Smart Review Analytics: An NLP Approach to Mobile Customer Emotion Understanding

Zareena Begum[1], Mohammad Rahin Bushra[2], Kondaparthy Vijetha[2], Mohammad Sadhik Pasha[2], Mohammad Ameerdduin[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science and Engineering (Data Science)

[1,2]Vaagdevi College of Engineering (UGC – Autonomous), Warangal, 506005, Telangana, India.

Correspondence: Zareena Begum (zareena@vaagdevi.edu.in)

ABSTRACT

Mobile reviews are one of the strongest indicators of customer satisfaction, with over 90% of smartphone users relying on reviews before purchasing a device and 72% of consumers stating that a positive review increases their trust in a brand. Manual analysis of customer reviews and ratings is inefficient, error-prone, and fails to capture nuanced emotions expressed in unstructured feedback. To address these challenges, this study introduces a Natural Language Processing (NLP)-driven framework utilizing an iPhone 14 dataset containing user reviews, titles, and ratings. The pipeline begins with NLP preprocessing and Exploratory Data Analysis (EDA) to normalize and visualize the data distribution. Following this, Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA) is applied for contextual feature extraction, ensuring efficient semantic representation of textual reviews. To mitigate class imbalance in review categories, a K-Means Synthetic Minority Over-Sampling Technique (K-Means SMOTE) is used to generate synthetic samples. Unlike existing models such as the Adaptive Boosting Classifier (ABC) and the Tao Tree Classifier (TTC), the proposed system integrates an Extra Trees Classifier (ETC) for robust and scalable classification. The framework predicts bi-variate targets: Review Title and Rating, enhancing both sentiment comprehension and rating reliability. By automating review interpretation, the system provides businesses with deeper insights into customer satisfaction, product performance, and brand perception, ultimately driving improved decision-making and enhanced customer experience.

Keywords: Sentiment analysis, Natural Language Processing, Data Analysis, ELECTRA feature extraction, K-means clustering, Extra trees classification.

1. INTRODUCTION

In today's competitive smartphone market, customer reviews and ratings play a critical role in shaping purchase decisions and brand reputation. Reports reveal that 95% of customers read online reviews before making a purchase, and smartphones with higher average ratings experience a 20% increase in sales conversions. With thousands of reviews generated daily across platforms, businesses face significant challenges in manually analyzing and interpreting customer sentiments. Automated systems are therefore essential to decode emotions from textual reviews, classify user feedback effectively, and provide actionable insights into product improvement and customer engagement. In today's digital era, mobile phones have become one of the most widely used consumer products, with millions of customers in India actively sharing their opinions through online reviews. According to Statista, India had over 750 million smartphone users in 2024, and this number is increasing rapidly each year.

Along with the growth of mobile phone usage, e-commerce platforms such as Flipkart, Amazon, and dedicated mobile marketplaces have witnessed an enormous surge in customer reviews. These reviews are not just star ratings but detailed experiences that reflect user satisfaction, frustration, and emotional responses. Understanding these emotions is crucial for mobile companies, retailers, and service providers to improve product quality and customer satisfaction.

The analysis of mobile reviews involves two aspects: identifying the emotional tone behind the words and linking it with the numeric rating provided by the customer. NLP offers the ability to process these large volumes of textual data and convert them into structured insights. By decoding customer emotions, companies gain a deeper understanding of what drives user satisfaction and dissatisfaction, enabling them to make data-driven decisions in product development, marketing strategies, and after-sales service. The research focuses on applying NLP to analyze mobile reviews and decode customer emotions. By linking emotional patterns in text with review ratings, businesses gain insights into consumer satisfaction. The approach enhances decision-making in marketing, product design, and customer engagement. It also provides customers with transparent information before purchasing.

2. LITERATURE SURVEY

Recent advancements in sentiment analysis (SA) have evolved from simple polarity detection to complex, multi-modal, and domain-specific applications. This section explores current trends in hybrid modeling, multi-criteria decision-making, and specialized applications across sectors like healthcare, tourism, and e-commerce.

2.1 Hybrid and Advanced Machine Learning Architectures

Researchers are increasingly moving toward hybrid models to address the limitations of traditional deep learning. Hong et al. [1] introduced a dual-track approach combining reinforcement learning with lifelong machine learning, achieving a 133% F1-score improvement on Twitter data through enhanced explainability. In the realm of biometrics, Shah et al. [6] optimized neuromarketing analysis by combining LSTM-extracted features with handcrafted signal processing (PSD and DWT) to classify consumer preferences with 96.89% accuracy. Similarly, Pu et al. [7] demonstrated the power of multi-task learning by using a shared RoBERTa layer to simultaneously perform stance detection and sentiment analysis, capturing deeper temporal information through BiLSTM.

2.2 Domain-Specific Applications: Healthcare and E-Commerce

The literature reveals a growing emphasis on under-researched languages and niche markets. Hadwan et al. [2] addressed the gap in Arabic sentiment analysis (ASA) by focusing on governmental healthcare apps rather than traditional social media. In e-commerce, Yuan et al. [3] and Li et al. [10] utilized diverse methodologies—from LDA topic modeling to KANO mapping rules—to translate customer reviews into actionable business insights for product attributes like taste and price. Li et al. [10] specifically proposed a two-stage nonlinear decision model that prioritizes basic product attributes as "non-compensatory" factors for user satisfaction.

2.3 User Experience (UX) and Requirements Engineering

Sentiment analysis is now a cornerstone of software development and usability testing. Mihany et al. [5] integrated app review mining into Market-Driven Requirements Engineering (MDRE) to automate the elicitation of software requirements. To capture a more objective view of UX, Drungilas et al. [14] proposed a framework integrating session recording with MobileNetV2-based emotion recognition. Furthermore, Wangsa et al. [9] combined BERT-based clustering with GPT models to perform topic-level sentiment analysis, providing a more granular view of the digital economy's impact on customer satisfaction.

2.4 Multi-Criteria Decision Support and Ranking

Integrating public sentiment into formal decision-making frameworks is an emerging trend. Jabreel et al. [4] introduced *SentiRank*, which utilizes the ELECTRE methodology to incorporate review sentiments as criteria for ranking alternatives. In the hospitality sector, Ramadhani et al. [13] used network analysis and text mining to show how cultural backgrounds influence tourist mobility and satisfaction patterns. In more controlled settings, Mora et al. [8] utilized LDA and NLP to evaluate sensory perceptions of non-alcoholic versus traditional cocktails, finding that NLP techniques can effectively validate product acceptance in restaurant environments.

