**A**

**INDUSTRIAL TRAINING SYNOPSIS**

**ON**

**Restaurant Review Analysis**



SHRI RAM MURTI SMARAK COLLEGE OF ENGINEERING, TECHNOLOGY & RESEARCH, BAREILLY

DR. A. P. J. ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW

2020-2021

Department Name: Computer Science and engineering.

**SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR**

**COMPLETION OF DEGREE**

**BACHELOR OF COMPUTER SCIENCE**

Submitted To: Submitted by:

Naman Singhal (1845010039)

SHRI RAM MURTI SMARAK COLLEGE OF ENGINEERING, TECHNOLOGY &RESEARCH, BAREILLY



**CERTIFICATE**

This is to certify that the project report entitled “**Restaurant Review Analysis**”is a bonafide record of the project work done by the student:

**Naman Singhal (1845010039)**

In partial fulfilment of the requirement for the 5th semester of COMPUTER SCIENCE & ENGINEERINGin **SHRI RAM MURTI SMARAK COLLEGE OF ENGINEERING, TECHNOLOGY &RESEARCH, BAREILLY**

**----- -------**

**(Mini Project Incharge) (Mini Project Guide)**

**ACKNOWLEDGEMENT**

I am pleased to present this Project report entitled Restaurant Review Analysis. It is indeed a great pleasure and a moment of immense satisfaction for me to express my sense of profound gratitude and indebtedness towards my guide **Aditya sir** whose enthusiasm is the source of inspiration for me. I am extremely thankful for the guidance and untiring attention, which he bestowed on me right from the beginning. His valuable and timely suggestions at crucial stages and above all his constant encouragement have made it possible for me to achieve this work. Would also like to give our sincere thanks to **Er.Anubha Dhaka (Mini Project Incharge) Computer Sc.& Engg. Dept.SRMSCET&R, BAREILLY** for necessary help and providing me the required facilities for completion of this project report. I would like to thank the entire Teaching staﬀ who are directly or indirectly involved in the various data collection and software assistance to bring forward this seminar report. I express my deep sense of gratitude towards my parents for their sustained cooperation and wishes, which have been a prime source of inspiration to take this project work to its end without any hurdles. Last but not the least, I would like to thank all our B.Tech. Colleagues for their co-operation and useful suggestion and all those who have directly or indirectly helped me in completion of this project work.

**Naman Singhal (1845010039)**

**PREFACE**

Restaurants nowadays prefer taking online orders. It not only helps in getting effective customer feedback but also useful for managing orders easily. We are moving towards an automated and digital world. Having a significant online presence is necessary for any restaurant to be successful and prosperous.

Getting customer feedback and analyzing them in an effective manner makes the difference. This study analyses the restaurant reviews and presents useful information that the ratings do not consider or overlook. Combined research is done using two different datasets of restaurant reviews in this paper.

Machine learning algorithms like Naïve Bayes and Logistic regression is used for first classifying the reviews in proper aspects then performing sentiment analysis on them. Summarization is done using gensim and results are displayed using effective visualization techniques.

Future work is also discussed so that an efficient analysis system can be developed utilizing the potential of reviews.

Keywords- Support Vector Machine (SVM), Naive Bayes classifier, Sentiment Analysis, Topic Modelling, aspect classification.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| 1. **No.** | **Title** | **Page Number** |
| 1. | Project Description | 1 |
| 2. | Modules | 2 |
| 3. | Data flow diagram | 3 |
| 4. | Language/Libraries/Tools | 4 |
| 5. | References | 5 |

**1. DESCRIPTION OF PROJECT**

Goal of this project is to learn predicting whether restaurant review is positive or negative. A great restaurant review can point us toward our favorite spot. Some reviews are positive, but are so vague that you question their legitimacy. Some might have helpful information, but are so poorly written they’re unintelligible. And other submissions read more like a personal rant than a restaurant review.

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 predictive algorithms Multinomial Naive Bayes, Bernoulli Naive Bayes and Logistic Regression. In the end, we hope to find a "best" model for predicting the review's sentiment.

**Natural language processing (NLP)** is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data. It is the branch of machine learning which is about analyzing any text and handling predictive analysis.

[Scikit-learn](https://tutorialspoint.dev/slugresolver/learning-model-building-scikit-learn-python-machine-learning-library/" \t "_blank) is a free software machine learning library for Python programming language. Scikit-learn is largely written in Python, with some core algorithms written in Cython to achieve performance. Cython is a superset of the Python programming language, designed to give C-like performance with code that is written mostly in Python.

#### Steps involved:-

**Step 1:** Import dataset with setting delimiter as ‘ ’ as columns are separated as tab space. Reviews and their category(0 or 1) are not separated by any other symbol but with tab space as most of the other symbols are is the review (like $ for price, ….!, etc) and the algorithm might use them as delimiter, which will lead to strange behavior (like errors, weird output) in output.

**Step 2:** Text Cleaning or Preprocessing

**Remove Punctuations, Numbers**: Punctuations, Numbers doesn’t help much in processing the given text, if included, they will just increase the size of bag of words that we will create as last step and decrease the efficiency of algorithm.

**Convert each word into its lower case**:

For example, it useless to have same words in different cases (eg ‘good’ and ‘GOOD’).

1

**2.Modules:**

**Modules:**

1. **Login :** In this module HR enter the User id and password is checked andonly valid user id and password will get entry into search zone. This is a security feature to avoid entry of unauhorized users.
2. **Reset:** Through this HR can enter its start-up procedure as if you had turned the login wrong and then login again.
3. **Submit:** Through this HR can automatically submits a subject to a particular process or to the authority to present review consideration.
4. **Back:** Back button allows an end user to navigate to previously viewed web page.

