Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd
Anatolian Science

ISSN,e-ISSN: 2548-1304
Volume:9, Issue:1, pp:51-61, 2024
https://doi.org/10.53070/bbd.1409094

Research Paper

A Model for Customer Opinion Mining and Sentiment Classification Using a Mixture of Experts Machine Learning Model

V.I.E. Anireh [1], E.N. Osegi[2*], A.A. Silas[1]

[1]Department of Computer Science, Rivers State University, Nkpolu, Rivers State, Nigeria

[2]Department of Computer Science, National Open University of Nigeria (NOUN)

(anireh.ike@ust.edu.ng. emmaosegi@gmail.com, awajimijaans@gmail.com)

Received:Feb.06, 2024 Accepted:Mar.20, 2024 Published: Jun.01, 2024

Abstract— This paper presents a customer opinion mining technique based on a Mixture of Experts (MoE) machine learning model. The approach allows a corpus from open source data repositories to be classified into positive, negative, and neutral sentiments as the case may be, and in a predictive manner. The results of simulations showed that the proposed MoE approach can effectively be used as a core tool in opinion mining and also serve in decision-making by appropriate categorizations. In particular, it was found that the use of higher epoch sizes and larger data greatly enhances the performance of the MoE by reducing perplexity and error cost margins to appreciable levels. Thus, the MoE presents a promising candidate for customer opinion mining particularly in business product development environments.

Keywords: *Classification, machine learning, opinion, sentiments.*

1. Introduction

Analyzing the enormous amount of data generated by potential customers for a given business can serve as a useful tool in achieving customer confidence and hence satisfaction. This is because their opinions can be effectively heard and the organization's goal(s) can be aligned to match their needs and expectations. In every business, there is the desire to know a customer's thoughts/feelings or public opinions about the quality of products and services offered. Also, a large number of users on the social web express their opinions about many things through media such as blogs, social networks, chats, emails, etc. The unique ability and capability of Machine Learning (ML) systems to identify different opinions about human experiences/activities, product service offerings from a variety of manufacturing firms, etc have become a key fundamental in the operation of modern business. Indeed, it is important to comprehend how user-generated content as obtained from the dynamic interactions that occur in social networks and other similar forums operates. Sentiment analysis refers to the automatic process of mining or extracting unique facts such as opinions and emotions from regular media such as text data, speech data, and social databases through the specific use of Natural Language Processing (NLP) techniques (Shelke et al., 2012). Generally, the primary objective of a sentiment analysis scheme is to extract useful sentiments from a unit (or corpus) of text. It involves the rating of comments through social forums or posts done on the internet. We need to develop automated methods of classifying these sentiments as it is very useful and crucial to social media, the product marketing industry, entertainment, administration, hospitals, etc. to monitor automatically and characterize the total feeling or mood of anyone towards a brand, company, school administration, medical services, politics, etc. and identify whether they are viewed as positive or negative (Bannister & Meriac, 2015; Williams et al., 2015). The study on sentiment analysis using an ensemble of predictive experts reveals the need to use more sophisticated techniques in other to understand and capture the broad range of emotions expressed by humans. Organizations stand to gain more because they become fully aware of the applications of sentiment analysis tools/techniques within their marketplace, enhancing the growth of industry-sector-specific goods and services. A good example of an application area is the use of intelligence tools that assist in decision-making for financial traders and analysts.

There is a need to measure customers' satisfaction levels to ascertain the effect of services rendered or the performance of the products to enhance high productivity. Sentiment analysis through the application of ensemble techniques produces very efficient and accurate results. Also, the need to extract insights from social data is relevant for smooth online interaction and online sales of goods. The application of systems properly developed with predictive experts that enable opinion mining allows for a quick understanding of the attitudes of consumers and how one reacts to them. Hence the application of sentiment analysis is broad and very powerful and much needed in the business industry.

This research study is aimed at developing a predictive system for sentiment classification using an ensemble of experts – Mixture of Experts (MoE) ML (MoE-ML) model. Furthermore, the research seeks to apply and validate the MoE-ML model to a sample open-source dataset. This research study will also improve the ability to build an intelligent expert system that can learn to predict the sentiment or determine the likelihood that a sentiment word or phrase is positive, neutral, or negative. Also, the ability to interpret the particular tone of a piece of writing in addition to inferring the possible opinions about certain things which are very important to information mining systems will thereby be enhanced.

2. Related Works

In recent times, decision-making has become a core part of man's day-to-day activities. A large number of persons could conveniently ask a group of related persons (friends) to advise on the most appropriate grocery shop to visit, explain who they were planning to vote for in the presidential elections, determine which movie is having more buyers, or decide which academic or health institution is best, etc. These and many more have rekindled interest in the development of new systems that deal with opinion mining because of the potential influence that opinions wield on products and services. This area of study that focuses on the opinions and experiences of persons is called sentiment analysis. Liu (Liu, 2012; Liu, 2020; Liu 2022) stated that analyzing sentiments represents an emerging and very popular field of research in the area of NLP. It involves the automatic analysis and evaluation of human-encoded text data, its corresponding tracking, and predictive judgment within a domain and is very suitable for opinion mining.

The origins of the first experiments on sentiment classifications can be traced to the works of Pang (Pang et al., 2002) where simplistic bag-of-words techniques based on Naïve-Bayes (NB), Maximum Entropy (MaxEnt) models or Support Vector Machines (SVM) was shown to be insufficient for the prediction of the sentiment of worded-documents while they demonstrated efficient performance for general purpose topic oriented document classifiers.

