

**MSc Project**

***7 COSC012W.3***

***Project Proposal***

**Comparing Sentiment Analysis on Social Media and E-Commerce Platform:**

**Apple Products**

**Hasan Enes Guray**

**Student ID: 19489124**

**MSc Data Science and Analytics student**

**University of Westminster**

**Supervisor: Amit Banik**

ABSTRACT

This MSc project will compare sentiment analysis of Apple products on social media and e-commerce platforms. Using advanced data science techniques, the project will extract insights from unstructured data. Data from YouTube, Reddit, Amazon, and eBay will be collected and analysed, revealing trends and patterns for strategic insights. The research will contribute to sentiment analysis and offer practical implications for the evolving tech industry. The project aims to fill a critical gap in the literature, benefiting academia and industry practices.

**Keywords:** Sentiment analysis, Apple products, Social media, E-commerce websites.

# Background & the problem domain:

Social media and e-commerce have revolutionized business-customer interaction, altering the way sentiments are harnessed for strategic insights. The evaluation of user sentiment from these platforms is essential, enabling comparisons between buyer impressions and wider audiences. Apple Inc., engages extensively with its online customer base, making understanding customer sentiment imperative for its competitive edge.

This endeavor holds academic significance by employing advanced data science on vast unstructured text data. Sentiment analysis, within natural language processing (NLP), bridges qualitative opinions with quantitative insights. This not only advances NLP but also unveils real-world interpretations of intricate customer sentiments.

With direct implications for industrial strategies, businesses recognize the value of customer sentiment in driving innovation and satisfaction. This initiative addresses the intricate task of deciphering sentiments across diverse platforms—spanning social media to e-commerce reviews—effectively addressing a genuine business challenge.

The project's core challenge lies in comprehending customer sentiments and contrasting sentiments across platforms for nuanced comprehension. Traditional sentiment analysis struggles with diverse sources like social media and e-commerce reviews, potentially obscuring the true sentiment landscape. This project employs advanced sentiment analysis to aggregate data from platforms like YouTube, Reddit, Amazon, and eBay, aiming to unveil subtle customer sentiment disparities and patterns that inform Apple's strategic decisions.

Sentiment analysis holds universal importance, prompting various projects. Dupinder Kaur's tweet sentiment study noted Apple's positivity dominance via Naive Bayes (<https://www.researchgate.net/publication/353072253_Sentimental_Analysis_on_Apple_Tweets_with_Machine_Learning_Technique> accessed on 10 August 2023). Jasmina Smailović forecast Apple stock movements through Twitter (<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ba53a72840a5e9dd5787235007a873984d3a4f3d> accessed on 10 August 2023). Hassan Saif's research improved sentiment prediction accuracy using semantic concepts (<https://link.springer.com/chapter/10.1007/978-3-642-35176-1_32> accessed on 10 August 2023). Rubi Gupta enhanced stock price prediction with microblogging sentiments (<https://ieeexplore.ieee.org/abstract/document/9175549> accessed on 10 August 2023). Elly Indrayuni optimized SVM for Apple product review sentiment classification (<https://www.researchgate.net/publication/340674251_Optimizing_Genetic_Algorithms_for_Sentiment_Analysis_of_Apple_Product_Reviews_Using_SVM> accessed on 10 August 2023). Though extensive research covers Apple sentiment, a direct e-commerce and social media comparison is missing. This project aims to bridge this gap, pioneering new territory. Thorough data collection, preprocessing, and analysis enrich academia and offer strategic guidance for businesses in digital environments.

# objectıves

* Perform an in-depth comparative sentiment analysis on discussions surrounding Apple products, encompassing both social media and e-commerce platforms, utilizing advanced data science techniques.
* Develop customized sentiment analysis methods specifically designed for the distinct characteristics of social media and e-commerce text, ensuring accurate sentiment interpretation considering platform dynamics, demographics, and product attributes.
* Analyse sentiment trends across both platforms, accounting for evolving dynamics, demographic shifts, and changes in product features, to discern and interpret sentiment patterns over time.
* Uncover recurring sentiment patterns, prominent themes, impactful keywords, and significant topics to identify the underlying factors driving Apple product sentiment on social media and e-commerce platforms.
* Evaluate the strategic implications of sentiment analysis for business strategies, encompassing innovation, marketing decisions, and overall customer satisfaction, with a focus on providing actionable insights for companies like Apple.
* Contribute to the advancement of sentiment analysis methodologies within the context of dynamic digital ecosystems, enriching academic discussions on natural language processing and sentiment analysis techniques.
* Address data quality challenges using robust preprocessing techniques, and manage bias through diverse training data and debiasing methods.

# Project

## Data Collection and Preprocessing

Data collection will start from platforms like YouTube, Reddit, Amazon, and eBay, employing web scraping to extract relevant user content like comments and reviews about Apple products. Metadata including timestamps, user profiles, and platform-specific features will also be collected to contextualize sentiment analysis.

## Sentiment Analysis Techniques

To capture sentiment complexity across platforms, a blend of adapted methods will be used. This includes lexicon-based approaches assigning sentiment scores to words and phrases, as well as ML models like Naive Bayes, Support Vector Machines (SVM), Recurrent Neural Networks (RNN), Transformer-based models like BERT, and LexMo for emotion analysis.

