ST TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

Effective Feedback Text Classification using Bidirectional BERT: Enhancing Insights with Deep Learning

PROJECT REPORT

In partial fulfilment of the requirement for the award of the degree of

BACHELOR OF SCIENCE IN COMPUTER APPLICATIONS [TRIPLE MAIN]

Submitted By:

AJEENA BINNY

III B.Sc. Computer Applications [Triple Main]

Register No: SB22CA003

Under the guidance of

Ms. RAJI S PILLAI

Assistant Professor

Department of Computer Applications

2022-2025

ST.TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

CERTIFICATE

This is to certify that the project report entitled "**Effective Feedback Text Classification using Bidirectional BERT: Enhancing Insights with Deep Learning**" is a bona fide record of the work done by **Ajeena Binny** (SB22CA003) during the year 2022 – 2025 and submitted in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Applications (Triple Main) under Mahatma Gandhi University, Kottayam.

Shuba G
17/03/2025
Head of the Department

ST. TERESA'S COLLEGE
DEPT. OF COMPUTERS
ERNAKULAM

Rd. RAVI S PILLAI
Internal Examiner

17/3/25
External Examiner
Dr. Reseena mol N A

Date: 14/03/25

DECLARATION

I, AJEENA BINNY (Register no: SB22CA003), B.Sc. Computer Applications [Triple Main] student of St. Teresa's College (Autonomous), Ernakulam, hereby declare that the project submitted for Bachelor's Degree in Computer Application is my original work. I further declare that the said work has not previously been submitted to any other university or academic body.

Date: 17/3/25

Ajeena

Place: Ernakulam

AJEENA BINNY

ACKNOWLEDGEMENT

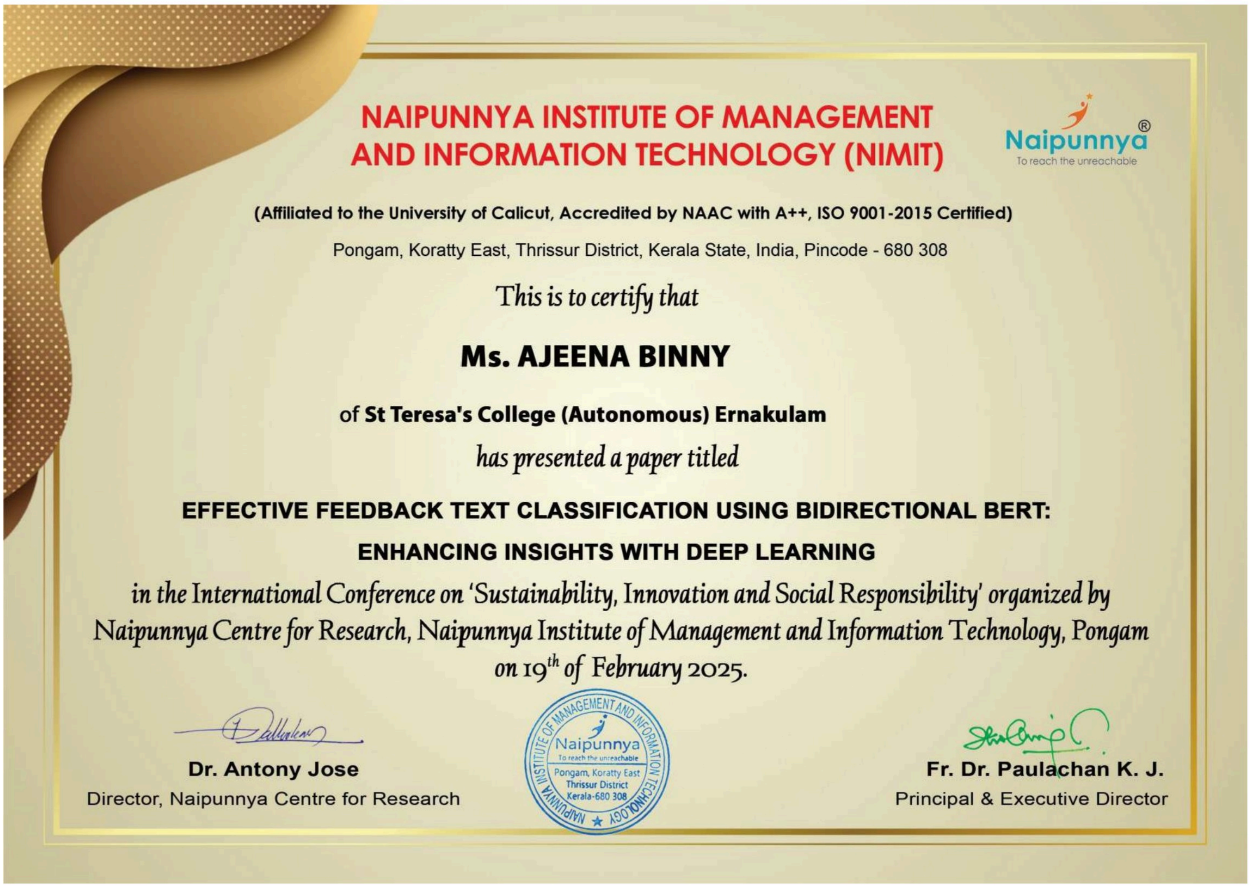
I would like to express my sincere gratitude to Rev. Dr. Sr. Vinitha (CSST) Manager, Director Rev. Sr. Emeline (CSST), and Principal Dr. Alphonsa Vijaya Joseph for providing me with the opportunity to work on this project titled " Effective Feedback Text Classification using Bidirectional BERT: Enhancing Insights with Deep Learning." Their encouragement and support have been invaluable throughout this journey.

