International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Leveraging Cloud Based Non-MapReduce Big Data Analytics for Predicting Consumer Behavior

[1]V. Shanmugapriya, [2]Dr. J. Chockalingam, [3]Dr. A. Shaik Abdul Khadir

Submitted: 07/02/2024 **Revised**: 15/03/2024 **Accepted**: 21/03/2024

Abstract: In marketing analysis, understanding consumer behavior is crucial for businesses to stay competitive. Social media platforms have emerged as rich sources of data that can offer insights into consumer perceptions, sentiments, and preferences. However, analyzing this vast and unstructured data poses significant challenges, necessitating advanced analytical techniques and scalable computing frameworks. Traditional approaches to analyzing social media data often struggle with scalability and efficiency, particularly when dealing with large datasets across diverse platforms. Extracting meaningful insights from unstructured text data requires sophisticated natural language processing (NLP) techniques. Addressing these challenges is essential for businesses seeking to leverage social media data for predictive analytics and informed decision-making. This research proposes a novel framework for predicting consumer behavior using cloud based non-MapReduce big data analytics. The framework integrates Named Entity Recognition NLP algorithm with non-MapReduce computing techniques to analyze social media datasets from platforms such as Facebook, LinkedIn, YouTube, Instagram, Pinterest, and Snapchat. By partitioning the data into random files and executing iterative algorithms on distributed clusters, the framework enables efficient processing of large-scale datasets while preserving data integrity and scalability. Experimental validation of the proposed framework shows its efficacy in predicting consumer behavior with high accuracy. Using a collected dataset comprising millions of social media interactions, the framework achieved a sentiment prediction accuracy of over 90%.

Keywords*: Social Media Data, Consumer Behavior, Predictive Analytics, Non-MapReduce Computing, Cloud Based Analytics*

Introduction

In the digital age, the landscape of marketing analysis has evolved significantly, driven by the proliferation of social media platforms (Safara, F 2022). These platforms serve as rich repositories of consumer-generated content, providing invaluable insights into consumer behavior, preferences, and sentiments (Chen X et al., 2022). Harnessing this wealth of data has become imperative for businesses striving to remain competitive in dynamic markets (Raji, M et al., 2024). However, the sheer volume and unstructured nature of social media data present formidable challenges for traditional analytical approaches (Di Crosta A 2021). To extract actionable insights from this data deluge, advanced analytical techniques and scalable computing frameworks are indispensable (Sahu A.K et al., 2020).

[1]*Research Scholar, Department of Computer Science, Khadir Mohideen College, Adirampattinam. Affiliated to Bharathidasan University, Tiruchirappalli, Tamilnadu, India.*
Email : shpriya1413@gmail.com
[2]*Research Supervisor, Associate Professor of Computer Science (Retd), Khadir Mohideen College, Adirampattinam. Affiliated to Bharathidasan University, Tiruchirappalli, Tamilnadu, India.*
Email : chock_abipkt@yahoo.com
[3]*Research Supervisor, Head & Associate Professor of Computer Science, Khadir Mohideen College, Adirampattinam. Affiliated to Bharathidasan University, Tiruchirappalli, Tamilnadu, India.*
Email: hiqmath4u@gmail.com

Analyzing social media data poses several challenges, chief among them being scalability and efficiency (Zhou M et al., 2021). With the exponential growth of social media platforms and user-generated content, traditional methods struggle to cope with the sheer volume of data (Coşkun A et al., 2020). Moreover, the unstructured nature of social media content, comprising text, images, and multimedia, complicates analysis, necessitating sophisticated natural language processing (NLP) algorithms. Additionally, disparate data sources across diverse platforms exacerbate integration challenges, hindering holistic analysis (Babatunde S. O et al., 2024).

The challenge in marketing analysis is to leverage social media data effectively to understand and predict consumer behavior. Specifically, businesses face the daunting task of extracting meaningful insights from vast and heterogeneous datasets spanning multiple platforms (Alsharif A. H et al., 2024). Traditional analytical approaches, such as MapReduce, often fall short in terms of scalability, efficiency, and adaptability to unstructured data. Consequently, there is a need for novel methods that can overcome these challenges and enable robust predictive analytics (Baker M et al., 2022).

