E-ISSN 2581 – 7957
P-ISSN 2277 – 3916
CVR Journal of Science and Technology, Volume 23, December 2022
DOI: 10.32377/cvrjst2316

Automatic Aspect-Based Sentiment Analysis for Motor Vehicle Sales Forecasting

C. Raghavendra
Assoc. Professor, CVR College of Engineering/ CSIT Department, Hyderabad, India
Email: drraghavendra@cvr.ac.in

Abstract: **For evaluating the vehicle's services and sales, the online review offers a bunch of information. The study looks at how a customer feels about the vehicle, which affects how many cars are sold. The brand image is damaged, and sales are affected when a review is wrong about a car, while positive reviews contribute to increased sales. The sales forecast process considers the online review data and previous sales data as potential sources for more precise sales predictions. Machine learning might speed up the time-consuming process of forecasting sales and understanding market trends. This method makes it easier to find the relevant words in postal correspondence relating to finance, automobile rules, environmental regulations, and customer service. In this study, we used a BERT (Bidirectional Encoder Representation Transformer) to collect consumer reviews of automobiles. The technique is made more visible and understandable by the ML algorithm, which also facilitates the creation and management of such assessments. So, the sales staff will have access to accurate information about the automotive industry, which will help them predict sales.**

Index Terms: **BERT; Sentiment Analysis; Machine Learning**

I. Introduction

Long-term viability in any manufacturing business is ensured by understanding past performance, current events, and most crucially, anticipating future performance. Forecasting the company's future performance will calculate fortune and demand. Primary sales data must be compiled and reviewed to estimate future sales, fulfill the need, or reduce risk. The discipline of foreseeing the future is called predictive analytics, a subfield of data science. Predictive analysis requires a statistical technique, mathematical procedures, and visualization. We propose a fully automated sentiment analysis model (AABSA) based on these aspects. The model can generate aspect-based sentiment scores from online reviews without human intervention or domain knowledge. We used K-means clustering to group sentence embedding in the AABSA model, then selected the center words from each cluster as aspects.

The choice of a vehicle for a consumer is heavily influenced by its performance, appearance, efficiency, and safety [8]. To expand the brand's client base, the manufacturer must consider the feedback. The sentiment analysis of the brand and product released might help the firm enhance its sales predictions even further [9].

Products, territories, clientele groups, and other relevant information is used to examine prior sales experience. Managers must also look back over a long period to find trends and patterns in the growth and decrease of sales volume [10].

The capability of a company's manufacturing, marketing, funding, leadership, and ability to react to optimize profit potential all affect how well it can respond to sales forecast outcomes. The position of sales in the market is considered

The intention is to apply aspect-based sentiment analysis to comprehend market trends, the state of the country, and brand perception. But first, maintaining the course and manually identifying components for each newly established category and new customer requirements [1].

Businesses find it difficult to derive insights from the vast volumes of unstructured review information regarding their own and rivals' products. A one-dimensional number cannot adequately represent the many facets of a product [2].

To identify word similarities, some subsequent studies used word embedding methods such as word2vec and wordnet clustering [3]. These efforts, however, fall short in capturing the intricate semantic linkages of aspect keywords.

Based on the features of the cars, the review system assigns a rating. The evaluation of the product's effectiveness, the success of the previous vehicle models, safety, performance, and the vehicle's appearance and feel is calculated. For sales forecasting, the exact numbers are employed [4]. Every client considers these essential aspects before purchasing a car; therefore, considering them when predicting sales would result in more accurate predictions.

This work makes several significant and methodological contributions. This paper proposes a hierarchical framework, which makes several substantive and methodological advances. We offer a novel approach to determine product qualities by submitting a hierarchical framework. We demonstrate three comparative benefits of the proposed method over benchmark techniques: 1) greater precision, 2) more accurate sales forecasting and 3) full automation without laborious hand-coding to increase prediction accuracy; this review rating is used to the other sales forecasting characteristics.

II. Statistical Approach

Data analysis using statistical approach can provide insight into future business performance. For instance, in work [13], weather-based Machine Learning Technique for Day-Ahead wind power forecasting, wind power is forecasted statistically using historical data. Similarly, the article Intelligent Sales Prediction the Gradient Boost Algorithm, a machine learning technique, anticipates predicting sales turnover [14].

The study [12] demonstrates how businesses generate competitive information using data analytics. The business organization can select the Analytics type based on its requirements and available resources [11].

