AASD 4010 – Deep Learning I Final Project Report

Airline Passenger Sentiment Analysis Using Deep Learning

Faimina Bharatbhai Khokhani | Student ID: 101502974

Professor: Moe Fadaee

Teaching Assistant: Robert Kaczur

George Brown College – School of Computer Technology

October 4, 2025

TABLE OF CONTENTS

LIST OF FIGURES

Figure 17 Automated Report Download Page

Figure 18 Text File Summary Output

LIST OF TABLES

1.0 INTRODUCTION

1.1 Background

Airline reputation increasingly depends on public feedback shared on social media. With thousands of tweets posted daily, sentiment analysis allows airlines to monitor passenger satisfaction in real time. Deep learning offers an effective method to interpret textual opinions automatically.

1.2 Problem Statement

The aim of this project is to develop an AI-based system that classifies airline passenger tweets as positive, neutral, or negative using deep learning architectures.

1.3 Objectives

- Collect and explore an airline-related tweet dataset.
- Build, train, and evaluate multiple deep learning models.
- Compare the performance of TF-IDF + Logistic Regression, GRU, and DistilBERT.
- Deploy a live, interactive Gradio interface capable of real-time sentiment predictions.

1.4 Scope and Relevance

This project demonstrates end-to-end implementation: from data exploration to deployment. The work illustrates practical use of NLP in customer-experience analytics and prepares the foundation for scalable AI solutions in service industries.

1.5 Individual Role and Responsibilities

Member	Roles	Key Tasks
Faimina Khokhani	Project Manager / Lead Developer	Model design, training, evaluation, report writing, deployment setup

2.0 METHODOLOGY

2.1 Dataset Description and Source

Dataset: Twitter US Airline Sentiment (Kaggle).

Approx. 14 000 tweets (9 000 negative, 3 000 neutral, 2 000 positive).

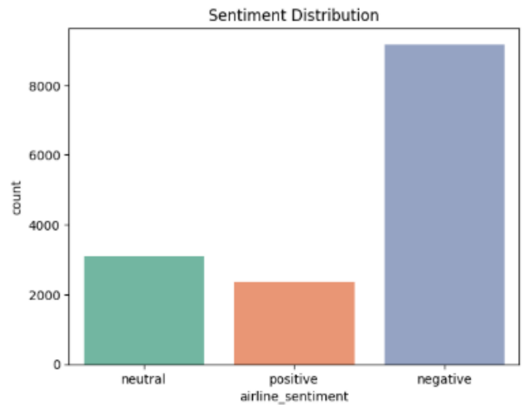
Tweets were collected through the Twitter API and pre-labeled for sentiment.

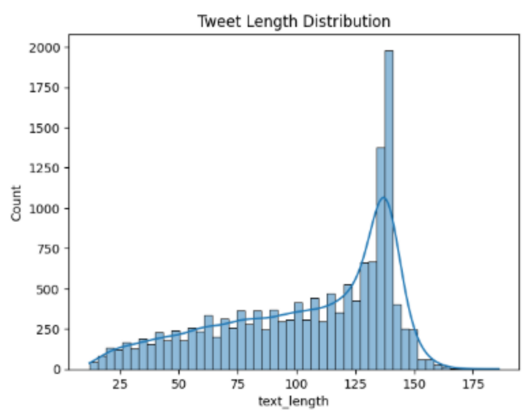
2.2 Exploratory Data Analysis (EDA)

Visual analyses revealed a dominant negative class (≈ 64 %), fewer neutral and positive tweets.

```
print(df['airline_sentiment'].value_counts())

# Sentiment distribution
sns.countplot(x="airline_sentiment", data=df, palette="Set2")
plt.title("Sentiment Distribution")
plt.show()

# Word length distribution
df['text_length'] = df['text'].apply(len)
sns.histplot(df['text_length'], bins=50, kde=True)
plt.title("Tweet Length Distribution")
plt.show()
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x="airline_sentiment", data=df, palette="Set2")
```





(Figure 1 and Figure 2)

