Sentiment Analysis on Amazon Product Reviews using Deep Learning Models.

MSc Research Project
Data Analytics

Annjoys Robert
StudentID: 22137459

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Annjoys Robert
Student ID:	22137459
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	14/12/2023
Project Title:	Sentiment Analysis on Amazon Product Reviews using Deep Learning Models
Word Count:	6214
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Annjoys Robert
Date:	14th December 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Sentiment Analysis on Amazon Product Reviews using Deep Learning Models

Annjoys Robert
22137459

Abstract

This study introduces an approach, to analyzing customer sentiments in the realm of e commerce titled "Sentiment Analysis on Amazon Product Reviews using Deep Learning Models". The main focus of this research is to extract insights from customer reviews by incorporating deep learning techniques. By utilizing a dataset consisting of over 568,000 reviews this study applies Convolutional Neural Networks (CNN) Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) networks to classify sentiments. The LSTM model achieves an accuracy rate of up to 93.12%. Furthermore the project explores the effectiveness of BERT (Bidirectional Encoder Representations from Transformers) for sentiment analysis. Highlights its ability to predict sentiment in context despite its intensity. This research not contributes significantly to sentiment analysis in e commerce. Also sheds light on the application and limitations of various deep learning models in natural language processing tasks.

1 Introduction

The rise of retail has revolutionized how consumers shop by providing access to a wide range of products and services. Central to this shift is the prevalence of user generated product reviews, on e commerce platforms. These reviews serve a purpose; sharing user experiences and influencing the purchasing decisions made by others.
Among the platforms available Amazon stands out as one of the largest global marketplaces. It hosts a range of product reviews, from a customer base covering numerous product categories. These reviews contain customer opinions and sentiments offering insights into consumer preferences and experiences. According to an article published by Forbes titled "The Power of Online Reviews in 2022 " 95% of shoppers read reviews before making a purchase highlighting the significant influence of user generated content on consumer behavior.

However the sheer quantity and unstructured nature of these Amazon reviews pose a challenge. For both consumers looking to make informed buying decisions and sellers aiming to understand customer feedback navigating through such an amount of data can be overwhelming. To tackle this issue our research focuses on utilizing Sentiment Analysis—a technique within natural language processing. As detailed in the report "Making Unstructured Data Usable" by Harvard Business Review Sentiment Analysis plays a role in transforming text into actionable insights.

Sentiment Analysis is designed to analyze product reviews at a level by breaking them down into aspects or features. These aspects can encompass factors such, as quality, price, durability and usability.

Sentiment Analysis evaluates the sentiment attached to each aspect determining whether it is positive, negative or neutral. This approach provides information compared to traditional methods as highlighted in a study called "Advancements, in Sentiment Analysis for Business Intelligence" published in the Journal of Big Data. It allows consumers to gain a nuanced view of customer feedback by not understanding the sentiment towards a product but also pinpointing specific attributes that receive appreciation or criticism.

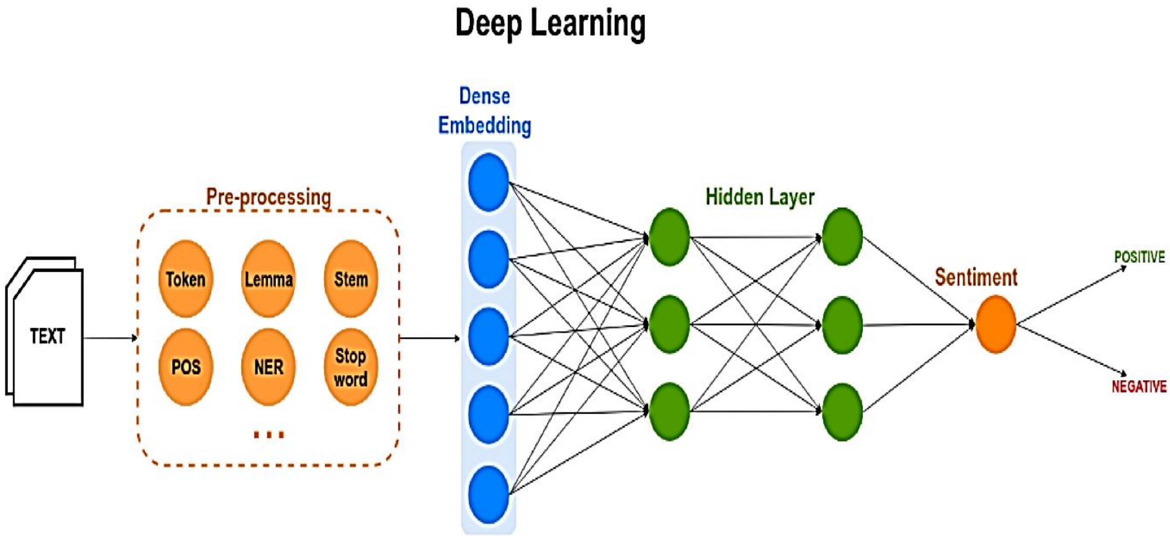

Figure 1: Phases of Sentiment Analysis.

From a business perspective Sentiment Analysis offers benefits for sellers and companies operating on Amazon. By identifying aspects that generate negative sentiments businesses can tailor their strategies for product improvement, marketing and customer engagement. This targeted approach contributes to customer satisfaction and loyalty while providing businesses with an edge in the ever evolving online marketplace. These insights are highlighted in an article by McKinsey & Company titled "Leveraging Consumer Sentiment Analysis, in Retail".