Received on 29-07-2022, Revised on 01-10-2022, Accepted on 05-11-2022.

TABLE I.
CUSTOMER REVIEW DATA ON THEIR VEHICLE

Sl. No.	Review Date	Name	Vehicle Name	Review Title	Review	Rating
1	04/14/08 10:47 PM (PDT)	Thomas	1997 Maruthi Passenger vehicle Minivan All- Trac 3drMin...	Best Minivan ever	My 1997 Passenger vehicle is the third one that I	5
2	11/12/08 05:31 PM (PST)	Susruta	1997 Maruthi Passenger vehicle Minivan All- Trac 3drMin...	My Favorite car Ever	I have owned lots of vans, and the Passenger vehicle is ...	4.8
3	04/14/10 07:43AM (PDT)	Andra	1997 Maruthi Passenger vehicle Minivan 3dr Minivan	Mom's Taxi Babies Ride	Sold 86 MaruthiVan 285K miles tobe replaced ...	5
4	12/17/16 04:40 PM (PST)	Nitish	1997 Maruthi Passenger vehicle Minivan All- Trac 3drMin...	My 4th Passenger vehicle, the best van ever made!	1st 95went over 300k beforebeing totaled	5
5	02/02/17 07:53 PM (PST)	Devender	1997 Maruthi Passenger vehicle Minivan	Great Vehicle, Maruthi's best design ever. Thankyou	There isno way back. Enjoy what you have.	5

The Automatic Aspect-Based Sentiment Analysis (AABSA) method was designed to extract hierarchical product attributes from online consumer reviews [5]. Table1. Based on these factors, we provide a completely automated sentiment analysis technique (AABSA). The algorithm can produce aspect-based sentiment scores from web reviews without assistance from a person or specialized knowledge. In the AABSA model, we grouped phrase embeddings using k-means clustering and selected the center words from clusters as attributes [6], [7]. A significant benefit is that since k-means clustering is an unsupervised learning model, we don't need to select the characteristics beforehand manually. Instead, AABSA may automatically identify several elements as well as aspect structures [16].

Furthermore, labels are not required throughout the learning process. As a result, it may independently determine the format of its information. The concept eliminates labeling and automatically allows us to identify aspects.

III. BIDIRECTIONAL ENCODER REPRESENTATION TRANSFORMER (BERT)

BERT is a pretraining approach for language representation to create open-source NLP models that can be downloaded and utilized [15]. These algorithms may be adjusted to deliver state-of-the-art predictions for a particular profession using your data or to extract top-notch linguistic traits from text data.

The aspect-based sentiment analysis uses the processes listed below to get the sentiment score.

1. Preprocess reviews.
2. Train word embeddings.
3. Train sentence embeddings.
4. Select hypernym candidates.
5. Hyponyms to assist in the subsequent
6. Sentiment analysis.
7. Merge hypernyms and hyponym candidates.
8. Match hypernyms to reviews
9. We use the Maximum Entropy (MaxEnt)

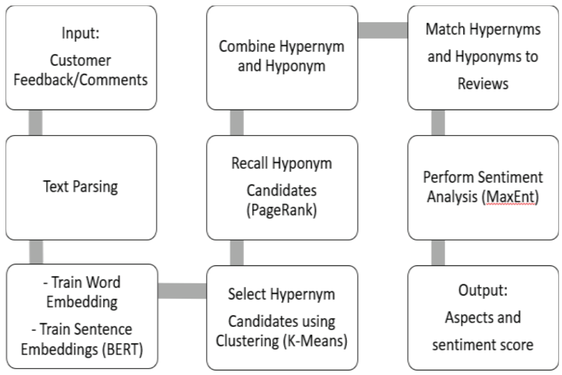

Figure 1. AABSA Framework

A multi-layer bidirectional Transformer encoder is called a BERT. BERT was developed by Google researchers in 2018 and has been shown to be cutting-edge for various natural language processing tasks such as text categorization, text summarization, and text production. Furthermore, a new disclosure claims that BERT plays a substantial role in Google's search algorithm, which helps it better evaluate queries. The study introduces two models, as seen in Figure 2.

BERT base – 12 layers (transformer blocks), 12 attention heads, and 110 million parameters.