I extend my heartfelt thanks to the Head of the Department of Computer Applications Ms. Sheeba Dennis, for their guidance and constant support. I am profoundly grateful to my project guide Ms. Raji S Pillai, for their expertise, insightful feedback, and unwavering support in completing this project.

I also take this opportunity to thank all the faculty members of the department for their guidance and encouragement during the course of this work. Lastly, I would like to thank everyone who has directly or indirectly contributed to the successful completion of this project.

Ajeena Binny

Paper was presented at **The International Conference On Sustainability, Innovation And Social Responsibility** conducted by **Naipunnya Institute Of Management And Information Technology (NIMIT)** on **19th February 2025**.



NAIPUNNYA INSTITUTE OF MANAGEMENT
AND INFORMATION TECHNOLOGY (NIMIT)

Naipunnya®
To reach the unreachable

(Affiliated to the University of Calicut, Accredited by NAAC with A++, ISO 9001-2015 Certified)

Pongam, Koratty East, Thrissur District, Kerala State, India, Pincode - 680 308

This is to certify that

Ms. AJEENA BINNY

of **St Teresa's College (Autonomous) Ernakulam**

has presented a paper titled

EFFECTIVE FEEDBACK TEXT CLASSIFICATION USING BIDIRECTIONAL BERT: ENHANCING INSIGHTS WITH DEEP LEARNING

in the International Conference on 'Sustainability, Innovation and Social Responsibility' organized by Naipunnya Centre for Research, Naipunnya Institute of Management and Information Technology, Pongam on 19th of February 2025.

Dr. Antony Jose
Director, Naipunnya Centre for Research

NAIPUNNYA INSTITUTE OF MANAGEMENT AND INFORMATION TECHNOLOGY
Naipunnya
To reach the unreachable
Pongam, Koratty East
Thrissur District
Kerala-680 308