The studies on sentiment classification studies include a determination of the objective or subjective nature of the text containing the sentiments. In particular, it seeks to verify whether a given objective or subjective text document contains positive, negative, or both positive/negative sentiment text data or possess the likelihood for such. It includes several characteristic parts such as the different tasks, feature sets, approaches, and application domains that impact the level of sentiment employed. It is also regarded as a type of text categorization. Text categorization is the original text problem. It uses either a supervised or an unsupervised technique to extract the different categories to which a text document belongs. An example of a web page news is a ₦200 financial statement, restaurants or services, etc.

Xia et al. (2011) investigated several linguistic features of text-data documents including parts-of-speech, word-relations, and Term Frequency-Inverse Document Frequency (TF-IDF) together with some ensemble learning techniques for sentiment classification task processing. The scheme included various Machine Learning (ML) approaches such as Naïve Bayes, maximum entropy, and support vector machines where a combination of fixed rules, meta-classifiers, and weighted rules were employed.

Socher et al. (2011) developed a novel ML framework using the Recursive Auto Encoders (RAE) model. Autoencoders are efficient in the learning of feature encodings which are in turn very useful for sentiment classification as they incorporate the recursive interaction between context and polarity words in a set of sentences. Indeed these sentences are formed into a unified framework while at the same time, the RAE learns the basic features needed to make accurate predictions. Polarity is a binary value either positive or negative.

Wang et al. (2014) did a similar work by carrying out empirical analysis of the performance of several different linguistic representations of text document together with bigram, unigram, term frequency, and the term frequency-inverse document frequency (TF-IDF) in union with tripartite ensemble learning techniques such as bagging, random subspace and boosting and including five classification algorithms (Naïve Bayes, Maximum entropy, decision tree, K-nearest neighbor and support vector machines). They were able to show that efficient classification models are necessary for sentiment analysis.

Ezenkwu et al (2015) targeted customer segmentation for a group of retail customers based on the k-means clustering algorithm. They were able to identify 4 customer segments or clusters including High Buyers-Regular-Visitors (HBRV), High-Buyers-Irregular-Visitors (HBIV), Low-Buyers-Regular-Visitors (LBRV), and Low-Buyers-Irregular-Visitors (LBIV).

Attention networks have also been proposed for sentiment classifications by Wang et al (2016) using the Long Short Term Memory (LSTM) neural networks and by Ma et al (2017) based on the set of interaction LSTMs.
Recent studies in this can of study can also be found in the research by Wang (Wang et al., 2024) using a Hierarchical Fusion Multimodal Sentiment Analysis Multitask-Learning Method (HFMSAM) for tripartite sentiment classifications and (Wang, 2024) using an Empirical Model (EM) in determining sentiments in stock market returns.

3. Materials and Methods

3.1. Data Set

A wide spectrum of human emotion and sentiment data has typically been captured using the Experienced Project (EP) dataset as provided in (Sudhof et al., 2014). The EP dataset contains very personal confessions made by persons. The confessions are encoded with a label set of five reactions by several users. The reaction labels employed include: you rock (approved), teehee (amusement), I understand, sorry, hugs, and wow, just wow (exhibiting shock). The labels with the highest vote as well as considering a full distribution over sentiment classifications are predicted.

Also, datasets such as movie reviews containing distributions over a wide range of human emotions were introduced. The datasets consisted of user stories that explain with mult-labels which when aggregated into a multi-distribution for capturing emotionally reactive sentiments. It was observed that the technique learns from a representation of vector spaces for *n*-word phrases and could successfully perform sentiment prediction tasks hence it outperforms competitive baselines.

The research study materials will employ the data obtained from the EP for validation purposes of the proposed model. A description of the key attributes is presented in Table 1.

Table 1. Descriptive Statistics of EP Dataset

S/No	Attribute Label	Type	Description
1	You Rock	Varchar	Indicating approval, congratulations
2	I Understand	Varchar	Show of empathy
3	Teehee	Varchar	The user found the storyline amusing.
4	Sorry, Hugs	Varchar	Sorry, Hugs
5	Wow, Just Wow	Varchar	Expressing shock and surprise
6	ID	Number	Train and test ID

3.2. Proposed Systems Approach

The proposed system (see Figure 1) performs sentence/document level sentiment classification based on an online web repository called Experience Project (EP) and the University of California dataset both of which have a broad spectrum of human personal confession that touches different problem areas. This collection is downloaded and saved in a directory for easy retrieval during analysis and processing.

As shown in Figure 1, the data is collected from a web repository which is the sentiment data. The key features in this data are then obtained from a feature generator that parses feature data both to an emulator interface and an optimized training set module. Thereafter the sentiments collected are classified into positive, negative, and neutral using the ensemble classifiers comprising of the Naive Bayes (NB), the Extreme Learning Machine (ELM), and the Long Short-Term Memory (LSTM). The output is displayed on a graphic user interface which the product analysts use to arrive at conclusions based on the data at hand.