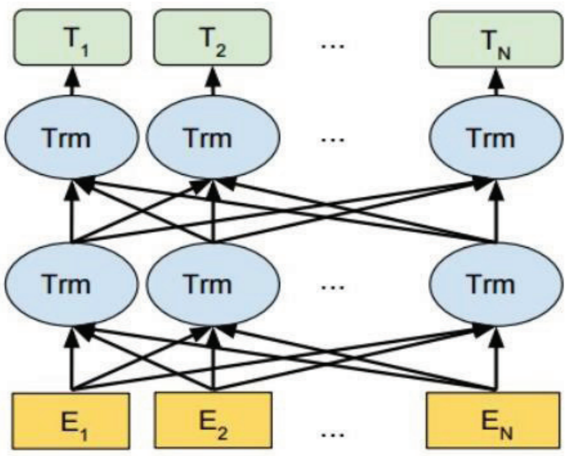

Figure 2. BERT Layers

BERT Large – 24 layers, 16 attention heads, and 340 millionparameters.

CVR Journal of Science and Technology, Volume 23, December 2022
DOI: 10.32377/cvrjst2316

Position Embeddings: To represent the placement of words in a sentence, BERT learns and uses positional embeddings. These are intended to make up for the fact that, Transformer cannot record "sequence" or "order" information Figure 3.

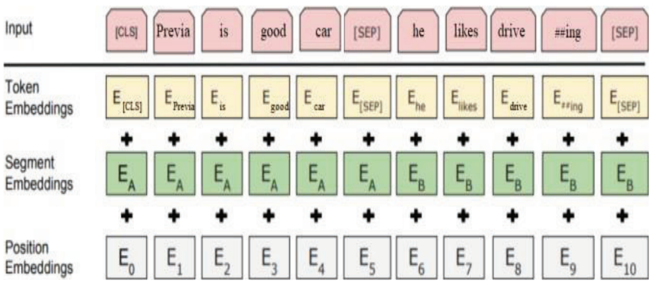

Figure 3. BERT sentiment analysis

Segment Embeddings: BERT can now accept sentence pairs as task inputs (Question-Answering). It learns a unique embedding for each to help the model distinguish between the first and second sentences. All the tokens marked as EA in the above example belong to sentence A.

TABLE II.
DATASET WITH A SENTIMENT SCORE

Month	Year	Brand	Sales	Inflation	Vehicle Sentiment
1	2021	Maruthi	1,36,591	1.10	1
2	2021	Maruthi	1,51,625	1.51	1
3	2021	Maruthi	2,00,801	2.22	1
4	2021	Maruthi	2,01,491	4.43	1
5	2021	Maruthi	2,03,009	5.21	1
6	2021	Maruthi	1,73,609	5.52	1
7	2021	Maruthi	1,54,731	5.61	1
8	2021	Maruthi	1,56,073	5.13	1
9	2021	Maruthi	1,28,495	5.34	1
10	2021	Maruthi	1,19,400	6.13	1
11	2021	Maruthi	1,24,506	6.52	1
12	2021	Maruthi	1,42,591	7.14	1

These are the embeddings for a particular token from the Word Piece token vocabulary that has been learned.

There are a lot of transformer encoders available from BERT. BERT assists with self-attention in both directions and is bidirectional.

The many preprocessing techniques are the source of BERT's adaptability. Thus, we can train the model on various NLP tasks without substantially altering its design Figure 1.

BERT uses WordPress tokenization. The vocabulary first starts with all of the constituent characters of the language, and then the most frequent/likely word combinations are repeatedly added.

TABLE III.
CORRELATION OF INPUT COLUMNS

	Month	Year	Total sales	Inflation	semiconductor Ship shortage	Vehicle Sentiment
Month	1.0	0.0	0.2	0.2	0.0	0.1
Year	0.0	1.0	0.3	0.5	0.6	0.8
Total Sales	0.2	-0.3	1.0	0.2	0.3	0.5
Inflation	0.1	0.5	0.2	1.0	0.8	0.4
Semiconductor Ship shortage	0.0	0.6	0.3	0.8	1.0	0.7
Vehicle Sentiment	0.1	0.7	0.5	0.4	0.7	1.0

Table 3 shows that the cell values with minus signs represent the positive and negative correlation. We can observe that the epidemic has a stronger association with decreased sales. By determining the external factors correlated with the overall sales, we might arrive at a reasonable sales projection.

IV. METHODOLOGY

The prediction model must include the following stages to create a realistic machine learning model to anticipate automobile sales.

