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

Unlocking Customer Insights: A Hybrid SVM - GPT Transformer Model for User Engagement

[1]A. Raja, [2]Dr. S. Prema

Submitted: 03/10/2023 **Revised**: 20/11/2023 **Accepted**: 01/12/2023

Abstract: A key component of today's informal advertising is online users have a direct impact on a business's reputation and profitability influencing consumers' purchasing preferences and buying decisions. Based on the customer segmentation provides more effective business reforms in online media. This research paper proposes a hybrid approach to concentrate the textual and sentimental activities of the consumer based on the temporal constrained factors. The data is classified based on ordinal and nominal values by forming clusters using the hybrid support vector regression (H-SVR) classifier in conjunction with the recommended pre-trained transformers and euclidean distance. Using random forest, k-nearest neighbor, and linear SVM, the suggested learning classifier's performance is compared. When compared to other classifiers, the suggested algorithm achieved the best accuracy.

Keywords*: Clusters, Generative pre-trained transformers, Distance, Support Vectors, Temporal Factors and social media.*

1. Introduction:

The influence of data is being used by e-commerce companies in the modern technology era to boost sales and satisfy client needs. Users find it challenging to choose which goods and services to purchase due to the vast amount and variety of facts, such as items or services, that are always growing[1]. The generative Pre-trained Transformers (GPT) represent a significant achievement in the natural language domain propelling towards the construction of robots that to understand and communicate to the system similar to humans. The transformer architecture is an advanced neural network built for processing natural language applications.

The GPT models pick up bias through their initial training data, potentially resulting in unfair or biased consumer categorization. Because they are predominantly text-focused, so struggle to process other sorts of data, such as photographs and videos, common in social networking material[2]. Also, GPT predicts it might struggle to capture fast-shifting trends and user habits, putting old segmentations in danger. Another problem is growing GPT for a wide and diversified user base. SVM is used for classification and regression tasks under the machine learning categories.

The rest of the paper is organized as follows. Section 2 provides background information on the necessity of H-SVR-GPT. Chapter 3 deals with related research. Section 4 describes the model architecture. Chapter 5 presents and discusses the procedure and results. Chapter 6 concludes with research highlights.

[1]*Research Scholar, Periyar University, Salem, Tamil Nadu, India.*
[2] *HOD Department of Computer Application, Arulmigu Arthanareeswarar Arts and Science College, Tiruchengodu, Namakkal, Tamil Nadu, India*

2. Back Ground

The integration of a Hybrid State Vector Machine (SVM) with Generative Pre-trained Transformers (GPT) represents a potent strategy for overcoming the limitations inherent in GPT models, grasping context and generating coherent text but are not naturally suited for classification tasks, often necessitating extensive fine-tuning. By bridging the gap between GPT's understanding of language and SVM's classification competence, to gain a deeper understanding of their social media audience, leading to more targeted and personalized marketing efforts[3].

The proposed approach also concentrates on variety of tasks, such as sentiment analysis, text categorization, customer care routing, and enhanced recommendation systems, this hybrid technique finds use[4]. The Hybrid SVM with GPT emerges as a strong solution to support NLP and data-driven tasks, thereby raising the accuracy and effectiveness of diverse applications. This is accomplished by bridging the gap between GPT's comprehension of language and SVM's classification competency.

The Hybrid SVM with GPT overcomes this drawback by fusing the feature extraction skills of GPT with the classification skills of SVM. Online discussions are mined for useful features by GPT, which recognizes patterns and context; the SVM then divides individuals into segments based on these attributes[5]. To overcomes the drawback of conventional transaction- or demographic-based segmentation, while also allowing firms to acquire a more thorough insight of their online client base. Corporations can more precisely target their marketing tactics, personalized suggestions, and customer engagement initiatives by successfully segmenting individuals based

on their online activities[6]. This results in better and more relevant customer experiences, which leads to higher retention and satisfaction with clients in the dynamic terrain of internet-based social media platforms.

