JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Comparative analysis of different Deep learning and Machine learning algorithms of User Rating Classification and Sentiment analysis

Krishan Kumar[1*], Dr.Randeep Singh[2]

[1*]Research Scholar Department of Computer Science & Engineering, IEC University Baddi Solan (HP), India,
[2]Professor Department of Computer Science & Engineering, IEC University Baddi, Solan (HP), India

Abstract*—Sentiment analysis of user reviews is crucial for understanding customer opinions and feedback in online marketplaces like Amazon. In this study, a Deep Belief Recurrent Neural Network (DBRNN) based approach is proposed for user rating classification of Amazon food reviews. The process is initiated by pre-processing the raw text data, converting it into numerical vectors using word embedding. The DBRNN architecture is then employed to automatically learn hierarchical representations from the review data, capturing complex patterns and interactions among words and phrases. To evaluate the performance of the approach, a large dataset of Amazon food reviews is utilized and assess the classification accuracy. The experimental results demonstrate that the DBRNN-based approach achieves competitive performance in accurately classifying user ratings. Also the contributions of this work lie in providing a comprehensive comparison of deep learning and machine learning algorithms such as Support vector machine (SVM) and Long Term Short Memory (LSTM) for sentiment analysis in the context of Amazon food reviews. The Hybrid approach shows more accuracy compared to SVM and LSTM.

Index Terms*—SVM, LSTM, DBN, DBRNN, sentiment analysis.

I. INTRODUCTION

Sentiment analysis and user rating classification are computational approaches aimed at tackling the task of extracting subjective information, such as opinions and sentiments, from natural language texts. Various methodologies have been employed to address this issue, encompassing disciplines such as natural language processing, text analysis, computational linguistics, and biometrics. In recent years, there has been a surge in the popularity of machine learning algorithms for semantic and review analysis due to their simplicity and accuracy. The primary objective of sentiment classification is to assess textual reviews provided by users and categorise them as either expressing positive or negative sentiments. This approach does not necessitate a comprehensive understanding of the semantic nuances present in individual phrases or documents [1-3]. Sentiment categorization has been explored across various domains, including but not limited to movie reviews, travel location reviews, and product reviews [4-5].

The food reviews on Amazon frequently comprise textual depictions authored by customers, accompanied by star ratings or numerical evaluations that denote the overall degree of contentment with the product. The reviews encompass a spectrum of brevity, ranging from concise remarks to comprehensive narratives that outline the customer's encounter with the culinary item.

The examination of food reviews on Amazon can yield advantages for organisations, as it offers useful insights from customers, aids in product enhancement, and contributes to heightened customer contentment. Furthermore, prospective consumers can enhance their decision-making process by taking into account the firsthand experiences of other individuals prior to acquiring a specific food product on the Amazon platform.

The advent of social media and online reviews has led to the significance of sentiment analysis as a prominent field of study in the domain of Machine Learning. Sentiment analysis has the potential to enhance various domains such as product development, marketing strategies, clinical medicine, and service delivery

in urban areas. The task of sentiment analysis presents a captivating endeavour because to its requirement for computational systems to engage with human language, specifically in the realm of natural language processing. This entails the ability to discern and extract subjective information. Sentiment analysis often employs several methods such as Naive Bayes, Support Vector Machine, and Maximum Entropy. In light of recent advancements in recurrent neural networks, Deep Learning has emerged as a superior approach compared to previous methods. This can be attributed to its capacity to construct a comprehensive sentence representation by leveraging the underlying structure of the sentence.

The task at hand pertains to supervised learning, wherein the objective is to forecast the binary target variable, indicating either a positive or negative sentiment, for each individual review. The objective is to optimise the precision of this categorization. The model would undergo training using a dataset consisting of a large number of reviews that are represented in the form of unstructured text. Every review would be categorised as either favourable or negative. In order to address this issue, a series of tasks would be executed.

1. Pre-process the data.
2. Train and tune the hyper parameters of the recurrent neural network.
3. Test the accuracy of the model on the testing set.

Metrics:

To assess the performance of the model, the F1 score would be employed as a metric to measure accuracy. Both precision and recall are given equal importance in this approach. When evaluating fine cuisine reviews, it is logical to assign equal importance to both precision and memory. Nevertheless, when it comes to the field of clinical medicine, it is imperative to place a greater emphasis on precision. The optimal F1 score is 1, representing perfect precision and recall, whereas the lowest attainable F1 score is 0, indicating complete absence of precision and recall.

F1 = 2 * precision*recall / (precision + recall)

II. LITERATURE REVIEW

Wu and Ji [6] conducted a study whereby they showcased the application of Recursive Neural Network inside the domain of sentiment analysis tasks. A methodology has been proposed and executed to analyse binary trees by utilising the Stanford NLP Parser in order to handle the unprocessed textual information extracted from Amazon Fine Food Reviews. Additionally, the researchers proposed a novel approach for assigning labels to tree nodes, aiming to provide the requisite level of supervision for Recurrent Neural Networks (RNNs). The present methodology was devised in order to tackle the problem of inadequate annotation within the original dataset. The researchers have recently presented a new model known as RNNMS (Recursive Neural Network for Multiple Sentences), which exhibits enhanced performance in comparison to their baseline model across all assessed criteria.

