ISSN 2277-2685
IJESR/Jul-Sep. 2024/ Vol-14/Issue-3/435-445
Tafazul Khan *et. al.,* / International Journal of Engineering & Science Research

AN INTELLIGENT APPROACH FOR FOOD RECIPE RATING PREDICTION USING MACHINE LEARNING

Tafazul Khan[1], Abdullah Ibrahim Hassan[2], Syed Nabeeluddin Ahmed[3], Neha Hasan[4]
[1,2,3] B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
nehahasan@lords.ac.in

Abstract: *In recent times, many studies and systems focus on restaurant or individual food ratings, but rating recipes using Artificial Intelligence remains rare. This study aims to evaluate recipes based on various attributes using different machine learning algorithms. It compares the performance of different classifiers in rating recipes according to various performance criteria. This approach can be economically beneficial for restaurants by helping them improve their recipes and attract more customers. It can also be useful at a personal level to enhance household recipes and assist customers in choosing restaurants based on the quality of their recipes for specific dishes.*

I. Introduction

The popularity of food tasting and rating is on the rise, as evidenced by numerous online and social media platforms. Food quality can be evaluated from various perspectives. In this paper, we focus on rating food recipes based on different attributes. A recipe rating system that considers preparation methods and user reviews can be highly beneficial for the food industry. Restaurants can use such a system to rate their recipes and make adjustments to increase customer satisfaction and acceptance.

This approach can also be valuable on a personal level for individuals who enjoy cooking, experimenting with different recipes, and are interested in seeing how their creations are rated. Many factors influence a person's food choices. While reviews of specific restaurants, including overall service and food quality, are widely available, detailed information about individual dishes or recipes is less accessible. For someone interested in a particular dish or recipe, it can be challenging to determine its quality. Different restaurants may prepare the same dish using different recipes. This system can help them compare the ratings of their dishes against others.

II. Literature Survey

There have been several studies focused on food rating, restaurant reviews, calorie estimation, and food quality assessment based on various attributes such as taste and health. Here is a list of such works:

1. The paper [1] provides reviews of Thai restaurants worldwide. It extracts reviews from social media and classifies them using neural networks.

2. Paper [2] discusses the main themes in the food and restaurant industry, including taste, customer experience, and location value, based on 294,034 reviews from Yelp.com using Latent Dirichlet Allocation (LDA). It evaluates both positive (taste) and negative (value) aspects, demonstrating a robust classification algorithm using support vector

machines (SVM) and Fuzzy Domain Ontology (FOD) that outperforms traditional algorithms like Naïve Bayes (NB) and SVM in predicting the quality of online reviews.

3. The paper [3] applies deep learning to food science, using it as an advanced data mining tool for food sensory research. Their survey shows that deep learning outperforms conventional machine learning algorithms in food science applications.

4. Paper [4] focuses on rating individual food items from restaurants rather than the entire restaurant, collecting data from online reviews. They use named entity recognition (NER) techniques to identify individual food names in restaurant reviews.

5. The paper [5] employs a recursive neural network for sentiment analysis on 'Amazon Fine Food Reviews,' parsing binary trees using Stanford NLP parsers.

6. Paper [6] discusses the attributes that most affect online restaurant reviews and star ratings, identifying food, service, and context as the three main contributing factors.

7. Finally, paper [7] classifies fast food restaurant reviews, quantifying textual reviews based on the overall opinion of each online review.

III. System Analysis

One existing system that addresses the evaluation and rating of food recipes and restaurant dishes is the Yelp Fusion API. This API provides developers with comprehensive access to a vast repository of restaurant-related data, including detailed information on dishes, user reviews, ratings, and photographs. Through the Yelp Fusion API, developers can query and retrieve specific details about restaurants, enabling them to analyze user-generated reviews and ratings for individual dishes. This functionality allows for the extraction of valuable insights regarding the reception of particular recipes based on factors such as taste, presentation, and overall customer satisfaction. Moreover, the API facilitates the aggregation of ratings across various dishes and restaurants, empowering developers to compare and assess the popularity and quality of different culinary offerings. This capability is particularly useful for building applications or conducting research focused on recipe rating systems, restaurant recommendation engines, and food quality assessments. By integrating the Yelp Fusion API into custom applications or analytical frameworks, developers can leverage real-time updates and a wealth of user-generated content to enhance understanding and decision-making in the food and restaurant industry. Overall, the Yelp Fusion API serves as a versatile toolset for exploring and utilizing food-related data to support innovative solutions in culinary evaluation and customer satisfaction analysis.

