Teesside University

Computing Masters Project

NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING-BASED SENTIMENT PREDICTOR FOR AMAZON DEVICES

Presented by: Tran Thuy Vy
Student ID: S3369268
Supervisor: Dr. Francis Ho
Program: Master of Science in Data Science

Contents

Introduction

- **Project Focus:**
  - Sentiment analysis of Amazon device reviews (Alexa, Echo, Kindle, Fire TV, Kindle Tablet…).
  - Techniques: Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL).
  - Categorizes reviews into positive, neutral, or negative sentiments.
- **Key Features:**
  - Web-based tool for real-time sentiment classification.
  - Power BI dashboard for visualizing sentiment trends and product performance.
- **Motivation and Goal:**
  - Help businesses refine products and improve customer satisfaction.
  - Provide a reliable and practical solution for sentiment analysis.

Problem Statement

- **Challenges with Manual Review:**
  Manually analyzing large volumes of customer reviews is **slow, error-prone, and costly**.
- **Limitations of Traditional ML Models:**
  Traditional ML models fail to handle **complex language patterns** like sarcasm, ambiguity, and subtle emotions, leading to inaccurate sentiment classification.
- **Opportunity with Amazon Device Reviews:**
  Devices like **Alexa, Echo, and Fire TV** generate thousands of reviews daily. While these reviews provide valuable feedback, businesses require tools to process and visualize this data effectively.

- Solution:
  - ✓ Develop a **web-based sentiment analysis application** using **NLP, ML and DL techniques**. The tool will classify reviews efficiently into **positive, neutral, or negative** sentiments.
  - ✓ Additionally, a **Power BI dashboard** provides key statistics, such as the distribution of sentiments, the number of reviews analyzed, top-performing products based on sentiment and key sentiment factors. This visual representation makes the results clearer and supports data-driven decisions.

Research Questions

RESEARCH QUESTION 1

What is the most effective NLP, Machine Learning, and Deep Learning techniques for analyzing sentiment in Amazon device reviews?

RESEARCH QUESTION 2

How do Machine Learning models compare with Deep Learning models like LSTM-based RNNs in terms of sentiment classification accuracy and F1-score?

RESEARCH QUESTION 3

What are the key drivers of positive and negative sentiment for Amazon devices?

RESEARCH QUESTION 4

How can sentiment analysis insights improve product recommendations and customer satisfaction?

RESEARCH QUESTION 5

What challenges arise in building sentiment analysis tools, and how can these challenges be mitigated?

Literature Review

Key findings from the literature that provide the foundation for this project. The review focuses on sentiment analysis techniques, challenges in the field, and the effectiveness of both traditional machine learning models, including ensemble methods like LightGBM, and deep learning models such as LSTM:

Sentiment Analysis Overview

"Sentiment analysis, also called opinion mining, refers to the use of natural language processing and machine learning to extract subjective information from text, enabling automated understanding of opinions and emotions."

Pang and Lee (2008)

Traditional ML for Text Classification

"Support Vector Machines have proven to be highly effective for text categorization tasks, particularly when dealing with high-dimensional feature spaces like textual data."

Joachims (1998)

Challenges in Sentiment Analysis

"Despite advances in sentiment analysis, challenges such as sarcasm, negations, and ambiguous expressions remain significant barriers to achieving accurate sentiment classification."

Cambria et al. (2013)

Advances in Ensemble Methods

"ML techniques, including ensemble methods, have shown strong performance in tackling the challenges of sentiment analysis by automating feature extraction and improving accuracy."

Medhat, Hassan, and Korashy (2014)

LSTM for Sequential Data

"Long Short-Term Memory networks effectively capture long-term dependencies in sequential data, making them particularly well-suited for tasks involving natural language processing"

Hochreiter and Schmidhuber (1997)

Methodology

Teesside University

Data Scraping

- BeautifulSoup
- Selenium

Amazon Reviews

Data Preprocessing

1. Convert to lower case
2. Decode HTML entities
3. Tokenization
4. Lemmatization
5. Remove extra Whitespace
6. Remove Punctuation
7. Remove Number
8. Remove Stopwords

ML Model Data Preparation

- Feature Extraction: TF-IDF
- Handling Imbalance: SMOTE

DL Model Data Preparation

- Tokenizer
- Padding
- Truncating

Train – Test Splitting

Model Development

1. Logistic Regression
2. Naïve Bayes
3. Random Forest
4. Linear Support Vector Machine (LinearSVC)
5. Light Gradient Boosting Machine (LightGBM)
6. Long Short – Term Memory (LSTM)