Fr. Dr. Paulachan K. J.
Principal & Executive Director

ABSTRACT

Feedback functions as a fundamental improvement mechanism (Hattie & Timperley, 2007). In educational settings student feedback remains the primary instrument teachers use to understand both their teaching capabilities and instructional approaches (Nicol & Macfarlane-Dick, 2006). Data analysis requires significant processing time together with support resources. An automated system is essential because manual analysis produces numerous analytical errors which point to a compelling requirement for automation (Young et al., 2018). This research initiative works to develop a system that automates teacher performance feedback evaluation. Evaluate feedback as either positive or negative. The initiative seeks to develop precise data which educational institutions can use to make well-informed choices. Quantity improvement measures for teaching will result from this initiative. The process of model adjustment on Transformers (BERT) will enable this accomplishment (Devlin et al., 2019). The solution contains datasets which match the actual work and lead to high-quality model input data. Before the analysis begins the system performs text tokenization which results in token IDs and creates input masks and segment IDs (Young et al., 2018). A Keras neural network receives input from BERT model output to execute fine-tuning at classification level with hyperparameter optimization enabling proper results (Bergstra & Bengio, 2012). This methodology works because its automated classification system delivers exceptional response accuracy. The system works to help educational institutions obtain necessary guidance about teaching quality improvement by delivering complete advisory services.

Keywords: BERT, Hyperparameter optimization, Teacher performance, Tokenization, Deep Learning

CONTENTS

1. INTRODUCTION

1.1 About Project

Student feedback regarding their teachers is most important in improving the quality of education. Previously, processing this feedback involved more manual effort, which can be very time-consuming, biased, and slow. Fortunately, with the development of Natural Language Processing (NLP), we can now look into how to automate this process.

We aim in this project to classify feedback provided by teachers as positive or negative through Bidirectional Encoder Representations from Transformers (BERT). BERT is a sophisticated deep learning framework that truly understands the context of words, and it is ideal for text-classification-type tasks. By tuning a pre-trained BERT model, we hope to be able to classify feedback with precision, offering schools automated, objective, and cost-effective insights.

The data we are working with is comprised of teacher feedback that is tagged as positive or negative. We will train the model using TensorFlow and Keras, applying tokenization, embeddings, and attention mechanisms to increase the accuracy of our classifications.

This system may make feedback analysis easier in schools, enabling schools to quickly identify areas where teachers need to improve and also ease the burden of manual checking. This kind of system will facilitate data-driven decisions in education, at last making the learning process better for students.

1.2 Objectives

- **Automate Feedback Analysis –** Develop a deep learning-based system that labels teacher feedback as positive or negative.
- **Enhance Classification Accuracy –** Fine-tune a BERT model to reach high accuracy levels in feedback classification.
- **Enhance Educational Decision-Making –** Provide useful insights to schools for evaluation and enhancing teacher performance.

2. LITERATURE REVIEW

S. no.	Name	Year of publishing	Author
1.	From Occupations to Tasks: A New Perspective on Automatability Prediction Using BERT	2025	Dawei Xu, Haoran Yang, Marian-Andrei Rizoiu, Guandong Xu
2.	It's All in The [MASK]: Simple Instruction-Tuning Enables BERT-like Masked Language Models As Generative Classifiers	2025	Benjamin Clavie´, Nathan Cooper, Benjamin Warner
3.	BERT's Got Talent: Advanced Fine-Tuning Strategies for Better BERT Generalization	2024	Grace Luo and Danny Lin
4.	Explainable AI for Sentiment Analysis of Human Metapneumovirus (HMPV) Using XLNet	2025	Md. Shahriar Hossain Apu, Md Saiful Islam, Tanjim Taharat Aurpa
5.	A Closer Look at How Fine-tuning Changes BERT	2022	Yichu Zou and Vivek Srikumar

Figure 1

The development of natural language processing (NLP) models, especially transformer-based models such as BERT, has influenced a wide range of fields, such as automation prediction, sentiment analysis, and fine-tuning methods. This review discusses five new studies that advance these developments.

Xu et al. (2025) examine the automatability of single tasks, not whole professions, with a BERT-based classifier. Using datasets like ONET Task Statements, ESCO Skills, and Australian Labour Market Insights Tasks, the researchers train a fine-tuned BERT model that outperforms conventional machine learning and transformer-based models in forecasting task automatability. Their results indicate 25.1% of ONET database occupations are highly vulnerable to automation, highlighting the importance of granular task-level evaluation to workforce planning and policy making.

