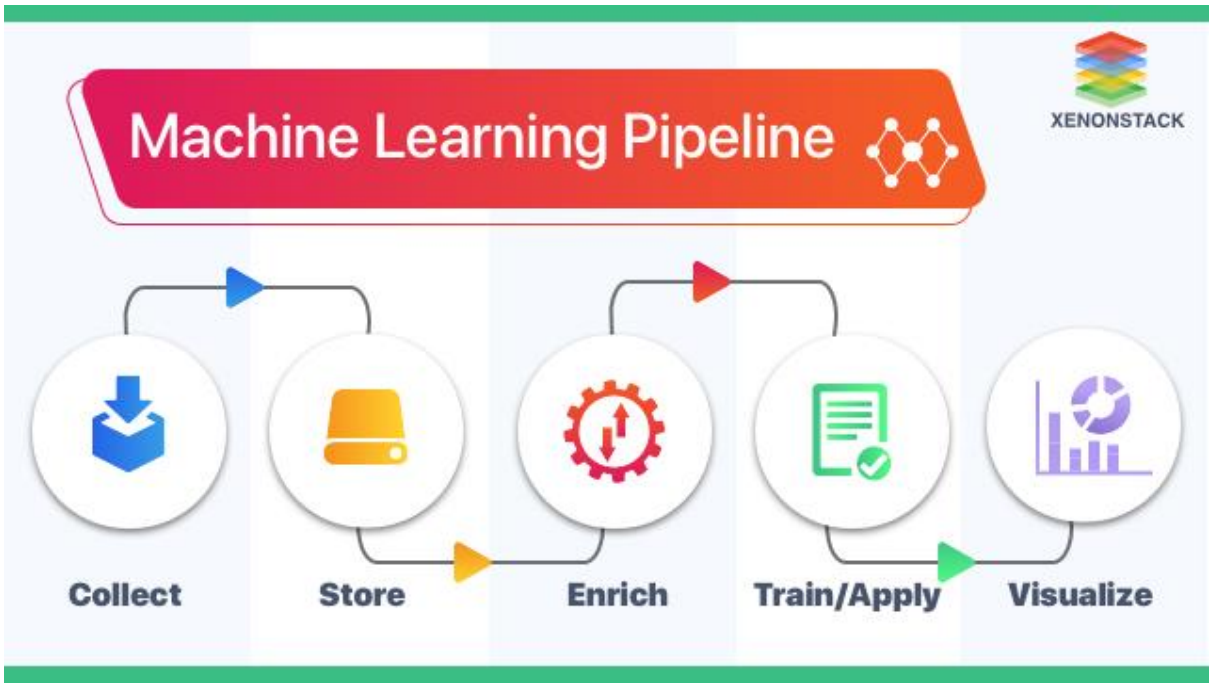
Exploratory Data Analysis: Performed initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

Data Visualization: Using data visualization, I summarized the data with graphs, pictures and maps, so that the human mind has an easier time processing and understanding the given data. Data visualization plays a significant role in the representation of both small and large data sets, but it is especially useful when we have large data sets, in which it is impossible to see all of our data, let alone process and understand it manually.

Training and Testing: In this project, datasets are split into two subsets. The first subset is known as the training data - it's a portion of our actual dataset that is fed into the machine learning model to discover and learn patterns. In this way, it trains our model. The other subset is known as the testing data.

Train and Evaluate Linear Support Vector Classifier (SVC): The Linear Support Vector Classifier (SVC) method applies a linear kernel function to perform classification and it performs well with a large number of samples. If we compare it with the SVC model, the Linear SVC has additional parameters such as penalty normalization which applies 'L1' or 'L2' and loss function

Train and Evaluate Pipeline in machine learning:



13

4.4 Screenshots

```
[1] import pandas as pd

[2] df = pd.read_csv('/content/Restaurant_Reviews.csv')

[42] df.head(10)
```

	Review\tLiked	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Review\tLiked.1
0	Wow... Loved this place.\t1	NaN	NaN	NaN	NaN	Wow... Loved this place.\t1
1	Crust is not good.\t0	NaN	NaN	NaN	NaN	Crust is not good.\t0
2	Not tasty and the texture was just nasty.\t0	NaN	NaN	NaN	NaN	Not tasty and the texture was just nasty.\t0
3	Stopped by during the late May bank holiday of...	NaN	NaN	NaN	NaN	Stopped by during the late May bank holiday of...
4	The selection on the menu was great and so wer...	NaN	NaN	NaN	NaN	The selection on the menu was great and so wer...
5	Now I am getting angry and I want my damn pho.\t0	NaN	NaN	NaN	NaN	Now I am getting angry and I want my damn pho.\t0
6	Honestly it didn't taste THAT fresh.)\t0	NaN	NaN	NaN	NaN	Honestly it didn't taste THAT fresh.)\t0
7	The potatoes were like rubber and you could te...	NaN	NaN	NaN	NaN	The potatoes were like rubber and you could te...
8	The fries were great too.\t1	NaN	NaN	NaN	NaN	The fries were great too.\t1
9	A great touch.\t1	NaN	NaN	NaN	NaN	A great touch.\t1

