Sentiment Analysis of MacBook Amazon Reviews Using NLP Techniques

Data 690: Introduction to NLP
Professor: Darin Johnson

AMRITA CHANDRASEKAR(NH53017)

OUTLINE

- Introduction
- Data Preprocessing
- Exploratory Data Analysis
- Models used for Sentiment Analysis
- Model Evaluation
- Topic Modeling
- Aspect Based Frequency
- Aspect Based Sentiment Analysis
- Conclusion
- Future Scope
- References

INTRODUCTION

Objectives

- To deploy sentiment analysis of reviews to extract detailed sentiments regarding the Macbook
- Compared various classification models and techniques including Logistic Regression, Naïve Bayes, SVM, BERT in analyzing text data to determine the most accurate methods for sentiment classification.
- To implement Topic Modelling to determine the most common words in reviews
- To implement Aspect based analysis by extracting aspects from review and then perform aspect frequency analysis and aspect-based sentiment analysis

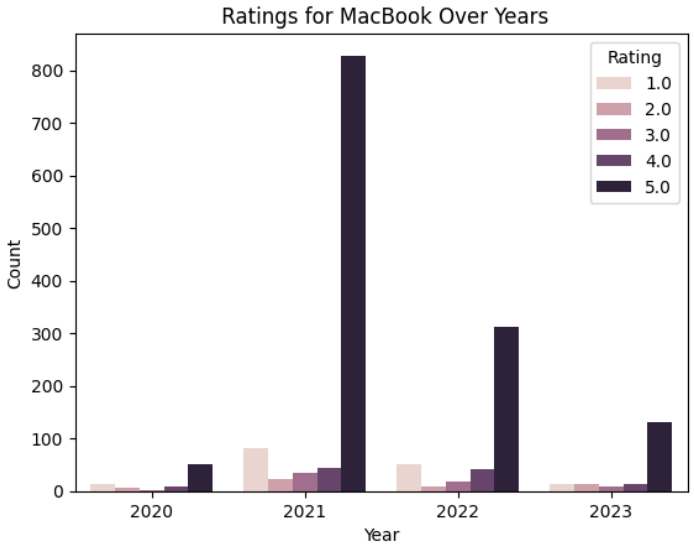
Data Collection

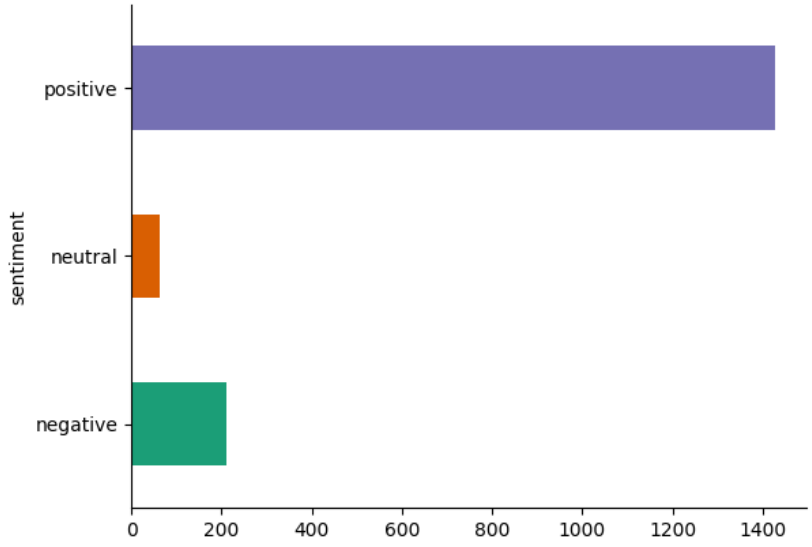
- Dataset is collected from Kaggle https://www.kaggle.com/datasets/poojaparab/amazon-product-review-dataset-apple-mac-air
- The dataset has the information about the product name(product), review title(title), date of the review(date), user rating out of 5(rating), and the actual review(body) of the customers.