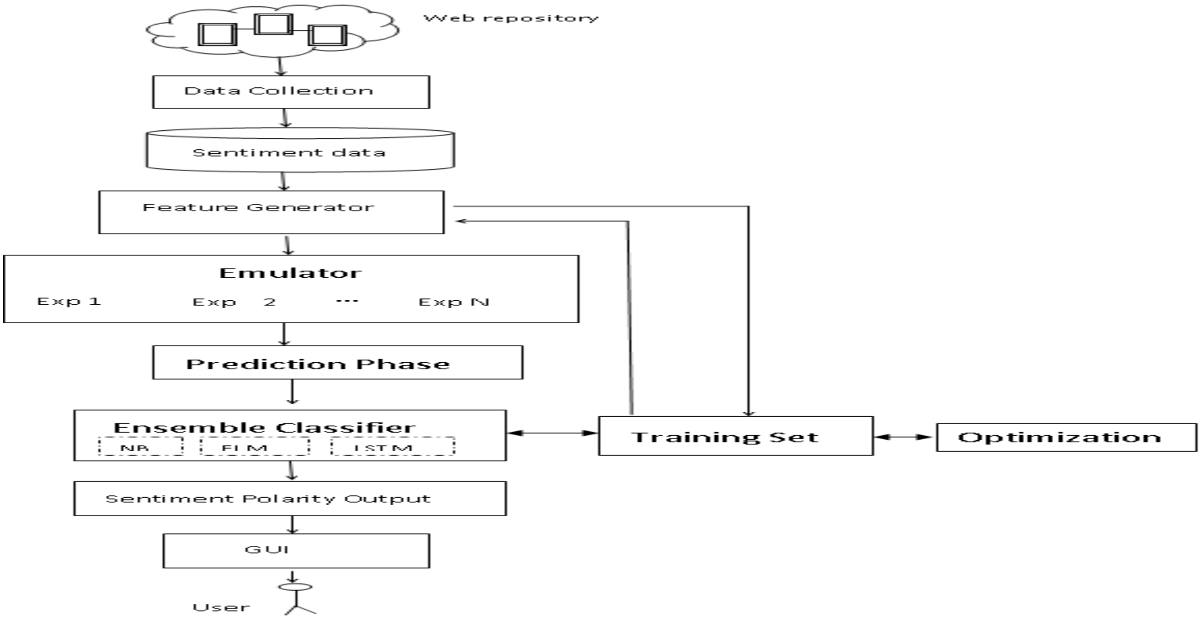

Figure 1: Cyclical process model of CRISP-DM.

3.3. Mixture of Experts Model

The proposed system (see Figure 1) performs sentence/document level sentiment classification based on an online ensemble Machine Learning (ML) model called Mixture of Experts (MoE). The MoE comprises a set of n combined “expert networks" $E_1, E_2, \cdots, E_k$, and a “gating network" G the output of which is a sparse *n-dimension* vector (see Figure 2). These experts fundamentally represent the neural networks defined earlier (sub-section 3.2) where each has its own unique set of operational parameters. In principle, these experts require that similar-sized inputs be fed into the combined system to produce the same or equivalent set of sized outputs. In this particular context, the models are restricted to feed-forward networks with identical architectures, but with separate or unique parameters.

If we denote the terms *G(x)* and *Ei(x)* as the output of the sparse-gating network and *i-th* expert-processor network respectively; then for a given input *x*, the corresponding output state say, y, of the MoE module can then be realized from the expression:

$$\sum_{i=1}^{n} G(x) E_i(x) \tag{1}$$

The expression in (1) implies we can save computational resources using *G(x)*. Thus, when *G(x)* = 0, *Ei(x)* = 0. For a large number of experts, this becomes particularly a useful model as we can select a sparse-weighted set of “experts", comprising a secondary MoE with each having its unique gating network.

The gating network employs a softmax squashing function to achieve the desired range as provided by:

$$G_\sigma(x) = soft\max(x, W_g) \tag{2}$$

The gating network can be trained by simple backpropagation along with any other models.

where,

W_g = weighting factor

X = sentiment data block

y_o = the output of the expert system

$\sum$ = the summation operator

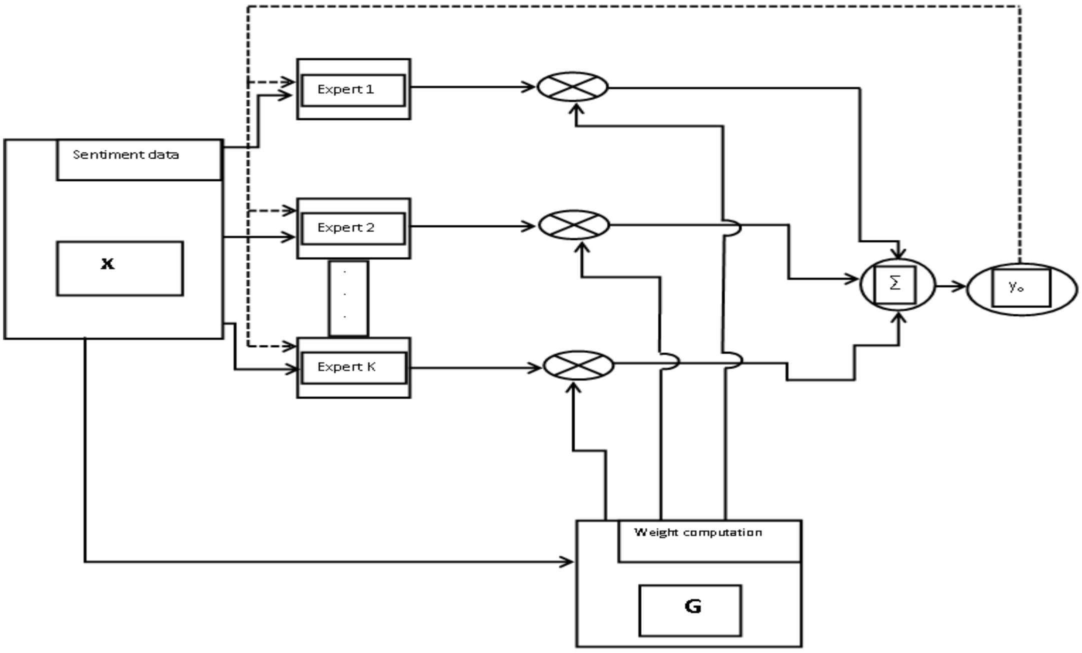

Figure 2: Mixture of Experts (MoE) Model.

A detailed nomenclature of the MoE model as depicted in Figure 2 is shown in Table 2.