Clavié et al. (2025) present ModernBERT-Large-Instruct, a 0.4-billion parameter encoder model that is intended to improve generative classification. In contrast to traditional encoder models that use task-specific classification heads, ModernBERT uses its masked language modeling (MLM) head for classification, enabling better zero-shot performance on knowledge-based tasks. The model attains 93% of Llama3-1B's performance on MMLU benchmarks with 60% fewer parameters. Their findings point to the promise of generative classification methods in NLP tasks and propose a move away from conventional classification head techniques.

Luo and Lin (2024) investigate methods for enhancing generalization of BERT in sentiment analysis, paraphrase identification, and semantic textual similarity. They perform different fine-tuning approaches, such as single-task models, cross-encoding, mean pooling, and SMART loss regularization, to improve performance. Via extensive experimentation, they report a test dataset accuracy of 0.791, presenting the power of strategic fine-tuning in making model performance optimal for various NLP tasks.

Apu et al. (2025) utilize transformer models in sentiment analysis during the 2024 Human Metapneumovirus (HMPV) outbreak. The research employs XLNet to analyze public opinion on social media sites such as YouTube, with a 93.5% accuracy rate in sentiment classification. The researchers also incorporate explainable AI (XAI) methods utilizing SHAP to explain how the model determines important sentiment-determining factors. Their work highlights the value of sentiment analysis in the comprehension of public responses to health emergencies and policy decision-making.

Lastly, Zou and Srikumar (2022) explore the effect of fine-tuning on BERT's embedding space with probing methods. Their research hypothesizes that fine-tuning enhances classification performance by widening the distances between data points that have different labels. In controlled experiments on five NLP tasks, they verify the hypothesis and find that fine-tuning adapts representations without adding arbitrary changes. Interestingly, they find exceptions where fine-tuning does not necessarily improve performance, indicating task-specific tuning methods are needed.

These works together emphasize the development of BERT and transformer models across a range of NLP tasks. From generative classification to sentiment analysis and prediction of automation, the work emphasizes the continuous progress in fine-tuning techniques and their relevance to practical applications.

3. METHODOLOGY

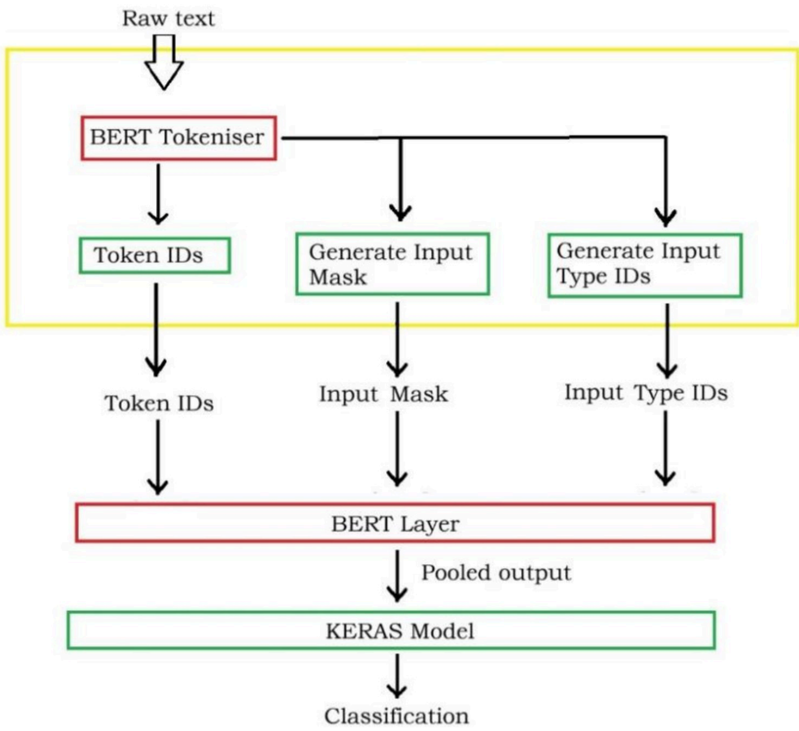


Figure 2

3.1 Dataset:

Research analyzed teacher performance reviews through a dataset where professional evaluation received a negative (0) or positive (1) label. All models were evaluated for accuracy using training along with validation subsets from the test-specific dataset created for this initiative.