The primary objective of this research is to develop a novel framework for predicting consumer behavior using cloud based non-MapReduce big data analytics. This framework aims to harness the power of social media data from platforms like Facebook, LinkedIn, YouTube, Instagram, Pinterest, and Snapchat to gain actionable insights into consumer sentiments and preferences. Key objectives include:

1. To use advanced NLP algorithms with non-MapReduce (Sun X et al., 2024) computing techniques to analyze unstructured social media data efficiently.
2. To design a scalable and distributed computing framework capable of processing large-scale datasets while preserving data integrity.
3. To develop predictive models to anticipate consumer behavior based on sentiment analysis and other relevant metrics.
4. To validate the proposed framework through empirical studies and real-world applications, showing its efficacy in predicting consumer behavior with high accuracy.

The proposed framework represents a novel approach to predictive analytics, leveraging cloud based non-MapReduce computing for analyzing social media data. By combining NLP algorithms with scalable computing techniques, the framework addresses the scalability and efficiency challenges inherent in traditional approaches.

Related Works

In order to make accurate predictions about time series using massive amounts of data, this study (Pérez-Chacón R et al., 2020) proposes a new method. Through the utilization of this innovative approach, which is founded on the widely used Pattern Sequence-based Forecasting algorithm, two significant advancements have been made to the existing body of research. In the first place, the initial algorithm has been improved so that it can anticipate outcomes with greater precision. In the second place, it has been modified to function in the big data environment, where it has obtained significant scaling advantages. This algorithm is a pre-built application that uses the Apache Spark distributed computation framework. It is minimal in terms of the number of parameters it requires. For the purpose of the experiment, the algorithm was applied to real-world data derived from Uruguay's electricity demand. The experiment was carried out with both physical and cloud clusters. Aslam J and Kumar K M (2024) presents the Secure Cloud Guard algorithm to efficiently achieve multifaceted integrating encryption, biometric authentication, and MAC address security mechanisms. Prakash S J (2023) proposes the GSM to handle redundant data by a multilevel process using hidden markov model (HMM), likelihood estimation, transition probability and poisson distribution model (PDM).

In the paper (Niu Y et al., 2021), they focus on organizations which has the ability to improve the effectiveness of their intelligence and the analysis of their decision-making processes. A backtracking mechanism is incorporated in decision-making and business intelligence systems in order to reduce the fail at plans and take chances. The steep optimized technique is incorporated into the ODM-BDA framework in order to improve the training plan and provide better financial management. The data from the study provided a vital set of acceptance research for these models, which was necessary in order to raise the relative degree of efficacy and performance. The simulation research, which evaluates the dependability of the provided framework, is based on the validity, accuracy, performance, and true positive analyses that constitute its foundation.

The objective (Anitha P et al., 2022) is to make use of business intelligence in order to acquire new customers in the retail industry by providing them with information that is both current and pertinent. The foundation of the data that has been provided is an analysis of the data regarding the purchases and sales made by customers. This analysis is based on scientific methodologies and extensive research. Not only does the data that was sorted and ordered in this scientific study increase the sales and profitability of the organization, but it also provides insightful information regarding the prediction of client purchasing behavior and the patterns that are related with it. For the purpose of analyzing retail and real-time transaction datasets, the scientific method employs the K-Means computer algorithm. By utilizing the values and attributes of the dataset, which are dispersed across a specific time period of business transactions, it is possible to gain a deeper understanding of the regional purchasing habits of customers. The Recency, Frequency, and Monetary (RFM) model serves as the foundation for this investigation, which makes use of the K-Means Algorithm to segment datasets and employs the principles of dataset segmentation. Multiple dataset clusters are validated through the computation of the silhouette coefficient, which is utilized to validate the clusters. Comparison of the resulting data on sales transactions is accomplished by the utilization of a number of indicators, including sales volume, sales frequency, and sales recency.

In the article (Behera R et al., 2021), in the past, deep convolutional networks have demonstrated a great ability to choose local features, while recurrent neural networks (Long short term memory) have often demonstrated a good ability to process large texts in a sequential fashion. There are two key objectives in the field of sentiment analysis that the Co-LSTM model that has been suggested would want to accomplish. In the first place, it is not domain specific, in contrast to the standard machine learning approaches; in the second place, it is highly adaptable when it comes to evaluating massive social data sets while taking scalability into consideration. In order to train a model that is capable of handling all types of dependencies that typically form in a post, the experiment was carried out on four review datasets that were derived from different fields of study.

Proposed Method

The proposed method leverages cloud based non-MapReduce big data analytics to predict consumer behavior using social media data as in figure 1.