The study conducted by Aravindan et al. [7] investigated various approaches for sentiment analysis using large datasets of Amazon Fine Food reviews. The researchers utilised the Apache Spark data processing engine and MLlib, the machine learning package designed for Apache Spark, to investigate three distinct methodologies: Linear Support Vector Classification (SVC), Logistic Regression, and Naive Bayes. The aforementioned methodologies attained accuracy rates of 80%. The study revealed that linear support vector classification (SVC) had greater performance in comparison to naïve Bayes (NB) and logistic regression. The researchers conducted an empirical investigation and found that sentiment analysis conducted within a big data framework demonstrates enhanced computing efficiency when compared to sentiment analysis conducted without the use of the map-reduce framework.

Paknejad (2018) [8] presented a thorough examination of the core principles associated with the examination and classification of customer reviews on the Amazon platform. Machine learning approaches, such as Support Vector Machine and Naive Bayes, are utilised in the implementation of a classification technique. The process of categorising is conducted to classify reviews into positive and negative categories, with the aim of assisting users in gaining a more comprehensive understanding of the products. Supervised learning methodologies are also utilised inside this particular situation. This study activity involves a comparison analysis aimed at identifying the most effective algorithms for classifying reviews in a manner that exhibits progressive improvement.

Fang and Zhan (2019) [9] employed the natural language processing methodology to analyse the reviews and arrange them in a predetermined sequence. This work efficiently addresses the challenge of categorising the polarity of the reviews. The data utilised in this evaluation consists of online product reviews collected

from Amazon.com. Promising results have been obtained from tests conducted on both sentence-level categorization and survey-level layout. Ultimately, an enhanced comprehension of the process is also acquired. This report additionally elucidates the various impediments and complexities associated with the present research endeavour.

The distant supervision methodology was utilised by Go et al. [10] in their investigation. Furthermore, many machine learning techniques, including Naive Bayes, Support Vector Machine, and Maximum Entropy, are utilised for the aim of evaluation and classification. The preceding messages can be classified as either positive or negative in relation to a query word. This particular characteristic proves advantageous for consumers who desire the chance to evaluate the sensory encounter of products before committing to a purchase, as well as for firms aiming to examine the general public's view of their brands. The aforementioned examination demonstrates a significant degree of precision, with an estimated accuracy rate of roughly 80 percent.

Abah (2011) [11] conducted a study with the objective of comprehending consumers' viewpoints on reviews and obtaining insight into their cognitive processes. To ensure the analysis is conducted properly, the author employs deep learning techniques, specifically long- and short-term memory and convolutional neural networks. The objective of this study is to methodically distinguish and classify the level of opinion extremism within the dataset of Amazon Electronic reviews. The inflammatory message is converted into its respective word vector representation with GloVe Embeddings. The selected models are assessed utilising parameters such as Exactitude, Precision, Recall, and F1 Score. Both models have a level of accuracy over 90 percent. The findings indicate that the models possess the ability to effectively depict the customer reviews with a high degree of accuracy.

In this study, Minaee et al. (2012) [12] present a thorough examination of the utilisation of deep learning methodologies within the domain of review and text categorization. The aforementioned models have exhibited enhanced efficacy in comparison to traditional machine learning methods across several tasks pertaining to message categorization, encompassing analysis, news organisation, question answering, and natural language production. This study presents a detailed survey of various deep learning models utilised for text classification, which have emerged in the past few years. The authors engage in a comprehensive analysis of the technical contributions, commonalities, and merits of these models. Ultimately, the authors provide a comprehensive quantitative review of the efficacy of various deep learning models on established benchmarks, while also delving into prospective areas for further investigation.

In this investigation, the researchers Shehu et al. (2013) [13] utilised Deep learning algorithms integrate the utilisation of both long-term and short-term memory. This work utilises an Optimisation Algorithm to perform review classification and sentiment analysis, making a unique contribution to the field of research. The applied model was utilised to extract consumer comments that were acquired from product reviews on the Amazon platform. The display of the produced collaborative model exhibits a notable degree of precision, correctness, review, and F1 ratio. However, when compared to the LSTM model on an individual basis, the latter demonstrates a significantly favourable performance rate.

Zola et al. (2014) [14] proposed a methodology for categorising reviews that can be utilised in many fields and from diverse sources. The author introduces an innovative methodology for conducting a thorough examination of multiple sources across various fields. This methodology entails harnessing cross-domain data from significant internet platforms, such as Amazon, and employing deep learning algorithms to construct structured models. Subsequently, these models are assessed by employing unannotated user-generated content derived from widely used social media platforms such as Facebook and Twitter. The researchers conducted an investigation into a three-pronged methodology, wherein they investigated various strategies for adjusted training, text pre-processing, and machine learning. This investigation was carried out using two different languages. The individuals under consideration hold citizenships of both England and Italy. The usage of under-sampling training with a Convolutional Neural Network yielded the most optimal outcomes. In contrast to the process of integration, deep learning algorithms often exhibit higher performance in comparison to other tactics.

