# Rating food review using sentimental analysis

A Project Report in partial fulfillment of the degree

**Bachelor of Technology**

in

**Computer Science & Engineering & Engineering/ Electronics**

**By**


19K41A05C8 G.Kavya

19K41A05D0 K.Sai Yamini

19K41A05D6 R.Mahitha


**Under the Guidance of**

**Mr Ramesh**

**Submitted to**


**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**S.R.ENGINEERING COLLEGE(A), ANANTHASAGAR, WARANGAL**

**(Affiliated to JNTUH, Accredited by NBA) n o v -2022.**

# DEPARTMENTOF COMPUTER SCIENCE &ENGINEERING

## CERTIFICATE

This is to certify that the Project Report entitled "Rating food recipe using sentimental analysis" is a record of bonafide work carried out by the student(s) G.Kavya, K.Sai Yamini , R.Mahitha, Roll No(s) 19K41A05C8, 19K41A05D0, 19K41A05D6 during the academic year 2021-22 in partial fulfillment of the award of the degree of ***Bachelor of Technology*** in **Computer Science & Engineering/Electronics** by the Jawaharlal Nehru Technological University,Hyderabad.

.

**Supervisor** **Head of the Department**

**External Examiner**

# ABSTRACT

With the growth of web we can find a number of reviews or opinions on any product in many website. Customer spends a lot of time looking for the right product based on the feedback the knowledgeable people share. So, we've created a model that will rate the review given recipe accordingly and it is easy to make decisions. In this paper we apply sentiment analysis on food comments using LSTM algorithm and word to vector. We use sentiment analysis for analyzing comments or reviews into rating. To analysis we will collect comments from web sites and perform pre-processing using natural language processing and apply LSTM algorithm to find class probability to each unique word. This model helps users to select best recipe by visualizing graph shown for a recipe without spending much time in analyzing them.

**Key Words**: food reviews, sentiment analysis, LSTM algorithm, natural language processing, word to vector

# Table of Contents


| S.NO | Content | Page No |
|---|---|---|
| 1 | Introduction | 1 |
| 2 | Literature Review | 2 |
| 3 | Design | 4 |
| 4 | Dataset | 6 |
| 5 | Pre-processing | 7 |
| 6 | Methodology | 9 |
| 7 | Results | 12 |
| 8 | Conclusion | 13 |
| 9 | References | 14 |

# 1. INTRODUCTION

As per today's internet world we can find hundreds of reviews for any product. Customers want to select best product. To select best they will analyse opinions of experienced people that is how many of them are saying that the product is good and bad. The time taken to analyse each and every product is very high. Even though there are star rating it may not be trusted and we will not know the reason for the best or worst. There are several food websites with recipes on how to cook. In this websites people share their experience about each recipe after cooking. Some people accept food is tasty and others may not. If there is a model which will automatically analyse user comments based on rating will be very useful to the customers to select best recipe in less time. The opinions or reviews given by user are in natural language which is not understood by the machine. Sentiment analysis is a technique which makes machine to understand the human language. Sentimental analysis is a process of determining a piece of writing into positive, negative and neutral. Sentimental analysis helps large-scale data analysts collect public opinion, perform market research, track brand and product credibility and appreciate client experience.

Sentimental analysis is a process of determining a piece of writing into positive, negative and neutral. Sentimental analysis helps large-scale data analysts collect public opinion, perform market research, track brand and product credibility and appreciate client experience. "Opinion mining" is also known as emotional research. The sentimental research has three different levels of reach. Sentiment analysis can be implemented through lexicon based, machine learning, hybrid based method. In lexicon based is one of the two main approaches to sentiment analysis and it involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text. The approach to machine learning makes use of supervised learning techniques. Supervised learning uses labeled data to rate test data positively or negatively. The combination of both lexicon and machine learning approach is hybrid based method. Hybrid method gives more accurate results.

