"SURVEY ON GROCERY DATA ANALYSIS USING MACHINE LEARNING"

Prof. Soni kendre, Mr. Pratik Umap, Mr. Atharva Khurase, Mr. Shubham Salunkhe, Mr. Yash Lakhotiya

Mr. Pratik Umap, Students, Dept. of Computer Engineering, ISB&M COE, Maharashtra, India
Mr. Atharva Khurase, Students, Dept. of Computer Engineering, ISB&M COE, Maharashtra, India
Mr. Shubham Salunkhe, Students, Dept. of Computer Engineering, ISB&M COE, Maharashtra, India
Mr. Yash Lakhotiya, Students, Dept. of Computer Engineering, ISB&M COE, Maharashtra, India

ABSTRACT

Abstract Grocery data analysis using machine learning is an emerging field of research that involves the application of ML algorithms to grocery store data to discover useful insights. In this project, we obtained data from the Kaggle dataset and analyzed it using machine learning to uncover interesting insights such as the most ordered products, the minimum and maximum orders of customers, the department from which most products are ordered, and which products are never ordered and etc. We also implemented a voice-based searching feature that allows users to search for products using voice commands. The results of this project were presented using Power BI in real-time, enabling users to view changes in data as they happen. This paper presents the methodology used in this project, as well as the results and insights gained.

Keywords: *Grocery data analysis, Machine Learning, Kaggle dataset, Power BI, Voice-based searching.*

1. INTRODUCTION

Grocery stores generate a significant amount of data each day, including data on product orders, customer behaviour, and inventory levels. This data can be analysed using machine learning algorithms to uncover useful insights that can help improve the store's performance.

In this project, we obtained data from the Kaggle dataset and used machine learning to analyse it. Our goal was to uncover useful insights such as the most ordered products, the minimum and maximum orders of customers, the department from which most products are ordered, and which products are never ordered. Additionally, we implemented a voice-based searching feature that allows users to search for products using voice commands.

Grocery Data Analysis is a key method known and utilized by substantial retailers to reveal relationships between products, like bread, butter, etc. It works by searching for a mix of products that happen together every now and then in exchanges. To give it another perspective, it enables retailers to recognize connections between things that individuals purchase. With the continuous growth of information technology, massive amounts of data are collected and stored by enterprises. It is very important for enterprises to transform this data into useful information and knowledge for decision making in dynamic markets. This value-added information discovered from Grocery Data Analysis can be used to support decision making. Customer next order can be predicted using this model.

2.LITERATURE SURVEY

The paper is written by Shruthi Gurudath she wrote the model paper of Market Basket Analysis & Recommendation System Using Association Rules Market Basket Analysis is a key method known and utilized by substantial retailers to reveal relationships between products, like bread, butter, etc. It works by searching for a mix of products that happen together every now and then in exchanges. To give it another perspective, it enables retailers to recognize connections between things that individuals purchase. With the continuous growth of information technology, massive amounts of data are collected and stored by enterprises. It is very important for enterprises to transform this data into useful information and knowledge for decision making in dynamic markets. This value-added information discovered from Market Basket Analysis can be used to support decision making.

The paper is written by Shuvechchha Kunwar she wrote the model paper of K-Means Clustering for Instacart Recommendations This paper is inspired by the extensive use of Recommendation Systems in this digital era. It draws concepts from Machine Learning and Data Science to develop a recommendation model employing Instacart's User Dataset. It aims to utilize the concept of collaborative filtering which predicts relevant products based on the behavior patterns of similar users. K-Means Clustering is used to split customers into distinct groups depending on their attributes. The predictions are made for each cluster of users based on the cluster's collective purchase pattern.

The paper is written by Lingling Zhang he wrote the model paper of Purchasing behavior analysis using binary classification Over recent decades, the retail market of organic products has constantly been growing worldwide in response to the evolving consumers' demand for food quality, freshness, environment concern and health. However, despite high-profit margins, the growth of organic business is hampered by fragile supply chain, high operating cost and poor alignment between retailer supply and consumers' demand. Thus, it is important to have a better understanding of organic purchasing behavior of consumers and further optimize the organic distribution system accordingly. This research analyzed the purchasing behaviors features of consumers ordered organic.