## Comparative Analysis and Insights

The project's essence is a thorough sentiment comparison between social media and e-commerce platforms. It involves tracking sentiment trends over time, exploring variations, and aspect-based comparisons for nuanced understanding.

## Output and Project Management

Project outcomes comprise sentiment analysis models for precise Apple-related sentiment evaluation across platforms. Visualizations will illustrate sentiment trends, aiding interpretation. Findings and methods will be shared through a research paper and presentations, fostering collaboration with the University of Westminster and Apple.

The project's management will be organized through a series of milestones. The initial phase involves data collection and preprocessing, spanning Weeks 1 to 4. This phase aims to collect and clean data from different platforms, ensuring data compatibility and quality.

The subsequent phase, spanning Weeks 5 to 6, focuses on sentiment analysis model development. During this period, the project will experiment with various sentiment analysis techniques, refine the models, and ensure their robust performance on both social media and e-commerce data.

The third phase, Weeks 7 to 9, involves comparative analysis and pattern recognition. This stage is dedicated to conducting a detailed comparative analysis, identifying sentiment patterns, and extracting valuable insights from the sentiment data.

The final phase, Weeks 10 to 12, centers on documentation and dissemination. During this period, the project will compile research findings, create visualizations, draft the research paper, and prepare presentations to share outcomes with relevant stakeholders.

## Alternative Approaches and Likelihood of Success

### Data Quality and Consistency:

**Challenge:** Gathering data from diverse platforms can introduce format variations, noise, and inconsistencies. Various structures exist for comments across platforms.

**Solution:** Employ robust data preprocessing to standardize data. Design scripts for format handling, and cleansing text by removing noise, irrelevant info, and special characters. Leverage regular expressions and NLP libraries for uniformity.

### Bias and Generalization:

**Challenge:** Sentiment models can replicate biases from training data, causing inaccurate or unfair predictions. Limited generalization across platforms or demographics may occur.

**Solution:** Utilize diverse, representative training data to counter bias. Assess model performance with a distinct validation set. Apply transfer learning for platform adaptation. Use debiasing techniques to minimize unwanted prediction bias.

### Model Complexity and Performance:

**Challenge:** Complex deep-learning sentiment models can be hard to interpret and computationally demanding. Generalization with limited data could suffer.

**Solution:** Begin with simpler models, and escalate complexity if required. Feature importance analysis highlights influential words/features. Monitor performance, and fine-tune hyperparameters for optimal results.

# resources

## Hardware Resources:

A robust personal computer will support diverse tasks, including data preprocessing, model development, and analysis. Sufficient storage is vital for extensive datasets, files, trained models, and intermediate results.

## Software Resources:

The project relies on software tools and libraries for efficient development and analysis. Python is key for data processing, sentiment analysis, and model creation. Web scraping tools like Praw, Google API Client, BeautifulSoup, and Selenium gather data from platforms. Machine learning libraries (scikit-learn, TensorFlow, PyTorch) aid model creation. NLP libraries (NLTK, spaCy) assist text preprocessing and linguistic analysis. VADER and BERT enable sentiment analysis. Data visualization tools (Matplotlib, Seaborn) show sentiment trends. Version control (GitHub) supports collaboration. IDE (Jupyter Notebook) facilitates code development.

## Library and Research Resources:

Regular access to academic journals, papers, and online resources is vital for staying current with sentiment analysis methods. Platforms like arXiv, IEEE Xplore, and ResearchGate offer research materials. Medium, Kaggle, and Towards Data Science will be utilized for comprehensive resource exploration and method analysis, guiding project decisions in line with best practices and field innovation.

## Data Resources:

Data collection is pivotal for project success. Access to relevant APIs is essential to gather data from social media (e.g., YouTube, Reddit) and e-commerce (Amazon, eBay) platforms. Custom web scraping scripts will extract text from platforms lacking APIs, ensuring comprehensive data acquisition.

## Stakeholder Collaboration:

Partnering with stakeholders is essential to refine project outcomes. Collaboration with Apple Inc. could provide valuable insights, enhancing research implications. University of Westminster resources, including facilities and expertise, enrich project quality.

## Project Timeline:

A well-structured timeline is vital for steady progress. Allotting time for data collection, preprocessing, model development, analysis, and documentation ensures thorough execution. Regular milestone checks ensure timely completion and flexibility for adjustments (Appendix A).

# references

Gupta, R. and Chen, M. (2020) ‘Sentiment analysis for stock price prediction’, *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* [Preprint]. doi:10.1109/mipr49039.2020.00051.

Indrayuni, E. and Nurhadi, A. (2020) ‘Optimizing genetic algorithms for sentiment analysis of Apple product reviews using SVM’, *SinkrOn*, 4(2), p. 172. doi:10.33395/sinkron.v4i2.10549.

Kaur, D. (2017) ‘Sentimental Analysis on Apple Tweets with Machine Learning Technique’, *JCSET(www.ijcset.net)*, 7(9), pp. 76–78.

Saif, H., He, Y. and Alani, H. (2012) ‘Semantic sentiment analysis of Twitter’, *The Semantic Web – ISWC 2012*, pp. 508–524. doi:10.1007/978-3-642-35176-1\_32.

Smailović, J., Grčar, M. and Žnidaršič, M. (no date) *Sentiment analysis on tweets in a financial domain* [Preprint].

# Appendices

## Appendix A – Project Schedule

A screenshot of a spreadsheet

Description automatically generated