Twitter Airline Sentiment

Group 5: Ange Olson, Pavani Samala, Yaxin Zhuang

Overview

- EDA/Description of Dataset
- Models and Experimental Setup
- Results
- Model Interpretation(Integrated Gradient)
- Summary and Conclusion
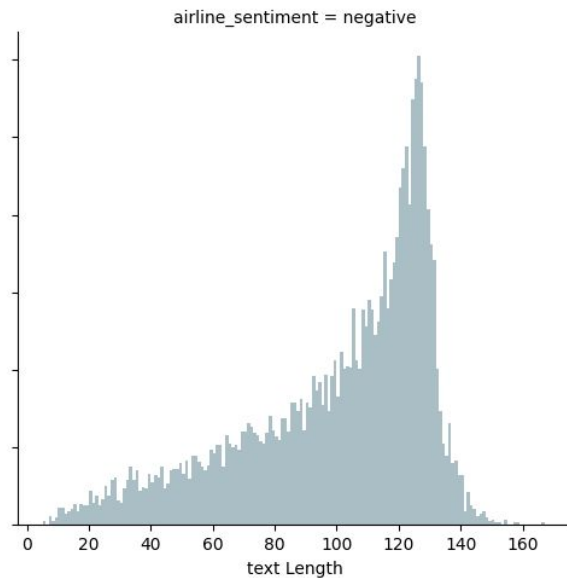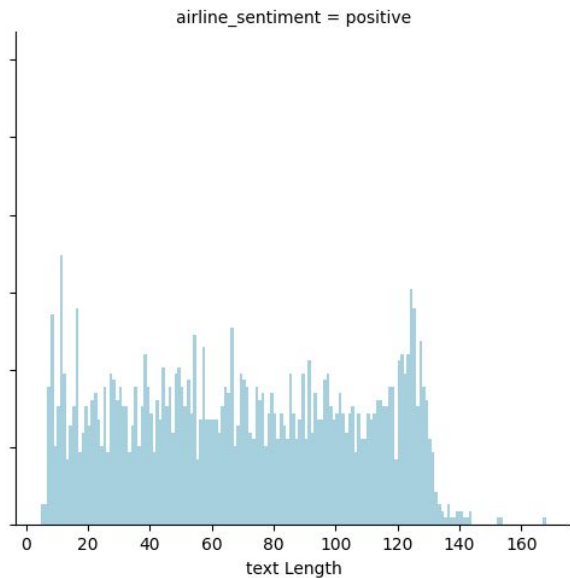- References

Description of Dataset

- Twitter data with 14,000 data points and 15 variables

Variable	Type	Description	Values
text	String	Twitter contents	Strings
airline_sentiment	Categorical	Travelers' feelings for six US airlines in February 2015	Positive, neutral, negative