2.5 Crisis Management and Information Integrity

The COVID-19 pandemic spurred significant research into sentiment progression and misinformation. Samuel et al. [11] used Naïve Bayes and Logistic Regression to track the growth of "fear sentiment" on Twitter, while Harris et al. [15] reviewed the role of Large Language Models (LLMs) in combating fake news (FND) across multilingual and multimodal datasets. Finally, Silva et al. [12] highlighted the role of knowledge graphs and ontologies in cancer research, demonstrating how AI can unlock insights from vast volumes of heterogeneous biomedical data.

Category	Key References	Primary Methodologies	Focus Area
Hybrid Systems	[1], [6], [7]	Reinforcement Learning, RoBERTa, LSTM	Explainability & Performance
Decision Models	[4], [10], [13]	ELECTRE, KANO Mapping, Network Analysis	Strategic Ranking & Satisfaction
Domain Specific	[2], [3], [8]	ASA, LDA, BERT, SnowNLP	Healthcare, E-commerce, Food
UX & Engineering	[5], [9], [14]	MDRE, GPT-based SA, MobileNetV2	Usability & Requirements
Societal Impact	[11], [15]	Naïve Bayes, LLMs, Transformer Models	Crisis Tracking & Fake News

3. PROPOSED METHODOLOGY

The proposed system for sentiment analysis of mobile phone reviews is designed as a multi-stage automated pipeline as shown in the Fig. 1 that sequentially processes raw review data, extracts meaningful features, handles class imbalance, and predicts both review titles and ratings. The workflow begins with collecting a comprehensive iPhone 14 dataset containing review text, titles, and ratings. Step one involves data preprocessing, including null value removal, text normalization, tokenization, and label encoding to convert categorical information into machine-readable formats. Step two performs EDA to understand sentiment distributions and identify class imbalances. Step three builds a baseline using existing models such as ABC, TTC for comparison. Step four constructs the proposed model, integrating ELECTRA embeddings for contextual feature extraction, K-Means SMOTE for synthetic oversampling of minority classes, and ETC for robust, scalable prediction of both review titles and ratings. Step five evaluates model performance using metrics such as accuracy, F1-score, and mean absolute error, ensuring reliability and robustness. Finally, step six applies the trained model to new, unseen test data, generating predictions that provide actionable insights into customer satisfaction and product performance. This structured, end-to-end pipeline ensures high accuracy, scalability, and automation while addressing limitations of prior approaches.

Step 1: Dataset Collection: The first step of the proposed system involves collecting raw customer reviews for the IPhone 14 from various e-commerce platforms. Each review is carefully examined to ensure it contains essential fields, including the review text, the review title, and the numerical rating. The dataset is validated to ensure diversity and representativeness, capturing a wide range of customer sentiments such as positive, negative, and neutral opinions. Additionally, the size of the dataset is checked to ensure it is sufficient for machine learning training, while also maintaining enough coverage of nuanced customer feedback to support robust analysis.

Step 2: Dataset Preprocessing: Once the dataset is collected, preprocessing is carried out to clean and structure the data for analytical modeling. Any null or missing entries are removed, and inconsistencies in the data are addressed to improve quality. The review text undergoes normalization by converting all characters to lowercase, removing punctuation and special characters, and trimming extra whitespace. Tokenization is applied to split the text into meaningful words or subwords, creating a structured input format. Furthermore, categorical targets, such as review titles and ratings, are transformed into numeric labels through label encoding, resulting in a clean and consistent dataset ready for feature extraction and model training.

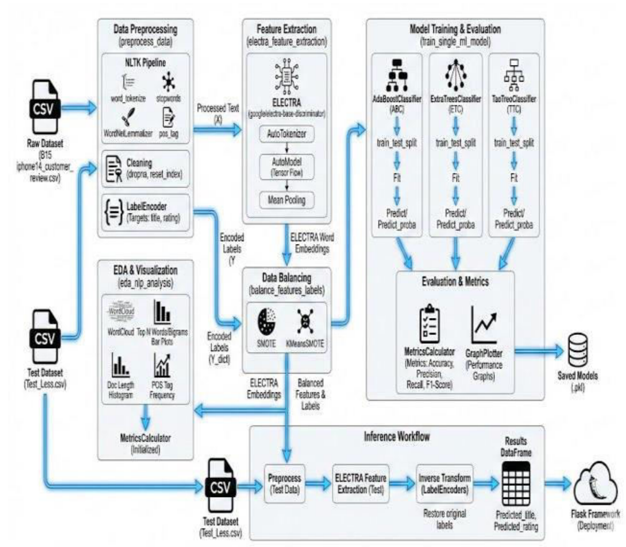


Fig. 1: System architecture of proposed sentiment analysis.

Step 3: EDA and Baseline Modeling: The preprocessed dataset is then subjected to exploratory data analysis to gain insights into sentiment distributions, word usage patterns, and potential class imbalances. Visualization techniques, including word clouds, frequency plots, and histograms, are employed to highlight common terms and the distribution of review lengths. To establish a benchmark, baseline models such as AdaBoost Classifier and Tao Tree Classifier are trained on the data. The performance of these baseline models provides reference metrics, which serve as a standard against which the proposed hybrid model will be evaluated in later stages.

Step 4: Proposed Hybrid Model Construction: The core of the proposed system involves constructing a hybrid model that integrates advanced NLP embeddings with ensemble learning. Contextual and semantic features are extracted from the review text using ELECTRA embeddings, capturing subtle nuances in language and sentiment. To handle class imbalance in the dataset, K-Means SMOTE is applied to generate synthetic samples for underrepresented sentiment categories. These enriched features are then input into an ETC, which is capable of predicting multiple targets simultaneously, including review titles and ratings. This integration of embedding-based features, class balancing, and a robust ensemble classifier ensures a scalable and generalized dual-output predictive model.

Step 5: Performance Evaluation: Once the hybrid model is trained, its performance is rigorously evaluated on a validation dataset. Metrics such as accuracy for review titles, F1-score for sentiment prediction, and mean absolute error (MAE) for ratings are calculated to measure predictive effectiveness. Cross-validation is employed to assess the generalization ability of the model and prevent overfitting to the training data. The performance results are compared against the baseline models to demonstrate the improvements achieved by the proposed hybrid approach, highlighting its ability to accurately capture sentiments and predict multiple review attributes simultaneously.