1. **Data Limitations and Bias**: Yelp's data may not cover all restaurants equally, leading to geographic biases and incomplete representations of culinary diversity.
2. **Quality and Reliability**: User-generated reviews vary in accuracy and reliability, potentially impacting the trustworthiness of insights derived from the API.
3. **Limited Customization**: The API provides predefined endpoints and functionalities, limiting the ability to customize data queries and analysis methods according to specific research or business needs.

IV. Proposed System

A proposed system for evaluating food recipes and dishes would incorporate several key components to ensure comprehensive and accurate ratings. First, it would gather data from diverse sources such as user reviews, social media platforms, and expert opinions, ensuring a broad dataset for analysis. Natural language processing (NLP) techniques would be utilized to extract essential features from reviews, including factors like taste, presentation, and nutritional value. Machine learning models, such as sentiment analysis and topic modeling using methods like Latent Dirichlet Allocation (LDA), would be employed to analyze and rate recipes based on these extracted features. This approach aims to provide nuanced evaluations that consider multiple aspects of recipe quality. To enhance user engagement and system accuracy, an interactive interface would allow users to provide feedback on recipes, influencing ratings and contributing to ongoing improvements in the system's performance. The system would be designed with scalability in mind, enabling integration across various platforms like mobile apps and websites, and facilitating future expansions in data sources and functionalities. Crucially, strong measures would be implemented to protect user privacy and ensure compliance with data protection regulations, safeguarding personal information collected through the system. By integrating these elements, the proposed system seeks to establish a robust framework for assessing food recipes effectively, supporting informed decision-making for both consumers and professionals in the culinary industry.

- **Comprehensive Data Collection**: Gathering data from diverse sources like user reviews and social media ensures a wide-ranging dataset, enriching the system's ability to provide thorough recipe evaluations.
- **Advanced Feature Extraction**: Using NLP techniques extracts essential recipe attributes (e.g., taste, presentation) from text data, facilitating detailed analysis and rating.
- **Effective Machine Learning Models**: Algorithms such as sentiment analysis and SVM enable accurate assessment of recipes based on extracted features, improving precision in evaluating recipe quality.
- **User Interaction and Feedback**: An interactive interface allows users to provide input on recipes, refining ratings and enhancing system accuracy over time.

V. Classification Study

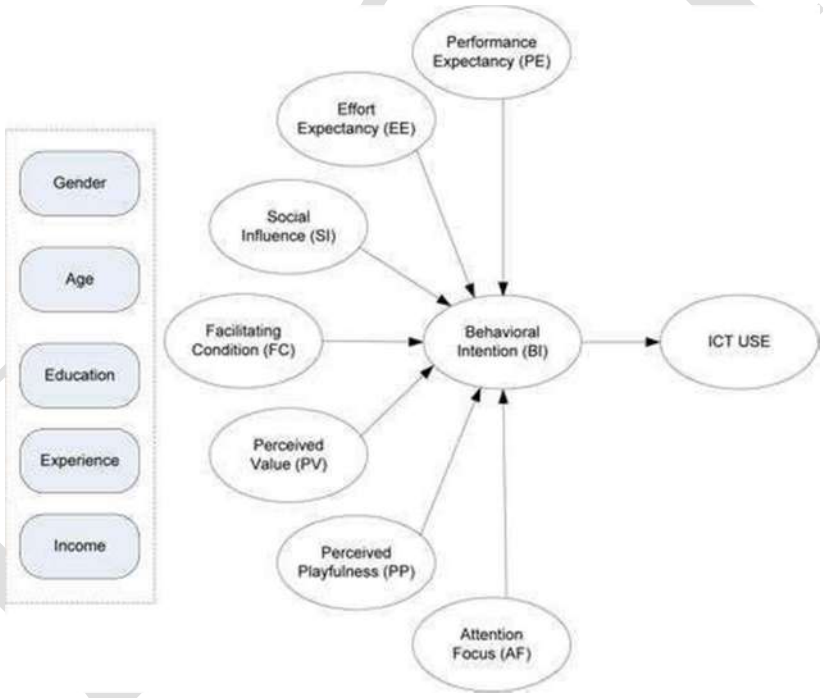
The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.Three key considerations involved in the feasibility analysis are, Economical Feasibility: This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility: This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility: The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