- **Model Evaluation**
- **Model Selection**
- **Hyperparameter Tuning**
Final Model

Model Development

- Flask API backend
- Web Interface

Data Visualization

Power BI dashboard

Evaluation and Results

Model	Accuracy	F1-Score
LightGBM	0.79	0.79
Naive Bayes	0.77	0.78
Logistic Regression	0.75	0.75
LSTM	0.75	0.75
Random Forest	0.74	0.74
LinearSVC	0.74	0.74

The **LightGBM model** underwent final hyperparameter tuning

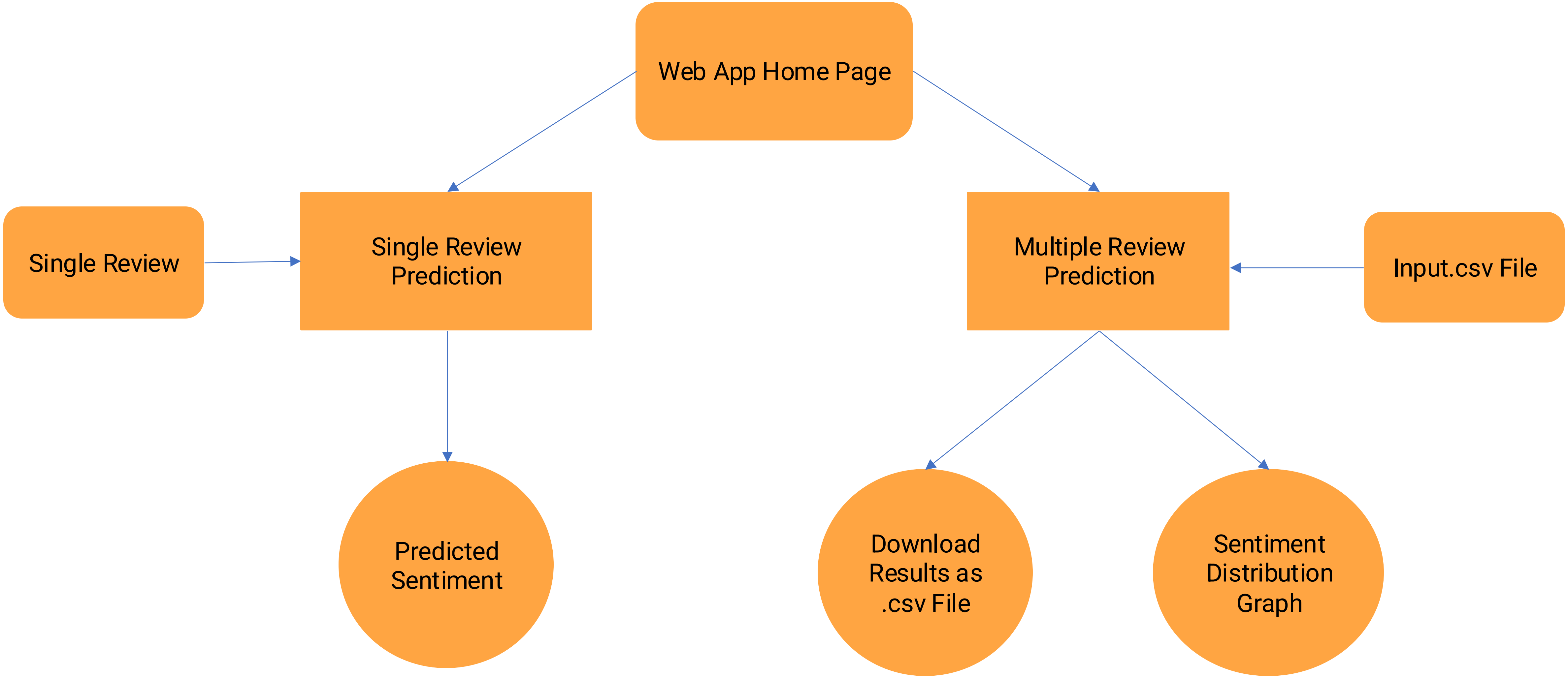
Metric	Before Tuning	After Tuning
Negative F1-Score	0.74	0.75
Neutral F1-Score	0.7	0.71
Positive F1-Score	0.86	0.87
Accuracy	0.79	0.8
Macro Avg F1-Score	0.77	0.77
Weighted Avg F1-Score	0.79	0.8