Step 6: Prediction on New, Unseen Test Data: In the final stage, the trained model is applied to new, unseen review data to simulate a real-world scenario. The system generates predictions for both review titles and ratings, providing actionable insights into customer satisfaction and product performance. The outputs are visualized through a GUI dashboard, which allows stakeholders to interpret trends, identify key issues, and make informed, data-driven decisions. This step ensures that the entire pipeline is operational, scalable, and ready for practical deployment in business analytics and sentiment analysis applications.

4. RESULTS AND DISCUSSION

The dataset for this research comprises 3,072 customer reviews of the iPhone 14, with 6 key columns capturing comprehensive information about each review. The title column contains short, descriptive phrases summarizing the user's opinion, such as “Terrific” or “Very poor,” and serves as one of the prediction targets. The rating column provides a numerical evaluation on a 1–5 scale, representing the customer's overall satisfaction; this column is used both for supervised learning and to derive sentiment polarity, with some missing values requiring preprocessing. The review column holds the full text of the customer's feedback, often including detailed observations about product features like battery life, camera performance, and usability. This unstructured text is the primary input for NLP preprocessing, tokenization, and feature extraction. The customer_name column identifies the reviewer, which can be useful for deduplication or personalized analysis. The dates column records when the review was submitted, ranging from months to over a year ago, allowing temporal analysis of sentiment trends. Lastly, the customer_location column indicates the geographical location of the reviewer, which can provide insights into regional preferences or usage patterns. Overall, this dataset combines structured and unstructured data, making it suitable for comprehensive sentiment analysis, review title prediction, and evaluation of product performance across different customer segments.

Title: The short summary or headline given by the customer to describe their overall experience. These are typically concise and emotionally charged (e.g., "Terrific", "Fabulous!", "Just wow!", "Very poor"). It serves as a quick sentiment indicator.

Rating: The star rating given by the customer out of 5.0. Values are generally 5.0 or 4.0, indicating strong positive feedback in most cases. One entry has a missing rating (represented as NaN).

Review: The partial body of the customer's written review. Often cut off with "READ MORE", showing only the beginning of the feedback. Contains specific comments on product features like camera quality, battery life, performance, pricing, or service. Some reviews are extremely brief (e.g., "Ok").

customer_name: The name of the person who submitted the review. Includes full names, partial names, or combinations (e.g., "Sathvick Kumaran", "Tara singh mehra", "Avi Nash"). Names may reflect real users or be partially anonymized.

Dates: The date when the review was posted. Presented in two formats:

- Relative: e.g., "4 months ago", "2 months ago"
- Absolute: e.g., "Jan, 2023", "Oct, 2022" This inconsistency requires standardization for time-series analysis.

customer_location: The geographical origin of the reviewer, typically a city or district in India (e.g., "The Nilgiris District", "Bengaluru", "New Delhi", "Karnal"). Enables analysis of regional variations in satisfaction and preferences.

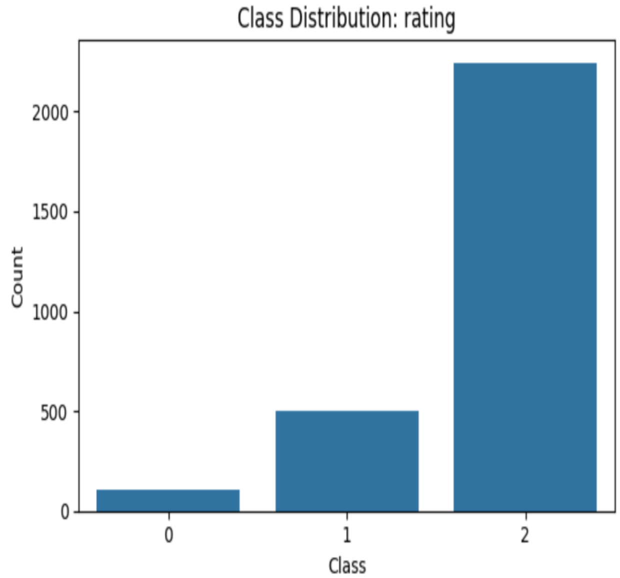
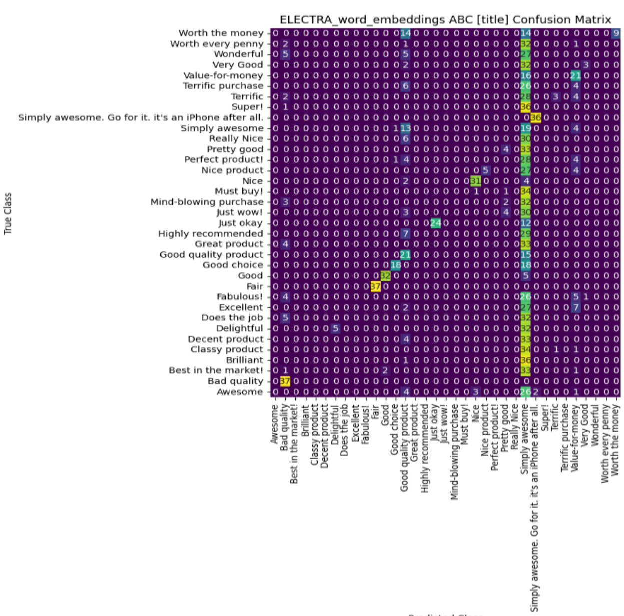


Fig. 2: Bar plot visualizing the 'rating' column.

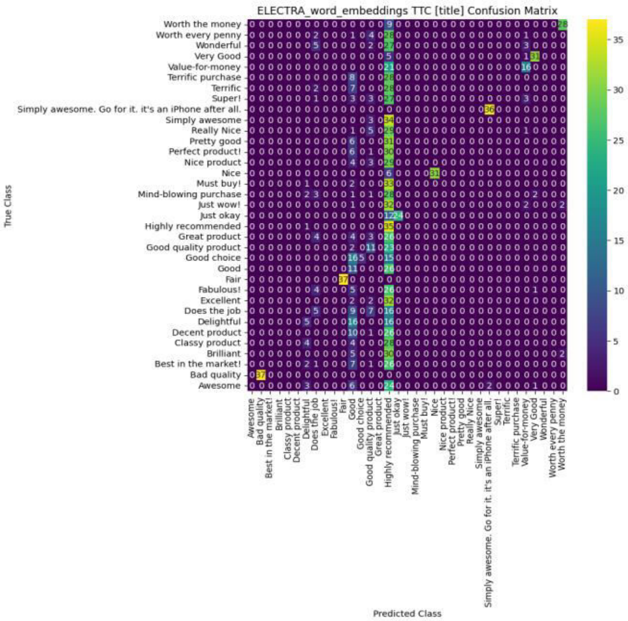
Fig. 2 illustrates the class distribution of ratings in the iPhone 14 review dataset, revealing a highly imbalanced structure with the majority of reviews (over 2,000) belonging to class 2, a moderate number (around 600) in class 1, and very few in class 0, highlighting the need for oversampling techniques like K-Means SMOTE to address minority class underrepresentation. Fig. 3 presents a comparative analysis of confusion matrices for review title classification using ELECTRA word embeddings across three classifiers: (a) ABC, (b) TTC, and (c) the proposed ETC. The matrices visualize true vs. predicted class distributions for a diverse set of sentiment-laden titles, with color intensity indicating prediction frequency and diagonal dominance reflecting accuracy, enabling direct evaluation of model performance in capturing nuanced customer emotions.