Table 2. MoE Nomenclature Details

Id	Symbol	Description	Functions
1	$\otimes$	Product Operator block	It multiplies the coefficients.
2	Expert	Expert block	Allows each predictive algorithm to be used
3	y_o	Output of Expert system	Allows the result of predictions of the expert system to be obtained.
4	**G**	Weighting factor	Each prediction of the experts is weighted numerically.
5	**X**	Sentiment data block	Allows the entry of sentiment data either manually or automatically when the algorithm requests for data.
6	Σ	Summation operator block	Sums up the mixture of experts' prediction results.
7	$\dashrightarrow$	Feedback loop	Shows the result is feedback to the system as new input making the system robust.

4. Results and Discussions

The results were obtained in two parts: first, a user interface model with a web control server developed in the PHP programming language generates the sentiment data including initial computations based on the MoE model as developed earlier in Section 3. Then, further graphical simulations were performed using the MATLAB simulation language.

4.1. Graphic Interface and Initial Results Capture

The user interface employed in sentiment opinion mining of customer data is shown in Figure 3. To view the interface, the user has to ensure that a local server is installed and activated; also a convenient web browser such as Mozilla is employed by typing the URL site: http://localhost/seanalyzerv1/ in the browser address bar. This accesses the index.php file or home page and displays the content on the web browser (Figure 3)

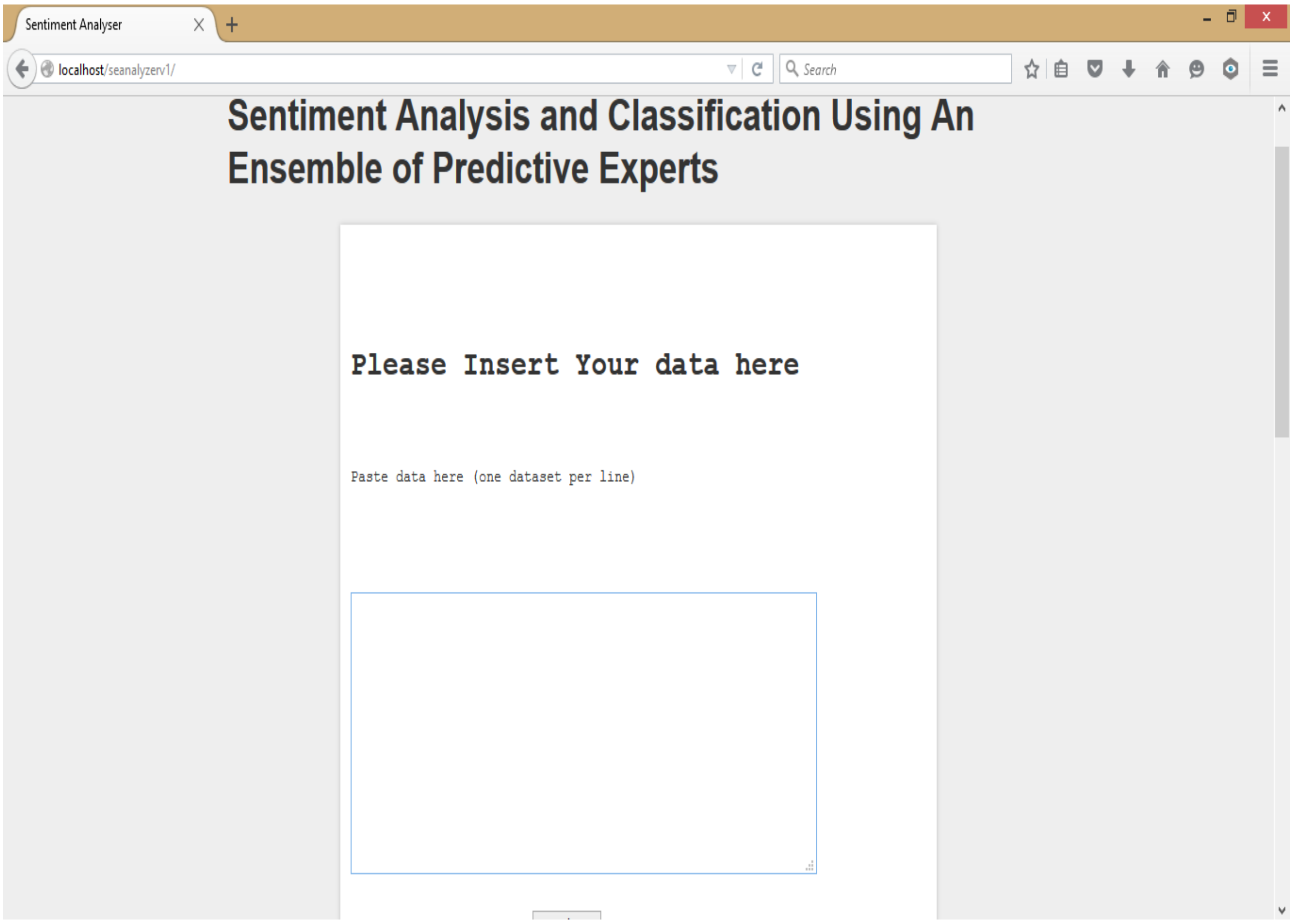

Figure 3: The Sentiment Analysis Classification User Interface.

Furthermore, Figures 4 and 5 show the entry of sample data and the analysis of the results. In Figure 4, the learning rate reveals that 25 distinct characters are obtained. These features including white space, removal of redundancy, etc. are necessary for the smooth running of the system. The system may be paused, resumed, or restarted as indicated on the menu

In Figure 5 the model samples that are being analyzed at different times are shown. In evaluating the performance of the classifiers, the epoch, perplexity, and cost metrics are noted (Karpathy, 2015).