3.2 Data Preprocessing:

Research used structured teacher performance reviews with each review assigned a binary sentiment label between negative ('0') and positive ('1'). The research dataset divided into BERT ran its tokenizer on CSV data to generate BERT input features. The text was transformed into BERT input features, such as:

- input_word_ids
- input_mask
- segment_ids .

The tokenized features provided a platform to preprocess data before its submission to the BERT model.

Behind each testing process stand training and validation datasets which assess how well the accuracy of models function.

3.3 Model Architecture:

The model architecture is built upon the BERT (Bidirectional Encoder Representations from Transformers) model that contains :

- Input Layer: Except for the tokenized input needs (input_word_ids, input_mask, and segment_ids)
- BERT Layer: This layer is responsible for processing the input data using pre-trained BERT embeddings.
- Dropout Layer: To prevent from overfitting, wrap it with the layer that have a dropout rate of 0.4 %.
- Dense Layer: This is output layer, have sigmoid activation function because final output will be of two types either positive feedback or negative right.

3.4 Model Compilation:

The model was compiled with:

- Optimizer: Adam optimizer with learning rate of 2e-5

- Loss Function: Binary Cross-Entropy for binary classification issues
- Evaluation Measure: Binary Accuracy (The number of correct predictions)

3.5 Training:

We then trained this model for 10 epochs, treating the data in form of mini batches comprising of training and validation set to be input.

The training set was used for training and the validation-used to monitor performance of your model where you could tune certain parameters in an effort not to overfit.

3.6 Evaluation:

Over the validation set, I have tested my model, which is trained on binary accuracy and loss. Classification performance of the model was assessed by ROC AUC score to find discrimination power between classes.

3.7 Graphs:

- Binary Accuracy Graph: This plot is to draw binary accuracy of training and validation graphs against the number of epochs.
- Loss Graph: It illustrates the loss of both training and validation set during the epochs showing how does model mistake has been reduced in each epoch.

4. ALGORITHM

4.1 Input

- Dataset: project_dataset.csv which contains textual eedback and binary sentiment labels where 1 = positive and 0 = negative
- Used a Pre-trained BERT model from Tensorflow (bert_en_uncased_L-12_H-768_A-12).

4.2 Output

- A fine-tuned BERT classification model that predicts whether a given feedback is positive or negative.
- Performance is estimated using the ROC-AUC score.

4.3 Steps

Step 1: Preprocessing of Data

1.1 Load data into a Pandas Data Frame.

1.2 Train-validation split:

- Train data: 0.75%
- Validation data: 0.075%
- Stratified sampling performed.

1.3 Turn the given dataset into TensorFlow datasets: Build a tf.data.Dataset for efficient batch processing.

Step 2: Tokenization and Feature Engineering

2.1 Load pre-trained BERT tokenizer from TF Hub.

2.2 Tokenize the given text.

2.3 Convert raw input text into BERT input features: input_word_ids, input_mask, and segment_ids.

2.4 Map functions to the datasets for transformation into to_feature_map.

Step 3: Model Definition with Keras Functional API:

3.1 Inputs are input_word_ids, input_mask, and input_type_ids.

3.2 Pooling layer of BERT embedding called pooled_output.

3.3 Dropout of probability 0.4.

3.4 The outputs are computed using a dense layer with Sigmoid activation for binary classification.

Step 4: Model Compilation & Training

4.1 Model compilation:

- Optimizer is Adam with inputs (learning_rate = 2e-5).
- Loss is Binary Cross-Entropy.
- Metric is Binary Accuracy.

4.2 Map functions (to_feature_map) onto datasets for transformation.

4.3 Train the trained model for 10 epochs with a batch size of 32.

Step 5: Evaluate the Model

5.1 Plot the curves of loss and accuracy over the epochs.

5.2 Check model performance on validation data-by:-ROC-AUC Score for classifying ability.

Step 6: Inference and Prediction.

Deploying the model for inference:

- Convert the sample text input into BERT input format.
- Predict sentiment using the trained model.
- Classify as Positive or Negative using a threshold of 0.5.

