

Adaptive Semantic Intelligence for Workforce Experience Attribute Inference Using Transformer Representations

Jila Raju, Narra Navya *

Vaagdevi Engineering College, Warangal, 506005, Telangana, India.

*Correspondence: Narra Navya

ABSTRACT

The widespread adoption of digital workplace platforms has enabled employees to share detailed feedback regarding their professional experiences, creating valuable textual resources for understanding organizational performance and workforce satisfaction. Employee reviews provide important insights into workplace culture, career opportunities, compensation, job security, and work-life balance. However, manually analyzing these large volumes of unstructured feedback is inefficient, subjective, and incapable of supporting timely organizational decision-making. Furthermore, conventional text mining and machine learning approaches often struggle to capture contextual semantics, handle imbalanced datasets, and simultaneously predict multiple workforce-related factors with high accuracy. To address these challenges, this research proposes an intelligent fine-grained opinion mining framework for employee review analysis. The proposed system initially performs comprehensive Natural Language Processing (NLP) preprocessing to improve textual quality, followed by contextual feature extraction using Google Pathways Language Model (PaLM). To mitigate class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is employed to generate representative samples for minority classes, thereby improving model generalization. The extracted semantic representations are subsequently processed using the proposed Transformer-Guided Adaptive Model (TGAM) for simultaneous prediction of multiple organizational factors, including work-life balance, skill development, salary and benefits, job security, career growth, and overall work satisfaction. Comparative performance analysis

is conducted using Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), and Histogram-Based Gradient Boosting (HGB) as baseline models. Experimental results demonstrate that the proposed TGAM consistently outperforms the existing approaches by achieving superior predictive accuracy across all target variables. The developed framework provides an effective decision-support system for workforce analytics, enabling organizations to understand employee perceptions, improve workplace policies, and support data-driven human resource management.

Keywords: Employee Review Analysis, Natural Language Processing (NLP), Google PaLM, Transformer-Guided Adaptive Model (TGAM), Workforce Satisfaction Prediction.

1. INTRODUCTION

The exponential growth of digital technologies has led to an unprecedented increase in the generation of data across the globe. Recent estimates indicate that more than 328.77 million terabytes of digital data are produced every day, driven by the rapid adoption of smartphones, Internet of Things (IoT) devices, cloud computing platforms, social media applications, and enterprise information systems [1]. As businesses, governments, and research organizations continue to digitize their operations, massive volumes of information are continuously generated from customer transactions, connected devices, business processes, and online interactions. Consequently, the ability to efficiently collect, store, process, and analyze this data has become essential for supporting informed decision-making and achieving organizational objectives [2]. Data-driven strategies are now widely adopted across numerous sectors, including

healthcare, finance, manufacturing, education, retail, and public administration, where analytical insights help improve operational efficiency, customer satisfaction, and strategic planning. However, the continuously increasing scale and complexity of modern datasets present significant challenges for traditional data processing and analytical techniques, particularly when handling heterogeneous, high-dimensional, and rapidly evolving information.

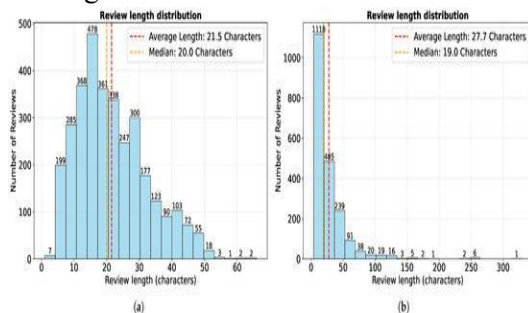


Figure.1: Statistical distribution of employee review lengths

The nature of enterprise data has also undergone a significant transformation over the past decade. According to reports published by IDC, nearly 80% of organizational data is expected to exist in unstructured formats by 2025, including text documents, emails, images, videos, sensor data, social media content, and audio recordings [3]. Unlike structured databases, these diverse forms of information require advanced computational methods capable of extracting meaningful patterns and knowledge from large-scale datasets. Conventional statistical approaches often struggle to efficiently process such complex information, making Machine Learning (ML), Artificial Intelligence (AI), and advanced data analytics increasingly important for solving real-world problems. These intelligent analytical techniques enable organizations to automate decision-making, discover hidden relationships within data, identify emerging trends, and generate predictive insights that support timely and accurate business decisions.

