

**MSc Project**

***7 COSC012W.3***

***Project Proposal***

**Comparing Sentiment Analysis of Apple Products on Social Media and E-Commerce Platforms**

**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 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 transformed business-customer interaction. Direct engagement and sentiment analysis from these platforms offer insights for strategic decisions, reflecting a significant shift in understanding and utilizing customer sentiments.

In this context, evaluating sentiments from user comments on social media and e-commerce is vital. It allows comparing impressions of both buyers and the broader audience. Apple Inc., known for innovation, interacts with a wide online customer base. Grasping customer sentiments on these platforms is vital for Apple's success in a competitive market.

This project's academic value stems from employing advanced data science on extensive unstructured text data. Sentiment analysis, a natural language processing subset, bridges qualitative customer opinions and quantitative insights. By utilizing sentiment analysis, the project advances natural language processing (NLP) and showcases real-world decoding of intricate customer sentiments.

This project directly impacts industrial strategies and decisions. Businesses acknowledge the importance of using customer sentiments to innovate, market, and satisfy. It addresses deciphering complex customer sentiments across platforms, spanning social media to e-commerce reviews, a genuine business challenge.

This project tackles a dual challenge: understanding customer sentiments amidst data overflow and comparing sentiments across platforms for nuanced insights. Common sentiment analysis methods struggle with diverse sources like social media and e-commerce reviews. The discrepancy between platforms can obscure the true sentiment landscape, highlighting the project's pivotal problem domain.

This project gathers and processes data from social media (YouTube, Reddit) and e-commerce (Amazon, eBay) using advanced sentiment analysis. It seeks to reveal nuanced customer sentiment differences across platforms. Aggregating sentiment data and recognizing recurring themes and products aims to inform Apple's strategic decisions with valuable patterns.

This project enriches sentiment analysis and NLP academically while furnishing actionable insights for firms like Apple. By comprehensively assessing sentiments across social media and e-commerce, it empowers businesses to transform customer opinions into informed strategies, enhancing satisfaction and fostering innovation in a competitive environment.

As the field of sentiment analysis garners universal importance and continued development, numerous studies and projects have materialized. While Dupinder Kaur was trying to explain the sentiment analysis of tweets about Apple with machine learning via the Naive Bayes classifier, he found that positive sentiment predominates among tweets (<https://www.researchgate.net/publication/353072253_Sentimental_Analysis_on_Apple_Tweets_with_Machine_Learning_Technique> accessed on 10 August 2023). Jasmina Smailović et al. explored the use of sentiment analysis on Twitter data to determine significant events and forecast stock price changes, demonstrating success in predicting Apple stock movements based on tweet volume and sentiment polarity (<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ba53a72840a5e9dd5787235007a873984d3a4f3d> accessed on 10 August 2023). Hassan Saif et al. paper discussed using sentiment analysis on Twitter to monitor public sentiments towards organizations including Apple, introducing a novel approach of incorporating semantic concepts as additional features for sentiment prediction, leading to improved accuracy in identifying both negative and positive sentiments compared to traditional methods (<https://link.springer.com/chapter/10.1007/978-3-642-35176-1_32> accessed on 10 August 2023). Rubi Gupta et al. explored using sentiment expressed on StockTwits, a microblogging platform, to enhance stock price prediction by analyzing tweet contents and studying the correlation between aggregated daily sentiment and daily stock price movement for companies including Apple (<https://ieeexplore.ieee.org/abstract/document/9175549> accessed on 10 August 2023). Elly Indrayuni et al. examined Apple product reviews to classify sentiments using the Support Vector Machine algorithm and genetic algorithms for optimization, resulting in accuracy improvements from 70.00% to 85.76% and AUC enhancements from 0.924 to 0.945 compared to the SVM model without optimization. While existing research covers various aspects of sentiment analysis related to Apple, a direct comparison between e-commerce and social media remains absent. This project addresses this gap, aiming to pioneer in uncharted territory. By collecting, preprocessing, and analyzing data meticulously, it not only enriches academia but also guides strategic decisions for companies navigating dynamic digital landscapes.

# objectıves

The primary goal of this MSc project is to conduct a comprehensive comparative sentiment analysis of discussions related to Apple products on both social media and e-commerce platforms. The intention is to bridge the existing gap in knowledge by investigating the subtleties of sentiment expressions within these distinct digital landscapes. This research endeavours will employ advanced data science techniques to extract meaningful insights from unstructured textual data, making contributions to both the academic field of sentiment analysis and the practical world of business decision-making.

To begin, the project aims to assemble and curate a substantial dataset sourced from diverse platforms, encompassing social media outlets such as YouTube and Reddit, as well as e-commerce giants Amazon and eBay. This data will undergo meticulous preprocessing procedures to ensure its cleanliness and suitability for subsequent sentiment analysis.