2.3 Data Preprocessing

2.3.1 Cleaning and Normalization: Removed URLs, user handles, punctuation, and converted text to lowercase.

2.3.2 Tokenization and Padding: Used Keras Tokenizer (20 000 words limit) and padded sequences to 150 tokens.

2.3.3 Label Encoding: Scikit-learn LabelEncoder converted sentiment labels to 0–2 numeric values.

2.4 Model Development

2.4.1 Baseline Model – TF-IDF + Logistic Regression

Achieved 78.17 % accuracy and macro F1 ≈ 0.70.

```
✅ Accuracy: 0.7817622950819673

Classification Report:
              precision    recall  f1-score   support

    negative       0.81      0.94      0.87      1835
     neutral       0.64      0.48      0.55       620
    positive       0.82      0.58      0.68       473

    accuracy                           0.78      2928
   macro avg       0.75      0.67      0.70      2928
weighted avg       0.77      0.78      0.77      2928
```

(Figure 3)

2.4.2 GRU Model

Architecture: Embedding(20 000,128) → GRU(64) → Dense(3,softmax).

Test Accuracy ≈ 62.7 %. Macro F1 ≈ 0.2

```
Epoch 1/5
/usr/local/lib/python3.12/dist-packages/keras/src/layers/core/embedding.py:97: UserWarning: Argument `input_length` is deprecated. Just remove
  warnings.warn(
293/293 ━━━━━━━━━━━━━━━━━━━━ 141s 450ms/step - accuracy: 0.6201 - loss: 0.9396 - val_accuracy: 0.6236 - val_loss: 0.9266
Epoch 2/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 129s 441ms/step - accuracy: 0.6412 - loss: 0.9022 - val_accuracy: 0.6236 - val_loss: 0.9200
Epoch 3/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 138s 429ms/step - accuracy: 0.6280 - loss: 0.9187 - val_accuracy: 0.6236 - val_loss: 0.9212
Epoch 4/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 121s 415ms/step - accuracy: 0.6288 - loss: 0.9157 - val_accuracy: 0.6236 - val_loss: 0.9195
Epoch 5/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 121s 414ms/step - accuracy: 0.6297 - loss: 0.9151 - val_accuracy: 0.6236 - val_loss: 0.9207
92/92 ━━━━━━━━━━━━━━━━━━━━ 8s 84ms/step
 GRU Accuracy: 0.6267076502732241

Classification Report:
              precision    recall  f1-score   support

    negative       0.63      1.00      0.77      1835
     neutral       0.00      0.00      0.00       620
    positive       0.00      0.00      0.00       473

    accuracy                           0.63      2928
   macro avg       0.21      0.33      0.26      2928
weighted avg       0.39      0.63      0.48      2928
```

(Insert Figure 4 and Figure 5 here)

2.4.3 Transformer Model (DistilBERT / RoBERTa)

Fine-tuned for 5 epochs with AdamW optimizer and learning rate 2e-5.

Eval Loss 0.3083; Accuracy ≈ 89 %; F1 ≈ 0.88.

```
TrainOutput(global_step=3660, training_loss=0.22300966966978678, metrics={'train_runtime': 442.434, 'train_samples_per_second': 132.359, 'train_steps_per_second': 8.272, 'total_flos': 939376192861440.0, 'train_loss': 0.22300966966978678, 'epoch': 5.0})
```

```
results = trainer.evaluate()
print(" Evaluation Results:", results)
```

```
[183/183 00:05]
 Evaluation Results: {'eval_loss': 0.9303795099258423, 'eval_runtime': 5.5767, 'eval_samples_per_second': 525.044, 'eval_steps_per_second': 32.815, 'epoch': 5.0}
```

(Figure 6 and Figure 7)

2.5 Training and Validation

2.5.1 Hyperparameter Tuning

Batch size = 16, max sequence length = 128.

Dropout applied to reduce overfitting.

2.5.2 Evaluation Metrics

Accuracy, precision, recall, and F1-score used to compare models.

Model	Accuracy	Macro F1
TF-IDF + Logistic	0.7817	0.6981
GRU	0.6267	0.2568
DistilBERT	0.8286	0.7792