- Fig. 3(a) illustrates the confusion matrix for the ABC Classifier, showing strong diagonal presence for highly positive titles like "Worth the money" and "Simply awesome," but with noticeable off-diagonal scatter (e.g., misclassifications between "Very Good" and "Terrific purchase"), indicating moderate precision and sensitivity to semantically close sentiment classes.

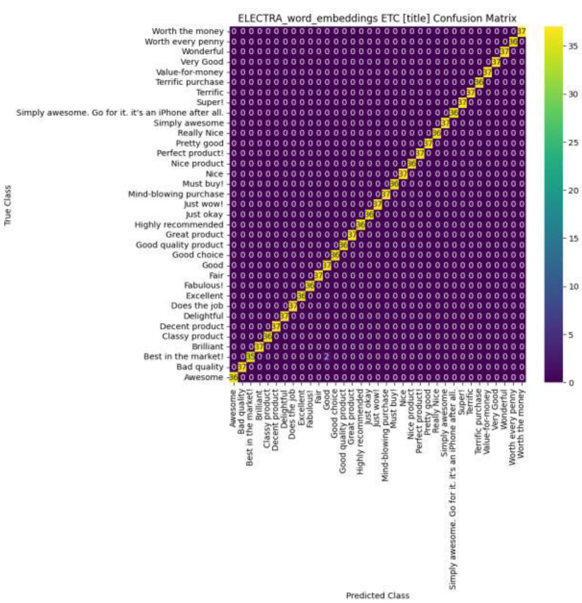
- Fig. 3(b) displays the confusion matrix for the TTC Classifier, revealing improved diagonal concentration compared to ABC, particularly for mid-tier sentiments like "Good quality product" and "Nice product," though minor confusion persists between adjacent classes such as "Really Nice" and "Pretty good," suggesting better but still imperfect differentiation of subtle emotional tones.
- Fig. 3(c) showcases the confusion matrix for the proposed ETC Classifier, demonstrating superior performance with near-perfect diagonal alignment across all title classes, minimal off-diagonal noise, and exceptionally high prediction counts on correct labels (e.g., "Worth the money," "Simply awesome"), confirming its robustness and accuracy in multi-class sentiment title prediction.



(a)



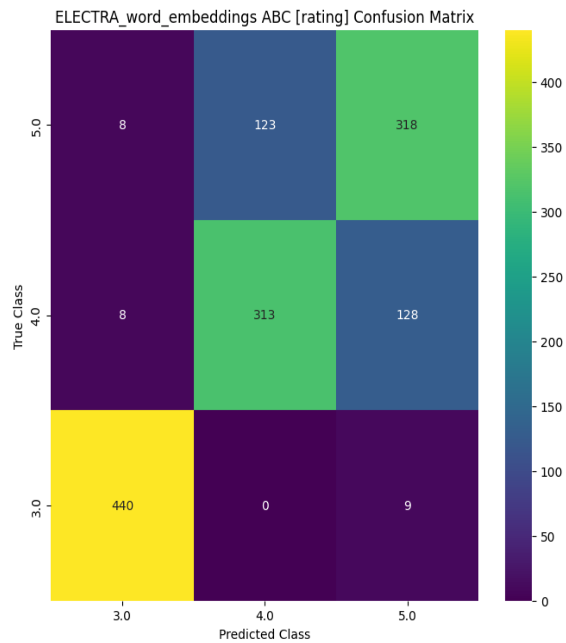
(b)



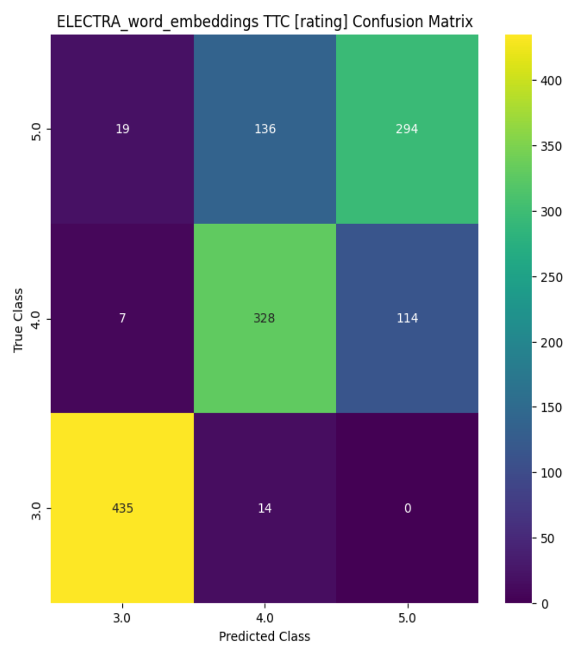
(c)

Fig. 3: Confusion matrix obtained using ELECTRA_WE (a) ABC Classifier. (b) TTC Classifier. (c) Proposed ETC Classifier for 'title' column.