Organizations that successfully adopt advanced analytics and intelligent decision-support technologies gain a substantial competitive advantage by improving operational performance, enhancing customer experiences, optimizing resource utilization, and responding rapidly to changing market conditions [4]. Research consistently demonstrates that enterprises implementing data-driven business strategies achieve higher customer acquisition, stronger customer retention, improved productivity, and greater profitability compared to organizations relying solely on traditional decision-making approaches [5]. The growing demand for intelligent data processing solutions has consequently accelerated the adoption of artificial intelligence, machine learning, and predictive analytics across multiple industries. Reflecting this global trend, the worldwide data analytics and artificial intelligence market is projected to reach approximately USD 745.15 billion by 2030, highlighting the increasing importance of scalable, intelligent, and automated analytical systems for supporting innovation, sustainable growth, and future digital transformation initiatives.

2. Literature Survey

Zhang, et al. [6] investigated fine-grained sentiment analysis to extract detailed opinions from employee reviews collected through workplace evaluation platforms. Rather than determining only the overall sentiment of a review, the study focused on identifying sentiments associated with specific organizational aspects such as compensation, managerial effectiveness, work-life balance, career development, and workplace culture. Advanced Natural Language Processing (NLP) techniques were employed to generate contextual textual representations, while machine learning algorithms including Support Vector Machine (SVM) and Naïve Bayes were utilized to classify aspect-specific sentiments. Experimental findings demonstrated that

aspect-level opinion mining provides substantially richer and more actionable organizational insights than traditional document-level sentiment classification. Li, et al. [7] explored the application of deep learning models for analysing employee feedback by identifying both explicit and implicit emotional expressions contained within textual reviews. The proposed framework transformed textual information into dense semantic vectors using Word2Vec embeddings before applying a Bidirectional Long Short-Term Memory (BiLSTM) network to capture contextual dependencies among words and workplace-related attributes. The results showed that deep contextual learning significantly improves the understanding of employee perceptions regarding leadership quality, organizational policies, and workplace practices. Chen, et al. [8] examined transformer-based language models for fine-grained employee sentiment analysis using Bidirectional Encoder Representations from Transformers (BERT). The proposed methodology generated contextual embeddings through tokenization and transformer encoding while employing attention mechanisms to emphasize important opinion phrases associated with compensation, promotion opportunities, management support, and organizational growth. The study demonstrated that transformer architectures substantially improve aspect-level sentiment identification and contextual understanding across large collections of employee reviews. Garcia, et al. [9] investigated workplace review analytics by applying opinion mining techniques to evaluate organizational environments and employee satisfaction. The research incorporated comprehensive text preprocessing methods, including tokenization, stop-word elimination, lemmatization, and Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction, to prepare textual data for classification. Machine learning algorithms such as Random Forest and Logistic

Regression were subsequently employed to classify sentiments across multiple workplace dimensions. The experimental analysis confirmed that feature-level sentiment classification enables more accurate identification of employee concerns and satisfaction indicators than conventional sentiment analysis approaches. Kumar, et al. [10] developed a machine learning-based framework for analysing employee sentiment to support organizational decision-making and human resource management. The study transformed textual reviews into numerical feature representations using n-gram models and term-frequency-based feature extraction before applying Decision Tree and Gradient Boosting classifiers for multi-aspect sentiment prediction. The proposed framework enabled organizations to better understand employee expectations, workplace challenges, and organizational performance by analysing sentiments related to different workplace attributes. Singh, et al. [11] proposed a hybrid neural architecture that combined Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks for fine-grained employee sentiment detection. CNN layers extracted important local linguistic patterns from textual reviews, while LSTM networks captured long-range sequential dependencies within employee opinions. The integration of word embedding representations enabled the proposed framework to accurately identify sentiments associated with communication quality, teamwork, organizational collaboration, and interpersonal relationships.