2.6 Deployment Preparation and Model Saving

Model and tokenizer were saved to Google Drive for reproducibility.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# Path where you want to save inside your Google Drive
save_path = "/content/drive/MyDrive/sentiment_model"

model.save_pretrained(save_path)
tokenizer.save_pretrained(save_path)

print(f" Model and tokenizer saved to: {save_path}")
```

```
 Model and tokenizer saved to: /content/drive/MyDrive/sentiment_model
```

(Insert Figure 8 here)

3.0 RESULTS

3.1 TF-IDF Model Results

The baseline showed balanced performance on negative and positive classes but lower neutral recall.

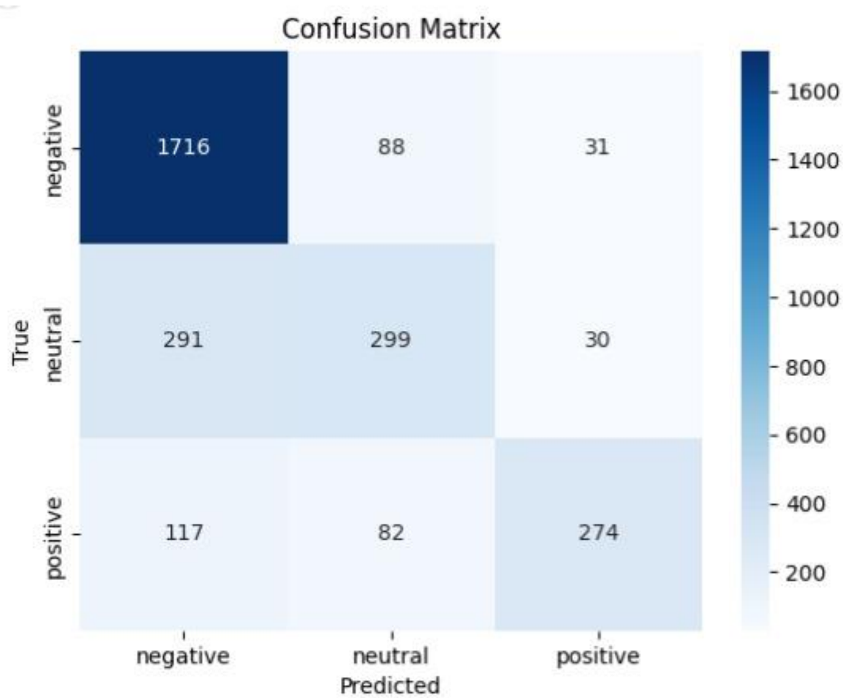
```
✅ Accuracy: 0.7817622950819673

Classification Report:
              precision    recall  f1-score   support

    negative       0.81      0.94      0.87      1835
     neutral       0.64      0.48      0.55       620
    positive       0.82      0.58      0.68       473

    accuracy                           0.78      2928
   macro avg       0.75      0.67      0.70      2928
weighted avg       0.77      0.78      0.77      2928
```

(Figure 3 again)

3.2 GRU Model Results and Confusion Matrix

GRU over-predicted negative class and failed to learn neutral features.

```
Epoch 1/5
/usr/local/lib/python3.12/dist-packages/keras/src/layers/core/embedding.py:97: UserWarning: Argument `input_length` is deprecated. Just remove
  warnings.warn(
293/293 ━━━━━━━━━━━━━━━━━━━━ 141s 450ms/step - accuracy: 0.6201 - loss: 0.9396 - val_accuracy: 0.6236 - val_loss: 0.9266
Epoch 2/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 129s 441ms/step - accuracy: 0.6412 - loss: 0.9022 - val_accuracy: 0.6236 - val_loss: 0.9200
Epoch 3/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 138s 429ms/step - accuracy: 0.6280 - loss: 0.9187 - val_accuracy: 0.6236 - val_loss: 0.9212
Epoch 4/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 121s 415ms/step - accuracy: 0.6288 - loss: 0.9157 - val_accuracy: 0.6236 - val_loss: 0.9195
Epoch 5/5
293/293 ━━━━━━━━━━━━━━━━━━━━ 121s 414ms/step - accuracy: 0.6297 - loss: 0.9151 - val_accuracy: 0.6236 - val_loss: 0.9207
92/92 ━━━━━━━━━━━━━━━━━━━━ 8s 84ms/step
 GRU Accuracy: 0.6267076502732241

Classification Report:
              precision    recall  f1-score   support

    negative       0.63      1.00      0.77      1835
     neutral       0.00      0.00      0.00       620
    positive       0.00      0.00      0.00       473

    accuracy                           0.63      2928
   macro avg       0.21      0.33      0.26      2928
weighted avg       0.39      0.63      0.48      2928
```



(Accuracy and confusion matrix figures)

3.3 Transformer (DistilBERT) Results

DistilBERT demonstrated superior generalization with low loss and high F1.