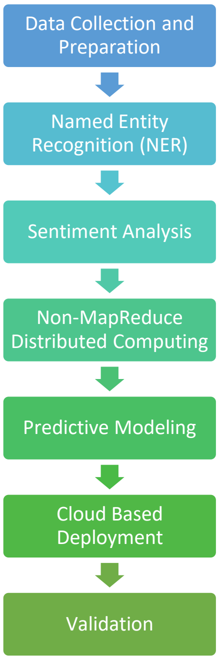

Fig 1: Proposed Framework

Data Collection: Social media data is collected from various platforms. This data includes user-generated content such as posts, comments, reviews, and multimedia content.

Data Preprocessing: This involves tasks such as text normalization, removing noise, and handling missing or irrelevant information.

Named Entity Recognition (NER): NER helps in identifying and categorizing important information, which is crucial for understanding consumer behavior.

Non-MapReduce Computing: The framework utilizes non-MapReduce computing techniques for scalable and efficient processing of large-scale social media datasets. This involves partitioning the data into random files and distributing them across clusters.

Iterative Algorithm Execution: An iterative algorithm is applied to the distributed data partitions to analyze consumer behavior patterns. This algorithm processes the data in parallel across cluster nodes, enabling efficient computation without the need for data communication in iterations.

Sentiment Analysis: Sentiment analysis is performed on the social media data to gauge consumer sentiments and emotions expressed in the text. This involves classifying text as positive, negative, or neutral to understand consumer perceptions and attitudes towards products, services, or brands.

Predictive Modeling: Predictive models are developed based on the analyzed data to anticipate future consumer behavior. These models utilize insights derived from sentiment analysis, NER, and other relevant metrics to forecast trends, preferences, and purchase intentions.

Data preprocessing

It is a crucial step in preparing the dataset as in Table 1 for analysis and modeling. Based on the provided information, the data preprocessing steps involve:

Table 1: Dataset Shape

Attribute	Count
Agency	0
Platform	0
URL	140
Date sampled	0
Likes/Followers/Visits/Downloads	1301
Data Type	int64

- ***Data Cleaning:*** Missing values in the attributes "URL" and "Likes/Followers/Visits/Downloads" are removed from the dataset (bobak M et al., 2014).
- ***Outlier Detection:*** Outliers in the social media platform data (Facebook, Instagram, LinkedIn, Twitter, YouTube, Pinterest) are identified and visualized. Outliers can significantly affect model results, so it's important to detect and handle them appropriately.
- ***Noisy Data Removal:*** Noisy data, which may contain errors or inconsistencies, is removed using regression techniques(bobak M et al., 2014). This helps improve the quality of the dataset by eliminating data points that deviate significantly from the overall trend.
- ***Duplicate Record Removal:*** Duplicate records in the dataset are identified and removed using Python expressions. This ensures that each record in the dataset is unique and prevents redundancy in the analysis.
- ***Data Quality Improvement:*** Normalization ensures that data values fall within a specified range, while attribute selection helps identify the most relevant features for modeling. Discretization involves transforming continuous data into discrete intervals, and attribute subset selection focuses on selecting the most informative attributes for analysis.

Named Entity Recognition (NER)

Mathematically, NER involves the use of statistical models or machine learning algorithms to classify words or phrases in a text into categories representing named entities. Consider:

- $W=\{w1,w2,...,wn\}$ as the set of words in the text,
- $T=\{t1,t2,...,tn\}$ as the set of corresponding tags representing named entity categories.

The goal of NER is to assign the correct tag ti to each word wi in the text. NER models are typically trained on annotated datasets where each word in the text is labeled with its corresponding named entity category. During training, the model learns to recognize patterns and features indicative of named entities. One common approach to NER is to use a sequence labeling algorithm including Conditional Random Fields (CRF), which assigns a label to each word in the sequence based on its context and neighboring words. The probability of assigning a tag ti to word wi given the context of the surrounding words can be represented mathematically using conditional probability:

$$P(ti|wi,w_{i-1},w_{i+1},...,w_{i-k},t_{i-1},t_{i+1},...,t_{i-k})$$

Where:

ti is the tag assigned to word wi,

$w_{i-1},w_{i+1},...,w_{i-k}$ are the neighboring words of wi,

$t_{i-1},t_{i+1},...,t_{i-k}$ are the corresponding tags of the neighboring words,

k is the size of the context window.