EDA of Dataset
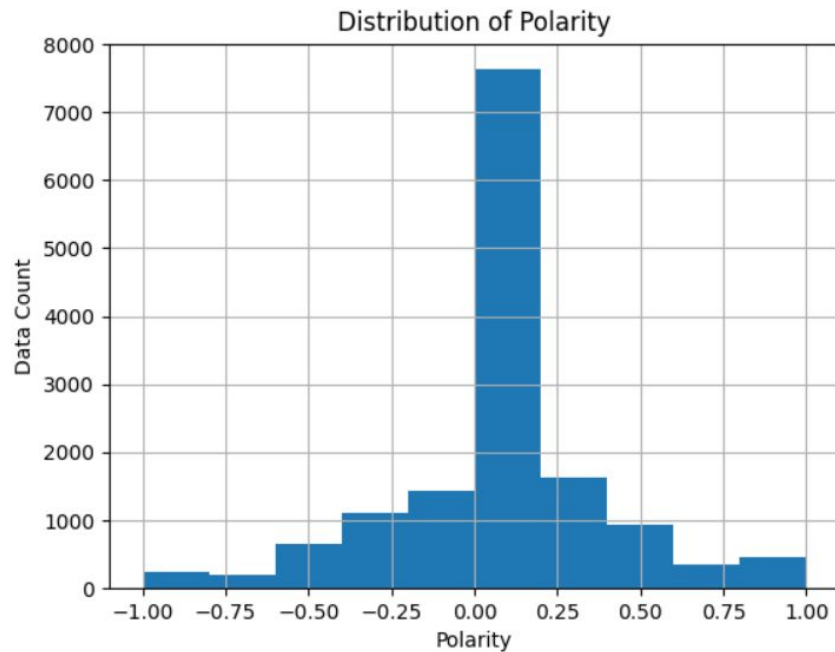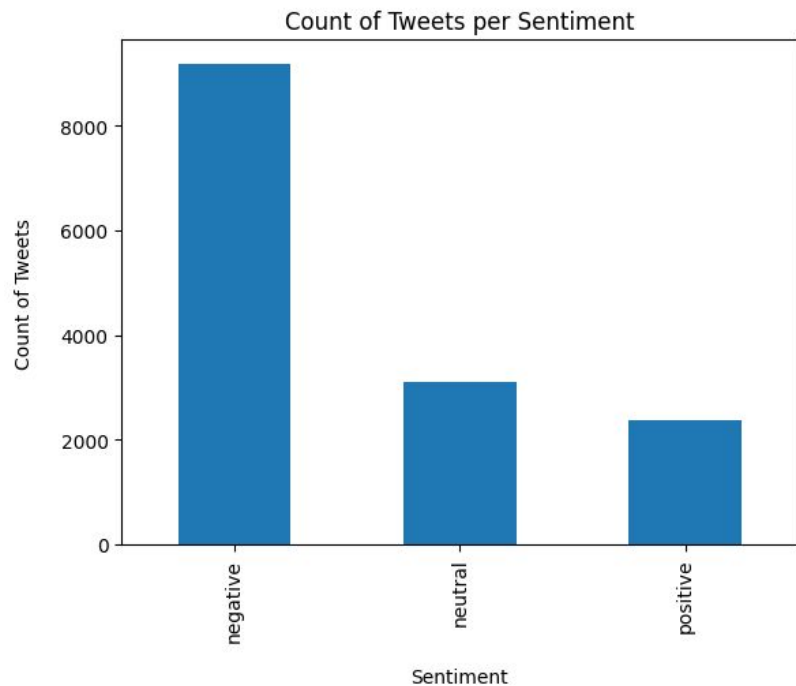
EDA of Dataset