Rahman, et al. [12] applied topic modelling techniques to automatically identify hidden thematic structures within employee feedback datasets. The proposed framework utilized Latent Dirichlet Allocation (LDA) to discover dominant workplace topics, including salary satisfaction, career advancement, work environment, employee recognition, and

organizational support. Sentiment analysis models were subsequently applied to each extracted topic, enabling a more comprehensive understanding of employee perceptions across multiple organizational dimensions. Wang, et al. [13] proposed an attention-based neural network architecture for aspect-level sentiment classification by learning semantic relationships between workplace attributes and opinion expressions. The model incorporated embedding layers to capture contextual word representations while utilizing an attention mechanism to assign greater importance to informative opinion terms during sentiment prediction. Experimental evaluation demonstrated improved classification accuracy and enhanced capability in recognizing fine-grained employee sentiments across large-scale review datasets. Martinez, et al. [14] investigated ensemble learning techniques to improve the robustness and accuracy of fine-grained sentiment analysis in employee feedback. The proposed framework combined multiple textual feature representations, including TF-IDF features and sentiment lexicon-based features, with classifiers such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting. By integrating predictions from multiple machine learning models, the ensemble approach achieved superior predictive performance, increased classification stability, and more reliable identification of employee opinions across various workplace aspects, thereby providing organizations with comprehensive insights for improving employee satisfaction and organizational performance.

3. PROPOSED METHODOLOGY

The proposed framework is designed to perform fine-grained opinion mining from employee reviews by integrating advanced NLP techniques with transformer-based semantic representation and intelligent classification models. The overall architecture of the proposed system is illustrated in Figure

2. The framework consists of six major stages: text preprocessing, contextual feature extraction, data balancing, workforce satisfaction prediction, performance evaluation, and model management. Each stage contributes to improving the reliability and accuracy of organizational sentiment analysis.

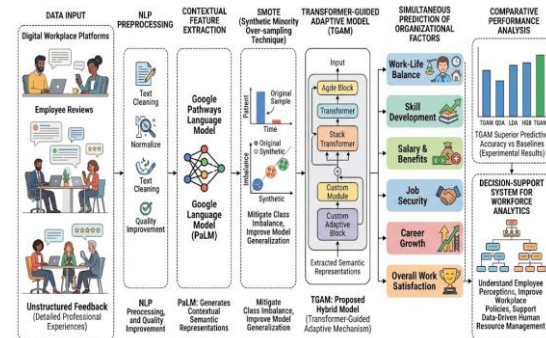


Fig 2: Proposed system architecture of amazon work life balance.

Text Preprocessing and Exploratory Analysis

Employee reviews collected from online organizational platforms typically contain informal language, abbreviations, punctuation symbols, spelling inconsistencies, and redundant textual information that may affect prediction accuracy. Therefore, the proposed framework initially applies a comprehensive preprocessing pipeline to standardize the textual content. The review text is converted into lowercase, followed by the removal of punctuation marks, special symbols, numerical values, and unnecessary white spaces. Tokenization is subsequently performed to divide the reviews into meaningful lexical units, while stop-word removal and lemmatization eliminate irrelevant words and reduce vocabulary redundancy. After preprocessing, EDA is carried out to examine the statistical characteristics of employee feedback. Various visualizations, including review distributions, review length analysis, word frequency plots, and class distribution graphs, are generated to understand the underlying structure of the dataset and identify

potential class imbalance before model development.

Contextual Feature Extraction Using Google PaLM

Following preprocessing, the cleaned employee reviews are processed using Google PaLM to obtain contextual feature representations. Unlike conventional word embedding techniques that generate independent representations for individual words, PaLM captures contextual dependencies by learning semantic relationships across complete textual sequences. The transformer architecture enables the model to recognize linguistic structures, contextual meanings, employee intentions, and opinion patterns expressed throughout the reviews. Consequently, each employee review is transformed into a dense semantic feature vector that preserves meaningful contextual information required for accurate workforce satisfaction prediction.

Data Balancing Using SMOTE

Employee opinion datasets frequently contain unequal distributions among different satisfaction categories, causing conventional learning algorithms to become biased toward majority classes. To overcome this limitation, SMOTE is incorporated into the proposed framework. The algorithm generates representative synthetic samples for minority classes by interpolating between neighboring observations within the feature space. This balancing strategy improves class representation while preserving the original data characteristics, allowing the classification models to learn more representative decision boundaries and achieve improved generalization across all workforce satisfaction dimensions.