LightGBM Prediction on Unseen Reviews

	Review	Prediction	Negative proba	Neutral proba	Positive proba
0	This product is perfect!	Positive	0.0750	0.0670	0.8580
1	I don't recommend this product, it doesn't work.	Negative	0.6954	0.2637	0.0410
2	The laptop is okay, nothing special.	Neutral	0.0878	0.8020	0.1102
3	I love my blink cameras and it's handy having our subscription plan setup through Amazon to make it easy to purchase the plan!	Positive	0.0018	0.0076	0.9906
4	Terrible experience with this device.	Negative	0.8524	0.1057	0.0419
5	The product is ok. Camera can be better. Little heavy.	Neutral	0.0646	0.8570	0.0784
6	I love it! Best purchase ever. I like it because I can see all around my front yard.	Positive	0.0084	0.0125	0.9791
7	Product is ok ok only. Main cons is volume is low for calls . Especially in MI phones. So consider this	Negative	0.6469	0.1717	0.1814
8	Breakdown after 3 weeks. Don't buy, really lousy customer service, no refund or gift code!!!	Negative	0.9044	0.0272	0.0684
9	Pc that supports games like League Of Legends with the RTX 3080 ti GE and the ryzen 9 3900X.	Positive	0.1248	0.3833	0.4919

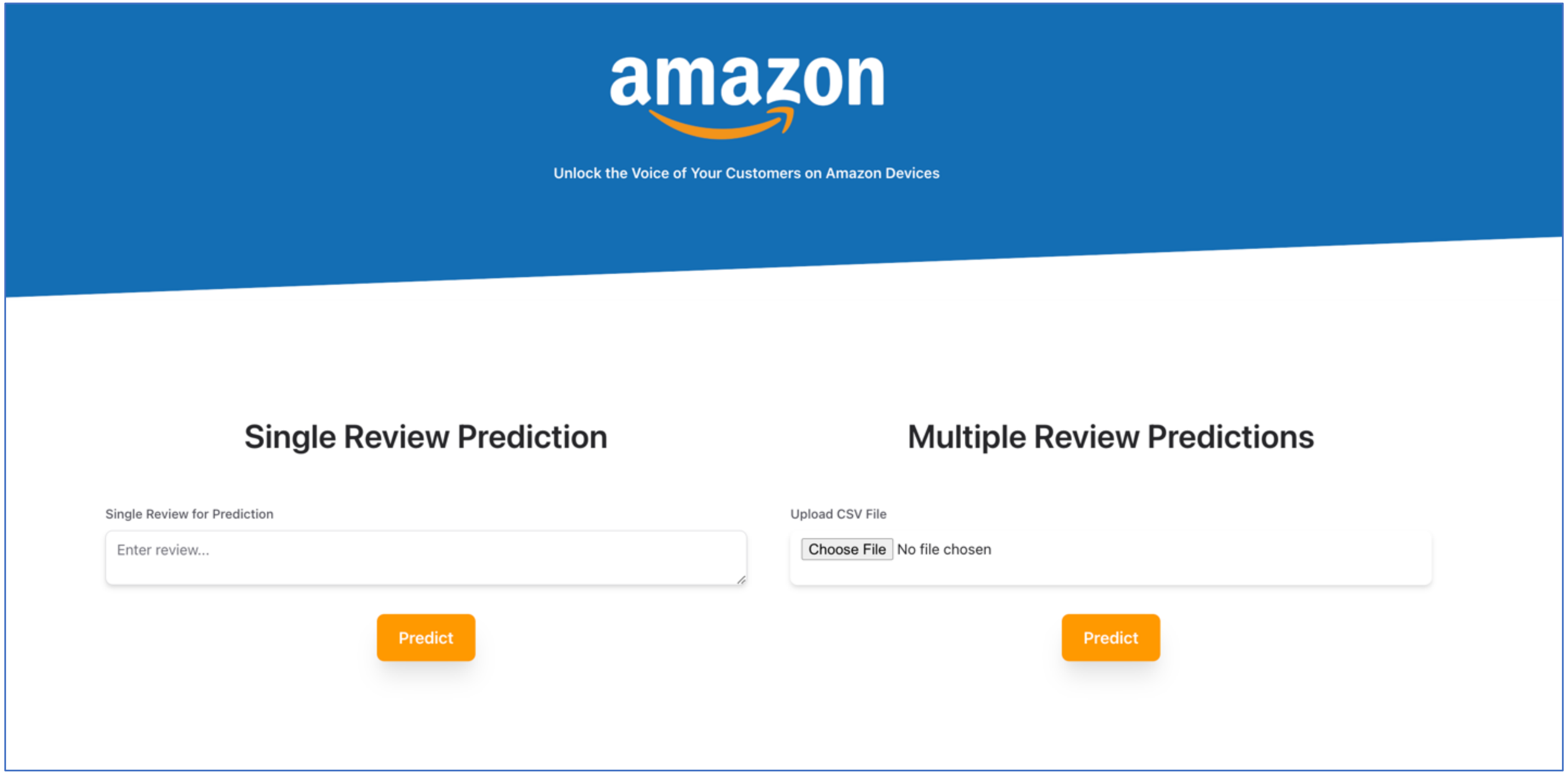
- Both F1-score and accuracy are chosen as primary metrics for evaluating and selecting the best-performing model.
- LightGBM achieved the highest accuracy (0.79) and F1-score (0.79), making it the best-performing model.
- After final hyperparameter tuning, accuracy increased from 0.79 to 0.80, the Weighted Average F1-Score improved to 0.80.
- The model performed well in predicting extreme and clear sentiments especially for highly **positive** or **negative** reviews.
- It effectively identified and balanced **neutral** reviews.
- However, it struggled **with mixed sentiment** reviews, where the presence of both pros and cons posed challenges for accurate classification. (Review #7)