VI. System Design

System Architecture:

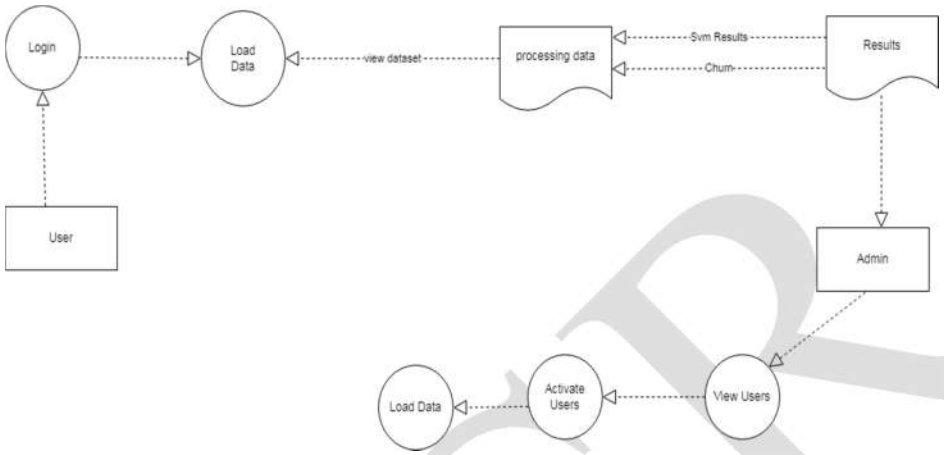


The bubble chart is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

The data flow diagram is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

It shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

It may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.[8]



UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artefacts of software system, as well as for business modelling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

Goals

The Primary goals in the design of the UML are as follows:

Provide users a ready-to-use, expressive visual modelling Language so that they can develop and exchange meaningful models.

Provide extendibility and specialization mechanisms to extend the core concepts.

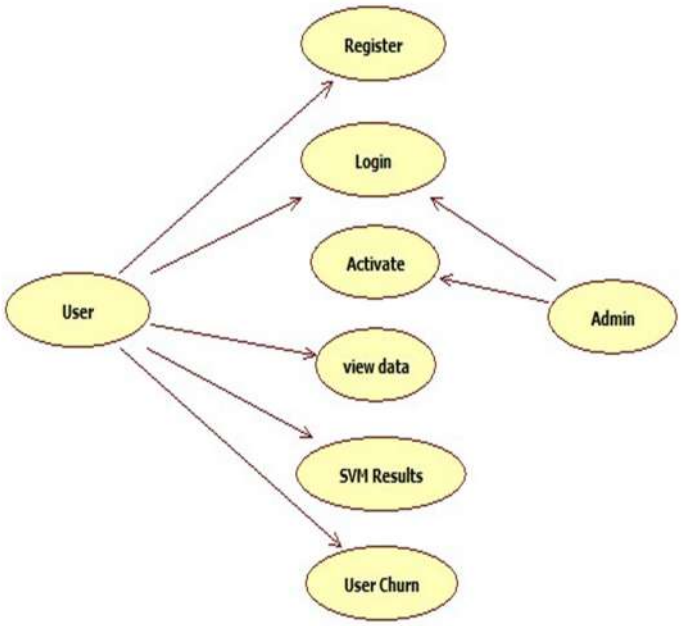
Be independent of particular programming languages and development process.

Provide a formal basis for understanding the modelling language.

Encourage the growth of OO tools market.

Support higher level development concepts such as collaborations, frameworks, patterns and components. Integrate best practices.