Coyne and colleagues (2015) [15] undertook a research study with the objective of assessing the suitability of sentiment analysis algorithms for analysing product reviews obtained from the online marketplace Amazon.com. This research entails the examination, instruction, and evaluation of diverse machine learning algorithms utilising a dataset consisting of product assessments obtained from Amazon.com. The dataset was selected in a random manner from a larger dataset available on Kaggle, which consists of a comprehensive collection of 4 million reviews. A comparative analysis was undertaken to evaluate the efficacy of three unique algorithms, namely Multinomial Naive Bayes (MNB), Linear Support Vector Machine (LSVM), and

Long short-term memory network (LSTM). The Long Short-Term Memory (LSTM) model exhibited exceptional performance, with an accuracy of 0.90 and an area under the receiver operating characteristic curve (AUROC) of 0.96. Following that, the Long Short-Term Memory (LSTM) model was employed to evaluate the remaining 3.94 million reviews gathered from the Kaggle dataset, along with a recently acquired dataset from Amazon.com that included reviews related to different product categories.

Chen and Hendry (2016) [16] proposed a novel approach for sentiment analysis by employing the Deep Belief network along with a noise reduction methodology. The rating prediction technique relies on user feedback and is executed through three distinct stages. The first step in the procedure entails the elimination of noise using pre-processing techniques. Furthermore, the model utilises Sentiword for the aim of extracting features. In addition, it utilises a deep learning methodology known as Deep Belief Networks (DBN). The system has a significant accuracy rate of 72.83%, surpassing the performance of the baseline models MLP, CNN, and LibSVM.

Xiao et al. (2017) [17] presented a methodology aimed at tackling the sentiment classification difficulty in imbalanced datasets by employing transfer learning techniques. The proposed methodology entails the utilisation of under-sampling approaches and a pre-trained text model that is based on transfer learning, with the aim of fine-tuning. The efficacy of the approach hinges upon the utilisation of a dataset for sentiment analysis that accurately reflects real-life situations. The applicability of the model, which incorporates transfer learning with under sampling for sentiment analysis and operates through post verification, is yet undetermined in its usefulness for various text categorization tasks.

III. PROPOSED WORK

Data Exploration

The dataset comprises a total of 568,454 food reviews provided by Amazon users, spanning the time period from October 1999 to October 2012. Figure 1 shows a description of attributes used in dataset.

Name	Description
ProductId	unique identifier for the product
UserId	unique identifier for the user
HelpfulnessNumerator	number of users who found the review helpful
HelpfulnessDenominator	number of users who indicated whether they found the review helpful
Score	rating between 1 and 5
Time	timestamp for the review
Summary	brief summary of the review
Text	text of the review

Figure 1 Dataset Description

The textual content pertaining to the evaluation of a particular product as provided by an individual user is maintained under the Text column. The Text column comprises unstructured textual data. In addition to the representation of the review through the Text feature, the identification of the product can be achieved by utilising the ProductId feature. Furthermore, the Amazon user associated with the review can be determined by examining the UserID feature. The Time column provides information regarding the date when the review was posted, while a concise summary of the review can be extracted from the Summary variable. In addition, the review includes two variables, Helpfulness Numerator and Helpfulness Denominator, which indicate the level of helpfulness attributed to the review.

Significantly, in relation to each review, there exists a variable denoted as "Score" which signifies a rating ranging from 1 to 5. It is noteworthy that a score of 1 indicates a negative assessment, while a score of 5 signifies an exceptionally positive review.

Sentiment Score:

Figure 2 shows the score distribution. Here is the distribution of the score:

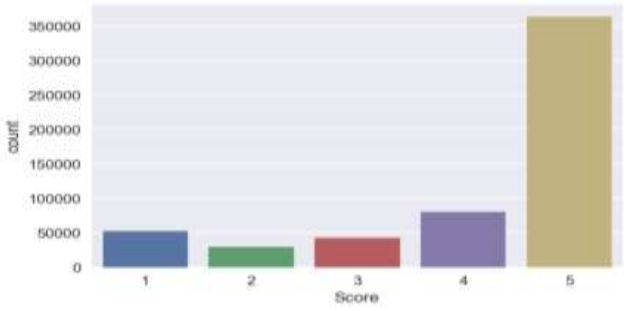

Figure 2 Score Distribution

It is evident from the data that a significant proportion of scores fall within the range of 4 and 5, with an average score of 4.18. Due of the highly left-skewed distribution, a binary prediction would be employed. Negative reviews can be classified as having a score ranging from 1 to 3, while positive reviews can be categorised as having a score of either 4 or 5.

Upon the establishment of a novel variable denoted as "Sentiment," which assumes the value of "NEGATIVE" when the Score falls within the range of 1 to 3, and "POSITIVE" when the Score is either 4 or 5, a subsequent distribution emerges. Figure 3 shows the analysis of positive sentiments and negative distribution.

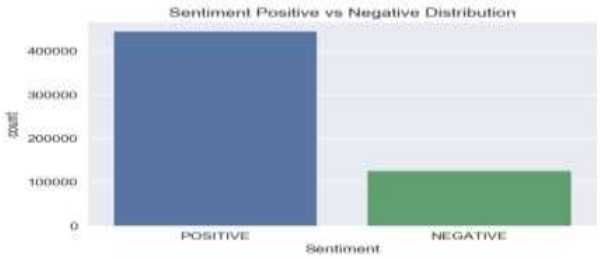

Figure 3 Sentiment positive vs Negative Distribution

With this new classification, 78% of fine food reviews are considered positive and 22% of them are considered negative.

Text Review

Now let's explore the input variable "text" which contains the review of the product. Figure 4 shows the review of text as positive and negative.