Evaluation and Results – Web App Performance



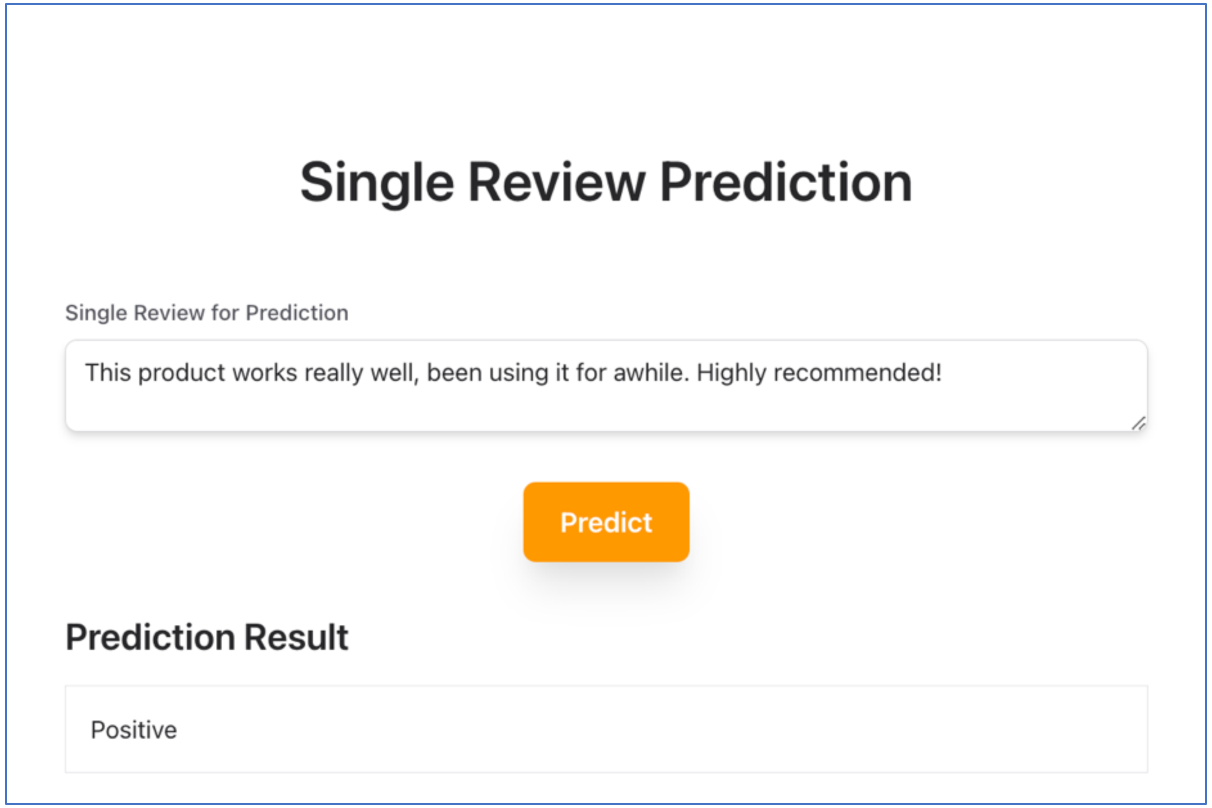
Web Interface Workflow for Sentiment Prediction

