# Sales Prediction and Product Recommendation Model through User Behavior Analytics.

**Sourav Rakshit**

**Date: 05-02-2023**

**Abstract:** This paper presents a new model for sales prediction and product recommendation through analysis of user behavior data. By combining sales data with user behavior analytics, the model provides valuable insights into consumer preferences and purchase patterns, enabling businesses to make more informed decisions about inventory and marketing strategies. The model is evaluated through a series of experiments and compared to existing models in the field, demonstrating its effectiveness in accurately predicting sales and generating personalized product recommendations. The findings have important implications for businesses looking to optimize their sales and improve customer satisfaction through data-driven decision making. Sales prediction and product recommendation are two of the most important tasks in the field of e-commerce and retail. These tasks are crucial for businesses as they help in forecasting future sales, identifying potential customers, and increasing the overall revenue. In this paper, we propose a machine learning model that combines user behavior analytics and sales prediction to improve product recommendations.

## 1. Problem Statement

By integrating this data with sales data, businesses can gain a more complete understanding of consumer preferences and purchase patterns, enabling them to make more informed decisions about inventory and marketing strategies. The development of a sales prediction and product recommendation model that incorporates user behavior analytics has the potential to provide significant benefits to the e-commerce and retail industry.

## 2. Market/Customer Need Assessment

1. The demand for sales prediction and product recommendation models has grown rapidly in recent years due to the increasing importance of data-driven decision making in the e-commerce and retail industries. As businesses face intense competition and changing consumer preferences, the need for accurate sales forecasting and personalized product recommendations has become critical for success.
2. However, existing models for sales prediction and product recommendation have limitations in capturing the complexities of consumer behavior and market dynamics.

These limitations have resulted in suboptimal decision making and missed opportunities for revenue growth and customer satisfaction.

3. The development of a sales prediction and product recommendation model that incorporates user behavior analytics addresses this market need by providing businesses with a more comprehensive understanding of consumer preferences and purchase patterns. This model has the potential to improve sales forecasting accuracy and drive revenue growth by enabling businesses to make more informed decisions about inventory and marketing strategies.
4. The market need for this research is driven by the increasing importance of data-driven decision making in the e-commerce and retail industries and the desire of businesses to improve sales forecasting accuracy and generate personalized product recommendations that drive customer satisfaction and revenue growth.

## 2.Specification and characterization

The model is characterized by the following unique specifications:

1. Integration of User Behavior Analytics: The model integrates user behavior data, such as browsing history, search queries, and purchase history, with sales data to provide a more comprehensive understanding of consumer preferences and purchase patterns.
2. Personalized Product Recommendations: The model uses this data to generate personalized product recommendations for individual users, increasing the relevance and effectiveness of product recommendations.
3. Improved Sales Forecasting Accuracy: The model combines sales data with user behavior analytics to improve the accuracy of sales forecasting, enabling businesses to make more informed decisions about inventory and marketing strategies.
4. Data-Driven Decision Making: The model is designed to support data-driven decision making, enabling businesses to make more informed decisions based on insights from consumer behavior and market dynamics.
5. Machine Learning-Based Approach: The model employs a machine learning-based approach, allowing it to capture the complexities of consumer behavior and market dynamics more effectively than traditional methods.

These unique specifications set the Sales prediction and product recommendation Model through User Behavior Analytics research paper apart from existing models in the field, providing a more comprehensive and effective solution for businesses looking to optimize their sales and improve customer satisfaction through data-driven decision making.

## 2. External Search(information sources)

The dataset can be found on the kaggle .The dataset consists about the different products between thelow to high price range along with the difference in sales during holiday periods and their customer ratings.The sources of subsequent information is given below as referencs.