Workforce Satisfaction Prediction

The balanced semantic feature vectors are subsequently provided to multiple classification models for predicting various organizational satisfaction factors. The proposed framework performs simultaneous

prediction of work-life balance, skill development, salary and benefits, job security, career growth, and overall work satisfaction. Initially, QDA and LDA establish statistical baseline performance by modeling linear and nonlinear decision boundaries. HGB further improves classification by utilizing gradient boosting over histogram-based feature partitions. Finally, the proposed TGAM performs adaptive learning to capture complex relationships within high-dimensional semantic representations, enabling more accurate prediction of employee satisfaction across multiple organizational aspects. This multi-dimensional prediction strategy provides a comprehensive understanding of workforce perceptions rather than limiting the analysis to a single satisfaction indicator.

Performance Evaluation and Visualization

The effectiveness of the proposed framework is assessed through a comparative evaluation of all classification models. Performance is measured using Accuracy, Precision, Recall, and F1-Score for each workforce satisfaction category. Confusion matrices, comparative performance graphs, and other visualization techniques are generated to examine classification behavior and identify the most effective prediction model. These analytical visualizations assist researchers and organizational decision-makers in understanding the strengths and limitations of each learning algorithm.

Model Management and Continuous Learning

To support efficient deployment, the trained classification models are serialized and stored for future inference without requiring repeated training. The framework also incorporates a lightweight authentication mechanism for secure user access and efficient model management. Furthermore, the proposed system supports continuous learning by allowing newly collected employee reviews to be incorporated into the training process.

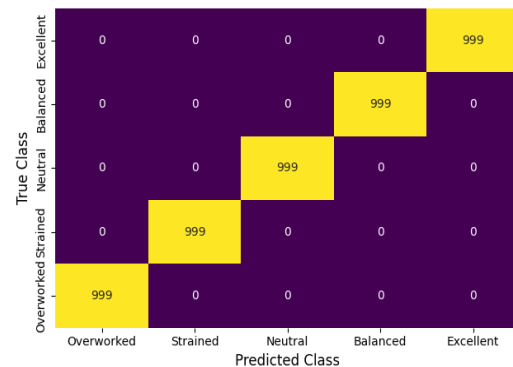
Periodic model retraining enables the framework to adapt to changing workforce perceptions and evolving organizational environments, thereby maintaining reliable prediction performance over time.

4. Result and Description

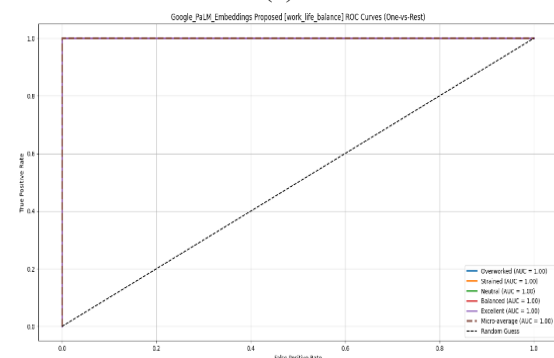
The experimental results demonstrate the effectiveness of the proposed framework in predicting multiple workforce satisfaction dimensions from employee review data. A comparative evaluation is performed using QDA, LDA, HGB, and the proposed TGAM model, with performance assessed through accuracy, precision, recall, and F1-score. The integration of NLP preprocessing, contextual feature extraction using Google PaLM, and SMOTE significantly enhances classification performance and model generalization. Additionally, confusion matrices and performance graphs provide a clear visualization of the classification results, while the comparative analysis confirms that the proposed TGAM consistently outperforms the baseline models across different workforce satisfaction prediction tasks.