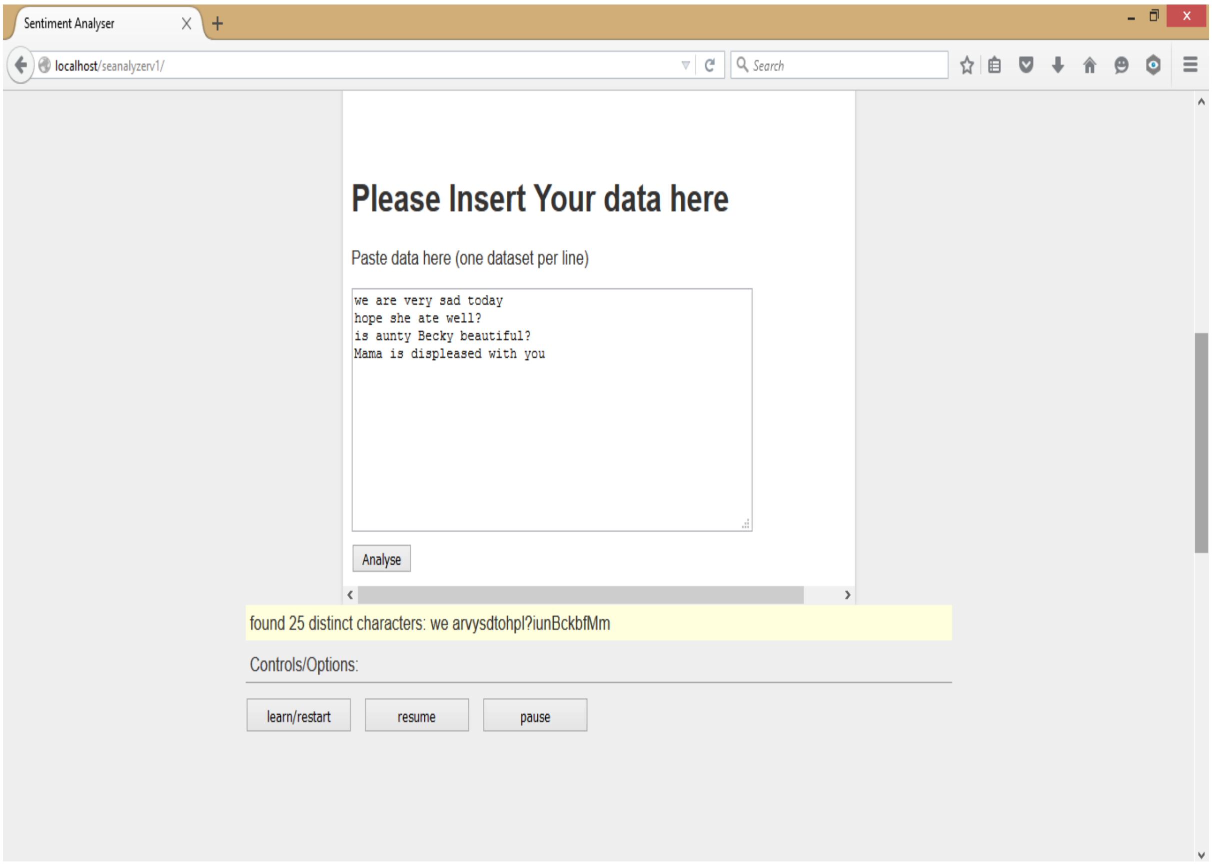

Figure 4: Sample Input Dataset Showing the Learn Rate.

