

# Food Reviews Classification using multi-label convolutional neural network text classifier

*Krutuja S Lasne<sup>1</sup>, Sejal S Nandrekar<sup>2</sup>, Ashraf A Khan<sup>3</sup>, Tushar Ghorpade<sup>4</sup>*

Department of Computer Engineering  
Ramrao Adik Institute of Technology  
Nerul, Navi Mumbai, India.

[<sup>1</sup>krutuja.lasne999@gmail.com](mailto:krutuja.lasne999@gmail.com), [<sup>2</sup>sejalnandrekar@gmail.com](mailto:sejalnandrekar@gmail.com), [<sup>3</sup>aashrafakhan@gmail.com](mailto:aashrafakhan@gmail.com), [<sup>4</sup>tushar.ghorpade@rait.ac.in](mailto:tushar.ghorpade@rait.ac.in)

**Abstract**—Most commercial websites, such as Amazon, encourage users to leave reviews of the goods and services they get after purchasing them. For certain consumers, this analysis is critical when determining whether or not to buy a product. Understanding the consequences of feedback and correctly classifying their utility may therefore be an advantageous method for websites. The classification results can also be used as a review and recommendation program for ongoing success. Nowadays people visit several restaurants on various occasions. They get confused most of the time after having a look at the food menu. Based on the ratings and reviews of the dish it becomes easier for them to decide the dish they wish to order. However they unable to read each review of the previous customers. So to overcome this issue, we have proposed NLP (Natural Language Processing) technique and Spacy CNN (Convolutional Neural Network) pipeline system which will classify all the reviews in a single rating. Each review is labelled with a reviewer's score indicating the sentiment of the reviewers. Our task is to predict a reviewer's score on a scale of 0 or 1. Where 1 indicates the users like the dish while 0 indicates that the reviewers were not satisfied with the dish.

**Index Terms**—Summarization, Sentiment, Review Recommendation, Machine Learning, CNN model.

## 1. Introduction

People visit several restaurants on various occasions. They get confused most of the time after having a look at the food menu. Based on the ratings and reviews of the dish it becomes easier for them to decide the dish they wish to order. But it is impossible for them to read each and every review of the previous customers [6],[9]. This proposal aims to create an automatic text-based classification model that can forecast the usefulness of Zomato, Swiggy and Reddit feedback accurately. The challenge at hand is to perform a binary classification using a mixture of text-based functionality and machine learning algorithms. The binary classes will be described as "like" and "dislike," with "1" denoting "like" and "0" denoting "dislike". Text encoding, tokenization, translating uppercase to lowercase, lemmatization, and other text-based features were used to create this model [8].

Further paper breakdowns are as follows, section 2 describes the Literature Survey, section 3 describes the problem

definition, section 4 describes the Proposed System and evaluation parameters for implementation, section 5 concludes the paper along with section 6 which shows the future work.

## 2. Literature Survey

This section describes the methodology adopted for the literature review. This paper represents an exploration of the contributions that have already been made in the academic field.

In this paper author Oman Somantri et. al. [1] proposed a hybrid method called PSO-IG means Particle Swarm Optimization and Information Gain. Author have used four different algorithm as K-NN, Support Vector Machine, Naive Bayes and Decision Tree. The main purpose of this hybrid model is to classify food reviews using sentiment analysis approach of culinary food which will be very helpful to culinary shopkeepers and seekers.

In this paper author Sindhu B Hegde et. al. [2] suggest a system of recommendations feedback from customers in forecasting star ratings for food classification for restaurant business. Positive, unfavourable, moderate, and divisive user feedback are grouped into four groups. SVM grouping is used to determine training data in accordance with divisions. The proposed framework is used two main tasks four course meal and sentiment food classification.

The author Hemalatha S et. al. [3] have proposed feature based sentiment analysis for restaurants. They have used ambiance, price, service and food quality features for classifying reviews as neutral, negative and positive. As part of the Yelp Database Challenge, highlights the introduction of machine learning algorithms that process textual and mathematical data presented in the Yelp database. The aim was to categorise business feedback so that they could be sorted consistently from negative to positive. Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, and Linear SVC are the algorithms used (Support Vector Clustering).

A data analysis or statement on the TripAdvisor’s web page about products in restaurants in Surabaya is used as a source in this paper [4]. The author Masrur Adnan et. al. used Web Harvy software to gather the results. The NLTK (Natural Language Tool Kit) library was used to process the data in order to promote the Python programming language. Wrapping events, removing marks, token-ups, altering sentences, suspending deletions, and stopping are all moves in the progression process. Using the Decision Tree, it figures out how many times each word appears in the text.