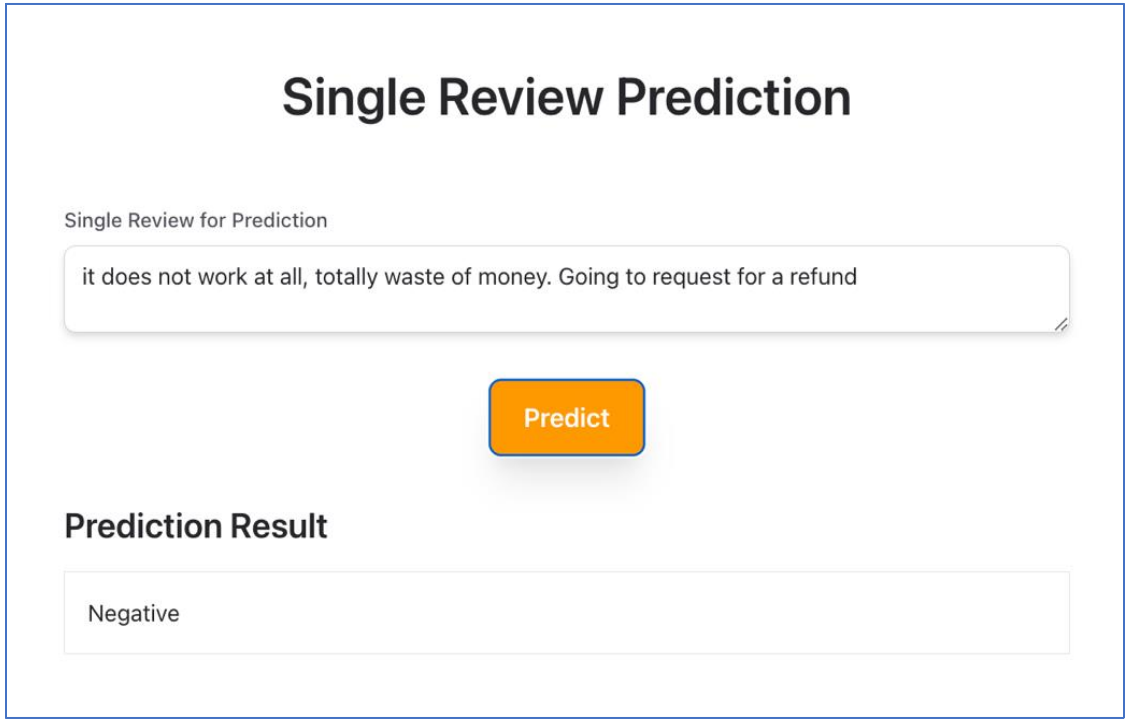
Evaluation and Results – Web App Performance



Home Page

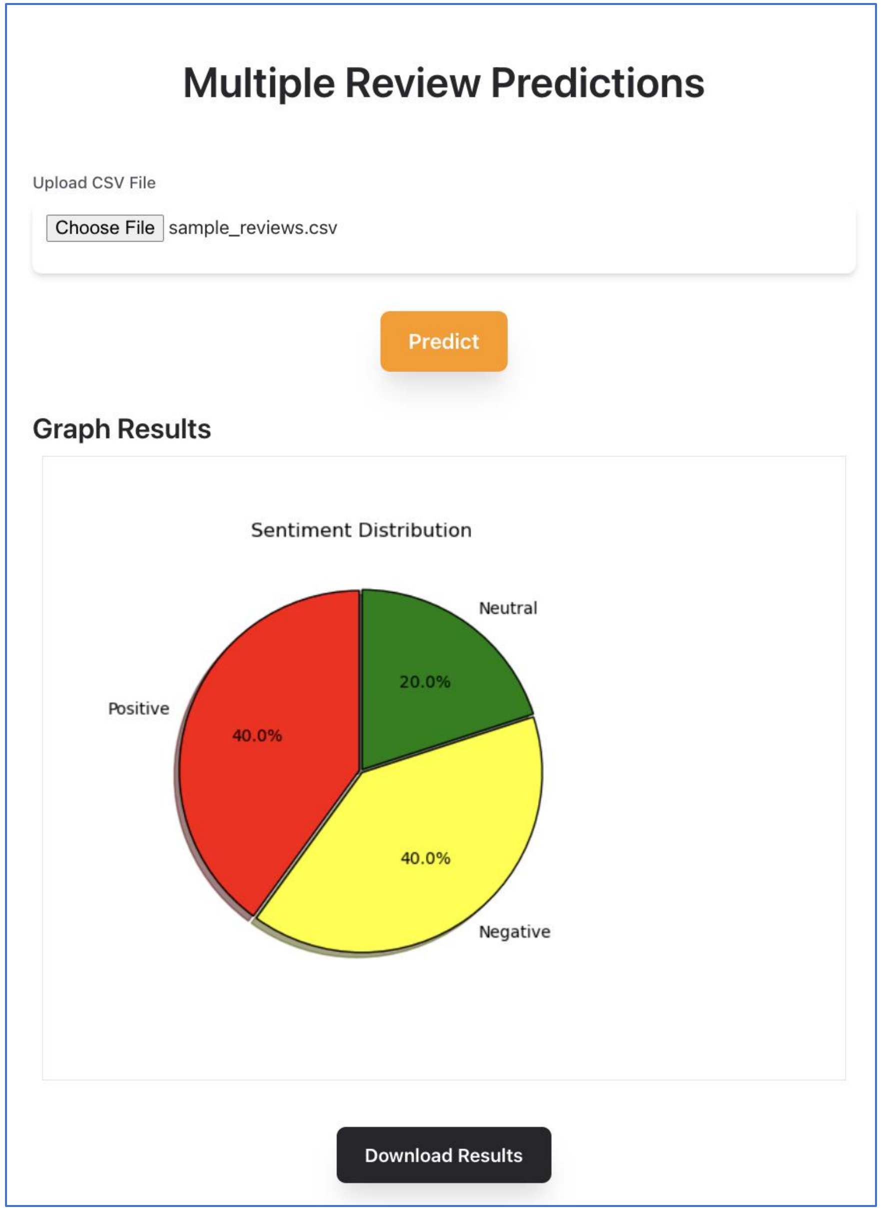
Evaluation and Results – Web App Performance





