# Comparison of Model Architectures for Sentiment Analysis

Group 14

Alec Ibarra
adi220000

Annette Llanas
ajl200006

Ashlee Kang
ajk200003

Syed Kabir
snk210004

01

# Research problem

"How do different model architectures compare when applied to sentiment analysis tasks?"

# Introduction

In this project, we aim to compare multiple model architectures for sentiment analysis. We'll look at traditional machine learning approaches as well as modern deep learning models and assess their strengths and weaknesses in terms of performance, training time, and complexity.

# Background

Sentiment analysis is a key task in Natural Language Processing (NLP) that identifies and extracts subjective information from text. With the rise of social media and online reviews, there's an increasing need for efficient models to analyze customer sentiment in real-time.

# Datasets

| Amazon Reviews | Twitter Airline Sentiment | YouTube Comments |
| --- | --- | --- |
| **50,000** product reviews. | **14,600** airline tweets. | **18,400** video comments. |
| Balanced classes, subset of larger dataset. | Naturally imbalanced, ~60% negative. | Naturally imbalanced, ~60% positive. |
| Source: Xiang Zhang's Google Drive | Source: Kaggle (crowdflower). | Source: Kaggle (atifaliak). |

Dataset Classes: **Positive**, **Neutral**, **Negative**

02

# Model Architectures

A brief overview of each architecture to be tested.

# Model Architectures

### Logistic Regression with TF-IDF

Simple linear model using word frequency features.

### LSTM with Pretrained Embeddings

Sequential model using GloVe to capture context.

### Text CNN

Convolutional model that learns local text patterns.

### TinyBERT (Transformer-Based)

Distilled pretrained transformer with deep contextual embeddings.

03

# Evaluation & Results

A comparison of model performance using key metrics.

# Logistic Regression with TF-IDF

| Dataset | Precision | Recall | F1-Score |
|---|---|---|---|
| Amazon Reviews | 0.64 | 0.64 | 0.64 |
| Airline Tweets | 0.79 | 0.80 | 0.79 |
| YouTube Comments | 0.75 | 0.76 | 0.75 |

Average F1-Score across all datasets: **0.7267**
Training time: **40ms** on average

# LSTM with Pretrained Embeddings

| Dataset | Precision | Recall | F1-Score |
|---|---|---|---|
| Amazon Reviews | 0.63 | 0.62 | 0.62 |
| Airline Tweets | 0.81 | 0.79 | 0.80 |
| YouTube Comments | 0.77 | 0.73 | 0.74 |

Average F1-Score across all datasets: **0.7200**
Training time: **725.7s** on average

# Text CNN

| **Dataset** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|
| Amazon Reviews | 0.60 | 0.59 | 0.59 |
| Airline Tweets | 0.69 | 0.70 | 0.70 |
| YouTube Comments | 0.62 | 0.59 | 0.60 |

Average F1-Score across all datasets: **0.6300**
Training time: **53.6s** on average

# TinyBERT (Transformer-based Model)

| Dataset | Precision | Recall | F1-Score |
|---|---|---|---|
| Amazon Reviews | 0.69 | 0.69 | 0.69 |
| Airline Tweets | 0.83 | 0.84 | 0.84 |
| YouTube Comments | 0.82 | 0.82 | 0.82 |

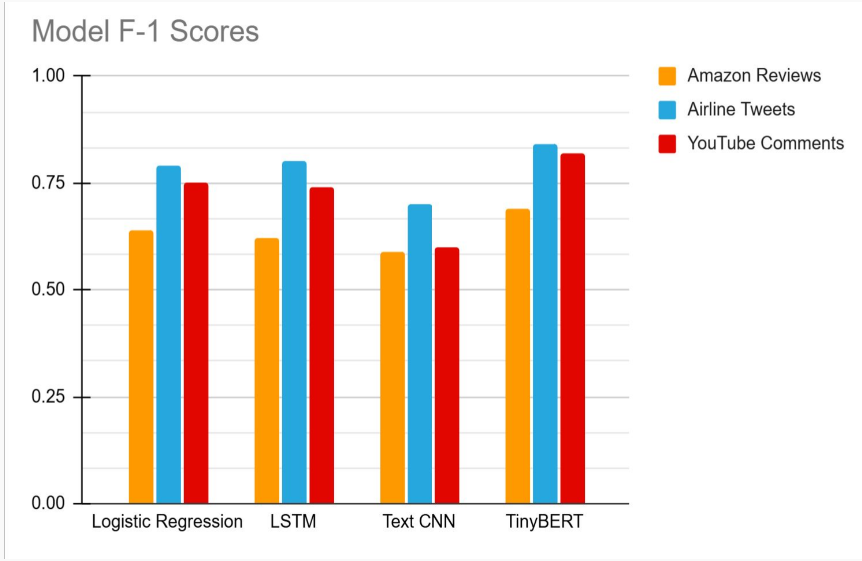
Average F1-Score across all datasets: **0.7800**
Training time: **571.5s** on average