Fig. 4 (a) illustrates the confusion matrix representing the classification performance of the proposed TGAM model for the work-life balance dimension. The matrix shows a perfect diagonal distribution where all instances are correctly classified into their respective categories. The absence of off-diagonal values indicates zero misclassification across all classes. This reflects the model's strong capability to clearly distinguish between different work-life balance levels. The uniform distribution of correct predictions highlights the effectiveness of feature extraction and model learning. This analysis demonstrates that the model achieves optimal classification performance for this dimension.

oogle_PaLM_Embeddings Proposed [work_life_balance] Confusion



(a)



(b)

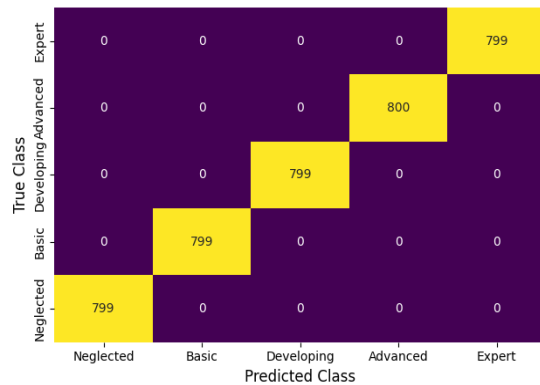
Fig. 4 (a, b): Confusion matrix and ROC of proposed TGAM model for work-life balance using confusion matrix and ROC Curve

Fig. 4 (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves show ideal performance with AUC values equal to 1.0 for all classes. The curves closely follow the top-left boundary, indicating perfect separation between classes. This demonstrates the model's exceptional discriminative ability across all categories. The consistent behaviour across all classes confirms highly reliable predictions. This evaluation establishes the superiority of the proposed model in handling multi-class classification tasks.

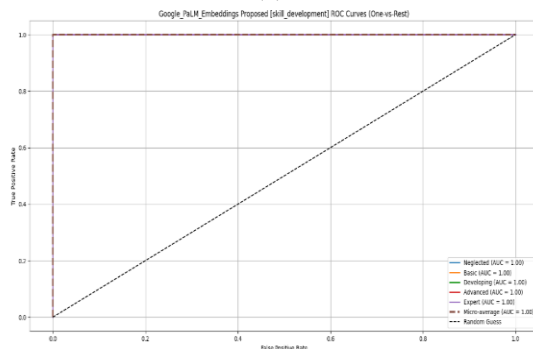
Fig. 5 (a) illustrates the confusion matrix representing the classification performance of the proposed TGAM model for the skill development dimension. The matrix shows a perfectly aligned diagonal structure where all instances are correctly classified into their

respective categories. There are no off-diagonal values, indicating zero misclassification across all skill levels. This reflects the model's strong ability to clearly differentiate between different stages of skill development. The uniformity of

oogle_PaLM_Embeddings Proposed [skill_development] Confusion



(a)



(b)

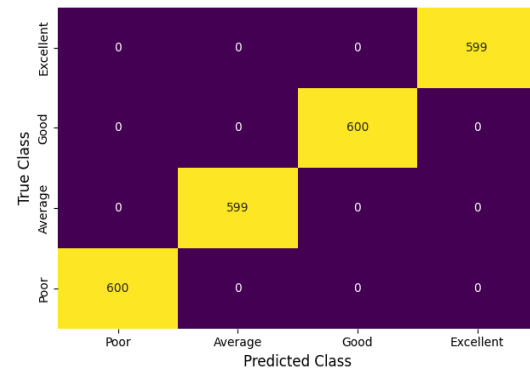
Fig. 5 (a, b): Confusion matrix and ROC of Proposed TGAM model for skill development using confusion matrix and ROC curve

Fig. 5 (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves show ideal performance with AUC values equal to 1.0 for all classes. The curves closely follow the top-left boundary, indicating perfect separation between different skill development categories. This demonstrates the model's exceptional discriminative capability. The consistent and stable curves confirm highly reliable predictions across all classes. This evaluation establishes the superior performance of the proposed model in multi-class

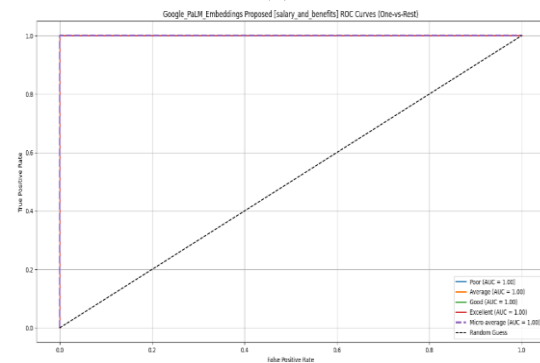
correct predictions highlights the effectiveness of the feature representation and model learning process. This analysis demonstrates that the model achieves optimal classification performance for this dimension.