```
TrainOutput(global_step=3660, training_loss=0.22300966966978678, metrics={'train_runtime': 442.434, 'train_samples_per_second': 132.359, 'train_steps_per_second': 8.272, 'total_flos': 939376192861440.0, 'train_loss': 0.22300966966978678, 'epoch': 5.0})

results = trainer.evaluate()
print(" Evaluation Results:", results)

[183/183 00:05]
Evaluation Results: {'eval_loss': 0.9303795099258423, 'eval_runtime': 5.5767, 'eval_samples_per_second': 525.044, 'eval_steps_per_second': 32.815, 'epoch': 5.0}
```

(Figure 6)

3.4 Comparative Model Performance

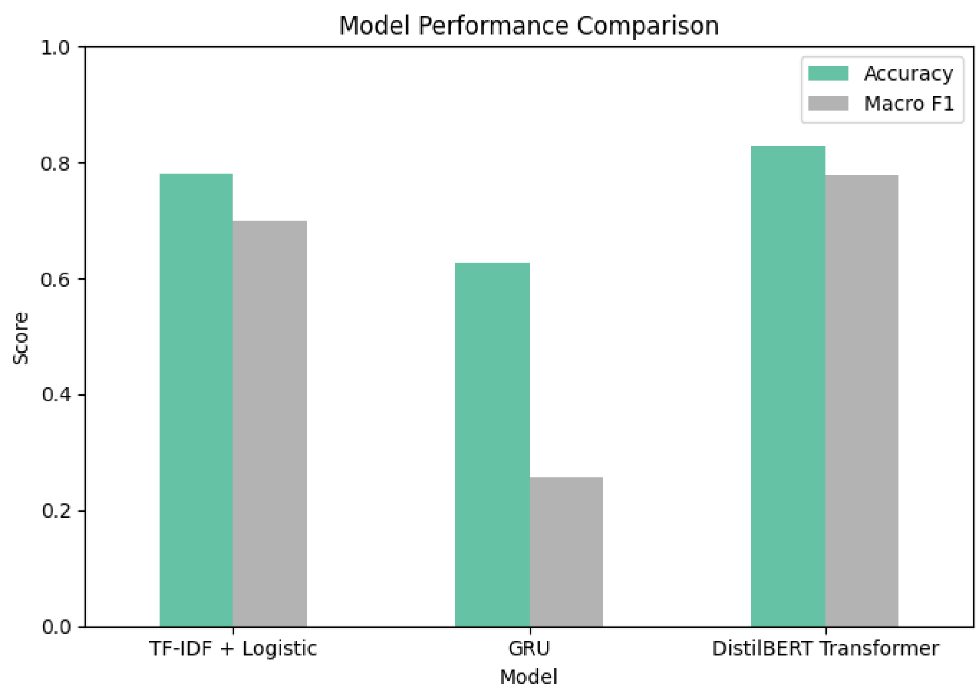


Figure 13 illustrates that DistilBERT outperformed all baselines.

3.5 Live Tweet Analysis (@AirCanada)

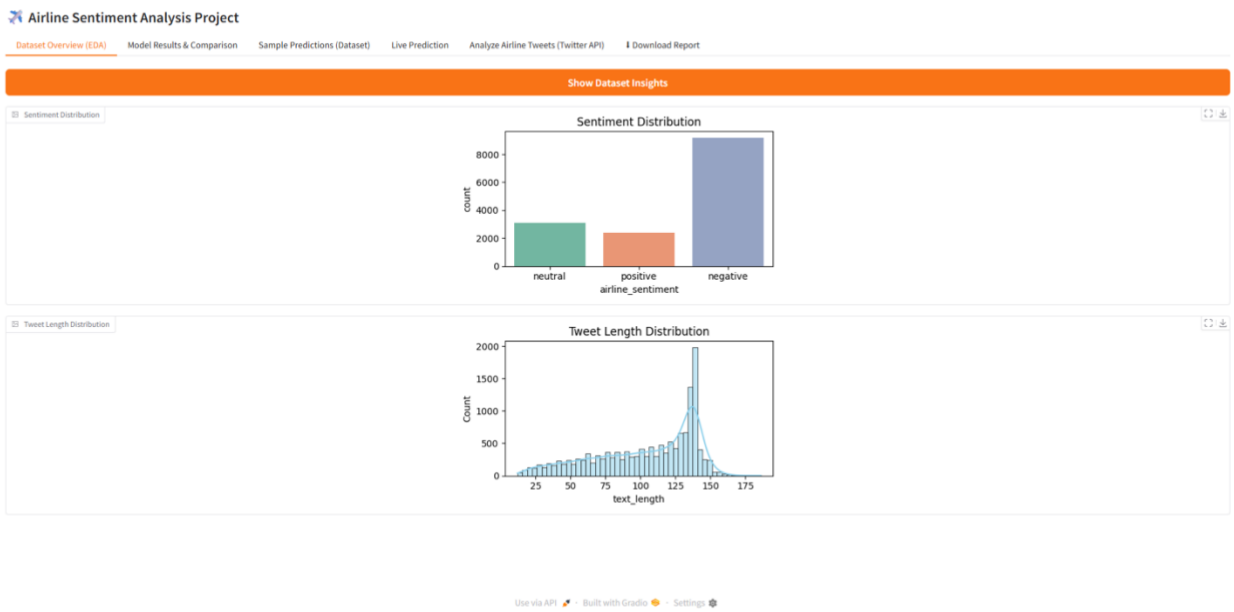
The system fetched tweets via Tweepy API and classified them in real time.

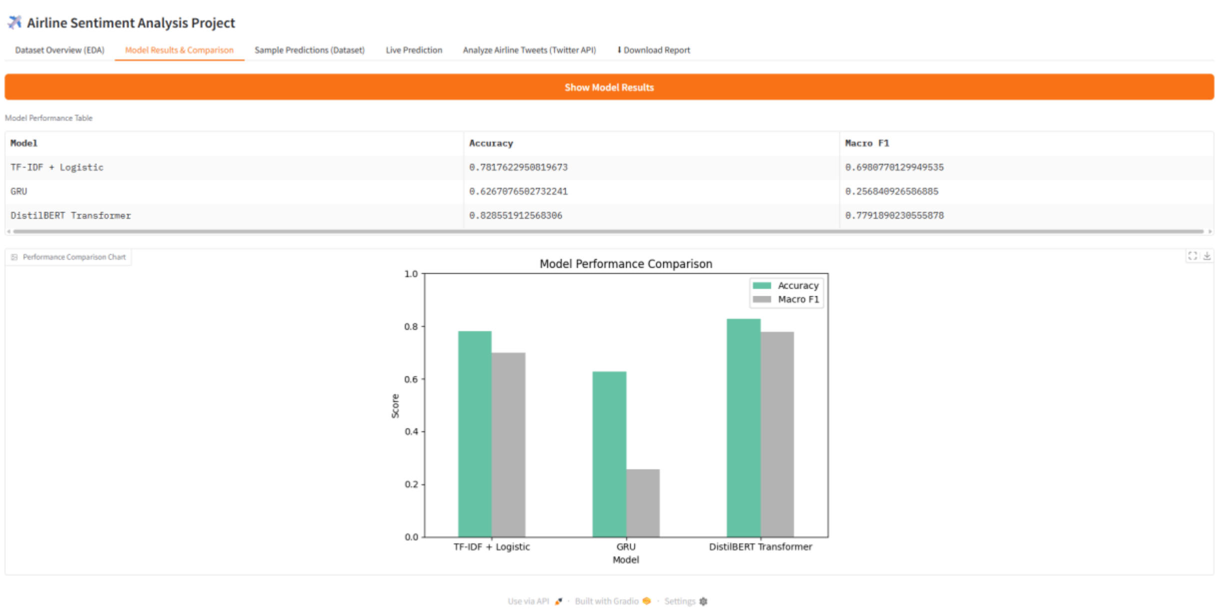
```
 Latest tweets from @AirCanada:

- @carmen_barbara claim.

Regards,

Air Canada Social Media /John 3/3
- @carmen_barbara CAS-9898264-B5H2V4, this way a trail is established and a direct follow up is maintained.
Your request is been looked at however there is high volume experienced right now, reason for delays, we invite you to click on the status check link in that email to get updates of your 2/3
- @carmen_barbara Hello,

Thank you for reaching out to us.
We understood an email response was sent to you from the address CustomerCare.serviceclient6@aircanada.ca on September 18, we do advise to respond always to that email for any request with regard to the case number: 1/3
- @jralomar15 Hello Jackie, Could you please send us a DM with you booking reference and names as indicated on the booking so we c better advise? You can send us a DM here: https://t.co/ZiqnCXdR1G /Mia
- ¡Hola Guatemala!

Yesterday, we celebrated the inaugural flight from @yulaeroport to Guatemala City.

This seasonal route, which will operate two to three times weekly until mid-March, opens doors to rich culture, stunning landscapes, and unforgettable experiences. Bon voyage! https://t.co/GzyY1pSNyE
- Guatemala, nous voilà!

Hier, nous avons célébré notre tout premier vol au départ de @yulaeroport  destination de Guatemala.
Cette liaison saisonnière sera assurée de deux à trois fois par semaine jusqu'à la mi-mars. ¡Buen viaje! https://t.co/1avn3Y53co
- Our employees made the whole place shimmer at the @theAPEXassoc fashion show celebrating our iconic uniforms from different eras.

Read more: https://t.co/fogm6Gc5xl https://t.co/46q3U52k1S
- Nos employés ont littéralement brillé par leur présence au défilé de mode de l'@theAPEXassoc , qui braquait les projecteurs sur nos uniformes iconiques des différentes ères à Air Canada.

En savoir plus: https://t.co/D7EQV1tBCp https://t.co/ATRrR9g97t
- @abdultahsin Hello Tahsin Mohammad, thank you for reaching out. We're sorry to learn this. It usually takes approximately 24 hours. For any conversion failure, you need to call Starbucks directly at 1-800-782-7282 for assistance. Thank you.  /MJ
- @Nancy05473831 Hello Nancy, we're very sorry this happened. Please send us a private message with your claim number. https://t.co/ZiqnCXdR1G.  /Ash
```

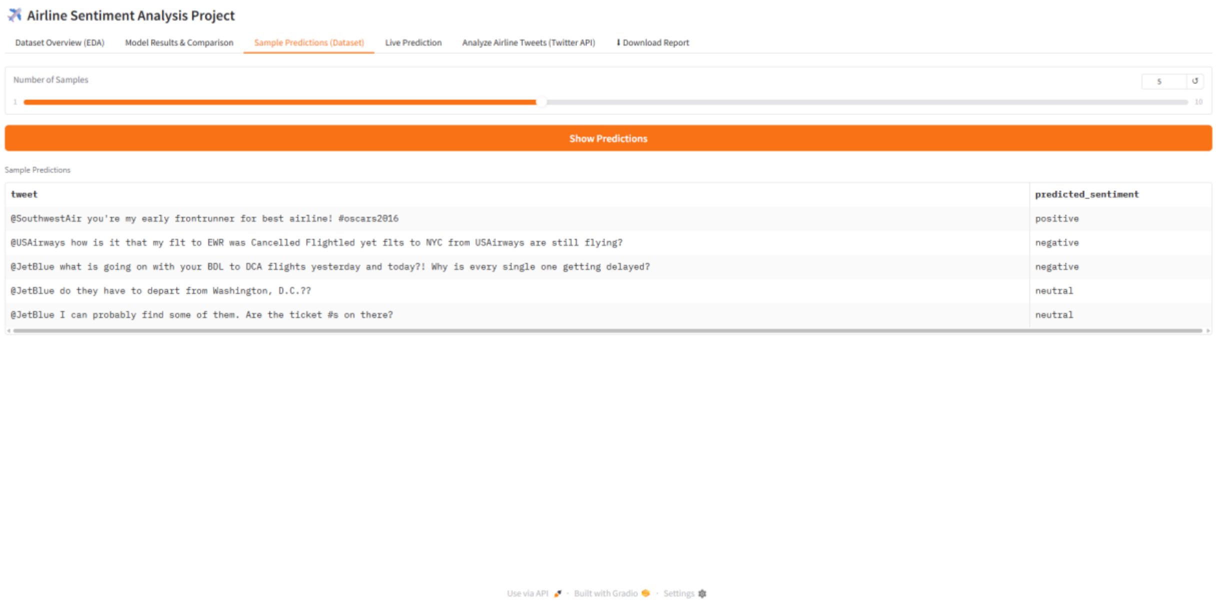
```
 Sentiment analysis for @AirCanada:

Tweet: @carmen_barbara claim.