## 2. Literature Survey

Hybrid method gives more accurate results Anshuman, shivani rao and misha kakkar [1]have used sentiment analysis to sort the recipes when an ingredient name is given as input. To sort the recipe they have used sentiment analysis of lexicon-based method. The reviews for number of recipes from various different sites were fetched out and through lexicon-based approach they were analysed. A bag of positive and negative words were used to rate the reviews based on word score comparision.
Reviews that has highest score was ranked first position and so on.

In this paper [2] has done lexicon based sentiment analysis on food recipe comments. In this paper they have classified food recipe comments from a community into positive, negative and neutral. Classification is done by identifying the polarity words from the sentence and by calculating polarity score. Using this method the accuracy score for positive comments is 90% and for negative 70%.

Here Sasikala and Mary inmaculate sheela[3] has done sentiment analysis using lexicon based method on food reviews based on customer rating. They have implemented it using r programming. The opinion word or polarity word from the sentence they have performed pre-processing. All the opinion words and its count are represented in matrix format. Any machine learning algorithm can be used to get the expected result.

Kavya suppala and narasinga rao[4] has used sentiment analysis of LSTM classifier on tweet data to compare between different tweets. In this they have collected tweets of previous data to train the model and using this labelled data they have predicted test data.

[5] In this experiment we have used amazon fine food reviews to train the model. It contains 5,68,454 reviews for 74,258 recipes. For experiment purpose we have used 2,00,000 comments for 3000 recipes. We have created a user interface where user select a recipe name for which we wants to observe the analysis of reviews. After submitting they can observe a pie graph with number of positive and negative comments for that recipe.

[6] in this paper propose a method for ranking recipes of a kind of food according to the “classicality” of each recipe.in first set to remove inconsistent which is in the system.in next step recognize the ingredients. after that algorithms on dataset.

In paper (7) authors proposed a tool that analyses much content on comments text. This helps the user to make their decision about the food recipe. YouTube is a platform where users can upload, rate, view, share, report, add to favorites, comment on the videos and subscribe to the channels. In this system SVM, Naïve Bayes algorithms, K-Nearest Neighbor, Deep learning are used also used some machine learning algorithms for get the better result. The first step is text pre-processing after that system recognize that text. After in next step apply machine learning algorithms and sentiment analysis on the dataset. At last users get the result based on the reviews of recipe.

In paper (8) Opinion Mining and Sentiment Analysis are critical tools for information-gathering to find out what people are thinking. In this system first user will enter the recipe name on YouTube and search the recipe exactly they want. After that using the of YouTube videos of the cooking recipe entered in the field. Then, the system collects the generated comments on these recipes videos and stores them in the database. In the next step splits the comments in some fragments according to the previous steps. In the next step system extract the polarity according to the features and at last users get the result and it shows the overall percentage, and the number of the Likes, Dislikes, and Views of each recipe video.

In paper (9) author create Foodoholic application and performed sentiment analysis on food recipe by developing an application. The objective of the application is to rank various recipes having core ingredient based on reviews. This saves time of the users searching for the best recipe for a particular ingredient. Sentiment Analysis is widely applied to reviews or social media to discover how people or customer feels about some text in a document or a sentence. whether the expressed opinion in document or sentence is positive, negative or neutral. In the next step the target of Sentiment Analysis is to find the opinions, feedback or review, and then identify the sentiment they want to express and at last get the result.

In paper (10) author use various tools which extracts YouTube cooking recipes comments automatically. In this system first user will enter the recipe name on YouTube and search the recipe exactly they want. After that using the YouTube APIs videos of the cooking recipe entered in the field. Then, their tool collects the generated comments on these recipes videos and stores them in a database. In the next step feature selection and preprocessing of dataset. After that various algorithms apply and the system filter automatically opinions reviews from the generated comments, and eliminating texts that bear no opinion by classifying the generated comments in the classes (opinion, other) using SVM classifier.

.