Our research aims to leverage the capabilities of Sentiment Analysis when analyzing Amazon product reviews. Through the implementation of learning models and techniques we seek to achieve a comprehensive understanding of customer sentiment. This involves the process of evaluating the strengths and weaknesses of products based on feedback, from consumers as highlighted in an article called "Deep Learning and Customer Sentiment" by MIT Technology Review. Our main focus is to contribute towards an and empowered online shopping experience on Amazon. Through insights into customer sentiment our research not assists individuals in making better purchasing decisions but also enables sellers to refine their products and strategies thereby improving the overall quality of the e commerce ecosystem. This aligns with the vision presented in the report titled "Shaping the Future of Retail for Consumer Industries" by the World Economic Forum.

2 Research Question

"How can deep learning models effectively analyze and categorize sentiments within product reviews across categories on Amazon?"

The core research question addressed in this project revolves around leveraging learning techniques to enhance the accuracy and efficiency of sentiment analysis in e commerce. The objectives involve employing and comparing deep learning models including CNN, RNN, LSTM and BERT using a dataset consisting of, over 568,000 reviews from e commerce platforms. The hypothesis states that these advanced models will provide nuanced and accurate sentiment analysis compared to methods.

This research paper makes a contribution, to the literature by showcasing how deep learning models can be applied in sentiment analysis within the e commerce industry. This is particularly important as consumer feedback plays a role in this sector. The study also offers insights into the challenges and computational requirements of implementing these models.

3 Related Work

3.1 Sentiment Analysis in the field of E-commerce

Sentiment Analysis has gained attention within the realm of e commerce. The research has progressed from general sentiment analysis to a nuanced approach called Aspect Based Sentiment Analysis (ABSA) which aims to extract sentiment associated with aspects from user reviews. For example Liang et al. (2020) presented an ABSA model utilizing machine learning techniques effectively identifying product aspects and their corresponding sentiments in reviews. However their approach faced limitations in terms of scalability across product categories thereby restricting its applicability in e commerce settings.

Another notable study conducted by Al Ghuribi et al. (2020) employed deep learning algorithms for ABSA achieving accuracy in sentiment prediction. Their model excelled in handling datasets. Encountered difficulties when detecting implicit expressions of sentiment which are common occurrences in customer reviews. This observation underscores the need for language processing models capable of comprehending nuanced expressions.

3.2 Deep Learning Techniques in Sentiment Analysis

Deep learning has had a significant impact, on sentiment analysis by providing more precise and context aware insights. A study conducted by Wang, Y., Wang, Z., Zhang, D., & Zhang, R. (2019) utilized Convolutional Neural Networks (CNNs) to classify sentiment and achieved results in terms of precision and recall. However their model proved to be less effective when dealing with texts where context plays a role.

On the hand RNNs and LSTMs have shown promise in preserving information, in sequences. Vo, A., Nguyen, Q., & Ock, C. (2018) conducted research that demonstrated the effectiveness of LSTMs in capturing long term dependencies in text, which's vital for understanding sentiment. While these models excel at retaining context they often face challenges due to complexity. Longer training times.

Recent advancements include the utilization of Transformer models like BERT for sentiment analysis. Studies have indicated that BERTs bidirectional approach provides an understanding of context and semantics (Roul, R. & Arora K., 2019). However one primary limitation of BERT and similar models is their resource nature as they require computational power. This can pose a barrier to adoption.

3.3 Summary and Justification for Research

In summary and justification for research, on Sentiment Analysis existing literature showcases advancements in deep learning techniques application. While previous studies have provided a foundation they do have some limitations. These include issues, with scalability, difficulties in interpreting sentiments and computational inefficiencies. Our research aims to address these limitations by developing an efficient ABSA (Aspect Based Sentiment Analysis) model that is also contextually aware. To achieve this we propose an approach that combines the strengths of deep learning techniques. We leverage the awareness of LSTMs and RNNs along with the efficiency and scalability of CNNs and the advanced semantic understanding of BERT. By integrating these methods our model aims to offer a nuanced analysis of customer sentiments in Amazon product reviews. This research contributes to the existing knowledge in this field. Has applications within the e commerce domain.

4 Methodology

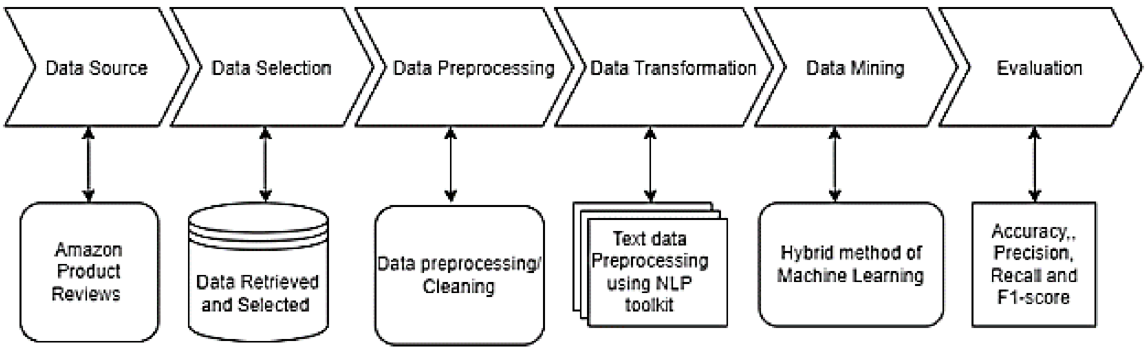

Figure 2: Process flow for Sentiment Analysis

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Databases (KDD) is a data mining method. It involves stages; gathering data cleaning it transforming it and applying algorithms to uncover meaningful patterns. Figure 2 outlines how this KDD process flow is specifically tailored for sentiment analysis. It demonstrates how data progresses from its form to extracting insights.