Author Jiayu Wu and Tianshu Ji [5] tested whether recursive neural network is effective with insufficient tree labelling or not. They proposed Recursive Neural Network for multiple sentences, to handle multiple sentences at once. However the accuracy was challenging.

Present Systems are totally random review based if we rely on star based review and does not contribute towards understanding the salient features of the dish. Star based system is totally depend on the mood of the user. If the consumer is angry it will result in decrease of average rating on the food. The consumer way change it not depending upon the environment and may lead to rating the dish thing of other parameters like waiter behaviour, environment of restaurant etc. which affect the average rating of dish.

### 3. Problem Definition

This proposal aims to create an automatic text-based classification model that can forecast the usefulness of Zomato, Swiggy, and Reddit feedback accurately. The challenge at hand is to perform a binary classification using a mixture of text-based functionality and machine learning algorithms. The binary classes will be described as "like" and "dislike," with

"1" denoting "like" and "0" denoting "dislike". Text encoding, tokenization, translating uppercase to lowercase, lemmatization, and other text-based features were used to create this model.

## 4. Proposed Methodology

### A. Proposed Work :

This proposal is aimed at the development of an application system through which the users enter the food review which gets converted into tokens and is passed through all methods which have been implemented in "Fig. 2"

The primary focus of the system is to understand the user vocabulary review and rate the dish automatically without relying on star-based system.

The proposed methodology is quite simple. The food or dish is to be rated and it is important to understand the vocabulary of the review. Most rating systems are inefficient to perform this task. In the new approach, the author is using spacy and food review dataset from Reddit.

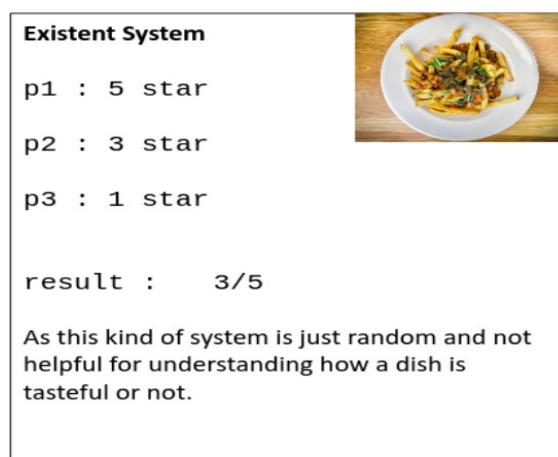


Fig. 1 Existing System

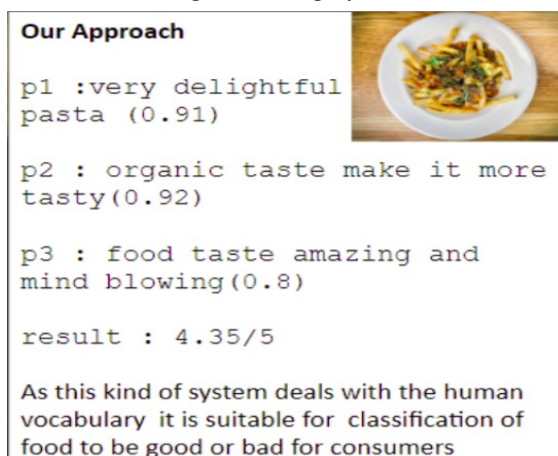


Fig.2 Proposed System

In this system, we learned how Spacy works and understood how to use this technology to understand the user review to predict the rating of the dish. As we are using spacy library for advanced NLP we are considering all parameters of the consumer review which may lead to an optimal solution.

**B. Implementation Details:**

The model is trained on a heavy datasets and spacy library in python is used to develop this system. This model is trained to understand a review from its core like understanding noun-adjective, adverb, verb, etc. So by using this module it will be able to achieve the score while classification of the review.

Data is taken from Kaggle and also some scraped data from an online food review website is used. As this proposal is based on supervised learning the reviews were labelled. By pre-processing the user review and passing the review through spacy filter, tokens will be created to understand grammar, nouns from the sentences.

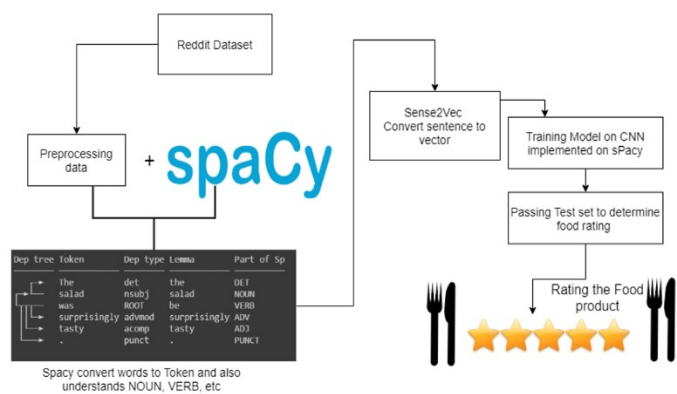


Fig. 3 System Design

As this machine can understand sentences so the tokens that are generated will be converted to vector using sense2vec. This will create vector for of every words and it understand every word and word related to another word.