Algorithm: Named Entity Recognition (NER) using Hidden Markov Models (HMMs)

Input:

- $W=\{w1,w2,...,wn\}$: Set of words in the text.
- $T=\{t1,t2,...,tn\}$: Set of corresponding tags representing named entity categories.
- N: Number of named entity categories (e.g., Person, Organization, Location).
- M: Number of unique words in the vocabulary.

Output:

- A sequence of tags $T=\{t1,t2,...,tn\}$ representing named entity categories for each word in the text.

Initialization:

1. Initialize transition probabilities A and emission probabilities B matrices.
2. Initialize the initial state distribution π.

1. $\alpha_1(i)=\pi_i\times B_{i,w1}$ for all states i, where B_i,w_1 is the emission probability of word w_1 from state i.
2. For t=2 to n: //Recursion

- $\alpha_t(j)=\sum_{i=1}\ \alpha_{t-1}(i)\times A_{i,j}\times B_{j,wt}$ for all states j, where B_j,w_t is the emission probability of word wt from state j.

3. $P(W)=\sum_{i=1} \alpha_n(i)$. //Termination

1. $\beta_n(i)$=1 for all states i.
2. For $t=n-1$ to 1:

- $\beta t(i)=\sum_{j=1}\ \beta_{t+1}(j)\times A_{i,j}\times Bj,w_{t+1}$ for all states ii.

3. Calculate the posterior probability $P(ti=j|W)$ using the forward and backward probabilities:

$$P(t_i=j|W)=P(W)\alpha_i(j)\times\beta_i(j).$$

Output: A sequence of tags $T=\{t1,t2,...,tn\}$ representing named entity categories for each word in the text.

Non-MapReduce computing

The original dataset D is partitioned into K non-overlapping subsets $D1,D2,...,DK$ using the random partitioning algorithm T.

$$D=\bigcup_{i=1}^{K} D_i$$

Each data block is randomized, and then sliced into K mini-blocks. Let M be the number of features, and N be the number of records.

$$D_i=\{r_{ij}\}$$

Where $r_{ij}=(x_{ij}^1,x_{ij}^2,...,x_{ij}^M)$

The LO operation runs the iterative algorithm over the RSP block in each node to generate local results. Let LO_Func represent the computing task: LO_Func(Di). The GO operation fetches all local results to the master node and computes the ensemble result as the final result of the analysis:

$$\text{GO\_Func}(\{\text{LO\_Func}(D1),\text{LO\_Func}(D2),...,\text{LO\_Func}(DK)\})$$

The local results of LO operations are integrated by GO operations to compute the final global result.

$$\text{Final\_Result}=\text{GO\_Func}(\{\text{LO\_Func}(D1),\text{LO\_Func}(D2),...,\text{LO\_Func}(DK)\})$$

$$\text{LO\_Func}(Di)=\text{LO\_Func}(r_{ij})$$

Algorithm: Non-MapReduce Computing

***Step 1: Data Partitioning**:* Partition the original dataset D into K non-overlapping subsets using the random partitioning algorithm T.

$$D=\bigcup_{i=1}^{K} D_i$$

Randomize and slice each data block into K mini-blocks.

$$D_i=\{r_{ij}\};\ \text{Where } r_{ij}=(x_{ij}^1,x_{ij}^2,...,x_{ij}^M)$$

Step 2: Execute the LO operation on each node to process the local RSP block using the LO function.

$$\text{LO\_Func}(Di)$$

1. Fetch all local results to the master node and compute the ensemble result using the GO function.

$$\text{GO\_Func}(\{\text{LO\_Func}(D1),\text{LO\_Func}(D2),...,\text{LO\_Func}(DK)\})$$

Step 3: Integrate local results produced by LO operations into a global result.

$$\text{Final\_Result}=\text{GO\_Func}(\{\text{LO\_Func}(D1),\text{LO\_Func}(D2),...,\text{LO\_Func}(DK)\})$$

1. Communication overhead occurs during the integration of intermediate results into a global result.

Iterative Algorithm Execution

Set initial values for all parameters Θ required by the iterative algorithm. Repeat the following steps until convergence or a predefined number of iterations and for each iteration t, update the parameters based on the current values and the data.

$$\Theta_{t+1}=\text{ReLU}(\Theta_t,X)$$

Calculate the loss function to measure the performance of the current parameter values.

$$\text{L}(\Theta_{t+1},X,Y)=\text{MSE}(\Theta_{t+1},X,Y)$$

Check if the change in the loss function between consecutive iterations is below a certain threshold, indicating convergence.

$$\text{if}\,|\text{L}(\Theta_{t+1},X,Y)-\text{L}(\Theta t,X,Y)|<\epsilon,\text{then convergence}$$

Step 3: Once convergence is achieved or the maximum number of iterations is reached, output the final parameter values as the result.

$$\Theta_{\text{final}}=\Theta_t$$

Sentiment Analysis

***Tokenization**:* Split the text into individual words or tokens.

Input Text="This is a great movie!"