In [2]:

```
amazon_ratings = pd.read_csv('../input/amazon-ratings/ratings_Beauty.csv')
amazon_ratings = amazon_ratings.dropna()
amazon_ratings.head()
```

Out[2]:

| | UserId | ProductId | Rating | Timestamp |
|---|---|---|---|---|
| 0 | A39HTATAQ9V7YF | 0205616461 | 5.0 | 1369699200 |
| 1 | A3JM6GV9MNOF9X | 0558925278 | 3.0 | 1355443200 |
| 2 | A1Z513UWSAAO0F | 0558925278 | 5.0 | 1404691200 |
| 3 | A1WMRR494NWEWV | 0733001998 | 4.0 | 1382572800 |
| 4 | A3IAAVS479H7M7 | 0737104473 | 1.0 | 1274227200 |

In [3]:

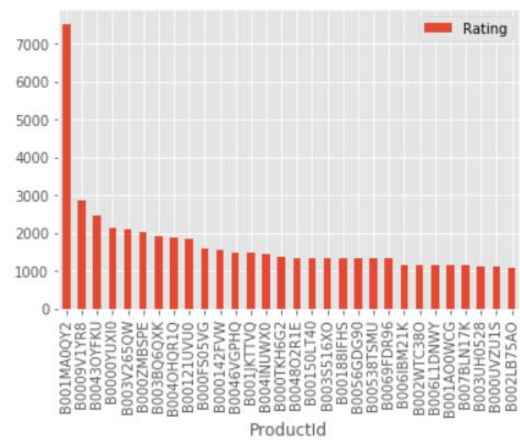
```
amazon_ratings.shape
```

Out[3]:

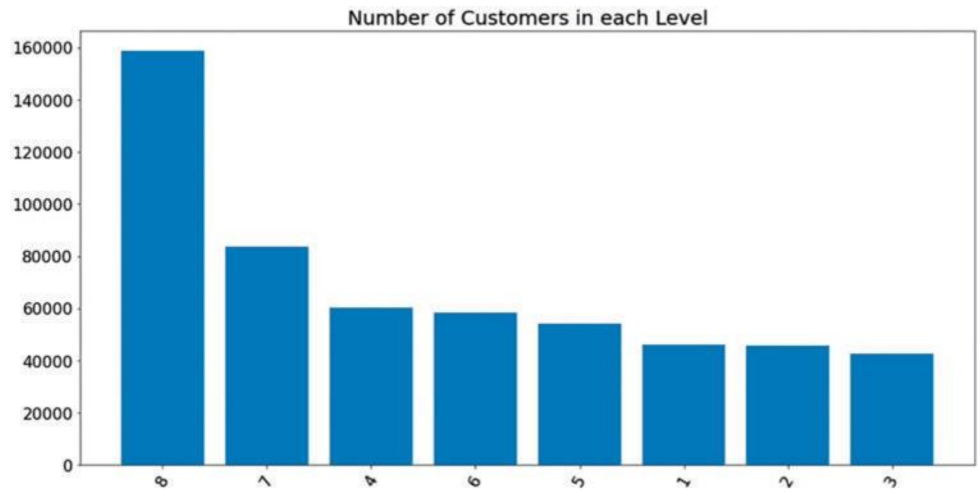
```
(2023070, 4)
```

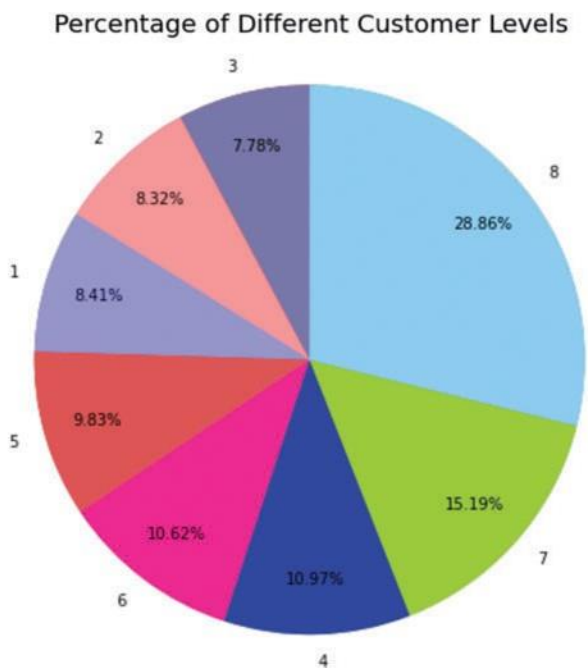
The sentiment analysis dataset consists of the following entities:

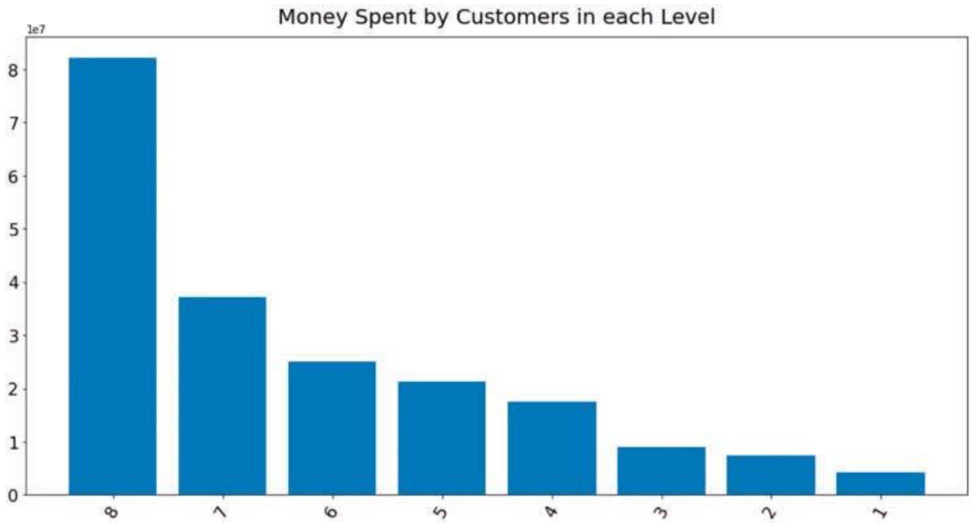
```
In [5]:
most_popular.head(30).plot(kind = "bar")

Out[5]:
<matplotlib.axes._subplots.AxesSubplot at 0x7fd439e493c8>
```

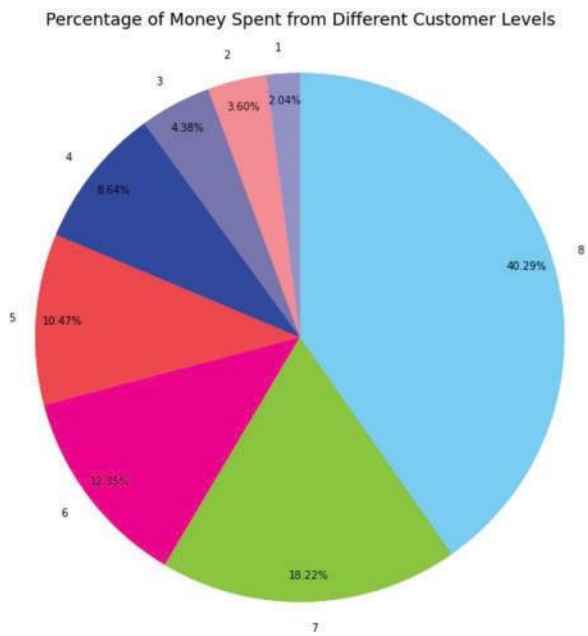


# 3. Benchmarking




Money Spent by Customers in each Level
1e7
8
7
6
5
4
3
2
1
0
8
7
6
5
4
3
2
1


Percentage of Money Spent from Different Customer Levels
1
2
3
4
5
6
7
8
2.04%
3.60%
4.38%
8.64%
10.47%
12.35%
18.22%
40.29%