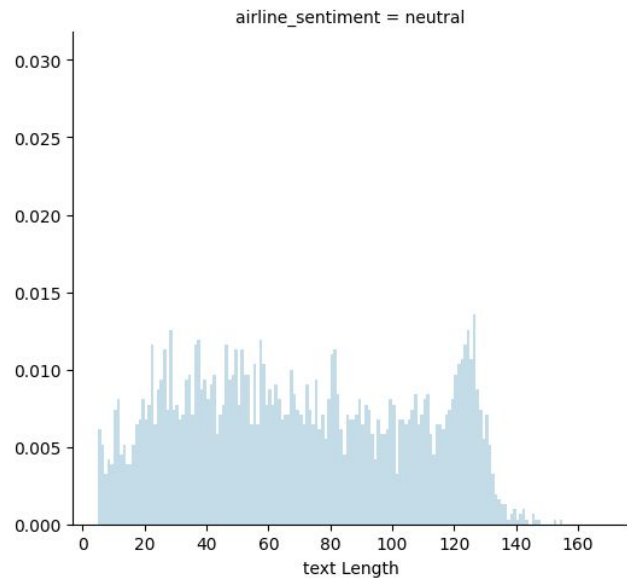
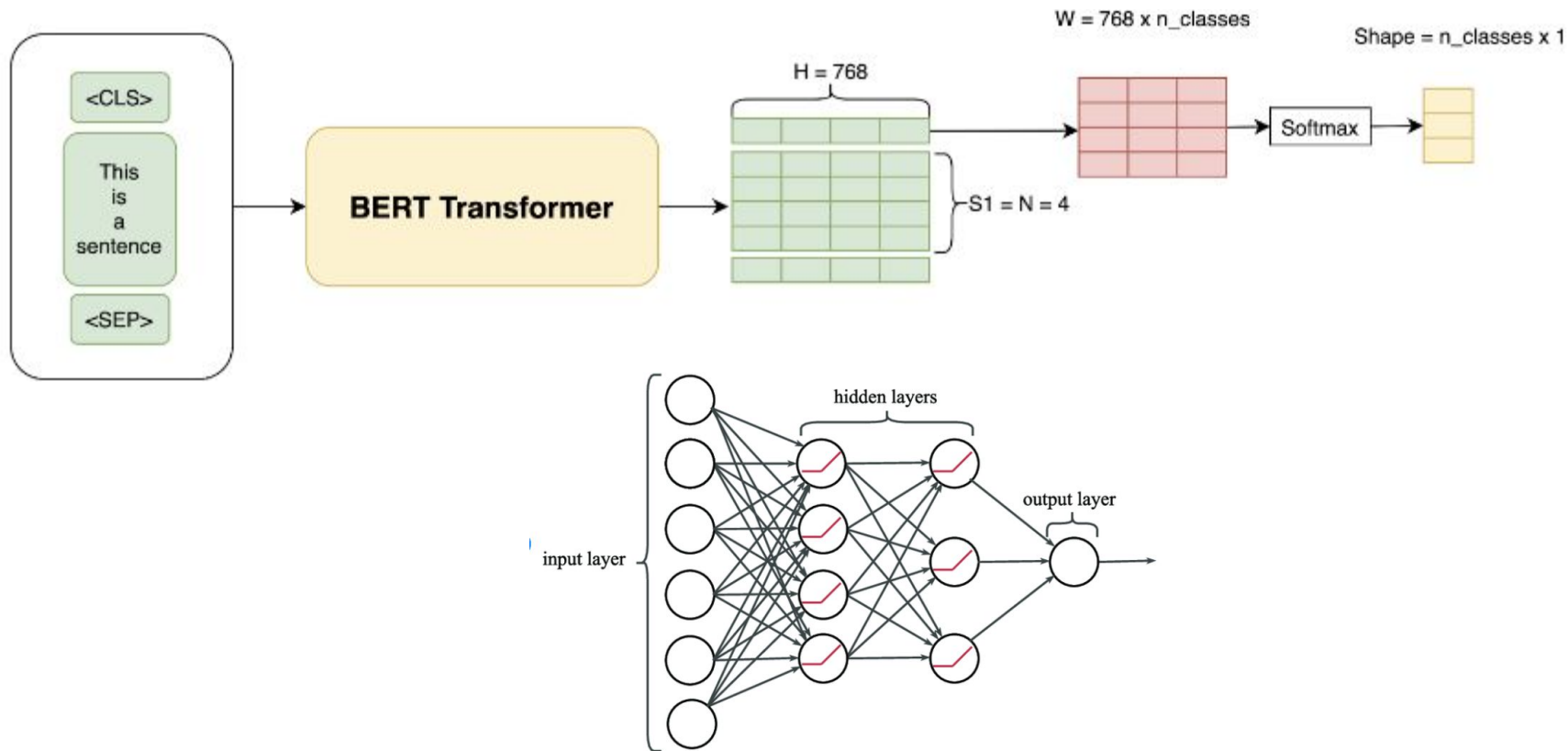# Data preprocessing