4.1 Data Collection

The data was collected from Amazon product reviews across categories. The dataset consists of reviews of lengths and complexities to ensure an analysis. To maintain privacy all personal information has been removed from the data. The dataset is structured in rows and columns, where each row represents a product review (see Figure 3).

Column	Description
Id	A unique identifier for each review.
ProductId	The specific identifier of the product being reviewed.
UserId	The identifier of the user who posted the review.
ProfileName	The profile name or nickname of the reviewer.
HelpfulnessNumerator	The number of users who found the review helpful.
HelpfulnessDenominator	The total number of users who rated the review's helpfulness.
Score	The rating given to the product by the reviewer, usually on a scale (like 1-5).
Time	The timestamp when the review was posted, likely in UNIX format.
Summary	A brief headline or summary of the review.
Text	The detailed text of the review.

Figure 3: Content of the Dataset

This structure enables, in depth analysis, including assessing user satisfaction through scores evaluating review helpfulness and identifying trends over time.

4.2 Data Preparation and Data Preprocessing

Data preparation and preprocessing are steps in any data analysis process especially when dealing with a dataset like the one described that focuses on sentiment analysis of product reviews. These steps involve procedures to clean and transform data into a suitable format for analysis. In this discussion I will provide insights into these steps as applied to the given dataset. The dataset contains reviews with attributes such, as identifiers, timestamps, textual content and user ratings. Understanding these aspects guides data preparation steps.

To perform this process effectively essential Python libraries like `numpy` `pandas` `nltk` `keras` `seaborn` and `matplotlib` are initially imported. These libraries are crucial, for a variety of tasks including manipulating data creating visualizations implementing machine learning algorithms and working with natural language processing. In order to work with the dataset, which is loaded into a pandas DataFrame from a CSV file we have columns such as `ProductId` `UserId` `ProfileName` `HelpfulnessNumerator` `HelpfulnessDenominator` `Score` `Time` `Summary` and `Text`. To get an understanding of the data we start with exploration techniques like using `.head()` `.info()` and `.isnull().sum()` to examine its structure types of data it contains and identify any missing values.

To ensure the accuracy and reliability of our analysis going forward it is crucial to clean the data by removing entries and addressing any missing values. These preprocessing steps are essential in establishing a foundation for applying machine learning and natural language processing models.

Additionally visualizing the data using tools like seaborn and matplotlib plays a role in gaining insights. Visualization helps uncover patterns and trends, in the dataset that may not be immediately evident when looking at data alone.
For example seaborn can be utilized to generate visuals while matplotlib provides greater flexibility in creating a wide range of static, animated and interactive visualizations.

By combining these tools and techniques we create a foundation, for data processing and model training. The organized and structured dataset allows us to effectively apply machine learning models like CNNs, RNNs, LSTMs and BERT for sentiment analysis. Our comprehensive approach involves importing the libraries loading and exploring the dataset thoroughly followed by data cleaning and deduplication. This ensures that the data is optimized to yield the accurate results from our sentiment analysis models.

Data Preprocessing

Within the sentiment analysis project scope a series of text cleaning functions are defined to preprocess the review data. These functions aim to refine and standardize the text to make it suitable for NLP models. The process includes removing HTML tags, symbols, hyperlinks while converting all text to lowercase for consistency. It also involves eliminating numbers and special characters to focus on content. Additionally we remove emojis through de emojifying as they may not be interpretable, by our models. Furthermore we unify spaces. Exclude stopwords – common words that typically do not significantly contribute to the sentiment of the text.

Stemming and lemmatization are used to reduce words to their base or root form. While stemming simply removes common word endings lemmatization is more advanced as it considers the context of the word to convert it into its base form. This approach helps capture the sentiment expressed in the reviews. The cleaning functions are systematically applied to both the 'review_text' and 'Summary' columns of the DataFrame ensuring that these text data fields are optimally prepared for analysis.

Additionally feature engineering is carried out to create variables such, as 'review_score,' which is derived from the 'Score' column and 'Num_word_review,' which counts the number of words in each review. These engineered features play a role in analysis and modeling by providing insights and contributing to the development of more accurate models. By incorporating these preprocessing and feature engineering steps this project ensures that not is the data clean and standardized but it also contains enriched features that enhance the effectiveness of sentiment analysis models.

4.3 Exploratory Data Analysis (EDA)

In performing Exploratory Data Analysis (EDA) for sentiment analysis various methods were employed to gain insights into underlying patterns and trends, within the data. We used Seaborn, a Python library, for visualizations to create representations of the review scores. This allowed us to get a understanding of how sentiments are spread in the data. We also generated word clouds for negative reviews, which helped identify the most frequently used words in each category. This approach not provided a visual grasp of the main sentiments expressed in the reviews but also highlighted the linguistic patterns associated with positive and negative sentiments.

Furthermore we conducted an analysis on review lengths and common words used in the reviews. This gave us an understanding of the characteristics such as average review length and commonly used vocabulary. These insights are crucial in developing sentiment analysis models.

Figure 4: Distribution of Review Scores

In Figure 4 you can see a count plot that visually represents the distribution of negative reviews. This plot helps us understand how sentiments are balanced within the dataset.

Figure 5: Word Cloud of the Positive words used in the Reviews.

Word Clouds: Word clouds are generated for the entire dataset, as well as separately for positive and negative reviews. These visualizations like figure 5 help in identifying the most frequent words in the reviews, giving an insight into common themes and topics.