**Restaurant Review Analysis:-**

# **1_JcYtLsVyrUtqm9S-0plXyQ**

2

# DATA FLOW DIAGRAM

The Data flow Diagram shows the flow of data. It is generally made of symbols given below:

(1) A **square** shows the Entity: -

(2) A **Circle** shows the Process: -

(3) An **open Ended Rectangle** shows the data store: -

(4) An **arrow** shows the data flow:-

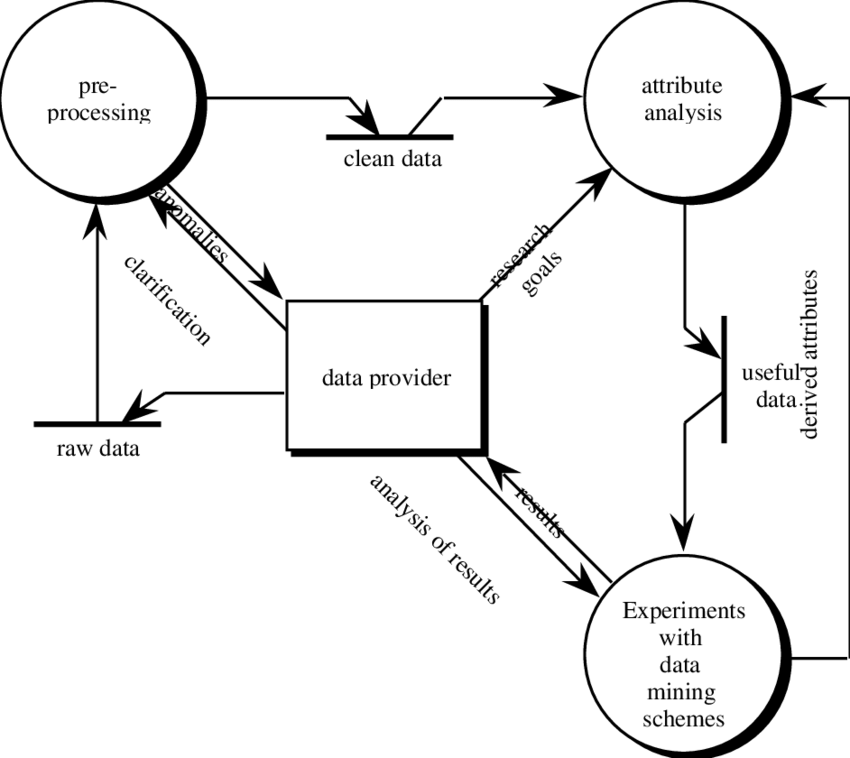
The DFD can be up to several levels.

The 0 level DFD states the flow of data in the system as seen from the outward in each module.

The first level DFD show more detail, about the single process of the 0 level DFD.

The second level DFD can show even more details and so on.

**Context Level DFD :-**



3

**4. Language/Libraries/Tools**

**Front End:-**

* Python sklearn
* Python tkinter

**Back End:** -

* Python and DataScience

**Libraries:-**

* tkinter
* CountVectorizer
* naive\_bayes
* pyinstaller
* MultinomialNB
* stop**\_**words
* todense
* feature vector

**Other S/W:-**

* Python3.x
* IDLE
* Anaconda(jupyter lab)

4

**References**

1. Basant, A., M. Namita, B. Pooja, and Sonal Garg. "Sentiment analysis using common-sense and context information." (2015).
2. Bykh, Serhiy, and Detmar Meurers. "Native Language Identification using Recurring n-grams– Investigating Abstraction and Domain Dependence-grams– Investigating Abstraction and Domain Dependence." In Proceedings of COLING 2012, pp. 425-440 (2012).
3. Westerski, Adam, Carlos Angel Iglesias Fernandez, and Fernando Tapia Rico. "Linked opinions:Describing sentiments on the structured web of data." (2011).
4. Yi, Jeonghee, and Neel Sundaresan. "A classifier for semi-structured documents." In KDD, pp. 340-344. (2000).
5. Moghaddam, Samaneh, and Martin Ester. "Opinion digger: an unsupervised opinion miner from unstructured product reviews." In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1825-1828. ACM, (2010).
6. Wiegand, Michael, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. "A survey on the role of negation in sentiment analysis." In Proceedings of the workshop on negation and speculation in natural language processing, pp. 60-68. (2010).
7. Flekova, Lucie, Daniel PreoţiucPietro, and Eugen Ruppert. "Analysing domain suitability of a sentiment lexicon by identifying distributionally bipolar words." In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 77-84. (2015).
8. Asghar, Muhammad Zubair, Aurangzeb Khan, Shakeel Ahmad, and Fazal Masud Kundi. "A review of feature extraction in sentiment analysis." Journal of Basic and Applied Scientific Research 4, no. 3 (2014): 181-186.
9. Lin, Yuming, Tao Zhu, Hao Wu, Jingwei Zhang, Xiaoling Wang, and Aoying Zhou. "Towards online antiopinion spam: Spotting fake reviews from the review sequence." In 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), pp. 261-264. IEEE, (2014).
10. Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. Part-of-speech tagging for twitter:Annotation, features, and experiments. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, (2010).
11. Cavnar, William B., and John M. Trenkle. "N-gram-based text categorization." In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, vol. 161175. (1994).
12. Aggarwal, Charu C. "Text Summarization." In Machine Learning for Text, pp. 361-380. Springer, Cham, (2018).

5