1. Importing data

```
[6] data=[]
    data1 = []
    for i in range(0,(value-1)):
      test = df['Review\tLiked.1'][i]
      dfn= test[0:-2]
      data.append(dfn)
      test1 = test.split()
      data1.append(test1[-1])
    print(data)
    print(data1)
```

```
['Wow... Loved this place.', 'Crust is not good.', 'Not tasty and the texture was just nasty.', 'Stopped by during the late May bank holiday off
['1', '0', '0', '1', '1', '0', '0', '0', '1', '1', '1', '0', '0', '1', '0', '0', '1', '0', '0', '0', '0', '1', '1', '1', '1', '1', '0', '1', '0',
```

2. Cleaning data.

```
[14] from sklearn.model_selection import train_test_split
     x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.05 ,random_state=0)
```

```
print(x_train,y_train,x_test,y_test)
```

```
'1' '1' '1' '0' '0' '1' '1' '1' '0' '0' '1' '0' '1' '1' '1' '0' '0' '1'
'0' '1' '0' '0' '1' '0' '0' '1' '0' '0' '0' '0' '1' '0' '0' '1' '0' '1'
'0' '1' '1' '0' '1' '0' '1' '1' '1' '0' '0' '0' '0' '0' '0' '1' '0' '1'
'0' '1' '1' '1' '0' '1' '1' '1' '0' '0' '1' '0' '1' '1' '0' '0' '0' '0'
'1' '0' '0' '0' '1' '1' '0' '1' '1' '1' '0' '0' '0' '0' '1' '0' '1' '1'
'1' '1' '0' '0' '1' '0' '1' '0' '1' '0' '1' '1' '0' '1' '1' '0' '0' '1'
'1' '0' '0' '0' '1' '0' '1' '0' '1' '1' '1' '1' '0' '0' '0' '0' '0' '0'
```

3. Dividing data set into training and testing data.

```
[17] from sklearn.feature_extraction.text import CountVectorizer

     vect = CountVectorizer(stop_words='english')
     x_train_vect = vect.fit_transform(x_train)
     x_test_vect  = vect.transform(x_test)
```

```
[18] from sklearn.svm import SVC
     model = SVC()
     model.fit(x_train_vect,y_train)

     SVC()
```

```
[19]
     y_pred = model.predict(x_test_vect)
     y_pred

     array(['1', '0', '0', '1', '0', '1', '0', '1', '1', '0', '0', '0', '1',
            '0', '0', '0', '0', '0', '1', '0', '0', '1', '1', '0', '1', '1',
            '0', '1', '0', '1', '1', '0', '1', '0', '0', '0', '0', '1', '1',
            '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '1'],
           dtype=object)
```

4. Fitting the model and predicting the output.

```
from sklearn.metrics import accuracy_score,classification_report
accuracy_score(y_pred,y_test)
```

```
0.8
```

5. Accuracy

```
[23] #for a highly imbalanced dataset, not only consider accuracy_score
```

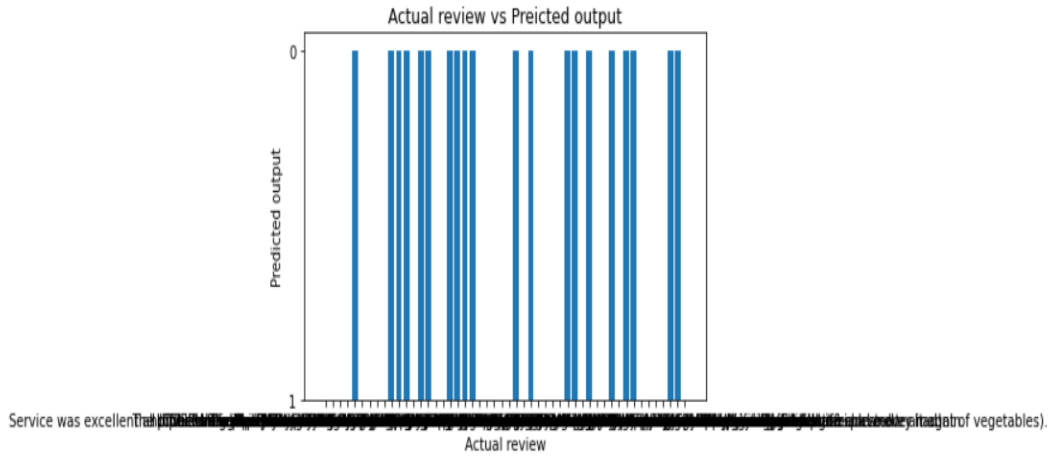
```
print(classification_report(y_pred,y_test))
```

```
              precision    recall  f1-score   support

           0       0.95      0.68      0.79        28
           1       0.70      0.95      0.81        22

    accuracy                           0.80        50
   macro avg       0.82      0.82      0.80        50
weighted avg       0.84      0.80      0.80        50
```