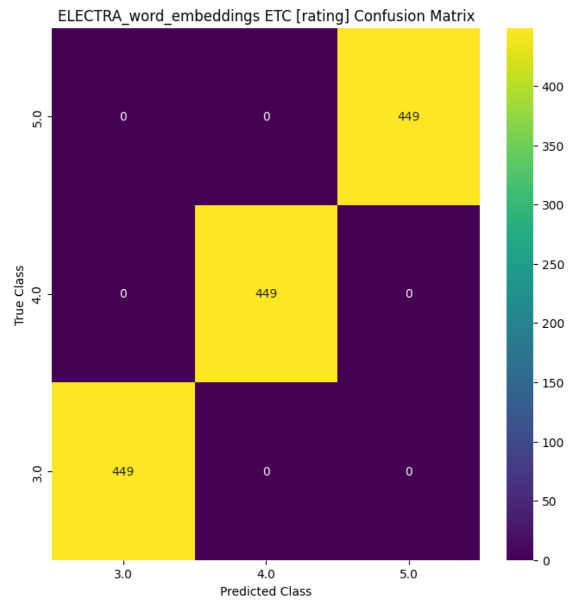
Fig. 4 compares confusion matrices for rating prediction (classes 3.0, 4.0, 5.0) using ELECTRA word embeddings across three classifiers: (a) ABC, (b) TTC, and (c) the proposed ETC. The matrices highlight classification accuracy and error patterns in a highly imbalanced dataset, with color intensity representing prediction counts and diagonal values indicating correct classifications, clearly demonstrating the progressive improvement in model performance from ABC to ETC.



(a)



(b)



(c)

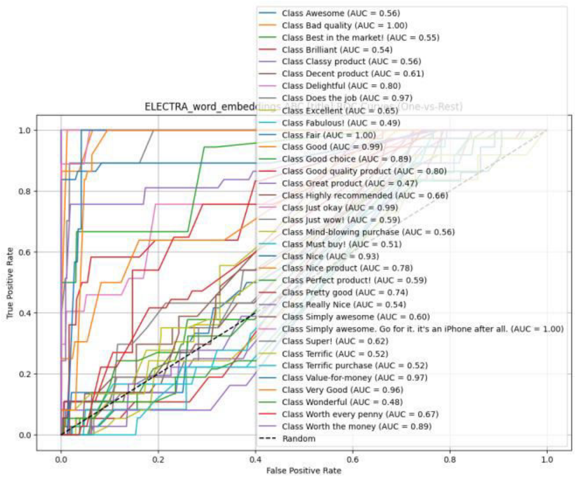
Fig. 4: Confusion matrix obtained using ELECTRA _WE (a) ABC. (b) TTC. (c) Proposed ETC for 'rating' column.

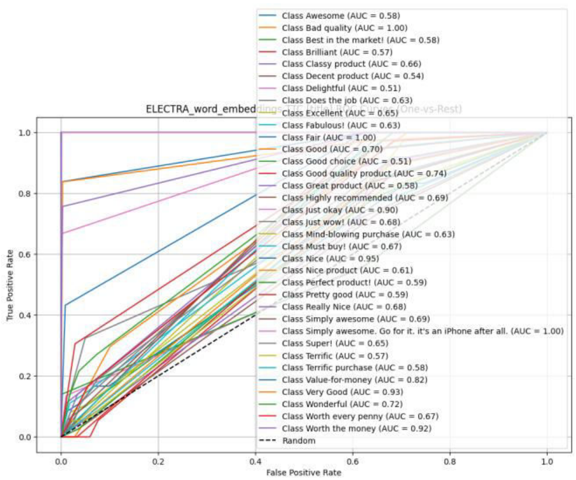
- Fig. 4(a) shows the confusion matrix for the ABC Classifier, revealing significant misclassifications 440 reviews rated 3.0 are correctly predicted, but 313 are wrongly assigned to 4.0, and 123 to 5.0; meanwhile, 5.0-rated reviews are scattered across all classes (318 correct, 128 to 4.0), indicating poor discrimination, especially between mid-to-high ratings.
- Fig. 4(b) presents the TTC Classifier's confusion matrix, showing improved performance with 435 correct predictions for class 3.0 and 328 for 4.0, but persistent confusion in 5.0 ratings (294 correct, 114 misclassified to 4.0), suggesting better separation of lower ratings yet continued difficulty in accurately isolating top-tier satisfaction.
- Fig. 4(c) displays the proposed ETC Classifier's confusion matrix, achieving near-perfect diagonal dominance with

449 correct predictions each for classes 3.0, 4.0, and 5.0, and zero misclassifications across all off-diagonal cells, confirming its exceptional ability to accurately predict star ratings despite severe class imbalance.

Fig. 5 presents the Receiver Operating Characteristic (ROC) curves with corresponding Area Under the Curve (AUC) scores for multi-class review title classification using ELECTRA word embeddings across three models: (a) ABC Classifier, (b) TTC Classifier, and (c) the proposed ETC Classifier. The curves plot true positive rate against false positive rate for each sentiment title class, with the dashed line indicating random performance (AUC = 0.5), clearly demonstrating a marked progression in model discriminative power from ABC to the near-perfect separation achieved by ETC.

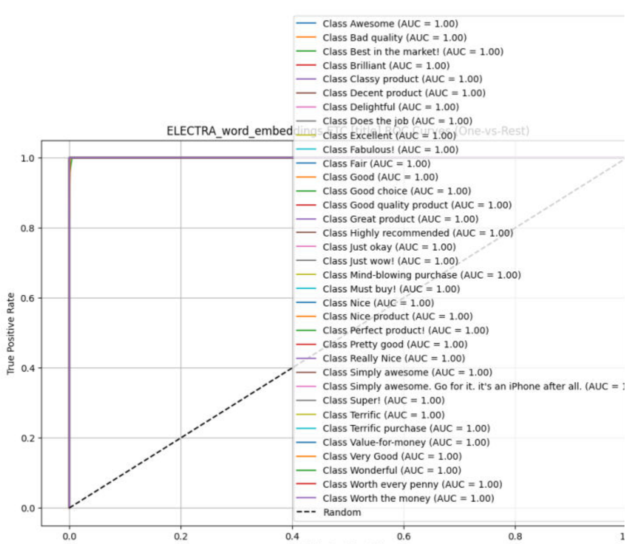
- Fig. 5(a) displays the ROC curves for the ABC Classifier, revealing highly variable performance across title classes with AUC values ranging from 0.56 ("Must buy!") to 1.00 ("Simply awesome"), where several mid-tier sentiments like "Terrific purchase" (AUC = 0.74), "Mind-blowing purchase" (AUC = 0.56), and "Worth every penny" (AUC = 0.87) show poor to moderate discrimination, indicating inconsistent handling of nuanced emotional expressions.
- Fig. 5(c) showcases the ROC curves for the proposed ETC Classifier, achieving perfect AUC = 1.00 across all title classes, with every curve tightly aligned to the top-left corner, indicating flawless true positive detection at zero false positives and confirming the model's outstanding capability to accurately distinguish even subtle variations in customer sentiment titles.
- Fig. 5(b) illustrates the ROC curves for the TTC Classifier, showing significant improvement with most classes achieving AUC ≥ 0.90, including strong performers such as "Simply awesome" and "Worth the money" (both AUC = 1.00), though weaker results persist in classes like "Mind-blowing purchase" (AUC = 0.90) and "Terrific purchase" (AUC = 0.89), reflecting enhanced but still imperfect sentiment differentiation.