# 3. DESIGN:

## 1. REQUIREMENT SPECIFICATION (S/W& H/W)

**Hardware Requirements**

- **System**
- **RAM**
- **Hard Disk**
- **Input**
- **Output**

: Pentium 4, Intel Core i3, i5, i7 and 2GHz
Minimum: 4GB or above
: 10GB or above

: Keyboard and

Mouse: Monitor or PC

**Software Requirements**

- **OS**
- **Platform**
- **Program Language**

: Windows 8 or Higher
Versions: Jupiter
Notebook/visual studio code
: Python

### 3.2 Flowchart

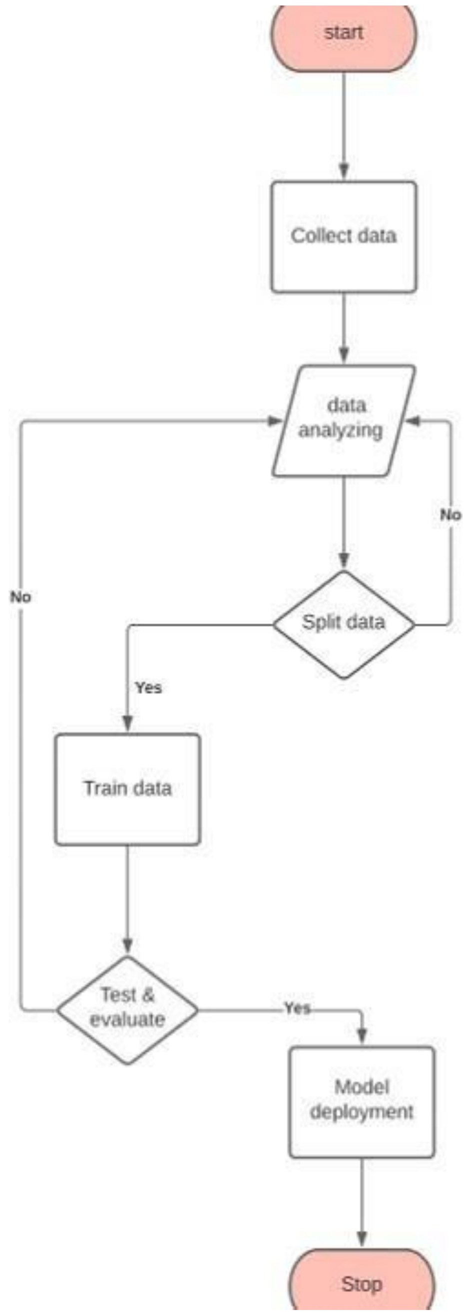
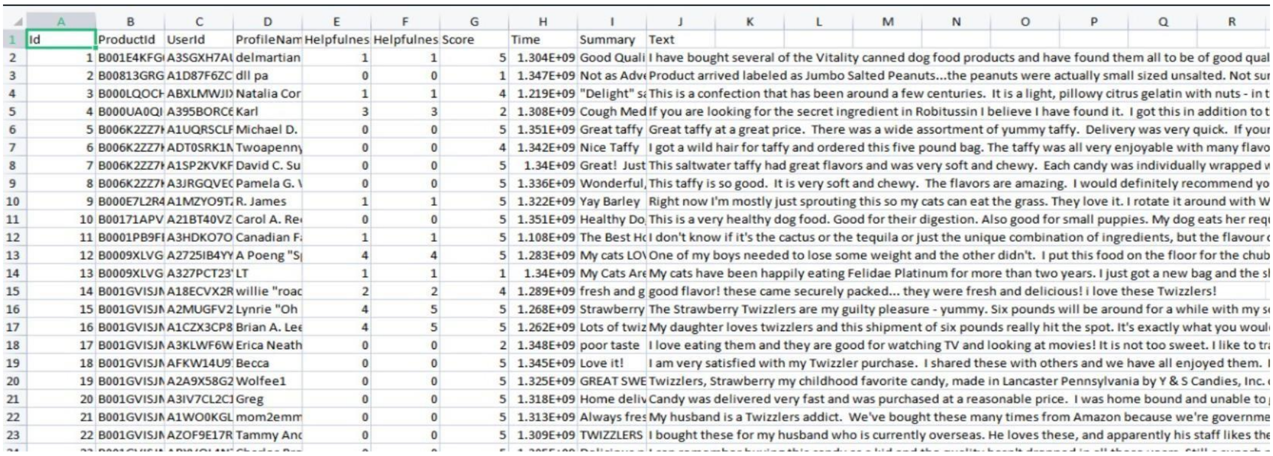