6. Report.

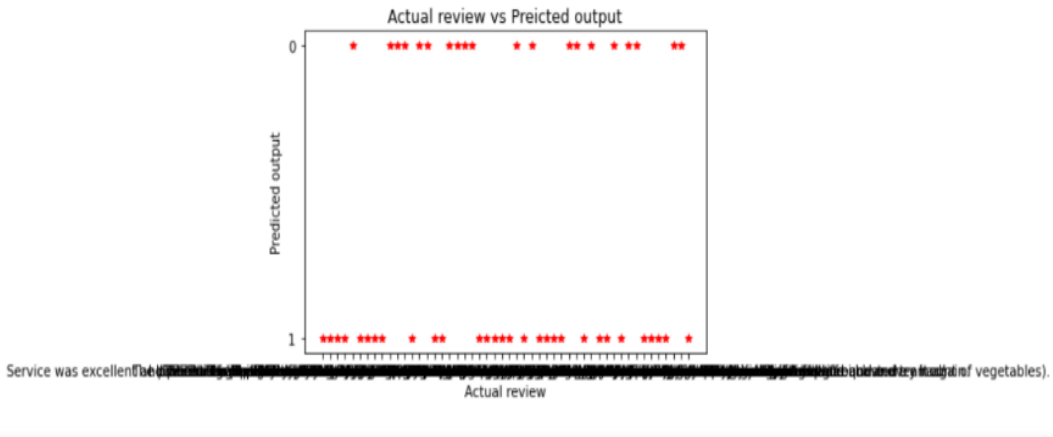
13

```
plt.bar(x_test,y_pred)
plt.xlabel("Actual review")
plt.ylabel("Predicted output")
plt.title("Actual review vs Preicted output")
plt.show()
```



7. Output in bar graph.

```
plt.scatter(x_test,y_pred, c= "r",marker="*",s=30,alpha=1 )
plt.xlabel("Actual review")
plt.ylabel("Predicted output")
plt.title("Actual review vs Preicted output")
plt.show()
```



8. Output in scatter plot.

CHAPTER – 4

CONCLUSION

This project helps to get the people satisfaction about existing restaurants of different areas in a city and analyses them to predict reviewing of the restaurant. This makes it an important aspect to be considered, before making a dining decision. Such analysis is essential part of planning before establishing a venture like that of a restaurant.

Lot of researches have been made on factors which affect sales and market in restaurant industry. Various dine-scape factors have been analysed to improve customer satisfaction levels. If the data for other citirs is also collected, such predictions could be made for accurate.

BIBLIOGRAPHY

- Chirath Kumarasiri, Cassim Faroo,"User Centric Mobile Based Decision-Making System Using Natural Language Processing (NLP) and Aspect Based Opinion Mining (ABOM) Techniques for Restaurant Selection". Springer 2018. DOI: 10.1007/978-3-030- 01174-1_4

- Shina, Sharma, S. & Singha ,A. (2018). A study of tree based machine learning Machine Learning Techniques for Restaurant review. 2018 4th International Conference on Computing Communication and Automation (ICCCA) DOI:/10.1109/CCAA.2018.8777649

- I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, 2017, pp. 542-546. doi: 10.1109/CIAPP.2017.8167276

- Rrubaa Panchendrarajan, Nazick Ahamed, Prakhash Sivakumar, Brunthavan Murugaiah, Surangika Ranathunga and Akila Pemasiri. Eatery – A Multi-Aspect Restaurant Rating System. Conference: the 28th ACM Conference

- Neha Joshi. A Study on Customer Preference and Satisfaction towards Restaurant in Dehradun City. Global Journal of Management and Business Research(2012) Link: https://pdfs.semanticscholar.org/fef5/88622c39ef76dd773fcad8bb5d 233420a270.pdf

- Bidisha Das Baksi, Harrsha P, Medha, Mohinishree Asthana, Dr. Anitha C.(2018) Restaurant Market Analysis. International Research Journal of Engineering and Technology (IRJET) Link: https://www.irjet.net/archives/V5/i5/IRJET-V5I5489.pdf

APPENDIX

Appendix A: Abbreviation

AI: Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

ML: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

KNN: The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.

Pipeline: pipeline is a means of automating the machine learning workflow by enabling data to be transformed and correlated into a model that can then be analyzed to achieve outputs. This type of ML pipeline makes the process of inputting data into the ML model fully automated.