DATA PREPROCESSING

- Converting body column to string type
- Removing punctuation and stop words
- Lowercasing
- Converting non string object to string in body column
- Tokenization
- Lemmatization

EXPLORATORY DATA ANALYSIS





EXPLORATORY DATA ANALYSIS

MODELS USED FOR SENTIMENT ANALYSIS

Logistic Regression

Naive Bayes

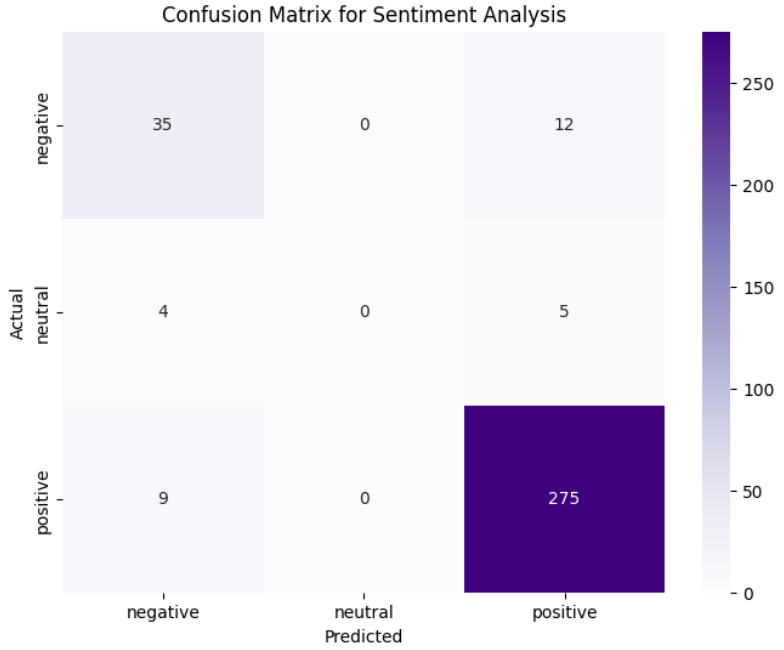
Support Vector Machine (SVM)

BERT (Bidirectional Encoder Representations from Transformers)

MODEL EVALUATION

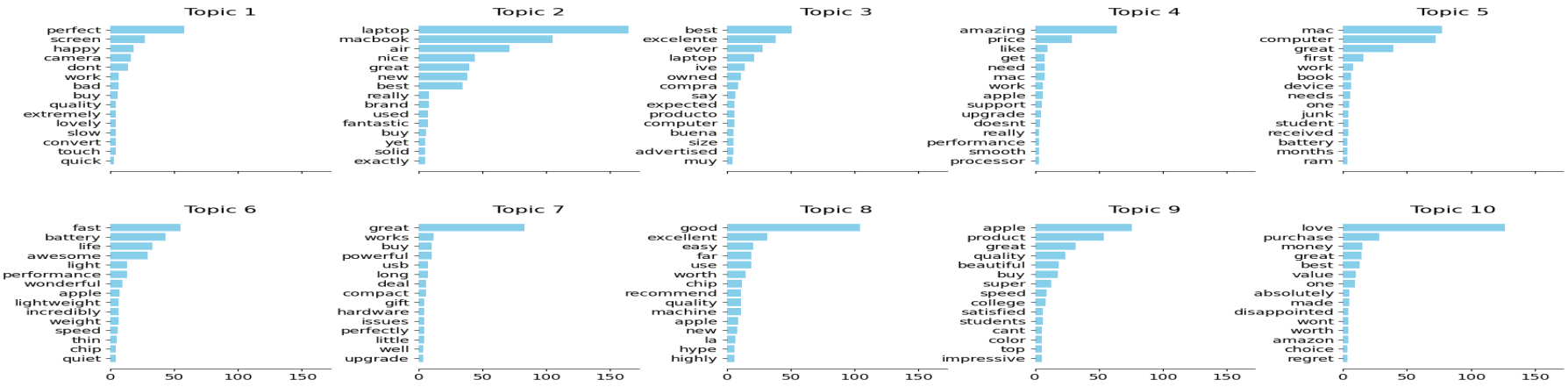
- **Logistic Regression**: Accuracy 86%
- **Naive Bayes**: Accuracy 86%
- **SVM**: Accuracy 89%
- **BERT**: Accuracy 91%

SENTIMENT ANALYSIS COFUSION MATRIX



TOPIC MODELLING

LDA (Latent Dirichlet Allocation) is used for topic modeling in order to automatically group related words into topics.

