

INTRODUCTION

1.1. Overview

The rise of e-commerce has transformed how consumers shop, with platforms like Amazon leading the way. Amazon Best Seller Rank (BSR) is a score assigned by Amazon to a product based on its sales volume and historical sales data, updated every hour. 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 BSR can provide businesses and sellers with important information to enhance their product positioning.

1.2. 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.

1.3. Research Gap

While prior studies have explored various aspects of e-commerce analytics, notable research gaps remain. Existing work on Amazon’s best-seller ranks focuses primarily on estimating sales from rank, rather than predicting rank from product attributes such as price, reviews, or ratings—limiting its utility for sellers seeking performance insights. Similarly, research into product return behaviour often neglects the influence of structured product features and lacks integration of user-generated content, such as review sentiment or star ratings, which may better predict return likelihood. Meanwhile, customer segmentation studies apply machine learning to classify user groups based on behaviour, yet they do not connect segmentation results to tangible business metrics like sales rank or product success. Furthermore, deep learning and interpretable modelling techniques remain underutilised across these domains. These limitations highlight the need for a data-driven approach that links structured product features to best-seller performance, offering both predictive insight and strategic value for e-commerce platforms.

1.4. Research Questions

- What product attributes most influence Amazon BSR?
- Predict based on review count, rating, and price?
- Which machine learning approaches are most suitable for predicting rank in such datasets?

1.5. Research Objectives

1. To explore and clean the dataset of Amazon BSR, 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 BSR 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.

1.6. Research Scope

This study will focus exclusively on a static dataset of Amazon BSR. The features under consideration include price, number of ratings, average star rating, and product rank, along with marketplace country. This study focuses on a static snapshot of Amazon best seller data. Longitudinal trends and category-level distinctions are excluded due to data unavailability.

1.7. Research Contribution

- **Theoretical:** Enhances understanding of factors driving product success on e-commerce platforms.
- **Methodological:** Demonstrates comparative analysis of machine learning techniques for BSR prediction.
- **Practical:** Offers actionable insights for online sellers aiming to improve product visibility.
- **Data & Tools:** Involves web scraping of Amazon data; analysis conducted using Python.