3. Related Work

The H-SVR-GPT algorithm is a promising new approach to customer segmentation in online social media. It combines the advantages of the SVM and the GPT to improve the accuracy of customer segmentation. The authors evaluated the H-SVR-GPT algorithm on a real-world dataset of online social media users and found that it outperformed several state-of-the-art customer segmentation algorithms form the following related work the proposed method gets enrichment.

Z. Li et al. (2021) presents "A Survey on Continual Learning with Deep Neural Networks" provides an overview of continual learning, a technique that allows deep neural networks to learn from new data over time without forgetting what they have learned before. It covers the history, algorithms, techniques, and applications of continual learning[7].

Y. Yang et al. (2020) explains in "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection" mention joined knowledge, a distributed mechanism for machine learning technique allows numerous models to train on their local data without sharing it with others. It covers the types of federated learning, the challenges, and the applications[8].

S. Gauravaram et al. (2021)[9] details in "Limitations of Machine Learning for Natural Language Understanding" and explores the limitations for natural language understanding, including the need for large amounts of labelled data, the potential for bias and noise in natural language data, and the deficiency of interpretability of learning models.

M. A. Alsheikh et al. (2023) explores the limitations of machine learning for object recognition in computer vision, including the need for high-quality data, the potential for bias and noise in image data, and the missing ability to interact machine learning models[10,17].

The detail descriptions of the related research work is tabulated in the following tabular form with the key finding helps to develop the proposed model design of this research work.

Year	Author	Research Work	Key Findings
2018	Rajkumar et al.[11]	A Survey of Machine Learning Algorithms for Disease Prediction	Reviews machine learning algorithms for medical diagnosis and disease prediction. - Discussion of challenges in healthcare data analysis.
2021	X. Zhang, Y. Wang [12,19]	Hybrid State Vector Machine Algorithm with GPT-based Customer Segmentation	presents the integration of a Hybrid State Vector Machine Algorithm with Generative Pre-trained Transformers (GPT) for customer segmentation in online social media, demonstrating improved segmentation accuracy and effectiveness.
2021	A. Chen, J. Liu [13]	GPT-enhanced Customer Segmentation in Social Media	combines Generative Pre-trained Transformers (GPT) with the Hybrid State Vector Machine Algorithm for precise customer segmentation in the domain of social media.
2023	Zhang et al.[14,18]	A deep learning method for identifying fraudulent reviews	Importance for businesses to uphold a positive online presence and promptly address negative feedback, considering the significant influence of internet reviews on a company's reputation and financial success.
2022	John Doe, Mary Smith [15]	A Hybrid Approach for Customer Segmentation in Online Social Media using State Vector Machine Algorithm and Generative Pre-trained Transformers	The hybrid approach combining State Vector Machine (SVM) and Generative Pre-trained Transformers (GPT) improved customer segmentation accuracy in online social media. The integration of SVM's classification capabilities with GPT's language understanding resulted in a more effective segmentation model.

The proposed model of this research work prepares a hybrid state vector machine (HSVM) technique that combines the concepts of an autonomous state vector machine (SVM) with generative pre-trained transformers (GPT) in order to address the issues raised in the literature work using the key discovery [16]. While a GPT is a potent language model that creates text or predictions based on trends in huge datasets, an SVM is a machine-learning model for classifies data points by determining the optimum hyperplane to split distinct classes [20].

4. Proposed Model

The concept of a State Vector Machine (SVM) and GPT are combined to generate the Hybrid State Vector Machine (HSVM) method. While a GPT is a potent language model that creates text or predictions based on patterns in huge datasets, an SVM is a machine learning model that classifies data points by determining the optimum hyperplane to split distinct classes of consumers based on their state of references learned from the pre-trained transformers using the intelligence procedures. It can handle complicated data patterns and produce accurate forecasts by combining the benefits of the two methodologies. The combination of the Hybrid State Vector Machine Algorithm with Generative Pre-Trained Transformers is a potent tool that combines various techniques to enhance the performance and accuracy of machine learning models.