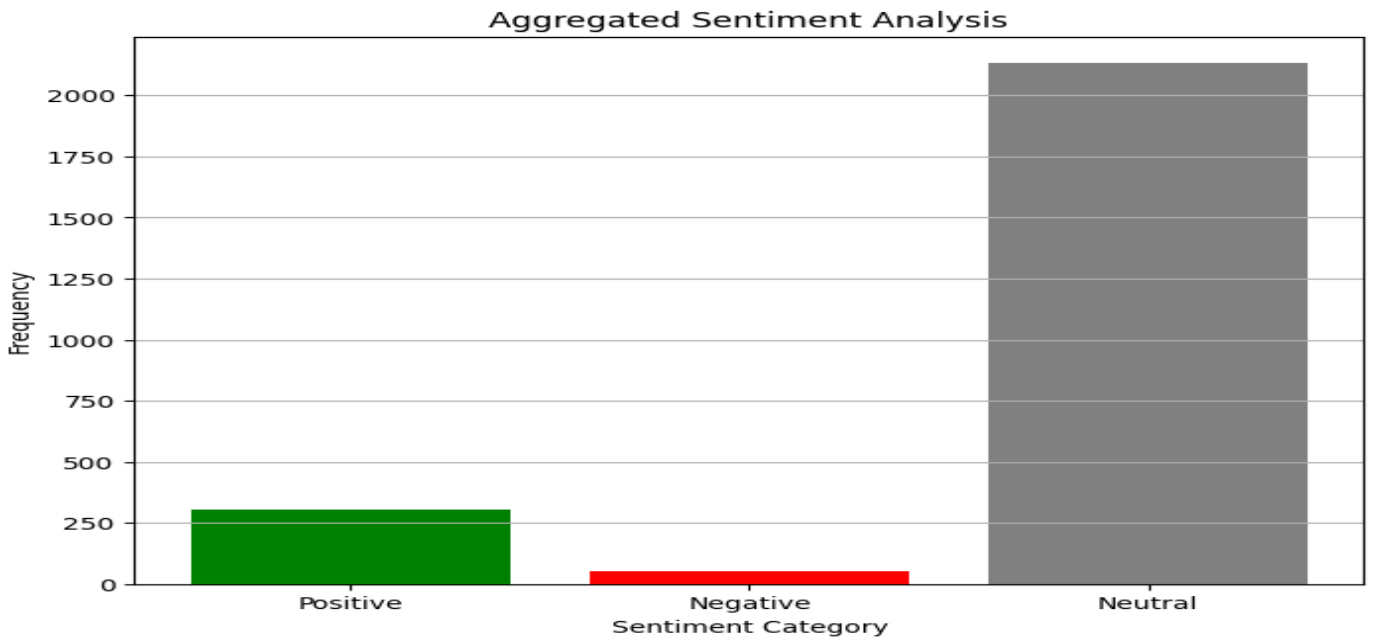


ASPECT FREQUENCY ANALYSIS

Top Aspect Categories by Frequency

- Macbook: 597
- Love: 183
- Battery: 84
- Purchase: 81
- Performance: 54
- Quality: 39
- Screen: 21
- Chip: 18
- Value: 16
- Camera: 16

ASPECT BASED SENTIMENT ANALYSIS



CONCLUSION

Successfully analyzed consumer reviews to extract detailed sentiments using advanced NLP techniques.

Various NLP models, including Logistic Regression, Naive Bayes, SVM, and BERT, were compared. BERT demonstrated superior performance in sentiment classification, highlighting the effectiveness of transformer-based models for this task.

Successfully extracted most commonly used words in reviews

Successfully implement aspect-based sentiment analysis

FUTURE SCOPE

- Expand Dataset for Greater Accuracy:
  - Increase the dataset size and diversity by collecting more reviews from various sources.
  - Improve model generalization and accuracy across different consumer opinions.
- Explore Advanced Deep Learning Models:
  - Investigate the use of advanced models like RoBERTa and GPT-3.
  - Explore ensemble methods for better performance in sentiment analysis and aspect extraction.

REFERENCES

- Diekson, Z. A., Prakoso, M. R. B., Putra, M. S. Q., Syaputra, M. S. A. F., Achmad, S., & Sutoyo, R. (2023). Sentiment analysis for customer review: Case study of Traveloka. Procedia Computer Science, 216, 682–690. https://doi.org/10.1016/j.procs.2022.12.184
- Selvakumar, B., & Lakshmanan, B. (2022). Sentimental analysis on user's reviews using BERT. Materials Today: Proceedings, 62, 4931–4935. https://doi.org/10.1016/j.matpr.2022.03.678
- Singh, A., & Kumar, P. (2022). Aspect based sentiment analysis for customer reviews. IEEE Conference Publication. https://ieeexplore.ieee.org/document/9761302

THANK YOU