1. Performance Comparison Summary

Precision, Recall, and F1-score
Based on the comparison charts, the LSTM model outperforms the Random Forest model in recall and F1-score for most sentiment classes. Specifically:

- LSTM demonstrates superior recall for the negative and neutral classes.
- LSTM achieves a higher F1-score across all classes except for the positive class, where both models perform comparably.
- Random Forest shows higher precision in the negative and neutral classes, but this is not complemented by equivalent recall, resulting in lower overall F1-scores for those classes.

LSTM Accuracy and Loss Trends
The training accuracy for LSTM increases steadily across epochs, while validation accuracy declines, indicating overfitting. Similarly, validation loss increases while training loss decreases, reinforcing the overfitting observation.

ROC Curve and AUC Scores
The ROC-AUC analysis shows:

- LSTM achieves better discrimination in the negative and positive classes, with AUC scores of 0.76 and 0.71, respectively.
- Random Forest achieves a slightly higher AUC in the neutral class (0.60) compared to LSTM (0.57).
- Overall, LSTM provides more balanced class-wise performance, particularly in classes with higher sentiment polarity.

Confusion Matrix Analysis

- LSTM correctly classifies a large portion of the positive samples (true positives = 8868), though it shows moderate misclassification in the negative and neutral classes.
- Random Forest classifies the positive class well (true positives = 11080) but performs poorly in classifying negative and neutral samples, showing high false positive and false negative rates for these categories.

2. Hyperparameter Tuning Summary

Random Forest Model Tuning
Hyperparameters tuned using Grid Search and Randomized Search included:

- Number of estimators (n_estimators)
- Tree depth (max_depth)

- Minimum samples required to split a node (min_samples_split)
  Cross-validation was applied to evaluate each parameter combination and reduce the risk of overfitting.

LSTM Model Tuning
Tuning focused on the architecture and training process, including:

- Number of LSTM layers and units per layer
- Embedding dimension and dropout rates
- Learning rate, batch size, and number of training epochs
  Despite tuning, overfitting persisted, as seen in the diverging training and validation performance curves.

3. Comprehensive Model Comparison and Selection

Model	**Accuracy**	**Best AUC**	**Overfitting**	**Inference Time**	**Ease of Deployment**
LSTM	~62%	0.76	Yes	Moderate to High	Requires GPU or optimized runtime
Random Forest	~60%	0.69	Less	Low	CPU-friendly and easy to deploy

Final Model Selection
The LSTM model is selected for its stronger overall performance in classifying sentiment, particularly in terms of recall, F1-score, and AUC for critical sentiment classes. Although the LSTM shows signs of overfitting, this can be addressed through techniques such as early stopping, regularization, and further architecture optimization.

Justification for Selection

- Superior recall and AUC in negative and positive sentiment detection, which are often more critical in sentiment classification tasks.
- Capable of capturing sequential dependencies in textual data, which traditional models like Random Forest cannot model effectively.

Documentation for Model Evaluation Report

- Include precision, recall, and F1-score comparisons across classes.
- Provide ROC curves and AUC summaries.
- Detail confusion matrix insights to reflect real-world classification behavior.
- Record all hyperparameter tuning strategies and final selected values.
- Justify model choice with clear evidence and note limitations of discarded models.