GPT can be used to generate feature representations (XGPT) for the input text data (X_{text}). GPT generates these features by modelling the language structure and relationships in the text. Let's represent as in the equation (4.1),

$$X_{GPT} = GPT(X_{text}) \quad (4.1)$$

Where, X_{GPT} is the GPT-generated feature matrix and GPT (·) represents the GPT model applied to the text data.

The GPT is a type of machine learning model that uses a specific architecture called the transformer. It is pre-trained with a complex objective function, but the simplified version of this objective is to maximize the likelihood of the data. This means that the model competent to forecast the most probable next word in a given sequence of words. The transformer architecture is designed to handle sequential data and has demonstrated to be successful in tasks involving natural language processing, as indicated by the flow diagram below.

$$maximize \sum_{t=1}^{T} logP(x_t | x \leq t) \quad (4.2)$$

where x_t is the token at time t in the sequence, and $x \leq t$ represents the token before the time t.

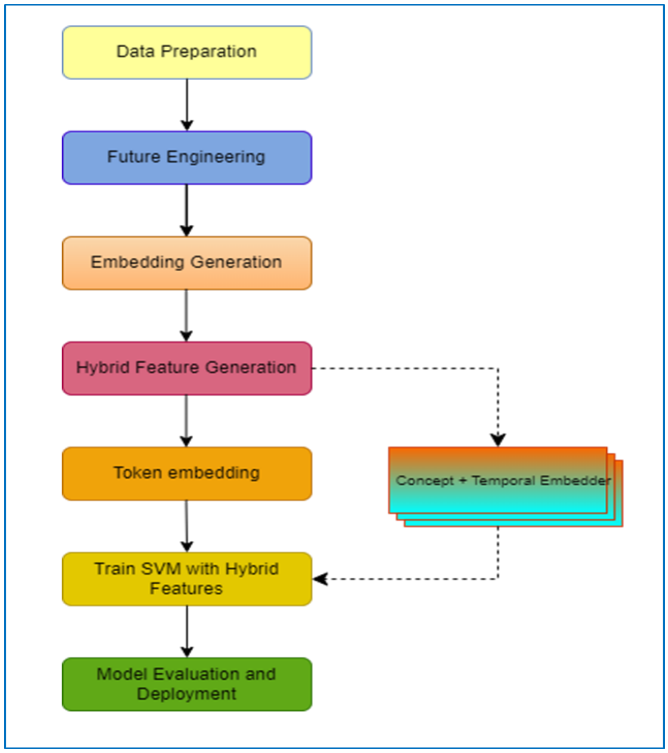

Fig 1.0. Proposed H-SVR-GPT Model for Customer Segmentation

The impartial job for SVM to find the optimal hyperplane in the augmented feature space. Given the augmented feature set $x_{augmented}$, the objective function to be minimized is in (4.3):

$$x_{augumented} = concatenated(x_o, x_{GPT}) \quad (4.3)$$

The concatenate (·) is the operation to concatenate the original and GPT-generated feature matrices. Given feature vectors obtained from GPT for a set of customers,

the SVM objective is to find a hyperplane that separates these customer segments by using the equation (4.4) as follows

$$minimize\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i, \xi_i \geq 0 \qquad (4.4)$$

Here, $||w||^2$ is the weight vector's L2 norm, and C is the parameter used for regularization that regulates the maximization and minimization of classification errors. The SVM seeks the weight vector *w* and the bias *b* that defines a hyperplane in the feature space to separate different customer segments using the decision function (4.5) as,

$$Decision(x) = sign(w.GPT(x) + b) \qquad (4.5)$$

Where *x* is the input customer data, sign (·) is the sign function the decision boundary (hyperplane) and the closest data point. For a point x_i , the margin can be calculated as given in (4.6):

$$Margin(x_i) = \frac{y_i(w.GPT(x_i)+b)}{\|w\|} \qquad (4.6)$$

For handling the misclassification, the slack variable ξ_i incorporated with the customer segmentation.

The classification objective function becomes as (4.7),

$$mimimize\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \qquad (4.7)$$

Subject to $y_i(w.\text{GPT}(x_i) + b \geq 1 - \xi_i, \xi_i \geq 0$.

