Amazon's Best Selling Books

1. Introduction:

Amazon is one of the largest e-commerce platforms in the world, and it offers a wide range of products including books. The collected dataset can include various features such as book title, author, publication year, price, customer rating, and book categories (Genre). The dataset is intended to provide insights into consumer behavior, book trends, and other factors that can influence sales on Amazon. And our aim is to develop a model that makes it possible to foresee how many books are being sold and what kind are most popular.

2. Methodology:

2.1 Data Set: This Kaggle dataset comprises a decade's worth of data in 550 rows and 7 columns the genre is the dependent column, And this allowing us to study insights, patterns and creating data-driven solutions for the book business and this dataset has helped us a lot to scale small business owners in order to maximize their profit.

2.2 Data Preprocessing: Many preprocessing steps must be completed before developing the model, such as filling in missing values and ensuring that each author's string length is the same throughout the dataset for the model to treat them equally. Additionally, since machine learning models only understands digits, categorical features must be encoded.

2.3 Model Description: We Implemented a wide range of machine learning algorithms to forecast Amazon best-selling books. Logistic Regression, Random Forest Classifier, SVC, KNN Classifier, Decision Tree Classifier, As more models are considered, the accuracy of the Random Forest classifier on this dataset has been rather good followed by Decision Tree and SVC.

2.4 Implementation: To execute the models, the Python programming language was utilized, and famous libraries like pandas, scikit-learn, seaborn, and matplotlib were utilized. These libraries made information preprocessing, model preparation, and assessment

a lot simpler to effectively perform. We had the option to guarantee an effective and fruitful course of execution by utilizing these different apparatuses.

2.6 Computational Requirements: Depending on the intricacy of the underlying method and the volume of data being examined, our models' processing needs vary. In order to achieve these requirements, we had to resort to using a computer with a sizable amount of computing power as well as memory capacity. This made guaranteed that the models could be trained and evaluated quickly without affecting the Amazons Best Books Seller performance or accuracy. This made it possible to train and assess the models effectively.

The computational resources required are determined by the quantity of the dataset and the complexity of the research. A normal laptop or desktop with at least 8GB of RAM and a current CPU would be sufficient for a dataset of our project. However, for advanced analyzing jobs or larger datasets, more powerful hardware or cloud computing resources may be required..

References to Research Papers:

As part of our project, we refered academic papers in-depth in order to gain knowledge and inspiration for creating our Amazon's best book seller system.

[1] Maity, Suman Kalyan, Abhishek Panigrahi, and Animesh Mukherjee. "Book reading behavior on goodreads can predict the amazon best sellers." *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. 2017.

[2] Bell, Reginald, et al. "An examination of differences between the most influential management books of the 20th century and amazon best sellers." *International Journal of Business Research and Information Technology* 3.1 (2016): 35-78.

We learned a lot from these references about the methodologies, algorithms, and evaluation strategies for Amazon's Best Book Seller prediction research.