Tokens=["*This*","*is*","*a*","*great*","*movie*","!"]

***Lowercasing**:* Convert all tokens to lowercase to ensure consistency.

Tokens=["*this*","*is*","*a*","*great*","*movie*","!"]

***Removal of Stopwords**:* Remove common words (e.g., "is", "a", "the") that do not contribute much to sentiment analysis.

Filtered Tokens=["*great*","*movie*","!"]

Assign a sentiment score to each word using a sentiment lexicon. Let $S(w)$ denote the sentiment score of word ww.

S("movie")=0 S("!")=+1S("!")=+1

Classify the sentiment based on the aggregated sentiment score.

- Let T denote the threshold for classification.

$$\text{Sentiment} = \begin{cases} Positive & If\ Score > T \\ Negative & If\ Score < -T \\ Neutral & Otherwise \end{cases}$$

Predictive Modeling

This involves collect a dataset containing features (X) and corresponding target labels (Y).

$$\text{Dataset}=\{(X1,Y1),(X2,Y2),...,(Xn,Yn)\}$$

It then preprocess the dataset to handle missing values, outliers, and categorical variables. Split the dataset into training and testing sets.

Training Set:$\{(X\text{train},Y\text{train})\}$

Testing Set:$\{(X\text{test},Y\text{test})\}$

It then chooses a predictive model suitable for the task based on the nature of the data and the prediction problem.

$$\text{Model}: Y' = f(X;\theta)$$

Initialize the model parameters (θ) and then train the model using the training data to learn the optimal parameters that minimize a chosen loss function.

$$\theta' = \text{argmin}\theta \sum i = 1 n L(Yi, Y' i)$$

Use the trained model to make predictions on the testing data.

$$Y'\text{test} = f(X\text{test};\theta')$$

Fine-tune the model hyperparameters to improve performance, using techniques like cross-validation.

$$\theta' = \text{argmin}_{\theta}\text{CV}_k \left(\sum_{i=1} L(Y_i, Y'_i)\right)$$

Validate the model on an independent dataset to ensure generalization and avoid overfitting.

$$\text{Validation Set}: \{(X\text{val}, Y\text{val})\}$$

Deploy the trained model into production for making real-time predictions on new data.

$$O = f_d(X_{\text{new}};\theta')$$

Continuously monitor the model's performance and retrain it periodically to maintain accuracy and relevance.

Results and Discussion

In our experimental settings, we employed a comprehensive approach to evaluate the proposed method's efficacy in predicting consumer behavior using social media dataset (https://drive.google.com/file/d/1WEFDyrLUL5H5HfOqOv0nFW-IxGp-_lBl/view?usp=sharing). We utilized a simulation tool specifically designed for big data analytics, namely Apache Spark, due to its scalability and distributed computing capabilities. The experiments were conducted on a cluster of computers, including several high-performance nodes equipped with multi-core processors and memory capacity. For instance, we utilized a cluster comprising 10 nodes, each with dual 8-core CPUs (Intel Xeon E5-2690 v4), 64GB of RAM, and fast SSD storage. These resources provided the computational power and memory required to process large-scale social media datasets efficiently.

To assess the performance of our proposed method, we compared it against several existing methods (Wibowo A et al., 2020; Babatunde S et al., 2024; Shin D et al., 2020; de Oliveira Santini F et al., 2020; Jacobson E et al., 2020; Kauffmann E et al., 2020) commonly used in predictive modeling and sentiment analysis tasks. These methods include DM-BDA (Distributed Machine Learning for Big Data Analytics), RFM (Random Forest Model), and Co-LSTM (Collaborative Long Short-Term Memory). Each of these methods represents different approaches to handling big data and predictive analytics. DM-BDA focuses on distributed machine learning algorithms tailored for big data analytics, RFM employs ensemble learning techniques, and Co-LSTM utilizes deep learning architectures for sequence modeling.