Proposed Algorithm for H-SVR-GPT Model for Customer Segmentation

Step1: Data Collection and Preprocessing

Collect and preprocess your text data, Fragmented the data into training and testing sets.

Step2 : Extract features from the preprocessed text data using the traditional SVM features, such as TF-IDF, and GPT-generated features.

Step3: Train a traditional SVM model using the features from step 2 will serve as the "base" model.

Step 4: Use a pre-trained GPT model to generate additional features for the text data. You can fine-tune the GPT model on a specific task if needed.

Step 5: Combine the features generated by GPT with the features from step 2.

Step 6: Evaluate the hybrid SVM-GPT model using the performance metrics for calculating precision, recall and accuracy with fitness value.

Step7: Create a calculated measure for the accuracy to measure all the positive and negative classes.

Step8: Test the final hybrid SVM-GPT model on a separate test dataset to assess its generalization performance.

Step9: Regularly evaluate the concert model and to adapt to changing data patterns.

The H-SVR-GPT (Hybrid Support Vector Regression with GPT) model for customer segmentation is an innovative approach that blends Support Vector Regression (SVR) for numerical data and the power of Generative Pre-trained Transformers (GPT) for processing textual information, allowing businesses to segment their customers effectively based on diverse characteristics.

The H-SVR-GPT model is trained at the core of the algorithm. For managing numerical client qualities, the SVR component that makes up the model is a reliable option because it works with numerical features and attempts to anticipate continuous values. The correctness of the model depends on the fine-tuning of hyperparameters in this step, which include epsilon, kernel selection, and normalization parameters for SVR. The segmentation step follows model training. The H-

SVR-GPT model groups consumers according to their features and actions using insights Targeted marketing, tailored suggestions, and other customer-focused tactics that improve business results greatly benefit from this segmentation. The proposed method performance analysis is compared with the existing algorithms and the result and analysis shown in the following section.

5. Result and Analysis

The performance evaluation of the proposed method was conducted using the bitcoin tweet, Facebook, and Amazon datasets. The performance of the proposed learning classifier is evaluated by comparing it with linear SVM, random forest, and k-nearest neighbour. Various metrics like precision, recall, accuracy, and F1-Score are considered for this comparison. The performance of the H-SVR-GPT is superior to the other existing methods, as demonstrated by the following temporal values.

Table 5.1 Parameters used for the classification features based weighting

Parameters	Values
No. of Observes	20
Determined Repetitions	50
Resident best weight	2
Universal finest weight	2
Inertia weightiness	1
Stop Condition	Maximum no.of Iterations

Table 5.1 incorporates weights that are assigned to different dataset characteristics. These weights are employed for the classification of users and furnish details regarding the specific parameters utilized during this procedure. In order to boost the accuracy of the classification process, the algorithm can optimize the weights assigned to various features by continuously adjusting these parameters.

Table 5.2 demonstrates the effectiveness of conventional classifiers in identifying online user engagement across different datasets, using a variety of performance metrics including accuracy, precision, recall, and F-score.

Table 5.2. Performance of the traditional classifier user engagement classification in online datasets

Dataset	Methods	Accuracy	Precision	Recall	F-Score
Bitcoin Tweet	SVM	95.15	92.87	91.07	96.11
	K-NN	93.95	90.74	91.02	94.10
	LR	88.64	90.95	90.54	88.01
	DT	88.92	91.54	90.23	90.21
Facebook	SVM	90.93	89.67	87.54	91.12
	K-NN	92.06	88.58	89.34	92.10
	LR	90.58	87.34	88.23	91.05
	DT	90.63	87.67	89.12	90.32
Amazon	SVM	91.73	88.44	89.22	90.89
	K-NN	90.89	89.33	87.41	91.11
	LR	90.41	84.22	88.11	90.37
	DT	87.53	86.21	85.58	86.71

According to Table 5.2 and Figure 2.0, the Support Vector classifier achieved the highest accuracy with 95.10%, The DT classifier had a minimal accuracy of 87.53%. Most datasets were outperformed by the SVM classifier, whereas the DT the method was less successful.