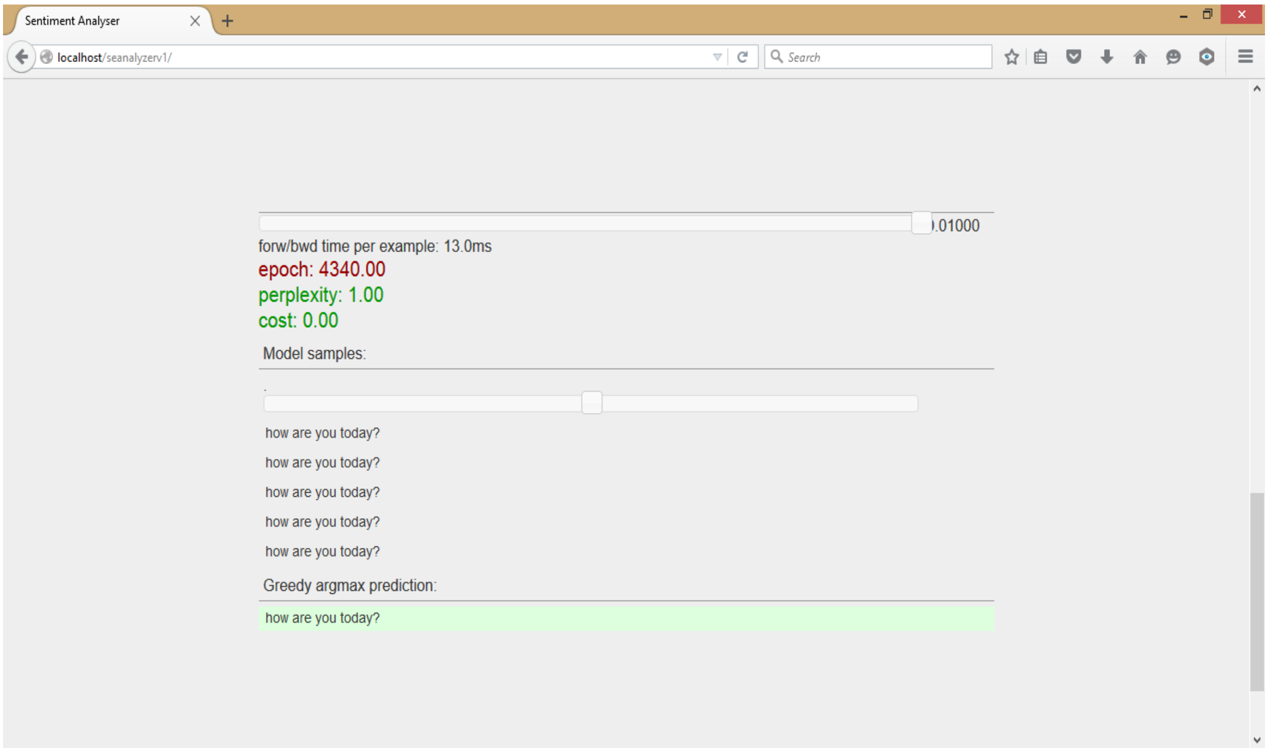

Figure 5: The User Interface Prediction Result Analysis.

In Figure 6, the sentiment classification report is shown. Here, the positive, negative, and neutral sentiments for a given document are successfully obtained based on the given sentiment scale: >2.5 for positive sentiment, < 2.5 for negative sentiment, and = 2.5 for neutral sentiment.

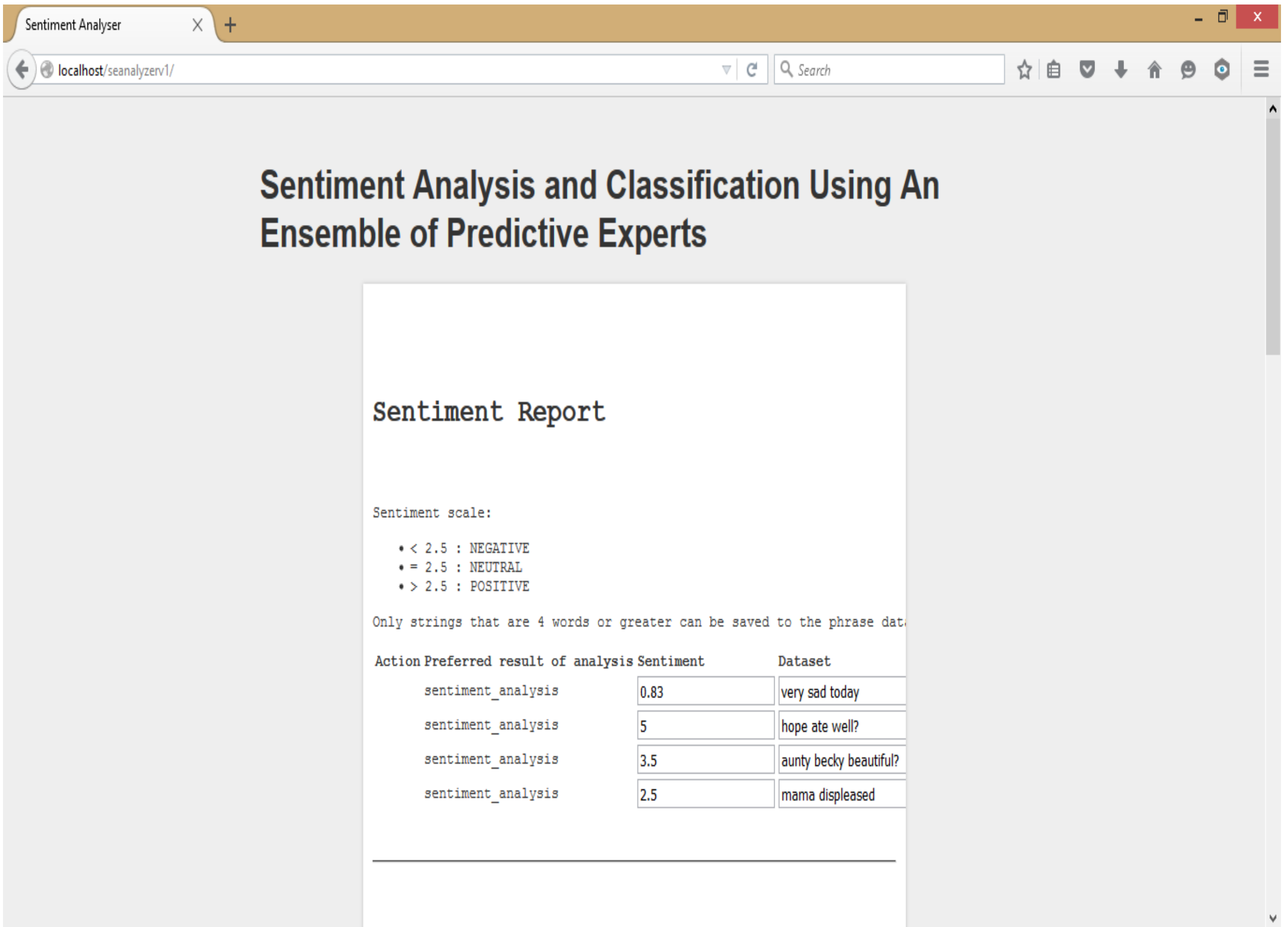

Figure 6: The Sentiment Classification Report.

4.2. The Learning Rates Effect on the MoE Model

The system learning rate must be obtained by various adjustments of the neuron in the program module to observe the changes after a given number of iterations. Where the learning rate is higher, the system is observed to contain too much kinetic energy and the parameter vector bounces around chaotically which makes it unable to settle down into deeper but narrower parts of the loss function. This will enable proper training of the model and evaluation of the time and convergence. Learning rate is a decreasing function of time. The sample experience dataset input into the system is (Sudhof et al., 2014):

I regularly shoplift.
I got caught once and went to jail, but I've found that this was not a deterrent.
I don't buy groceries,
I don't buy school supplies for my kids,
I don't buy gifts for my kids, we don't pay for trios los, arid
I don't buy most incidentals for the house (cleaning supplies, toothpaste, etc.)...

