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:

Amazon, as the world's leading online marketplace, offers an enormous range of products across various categories. Among these, only a few manage to earn the "Best Seller" badge—a designation that boosts visibility and drives further sales. While some might speculate that factors such as price, customer ratings, and review count contribute to a product's success, the actual mechanics behind Amazon's ranking system remain vague and complex.

For businesses and individual sellers, gaining insights into what contributes to sales performance is of immense practical value. Equally, for data scientists, the challenge of making sense of such real-world data presents a rich

opportunity for analysis and modelling. Despite the commercial importance of this area, limited academic work has systematically studied what differentiates top-ranking products from others.

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 primary aim of this project is to identify and model the key factors influencing a product's best-seller status on Amazon. By analysing features such as pricing, review count, and star ratings, the study seeks to explain the dynamics of sales rankings and build predictive models that estimate performance.

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 study will focus exclusively on a static dataset of Amazon best sellers. The features under consideration include price, number of ratings, average star rating, and product rank, along with marketplace country. The scope is limited to the provided attributes; category data and time-series information are currently unavailable and therefore fall outside the initial scope.

Expected Contribution of the Project:

● E-commerce Insight: The findings will offer a clearer understanding of the factors that contribute to sales success on Amazon.

● Predictive Framework: A data-driven model will be developed to estimate product performance in the best-seller rankings.

● Skill Development: The project reinforces core data science competencies such as data wrangling, statistical analysis, and machine learning.

● Practical Application: Sellers and marketers can benefit from the insights to improve their product strategies.

- Foundation for Future Work: The research can be extended to include temporal trends, product categories, or even comparisons with other e-commerce platforms.

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
anv):

....

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**