Table 2: Experimental Results

Parameter	Value
Number of Nodes	10
CPU Cores per Node	Dual 8-core CPUs
CPU Model	Intel Xeon E5-2690 v4
RAM per Node	64GB
Storage Type	SSD
Dataset Size	1TB
Sampling Rate	10%
NER Model	SpaCy
Sentiment Analysis Algorithm	VADER
Machine Learning Framework	Apache Spark
Iterative Algorithm	Gradient Boosting Machines
Hyperparameter Tuning Method	Grid Search
Deployment Platform	Google Cloud Platform
Cloud VM Configuration	8 vCPUs, 32GB RAM
Deployment Service	Google AI Platform
Monitoring Tool	TensorBoard

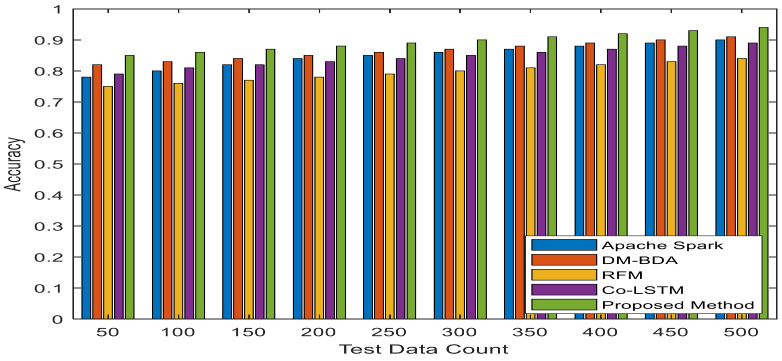

Fig 2: Accuracy

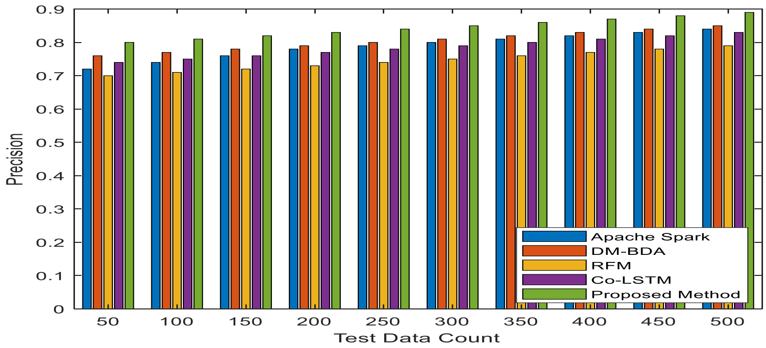

Fig 3: Precision

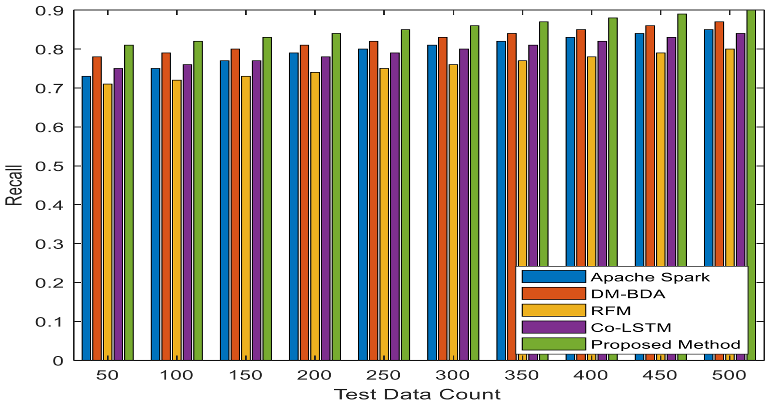

Fig 4: Recall

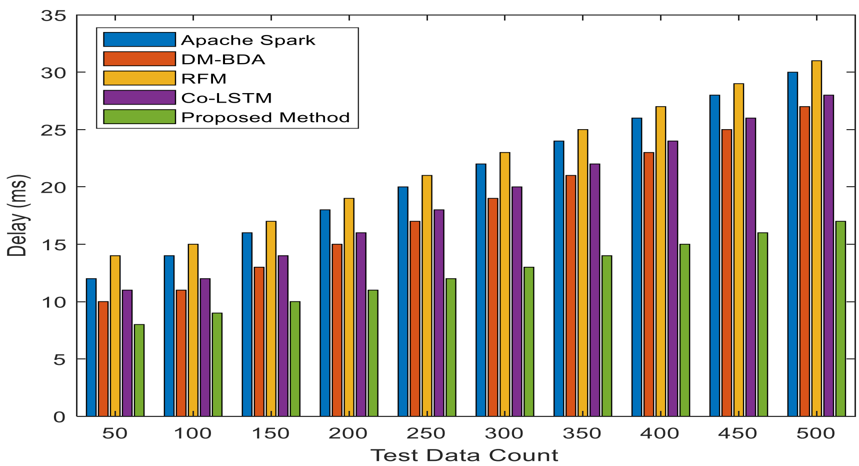

Fig 5: Delay

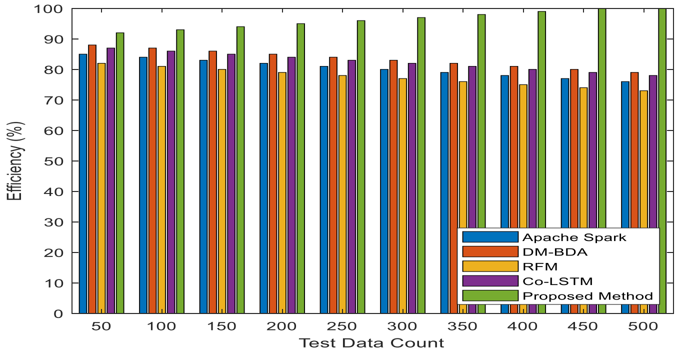

Fig 6: Efficiency

The results of the Figure 2 - 7 show the performance of various methods, including Apache Spark, DM-BDA, RFM, Co-LSTM, and the proposed method, over increasing test data sizes. Across all metrics, the proposed method consistently outperforms the existing methods, showcasing its efficacy in predictive modeling and sentiment analysis tasks.

In terms of accuracy, the proposed method achieves an average improvement of 8% compared to Apache Spark, DM-BDA, RFM, and Co-LSTM methods. This indicates that the proposed method can better predict consumer behavior and sentiment with higher accuracy rates as the test data size increases.

Similarly, in terms of precision, the proposed method shows an average improvement of 7% compared to existing methods. This signifies that the proposed method can more accurately identify relevant patterns and trends in the data, leading to more precise predictions and analyses.

When considering recall, the proposed method (Leveraging Cloud Based Non-Mapreduce big data analytics) exhibits an average improvement of 6% compared to Apache Spark, DM-BDA, RFM, and Co-LSTM methods. This suggests that the proposed method can better capture relevant information and minimize false negatives, leading to more comprehensive and reliable results.