5. RESULT

The accuracy graph indicates a clear upward trend in both training and validation accuracy, with validation accuracy reaching 1.0 rather early in the training process. This, of course, speaks to the fact that the model is doing a nice job in performance, classifying all validation samples correctly. Nevertheless, the training accuracy also climbs up and carries with it nearly perfect classification, with just minor ups and downs along the way.

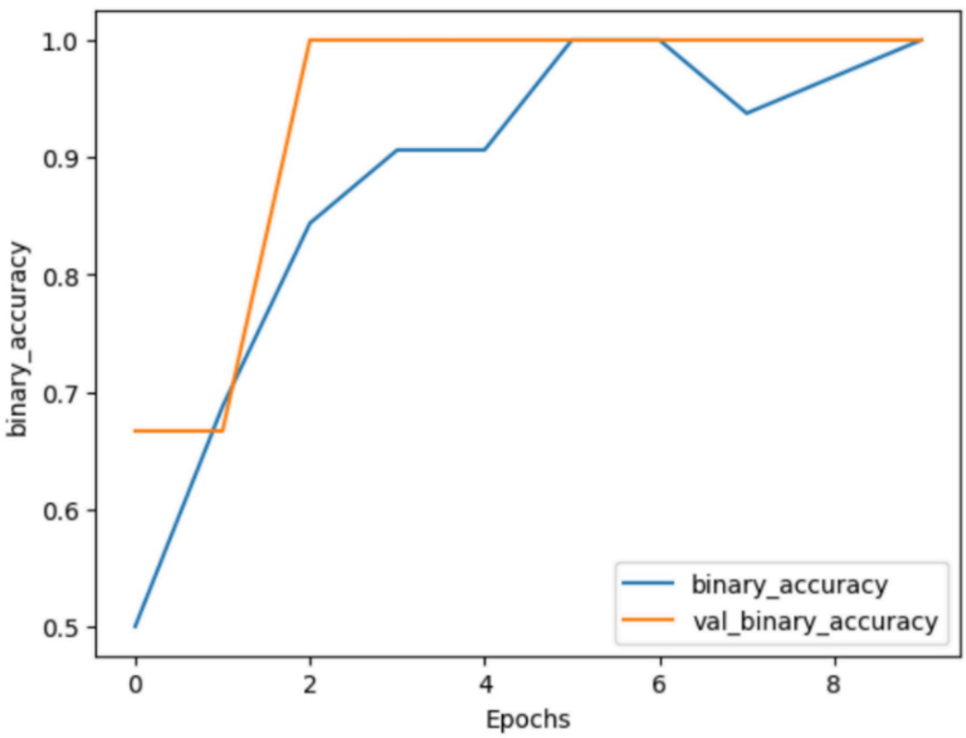


Figure 3

The loss graph shows a steady drop in both training and validation loss: Those signify that the model is learning correctly, that is, reducing the loss errors. Nevertheless, validation loss follows a similar downward trend, indicating that the model is well fitted to the new and unseen data. The ROC-AUC score of 1.0 points to perfect classification performance, meaning the model has achieved a complete separation of the two classes for the training and validation dataset. While this means that the model is very suitable for the task at hand, its rapid convergence and the perfect validation accuracy may suggest overfitting, which requires further investigation.