**Figure** 1 **-** flow chart

# 4. DATASET:



| Id | ProductId | UserId | ProfileNam | Helpfulnes | Helpfulnes | Score | Time | Summary | Text |
|---|---|---|---|---|---|---|---|---|---|
| 1 | B001E4KFG | A3SGXH7AU | delmartian | 1 | 1 | 5 | 1.304E+09 | Good Quali | I have bought several of the Vitality canned dog food products and have found them all to be of good qual |
| 2 | B00813GRG | A1D87F6ZC | dll pa | 0 | 0 | 1 | 1.347E+09 | Not as Adv | Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sur |
| 3 | B000LQOCH | ABXLMWJI | Natalia Cor | 1 | 1 | 4 | 1.219E+09 | "Delight" s | This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in t |
| 4 | B000UA0QI | A395BORC6 | Karl | 3 | 3 | 2 | 1.308E+09 | Cough Med | If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to t |
| 5 | B006K2ZZ7 | A1UQRSCLF | Michael D. | 0 | 0 | 5 | 1.351E+09 | Great taffy | Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If your |
| 6 | B006K2ZZ7 | ADT0SRK1N | Twoapenny | 0 | 0 | 4 | 1.342E+09 | Nice Taffy | I got a wild hair for taffy and ordered this five pound bag. The taffy was all very enjoyable with many flavo |
| 7 | B006K2ZZ7 | A1SP2KVKF | David C. Su | 0 | 0 | 5 | 1.34E+09 | Great! Just | This saltwater taffy had great flavors and was very soft and chewy. Each candy was individually wrapped |
| 8 | B006K2ZZ7 | A3JRGQVE | Pamela G. | 0 | 0 | 5 | 1.336E+09 | Wonderful, | This taffy is so good. It is very soft and chewy. The flavors are amazing. I would definitely recommend yo |
| 9 | B000E7L2R4 | A1MZYO9T | R. James | 1 | 1 | 5 | 1.322E+09 | Yay Barley | Right now I'm mostly just sprouting this so my cats can eat the grass. They love it. I rotate it around with W |
| 10 | B00171APV | A21BT40VZ | Carol A. Re | 0 | 0 | 5 | 1.351E+09 | Healthy Do | This is a very healthy dog food. Good for their digestion. Also good for small puppies. My dog eats her requ |
| 11 | B0001PB9FI | A3HDKO7O | Canadian F | 1 | 1 | 5 | 1.108E+09 | The Best H | I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavour |
| 12 | B0009XLVG | A2725IB4YY | A Poeng "S | 4 | 4 | 5 | 1.283E+09 | My cats LO | One of my boys needed to lose some weight and the other didn't. I put this food on the floor for the chub |
| 13 | B0009XLVG | A327PCT23 | LT | 1 | 1 | 1 | 1.34E+09 | My Cats Ar | My cats have been happily eating Felidae Platinum for more than two years. I just got a new bag and the s |
| 14 | B001GVISJN | A18ECVX2R | willie "roac | 2 | 2 | 4 | 1.289E+09 | fresh and g | good flavor! these came securely packed... they were fresh and delicious! i love these Twizzlers! |
| 15 | B001GVISJN | A2MUGFV2 | Lynrie "Oh | 4 | 5 | 5 | 1.268E+09 | Strawberry | The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be around for a while with my s |
| 16 | B001GVISJN | A1CZX3CP8 | Brian A. Le | 4 | 5 | 5 | 1.262E+09 | Lots of twiz | My daughter loves twizzlers and this shipment of six pounds really hit the spot. It's exactly what you woul |
| 17 | B001GVISJN | A3KLWF6W | Erica Neath | 0 | 0 | 2 | 1.348E+09 | poor taste | I love eating them and they are good for watching TV and looking at movies! It is not too sweet. I like to tr |
| 18 | B001GVISJN | AFKW14U9 | Becca | 0 | 0 | 5 | 1.345E+09 | Love it! | I am very satisfied with my Twizzler purchase. I shared these with others and we have all enjoyed them. |
| 19 | B001GVISJN | A2A9X58G2 | Wolfee1 | 0 | 0 | 5 | 1.325E+09 | GREAT SWE | Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylvania by Y & S Candies, Inc. |
| 20 | B001GVISJN | A3IV7CL2C | Greg | 0 | 0 | 5 | 1.318E+09 | Home deliv | Candy was delivered very fast and was purchased at a reasonable price. I was home bound and unable to |
| 21 | B001GVISJN | A1WO0KGL | mom2emm | 0 | 0 | 5 | 1.313E+09 | Always fres | My husband is a Twizzlers addict. We've bought these many times from Amazon because we're governme |
| 22 | B001GVISJN | AZOF9E17R | Tammy Anc | 0 | 0 | 5 | 1.309E+09 | TWIZZLERS | I bought these for my husband who is currently overseas. He loves these, and apparently his staff likes the |