Single Review Prediction Interface

Evaluation and Results – Web App Performance



Multiple Review Prediction Interface

How it works:

- **Input:** Users upload a **.csv** file with a column named **"Review"** containing the text reviews.
- **Output:**
  - **Downloadable Results:** A .csv file containing the original reviews along with their predicted sentiments.
  - **Sentiment Distribution Graph:** A pie chart showing the proportion of positive, neutral, and negative reviews.

Evaluation and Results – Web App Performance

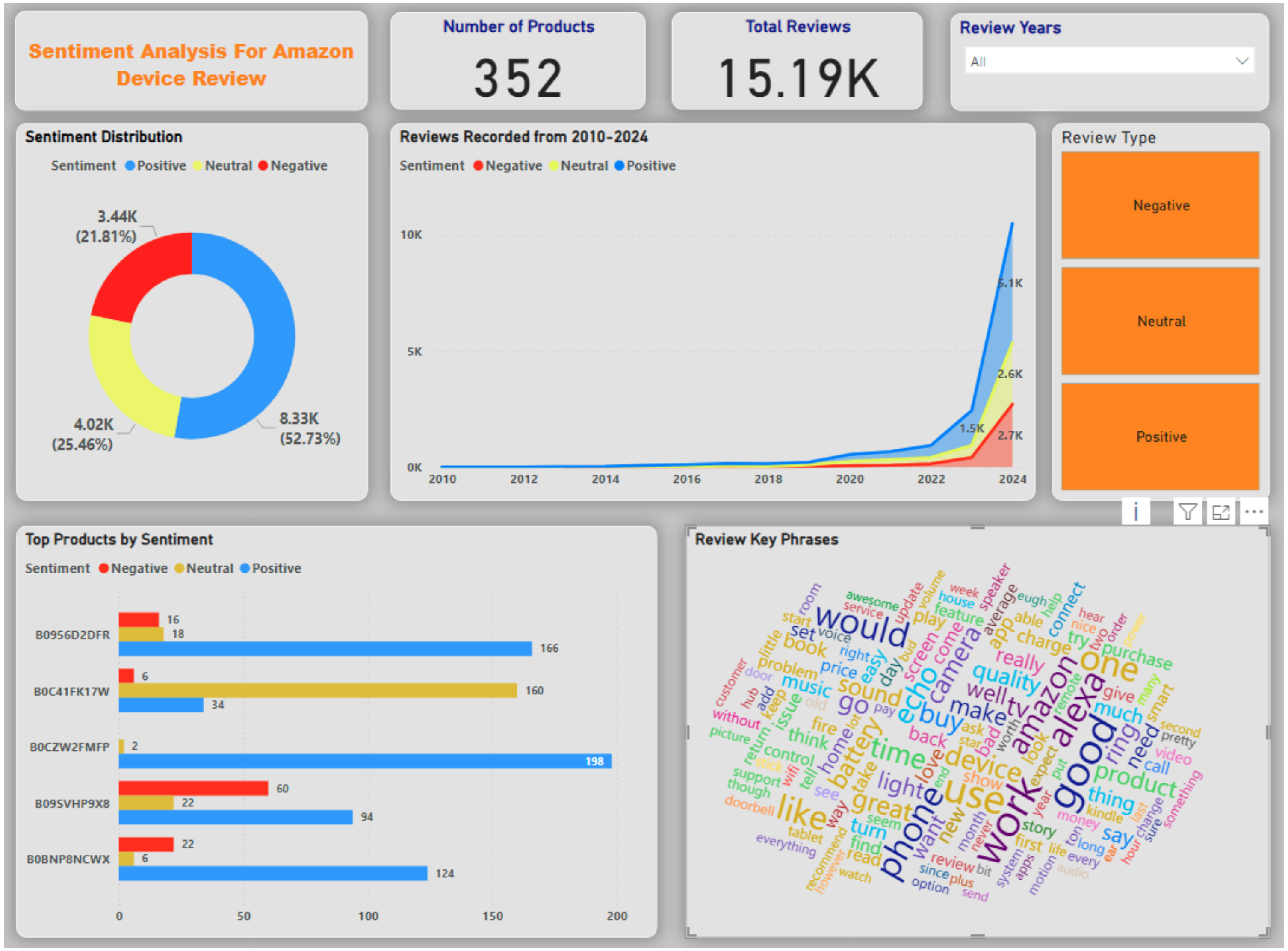
Input (*customer_reviews.csv*)