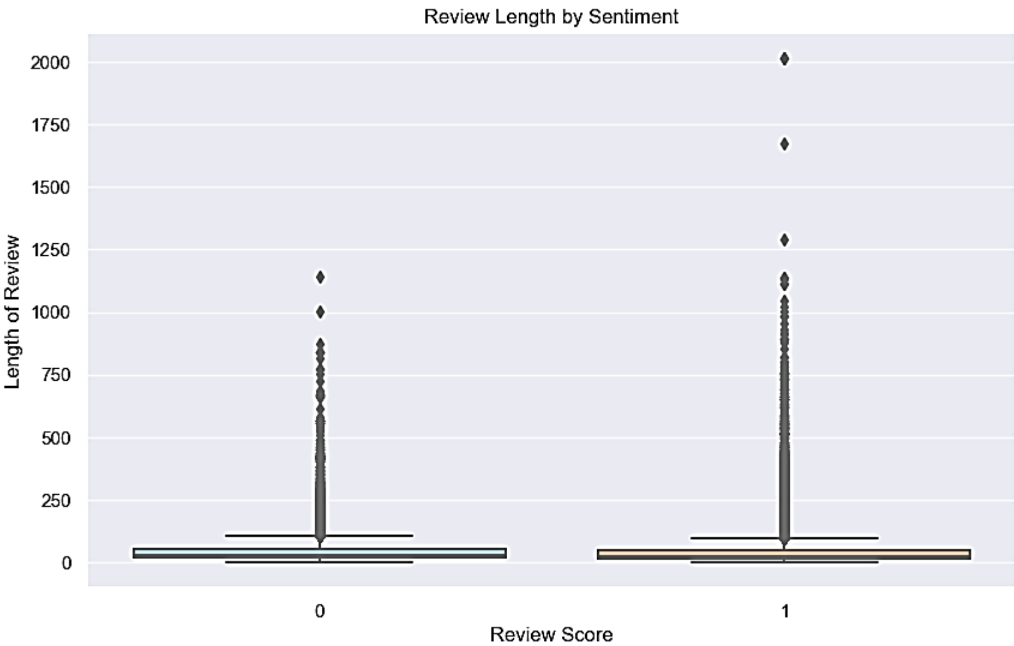

Figure 6: Review Length by Sentiment

Histogram of Review Lengths: The distribution of the length of the reviews (word count) is plotted using a histogram. This graph (figure 6) helps in understanding the verbosity of the reviews in the dataset.

Word Frequency Count: To identify and plot the used words in the dataset we utilize the Counter function from the collections library. This analysis is essential for grasping terms and phrases that frequently appear in these reviews.

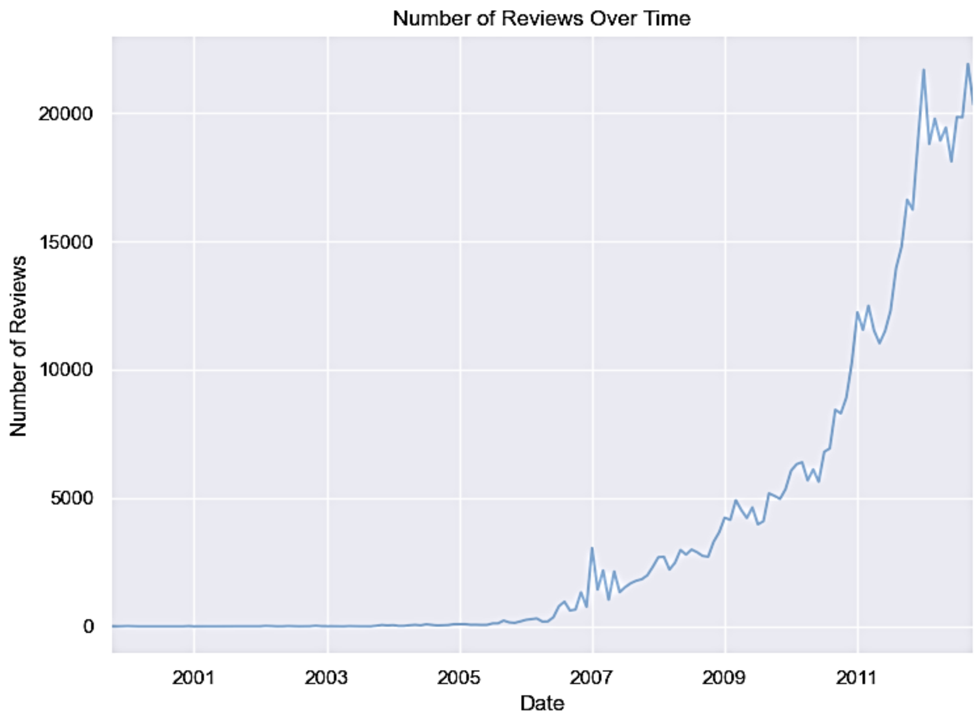

Figure 7: Number of Reviews Over time

Time-Series Analysis(Trend Over Time): The program converts the UNIX timestamp to a readable date format and then resamples the data to analyze the number of reviews and average review scores over time like in figure 7. This time-series analysis reveal trends and patterns across different time frames.

4.4 Model Implementations

When it comes to training and evaluating models on a dataset of product reviews there are steps we need to take to determine sentiment polarity. We have to preprocess the text by normalizing it and formatting the data in a way that optimizes model performance. Different models serve needs; CNNs are known for their speed and ability to detect key phrases and expressions making them ideal, for identifying important elements. On the hand RNNs and LSTMs excel at processing the nature of text capturing long term dependencies that are crucial for understanding the overall context and flow of sentiments in reviews. Lastly BERT offers understanding allowing for nuanced sentiment analysis but requiring significant computational resources.