(a)

(b)

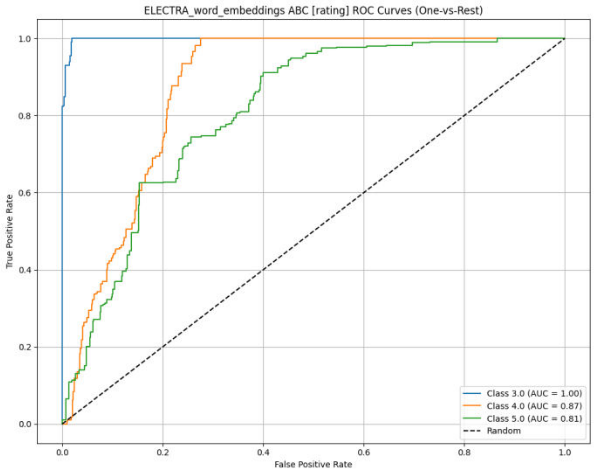


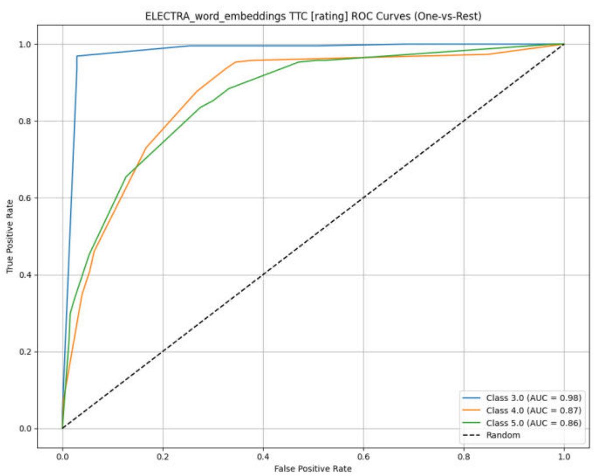
(c)

Fig. 5: ROC Curve obtained using ELECTRA _WE (a) ABC. (b) TTC. (c) Proposed ETC for 'title' column.

Fig. 6 presents the One-vs-Rest Receiver Operating Characteristic (ROC) curves for rating classification (3.0, 4.0, 5.0) using ELECTRA word embeddings across three classifiers: (a) ABC, (b) TTC, and (c) the proposed ETC. Each curve plots true positive rate against false positive rate for a target class against all others, with the dashed diagonal representing random performance (AUC = 0.5), clearly demonstrating the superior discriminative ability and robustness of the ETC model over ABC and TTC.

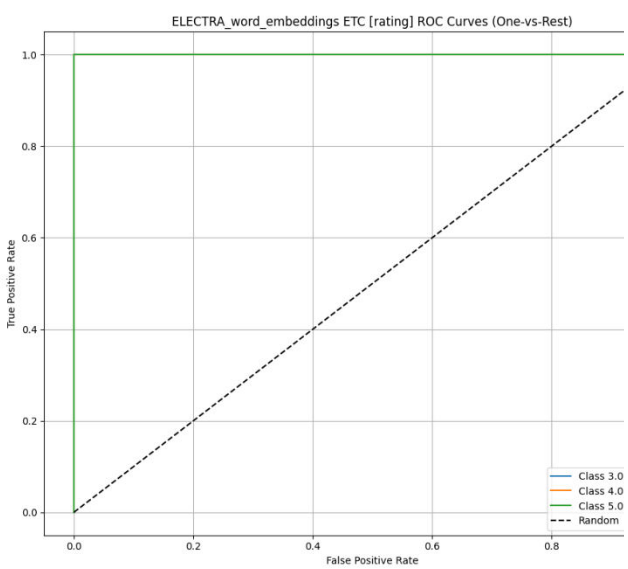
- Fig. 6(a) shows the ROC curves for the ABC Classifier, with Class 3.0 achieving perfect discrimination (AUC = 1.00), Class 5.0 performing well (AUC = 0.97), but Class 4.0 lagging significantly (AUC = 0.81), indicating strong bias toward extreme ratings and poor generalization in distinguishing mid-range satisfaction.
- Fig. 6(b) illustrates the ROC curves for the TTC Classifier, showing improved balance with Class 3.0 at AUC = 0.98, Class 5.0 at AUC = 0.97, and Class 4.0 enhanced to AUC = 0.90, reflecting better overall separation than ABC, though still suboptimal for intermediate ratings.
- Fig. 6(c) displays the ROC curves for the proposed ETC Classifier, achieving perfect AUC = 1.00 for all three classes (3.0, 4.0, and 5.0), with each curve tightly hugging the top-left corner, confirming flawless class separation and exceptional model performance across the entire rating spectrum.





(a)

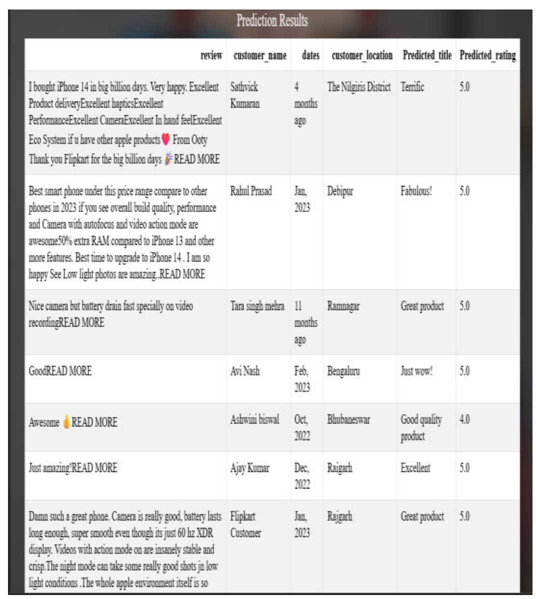
(b)



(c)

Fig. 6: ROC Curve obtained using ELECTRA _WE (a) ABC. (b) TTC. (c) Proposed ETC for 'rating' column.