The project will delve into the realm of sentiment analysis methodologies, investigating state-of-the-art techniques tailored to the specific characteristics of textual data extracted from social media and e-commerce platforms. The objective here is to adapt existing models or devise novel ones capable of accurately capturing the intricate nuances of sentiment expression.

An integral facet of the project is the in-depth comparative analysis of sentiment expressions between the distinct worlds of social media and e-commerce. The research will highlight notable similarities and discrepancies in sentiment trends, considering factors such as platform dynamics, user demographics, and the nature of the products under scrutiny.

Through the application of data aggregation techniques, the project aims to uncover recurrent sentiment patterns that emerge over time within these digital realms. By identifying prevalent themes, significant keywords, and recurring topics, the project intends to elucidate the drivers behind positive and negative sentiments for Apple products across both social media and e-commerce platforms.

The project's broader scope extends to the strategic implications these sentiment analysis outcomes hold for business decision-making. This involves interpreting the tangible impact on crucial aspects like product innovation, marketing strategies, and the enhancement of overall customer satisfaction. The project's aspiration is to offer actionable insights capable of steering companies like Apple toward refined strategies rooted in an astute understanding of customer sentiments.

While the primary focus remains on the practical applications of sentiment analysis, the project concurrently strives to enhance existing methodologies within the academic domain. By bolstering the discourse surrounding sentiment analysis and natural language processing, the project seeks to provide empirical evidence that illuminates sentiment variations across distinct digital platforms.

An essential component of the project is knowledge enrichment. The research endeavours to deepen understanding in the realm of advanced data science techniques and sentiment analysis, especially when applied in the real-world context of dynamic digital ecosystems. This aspect of the project is geared toward equipping the researcher with insights into the challenges and opportunities embedded in sentiment analysis across various online platforms.

Dissemination of the project's findings stands as another significant objective. The research aims to present its outcomes with Apple and the University of Westminster through academic papers, presentations, thus contributing to the broader research community's comprehension of challenges and solutions within the field of sentiment analysis.

The project underscores its practical applicability and relevance within the industry. By showcasing how sentiment analysis can be harnessed to refine business strategies, particularly within the technology sector epitomized by Apple, the research endeavors to highlight the capacity of sentiment insights to guide companies through the dynamic evolution of the digital landscape. The temporal framework of the project has been meticulously designed to ensure the successful attainment of these objectives, all within the constraints of the designated MSc project duration.

# Project

The proposed MSc project aims to conduct a comprehensive comparative sentiment analysis of discussions related to Apple products on both social media and e-commerce platforms. This section outlines the methods and techniques to be employed to achieve the project's objectives, while also discussing possible alternatives and outlining a manageable project plan.

## Data Collection and Preprocessing

The project will begin by collecting data from various platforms, including YouTube, Reddit, Amazon, and eBay. This will involve utilizing web scraping techniques to extract user-generated content such as comments, reviews, and posts relevant to Apple products. Additionally, metadata like timestamps, user profiles, and platform-specific features will be collected to provide context for the subsequent sentiment analysis.

## Sentiment Analysis Techniques

To address the complexity of sentiment expression within diverse platforms, a combination of existing sentiment analysis methods will be explored and adapted. These techniques may include lexicon-based approaches, where sentiment scores are assigned to individual words and phrases using sentiment dictionaries. Machine learning models, such as Naive Bayes, Support Vector Machines (SVM), or deep learning models like Recurrent Neural Networks (RNN) or Transformer-based models like BERT, will also be tried to develop. LexMo will be applied to conduct a respective emotion analysis.

## Comparative Analysis and Insights

The core of the project involves a detailed comparative analysis of sentiment expressions across social media and e-commerce platforms. This will be achieved by comparing sentiment trends over time, investigating potential variations, and performing an aspect-based comparison to understand differences in sentiment expression.

## Output and Project Management

The project's outcomes will encompass sentiment analysis models trained to accurately assess sentiments within Apple-related discussions on various platforms. Additionally, the project will produce insightful visualizations depicting sentiment trends, comparisons, and patterns to facilitate interpretation. The research findings and methodologies will be documented in a research paper and presentations for sharing with the University of Westminster and potential collaboration with 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

While the proposed approach focuses on adapting and combining existing sentiment analysis techniques, it's important to acknowledge potential alternatives and risks that may arise during the project's execution. These potential risks and alternative approaches are:

### Data Quality and Consistency:

**Risk:** Collecting data from various platforms may result in varying data formats, noise, and inconsistencies. Different platforms might have different structures for reviews, comments, and user-generated content.