1. Find the highly connected variables that affect the vehicle's sales.
2. Three methods are used to build the model: Random Forest, Support Vector Regression, and Multiple Linear Regression.
3. All three models' accuracy and sales are calculated using the Mean squared error approach while building the models.
4. With the sales and affecting factor, the identical procedure was performed, and the accuracy for the three models was determined using the Mean squared error technique.
5. The effectiveness of both approaches is determined by comparing the results.

Along with sentiment analysis, the dataset includes the values of previous sales and elements that may influence future sales—table 2.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = Mean Squared Error
n = number of data points
Y_i = Observed values
$\hat{Y}_i$ = Predicted values

A standard error metric for regression problems is the Mean Squared Error MSE. It is a loss function for algorithms that optimize or fit regression scenarios using the least-squares framework. Reducing the mean squared error between the anticipated and expected values is referred to as "least squares."

The squaring also inflates or amplifies significant errors; the more positive squared errors there are, the more significant the difference between the expected and predicted figures. Models receive a greater "punishment" for more severe mistakes when MSE is used as a loss function. Increasing the average error score when employed as a statistic also has the effect of "punishing" models.

The loss function is computed using the actual and expected values in Table 4. The outcomes of the three algorithms are then contrasted.

TABLE IV.
ACCURACY OF ML MODELS

Algorithm	Accuracy Without Reviewfactors	AccuracyWith the ReviewFactors
Linear Regression	2500	1800
SVM.SVR	1900	1500
Random Forest Regressor	1600	1200

Table 4 shows that the Model's Accuracy is increased for all three models with the review score. By comparing without review factors and review factors, the loss can be reduced by 72% of Linear Regression, 78.94% of SVM, and 75% of Random Forest Regression.

V. CONCLUSIONS

Using sentiment analysis to make predictions has increased prediction accuracy. The AABSA model improves sales forecasting's thoroughness and accuracy. As aspect identification is automated, it does not require labor-intensive manual coding. To anticipate sales, robust machine learning algorithms are also used.

REFERENCES

[1] Johnson, R., Zhang, T.: Supervised and semi-supervised text categorization using lstm for region embeddings. arXiv preprint arXiv:1602.02373(2016)

[2] Nguyen, T.H., Shirai, K.: Phrasernn: Phrase recursive neural network for aspect-based sentimentanalysis. In: Proceedings of the 2015 Conference on empirical methods in Natural Language Processing. pp. 2509{2514 (2015)

[3] Contiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.S., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al.: Semeval-2016 task 5: Aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016). pp. 19{30 (2016)

[4] Radford, A., Jozefowicz, R., Sutskever, I.: Learning to generate reviews and discovering sentiment. arXiv preprint arXiv:1704.01444 (2017)

[5] Timoshenko, A., Hauser, J.R.: Identifying customer needs from user-generated content. Marketing Science 38(1), 1{20 (2019)

[6] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., e, Q.V.: Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237 (2019)

[7] Cheriyan, S., Ibrahim, S., Mohanan, S., & Treesa, S.(2018, August). Intelligent Sales Prediction UsingMachine Learning Techniques. In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE) (pp. 53-58).IEEE.

[8] CyrilForopon, JayanthiRanjan Big Data Analytics in Building the Competitive Intelligence of Organizations, 2020, Elsevier

[9] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3).

[10] Gurnani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., & Bhirud, S. (2017, February). Forecasting of sales by using a fusion of machine learning techniques. In 2017 International Conference on Data Management, Analytics andInnovation (ICDMAI) (pp. 93-101). IEEE.

[11] Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., & Ekmis, M. A. (2019). An improved demand forecasting model using a deep learning approach and proposed decision integration strategy for the supply chain: *complexity*, *2019*.

[12] Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis.2017 - Elsevier

[13] Dolara, A., Gandelli, A., Grimaccia, F., Leva, S., & Mussetta, M. (2017, November). Weather-based machine learning technique for Day-Ahead windpower forecasting. In 2017 IEEE 6th international conference on renewable energy research and applications (ICRERA) (pp. 206-209). IEEE.

[14] Sai, B. K., & Sasikala, T. (2019, April). Predictive Analysis and Modeling of Customer Churn in Telecom using Machine Learning Technique. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 6-11). IEEE.

[15] Hoang M, Bihorac OA, Routes J. Aspect-based sentiment analysis using best. InProceedings of the 22nd Nordic Conference on Computational Linguistics 2019 (pp. 187-196).

[16] Baboota, Rahul, and Harleen Kaur. "Predictive analysis and modeling football results using machine learning approach for English Premier League." International Journal of Forecasting 35.2(2019): 741-755.