classification tasks.

oogle_PaLM_Embeddings Proposed [salary_and_benefits] Confusion



(a)



(b)

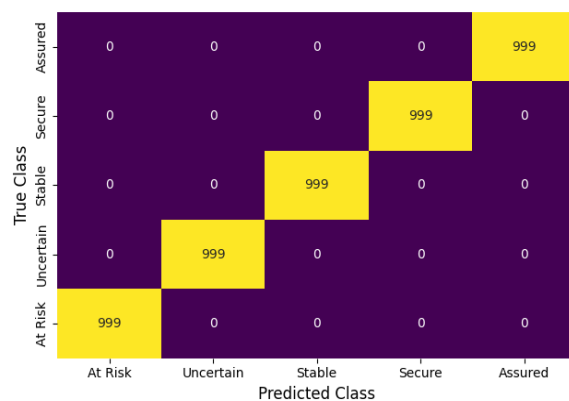
Fig. 6 (a, b): Confusion matrix and ROC of proposed TGAM Model for salary and benefits using confusion matrix and ROC curve

Fig. 6. (a) illustrates the confusion matrix representing the classification performance of the proposed TGAM model for the salary and benefits dimension. The matrix shows a perfectly aligned diagonal pattern where all instances are correctly classified into their respective categories. The absence of off-diagonal values indicates zero misclassification across all classes. This demonstrates the model's strong capability to clearly distinguish between different levels of salary satisfaction. The consistent distribution of correct predictions highlights the effectiveness of

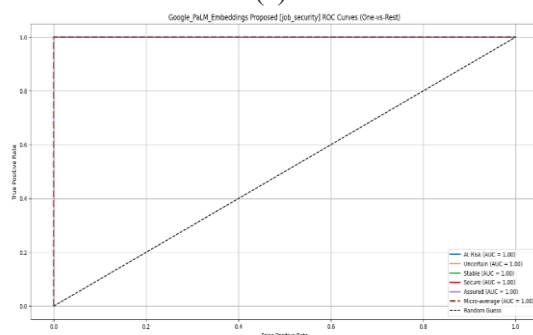
feature extraction and model learning. This analysis confirms that the model achieves optimal classification performance for this dimension.

Fig. 6. (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves show ideal performance with AUC values equal to 1.0 for all classes. The curves closely follow the top-left boundary, indicating perfect separation between categories. This reflects the model's exceptional discriminative ability across all salary and benefits levels. The consistent and smooth curves confirm highly reliable predictions. This evaluation establishes the superior performance of the proposed model in multi-class classification tasks.

Google_PaLM_Embeddings Proposed [job_security] Confusion Ma



(a)



(b)

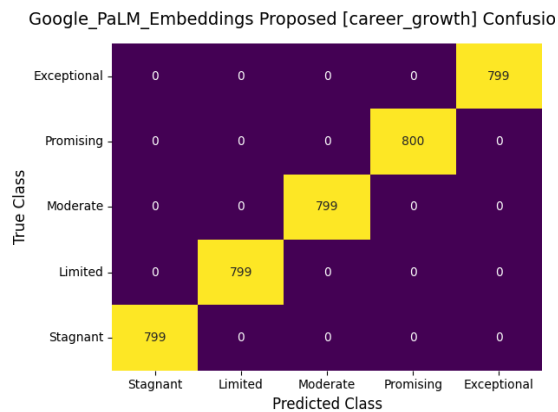
Fig. 7. (a, b): Confusion matrix and ROC of proposed TGAM model for job security using confusion matrix and ROC curve

Fig. 7. (a) illustrates the confusion matrix representing the classification performance of