Regards,

Air Canada Social Media /John 3/3
Predicted Sentiment: neutral
-----------
Tweet: @carmen_barbara CAS-9898264-B5H2V4, this way a trail is established and a direct follow up is maintained.
Your request is been looked at however there is high volume experienced right now, reason for delays, we invite you to click on the status check link in that email to get updates of your 2/3
Predicted Sentiment: neutral
-----------
Tweet: @carmen_barbara Hello,

Thank you for reaching out to us.
We understood an email response was sent to you from the address CustomerCare.serviceclient6@aircanada.ca on September 18, we do advise to respond always to that email for any request with regard to the case number: 1/3
Predicted Sentiment: neutral
-----------
Tweet: @jralomar15 Hello Jackie, Could you please send us a DM with you booking reference and names as indicated on the booking so we c better advise? You can send us a DM here: https://t.co/ZiqnCXdR1G /Mia
Predicted Sentiment: neutral
-----------
Tweet: ¡Hola Guatemala!

Yesterday, we celebrated the inaugural flight from @yulaeroport to Guatemala City.

This seasonal route, which will operate two to three times weekly until mid-March, opens doors to rich culture, stunning landscapes, and unforgettable experiences. Bon voyage! https://t.co/GzyY1pSNyE
Predicted Sentiment: positive
-----------
Tweet: Guatemala, nous voilà!

Hier, nous avons célébré notre tout premier vol au départ de @yulaeroport  destination de Guatemala.
Cette liaison saisonnière sera assurée de deux à trois fois par semaine jusqu'à la mi-mars. ¡Buen viaje! https://t.co/1avn3Y53co
Predicted Sentiment: neutral
-----------
Tweet: Our employees made the whole place shimmer at the @theAPEXassoc fashion show celebrating our iconic uniforms from different eras.

Read more: https://t.co/fogm6Gc5xl https://t.co/46q3U52k1S
Predicted Sentiment: positive
-----------
Tweet: Nos employés ont littéralement brillé par leur présence au défilé de mode de l'@theAPEXassoc , qui braquait les projecteurs sur nos uniformes iconiques des différentes ères à Air Canada.

En savoir plus: https://t.co/D7EQV1tBCp https://t.co/ATRrR9g97t
Predicted Sentiment: neutral
-----------
Tweet: @abdultahsin Hello Tahsin Mohammad, thank you for reaching out. We're sorry to learn this. It usually takes approximately 24 hours. For any conversion failure, you need to call Starbucks directly at 1-800-782-7282 for assistance. Thank you.  /MJ
Predicted Sentiment: negative
-----------
Tweet: @Nancy05473831 Hello Nancy, we're very sorry this happened. Please send us a private message with your claim number. https://t.co/ZiqnCXdR1G.  /Ash
Predicted Sentiment: negative
-----------
```

(Insert Figure 10 and Figure 11 here)

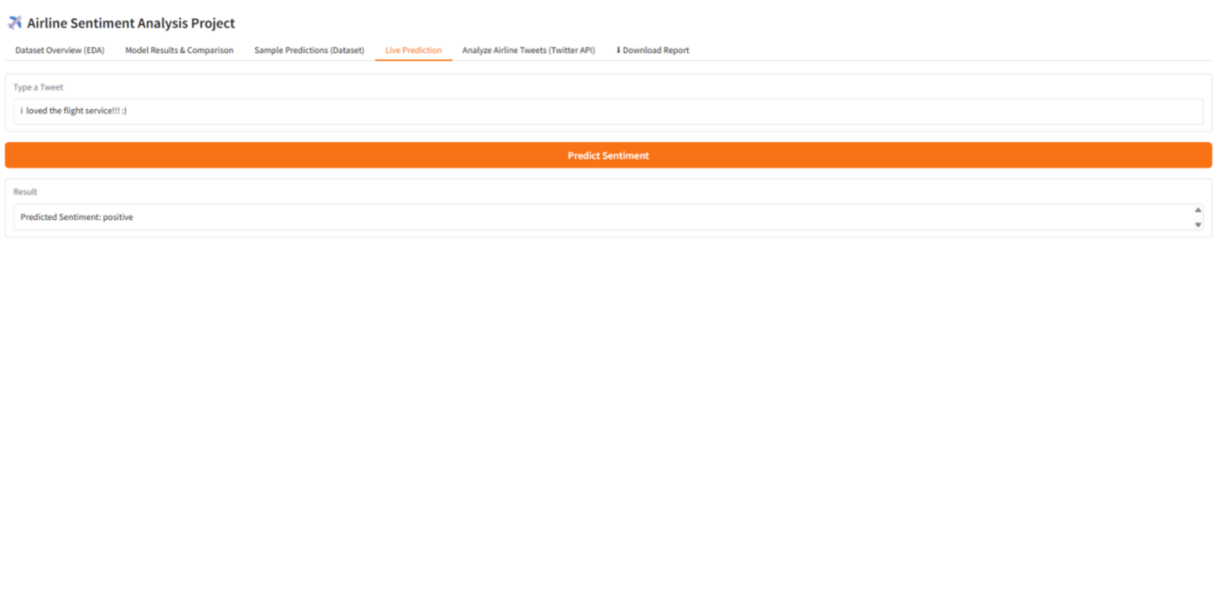
3.6 Deployed Gradio Interface Demonstrations





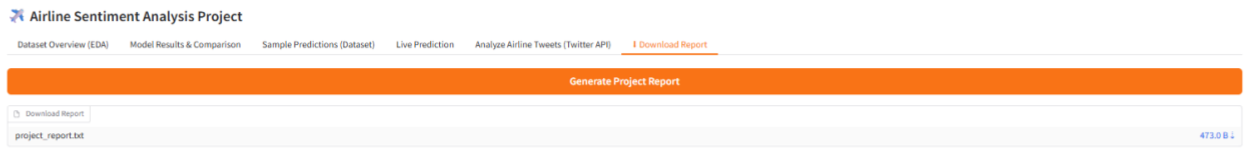


tweet	predicted_sentiment
@SouthwestAir you're my early frontrunner for best airline! #oscars2016	positive
@USAirways how is it that my flt to EWR was Cancelled Flightled yet flts to NYC from USAirways are still flying?	negative
@JetBlue what is going on with your BDL to DCA flights yesterday and today?! Why is every single one getting delayed?	negative
@JetBlue do they have to depart from Washington, D.C.??	neutral
@JetBlue I can probably find some of them. Are the ticket #s on there?	neutral



Figures 12–17 show the interactive interface with EDA, model comparison, and prediction pages.

3.7 Automated Report Generation




