Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7722264

E-COMMERCE BASED PROTOTYPE FOR PREDICTING THE PRODUCT RETURNS

Vidya Rajasekaran[1] Dr. Latha Tamilselvan[2*]
[1, 2*] Department of Information Technology, B.S. Abdur Rahman Crescent Institute of Science and Technology, India, latha.tamil@crescent.education

Abstract**—**Return To Origin (RTO) is a major issue faced by e-commerce businesses. We develop a prototype to predict the product return made by the customers. Our model predicts the return of the products in advance so that necessary measures can be carried out to minimize the overall operational and financial losses. The research focuses on obtaining a normalized return score using three entities: consumer's return behavior, product return rate from Vendor's, and the feedback submitted by the customers. A hypothesis is set with the conditional statements and the obtained return score is checked to the range in which it falls and conclusions are finally driven out. The part of the system is implemented in python using a real-time dataset. The overall intention of this research is to decrease the return rate. The increase in RTO affects the revenue of the e-commerce firm and also for the vendors, so this model can assist them in decision-making for minimizing their losses.

Index Terms— Consumer Behavior, Sentiment Analysis, Natural Language Processing, Polarity, Predictive Model, E-commerce

1. INTRODUCTION

THE growth of the e-commerce industry is continuously increasing expressly. This increases the sales of the firm and on the hidden side; the revenue of the e-commerce firm is greatly affected by the increased numbers of RTO. RTO is the return of a delivered product from the consumer, back to the origin. In a study, Barclayard has calculated that every year in the UK customers making shopping online is estimated to be returning £7 billion of purchases [5]. The returns made for online products are 30% when compared to the Brick-and-Mortar stores where the returns observed are 8.89% [11]. This kind of return causes greater trouble for the vendors and the firm, which in turn increases their operational costs. Strategies should be developed to minimize the return rates and then the internal process causing the returns is needed to be streamlined. The reason for return varies and each return needs to be studied carefully.

Several reasons are contributing to the returns and the major three reasons are observed as (i) the wrong item received other than ordered (ii) the received product looks different from the product shown in the catalog while placing the order (iii) the product received in damaged condition. The main focus of this research is to minimize the count on returns with a detailed examination of the cause of returns. There are several causes promoting returns that include low-quality images and minimum description on products, size and measurement issues, low customer service, improper packaging, late delivery, and so many other issues. By knowing the proper reason for returns the firms and vendors can build up the necessary measures to minimize the returns. The firms should develop standard return policies so that the customers are not disturbed. A prediction system can be developed to predict any kind of product returns in advance.

Several shreds of evidence are stating the returns and their issues. Based on the statistics [22], most of the customers check for the return option while making their purchases through e-commerce.

2. LITERATURE SURVEY

Rajagopalan and Ward [1] have developed a prediction model for analyzing the product return volume for a company. Their work has majorly focused on the quality issues contributing to returns and the measures and their impacts in improving the quality reduce the return volume. Their work has shown results with good prediction accuracy. They have also implemented the model across several machine learning techniques and have stated LASSO to be proven to yield better accuracy when compared to the rest of machine learning techniques. The manufacturers can also use this framework to track product returns but there is no guarantee that a product has been returned only because of the quality issue. There are several other numerous issues where a product return is made and the author has not put effort into including the other several related issues. Our study focuses on several historical parameters causing the returns and uses mathematical techniques to find the prediction on returns, so where the prediction accuracy can be improved more.

Dheeraj et al. [3] the research focuses on developing a customer-oriented personalized search engine based on the customers' search history. The profile of the customer is maintained and their search preferences are studied. The final output of the research delivers an e-commerce tool that serves in assisting the customers in selecting the e-commerce website based on the obtained ranking of websites for purchasing a specific product. In this work, the consumer's behavior is studied, and similarly, in our research work, we try to study the consumer behavior based on the purchases and returns made.

The research by Esmeli et al. [4] is carried out to make an earlier purchase prediction using machine learning models. Our work concentrates on the return aspect of the dispatched products.

The work in [6] by Tolga and Allan focuses on comparing the performance of return prediction models based on the statistical and economic measures and evaluating their performances. The research is involved in predicting the stock returns. The idea behind the research is to predict the returns in advance and our research also focuses on return prediction using different strategies.