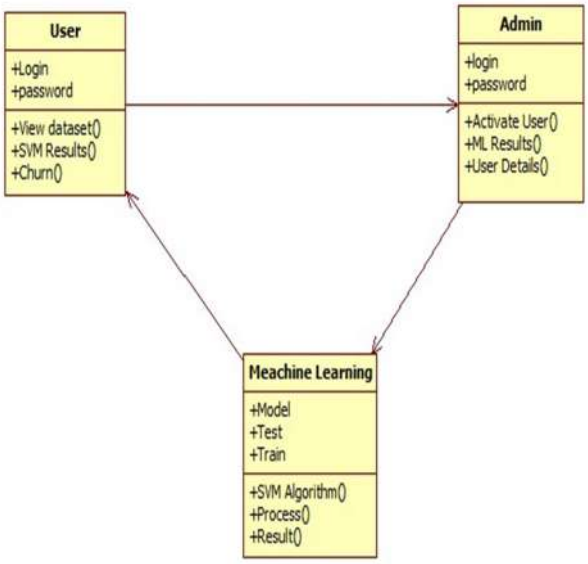
- Feature Extraction Component: One of the notable developments in any gathering- centered issues is highlighting extraction. Looks are the cornerstone for both purposes of planning and screening. This feature contains noteworthy image information that will be used to identify the Food.
- Identification unit for Food: See if the Food is considerate or hazardous.
- Input Attributes: For example, all noteworthy attributes, asymmetry, edge, concealment, distance, progression, etc. that have been expelled from the image are now provided as a dedication to Part II, which is the classifier part.
- Classifier engine: characterizes the images by grouping the calculation into one of the predefined Food.



A use case diagram in the Unified Modelling Language (UML) is a type of behavioural diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

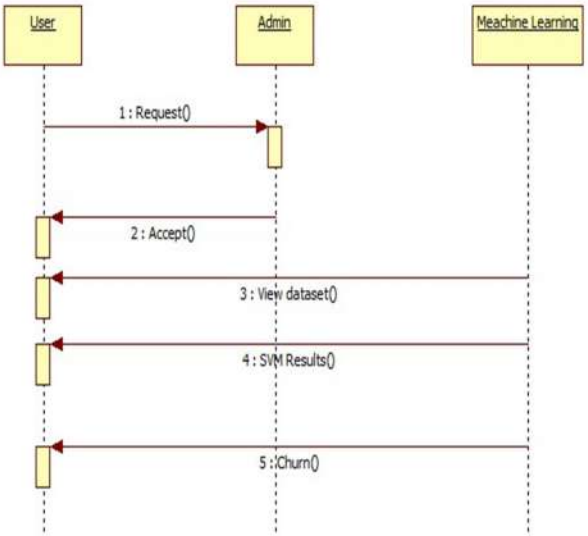
Class Diagram

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



Sequence Diagram

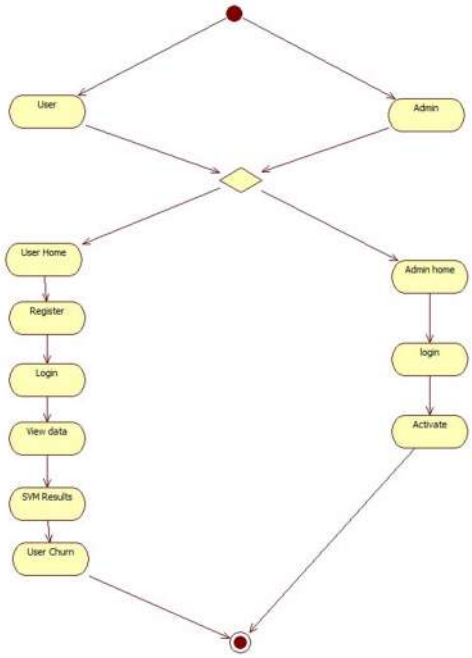
A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams



Activity Diagram:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the

business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



VII. Modules Description

User:

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in float format. Here we took Three Customer Behaviour dataset for testing purpose. User can also add the new data for existing dataset based on our Django application. User can click the Classification in the web page so that the data calculated Accuracy and F1-Score, Recall, Precision based on the algorithms. User can click Prediction in the web page so that user can write the review after predict the review that will display results depends upon review like positive, negative or neutral.

Admin:

Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. Admin can click the Results in the web page so calculated Accuracy and F1-Score, Precision, Recall based on the algorithms is displayed. All algorithms execution complete then admin can see the overall accuracy in web page.