Text:

- Data cleaning
- Limitations
- Tokenization/encoding through BERT

Labels:

- One-hot encoding for multiclass
- Applications of label encoding

NLP models

Experimental setup

- Parameter setup
- Performance Judgment

      -Train accuracy and loss

      -Post-hoc analysis

      -Model interpretation

Model Results
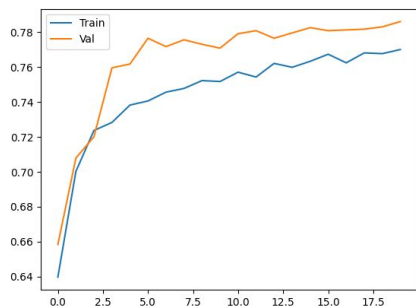
MLP Head:
    Epoch 20
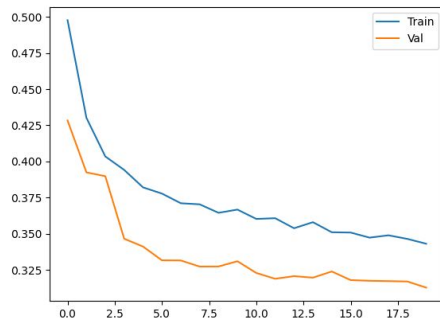   Train_loss: 0.34; Val_loss: 0.31
    Train_acc: 0.77; Val_acc:0.78

RNN Head:
    Epoch 20
    Train_loss: 0.10; Val_loss:0.41
    Train_acc: 0.94; Val_acc:0.81

Accuracy   Loss   Accuracy   Loss
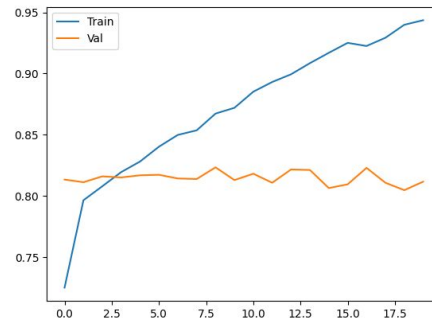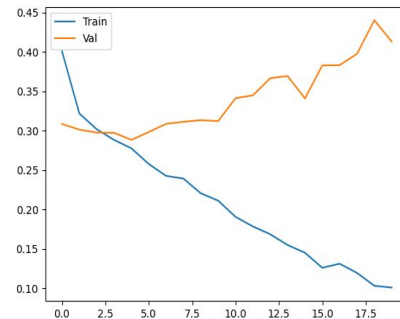
Post-Hoc Analysis

Classes/Metrics	Precision	Recall	F1-score
Negative	0.84	0.90	0.87
Positive	0.71	0.73	0.72
Neutral	0.65	0.49	0.56

	Negative	Neutral	Positive
Negative	1653	122	60
Neutral	234	302	84
Positive	85	43	345

Post-Hoc Analysis

@SouthwestAir how long does it take for my Rapid Rewards points to be credited to my account?

@USAirways can I bring a copy or photo of my baby's birth certificate for the flight? I feel uncomfortable traveling with the real thing

Model interpretation: Integrated Gradients

Baseline Image
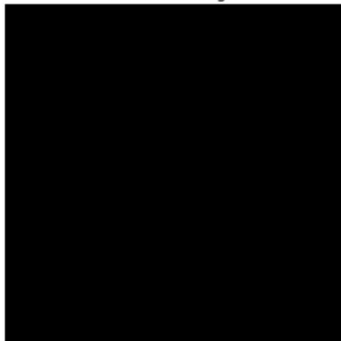
Original Image

IG Attribution Mask

Original + IG Attribution Mask Overlay

Goal: use Gradient calculation/approximation to attribute increases or decreases to the probability of belonging to a class by feature

$$IntegratedGrads_i(x) ::= (x_i - x'_i) \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha \times (x-x'))}{\partial x_i}$$