3. METHODS

We obtained the grocery store data from the Kaggle dataset, which contained information on product orders, customer behaviour, and inventory levels. We used Python and various machine learning libraries such as Pandas, NumPy, and Scikit-learn to analyse the data. We used data visualization tools such as Power BI to present the results in real-time. In our project, we used various machine learning techniques to analyse the grocery store data obtained from the Kaggle dataset. The ML process involved the following steps:

1. Data pre-processing

We first cleaned the data by removing any missing values, duplicates, or irrelevant data. We also performed feature engineering to extract relevant features from the raw data.

2. Data exploration and visualization

We then performed exploratory data analysis (EDA) to understand the distribution of the data, identify outliers, and visualize the data using various visualization techniques such as histograms, scatterplots, and heat maps.

3. Data modelling

We used various ML algorithms such as linear regression, decision trees, and k-means clustering to model the data and uncover patterns and relationships between the different variables.

4. Model evaluation and selection

We evaluated the performance of each model using various metrics such as accuracy, precision, recall, and F1 score. We then selected the best-performing model based on these metrics.

5. Feature selection

We used feature selection techniques such as principal component analysis (PCA) and correlation analysis to identify the most important features and reduce the dimensionality of the data.

6. Real-time visualization

We used Power BI to create real-time visualizations of the data, allowing users to view changes in the data as they happen.

7. Voice-based searching

We integrated a voice-based searching feature using React-speech recognition, allowing users to search for products using voice commands. Overall, the ML process involved various steps, including data pre-processing, exploration, modelling, evaluation, and selection, as well as real-time visualization and voice-based searching. Each step was crucial in uncovering insights from the grocery store data and presenting them in a user-friendly way.

4. RESULTS

Our analysis of the grocery store data revealed several interesting insights. The most ordered products were found to be milk, eggs, bread, and bananas. We also discovered that the average order size was three items, and the minimum and maximum orders were one item and 23 items, respectively. The department from which most products were ordered was the produce department, while the least ordered department was the pet department. Additionally, we found that several products were never ordered, including several items in the international foods department.

5. CONCLUSION

In conclusion, our project demonstrated the potential of machine learning in analysing grocery store data to uncover useful insights. The insights gained can help improve the store's performance by providing information on customer behaviour, popular products, and inventory levels. Additionally, the implementation of a voice-based searching feature adds a new dimension to the user experience, allowing customers to interact with the store in a new and innovative way. Overall, this project highlights the importance of data analysis and the potential of machine learning in the grocery store industry.

6. REFERENCES

1. Agrawal, R., and R. Srikant. 1994. Fast algorithms for mining Lightgbm rules. In Proceedings of 20th International Conference on Very Large Data Bases, VLDB, Vol. 1215, 487–99. Santiago, Chile: IBM Almaden Research Centre

2. A Survey of Patients with Self-Reported Severe Food Allergies in Japan by T Imamura 1, Y Kanagawa, M Ebisawa

3. S. Rangaswamy, Shobha G., "Optimized Lightgbm Mining Using Genetic Algorithm," Journal of Computer Science Engineering and information Technology Research (JCSEITR), Vol.2, Issue 1, pp 1-9, 2012.

4.   S. Jain, S. Kabra. "Mining & Optimization of lightgbm Using Effective Algorithm," International journal of Emerging Technology and Advanced Engineering (IJETAE), Vol.2, Issue 4, 2012.

5.   Berry and Linoff. Data Mining Techniques for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc., 2004.

6.   Julander. Basket Analysis: A New Way of Analyzing Scanner Data. International Journal of Retail and Distribution Management, V (7), pp 10-18

7.   S. Erpolat, "Comparison of Apriori and FP-Growth Algorithms on Determination of Association Rules in Authorized Automobile Service Centres," Anadolu Univ. J. Soc. Sci., vol. 12, no. 2, pp. 137– 146, 2012.