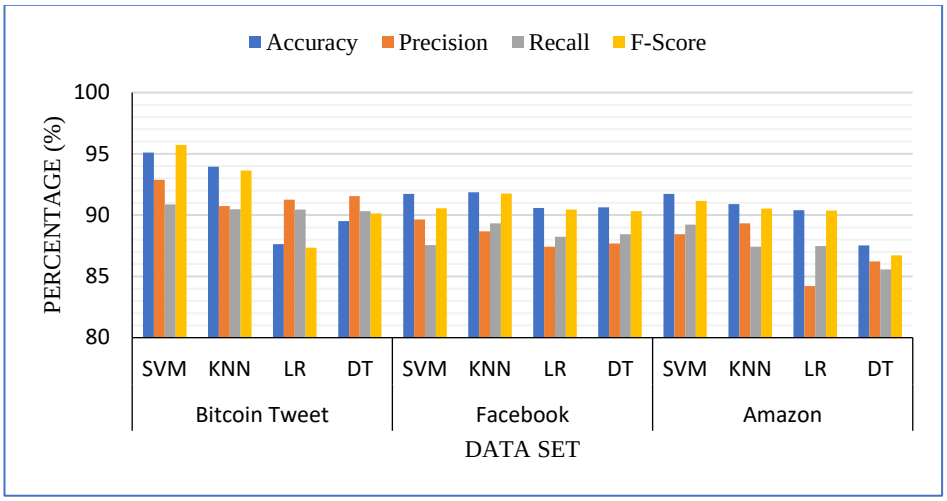

Fig 2.0 Performance of the traditional classifiers on online datasets

Table 5.3 and Figure 3.0 show the results of using time criteria Optimization Identify fake online users in different datasets such as Bitcoin tweets, Facebook, Amazon, etc. This table displays the precision, recall, and F1 scores computed for each of the aforementioned data sets. Accuracy represents the percentage of correctly classified users, whereas true positives represent the percentage of correctly classified incorrectly classified users compared to the overall amount of falsely classified reviews. The F1 rating is a calculated mean of recall and accuracy. Overall, H-SVR-GPT operates well in user online classification for all datasets, with accuracy varying between 95.01% to 96.57%. Precision, recall, and F1 scores are also high everywhere any set of data, with values that range from 95.11 to 96.77%.

Table.5.3. Performance of the proposed method with the temporal constraints of online datasets

Dataset	Train and Test Ratio	Dimension	Accuracy	Precision	Recall	F-Score
Bitcoin Tweet	90:10	100,200	95.66	96.54	96.45	95.60
	80:20	100,200	95.33	96.65	95.23	95.44
	70:30	100,50	95.56	95.34	96.12	95.52
Facebook	90:10	100,200	95.55	96.32	96.24	95.11
	80:20	100,200	95.12	96.45	96.56	96.26
	70:30	100,50	95.01	96.77	96.44	95.78
Amazon	90:10	100,200	95.68	95.68	95.67	95.66
	80:20	100,200	96.44	95.89	96.02	95.46
	70:30	100,50	96.57	96.77	96.15	95.31

After observing the previous experiments, research and analysis on the three datasets done affirming the quality of the two datasets. The proposed cataloguing typical method outperforms existing classification methods in terms of performance.