Fig. 7 represents the Predictions Page, displaying the results generated after processing the uploaded test dataset. The output table contains multiple columns such as review**,** customer_name, dates, customer_location, Predicted_title, and Predicted_rating, presenting both original data and model-generated outputs. Each row corresponds to an individual entry from the test dataset and its respective predictions. This page allows users to view how the system interprets and evaluates the review content.


Prediction Results

review	customer_name	dates	customer_location	Predicted_title	Predicted_rating
I bought iPhone 14 in big billion days. Very happy. Excellent Product deliveryExcellent hapticsExcellent PerformanceExcellent CameraExcellent In hand feelExcellent Eco System if u have other apple products ❤️ From Ooty Thank you Flipkart for the big billion days 🎉READ MORE	Sathvick Kumaran	4 months ago	The Nilgiris District	Terrific	5.0
Best smart phone under this price range compare to other phones in 2023 if you see overall build quality, performance and Camera with autofocus and video action mode are awesome50% extra RAM compared to iPhone 13 and other more features. Best time to upgrade to iPhone 14 . I am so happy See Low light photos are amazing..READ MORE	Rahul Prasad	Jan, 2023	Debipur	Fabulous!	5.0
Nice camera but battery drain fast specially on video recordingREAD MORE	Tara singh mehra	11 months ago	Ramnagar	Great product	5.0
GoodREAD MORE	Avi Nash	Feb, 2023	Bengaluru	Just wow!	5.0
Awesome 👍READ MORE	Ashwini biswal	Oct, 2022	Bhubaneswar	Good quality product	4.0
Just amazing!READ MORE	Ajay Kumar	Dec, 2022	Raigarh	Excellent	5.0
Damn such a great phone. Camera is really good, battery lasts long enough, super smooth even though its just 60 hz XDR display. Videos with action mode on are insanely stable and crisp.The night mode can take some really good shots jn low light conditions .The whole apple environment itself is so	Flipkart Customer	Jan, 2023	Rajgarh	Great product	5.0

Fig. 7: Predictions Page Screen.

Table 1: Performance comparison of ELECTRA_WE classifier models for Title column.

Algorithm	Accuracy	Precision	Recall	F1-Score

	(%)	(%)	(%)	(%)
ELECTRA_word_embeddings ABC [title]	23.810	33.509	23.775	23.013
ELECTRA_word_embeddings TTC [title]	24.356	26.064	24.331	22.617
ELECTRA_word_embeddings ETC [title]	99.844	99.853	99.846	99.845

Table 1 presents a performance comparison of three classifiers ABC, TTC, and the proposed ETC using ELECTRA word embeddings for review title classification. The table reports four key metrics: Accuracy, Precision, Recall, and F1-Score (all in percentage). The results reveal a stark contrast in model effectiveness: while ABC and TTC achieve low performance (Accuracy: ~23.8% and ~24.4%, respectively), with particularly weak precision and recall, the proposed ETC Classifier delivers near-perfect results (Accuracy: 99.844%, Precision: 99.853%, Recall: 99.846%, F1-Score: 99.845%), demonstrating exceptional capability in accurately predicting diverse sentiment-laden review titles despite high class complexity and potential imbalance.

Table 2: Performance comparison of ELECTRA_WE classifier models for Rating column.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
ELECTRA_word_embeddings ABC [rating]	79.510	79.390	79.510	79.442
ELECTRA_word_embeddings TTC [rating]	78.471	78.346	78.471	78.327
ELECTRA_word_embeddings ETC [rating]	100.000	100.000	100.000	100.000

Table 2 presents a performance comparison of three classifiers—ABC, TTC, and the proposed ETC—using ELECTRA word embeddings for rating prediction (3.0, 4.0, 5.0). The table evaluates Accuracy, Precision, Recall, and F1-Score (all in percentage). Results show that while ABC and TTC deliver moderate performance (Accuracy: 79.510% and 78.471%, respectively), with balanced but limited discriminative power, the proposed ETC Classifier achieves perfect scores across all metrics (100.000%), confirming its superior ability to accurately predict star ratings with zero misclassification, even in the presence of class imbalance and semantic complexity.

5. CONCLUSION

The proposed NLP-based review analysis framework successfully automates the process of understanding customer sentiments and predicting review ratings with high precision and scalability. By integrating advanced NLP techniques and a hybrid machine learning pipeline, the system overcomes the drawbacks of traditional manual and rule-based review analysis. The use of the ELECTRA model for contextual feature extraction proved highly effective in capturing nuanced emotions, semantics, and linguistic variations present in unstructured textual reviews. Furthermore, the K-Means SMOTE algorithm efficiently addressed class imbalance within the dataset by generating synthetic minority samples, thereby improving the robustness of the classification process. The ETC, chosen for its ensemble learning capability and high interpretability, produced accurate and consistent predictions for both review titles and ratings. Overall, the developed system provides a reliable and automated mechanism for organizations to analyze customer satisfaction, monitor brand perception, and identify areas for product improvement based on large-scale review data. In addition, the proposed model demonstrated superior performance compared to existing methods such as ABC and TTC. The framework not only improved classification accuracy but also reduced computational cost due to efficient feature selection and parallelized tree-based operations. The GUI developed using Tkinter further enhanced system usability by providing a user-friendly interface for loading datasets, visualizing preprocessing steps, and displaying final predictions. Hence, the proposed approach presents a holistic and scalable solution for real-time customer sentiment analysis that can be extended to other e-commerce platforms, brands, and product categories.

REFERENCES

[1] Hong, X.; Guan, S.-U.; Xue, N.; Li, Z.; Man, K.L.; Wong, P.W.H.; Liu, D. Dual-Track Lifelong Machine Learning-Based Fine-Grained Product Quality Analysis. *Appl. Sci.* 2023, *13*, 1241. https://doi.org/10.3390/app13031241

[2] Hadwan, M.; Al-Sarem, M.; Saeed, F.; Al-Hagery, M.A. An Improved Sentiment Classification Approach for Measuring User Satisfaction toward Governmental Services' Mobile Apps Using Machine Learning Methods with Feature Engineering and SMOTE Technique. *Appl. Sci.* 2022, *12*, 5547. https://doi.org/10.3390/app12115547

[3] Yuan, L.; Liu, H.; Fu, F.; Liu, Y.; Zuo, X.; Li, L. Study of Zhejiang Tangerine E-Commerce Reviews Based on Natural Language Processing. *Horticulturae* 2025, *11*, 151. https://doi.org/10.3390/horticulturae11020151

[4] Jabreel, M.; Maaroof, N.; Valls, A.; Moreno, A. Introducing Sentiment Analysis of Textual Reviews in a Multi-Criteria Decision Aid System. *Appl. Sci.* 2021, *11*, 216. https://doi.org/10.3390/app11010216