**Solution:** Implement rigorous data preprocessing and cleaning techniques to standardize the collected data. Develop scripts that handle various data formats and clean the text data by removing irrelevant information, special characters, and noise. Use regular expressions and NLP libraries to ensure consistency.

### Bias and Generalization:

**Risk:** Sentiment analysis models can inherit biases present in the training data, leading to inaccurate or unfair predictions. The models might not generalize well to different platforms or user demographics.

**Solution:** Employ a diverse and representative training dataset to mitigate bias. Regularly evaluate the model's performance on a separate validation dataset. Fine-tune the models with techniques like transfer learning to adapt to different platforms. Implement techniques like debiasing to reduce unwanted bias in predictions.

### Model Complexity and Performance:

**Risk:** Implementing advanced sentiment analysis techniques like deep learning models may lead to complex models that are difficult to interpret. These models might require significant computational resources and may not generalize well with limited data.

**Solution:** Start with simpler sentiment analysis models and gradually increase complexity as needed. Use techniques like feature importance analysis to understand which words or features contribute most to the model's predictions. Regularly monitor model performance and adjust hyperparameters to achieve optimal results.

Despite these potential challenges, the project's likelihood of success is bolstered by the availability of relevant data sources and established sentiment analysis techniques. The comprehensive milestone-based project plan ensures steady progress and manageable execution within the defined 12 weeks, aligning with the scope of an MSc level project. Adapting to potential challenges through the outlined solutions and alternative approaches enhances the project's prospects for achieving its objectives and contributing to the fields of sentiment analysis and data science.

# resources

To ensure the successful execution of this project, a variety of resources will be required, each serving a specific purpose in achieving the project's objectives.

## Hardware Resources:

A high-performance personal computer will serve as the backbone for various project tasks, including data preprocessing, model development, and analysis. Additionally, ample storage space is essential to accommodate the extensive datasets, preprocessed files, trained models, and intermediate results that will be generated throughout the project.

## Software Resources:

The project heavily relies on software tools and libraries that enable efficient development and analysis processes. The Python programming language will be the foundation for various tasks such as data processing, sentiment analysis, and model implementation. Web scraping tools, such as Praw, Google API Client, BeautifulSoup and Selenium, will be essential for data collection from platforms like YouTube, Reddit, Amazon, and eBay. Machine learning libraries like scikit-learn, TensorFlow, and PyTorch will facilitate the creation and experimentation of sentiment analysis models. Natural Language Processing (NLP) libraries like NLTK and spaCy will support text preprocessing and linguistic analysis. To implement sentiment analysis, prebuilt tools like VADER for lexicon-based analysis and Hugging Face Transformers for advanced models like BERT will be utilized. Data visualization tools, including Matplotlib and Seaborn, will be employed to convey sentiment trends visually. Version control through Git and platforms like GitHub will aid in collaborative development. An Integrated Development Environment (IDE) like Jupyter Notebook or Visual Studio Code will be used for code development and experimentation.

## Library and Research Resources:

Access to academic journals, papers, and online resources is crucial to stay informed about the latest sentiment analysis techniques and methodologies. Platforms like arXiv, IEEE Xplore, and ResearchGate will provide valuable research materials. Moreover, platforms such as Medium, Kaggle, and Towards Data Science, which serve as hubs for knowledge dissemination within the realm of data science, will be harnessed for comprehensive resource exploration and meticulous methodology analysis. These resources will guide the project's decisions, ensuring alignment with best practices and innovation in the field.

## Data Resources:

Data collection is fundamental to this project's success. To gather data from social media and e-commerce platforms, access to relevant APIs is necessary. Social media APIs from platforms like YouTube and Reddit will be used for scraping data from user-generated content. Similarly, e-commerce APIs from platforms like Amazon and eBay will provide access to product reviews and descriptions. For platforms without APIs, custom web scraping scripts will be developed to extract the required textual data.

## Stakeholder Collaboration:

Collaboration with stakeholders is vital for gaining domain-specific insights and refining the project's outcomes. Potential collaboration with Apple Inc. could offer invaluable insights and enhance the practical implications of the research findings. Additionally, the University of Westminster's resources, including research facilities, libraries, and academic expertise, will provide guidance and feedback, enriching the project's overall quality.

## Project Timeline:

To ensure a well-paced progress, a carefully designed timeline is essential. Allocating sufficient time to each project phase, including data collection, preprocessing, model development, analysis, and documentation, will enable a thorough and successful execution. Regular monitoring against predefined milestones will facilitate timely completion and provide room for adjustments as needed.

# references

# Appendices

## Appendix A – Project Schedule

A screenshot of a spreadsheet

Description automatically generated