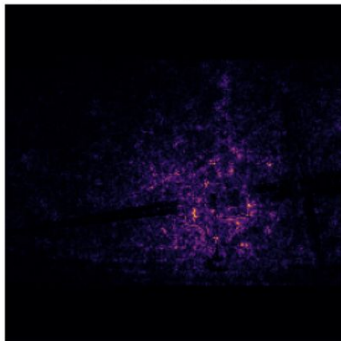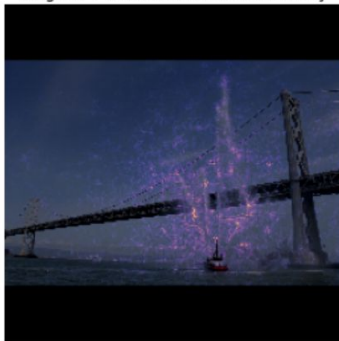# Model interpretation: Integrated Gradients

Legend: ■ Negative □ Neutral ■ Positive

True Label	Predicted Label	Attribution Label	Attribution Score	Word Importance
negative	negative (0.81)	0	0.35	[CLS] americana ##ir th ##x for losing my bag how hard is it to care for a bag w priority on it ? why do u con ##t to not care for ep s ? [SEP]
negative	negative (0.92)	0	0.69	[CLS] americana ##ir i even went to ticket counter and got no help [SEP]
positive	positive (0.86)	2	2.88	[CLS] jet ##bl ##ue good to hear th ##x for being responsive [SEP]
positive	positive (0.67)	2	3.35	[CLS] usa ##ir ##ways customer service at its best ! rachel s took great care of us at the ph ##x airport http // t co / h ##g ##7 ##ve ##q ##hg ##hy [SEP]
neutral	negative (0.79)	0	0.03	[CLS] jet ##bl ##ue 2 55 tomorrow from ric to bo ##s looking good or am i better res ##ched ##ulin ##g ? [SEP]
negative	negative (0.88)	0	0.75	[CLS] southwest ##air really ? all other carriers are staffed and you ve got a triple loop ##ed one and no employees in sight in ok ##c [SEP]
negative	neutral (0.71)	1	2.53	[CLS] southwest ##air have you considered adding the we ll call you back when we have someone free feature to your support line ? [SEP]
negative	negative (0.88)	0	0.72	[CLS] southwest ##air is your b ##wi s ##j ##d service seasonal ? wasn t part of extension called int ##l desk they didn t know want to fly in sept on sat [SEP]
negative	negative (0.92)	0	1.02	[CLS] americana ##ir it is now going to be reported to the police due to the sexual ass ##ult sad that you didn t care [SEP]
neutral	neutral (0.67)	1	2.01	[CLS] southwest ##air vin ##dict ##ive t ##k larry david works for southwest ? [SEP]

Summary and conclusion

- The MLP head out performed the RNN head (accuracy = ~79%)
- Model did a fair job of picking up the right context, but not consistently
- Improvements:
  - Auxiliary sentences
  - Domain specific text data

REFERENCES

- Twitter US Airline Sentiment | Kaggle
- https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment?datasetId=17&sortBy=voteCount
- https://medium.com/aiguys/bert-explainability-5b54cff01407
- https://captum.ai/tutorials/Bert_SQUAD_Interpret
- https://github.com/bentrevett/pytorch-sentiment-analysis/issues/87
- https://www.geeksforgeeks.org/fine-tuning-bert-model-for-sentiment-analysis/
- https://stackoverflow.com/questions/73911673/pytorch-dataloader-with-huggingface-transformer-getting-error-unable-to-creat
- https://github.com/pytorch/captum/blob/master/captum/attr/_utils/visualization.py
- https://www.theaidream.com/post/google-bert-understanding-the-architecture
- https://arxiv.org/abs/1810.04805
- https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html
- https://www.tensorflow.org/tutorials/interpretability/integrated_gradients
- https://arxiv.org/pdf/1703.01365.pdf
- https://github.com/pytorch/captum/blob/master/captum/attr/_core/layer/layer_integrated_gradients.py