The generated results sample snapshot at a hidden neuron setting of 20x20 is shown in Figure 7 while the results for different epochs are given in Figure 8 and as well (numerically) in Table 3.

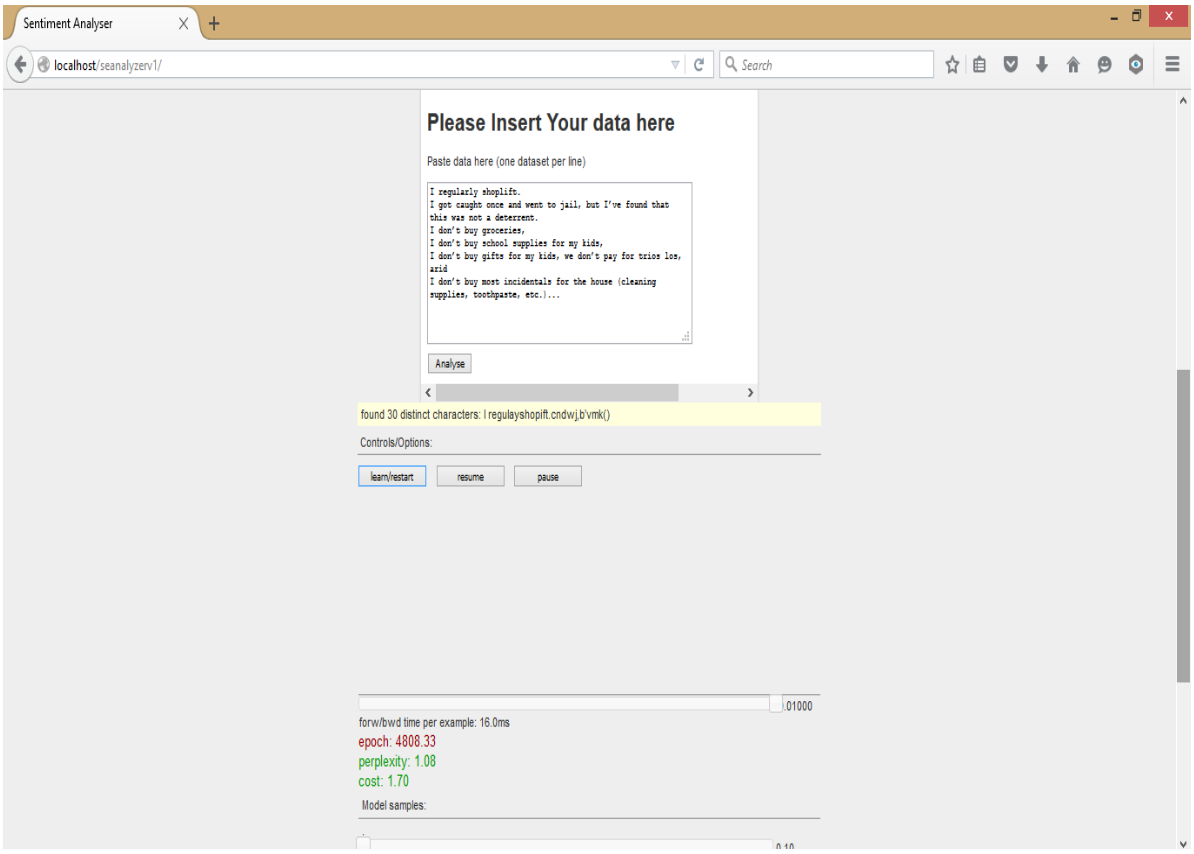

Figure 7: Screenshot of Sample Experience Dataset Used For Learn Rate Process (Neuron 20:20).