Data Preprocessing:

A dataset can be viewed as a collection of data objects, which are often also called as a records, points, vectors, patterns, events, cases, samples, observations, or entities. Data objects are described by a number of features that capture the basic characteristics of an object, such as the mass of a physical object or the time at which an event occurred, etc. Features are often called as variables, characteristics, fields, attributes, or dimensions. The data preprocessing in this forecast uses techniques like removal of noise in the data, the expulsion of missing information, modifying default values if relevant and grouping of attributes for prediction at various levels.

Machine learning, Knowing the nutrition content of the food that we are consuming helps in maintaining balanced diet. We have aimed with a variety of food categories, each containing thousands of images, and through machine learning training to achieve higher classification accuracy. Firstly, we have planned to train and optimize a CNN, state-of-art model using Tensorflow, we are using CNN as the convolution layers are tweak able and easy to implement. Second, we adapt our model with GUI features as well as nutrition analysis.

Based on the split criterion, the cleansed data is split into 60% training and 40% test, then the dataset is subjected to four machine learning classifiers such as Support Vector Machine (SVM). The accuracy, Precision, Recall, F1-Score of the classifiers was calculated and displayed in my results. The classifier which bags up the highest accuracy could be determined as the best classifier.

1.Acquisition of Image: Images are obtained either by lens or by secretly deleting them from the contraction. Whatever the source may be, it is very important that the image of the data is transparent and cautious. An incredible picture is needed for this.

2.Pre-Processing of Image: In this process, the photo is standardized by clearing the commotion, as it may confuse the evaluation. Similarly, the image given as the information may not be of standard size as required by the figure, so it is vital that the image size needed is obtained.

3.Data Storage Aspect To Preserve Information Images For Testing And Training: If controlled learning will occur, as is the case here, it is important to prepare data sets. The sample database is the images collected during the photo procurement process

4.Classifier to classify the FOOD: The classifier used here is the last layer of the system which gives the true probability of each experience. The project involves two major parts Image preparation unit and grouping unit. The object processing system enhances the image by removing the clatter and noisy bits. The Food image will then be isolated into different segments to isolate the image from running the mill after the image features are evacuated to check whether or not the Food is contaminated.

VIII. Conclusion

The proposed system for rating food recipes and dishes leverages advanced data collection, feature extraction through natural language processing, and effective machine learning models to provide comprehensive evaluations. User interaction and feedback enhance system accuracy over time, while scalability ensures adaptability to evolving

needs. This approach supports informed decision-making for consumers and culinary professionals alike, fostering trust through robust privacy measures. Ultimately, the system aims to elevate the quality assessment of recipes, contributing to enhanced culinary experiences and industry innovation.

IX. References

[1] Niphat Claypo and Saichon Jaiyen. Opinion mining for thai restaurant reviews using neural networks and mrmr featureselection.In2014 International Computer Science and Engineering Conference (ICSEC), pages 394–397. IEEE,2014

[2] Yi Luo and Xiaowei Xu. Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp. Sustainability, 11(19):5254.

[3] Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and YongHe. Application of deep learning in food: A review. Comprehensive Reviews in Food Science and Food Safety,18(6):1793–1811, 2019.

[4]Burusothman Ahiladas, Paraneetharan Saravanaperumal,Sanjith Balachandran, Thamayanthy Sripalan, and Surangika Ranathunga. Ruchi: Rating individual food items in restau-rant reviews.InProceedings of the 12th InternationalConference on Natural Language Processing, pages 209–214, 2015.

[5] Jiayu Wu and Tianshu Ji. Deep learning for amazon food review sentiment analysis.

[6] Qiwei Gan, Bo H Ferns, Yang Yu, and Lei Jin. A text mining and multidimensional sentiment analysis of online restaurant reviews.Journal of Quality Assurance in Hospitality & Tourism, 18(4):465–492, 2017.

[7] Lindsey Wright. Classifying textual fast food restaurant reviews quantitatively using text mining and supervised machine learning algorithms. 2018

[8] Thomas G Dietterich et al.Ensemble learning. The Handbook of brain theory and neural networks, 2:110–125,2002.

[9] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, and Wojna Z. “Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition” (2021).

[10] Mohammed A. Subhi and Sawal Md. Ali. “A Deep Convolutional Neural Network for Food Detection and Recognition” (2019).

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh,S. Ma, Z.Huang, A. Karpathy, A. Khosla, M. Bernstein "ImageNet Large Scale Visual Recognition Challenge" (2019).