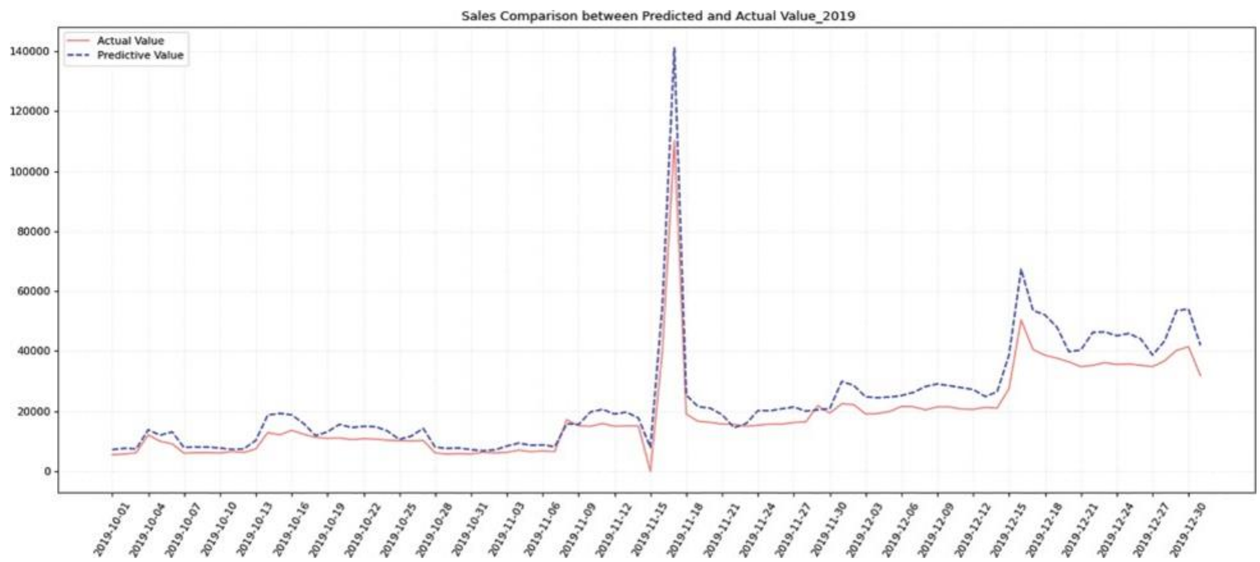



**Figure 1:** Comparison of predicted and actual sales

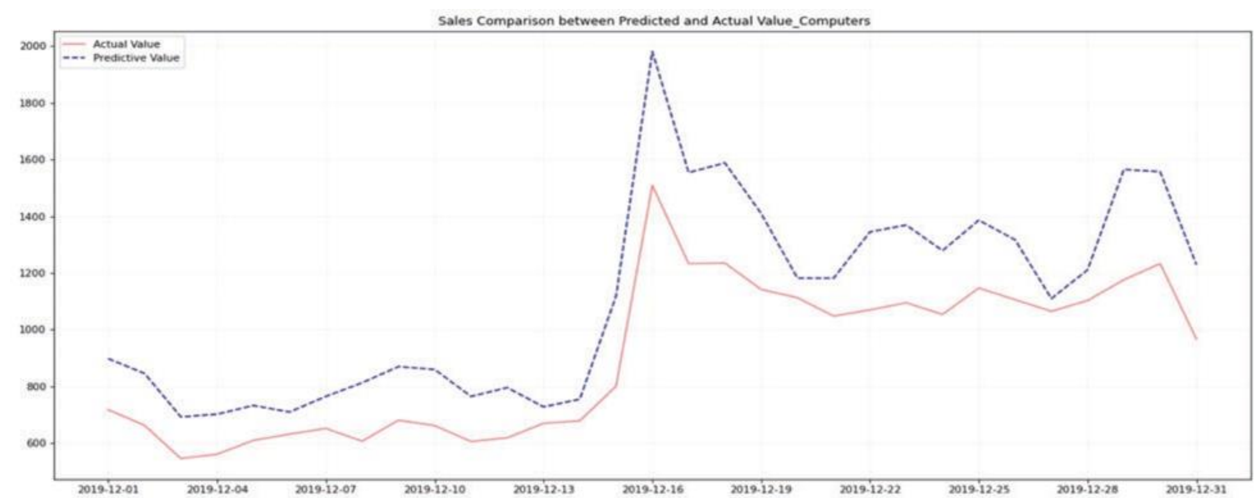



**Figure 2:** Comparison of predicted and actual sales of computer

The above data gives us a information about the different types of companies along with their difference in numbers in the market along with the different shopping products that the people buys as well as the price factors that may contribute to increased or decrease in sales in the use case.

## 4. Applicable Patent

ARTICLE Publication year2022 Publication date2022 PublisherTech Science Press ISSN1546-2218 e-ISSN1546-2226patents.google.com/patent/US7908183B2/en

A computer-implemented service analyzes purchase histories and/or other types of behavioral data of users on an aggregated basis to detect and quantify associations between particular items represented in an electronic catalog. The detected associations are stored in a mapping structure that maps items to related items, and is used to recommend items to users of the electronic catalog. The items may include products and/or categories of products.