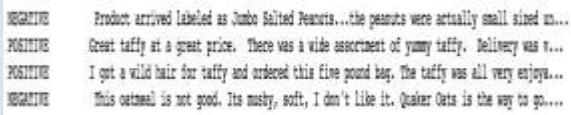

Figure 4 Review of Text

The input variable represents a text that lacks a predetermined structure. By examining only the initial 90 characters, one can promptly discern if the review has a favourable or unfavourable sentiment. In the context of negative reviews, individuals commonly employ negative terms such as "not good" or "don't like it." The aforementioned negative expressions can exhibit varying degrees of complexity, contingent upon the contextual factors, for instance, the phrase "were actually small sized." Positive evaluations are characterised by the use of positive expressions, such as "Great taffy," "great price," "wide assortment," "yummy taffy," "very enjoyable," and so on.

It is imperative to acknowledge that the provided text lacks structure, necessitating pre-processing steps such as punctuation removal and conversion of uppercase characters to lowercase.

Exploratory Visualization

One often employed method for differentiating positive reviews from negative reviews involves analysing the frequency of words used. It is conceivable that specific terms, such as "excellent" or "very tasty," have a higher frequency of occurrence in favourable evaluations compared to negative reviews. Let us endeavour to substantiate this notion.

The findings obtained from the calculation of the ratio between the frequency counts in the positive review divided by the frequency count in the negative review for each word are as follows:

When the ratio exceeds 1, it indicates that the occurrence of terms is more frequent in positive reviews compared to negative reviews. Figure 5 shows the Word cloud with highest positive/Negative Ratio.

Figure 5 Word cloud with highest positive/Negative Ratio

The word cloud depicted above exhibits several prominent terms, like "highly," "addictive," "soothing," and "enjoy," which convey a positive connotation. Specific ingredients or goods such as "chilli," "magnesium," and "almonds" are also emphasised. These terms are considered neutral in nature, but they are

likely to be frequently used in good reviews due to Amazon's inventory management practises. Amazon tends to prioritise the distribution of products, such as almonds or chilli, that consistently meet or exceed users' expectations.

Let us now direct the attention towards the lexical items that exhibit a higher frequency of occurrence in unfavourable reviews as compared to good evaluations. Figure 6 Word cloud with highest positive/Negative Ratio.

Figure 6 Word cloud with lowest positive/Negative Ratio

There exist a variety of negative terms, including "strange", "failed", "fooled", and "choke". Nevertheless, comprehending the rationale for the increased frequency of particular terms such as "wellness," "feet," and "buyer" in negative reviews is a greater challenge.

IV. ALGORITHMS AND TECHNIQUES

To ascertain the sentiment of a review, a Deep Belief Recurrent Neural Network is employed. Most sentiment analysis techniques use a word-level analysis, wherein each word is evaluated separately. Positive words are assigned positive points, while negative words are assigned negative points. These scores are then aggregated to determine the overall sentiment. The approach being discussed here is known as the lexical approach. Nevertheless, a drawback of this approach is its disregard for word sequence, potentially resulting in the omission of crucial information. This methodology possesses the capability to comprehend nuances due to its avoidance of superficial analysis of textual content. The system generates abstract representations of the acquired knowledge.

Deep Belief Recurrent Neural Networks (DBRNN)

The findings demonstrate the importance of analysing the sentiment of Amazon food reviews using deep learning techniques, specifically DBRNN. Due to its efficacy and accuracy in capturing complex patterns and long-term dependencies, DBRNN is a potential alternative for usage in practical applications in online markets. This comparison study concludes with convincing evidence that DBRNN is superior to SVM and LSTM for categorising Amazon food reviews based on user ratings [18].

Embedding

We'll first utilise an embedding layer to convert words into vectors. This approach is more effective than a single hot encoding, particularly in the real-world setting where the vocabulary size exceeds 240,000 words.

We'll employ the word2vec method to create these embeddings, so words with similar meanings would be clustered together as seen in figure 7.

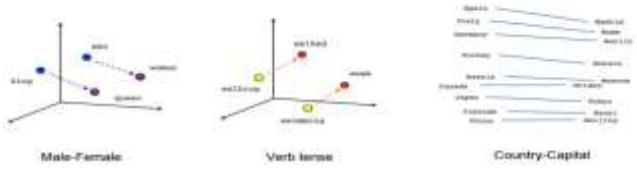

Figure 7 Word embeddings

The word2vec method is predicated on the idea that related words appear in related contexts. This strategy allows us to identify some semantic correlations, such as the gender differences between a woman and a queen or between a man and a king.

The embedded words are then sent to the DBN cells as the next step. You can see in the graph that the first, second, and third DBN cells are connected. These linkages allow us to convey information included in the word sequence, in contrast to other models that presume all inputs are independent of one another. The DBN cell's capacity to add or remove information from the cell state is one of its advantages.

Positive or negative results?

Finally, the sigmoid function is utilized to forecast whether the review is favourable or unfavourable. The function's advantage is that it can be used to represent a chance of success and is restricted to the range of 0 and 1. For instance, the likelihood is calculated that a review would be favourable [19].

