5. CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1. Research Summary

This research explored the application of supervised machine learning techniques to predict Amazon BSR using structured product-level data. The study utilised a dataset collected in real time from Amazon's best seller listings, focusing on software products across multiple countries. The research began with thorough data cleaning and exploratory data analysis to identify key patterns and variable distributions. Feature engineering was then conducted to construct new, informative variables such as review_density, weighted_rating, and log_num_ratings, enhancing the model's ability to capture product performance dynamics.

Three regression models were implemented and compared—Linear Regression, Decision Tree, and Random Forest. The Random Forest Regressor outperformed the others in terms of R^2, RMSE, and MAE, indicating its superior ability to model the complex and non-linear relationships in the data. Cross-validation results confirmed its stability and generalisation performance. Overall, the study demonstrated that ensemble learning methods, combined with well-crafted features, offer a robust approach to predicting BSR and can be used to inform marketing, pricing, and inventory decisions in e-commerce settings.

5.2. Research Objectives

The primary objective of this research was to develop and evaluate machine learning models capable of predicting Amazon BSR based on structured product attributes. To achieve this, the following specific objectives were formulated:

- Understand which product features most influence Amazon Best Seller Rank.

- Develop predictive models using machine learning techniques.

- Evaluate and compare the performance of different regression models.

- Identify patterns and category-level trends through exploratory analysis.

All objectives were achieved, with clear evidence that model accuracy improves with well-engineered features and non-linear approaches like Random Forest.

5.3. Research Contributions

This study makes valuable contributions to both the theoretical and practical domains of data science and e-commerce. By developing and evaluating machine learning models to predict Amazon Best Seller Rank (BSR), the research offers insights into model effectiveness, feature importance, and real-world applicability.

5.3.1. Theoretical Perspective

- From a theoretical standpoint, the research extends existing knowledge in predictive modelling by demonstrating the limitations of linear models in complex e-commerce environments. It highlights the importance of non-linear and ensemble methods—such as decision trees and random forests—in capturing the intricacies of customer behaviour and sales dynamics. Additionally, the study reinforces the significance of feature engineering, showing that derived features like review_density and weighted_rating offer greater explanatory power than raw attributes alone. The methodological design also supports the use of cross-validation as a robust approach to assess generalisability in supervised learning tasks.

5.3.2. Practical Perspective

Practically, this research provides a replicable and scalable framework for Amazon sellers, analysts, and digital marketers to forecast product ranking performance using accessible data. The final model identifies key drivers of BSR, such as price, review count, and star rating, enabling more informed decisions about pricing strategies, customer engagement, and market positioning. Moreover, the structured pipeline—ranging from data preprocessing to model evaluation—can be easily adapted for real-time applications, such as dashboards or automated product monitoring systems. In doing so, the study contributes tools and insights that are directly actionable in today's data-driven e-commerce landscape.

5.4. Research Limitations

While the study achieved its objectives and produced promising results, several limitations must be acknowledged:

A. Domain-Specific Dataset:

The dataset used in this study focused solely on software products from Amazon. As a result, the findings and model performance may not generalise to other product categories such as electronics, fashion, or books, where consumer behaviour and ranking dynamics could differ significantly.

B. Static Snapshot of Data:

Although BSR is updated hourly, the dataset represents a single snapshot in time. This limits the model's ability to capture temporal patterns or trends, such as seasonal fluctuations or the impact of promotions.

C. Limited Feature Scope:

The analysis relied entirely on structured data such as price, ratings, and number of reviews. It did not incorporate textual data like product descriptions or customer reviews, which could contain rich signals relevant to product success.

D. No External Validation:

The model's performance was validated using train-test splits and cross-validation within the dataset. However, the model was not tested on a completely independent dataset or deployed in a real-world environment, which limits the assessment of its robustness in practice.

E. Country-Level Bias:

Some countries in the dataset were underrepresented, which may have led to biased country-specific features. This imbalance could affect the model's ability to generalise across global marketplaces.

Despite these limitations, the research establishes a solid foundation for future studies and applications involving BSR prediction.

5.5. Future Work Recommendations

To build upon the findings of this research, future work should consider expanding both the depth and breadth of the analysis. One valuable direction would be to incorporate temporal data, allowing for time-series modelling that captures changes in BSR over time and better reflects the real-time nature of Amazon rankings. Additionally, future studies could benefit from applying the model across multiple product categories beyond software, enabling broader generalisation and uncovering category-specific trends. Enhancing the dataset with textual features such as product descriptions and customer reviews could also improve prediction, particularly through sentiment analysis and natural language processing techniques. Further optimisation of model hyperparameters using advanced tuning methods, such as grid search or Bayesian optimisation, may yield better performance and greater reliability. Moreover, deploying the model in a live, real-time environment would increase its practical utility for sellers and analysts. Exploring alternative machine learning approaches, including gradient boosting methods or deep learning architectures, could also enhance prediction accuracy and adaptability. Overall, future research should aim to create more robust, scalable, and generalisable BSR prediction systems suitable for dynamic and competitive e-commerce platforms.

5.6. Summary

This chapter has brought the research to its conclusion by revisiting the key components and outcomes of the study. It began with a summary of the research process, highlighting the use of supervised machine learning techniques to predict Amazon Best Seller Rank (BSR) from structured product data. The research objectives were clearly defined and successfully achieved, leading to valuable theoretical and practical contributions. While the study demonstrated that models like Random Forest can effectively capture non-linear relationships within e-commerce data, it also acknowledged several limitations, such as a narrow product focus and the exclusion of time-based and textual features. These limitations present opportunities for future improvement. Recommendations were offered to guide subsequent research efforts, including the incorporation of temporal data, broader product coverage, and deployment in real-time environments. Overall, the research provides a foundational approach for data-driven BSR prediction and contributes meaningful insights to both academia and industry.