For Eg: doc = spacy token (“peanut butter”)

[(('peanut butter', 'NOUN'), 1.0), (('peanut butter', 'NOUN'), 0.9349336624145508), (('almond butter', 'NOUN'), 0.9219189882278442), (('apple slices', 'NOUN'), 0.9212321043014526)]

As this all food belong to same category they have same value. The data is trained in the further step. As spacy has it’s own training algorithm called TextCategorizer it helps to train the model and store it in file so that it can be exported and can be stored for future use.

**C. Word vectors and similarity:**

Sentence to vector is a program that converts words into sentences. The aim is to create a model that is superior to the word to a vector model. The theory behind sense to vector is quite simple. If the problem is that duck in the sense of waterfowl and duck in the sense of crouch are two different concepts, the easy solution is to have just two entries, duck as noun and duck as a verb. Trask et al, (2015) presented a series of tests that demonstrated the concept’s viability. It assigns sections of speech tags to phrases, such as verb, noun, and adjective, which are then used to provide meaning.

- Please [VERB] book my ticket.
- Read the [NOUN] book.

Reddit has a lot of discussions about food, we can acquire strong similarity vectors for food goods.

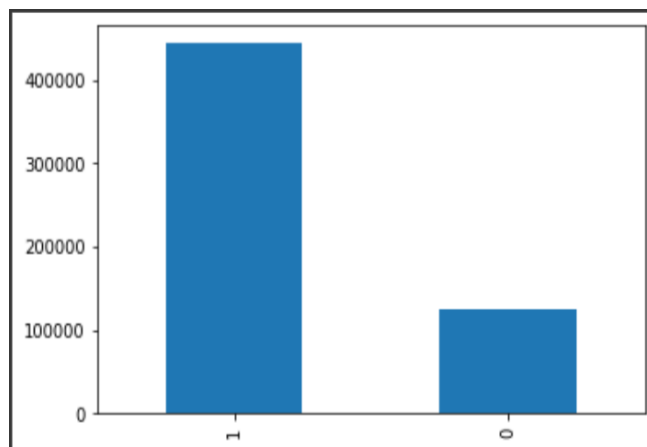


Fig. 4 Histogram plot on data set

```

Loading food reviews data...
Using 30000 examples (24000 training, 6000 evaluation)
    
```

Fig. 5 Number of data for training and testing

### D. Spacy Text Categorizer:

We have trained a multi-label convolutional neural network text classifier on our food evaluations using spacy’s newest text categorizer component.

Spacy provides a categorization paradigm that includes a number of labels that are not mutually exclusive. The model setup can be changed, but by default, the Text Categorizer class uses a convolutional neural network to assign position-sensitive vectors to each word in the document. To stop exchanging weights with other pipeline elements, the Text Categorizer uses its own CNN model. The text sensor is summarised after max and mean pooling, and a multilayer perceptron is used to predict an output vector of length no class, before elementwise logistic activation. The value of each output neuron represents the chance of any class being present.

## 5. Results And Analysis

### 5.1 Dataset:

Fig. 4 describes graph of fine cuisine reviews from Amazon are included in this collection. All 500,000 reviews up to October 2012 are included in the data, which covers more than a decade. In addition, reviews include product and use statistics, ratings, and a plain- language summary. Ratings from all of Amazon’s other categories are also included.

#### Data includes:

Since October 1999 to October 2012, the reviews are taken. The number ‘1’ is considered as a Positive review and number ‘0’ is considered as a negative review. The total number of reviews are 30000. Out of that 24000 reviews used for training and 6000 reviews used for validation.

### 5.2 Performance Measure:

As SpaCy is a powerful tool to detect sentiment analysis on the review it will understand all pros and cons of the review depending on the trained dataset. As Food is the primary need of a consumer we should understand what kind of food is most likely to have by a consumer and trending food needs to have a better ranking algorithm. As per the study of existing system from few reputed papers authors have uses different algorithms to do sentiment analysis [7] of review but complex algorithm takes more time to implement and take more time to predict as Spacy has an encrewebsm library spacy gets a plus to understand noun,pronoun,adjective etc which is very help to understand the review and rate the review.