Tuning hyper parameters

For the recurrent neural network following hyper parameters had to tuned :

Optimizer hyper parameters:

- Learning rate
- Minibatch size
- Number of epochs

Model hyper parameters:

- Number of hidden layers (DBN layers)
- Number of units in the DBN cells
- Dropout rate

V. METHODOLOGY

Data Pre-processing

To facilitate the training process of the model, it was necessary to convert the reviews into the appropriate format. The following steps were executed:

- Remove punctuations.
- Transform all the characters into lower case.
- Transform each review into a list of integers:
  - ➢ To begin, it is necessary to establish a dictionary that assign a unique numeric value to each term present in the lexicon of the reviews.
  - ➢ Next, convert each review into a sequence of integers by utilising the pre-established vocabulary.
- Create an array with 200 columns, and enough rows to fit one review per row.
- Next, proceed to store each review, which is composed of integers, into the array. In the event that the word count of the review falls below 200 words, it is possible to augment the list by appending additional zeros. In the event that the word count of the review surpasses 200 words, it is advisable to restrict the review to solely encompass the initial 200 words.

If the review is not constrained to a maximum length of 200 words, the training process for the Deep Bidirectional Recurrent Neural Network (DBRNN) would be excessively time-consuming. Figure 8 shows an analysis of the distribution of review length:

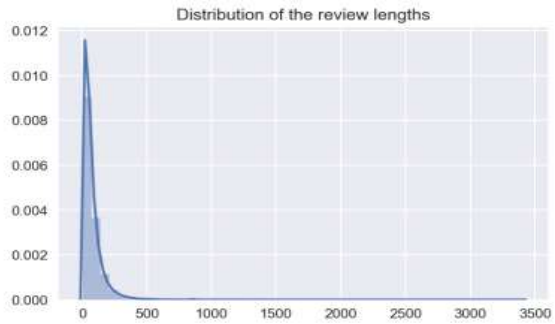

Figure 8 Distribution of the review lengths

The average and the 75th percentile of the length of the reviews are 79 and 97, respectively. Hence, by imposing a constraint on the review duration to a maximum of 200 words, the risk of significant information loss is minimised.

VI. IMPLEMENTATION

Split the dataset into a training, testing and a validation dataset.

In order to ensure the proper training and analysis of the DBRNN model, it is imperative to partition the data into separate datasets designated for training, validation, and testing objectives. The model would undergo training using a specifically selected collection of data for the purpose of refining its performance. The evaluation of the model's conclusions would be conducted on a different dataset specifically intended for

validation. In order to evaluate the applicability of the model, assessments is performed on the specifically selected testing dataset.

Build the neural network with Tensorflow

The last phase involves constructing the Deep Bidirectional Recurrent Neural Network (DBRNN) utilising the widely utilised deep learning framework known as Tensorflow. In order to provide input, labels, and keep_prob data to Tensorflow, it is necessary to initially establish placeholders.

The placeholder "inputs_" is defined as a tensor with an integer data type and a shape of [None, None]. It is

The placeholder "labels_" is defined as a tensor with an integer data type and a shape that can vary in both dimensions. It The placeholder "keep_prob" is defined as a TensorFlow float32 data type with the name 'keep_prob'.

After transmitting the input data, which comprises the reviews, to Tensorflow, the objective is to convert each integer that represents a word into a vector. The aforementioned technique is commonly referred to as embedding.The embedding matrix, named 'embedding_', is initialised with random values drawn from a normal distribution

The function `tf.nn.embedding_lookup` is used to retrieve the embeddings from the `embedding` tensor based on the indices specified in the `inputs_` tensor.

Next, it is necessary to generate the Deep Belief Network (DBN) cells and explicitly define the quantity of hidden units within each DBN cell.

The variable "dbn" is assigned the value of "tf.contrib.dbrnn".The BasicDBNCell class is instantiated with a parameter, dbn_size, which represents

In order to mitigate the issue of overfitting on the training set, it is advisable to incorporate a Dropout layer into the model architecture.

The TensorFlow library includes the module tf.contrib.dbrnn.The DropoutWrapper class is instantiated with the parameters dbn and output_keep_prob, where db

The utilisation of tf.contrib.dbrnn.The MultiDBRNNCell allows for the stacking of several Deep Belief Network (DBN) layers. Typically, in the field of deep learning, it is commonly observed that increasing the number of hidden layers in a neural network tends to improve the accuracy of predictions. In the above code snippet, a TensorFlow function is employed to create a multi-layer deep bidirectional recurrent neural network (DBRNN) cell, with the number of layers specified by the variable "dbn_layer" and the dropout rate set by the variable "drop". All states are initialised to zero.

The starting state is obtained by using the `zero_state` method on the `cell` object, passing in the `batch_size` and `tf.float32` as arguments.

Next, it is desired to propagate the input data across the Deep Belief Network (DBN) cells. The output of each Deep Belief Network (DBN) cell is propagated to the subsequent DBN cell. This technique facilitates the ability of the recurrent neural network to retain and recall information.

The function `tf.nn.dynamic_dbrnn` is used to compute the outputs and final state of a dynamic deep bidirectional recurrent neural network (DBRNN). It takes as input the cell, the embedded input sequence, and the initial state.

In order to obtain the forecast, a fully connected layer is employed, which incorporates a sigmoid function.

The predictions are obtained by applying a fully connected layer to the last element of the outputs tensor. The activation function used is the sigmoid function, and the weights are initialised using a truncated normal distribution with a standard deviation of 0.1.

In order to assess the precision of neural network, it is imperative to establish a well-defined cost function. In this particular situation, the mean squared error is employed.

The cost is calculated using the mean squared error loss function in TensorFlow, denoted as tf.losses.mean_squared_error(labels_, predictions).