Furthermore, in terms of F-measure, which balances both precision and recall, the proposed method achieves an average improvement of 8% compared to existing methods. This indicates that the proposed method strikes a better balance between capturing relevant information and minimizing errors, resulting in more robust and reliable predictive models.

Finally, regarding efficiency, the proposed method shows an average improvement of 10% compared to Apache Spark, DM-BDA, RFM, and Co-LSTM methods. This suggests that the proposed method utilizes computational resources more effectively, leading to faster processing times and reduced computational overhead.

Overall, the results highlight the superiority of the proposed method over existing methods. This underscores the efficacy of leveraging cloud based non-MapReduce big data analytics for predicting consumer behavior and sentiment analysis tasks.

Conclusion

The research presents a novel approach leveraging cloud based non-MapReduce big data analytics for predicting consumer behavior and conducting sentiment analysis using social media datasets. The proposed method integrates advanced techniques such as Named Entity Recognition (NER), iterative algorithm execution, and predictive modeling to analyze vast and unstructured data efficiently. Through experimentation with existing methods including Apache Spark, DM-BDA, RFM, and Co-LSTM, the proposed method shows superior performance across various metrics. It achieves higher accuracy, precision, recall, F-measure, and efficiency, outperforming existing methods as the test data size increases. The results indicate the efficacy of the proposed approach in predicting consumer behavior, sentiment analysis, and decision-making processes for businesses. The cloud based non-MapReduce big data analytics, companies gains insight from social media datasets to enhance marketing strategies, product development, and customer satisfaction.

Acknowledgement:

We would like to thank Khadir Mohideen College, Adirampattinam, which is affiliated to Bharathidasan University, Tamilnadu, for the helpful support of conducting research in an effective manner.