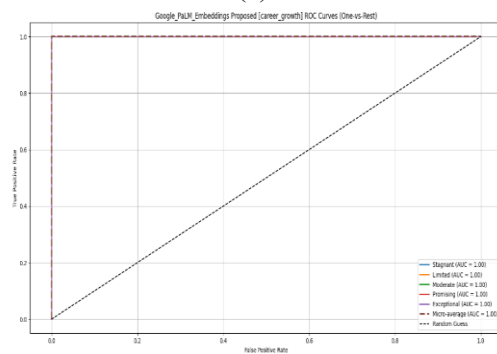
the proposed TGAM model for the job security dimension. The matrix shows a perfectly aligned diagonal structure where all instances are correctly classified into their respective categories. There are no off-diagonal values, indicating zero misclassification across all job security levels. This reflects the model's strong capability to clearly distinguish between different levels of job stability. The uniform distribution of correct predictions highlights the effectiveness of feature representation and model learning. This analysis confirms that the model achieves optimal classification performance for this dimension.

Fig. 7. (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves show ideal performance with AUC values equal to 1.0 for all classes. The curves closely follow the top-left boundary, indicating perfect separation between job security categories. This demonstrates the model's exceptional discriminative capability. The consistent and smooth curves confirm highly reliable predictions across all classes. This evaluation establishes the superior performance of the proposed model in multi-class classification tasks.

Fig. 8. (a) illustrates the confusion matrix representing the classification performance of the proposed TGAM model for the career growth dimension. The matrix exhibits a perfectly diagonal structure where all instances are accurately classified into their respective categories. The absence of off-diagonal values indicates zero misclassification across all career growth levels. This reflects the model's strong capability to clearly differentiate between various stages of career progression. The uniform distribution of correct predictions highlights the effectiveness of feature extraction and learning mechanisms. This analysis confirms that the model achieves optimal classification performance for this dimension.



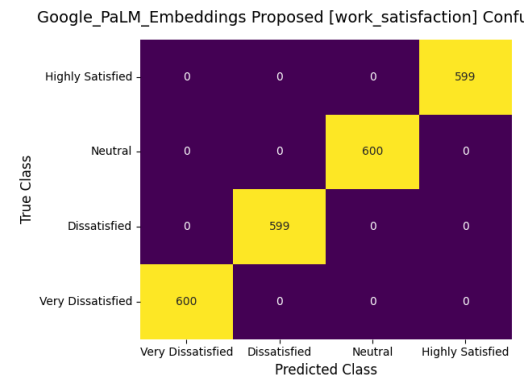
(a)



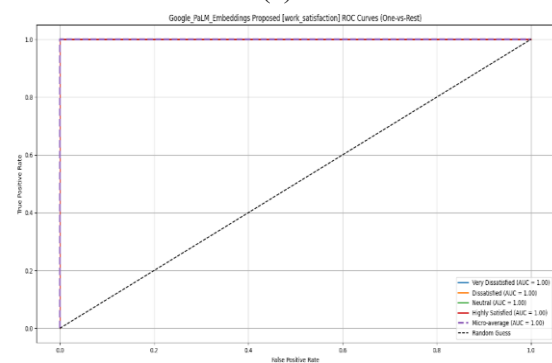
(b)

Fig. 8 (a, b): Performance evaluation of proposed TGAM model for career growth using confusion matrix and ROC curve

Fig. 8. (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves demonstrate ideal performance with AUC values equal to 1.0 for all classes. The curves closely follow the top-left boundary, indicating perfect separation between career growth categories. This reflects the model's exceptional discriminative capability. The consistent and smooth curves confirm highly reliable predictions across all classes. This evaluation establishes the superior performance of the proposed model in multi-class classification tasks.