Ultimately, the objective is to minimise the cost function with the aim of achieving a network output that closely approximates the desired values. The Adam optimizer is utilised for the implementation of this project.

Model Training

The subsequent phase entails the training of model. The stochastic gradient descent approach is employed. The approach involves the computation of an approximation of the cost function using tiny subsets of the dataset. The aforementioned methodology exhibits more scalability compared to the conventional gradient descent algorithm. If the model is through the learning process, it is expected that there would be a reduction

in the cost function and an improvement in the model's accuracy on the validation set for every minibatch.

Refinement

To refine the model, tensor board is utilized extensively.

Vocabulary reduction

Various techniques for vocabulary reduction have been employed in this study to mitigate data noise. These techniques include the elimination of stop words, as well as the removal of terms based on minimum frequency or polarity. Regrettably, none of the aforementioned methodologies yielded a notable acceleration in the convergence of the DBRNN, as evidenced by the graphical representation provided in figure 9. In this study, the comparative costs associated with language reduction is examined across varying numbers of iterations.

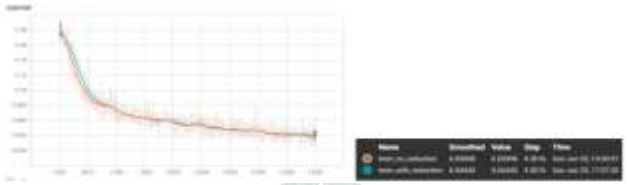

Figure 9 Vocabulary Reduction

The number of layers and the dimensions of the Deep Belief Network (DBN).

The graph depicted in figure 10 below demonstrates a comparative analysis of the cost per number of iterations across various combinations of DBRNN architecture, notably examining the variations in the number of layers and DBN size. It is evident that there is no reduction in cost when the dimension of the Deep Belief Network (DBN) is equivalent to 512. The attainment of convergence in Deep Bidirectional Recurrent Neural Networks (DBRNNs) is likely to be hindered by the substantial number of weights that require adjustment during the back-propagation of errors. Based on this reasoning, the option of incorporating a DBN layer with a dimensionality of 512 is excluded.

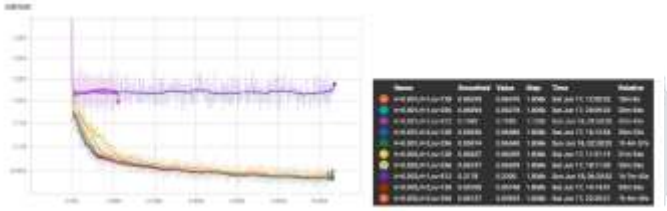

Figure 10 comparative analysis of the cost per number of iterations

Moreover, it is evident that when the learning rate is assigned a value of 0.002, the rate at which the cost diminishes per iteration exhibits a higher degree of unpredictability as shown in figure 11. This discovery implies that it can be prudent to decrease the learning rate.

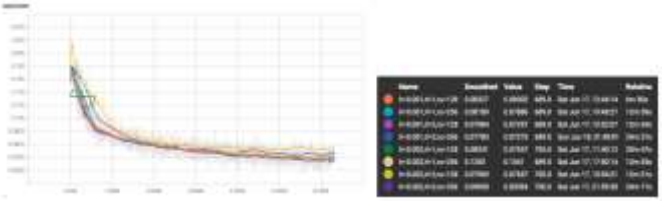

Figure 11 comparative analysis

When the learning rate is set to 0.001, the observed rate of decline demonstrates a more progressive and consistent trend as shown in figure 12.

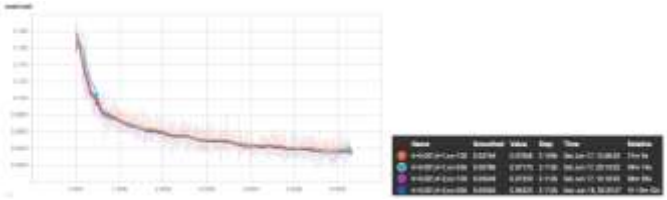

Figure 12 Comparative Analysis

Upon closer examination of iterations beyond 4000, it becomes evident that the cost is generally higher when utilising a single layer and a Deep Belief Network (DBN) with a size of 128 units. This architecture can be disregarded as shown in figure 13.

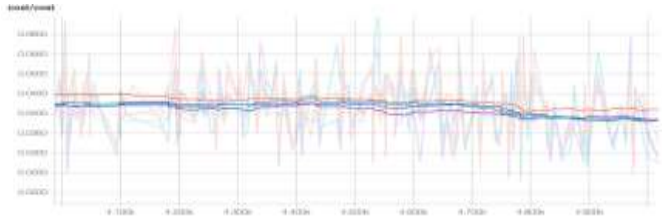

Figure 13 comparative analysis of various architecture

All of the remaining architectures possess identical costs. In contrast, a significant increase in time is shown in figure 14 when comparing the relative time required for a Deep Belief Network (DBN) with a size of 256 and 2 layers.