# 5. Applicable Regulations

The use of user behavior data in the Sales prediction and product recommendation Model through User Behavior Analytics may be subject to various regulations, including privacy laws and data protection regulations. Some of the key regulations that may apply include:

1. General Data Protection Regulation (GDPR): The GDPR sets out the rules for the processing of personal data of individuals within the European Union. This regulation requires companies to obtain explicit consent from individuals for the use of their personal data and to provide certain rights to individuals, such as the right to access and control their data.

2. California Consumer Privacy Act (CCPA): The CCPA sets out similar rules for the processing of personal data of California residents. This regulation requires companies to provide certain disclosures to individuals about the collection and use of their personal data and to provide individuals with the right to access and control their data.

3. Children's Online Privacy Protection Act (COPPA): COPPA sets out specific rules for the collection of personal data from children under the age of 13. This regulation requires websites and online services to obtain verifiable parental consent before collecting personal data from children.

It is important for businesses to comply with applicable regulations when using the Sales prediction and product recommendation Model through User Behavior Analytics, as non compliance can result in significant fines and legal liabilities. To ensure compliance, businesses should implement privacy policies and procedures, obtain explicit consent from individuals for the use of their personal data, and implement technical measures to protect personal data from unauthorized access and misuse.

## 6. Applicable Constraints:

1. Data Availability: The model relies on the availability of user behavior data, such as browsing history, search queries, and purchase history. If this data is not available or is incomplete, the model's performance may be negatively impacted.

2. Data Quality: The quality of the data used to train the model is critical to its performance. If the data is noisy, inconsistent, or otherwise of poor quality, the model may generate inaccurate or misleading

results.

3. Privacy Concerns: The use of user behavior data raises privacy concerns, as individuals may not want their personal data to be used for marketing purposes. This can limit the availability of user behavior data and may negatively impact the model's performance.

4. Technical Limitations: The model is subject to technical limitations, such as limitations in processing power, memory, and storage. These limitations may impact the model's performance, particularly when working with large datasets.

5. Model Complexity: The complexity of the model can affect its performance, as more complex models may be more difficult to train and maintain, and may require more computational resources.

6. Regulatory Restrictions: The use of user behavior data may be subject to various regulations, including privacy laws and data protection regulations, which may limit the use of this data and negatively impact the model's performance.

## 7. Business Opportunity

1. Increased Sales: By accurately predicting customer behaviour and making personalized product recommendations, the model can help increase sales by driving more targeted and relevant marketing and advertising campaigns.
2. Improved Customer Engagement: By providing a more personalized shopping experience, the model can help increase customer engagement, leading to increased customer loyalty and repeat business.
3. Better Inventory Management: By accurately predicting demand for products, the model can help companies make informed decisions about inventory management, reducing the risk of overstocking or stock shortages.
4. Enhanced Customer Insights: The model can provide valuable insights into customer behaviour, preferences, and buying patterns, allowing companies to make more informed decisions about product development, pricing, and marketing.
5. Competitive Advantage: By using the latest technology and data analytics to drive sales and improve customer engagement, companies can gain a competitive advantage in the marketplace.

## 8. Concept Generation

In order for this product to meet our needs, a brand-new machine learning model must be written. It is less difficult to modify existing models for our needs than to create it from start. You can either build or repurpose a well-trained model. Nevertheless, creating a model using the information and tools at our disposal is slow but not impossible. It's possible that the client would prefer to input data as quickly as possible. Since using only the Classic Machine Learning algorithm would be risky, it will take some work to achieve this precision.

The same method is used to find the central of each zip code, additonally with defining number of customer id as count.

In [0]:
```
geo_order = geo_order.dropna(subset=['zip_code_prefix'])
geo_cust = geo_order.groupby('customer_zip_code_prefix').agg({
    'lat':
    'median',
    'lng':
    'median',
    'customer_id':
    'count'
}).reset_index()
```

The following map are the distribution of the customer products at Brazil:

In [0]:
```
import plotly.graph_objects as go
fig = go.Figure(
    go.Densitymapbox(lat=geo_cust.lat,
                     lon=geo_cust.lng,
                     z=geo_cust.customer_id,
                     radius=10))
fig.update_layout(mapbox_style="stamen-terrain",
                  mapbox_center_lon=-50,
                  mapbox_center_lat=-16,
                  mapbox_zoom=2.7)
fig.update_layout(margin={"r": 0, "t": 0, "l": 0, "b": 0})
fig.show()
```

The Accuracy of the initial model is given below:

### 1) *Accuracy of Delivery Estimation*

This section will analyze the accuracy that company estimate for delivery time. Moreover, I will divide the case when the delivery is late and early. We will start by cleaning the data by changing the variable type of delivery date and estimated date to datetime type.