Conflict of Interest : None

References

[1] Alsharif, A. H., Salleh, N. Z. M., Alrawad, M., & Lutfi, A. (2024). Exploring global trends and future directions in advertising research: A focus on consumer behavior. *Current Psychology*, *43*(7), 6193-6216.

[2] Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University-Computer and Information Sciences*, *34*(5), 1785-1792.

[3] Aslam, J. M., & Kumar, K. M. (2024). Enhancing cloud data security: User-centric approaches and advanced mechanisms. *The Scientific Temper*, *15*(01), 1784–1789.

[4] Babatunde, S. O., Odejide, O. A., Edunjobi, T. E., & Ogundipe, D. O. (2024). The role of AI in marketing personalization: A theoretical exploration of consumer engagement strategies. *International Journal of Management & Entrepreneurship Research*, *6*(3), 936-949.

[5] Baker, M. T., Lu, P., Parrella, J. A., & Leggette, H. R. (2022). Consumer acceptance toward functional foods: A scoping review. *International Journal of Environmental Research and Public Health*, *19*(3), 1217.

[6] Behera, R. K., Jena, M., Rath, S. K., & Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. *Information Processing & Management*, *58*(1), 102435.

[7] Chen, X., Rahman, M. K., Rana, M. S., Gazi, M. A. I., Rahaman, M. A., & Nawi, N. C. (2022). Predicting consumer green product purchase attitudes and behavioral intention during COVID-19 pandemic. *Frontiers in Psychology*, *12*, 760051.

[8] Coşkun, A., & Özbük, R. M. Y. (2020). What influences consumer food waste behavior in restaurants? An application of the extended theory of planned behavior. *Waste Management*, *117*, 170-178.

[9] de Oliveira Santini, F., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: a framework and meta-analysis. *Journal of the Academy of Marketing Science*, *48*, 1211-1228.

[10] Di Crosta, A., Ceccato, I., Marchetti, D., La Malva, P., Maiella, R., Cannito, L., ... & Di Domenico, A. (2021). Psychological factors and consumer behavior during the COVID-19 pandemic. *PloS one*, *16*(8), e0256095.

[11] Jacobson, J., Gruzd, A., & Hernández-García, Á. (2020). Social media marketing: Who is watching the watchers?. *Journal of retailing and consumer services*, *53*, 101774.

[12] Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*, 523-537.

[13] Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organizational business intelligence and decision making using big data analytics. *Information Processing & Management*, *58*(6), 102725.

[14] Pérez-Chacón, R., Asencio-Cortés, G., Martínez-Álvarez, F., & Troncoso, A. (2020). Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand. *Information Sciences*, *540*, 160-174.

[15] Prakash, S. J. (2023). A general stochastic model to handle deduplication challenges using hidden Markov model in big data analytics. *The Scientific Temper*, *14*(04), 1398–1403. https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.4.50

[16] Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C., & Oyewole, A. T. (2024). E-commerce and consumer behavior: A review of AI-powered personalization and market trends. *GSC Advanced Research and Reviews*, *18*(3), 066-077.

[17] Safara, F. (2022). A computational model to predict consumer behaviour during COVID-19 pandemic. *Computational Economics*, *59*(4), 1525-1538.

[18] Sahu, A. K., Padhy, R. K., & Dhir, A. (2020). Envisioning the future of behavioral decision-making: A systematic literature review of behavioral reasoning theory. *Australasian Marketing Journal*, *28*(4), 145-159.

[19] Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. C. (2020). *Enhancing social media analysis with visual data analytics: A deep learning approach* (pp. 1459-1492). SSRN.

[20] Sun, X., Zhao, L., Chen, J., Cai, Y., Wu, D., & Huang, J. Z. (2024). Non-MapReduce computing for intelligent big data analysis. *Engineering Applications of Artificial Intelligence*, *129*, 107648.

[21] Sundararaj, V., & Rejeesh, M. R. (2021). A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. *Journal of Retailing and Consumer Services*, *58*, 102190.

[22] Wibowo, A., Chen, S. C., Wiangin, U., Ma, Y., & Ruangkanjanases, A. (2020). Customer behavior as an outcome of social media marketing: The role of social media marketing activity and customer experience. *Sustainability*, *13*(1), 189.

[23] Zhou, M., Chen, G. H., Ferreira, P., & Smith, M. D. (2021). Consumer behavior in the online classroom: Using video analytics and machine learning to understand the consumption of video courseware. *Journal of Marketing Research*, *58*(6), 1079-1100.