LOSS	P	R	F
327.039	0.836	0.851	0.843
200.148	0.852	0.864	0.858
165.254	0.855	0.869	0.862
138.109	0.858	0.867	0.862
123.523	0.866	0.864	0.865
112.838	0.865	0.875	0.870
103.065	0.866	0.872	0.869
94.570	0.870	0.865	0.867
89.310	0.869	0.865	0.867
82.493	0.864	0.870	0.867

Fig. 6. Training data for 10 epochs

The following table 1 and Fig. 7 shows the comparison of different models with evaluation parameters are precision, recall and f-score.

Table 1: Comparison between various models.

Evaluation Parameters	Proposed Model using Spacy CNN	Naive Bayes	RNNMS	Decision Tree
Precision	86.40 %	69.00 %	75.00 %	48.00 %
Recall	87.00 %	69.00 %	75.00 %	36.80 %
F-score	86.70 %	69.00 %	74.00 %	41.40 %

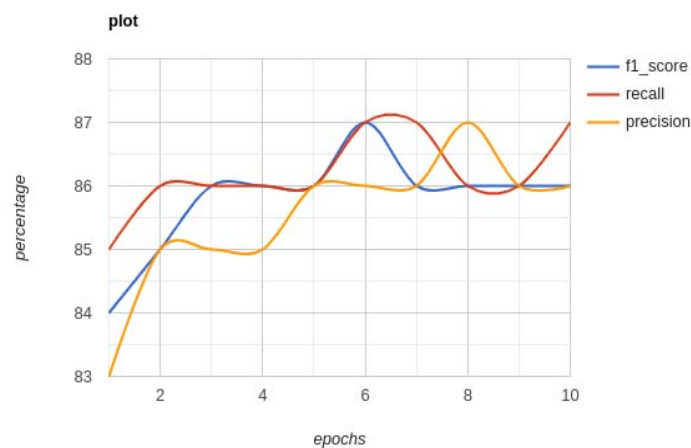


Fig. 7. Graph plot

The above graph plot is for Recall, Precision and F-score. It shows percentage for 10 epochs.

## 6. Conclusion

On the basis of the ratings and reviews of the dish, it becomes easier for customers to decide the dish they wish to order. This kind of system will help many online restaurant/online food markets to create a satisfying review process for the users and with this kind of mechanism the Hotel/food industries can create a food ranking page automatically. By reviews, the online food company will understand what user likes and what user doesn't like independent on where the users are located. We also can create a hybrid model where we will consider both the rating system and review rating system in the future. In Conclusion, the rating system don't work independently to provide how the product is useful to the consumer but with the help of review classification, we can generate rating with the help of users review.

## References

1. O. Somantri, D. A. Kurnia, D. Sudrajat, N. Rahaningsih, O. Nurdiawan and L. Perdana Wanti, "A Hybrid Method Based on Particle Swarm Optimization for Restaurant Culinary Food Reviews," Fourth Inter-national Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985842.(2019)
2. S. B. Hegde, S. Satyappanavar and S. Setty, "Sentiment based Food Classification for Restaurant Business," International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1455-1462, doi: 10.1109/ICACCI.2018.8554794.(2018)
3. H. S. and R. Ramathmika, "Sentiment Analysis of Yelp Reviews by Machine Learning," International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 700-704, doi: 10.1109/ICCS45141.2019.9065812.(2019)
4. M. Adnan, R. Sarno and K. R. Sungkono, "Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm," International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2019, pp. 121-126, doi: 10.1109/ISEMANTIC.2019.8884282.(2019)
5. Jiayu Wu, Tianshu Ji "Deep Learning for Amazon Food Review Sentiment Analysis"
6. Yinyang Sasikala P, L.Mary Immaculate Sheela, "Sentiment Analysis of Online Food Reviews using Customer Ratings," Volume 119 No. 15 2018, 3509-3514 ISSN: 1314-3395 (on-line version)(2018)
7. Feiran Y. M. Aye and S. S. Aung, "Enhanced Sentiment Classification for Informal Myanmar Text of Restaurant Reviews," IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA), 2018, pp. 31-36, doi: 10.1109/SERA.2018.8477231.(2018)
8. K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning", 21st International Arab Conference on Information Technology (ACIT), 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300098.(2020)
9. S. M. Asiful Huda, M. M. Shoikot, M. A. Hossain and I. J. Ila, "An Effective Machine Learning Approach for Sentiment Analysis on Popular Restaurant Reviews in Bangladesh," 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2019, pp. 170-173, doi: 10.1109/AiDAS47888.2019.8970976.(2019)
10. Tsung- A. Bhatt, A. Patel, H. Chheda, K. Gawande, — "Amazon Review Classification and Sentiment Analysis" ,IJCSIT, Vol. 6 (6) , 5107-5110(2015)