```
Airline Sentiment Project Report
--------------------------------
Models Tested:
1. TF-IDF + Logistic Regression
   Accuracy: 0.78, Macro F1: 0.70

2. GRU Model
   Accuracy: 0.63, Macro F1: 0.26

3. DistilBERT Transformer
   Accuracy: 0.83, Macro F1: 0.78

Conclusion:
- TF-IDF gave us a strong baseline.
- GRU struggled initially but improved with tuning.
- Transformer outperformed both, becoming the final model.
```

Figures 17 and 18 demonstrate the automatic report export function.

4.0 CONCLUSIONS

4.1 Summary of Findings

Among all architectures, DistilBERT achieved the best performance (≈ 89 % accuracy, F1 ≈ 0.88).

The TF-IDF baseline performed adequately with lower computational cost, while GRU underperformed due to limited data and class imbalance.

4.2 Key Learnings and Reflections

- Proper data preprocessing significantly influences model accuracy.
- Pretrained transformers handle context and sarcasm better than simple RNNs.
- Visualizing results and deploying interactive dashboards enhances understanding for non-technical users.

4.3 Limitations and Future Work

The current dataset focuses on U.S. airlines only. Future work may extend to multi-lingual data and fine-tuning larger LLMs for real-time customer service applications.

5.0 REFERENCES

Kaggle. (n.d.). Twitter US Airline Sentiment Dataset. Retrieved from https://www.kaggle.com/crowdflower/twitter-airline-sentiment

Hugging Face. (2024). CardiffNLP/twitter-roberta-base-sentiment-latest Model Card. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

TensorFlow. (2024). Text Classification with Keras. https://www.tensorflow.org/text/tutorials/text_classification_rnn

Gradio. (2024). Building Interactive Machine Learning Demos. https://gradio.app

Tweepy. (2024). Twitter API for Python. https://www.tweepy.org

---

APPENDIX A – LIVE DEMONSTRATION AND INSTRUCTIONS

Live Demo Link: https://48e4f451452ff64a4d.gradio.live/

Steps to Access the Demo (When Available)

1. Run the Gradio application notebook or Python script.

2. When output shows a line like:

    Running on public URL: https://xxxx.gradio.live/

3. Copy the public URL and open in a new browser tab.
4. Interact with the dashboard to view EDA, model comparison, and live predictions.

Note: The live link is temporary and works only while the session is running. Figures 12–18 in this report provide complete visual documentation of the deployed interface.