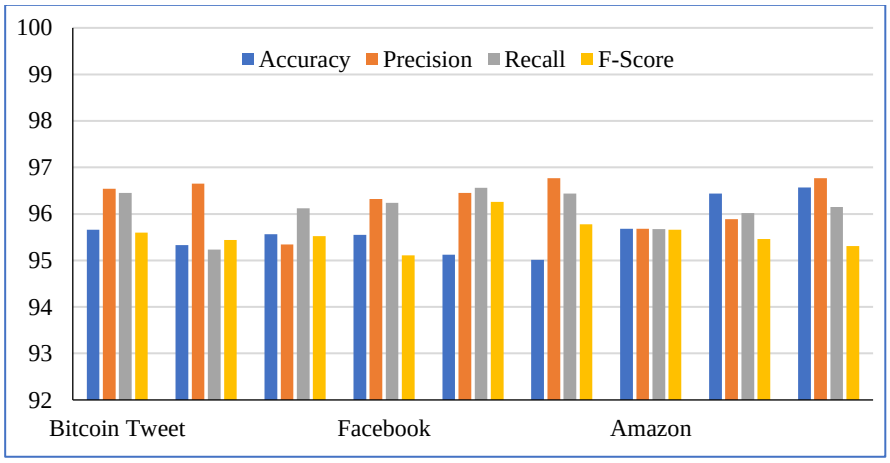

Fig 3.0 Performance of the proposed method with different datasets

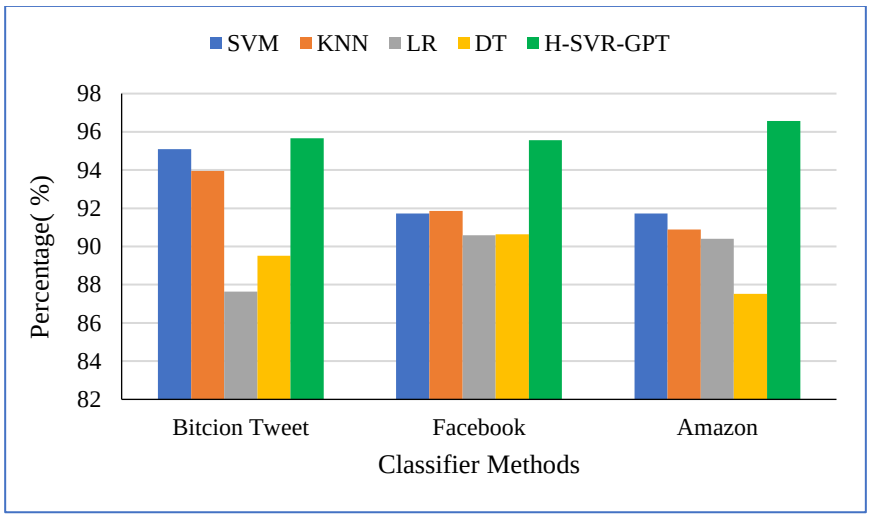

Fig 4.0 Performance Accuracy Comparison of the proposed H-SVR-GPT method

The proposed H-SVR-GPT model has shown promising outcomes in identifying customer engagement on various platforms. This highlights the effectiveness of the suggested methodology which employed a combination of learning-based techniques.

6. Conclusion

H-SVR-GPT, a Hybrid State Vector Machine with Generative Pre-trained Transformers, outperforms more conventional algorithms like SVM, k-Nearest Neighbours, Logistic Regression, and DT when it comes to consumer segmentation in online social media. By fusing SVM's classification capabilities with GPT's abilities to comprehend complex social media data, it guarantees a thorough approach to segmentation. H-SVR-GPT efficiently overcomes these drawbacks, providing accurate, context-aware consumer segmentation for online platforms. While SVM may be highly computational for huge databases, KNN depends on high-dimensions, LR lacks complexity handling, and DT struggles with context-rich data. In future exploring methods to incorporate with the proposed system connect the user feedback and preferences into the segmentation process can further refine and personalize segments.

References

[1] Ahmed H, Traore I, Saad S (2018) Detecting opinion spams and fake news using text classification. Secur Priv 1(1):e9. https://doi.org/10.1002/spy2.9

[2] Alimena, J., Iiyama, Y., and Kieseler, J. (2020). Fast convolutional neural networks for identifying long-lived particles in a high-granularity calorimeter. J. Instrument. 15, P12006–P12006. doi: 10.1088/1748-0221/15/12/P12006.

[3] Alsubari SN, Deshmukh SN, Aldhyani THH, Al Nefaie AH, Alrasheedi M (2023) Rule-based classifiers for identifying fake reviews in e-commerce: a deep learning system. In: Som T (ed) Interdisciplinary mathematics. Springer, Singapore. https://doi.org/10.1007/978-981-19-8566-9_14.