**Figure** 2 Visualizing attributes of the dataset

# 5. DATAPREPROCESSING:

**Stemming:**
Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming is important in natural language understanding (NLU) and natural language processing (NLP). Stemming is a part of linguistic studies in morphology and artificial intelligence (AI) information retrieval and extraction. Stemming and AI knowledge extract meaningful information from vast sources like big data or the Internet since additional forms of a word related to a subject may need to be searched to get the best results Recognizing, searching and retrieving more forms of words returns more results. When a form of a word is recognized it can make it possible to return search results that otherwise might have been missed. That additional information retrieved is why stemming is integral to search queries and information retrieval.

**Removing stop words:**
The words which are generally filtered out before processing a natural language are called stop words. These are actually the most common words in any language (like articles, prepositions, pronouns, conjunctions, etc) and does not add much information to the text. Examples of a few stop words in English are “the”, “a”, “an”, “so”, “what”. Stop words are available in abundance in any human language. By removing these words, we remove the low-level information from our text in order to give more focus to the important information. In order words, we can say that the removal of such words does not show any negative consequences on the model we train for our task. Removal of stop words definitely reduces the dataset size and thus reduces the training time due to the fewer number of tokens involved in the training.
NLP is one of the most researched areas today and there have been many revolutionary developments in this field. NLP relies on advanced computational skills and developers across the world have created many different tools to handle human language. Out of so many libraries out there, a few are quite popular and help a lot in performing many different NLP tasks.

**Tokenization**
Tokenization is the first step in any NLP pipeline. It has an important effect on the rest of your pipeline. A tokenizer breaks unstructured data and natural language text into chunks of information that can be considered as discrete elements. The token occurrences in a document can be used directly as a vector representing that document. This immediately turns an unstructured string (text document) into a numerical data structure suitable for machine learning. They can also be used directly by a computer to trigger useful actions and responses. Or they might be used in a machine learning pipeline as features that trigger more complex decisions or behavior.

## punctuation removal:

The punctuation removal process will help to treat each text equally. For example, the word data and data! are treated equally after the process of removal of punctuations. We need to take care of the text while removing the punctuation because the contraction words will not have any meaning after the punctuation removal process. Such as ‘don’t’ will convert to ‘dont’ or ‘don t’ depending upon what you set in the parameter. We also need to be extra careful while choosing the list of punctuations that we want to exclude from the data depending upon the use cases. As string. punctuation in python contains these symbols !"#$%&\'()*+,-./:;?@[\\]^_{|}~`

# 6. METHODOLOGY:

This section talks about the algorithms used for the project. We used LSTM(Long Short Term Memory) and Word to Vector.