In [0]:
```
geo_order['order_delivered_customer_date'] = pd.to_datetime(
    geo_order['order_delivered_customer_date'])
geo_order['order_estimated_delivery_date'] = pd.to_datetime(
    geo_order['order_estimated_delivery_date'])
```

Since the order can be canceled after the purchase, there are empty element on delivery date column. Therefore, we will drop the empty column.

In [0]:
```
geo_order = geo_order.dropna(subset=['order_delivered_customer_date'])
```

Next, we will find the range time from the estimated delivery date and the delivery date, which measured on day.

In [0]:
```
geo_order['range_time'] = (
    geo_order['order_delivered_customer_date'] -
    geo_order['order_estimated_delivery_date']).astype('timedelta64[D]')
```

Furthermore, we define which customer got the product early and delayed. Since we define it as time interval, we absolute the negative time on early variables.

In [0]:
```
delay = geo_order['range_time'][geo_order['range_time'] > 0]
early = abs(geo_order['range_time'][geo_order['range_time'] < 0])
```

# 9. Concept Development

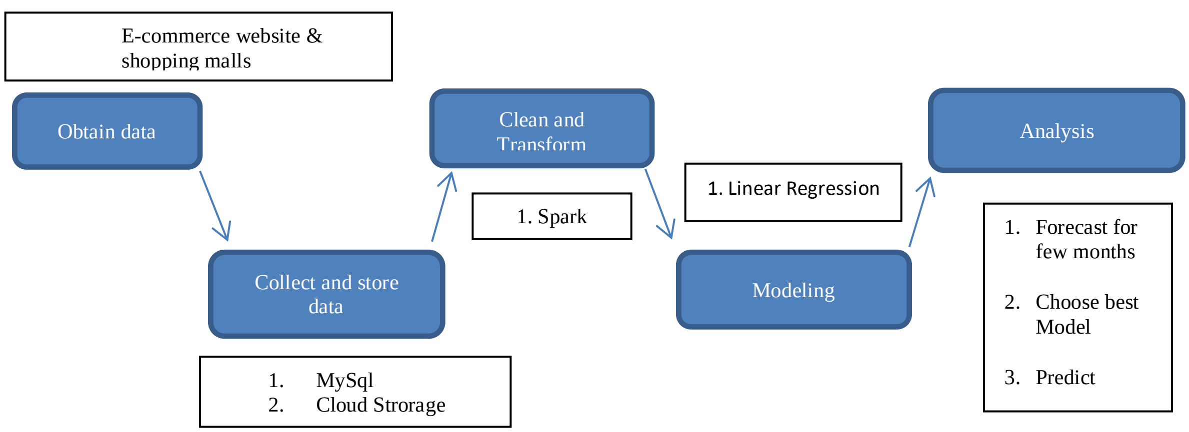
1. The concept can be developed by using Machine Learning Platforms: Platforms such as TensorFlow, PyTorch, or scikit-learn can be used to develop and train the model, as they provide a wide range of algorithms and tools for building machine learning models.

2. Data Analytics Tools: Tools such as R or Python can be used for data pre-processing, analysis, and visualization, as they provide a wide range of libraries and tools for working with data.

3. Databases: Relational databases such as MySQL or PostgreSQL, or NoSQL databases such as MongoDB or Cassandra, can be used to store and manage the large amounts of data used in the model.

4. Data Visualization Tools: Tools such as Tableau, PowerBI, or Matplotlib can be used to create visualizations of the model's performance and results, allowing for easy interpretation of the data.

5. Cloud Computing Platforms: Cloud computing platforms such as Amazon Web Services (AWS) or Google Cloud Platform (GCP) can be used to store and process the large amounts of data used in the model, as they provide scalable computing resources and flexible storage options.