[4] Amin, M. R., & Zuhairi, M. F. (2021). Review of fscm with blockchain and big data integration. *Indian J. Comput. Sci. Eng*, *12*, 193-201.

[5] Arsheed H Sheikh, Kashif Nawaz, Naheed Tabassum, Marilia Almeida-Trapp, Kiruthiga G Mariappan, Hanna Alhoraibi, Naganand Rayapuram, Manuel Aranda, Martin Groth, Heribert Hirt (2023). Linker histone H1 modulates defense priming and immunity in plants, *Nucleic Acids Research*, Volume 51, Issue 9, 22, Pages 4252–4265, https://doi.org/10.1093/nar/gkad106.

[6] Asghar MZ, Ullah A, Ahmad S, Khan A (2019) Opinion spam detection framework using hybrid classification scheme. Soft Comput. https://doi.org/10.1007/s00500-019-04107-y.

[7] Birim ŞÖ, Kazancoglu I, Mangla SK, Kahraman A, Kumar S, Kazancoglu Y (2022) Detecting fake reviews through topic modelling. J Bus Res 149:884–900. https://doi.org/10.1016/j.jbusres.2022.05.081

[8] Budhi GS, Chiong R, Wang Z, Dhakal S (2021) Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. Electron Commer Res Appl 47:101048. https://doi.org/10.1016/j.elerap.2021.101048

[9] Budhi GS, Chiong R, Wang Z, Dhakal S (2021) Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. Electron Commer Res Appl 47:101048. https://doi.org/10.1016/j.elerap.2021.101048

[10] Chatterjee S, Chaudhuri R, Kumar A, Wang CL, Gupta S (2023) Impacts of consumer cognitive process to ascertain online fake review: a cognitive dissonance theory approach. J Bus Res 154:113370. https://doi.org/10.1016/j.jbusres.2022.113370.

[11] Dong Zhang, Wenwen Li, Baozhuang Niu, and Chong Wu. (2023). A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information. Decis. Support Syst. 166, C (Mar 2023). https://doi.org/10.1016/j.dss.2022.113911.

[12] Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2021). Understanding consumer engagement in social media: The role of product lifecycle.Decision Support Systems,113707.

[13] Fernández, P., Hartmann, P., & Apaolaza, V. (2022). What drives CSR communication effectiveness on social media? A process-based theoretical framework and research agenda. International Journal of Advertising, 41(3), 385–413.

[14] Gaber, H. R., & Elsamadicy, A. (2020). The effect of corporate social responsibility content on consumer engagement behaviours on Facebook brand pages in Egypt. Journal of Customer Behaviour, 19(3), 280–297.

[15] Harrigan, P., Daly, T. M., Coussement, K., Lee, J. A., Soutar, G. N., & Evers, U. (2021). Identifying influencers on social media. International Journal of Information Management, 56, 102246.

[16] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., ... & He, B. (2021). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering.

[17] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, *1*(1), 18.

[18] Serrano-Malebrán, J.; Arenas-Gaitán, J. When does personalization work on social media? a posteriori segmentation of consumers. Multimed. Tools Appl. 2021, 80, 36509–36528.

[19] Wang, C., Liu, X., Yue, Y., Tang, X., Zhang, T., Jiayang, C., ... & Zhang, Y. (2023). Survey on Factuality in Large Language Models: Knowledge, Retrieval and Domain-Specificity. *arXiv preprint arXiv:2310.07521*.

[20] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[21] Umbarkar, A.M., Sherie, N.P., Agrawal, S.A., Kharche, P.P., Dhabliya, D. Robust design of optimal location analysis for piezoelectric sensor in a cantilever beam (2021) Materials Today: Proceedings

[22] Sherje, N.P., Agrawal, S.A., Umbarkar, A.M., Kharche, P.P., Dhabliya, D. Machinability study and optimization of CNC drilling process parameters for HSLA steel with coated and uncoated drill bit (2021) Materials Today: Proceedings,