### 6.1 LSTM(LONG SHORT TERM MEMORY)

deep learning model based LSTM(Long short term memory) method for spam detection in email.

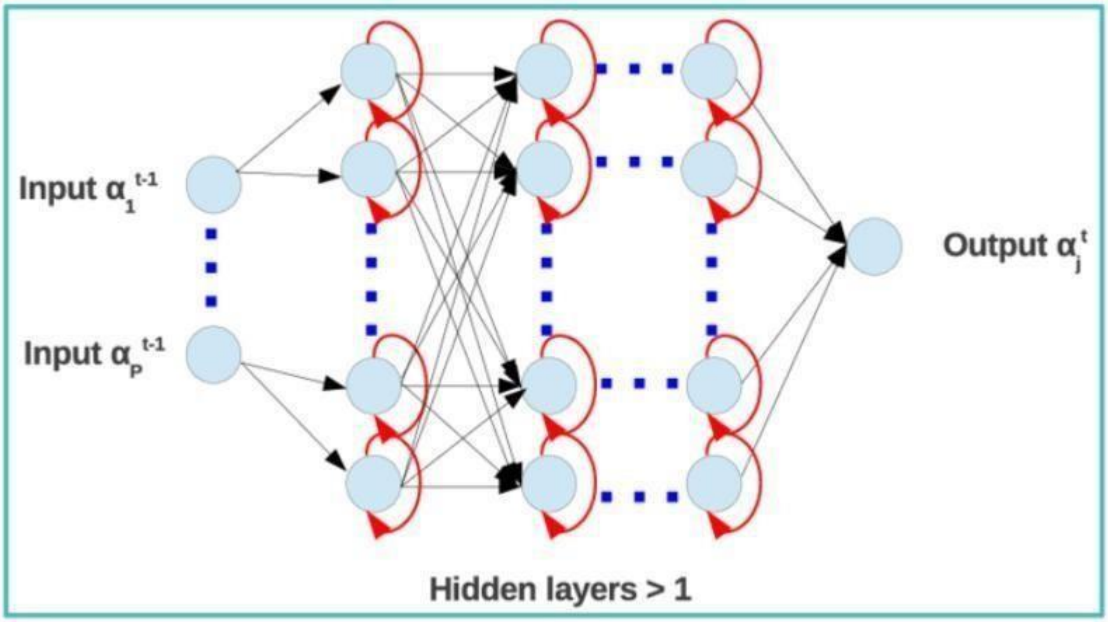



Fig 4:LSTM Model

LSTMs have three types of gates they are:

Input gates, forget gates, and output gates which controls the flow of information. The hidden layer output of LSTM includes the hidden state and the memory cell. Only the hidden state is passed into the output layer. The memory cell is entirely internal.

This challenge to address long-term information preservation and short-term input skipping in latent variable models has existed for such a long time. One of the earliest approaches to address this was the long short-term memory (LSTM). It shares many of the properties of the GRU. Interestingly, LSTMs have a slightly more complex design than GRUs but predates GRUs by almost two decades.