These models each have their strengths which highlight the approaches in sentiment analysis. They strike a balance between efficiency, model complexity and their ability to pick up on linguistic cues. However they also come with limitations that require us to consider trade offs in terms of demand, complexity and language processing intricacies. This emphasizes the importance of tailoring our approach based on requirements, for sentiment analysis tasks.

4.4.1 Convolutional Neural Network (CNN)

CNNs popular, in image processing have been successfully adapted for NLP tasks because they can extract position invariant features from text. The architecture of CNNs includes an embedding layer for converting text into vectors, followed by max pooling layers. This design is effective in identifying phrases and patterns in text data. However compared to RNN based models CNNs are generally less proficient at capturing long term dependencies within text. This is because CNNs primarily focus on features, which may overlook the sequential and interconnected nature of linguistic elements in longer texts. In sentiment analysis understanding the context and narrative flow is just as crucial as identifying indicators of sentiment.

4.4.2 Recurrent Neural Network (RNN) with SimpleRNN Cells

RNNs are specifically designed to handle data making them particularly effective for analyzing text where the order and flow of words matter. In the context of sentiment analysis RNNs can capture the subtleties and progression of sentiments within a text. They process data sequentially through cells that handle one element at a time. However these cells face challenges such, as vanishing or exploding gradient problems when dealing with sequences. This limitation makes them less effective, in understanding the relationships between parts of a text. The vanishing gradient problem specifically hampers the models capacity to learn and remember information from sections of the text as it progresses through the sequence. This aspect is crucial for comprehending the sentiment expressed in reviews or discussions.

4.4.3 Long Short-Term Memory (LSTM) Network

LSTMs, which are a type of RNNs excel at handling sequences of data commonly found in text. They enhance models by incorporating LSTM cells with gating mechanisms. These gates effectively. Preserve information over extended periods thereby addressing long term dependencies in textual data more effectively. However this increased complexity requires resources making LSTMs slower to train. This can be a drawback, with datasets where there is a higher risk of overfitting due to the deep and complex nature of the model. Striking a balance between LSTMs ability to process data and its computational intensity is an important consideration when applying it for sentiment analysis.

4.4.4 BERT

BERT (Bidirectional Encoder Representations from Transformers) represents a breakthrough in NLP due, to its understanding of language context and nuances. BERT differs from models by utilizing a transformer architecture, which allows it to analyze words in the context of a sentence rather, than sequentially. This approach enables an understanding of language subtleties. However BERT models require power making them less practical for simpler NLP tasks. Additionally their complexity poses challenges in terms of interpretability making it difficult to comprehend the decision making or classifications involved.

5 Design Specification

The sentiment analysis model has the objective of categorizing data into sentiments such as positive, negative or neutral. This model finds application in datasets like customer reviews, social media posts and survey responses providing insights into opinion and customer satisfaction.

Architecture and Framework

Data Preprocessing:

Within the realm of Natural Language Processing (NLP) data preprocessing plays a role in preparing raw text data for machine learning models. The process involves steps such as text normalization where text is broken down into units like words or phrases, through tokenization; lemmatization which reduces words to their base or dictionary form; and stemming that trims words to their root forms.
These steps are crucial, for simplifying the text and standardizing variations of the word, which helps the models process the text more effectively.

Another important aspect of preprocessing involves removing stop words. These are used words like 'the' 'is and 'in' that are often irrelevant to the intended meaning and can clutter the analysis. By eliminating these words we streamline the dataset allowing the model to focus on aspects of the text.

Additionally handling emojis and symbols plays a role in NLP tasks during preprocessing. Emojis and symbols can carry contextual significance especially in social media posts and product reviews. Properly understanding these symbols is essential for accurate sentiment analysis. Having clean and standardized input data serves as a foundation for model training. Models trained on noisy data may perform poorly since they can get confused by irregularities and inconsistencies, within the text. Preprocessing ensures that input data is consistently formatted enhancing the models ability to learn from it and make predictions.
Figure 8; Key Elements, in Data Preprocessing

Aspect	Description
Technique	Natural Language Processing (NLP)
Tools	NLTK for Python
Functionality	Text normalization (tokenization, lemmatization, stemming), stop-word removal, handling of emojis and symbols
Requirement	Clean and standardized input for effective model training

Figure 8: Key aspects in Data Preprocessing

Model Training and Evaluation:

In the field of machine learning sentiment analysis relies on types of models each with its own strengths. Convolutional Neural Networks (CNN) excel at extracting features from text which makes them efficient for identifying phrases and patterns. However their focus on context might limit their effectiveness in understanding the sentiment of longer texts. On the hand Recurrent Neural Networks (RNN) and their advanced version known as Long Short Term Memory Networks (LSTM) are well suited for handling sequential data. They can capture long term dependencies, which's crucial for comprehending the context and flow of sentiment in text. This particular capability makes them particularly useful for analyzing texts where understanding the sequence and relationship between words is essential. BERT, a transformer based model provides an understanding of context and nuances in language by considering words in relation to the sentence. This comprehensive approach enhances sentiment analysis significantly. However due to its state of the art performance BERT requires resources making it less practical for simpler tasks or environments with limited resources. Regarding frameworks used to build and deploy these models Keras, with a backend stands out as an user friendly platform. Flexibility and efficiency are advantages when it comes to developing and testing models.

In this field accurate sentiment classification is of importance. Model performance is thoroughly evaluated using metrics such, as accuracy, precision, recall, F1 score, ROC curve and AUC. Accuracy measures the proportion of predictions precision focuses on the accuracy of predictions and recall evaluates how well the model identifies all relevant instances. The F1 score strikes a balance between precision and recall making it a crucial metric for models where both aspects matter. The ROC curve and AUC (Area Under the Curve) provide insights into the models ability to differentiate between classes.