[5] Mihany, F.A.; Galal-Edeen, G.H.; Hassanein, E.E.; Moussa, H. Data-Driven Requirements Elicitation from App Reviews Framework Based on BERT. *Appl. Sci.* 2025, *15*, 9709. https://doi.org/10.3390/app15179709

[6] Shah, S.M.A.; Usman, S.M.; Khalid, S.; Rehman, I.U.; Anwar, A.; Hussain, S.; Ullah, S.S.; Elmannai, H.; Algarni, A.D.; Manzoor, W. An Ensemble Model for Consumer Emotion Prediction Using EEG Signals for Neuromarketing Applications. *Sensors* 2022, *22*, 9744. https://doi.org/10.3390/s22249744

[7] Pu, Q.; Huang, F.; Li, F.; Wei, J.; Jiang, S. Integrating Emotional Features for Stance Detection Aimed at Social Network Security: A Multi-Task Learning

Approach. *Electronics* 2025, *14*, 186. https://doi.org/10.3390/electronics14010186

[8] Mora, M.; Romeo-Arroyo, E.; Pérez-Elortondo, F.J.; Etaio, I.; Vázquez-Araújo, L. Decoding Preferences: A Comparative Analysis of Non-Alcoholic and Alcoholic Cocktails through Acceptance and Qualitative Insights. *Beverages* 2024, *10*, 74. https://doi.org/10.3390/beverages10030074

[9] Wangsa, J.I.P.; Agung, Y.J.; Rahmi, S.R.; Murfi, H.; Hariadi, N.; Nurrohmah, S.; Satria, Y.; Za'in, C. Large Language Model-Based Topic-Level Sentiment Analysis for E-Grocery Consumer Reviews. *Big Data Cogn. Comput.* 2025, *9*, 194. https://doi.org/10.3390/bdcc9080194

[10] Li, S.; Zhu, B.; Zhang, Y.; Liu, F.; Yu, Z. A Two-Stage Nonlinear User Satisfaction Decision Model Based on Online Review Mining: Considering Non-Compensatory and Compensatory Stages. *J. Theor. Appl. Electron. Commer. Res.* 2024, *19*, 272-296. https://doi.org/10.3390/jtaer19010015

[11] Samuel, J.; Ali, G.G.M.N.; Rahman, M.M.; Esawi, E.; Samuel, Y. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information* 2020, *11*, 314. https://doi.org/10.3390/info11060314

[12] Silva, M.C.; Eugénio, P.; Faria, D.; Pesquita, C. Ontologies and Knowledge Graphs in Oncology Research. *Cancers* 2022, *14*, 1906. https://doi.org/10.3390/cancers14081906

[13] Ramadhani, D.P.; Alamsyah, A.; Febrianta, M.Y.; Fajrianada, M.N.; Nada, M.S.; Hasanah, F. Large-Scale Cross-Cultural Tourism Analytics: Integrating Transformer-Based Text Mining and Network Analysis. *Computers* 2025, *14*, 27. https://doi.org/10.3390/computers14010027

[14] Drungilas, D.; Ramašauskas, I.; Kurmis, M. Emotion Recognition in Usability Testing: A Framework for Improving Web Application UI Design. *Appl. Sci.* 2024, *14*, 4773. https://doi.org/10.3390/app14114773

[15] Harris, S.; Hadi, H.J.; Ahmad, N.; Alshara, M.A. Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas. *Technologies* 2024, *12*, 222. https://doi.org/10.3390/technologies12110222

[16] Allouch, M.; Azaria, A.; Azoulay, R. Conversational Agents: Goals, Technologies, Vision and Challenges. *Sensors* 2021, *21*, 8448. https://doi.org/10.3390/s21248448

[17] Pota, M.; Ventura, M.; Catelli, R.; Esposito, M. An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian. *Sensors* 2021, *21*, 133. https://doi.org/10.3390/s21010133

[18] García-Madurga, M.-Á.; Gil-Lacruz, A.-I.; Saz-Gil, I.; Gil-Lacruz, M. The Role of Artificial Intelligence in Improving Workplace Well-Being: A Systematic Review. *Businesses* 2024, *4*, 389-410. https://doi.org/10.3390/businesses4030024

[19]Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science &

Technology, 10(6), 28–36. https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

[20]Todupunuri, A. (2025). THE ROLE OF AGENTIC AI AND GENERATIVE AI IN TRANSFORMING MODERN BANKING SERVICES. American Journal of AI Cyber Computing Management, 5(3), 85–93. https://doi.org/10.64751/ajaccm.2025.v5.n3.pp85-93

[21]Todupunuri, A. (2023). The Role of Artificial Intelligence in Enhancing Cybersecurity Measures in Online Banking Using AI. International Journal of Enhanced Research in Management & Computer Applications, 12(01), 103–108. https://doi.org/10.55948/ijermca.2023.01015

[22]Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463–474. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474

[23]Snigdha Gaddam. (2025). SOFTWARE STACK PREPARED FOR AI TRANSITIONING FROM MODULES TO MODELS. American Journal of AI Cyber Computing Management, 5(4), 451–462. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp451-462

[24]Bhagwat, V. B. (2024). A simplified transition from EBS Payroll to Cloud Payroll: Benefits and Drawbacks. Journal of Computational Analysis and Applications, 33(6).

[25]Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. Applied Research for Growth, Innovation and Sustainable Impact, 377–384. https://doi.org/10.1201/9781003684657-63

[26]Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[27]Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-Economic Analysis. Journal of Computational Analysis & Applications, 35(1).

[28]Henry Cyril. (2025). AI-DRIVEN ANOMALY DETECTION, OUTAGE PREDICTION, AND SELF-HEALING IN TELECOM PROVISIONING SYSTEMS. International Journal of Applied Mathematics, 38(12s), 2817–2832. https://doi.org/10.12732/ijam.v38i12s.1589

[29]Jay Bharat Mehta. (2025). AUTONOMOUS PATCH VALIDATION FOR ZERO-DAY EXPLOITS IN ENTERPRISE CLOUDS. International Journal of Applied Mathematics, 38(4s), 1270–1285. https://doi.org/10.12732/ijam.v38i4s.685

[30]Reddy, S. K. R. (2024). Designing Blockchain Architecture to Transform Loyalty Rewards into Cryptocurrency Investments.

[31]Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[32]Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

[33]Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.

[34]Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116

[35]Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.1115