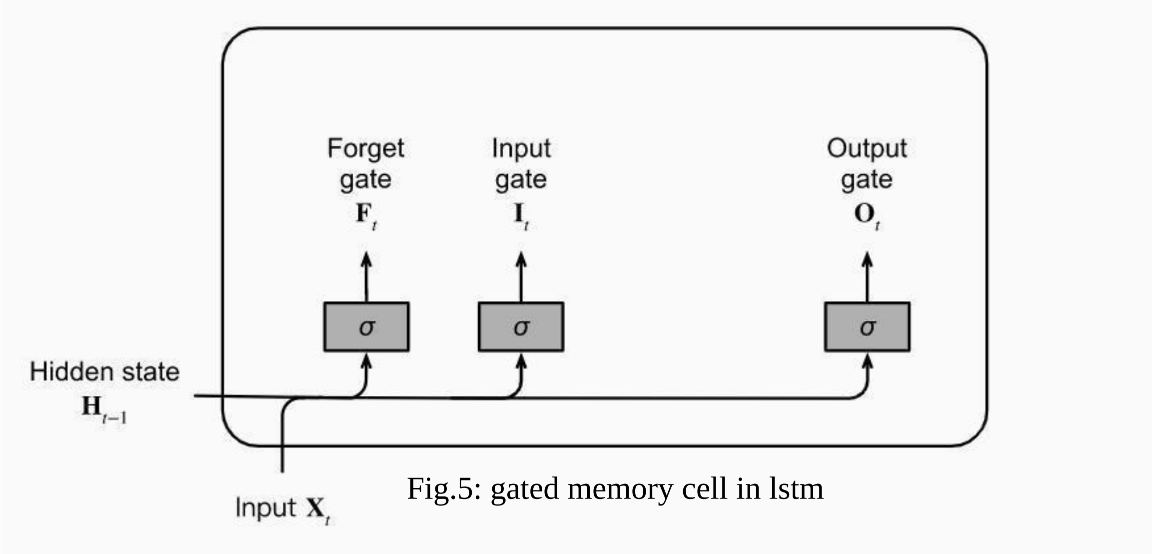


Fig.5: gated memory cell in lstm

## 6.2 WORD TO VECTOR

Word2Vec model is used for Word representations in Vector Space which is founded by Tomas Mikolov and a group of the research teams from Google in 2013. It is a neural network model that attempts to explain the word embeddings based on a text corpus.

These models work using context. This implies that to learn the embedding, it looks at nearby words; if a group of words is always found close to the same words, they will end up having similar embeddings. To label how words are similar or close to each other, we first fix the **window size,** which determines which nearby words we want to pick.

The General Flow of the Algorithm

- Step-1: Initially, we will assign a vector of random numbers to each word in the corpus.
- Step-2: Then, we will iterate through each word of the document and grab the vectors of the nearest n-words on either side of our target word, and concatenate all these vectors, and then forward propagate these concatenated vectors through a linear layer + softmax function, and try to predict what our target word was.
- Step-3: In this step, we will compute the error between our estimate and the actual target word and then backpropagated the error and then modifies not only the weights of the linear layer but also the vectors or embeddings of our neighbor's words.
- Step-4: Finally, we will extract the weights from the hidden layer and by using these weights encode the meaning of words in the vocabulary.
- Word2Vec model is not a single algorithm but is composed of the following two pre-processing modules or techniques:

- Continuous Bag of Words (CBOW)
- Skip-Gram.
- Both of the mentioned models are basically shallow neural networks that map word(s) to the target variable which is also a word(s). These techniques learn the weights that act as word vector representations. Both these techniques can be used to implementing word embedding using word2vec.
- Why Word2Vec technique is created?
- As we know that most of the NLP systems treat words as atomic units. In existing systems with the same purpose as that of word2vec, there is a disadvantage that there is no notion of similarity between

words. Also, those system works for small, simpler data and outperforms on because of only a few billions of data or less.

- So, In order to train the system with a larger dataset with complex models, these techniques use a neural network architecture to train complex data models and outperform huge datasets with billions of words and with vocabulary having millions of words.
- It helps to measure the quality of the resulting vector representations and works with similar words that tend to close with words that can have multiple degrees of similarity.
- **Syntactic Regularities:** These regularities refer to grammatical sentence correction.
- **Semantic Regularities:** These regularities refer to the meaning of the vocabulary symbols arranged in that structure.
- The proposed technique was found that the similarity of word representations goes beyond syntactic regularities and works surprisingly well for algebraic operations of word vectors.