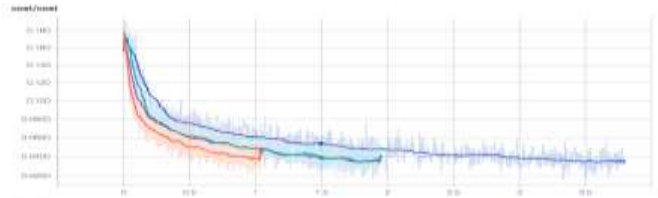

Figure 14 comparative analysis

In summary, it is observed that comparable results are obtained while employing a single layer with a Deep Belief Network (DBN) consisting of 256 units, as well as a Deep Bidirectional Recurrent Neural Network (DBRNN) with two layers and 128 units.

The dropout layer is a technique used in neural networks to In order to mitigate the issue of overfitting, one effective approach is to employ the random deactivation of a specific percentage of neurons throughout the training process.

The incorporation of a dropout layer is commonly employed to address the potential problem of overfitting that can occur during the training process of a model. In the present model framework, it is seen that the incorporation of the dropout layer produces minimal effects. One advantageous element is that the cost on the testing set remains consistent even when the model is trained with an excessive number of epochs. This suggests that the model exhibits the capacity for generalisation. The graph depicted in figure 15 presents a comparative analysis of the expenses associated with the training and testing datasets, with and without the incorporation of a dropout layer that possesses a probability of 0.5.

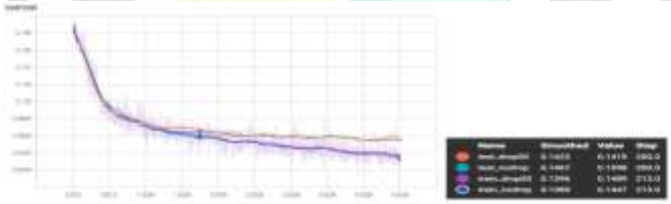

Figure 15 comparative analysis

VII. RESULTS

Model Evaluation and Validation

Based on the aforementioned observation, it has been concluded that the final model would encompass the complete vocabulary, feature a single Deep Belief Network (DBN) layer, have a DBN size of 256, and incorporate a dropout layer with a probability of 0.5.

In order to determine the applicability of the model, it is crucial to analyse its performance on the unmodified testing dataset. In order to evaluate the effectiveness of the binary classifier, it is feasible to construct a Receiver Operating Characteristic (ROC) curve as shown in figure 16. The construction of the receiver operating characteristic (ROC) curve involves plotting the true positive rate, which represents the proportion of positive occurrences correctly detected, against the false positive rate, which represents the proportion of negative instances wrongly identified. The ideal scenario arises when the curve is aligned with the left corner. This would suggest achieving a sensitivity rate of 100%, denoting the lack of false negatives, together with a specificity rate of 100%, denoting the absence of false positives.

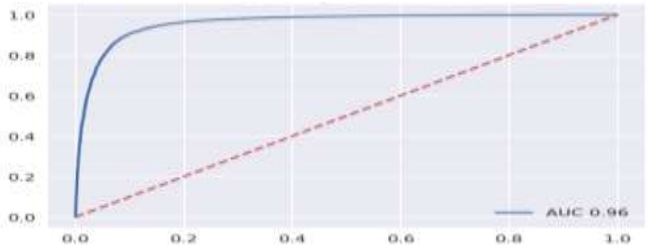

Figure 16 Receiver Operating Characteristic Curve

The value of the area under the curve (AUC) is clearly 0.96. Given the premise that the highest attainable AUC score is 1, it can be deduced that the model exhibits a noteworthy capacity to make accurate predictions for reviews that have not been encountered before. Furthermore, the model demonstrates a significant level of equilibrium in the proportions of false positives and false negatives.

Another noteworthy metric to consider is the F1 score, which encompasses both precision and recall. The analysis of the original dataset revealed that a majority of the evaluations, specifically 78%, demonstrated a positive sentiment. If an unsophisticated algorithm were utilised to forecast that all reviews include positive sentiment, the resultant accuracy would amount to 78%. It is advisable to employ the F1 score as a means of assessing the performance of the Deep Bidirectional Recurrent Neural Network (DBRNN). Figure 17 shows the results of various performance evaluation parameters on proposed model.

	precision	recall	f1-score	support
0.0	0.90	0.73	0.81	12363
1.0	0.93	0.98	0.95	44137
avg / total	0.92	0.92	0.92	56500

Figure 17 Performance Evaluation

The achieved F1 score is 0.92. Both the precision and recall values exhibit a high level of accuracy, with a score of 0.92. This suggests that the model exhibits a favourable balance between sensitivity and specificity. While there is room for further improvement, the DBRNN model has developed the capability to distinguish between negative fine cuisine ratings and positive reviews.

Comparison Analysis

Figure 18 shows the comparison of proposed model with earlier developed state of the art methods based on accuracy metrics. Table 1 shows the Comparison evaluation of DBRNN with SVM and LSTM.

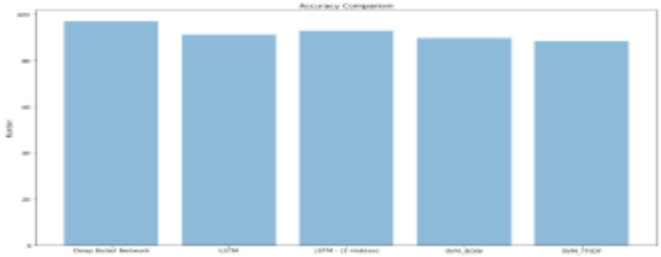

Figure 18 Comparison based on Accuracy

Table 1 Comparative analysis

Sl. No.	Algorithms	Accuracy
1	DBRNN	96%
2	LSTM(one layer)	92.60%
3	LSTM(two layers)	93.40%
4	SVM(BOW)	90.35%
5	SVM(TFIDF)	89.90%

The examination of user rating classification for Amazon food reviews through a comparative analysis of SVM, LSTM, and DBRNN has provided insights into the respective merits and limitations of each methodology. The objective of the study was to ascertain the optimal model for sentiment analysis inside this particular scenario. The findings unequivocally indicate the higher performance of Deep Belief Recurrent Neural Networks (DBRNN) in comparison to Support Vector Machines (SVM) and Long Short-Term Memory (LSTM).