Review
Highly recommend it to anyone looking for something reliable and well-made.
The quality of this item is absolutely awful.
It's okay, but I probably wouldn't buy it again.
Fantastic! I loved every aspect of this.
Terrible experience. I want a refund.
Not bad, but not great either.
Customer service was unresponsive and unhelpful. Very disappointing experience.
I can't believe how poorly this was made.
Decent product for the price.
This is the best purchase I've ever made!

Output (*prediction.csv*)

Review	Prediction
Highly recommend it to anyone looking for something reliable and well-made.	Positive
The quality of this item is absolutely awful.	Negative
It's okay, but I probably wouldn't buy it again.	Neutral
Fantastic! I loved every aspect of this.	Positive
Terrible experience. I want a refund.	Negative
Not bad, but not great either.	Neutral
Customer service was unresponsive and unhelpful. Very disappointing experience.	Negative
I can't believe how poorly this was made.	Negative
Decent product for the price.	Positive
This is the best purchase I've ever made!	Positive

Evaluation and Results – Power BI Dashboard



Research Questions - Answers

Teesside University

RESEARCH QUESTION 1

What is the most effective NLP, Machine Learning, and Deep Learning techniques for analyzing sentiment in Amazon device reviews?

Answer:

The most effective techniques for sentiment analysis in this project were a combination of **NLP preprocessing** and the **LightGBM model**. NLP methods like **tokenization**, **lemmatization**, and **TF-IDF Vectorization** effectively prepared the text data for analysis. Among the evaluated models, **LightGBM** achieved the highest accuracy (**0.80**) and F1-score (**0.79**), outperforming traditional models like Naive Bayes and Logistic Regression. While the **LSTM** model captured sequential dependencies, its performance was lower (**0.75** accuracy), likely due to the smaller and imbalanced dataset.

RESEARCH QUESTION 2

How do ML models compare with DL models like LSTM-based RNNs in terms of sentiment classification accuracy and F1-score?

Answer:

For this project, **LightGBM** (a ML model) outperformed **LSTM** (a DL model) in sentiment classification. LightGBM achieved the highest accuracy (**0.80**) and F1-score (**0.79**), while LSTM had a lower accuracy of **0.75**, similar to Logistic Regression. Traditional ML models like LightGBM were faster, more efficient, and better for the smaller, imbalanced dataset used in this project. On the other hand, LSTM excelled at capturing sequential patterns and context but required larger datasets and more computational resources to perform at its best.

RESEARCH QUESTION 3

What are the key drivers of positive and negative sentiment for Amazon devices?

Answer:

Positive sentiment is driven by factors such as **functionality**, **ease of use**, and **performance**, as reflected in frequent words like *"work"*, *"use"*, *"great"*, … Customers often express satisfaction with the seamless integration of devices and the overall quality, indicated by words like *"love"* and *"fantastic."* Negative sentiment primarily stems from issues related to **hardware reliability**, **battery life**, and **customer service**. Common words like *"phone,"* *"battery,"* and *"charge"* highlight technical problems, while terms such as *"refund"* and *"return"* suggest dissatisfaction with service policies or product failures.