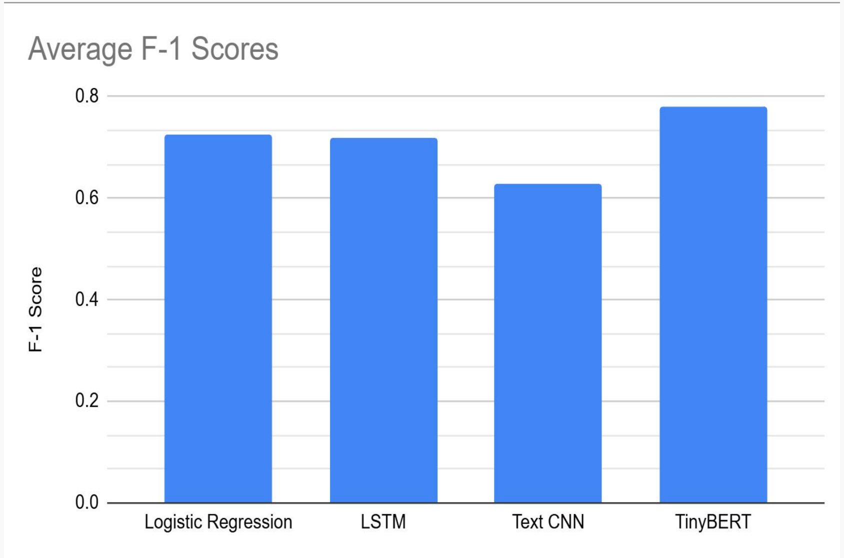
04

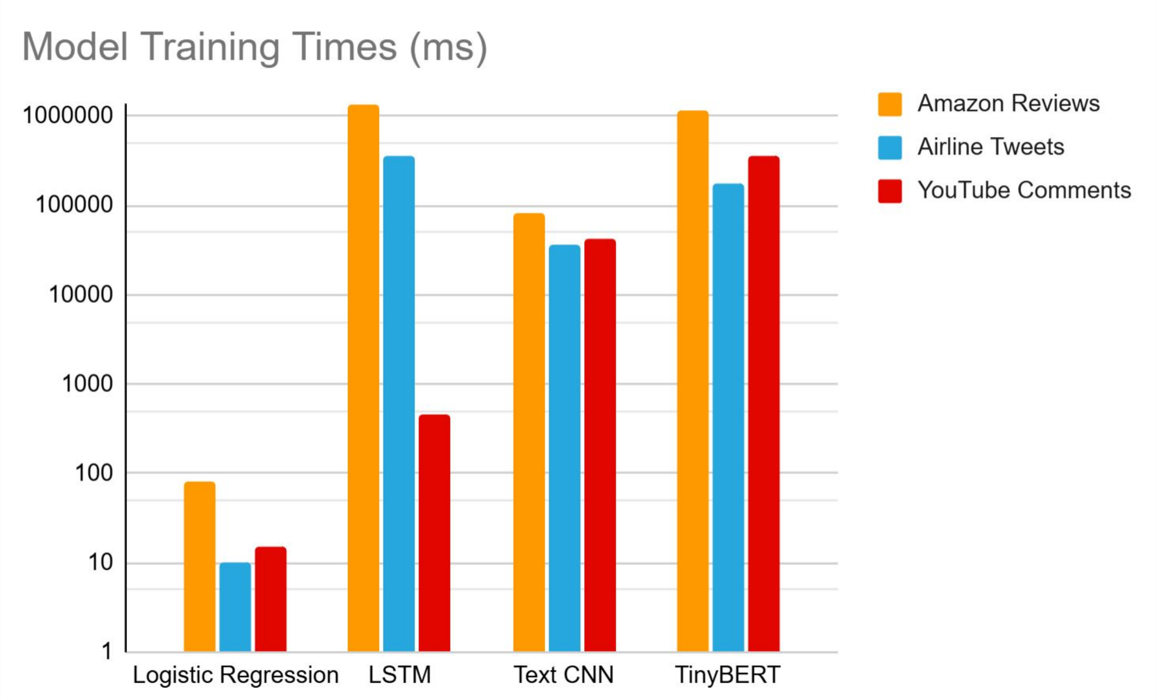
# Conclusions & Key Takeaways

A brief overview of each architecture to be tested.

# Results Analysis




Model Training Times (ms)
Amazon Reviews
Airline Tweets
YouTube Comments
1000000
100000
10000
1000
100
10
1
Logistic Regression
LSTM
Text CNN
TinyBERT

# Key Takeaways

| Lessons Learned | Future Work |
| --- | --- |
| | |

**Lessons Learned**

- Amazon review performance disparity
- Overall time and performance tradeoffs
- Simple models can match deep models on some datasets

**Future Work**

- Ensemble Methods
- Hyperparameter Tuning
- Domain-specific embeddings

# Thank you!

Any questions?