## 7.RESULTS:

```
CV 1/5; 67/80] END lr__C=100.0, lr__penalty=l1, lr__solver=sag;, score=nan total time=   0.0s
CV 2/5; 67/80] START lr__C=100.0, lr__penalty=l1, lr__solver=sag...............
CV 2/5; 67/80] END lr__C=100.0, lr__penalty=l1, lr__solver=sag;, score=nan total time=   0.0s
CV 3/5; 67/80] START lr__C=100.0, lr__penalty=l1, lr__solver=sag...............
CV 3/5; 67/80] END lr__C=100.0, lr__penalty=l1, lr__solver=sag;, score=nan total time=   0.0s
CV 4/5; 67/80] START lr__C=100.0, lr__penalty=l1, lr__solver=sag...............
CV 4/5; 67/80] END lr__C=100.0, lr__penalty=l1, lr__solver=sag;, score=nan total time=   0.0s
CV 5/5; 67/80] START lr__C=100.0, lr__penalty=l1, lr__solver=sag...............
CV 5/5; 67/80] END lr__C=100.0, lr__penalty=l1, lr__solver=sag;, score=nan total time=   0.0s
CV 1/5; 68/80] START lr__C=100.0, lr__penalty=l1, lr__solver=saga..............
CV 1/5; 68/80] END lr__C=100.0, lr__penalty=l1, lr__solver=saga;, score=0.894 total time=33.8min
CV 2/5; 68/80] START lr__C=100.0, lr__penalty=l1, lr__solver=saga..............
CV 2/5; 68/80] END lr__C=100.0, lr__penalty=l1, lr__solver=saga;, score=0.897 total time=32.7min
CV 3/5; 68/80] START lr__C=100.0, lr__penalty=l1, lr__solver=saga..............
```

**Figure** 8**.** Result of various models with the proposed model

The neural network deep learning algorithms that we used is LSTM(long-short term memory). This algorithms worked well on spam mail detection. We got 89.7% accuracy. LSTM has four layers 1)LSTM layer-1 2)LSTM layer-2 3)drop out layer 4)dense layer.

## 7. CONCLUSION:

The previous or existing spam mail detection systems used traditional text based machine learning models. The results highly rely on the crafted extracted features. The performances are unstable when detecting spam mails.

So, we propose a deep learning model based LSTM(Long short term memory) method for spam mail detection. The neural network deep learning algorithms that we used is LSTM(long-short term memory). This algorithms worked well on spam mail detection system. We got 89.7% accuracy.

During computation of long text mails which are far away, it is impossible to store which causes vanishing of gradient. In order to maintain we use LSTM(Long Short-Term Memory Network)

## 8. REFERENCES:


1) [1] B. Prasad Majumder, S. Li, J. Ni, and J. McAuley, “Generating personalized recipes from historical user preferences,” Empirical Methods in Natural Language Processing (EMNLP), 2019. 5
2) [2] M. Harvey, B. Ludwig, and D. Elswiler, “You are what you eat: Learning user tastes for rating prediction,” Lecture Notes in Computer Science, 2013.
3) [3] J. Freyne and S. Berkovsky, “Intelligent food planning: personalized recipe recommendation,” 2010.
4) [4] C.-Y. Teng, Y.-R. Lin, and L. Adamic, “Recipe recommendation using ingredient networks,” 2012.
5) [5] X. Mao, R. Yanghui, and Q. Li, “Recipe popularity prediction based on the analysis of social reviews,” 2013.
6) [6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in North American Chapter of the Association for Computational Linguistics, 2019.
7) [7] https://www.irjet.net/archives/V7/i3/IRJET-V7I3387.pdf
8) [8] https://www.researchgate.net/publication/348603412_Sentiment_Analysis_for_Rating_Cooking_Recipes_Based_on_User%27s_Feedback_Using_Data_Mining_TechniquesClassification
9) [9] https://nevonprojects.com/cooking-recipe-rating-based-on-sentiment-analysis/

[10] https://www.irjet.net/archives/V7/i3/IRJET-V7I3387.pdf