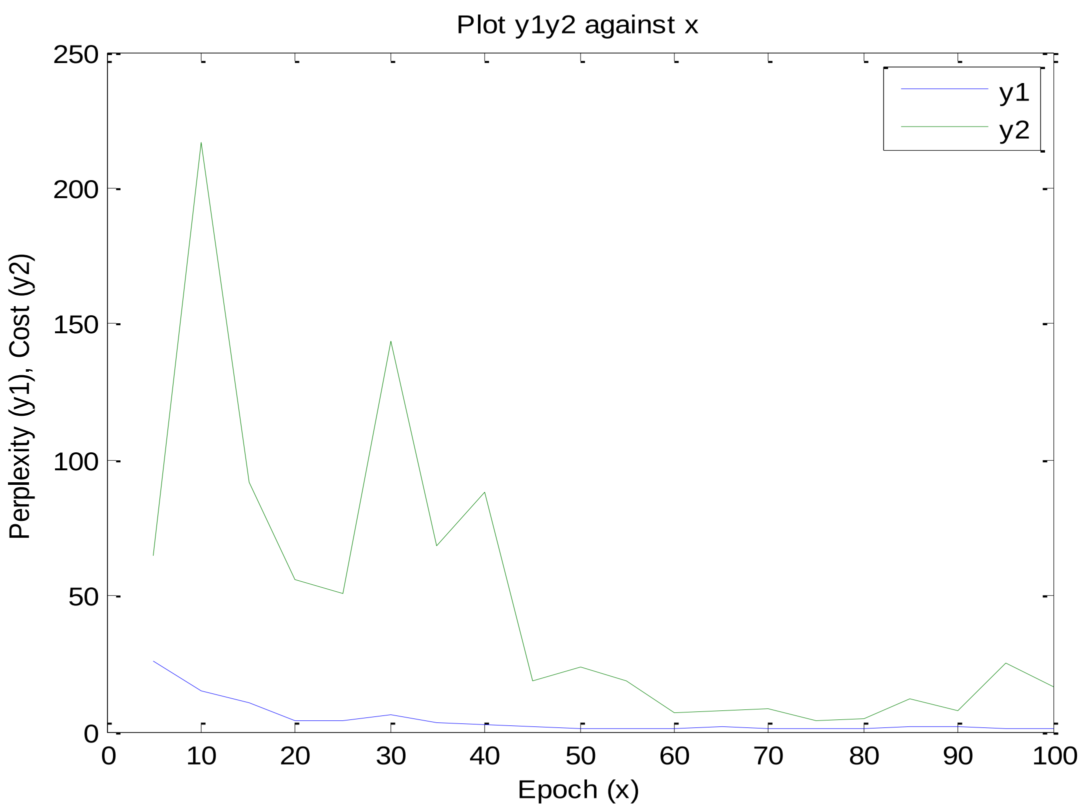

Figure 8: Simulation of Learn Rate of Experience Dataset (neuron 20:20).

Table 3. Model Learn Rates of Hidden Size for Neuron 20:20

Epoch (x)	Perplexity (y1)	Cost (y2)
5.00	25.70	64.44
10.00	15.06	216.97
15.00	10.58	91.99
20.00	4.21	56.08
25.00	3.66	50.62
30.00	6.04	143.93
35.00	3.00	68.20
40.00	2.92	87.89
45.00	1.62	18.72
50.00	1.34	23.73
55.00	1.25	18.39
60.00	1.40	6.71
65.00	1.48	7.88
70.00	1.25	8.75
75.00	1.19	3.67
80.00	1.12	4.42
85.00	1.81	11.87
90.00	1.47	8.04
95.00	1.37	25.23
100.00	1.30	16.22

As can be seen in Figure 8 and clearly from Table 3, increasing the number of epochs greatly enhances (reduces) the perplexity and generally reduces the error cost margin. Thus, the higher epochs are better. However, it might impact computational run times if much higher epochs are used. This remains a matter for further research.

4.3. Test Results on Real Dataset

The results using proposed sentiment analyser approach applied to a real world dataset – the Amazon product review dataset provided in (Kotzias et al., 2015) is as shown in Table 4 for first 3 epochs (5, 10 and 15).

Table.4. Model Learn Rates of Hidden Size for Neuron 20:20

Epoch (x)	Perplexity (y1)	Cost (y2)
5.00	5.53	58.14
10.00	9.26	191.43
15.00	5.74	48.93

As can be seen in Table 4, the perlexity and cost estimates are lower when compared to that of Table 3; this may be attributed to the larger dataset (approximately 1000 samples) thereby enabling the expert network to learn more efficiently. Thus, this result validates the bigger data better learning paradigm as is the research findings from many deep learning applications.

5. Conclusion

Customer opinion mining is of great importance since it allows a business to make better use of marketing budgets as well as to gain a competitive edge over rival companies. More importantly, it demonstrates better knowledge of various customer purchasing behaviors and patterns over time.
In this regard, the sentiments of existing customers will play a major role as a marketing strategy to attract potential customers. Sentiment analysis involves the extraction of sentiment from a unit of text. We have developed an automated method of classifying these sentiments based on a Mixture of Experts (MoE) as it is very useful and crucial to social media, the product marketing industry, entertainment, administration, hospitals, etc., to monitor automatically and characterize the total feeling or mood of anyone towards a brand, company,

school administration, medical services, politics, etc. and identify whether they are viewed as positive, negative or neutral.

Some of the challenges faced by e-commerce, stores, and supermarkets involve dealing with huge volumes of customers with different and similar wants, different and similar purchase prices, and buying patterns that can be circumvented using the proposed approach.

Furthermore, this research study will help:

- Improve promotional service offerings to potential customers and more useful personalized solutions for existing customers.

- Recognition and understanding of user behavior using machine learning.

- Demonstrate capacity in building analytical models using a machine learning algorithm.

Future studies should be geared towards improving the service categorization offerings and including the use of more progressive continual machine learning strategies that deal with short spans of timely sentiment data.

Acknowledgments

The SurePay Foundations Group is hereby acknowledged.

References

Banister, C. M., & Meriac, J. P. (2015). Political skill and work attitudes: A comparison of multiple social effectiveness constructs. *The Journal of Psychology*, *149*(8), 775-795.

Ezenkwu, C. P., Ozuomba, S., & Kalu, C. (2015). Application of K-Means algorithm for efficient customer : a strategy for targeted customer services. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, *4*(10), 40-44.

Kotzias, D., Denil, M., De Freitas, N., & Smyth, P. (2015, August). From group to individual labels using deep features. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 597-606).

Liu, B. (2012). Sentiment analysis: A fascinating problem. In *Sentiment Analysis and Opinion Mining* (pp. 1-8). Cham: Springer International Publishing.

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Shelke, N. M., Deshpande, S., & Thakre, V. (2012). Survey of techniques for opinion mining. *International Journal of Computer Applications*, *57*(13).

Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., & Manning, C. D. (2011, July). Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 151-161).

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision support systems*, *57*, 77-93.

Wang, L., Peng, J., Zheng, C., & Zhao, T. (2024). A cross-modal hierarchical fusion multimodal sentiment analysis method based on multi-task learning. *Information Processing & Management*, *61*(3), 103675.

Wang, W. (2024). Investor sentiment and stock market returns a story of night and day. *The European Journal of Finance*, 1-33.

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 606-615).

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., & Spasić, I. (2015). The role of idioms in sentiment analysis. *Expert Systems with Applications*, *42*(21), 7375-7385.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information sciences*, *181*(6), 1138-1152.