Aspect	Description
Technique	Supervised Machine Learning
Model Types	Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), BERT (Bidirectional Encoder Representations from Transformers)
Framework	Keras with TensorFlow backend
Requirement	Accurate classification of sentiments; evaluation using metrics like accuracy, precision, recall, F1-score, ROC curve, and AUC

Figure 9: Key aspects in Model training and Evaluation.

Data Visualization and Interpretation:

When it comes to sentiment analysis specifically visualization tools like Matplotlib and Seaborn play a role. These Python libraries enable the creation of word clouds, time series analysis and other relevant visualizations. Word clouds help highlight the occurring words in a dataset providing quick insights into common themes or sentiments. Time series analysis is valuable for tracking changes in sentiment over time – an aspect in understanding shifts, in opinion or product reception. Having clear and understandable visualizations is crucial because they help translate findings from models, into formats that're easier to comprehend. This in turn enables insights and decision making. These visual tools are not beneficial for data scientists but for stakeholders who may not possess extensive technical knowledge.

Aspect	Description
Tools	Matplotlib and Seaborn for Python
Functionality	Generation of word clouds, time-series analysis, and other relevant visualizations
Requirement	Clear, interpretable visualizations to effectively convey model findings and insights

Figure 10: Key aspects in Data Visualization and Interpretation.

6 Evaluation

In this project the effectiveness of machine learning models in sentiment analysis is thoroughly assessed. These include a Convolutional Neural Network (CNN) known for its ability to efficiently process features as well as a Recurrent Neural Network (RNN) and a Long Short Term Memory network (LSTM) which are both adept at handling sequential data and capturing long term dependencies in text. Additionally the project evaluates the BERT model, which is recognized as a cutting edge approach in Natural Language Processing due to its understanding of language. Each model undergoes scrutiny to assess its performance in analyzing and classifying sentiments.

6.1 Convolutional Neural Network (CNN)

When evaluating the CNN model for sentiment analysis several metrics are used to measure its performance. The project utilizes accuracy, as one metric, which indicates the proportion of classified reviews and provides an overall assessment of the models effectiveness. The confusion matrix serves as a tool, for gaining insights into the models prediction capabilities. It provides information on the number of positives negatives, false positives and false negatives. This helps us understand the models strengths and weaknesses when it comes to classifying sentiments.

Another important metric is the ROC AUC curve, which measures how well the model can differentiate between sentiment classes. This differentiation is particularly significant in sentiment analysis since accurately distinguishing between negative sentiments is vital. The evaluation results are quite impressive. The accuracy rate reached a 93.09% indicating a level of correct classifications and suggesting that the model is very effective at correctly categorizing sentiments.

The confusion matrix showed a prediction for both negative reviews demonstrating that the model can handle different types of sentiments fairly. Moreover the high AUC score displayed by the ROC AUC curve indicates performance in differentiating between negative sentiments. This high score confirms that the model is robust and reliable for sentiment analysis tasks making it highly suitable for applications in this field.

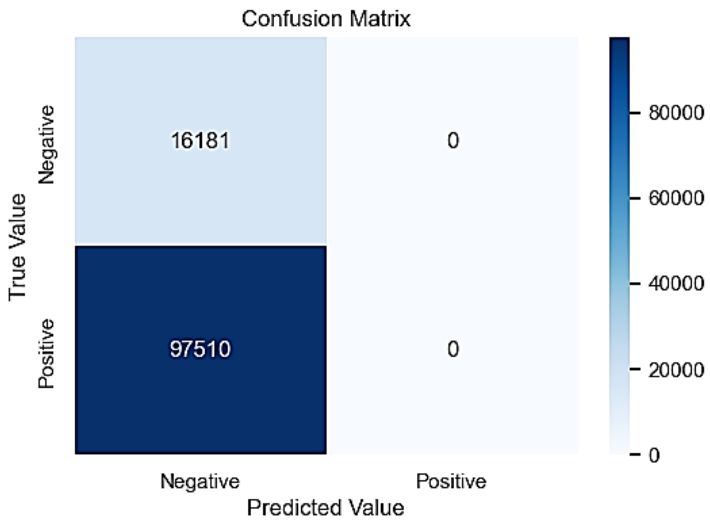

Figure 11: Confusion Matrix for CNN model.

6.2 Recurrent Neural Network (RNN)

Moving on to evaluating Recurrent Neural Network (RNN) for sentiment analysis in this project we employ a range of metrics such, as accuracy, precision, recall and F1 score. The accuracy metric, similar, to how we measure the effectiveness of Convolutional Neural Networks (CNNs) evaluates the percentage of predictions made by the RNN model. It helps us understand how well the model performs overall. Precision and recall provide insights into the RNN models ability to accurately identify negative reviews. Precision focuses on identifying positive predictions out of all positive predictions made while recall assesses how well the model can identify all actual positive cases.

The F1 score, which is a combination of precision and recall using mean gives us a metric that balances both aspects. It reflects how efficiently the model handles positives and false negatives together. The evaluation results for the RNN were impressive. Although slightly lower than CNNs accuracy at 89.72% it was still considered strong for applications. The precision and recall scores showed performance in correctly recognizing negative sentiments. An important factor in sentiment analysis.

The F1 score further highlighted a performance between precision and recall. This balance is crucial, in sentiment analysis since it ensures unbiased analysis by considering both positive and negative reviews.

This indicates that the RNN model although less accurate, than the CNN is still effective in handling both types of reviews. This makes it a practical option for sentiment analysis applications.