Research Questions - Answers

Teesside University

RESEARCH QUESTION 4

How can sentiment analysis insights improve product recommendations and customer satisfaction?

Answer:

Sentiment analysis insights can improve **product recommendations** by identifying frequently praised features (e.g., "Alexa integration" or "ease of use") and tailoring suggestions to customer preferences. Products with consistent positive sentiment can be highlighted in recommendations, while those with recurring complaints can be avoided or flagged for improvement. Additionally, addressing common negative feedback, such as technical issues or customer service problems, directly enhances **customer satisfaction**. Businesses can also use these insights to personalize customer interactions, prioritize product upgrades, and refine their marketing strategies, ultimately creating a more customer-centric approach.

RESEARCH QUESTION 5

What challenges arise in building sentiment analysis tools, and how can these challenges be mitigated?

Answer:

- **Imbalanced Dataset**: The dataset collected from Amazon reviews was imbalanced, with Negative and Neutral classes underrepresented.
  - ➢ **Mitigation**: Techniques like SMOTE was applied to ensure balanced training and improve the model's ability to predict underrepresented classes.
- **Limited Data Access**: Amazon restricts data scraping to **a maximum of 10 pages of reviews per product**, which constrained the dataset size.
  - ➢ **Mitigation**: Reviews were scraped from multiple products to expand the dataset and ensure sufficient data diversity for model training.
- **Mixed Sentiments in Reviews**: Mixed sentiments (reviews containing both pros and cons) and subtle language (sarcasm) posed classification challenges.
  - ➢ **Mitigation**: Text preprocessing techniques like tokenization, lemmatization, and TF-IDF vectorization were used to structure the data effectively, while the LightGBM model was chosen for its reliability in handling such cases.

Conclusion

Project Summary

This project successfully developed a **web-based sentiment analysis tool** for Amazon product reviews, enabling users to classify reviews as **positive, neutral, or negative**. It demonstrates the integration of **Natural Language Processing (NLP)** and **Machine Learning (ML)** techniques to automate sentiment analysis efficiently.

Key Achievements

- Successfully implemented and evaluated multiple models, with **LightGBM** selected for deployment due to its superior performance (Accuracy: **0.80**, F1-Score: **0.80**).
- Built a **user-friendly web application** for real-time sentiment classification.
- Designed a **Power BI dashboard** to provide an overview of sentiment distribution, product performance, and review trends, enhancing data visualization and user insights.

Model Insights

- Performed well in identifying clear positive and negative sentiments.
- Challenges remain in handling mixed or subtle sentiments, highlighting opportunities for further improvement.

Future Work

- **Dataset Expansion**: Scrape additional reviews across more products and explore external datasets to improve model generalization and reduce data limitations.
- **Explore Advanced Embedding Methods**: Implement **Word2Vec** or **GloVe embeddings** to improve feature representation by leveraging pre-trained semantic and syntactic information.
- **Higher-Order N-Grams**: Experiment with **trigrams** or higher-order n-grams to capture more contextual information in reviews, while managing dimensionality through term frequency reduction.
- **Improving Model Performance**: Address class imbalance through techniques like oversampling, class weighting, or synthetic data generation, and refine predictions for mixed sentiments.
- **Focus on Neutral Class**: Enhance feature representation and dataset balance for the neutral class to reduce confusion and improve precision.
- **Analyze Misclassifications**: Investigate misclassifications to identify and address model weaknesses effectively.

Ethical Considerations

Data Collection

Reviews were scraped from Amazon's website following ethical guidelines and adhering to its terms of service, including the limitation of scraping a maximum of 10 pages per product.

Data Privacy

Only publicly available reviews were used, and no personally identifiable information was collected or stored during the data scraping process.

Imbalanced Dataset

Addressed potential bias caused by an imbalanced dataset to ensure fair and accurate sentiment classification across all classes.

References

- Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. Foundations and Trends® in information retrieval, 2(1–2), pp.1-135.
- Joachims, T., 1998, April. Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Cambria, E., Schuller, B., Xia, Y. and Havasi, C., 2013. New avenues in opinion mining and sentiment analysis. IEEE Intelligent systems, 28(2), pp.15-21.
- Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), pp.1093-1113.
- Schmidhuber, J. and Hochreiter, S., 1997. Long short-term memory. Neural Comput, 9(8), pp.1735-1780.

thanks!