Although Support Vector Machines (SVM) demonstrated impressive accuracy and efficiency when dealing with high-dimensional data, it encountered difficulties in capturing the intricate sequential connections and contextual nuances present in natural language text. The Long Short-Term Memory (LSTM) model, due to its recurrent structure, demonstrated superior capability in capturing these dependencies. However, it necessitated a substantial amount of computer resources and training time.

On the other hand, Dynamic Bayesian Networks (DBN) shown notable efficacy in acquiring hierarchical representations from the dataset, successfully capturing detailed patterns and correlations among words and sentences. The aforementioned capabilities enabled DBN to outperform both SVM and LSTM in terms of accuracy, hence attaining superior outcomes in the classification of user ratings for Amazon food reviews.

VIII. CONCLUSION

The significance of utilising deep learning techniques, particularly DBRNN, for sentiment analysis in the domain of Amazon food reviews is emphasised by the research findings. The intriguing potential of utilising DBRNN in online marketplaces lies in its ability to effectively capture intricate patterns and long-term dependencies, hence enhancing efficiency and accuracy. In summary, this comparative analysis presents persuasive data supporting the superior efficiency of DBRNN in comparison to SVM and LSTM for the categorization of user ratings in Amazon food reviews.

REFERENCES

[1]. Bing Liu. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

[2]. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processingVolume 10, pages 79–86. Association for Computational Linguistics, 2002.

[3]. Peter D Turney and Michael L Littman. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), 21(4):315–346, 2003.

[4]. Jingjing Liu, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. Low-quality product review detection in opinion summarization. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007.

[5]. Qiang Ye, Ziqiong Zhang, and Rob Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert systems with applications, 36(3):6527–6535, 2009.

[6]. Wu, Jiayu and Tianshu Ji. “Deep Learning for Amazon Food Review Sentiment Analysis.” (2016).

[7]. Aravindan, T.R. & Vigneshwar, C. & Gopalswamy, Suganeshwari. (2023). Sentiment Classification for Amazon Fine Foods Reviews Using Pyspark. 10.3233/ATDE221293.

[8]. Paknejad, Sepideh. “Sentiment classification on Amazon reviews using machine learning approaches.” (2018).

[9]. Fang, X. and Zhan, J. (2015). Sentiment analysis using product review data, Journal of Big Data 2(1): 5. URL: https://doi.org/10.1186/s40537-015-0015-2.

[10]. Go, A., Bhayani, R. and Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision, p. 6.

[11]. Abah, J. O. (2021). Sentiment Analysis of Amazon Electronic Product Reviews using Deep Learning, PhD thesis, Dublin Business School. URL: https://esource.dbs.ie/handle/10788/4291.

[12]. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. and Gao, J. (2021). Deep Learning–based Text Classification: A Comprehensive Review, ACM Computing Surveys 54(3): 62:1–62:40. URL: https://doi.org/10.1145/3439726.

[13]. Shehu, S. A., Mohammed, A. D. and Abdullahi, I. M. (2021). An Optimized Customers Sentiment Analysis Model Using Pastoralist Optimization Algorithm (POA) and Deep Learning, 2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA), pp. 132–139.

[14]. Zola, P., Cortez, P., Ragno, C. and Brentari, E. (2019). Social Media Cross-Source and Cross-Domain Sentiment Classification, International Journal of Information Technology & Decision Making 18(05): 1469–1499. Publisher: World Scientific Publishing Co. URL: https://www.worldscientific.com/doi/abs/10.1142/S0219622019500305.

[15]. Coyne, Emilie & Smit, Jim & Güner, Levent. (2019). Sentiment analysis for Amazon.com reviews. 10.13140/RG.2.2.13939.37920.

[16]. R. Chen and Hendry, "User Rating Classification via Deep Belief Network Learning and Sentiment Analysis," in IEEE Transactions on Computational Social Systems, vol. 6, no. 3, pp. 535-546, June 2019, doi: 10.1109/TCSS.2019.2915543.

[17]. Z. Xiao, L. Wang and J. Y. Du, "Improving the Performance of Sentiment Classification on Imbalanced Datasets With Transfer Learning," in IEEE Access, vol. 7, pp. 28281-28290, 2019, doi: 10.1109/ACCESS.2019.2892094.

[18]. Ravi Chandra, & Basavaraj Vaddatti. (2021). Sentiment Analysis using Deep Belief Network for User Rating Classification. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 10(8), 87–91. https://doi.org/10.35940/ijitee.H9233.061082

[19]. G. Sathish S. V. Saravanan's. Narmadha S. U. Maheswari 2012. "Personal Authentication System using Hand Vein Biometric." International Journal of ComputerTechnology and Applications vol. 3(1) pp. 383-391 Jan-Feb 2012.