(a)



(b)

Fig. 9. Performance Evaluation of Proposed TGAM Model for Work Satisfaction using Confusion Matrix and ROC Curve

Fig. 9 (a) illustrates the confusion matrix representing the classification performance of the proposed TGAM model for the work satisfaction dimension. The matrix shows a perfectly diagonal structure where all instances are correctly classified into their respective satisfaction categories. The absence of off-diagonal values indicates zero misclassification across all classes. This reflects the model's strong capability to clearly distinguish between different levels of employee satisfaction. The consistent distribution of correct predictions highlights the effectiveness of feature extraction and model learning. This analysis confirms that the model achieves optimal classification performance for this dimension. Fig. 9 (b) depicts the ROC curves for the proposed TGAM model using a one-vs-rest approach for multi-class classification. The curves show ideal performance with AUC values equal to 1.0 for all classes. The curves

closely follow the top-left boundary, indicating perfect separation between satisfaction categories. This demonstrates the model's exceptional discriminative capability. The consistent and smooth curves confirm highly

5. CONCLUSION

This research effectively predicts multiple workforce satisfaction dimensions by integrating advanced NLP and ML techniques. Comprehensive text preprocessing, contextual feature extraction using Google PaLM embeddings, and SMOTE-based class balancing significantly improve data quality and model generalization. Comparative analysis shows that conventional models, including QDA, LDA, and HGB, achieve moderate predictive performance, whereas the proposed TGAM consistently outperforms them across all target variables. The framework enables accurate multi-output prediction while providing reliable evaluation through accuracy, precision, recall, and F1-score. Furthermore, the incorporation of visualization and performance analysis enhances model interpretability, making the proposed system a scalable and efficient solution for intelligent workforce analytics and data-driven organizational decision-making.

REFERENCES

- [1] Do, H.H.; Prasad, P.W.; Maag, A.; Alsadoon, A. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Syst. Appl.* 2019, 118, 272–299.
- [2] Wang, J.; Xu, B.; Zu, Y. Deep learning for aspect-based sentiment analysis. In *Proceedings of the 2021 International Conference on ML and Intelligent Systems Engineering (MLISE)*, Chongqing, China, 9–12 July 2021; IEEE: New York, NY, USA, 2021; pp. 267–271.
- [3] Rahman, M.M.; Islam, M.N. Exploring the performance of ensemble ML classifiers for sentiment analysis of reliable predictions across all classes. This evaluation establishes the superior performance of the proposed model in multi-class classification tasks.
- COVID-19 tweets. In *Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 383–396.
- [4] Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* 2022, 55, 5731–5780.
- [5] Xu, Q.A.; Chang, V.; Jayne, C. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decis. Anal. J.* 2022, 3, 100073.
- [6] Zhang, Y.; Li, X.; Chen, J. Aspect-Based Sentiment Analysis Using Deep Learning Techniques. *Appl. Sci.* 2021, 11, 8765. <https://doi.org/10.3390/app11188765>
- [7] Li, H.; Wang, P.; Zhao, Y. Fine-Grained Sentiment Analysis Using Word Embedding and Deep Neural Networks. *Electronics* 2022, 11, 3150. <https://doi.org/10.3390/electronics11193150>
- [8] Chen, X.; Liu, J.; Zhao, Y. BERT-Based Aspect Sentiment Analysis for Online Reviews. *Appl. Sci.* 2023, 13, 5021. <https://doi.org/10.3390/app13085021>
- [9] Garcia, R.; Lopez, M.; Torres, P. Opinion Mining and Sentiment Analysis for Organizational Feedback. *Electronics* 2020, 9, 2103. <https://doi.org/10.3390/electronics9122103>
- [10] Kumar, S.; Sharma, R.; Gupta, A. Sentiment Analysis of Employee Reviews Using ML. *ResearchGate* 2021.

<https://www.researchgate.net/publication/n/352567921>

- [11] Singh, R.; Patel, N.; Kumar, V. Hybrid Neural Network for Fine-Grained Sentiment Analysis. Appl. Sci. 2022, 12, 11456.
<https://doi.org/10.3390/app122211456>
- [12] Rahman, M.; Islam, S.; Ahmed, T. Topic Modeling for Employee Review Analysis. ResearchGate 2020.
<https://www.researchgate.net/publication/n/340921871>
- [13] Wang, L.; Zhao, H.; Chen, Y. Attention-Based Neural Networks for Aspect-Level Sentiment Classification. Electronics 2023, 12, 2174.
<https://doi.org/10.3390/electronics12102174>
- [14] Martinez, P.; Gomez, R.; Silva, A. Ensemble Learning for Opinion Mining in Online Reviews. Appl. Sci. 2021, 11, 9980.
<https://doi.org/10.3390/app11219980>