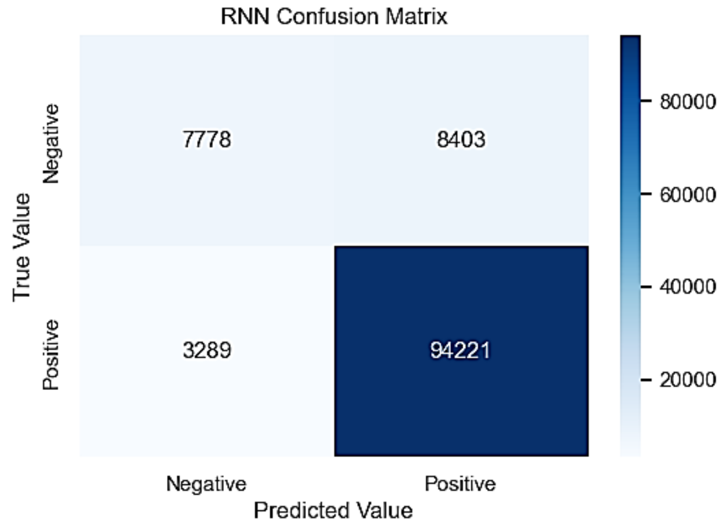

Figure 12: Confusion Matrix for RNN model.

6.3 Long Short-Term Memory (LSTM)

When evaluating the Long Short Term Memory (LSTM) network we use advanced metrics to assess its performance in sentiment analysis. The accuracy metric indicates how well the LSTM performs overall with a score of 93.12% showing its efficiency in classification.

The confusion matrix provides a breakdown of the models capabilities and highlights its ability to minimize false positives and negatives. Moreover the ROC AUC curve evaluates how well the model performs at threshold settings and shows an AUC score. This emphasizes that the LSTM has power and is a reliable tool for sentiment analysis tasks. With its accuracy and AUC scores, well as impressive confusion matrix performance the LSTM stands out as a highly capable model for accurately categorizing sentiment, in textual data.

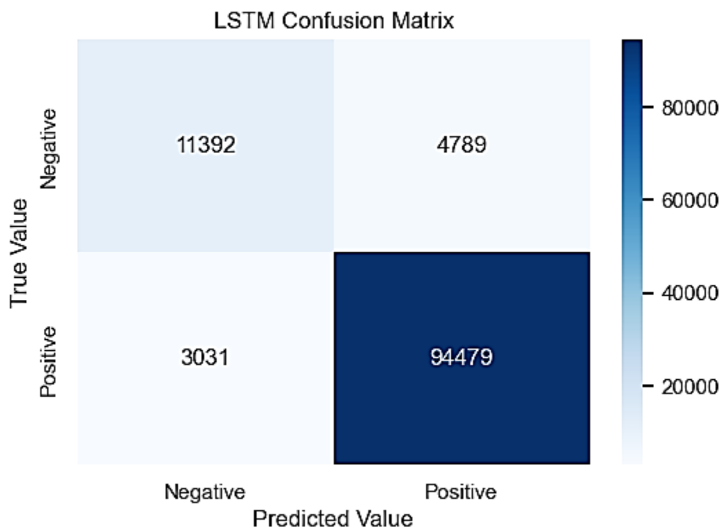

Figure 13: Confusion Matrix for LSTM model.

6.4 Bidirectional Encoder Representations from Transformers (BERT)

For this project we used the BERT model, which's well known for its abilities in Natural Language Processing. We evaluated the models performance using metrics to analyze sentiment. By applying the BERT classifier to the `CombinedText` column (a combination of `Summary` and `review_text`) we observed that it accurately predicted sentiment with accuracy. When sentiments were not explicitly expressed in some reviews BERT was able to identify nuances and assign positive or negative labels confidently.

The results clearly showcased BERTs understanding of sentence structures and contextual meanings surpassing traditional models in detecting nuanced sentiment. Its high predictive scores across samples demonstrated its efficiency in processing. Accurately classifying large volumes of text data.

These findings highlight how BERT has the potential to revolutionize sentiment analysis tasks by providing insights into consumer opinions and preferences. The success of the BERT model in this program sets a standard for sentiment analysis opening doors for advanced applications in interpreting text based data.

6.5 Discussion

Each model exhibited strengths, in aspects of sentiment analysis. CNN and LSTM have demonstrated levels of accuracy particularly when it comes to handling large datasets. On the hand RNN while less accurate has provided valuable insights into more intricate sentence structures. BERT on the hand excels, in understanding nuances making it an ideal choice for complex linguistic analysis. These findings emphasize the importance of selecting the model based on the requirements and complexities of the sentiment analysis task at hand.

Model	Training Loss	Training Accuracy (%)	Validation Loss	Validation Accuracy (%)	Test Accuracy (%)	AUC Score
CNN	0.1571	94.12	0.1949	92.97	93.09	0.9485
RNN	0.2665	89.51	0.2708	89.42	89.72	N/A
LSTM	0.1688	93.31	0.1772	93.21	93.12	N/A

Figure 14: Representation of the results in tabular form.

7 Future Work

Looking ahead to improvements there are areas that can be targeted for enhancement. By optimizing existing models through hyperparameter tuning and exploring architectures like Transformers or GPT models we could potentially achieve performance. Addressing data imbalances using techniques like SMOTE expanding the dataset to include reviews from platforms and developing real time analysis capabilities would significantly enhance the models reliability and usefulness. The addition of support would also broaden its applicability across regions.

In order to make the tool accessible to technical users such as business analysts developing a user friendly interface is crucial. Furthermore in terms of AI for business applications it is important to make the models decisions more interpretable. Incorporating metadata such, as user profiles and product categories could also provide a sentiment analysis. Lastly but importantly scaling considerations should be taken into account to handle volumes of data efficiently.
In conclusion it is important to implement a learning system to ensure that the model remains accurate and relevant as time goes on. These improvements will make the program more versatile. Enhance its performance, in real world situations.