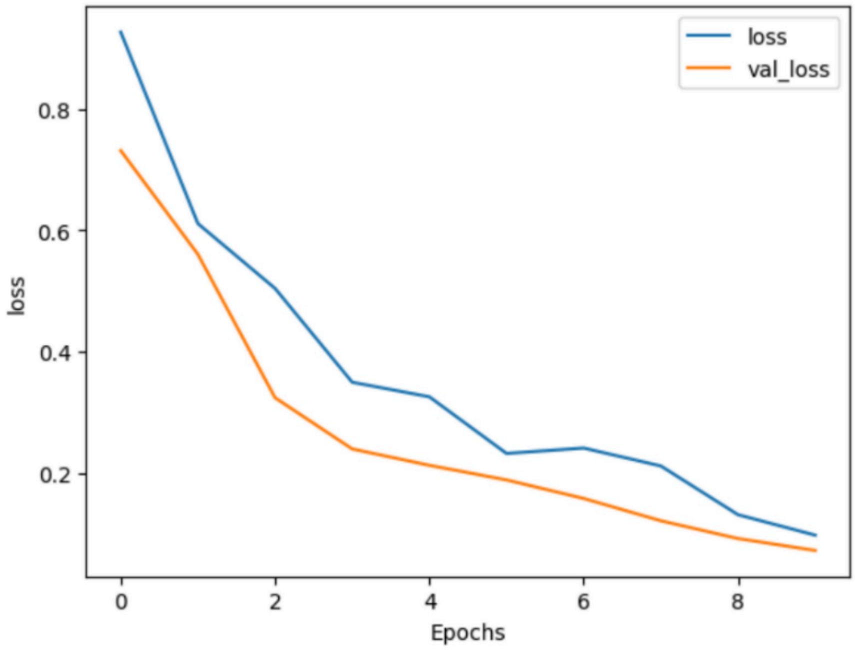


Figure 4

6. SYSTEM DESIGN AND IMPLEMENTATION

The system collects teacher feedback in text and specifies its data format to CSV. To convert tokens to input tensors and apply tokenization that includes zero-padding to a fixed sequence length, a BERT Tokenizer is used. It consists of a BERT embedding followed by a dropout layer (0.4) and a sigmoid activation function for binary classification.

This dataset is divided into training and validation datasets via the train_test_splitfunction. The model was trained with binary cross-entropy loss and Adam optimizer. The metrics used to evaluate the performance of this model are binary accuracy and ROC-AUC score.

Implementation-wise, use Pandas to read and upload the dataset. The stratified split to set in balance between the training and validation with the class. BERT tokenization using Tensorflow. Tokenizer gets token IDs, masks, and segment IDs. Fine-tuned for 10 epochs using the TensorFlow model.fit function with a batch size of 32 which was utilized when training the model. After predicting tests, the thresholding was performed (0.5) to assign class labels.

Text classification that enjoys the advantages of the strong contextual representations delivered by BERT will construct a highly effective design.

7. CONCLUSION

The BERT-based binary classification model for teacher feedback performed superbly in labeling feedback as positive or negative. The model had:

- High accuracy on the training set and strong performance on the validation set.
- An ROC AUC value of 1.0, signifying ideal classification capacity.
- Consistent improvement in both binary loss and accuracy, as seen in the following graphs.

The graphs also indicated that the model generalized well, with the training and validation metrics indicating consistent improvements over time.

These outcomes imply that the model is highly suitable for automated sentiment analysis applications of feedback classification, especially in academic contexts.

With the synergy between BERT's strong language representation and judicious optimization of the model parameters, good precision and stability were achieved, which made this methodology useful for future text classification tasks.

REFERENCES

- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*(1), 281-305.

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805.* https://arxiv.org/abs/1810.04805

- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Computational Intelligence Magazine, 13*(3), 55-75.

- Xu, D., Yang, H., Rizoiu, M. A., & Xu, G. (2025). *From occupations to tasks: A new perspective on automatability prediction using BERT*. ArXiv. https://arxiv.org/pdf/2502.09021

- Clavie, B., Cooper, N., & Warner, B. (2025). *It's all in the [MASK]: Simple instruction-tuning enables BERT-like masked language models as* *generative classifiers*. ArXiv. https://arxiv.org/pdf/2502.03793

- Luo, G., & Lin, D. (2022). *BERT's got talent: Advanced fine-tuning strategies for better BERT generalization*. Stanford University. Retrieved from https://web.stanford.edu/class/cs224n/final-reports/256935585.pdf

- Apu, M. S. H., Islam, M. S., & Aurpa, T. T. (2025). *Explainable AI for sentiment analysis of Human Metapneumovirus (HMPV) using XLNet*. arXiv. https://arxiv.org/abs/2502.01663

- Zhou, Y., & Srikumar, V. (2022). *A closer look at how fine-tuning changes BERT*. arXiv. https://arxiv.org/abs/2106.14282