SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043

Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	Chen Junhao
Metric Number:	MCS241030
Student Email & Phone:	chenjunhao@graduate.utm.my & 60-1168122912
Project Title:	Amazon Best Seller Rank Prediction Using Machine Learning
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

The rise of e-commerce has transformed how consumers shop, with platforms like Amazon leading the way. With countless products contesting for attention, the "Best Seller Rank" has emerged as a crucial measure of a product's success. This ranking not only indicates a product's popularity but also plays a significant role in influencing buyer trust and choices. Gaining insights into the elements that impact "Best Seller Rank" can provide businesses and sellers with important information to enhance their product positioning.

Problem Background:

In today's fast-paced digital economy, Amazon has established itself as the leading global e-commerce platform, offering millions of products across categories such as electronics, fashion, books, and software. Among these, only a select few products attain the coveted "Best Seller" status — a distinction that greatly enhances visibility, consumer trust, and ultimately, sales performance.

Although it is widely assumed that factors such as product price, star ratings, number of reviews, and customer satisfaction play a role in determining a product's success, the actual dynamics influencing best-seller rankings are complex and not fully understood. Furthermore, Amazon does not publicly disclose the exact algorithm used to assign these rankings, making it an intriguing topic for data-driven investigation.

Understanding what makes a product rise to the top is valuable not only for sellers seeking to optimise their listings and pricing strategies, but also for marketers, analysts, and data scientists aiming to uncover patterns and build predictive models. Despite the commercial importance of this knowledge, there remains a gap in academic literature and applied research that systematically analyses the features influencing best-seller status on Amazon.

This project seeks to apply data science methodologies to analyse a dataset of Amazon best sellers, with the aim of answering the following research questions:

- What product features are most strongly associated with high best-seller rankings?
- Can we predict a product's best-seller status based on its attributes such as price, ratings, and review count?
- What patterns or trends can be identified within and across different product categories (if applicable)?

Problem Statement:

There exists a limited comprehension of the impact that specific product attributes have on Amazon's best seller rankings. Although critical factors such as pricing and customer reviews are accessible, sellers encounter difficulties in accurately forecasting or manipulating their ranking. This project aims to address this issue by employing machine learning algorithms to model and predict a product's best seller rankings utilising publicly available data.

Aim of the Project:

The aim of this project is to apply data science techniques to analyse and model the key factors that influence a product's best-seller status on Amazon. By examining variables such as price, star rating, and number of customer reviews, the project seeks to uncover patterns and build predictive models that explain what contributes to a product's success in Amazon's ranking system.

Objectives of the Project:

1. To explore and clean the dataset of Amazon best seller rankings, addressing any missing or inconsistent values and preparing the data for analysis.
2. To perform EDA in order to identify patterns, trends, and relationships among key variables such as product price, customer ratings, and number of reviews.

3. To determine the most influential features contributing to a product's best seller rankings through correlation analysis and feature importance techniques.
4. To develop predictive models using supervised machine learning algorithms that estimate a product's likelihood of achieving a high rank based on its attributes.
5. To evaluate model performance using appropriate statistical metrics, and interpret the results to extract meaningful insights.

Scopes of the Project:

This project is focused on analysing a dataset of Amazon best-selling products using data science techniques to uncover the factors that contribute to high sales rankings. The scope includes data cleaning, exploratory data analysis, and predictive modelling based on product attributes such as price, customer star rating, and number of reviews.

The analysis is limited to the variables provided within the dataset and does not currently account for temporal data (e.g., seasonal trends) or product categories, as these fields are not included.

The scope also includes:

- Identifying significant patterns and correlations within the data
- Building and evaluating regression or classification models to predict ranking performance
- Generating data visualisations to communicate insights clearly and effectively

Expected Contribution of the Project:

1. Insights into Key Drivers:

Price sensitivity, rating thresholds, and feature preferences for top-selling products.

2. Predictive Framework for Ranking Performance:

The development of a predictive model will offer a foundation for estimating the likelihood of a product achieving best-seller status, which can be further developed or adapted for use in e-commerce analytics.

3. Actionable Insights for Sellers and Marketers:

Findings from the project may inform data-driven strategies for product pricing, customer engagement, and review management, offering value to Amazon sellers and digital marketers.

Project Requirements:

Software: Python

Hardware:

Technology/Technique/ Methodology/Algorithm: EDA, correlation and feature importance analysis, regression

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:

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**