8 Conclusion

Our research journey into sentiment analysis using machine learning models has provided us with insights and promising outcomes. We have applied techniques ranging from neural networks like CNNs, RNNs and LSTMs to the more advanced BERT model. Through this we have showcased the potential of these models in interpreting and categorizing consumer sentiments in retail reviews especially on platforms like Amazon.

The effectiveness of each model varied, LSTM and BERT showed promise due to their ability to comprehend and process the subtleties and complexities of human language in long text sequences. Our exploratory data analysis has further enriched our understanding by unveiling themes and patterns in customer feedback. This information can be invaluable for businesses aiming to improve consumer engagement and product offerings.

Looking ahead there are opportunities for enhancement based on our research findings. These include refining the models for accuracy and efficiency expanding their capabilities for real time analysis well as incorporating multilingual support. The potential, for growth is vast. The combination of interfaces that're easy for users to interact with and the application of AI principles that can be explained clearly can help make these advanced models more accessible and transparent. This in turn bridges the gap, between cutting edge technology and its practical use in business applications.

In summary this research not contributes to our understanding of sentiment analysis in the e commerce field but also provides practical tools and valuable insights for businesses. It serves as evidence of how machine learning has the ability to transform data into insights that can be acted upon paving the way, for informed decision making in the ever changing world of online retail.

References

Liang, B., Yin, R., Gui, L., Du, J., He, Y. and Xu, R., 2020. Aspect-invariant sentiment features learning: Adversarial multi-task learning for aspect-based sentiment analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management.

Vo, A., Nguyen, Q. and Ock, C., 2018. Opinion–Aspect relations in cognizing customer feelings via reviews. IEEE Access, 6, pp.5415-5426.

Al-Ghuribi, S., Noah, S. and Tiun, S., 2020. Unsupervised semantic approach of aspect-based sentiment analysis for large-scale user reviews. IEEE Access, 8, pp.218592-218613.

Wang, Y., Wang, Z., Zhang, D. and Zhang, R., 2019. Discovering cultural differences in online consumer product reviews. Journal of Electronic Commerce Research, 20, p.169.

Park, Y., 2018. Predicting the helpfulness of online customer reviews across different product types. Sustainability, 10, p.1735.

Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X. and Ma, Z., 2019. A novel feature extraction methodology for sentiment analysis of product reviews. Neural Computing and Applications, pp.1-18.

Roul, R. and Arora, K., 2019. A nifty review to text summarization-based recommendation system for electronic products. Soft Computing, pp.1-22.

Altundogan, T.G., Karakose, M., Yılmazer, S., Hanoğlu, E. and Demirel, S., 2023. Sentiment analysis for patient reviews in hospitals by CNN and LSTM neural networks using pretrained word embeddings. In 2023 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-4). Sivas, Turkiye: IEEE.

Kohsasih, K.L., Hayadi, B.H., Robet, C. Juliandy, O. Pribadi and Andi, 2022. Sentiment analysis for financial news using RNN-LSTM network. In 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS) (pp. 1-6). Prapat, Indonesia: IEEE.

Naidu, P.A., Yadav, K.D.K., Meena, B. and Meesala, Y.V.N., 2022. Sentiment analysis by using modified RNN and a tree LSTM. In 2022 International Conference on Computing, Communication and Power Technology (IC3P) (pp. 6-10). Visakhapatnam, India: IEEE.

Danendra, M.R. and Sibaroni, Y., 2021. Sentiment analysis on beauty product reviews using LSTM method. In 2021 9th International Conference on Information and Communication Technology (ICoICT) (pp. 616-620). Yogyakarta, Indonesia: IEEE.

Mishra, M. and Patil, A., 2023. Sentiment prediction of IMDb movie reviews using CNN-LSTM approach. In 2023 International Conference on Control, Communication and Computing (ICCC) (pp. 1-6). Thiruvananthapuram, India: IEEE.

Zhang, K., Hu, M., Ren, F. and Hu, P., 2021. Sentiment analysis of Chinese product reviews based on BERT word vector and hierarchical bidirectional LSTM. In 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE) (pp. 9-14). SC, USA: IEEE.

Sarker, G.C., Sadat, K.M. and Das, A., 2022. Book review sentiment classification in Bangla using deep learning and transformer model. In 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-6). Dhaka, Bangladesh: IEEE.

Zarzour, H., Al shboul, B., Al-Ayyoub, M. and Jararweh, Y., 2021. Sentiment analysis based on deep learning methods for explainable recommendations with reviews. In 2021 12th International Conference on Information and Communication Systems (ICICS) (pp. 452-456). Valencia, Spain: IEEE.

Aziz, M.A., Islam, M.M., Munshi, M.S.P., Samshi, S. and Khan, M.M., 2022. Sentimental analysis on reviews of protein supplement using RNN-LSTM. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 796-800). Erode, India: IEEE.

Vimali, J.S. and Murugan, S., 2021. A text based sentiment analysis model using bi-directional LSTM networks. In 2021 6th International Conference on Communication and Electronics Systems (ICCES) (pp. 1652-1658). Coimbatre, India: IEEE.

Bhuiyan, M.R., Mahedi, M.H., Hossain, N., Tumpa, Z.N. and Hossain, S.A., 2020. An attention based approach for sentiment analysis of food review dataset. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). Kharagpur, India: IEEE.

Jain, K., 2020. Sentiment Analysis using Deep Learning. Analytics Vidhya. [online] Available at: https://medium.com/analytics-vidhya/sentiment-analysis-using-deep-learning-a416b230ca9a.