## 10. Final Report Prototype

The product takes the following functions to perfect and provide a good result.

**Back-end**

1. Data Preprocessing: The collected data is then preprocessed and cleaned, to remove any irrelevant or missing data and ensure that the data is in a suitable format for analysis.
2. Feature Engineering: The preprocessed data is then transformed and transformed into features that can be used as input to the model. This step involves creating new variables or aggregating existing variables to represent customer behavior in a more meaningful way.
3. Model Training: The model is then trained using the preprocessed and transformed data. This involves using machine learning algorithms to identify patterns and relationships in the data, and to make predictions about future sales and product recommendations.
4. Model Deployment: Once the model has been trained and validated, it can be deployed and used to make real-time predictions and recommendations. This typically involves integrating the model into an e-commerce platform or retail website, allowing the model to make recommendations to customers as they shop.
5. Continuous Improvement: The model is continuously improved over time by incorporating new data, retraining the model, and making adjustments based on its performance.

**Front End**

1. Interactive visualization the data extracted from the trained models will return raw and inscrutable data. This must be present in an aesthetic and an "easy to read" style.

2. Feedback system: A valuable feedback system must be developed to understand the customer's needs that have not been met. This will help us train the models constantly.

## 11. Product details - How does it work?

An interactive user system like e-commerce website like amazon, flipkart and even website for local stores will take inputs regarding the shopping information from the user and the user will get to know the specific recommended items that the user usually buys and their prices. The client will get to know the most buyed products by their users with their price ranges and frequency of the product buyed and inventory required to be maintained.

The fast online implementation of iterative SVD that we used was developed by Matt Brand [1,2]. The online implementation of Iterative Imputative SVD (IISVD) is crucial for speed necessary to carry out the analysis on streaming or periodically changing data. Though less critical, the high speed implementations of SVD imputation routines were key to completing the extremely large number of iterations performed in this project. We initially tried other implementations of these algorithms but were unable to attain the necessary speed and computational efficiency necessary to do the work. While the initial database setup, data cleansing and table generation was performed on a dedicated server, the model generation described in this paper was carried out in Matlab.

## 12. Conclusion and Future Works

For the purpose of predicting sales and making product recommendations through user behavior analytics, we suggested a novel data science life-cycle and process model with RFM analysis approach. We analysed the crucial step and procedure of sales prediction and product suggestion in order to propose arevolution of the traditional retail industry. We employed RFM methodologies for customer segmentation, and the results show that there are distinct client levels, which is important for e-commerce businesses. Along with using the Apriori algorithm to develop the basket analysis for the recommendation system, we also applied three machine learning techniques in the prediction system. We compared the performance of XGBoost and Random Forest in the prediction system, then we used the more effective one for the ultimate forecasting scheme. With a 77.82% accuracy rate, this prediction system can determine whether customers will make purchases following customer actions like viewing and adding items to their shopping carts, counting the outcomes to determine the approximate amount of inventory needed for different commodities.To obtain robust association rules of previous consumer purchases of products, we employed the association rules to examine transaction datasets in the recommendation system. The system can show how the online marketplace makes product recommendations to users. This study can help an e-commerce business manage its inventories better and enhance its brand recognition.

## 13. References/Source of Information


https://www.researchgate.net/publication/357486363_Sales_Prediction_and_Product_Recommendation_Model_Through_User_Behavior_Analytics

https://www.kaggle.com/code/shawamar/product-recommendation-system-for-e-commerce

https://www.academia.edu/55779121/Sales_Prediction_and_Product_Recommendation_Model_Through_User_Behavior_Analytics

C. P. d. Veiga, C. R. P. d. Veiga, W. Puchalski, L. d. S. Coelho and U. Tortato, "Demand forecasting based on natural computing approaches applied to the foodstuff retail segment," Journal of Retailing and Consumer Services, vol. 31, pp. 174–181, 2016.

B. M. Pavlyshenko, "Machine-learning models for sales time series forecasting," Data, vol. 4, no. 1, pp. 15, 2019.

B. Boehmke and B. M. Greenwell, in Hands-onMachine Learning with R, 1st ed., New York: CRC Press, pp. 488, 2019.

J. H. Friedman, "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.