SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters in Data Science**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	CHAU AI VIN
Metric Number:	MCS241033
Student Email & Phone:	chauaivin@graduate.utm.my	018-7761223
Project Title:	Sentiment Analysis of Online Customer Reviews with Machine Learning Models (Logistic Regression and SVM) to Predict Product Success in E-Commerce Platforms
Supervisor 1:	
Supervisor 2 / Industry Advisor (if any):	

SECTION B: Project Proposal

Introduction:

Since the advent of e-commerce sites, the trend of consumer purchase decisions has changed (Ma et al., 2024). When purchasing a product previously, advertisements and recommendations from acquaintances were major sources of information. Now, with the development of the e-commerce industry, in the form of product reviews, the e-commerce websites also ask their consumers to provide feedback on the products they have bought. Online reviews, which have taken the place of traditional "word-of-mouth" with the phenomenal rise of electronic commerce, significantly influence customer purchasing behavior as well as product sales (Thakur, 2024). Sentiment analysis, especially through Natural Language Processing (NLP), helps companies understand the attitudes and views of customers based on these reviews. However, sentiment analysis alone as a predictor of the success of a product on an e-commerce site is not sufficient (Daza et al., 2024). By integrating sentiment analysis with conventional machine learning models such as Logistic Regression and Support Vector Machines (SVM), companies can better understand consumer behavior and make more informed product development and marketing strategy decisions (Haroon et al., 2024). This research seeks to investigate how to forecast the success of products on e-commerce websites by integrating sentiment analysis with these machine learning methods.

Problem Background:

With online shopping sites having increased at a rapid pace, customers upload their experience of a product online by posting reviews, and the view in reviews also plays a key role in affecting other potential buyers' purchasing decision (Thakur, 2024; Ma et al., 2024). Sentiment analysis (NLP) is now becoming a useful process to manage large quantities of review data, classifying reviews automatically into positive, negative or neutral. However, the existing sentiment analysis models, especially those that rely solely on NLP, are prone to fail to completely capture the subtlety and complexity of consumer sentiment, especially when handling large and diverse review data. For instance, sentiment analysis models typically struggle to achieve accurate and fine-grained capture of consumers' subtle variations in emotional expression from their reviews, especially in the presence of large volumes of diverse and diverse types of reviews and thus generate low model prediction accuracy. Apart from that, sentiment analysis usually focuses on examining the text itself, but not necessarily considering other features in the text, such as sentiment intensity, richness of vocabulary, and complexity of sentiment expression. All of these features are crucial to predicting the market performance of a product (Singh et al., 2022; Shad et al., 2024). In order to better forecast the market performance of products, it needs to include conventional machine learning algorithms such as Logistic Regression and SVM to enhance the predictive power of sentiment analysis models (Haroon et al., 2024).

Problem Statement:

According to Daza et al. (2024), although sentiment analysis is widely used in e-commerce, it is generally impossible to make full accurate predictions of product success solely using NLP technology, especially when the review content and emotional expressions are complex. For example, emotional expressions may vary due to cultural, linguistic, and vocabulary differences, so that the naive sentiment classification is not precise enough to make accurate predictions. Additionally, sentiment intensity, contextual details of comments, and the impact of different emotional expressions on product success are overlooked in the majority of existing sentiment analysis models. Thus, sentiment classification results generated by sentiment analysis models cannot truly reflect consumers' general evaluation of products, nor can they accurately predict the market performance of products, so, the prediction results are biased (Singh et al., 2022; Shad et al., 2024). Therefore, the problem statement of this study is the lack of accuracy of solely utilizing sentiment analysis in predicting the success of e-commerce products.

Aim of the Project:

The aim of this project is to predict product success in e-commerce platforms by combining sentiment analysis through Natural Language Processing (NLP) with Logistic Regression and Support Vector Machines (SVM).

Objectives of the Project:

1.	To examine the relationship between customer sentiment expressed in online reviews and product success on e-commerce platforms.
2.	To train and evaluate the performance of a logistic regression and SVM sentiment analysis model to predict product success.
3.	To provide actionable business recommendations to improve product performance and marketing efficiency by combining sentiment analysis and machine learning algorithms.

Scopes of the Project:

1. Focus on online customer reviews of e-commerce sites such as Amazon, eBay, Alibaba ad so on.
2. Utilize sentiment analysis along with machine learning algorithms such as logistic regression and SVM to predict successful products.
3. The programming language used is Python.
4. The research will be limited to publicly available product reviews on these online retail websites, forecasting product success based on sentiment data.

Expected Contribution of the Project:

The expected contribution of this project is to combine sentiment analysis, logistic regression and SVM to predict e-commerce product success more precisely. Besides, it is to provide business actionable intelligence derived from machine learning algorithms and sentiment analysis to enable them to make product development and marketing decisions that are well-informed as well as to augment the understanding of the relationship between customer sentiment and e-commerce product success and add to the knowledge base in the field of data science and business analytics.

Project Requirements:

Software:	Python
Hardware:	- Intel Core i3-10110U CPU - 8GB RAM
Technology/Technique/ Methodology/Algorithm:	-Sentiment Analysis through Natural Language Processing (NLP) -Logistic Regression -Support Vector Machines (SVM)

Type of Project (Focusing on Data Science):

[√]   Data Preparation and Modeling

[   ]   Data Analysis and Visualization

[   ]   Business Intelligence and Analytics

[√]   Machine Learning and Prediction

[√]   Data Science Application in Business Domain

Status of Project:

[√]   New

[   ]   Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

[√]   Myself

[   ]   Supervisor/Industry Advisor (   )

Student Name: CHAU AI VIN

CHAU AI VIN   17/4/2025
Signature   **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature   **Date**

Name of Supervisor 2 (if any):

Signature   **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:
[   ] FULL APPROVAL   [   ] CONDITIONAL APPROVAL (Major)*
[   ] CONDITIONAL APPROVAL (Minor)   [   ] FAIL*
* Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

....

Signature   **Date**

Name of Evaluator 2:

....

Signature   **Date**