This paper [7] by Huibing and Junchao introduces an ensemble model for repeated buyer's predictions. The purchase behavior of the customer is studied in this paper and our work concentrates on the return behavior of the customers which is also an important entity contributing to the losses and revenue of the e-commerce firm.

The research by Abdullah and Nur Amin [8] was carried out to study the order return through predictive modeling. The most common factor which influenced the return orders was also included in the research. The research focuses on the most common challenges causing returns. Our research is very similar to this work and differs in the state of prediction mechanisms used to predict the returns.

A Mahalanobis feature extraction technique is developed in this paper [9] by Kranz al. A decision support system is introduced to predict the product returns and in our research, we develop an alarming system to denote the return and make predictions for individual products. The classification techniques are applied to predict the shopping behavior of the customers in online mode. Machine Learning techniques are applied and the most appropriate technique is identified where our research focuses on studying the behavior of the consumers through the dataset of customer purchase and return history. The necessity of such kinds of research and measures to control and prevent the returns arises in recent days.

3. PROBLEM STATEMENT

The product return option increases the operational and financial losses by occupying more staff, their work times, and extra transportation load for the logistics department. This type of return option is the grant provided by the e-commerce firms for the customers to carry a risk-free purchase. Therefore this kind of return option becomes unavoidable and on the other hand, their only option is to take necessary measures to decrease the returns through prediction models and understand the reasons related to their returns. The reasons for returns can be read through several sources like customer feedback, customer behavior, and vendor return rates. The metrics obtained from these sources can be used to find the actual cause of return. After finding the cause of returns these data can be stored in machines and the machines can be trained using machine learning models for making predictions in the future. The data extracted through several sources will be in different formats and all the values are converted into numeric formats for calculating the prediction score. The resulting output score value is used to make the final decision on return prediction. Understanding the returns is more beneficial for e-commerce firms to streamline their process. The firms can also act accordingly to the situation and cause claiming returns and involve in measures to decrease the losses. The e-commerce sites have coded their return policies and consumers are dealt with according to them for making returns and claiming refunds. The reason associated with the return is an important asset that can be used in framing prediction models. Once the predictions are done and the cause of returns is found then necessary precautions can be carried out to decrease the return count. The necessary measures to combat the rise in returns are to provide the exact measurement information on the product, a very transparent return policy, increase the price of the product to cover the cost of the return mechanism, check with logistics to speed up the delivery process to avoid the delay services in delivery, a new scheme to resell the returned products, lengthen the return process and policy so that the customer decides to keep the product rather than returning.

4. PROPOSED WORK

In this paper, we have developed a prototype to predict the return of any product in advance. The flow of the entire process is summarized below. We first start with data cleaning and data preprocessing. Data transformation is carried out followed by normalization and NLP (Natural Language Processing) techniques are also implemented in the dataset. The recommendation, polarity, and rating scores are generated. Similarly, the customers' return behaviors are also studied in detail and the score value is generated according to their return behaviors. Also, the vendor's count on the return of products is calculated and finally, a mathematical model is framed and the trained dataset is generated. Using this model whenever a new order is placed the automated prediction prototype calculates and generates the return risk rate of a product so the seller can make decisions accordingly. This system needs to be checked with real-time live sessions on order placement and returns for an e-commerce firm to prove the accuracy of the model.

5. EMPIRICAL SETTING

3.1 ERPSO Model

The **E-**commerce-based **R**eturn **P**rediction using **S**entiment analysis and **O**ptimization techniques model (ERPSO) is the designed prototype to predict product returns. An e-commerce company provides a product return option for its customers. Most customers check for these kinds of options before placing an order. Even though this is a provision granted for the customers from the e-commerce firm they also sometimes make the route to major losses. If the count on the number of returns increases it affects the revenue of the business and leads to unnecessary functional and operational costs. The vendors are mostly affected by these returns because reselling the returned product is noted to be a hectic task. The reason for returning can be due to any kind of issue. The return cases and reasons for returns were studied and examined in detail for e-commerce firms. The key facts and highlighted issues for return depend either on the customer or the seller. A detailed analysis was made of the issues found and the focus was done to check whether those issues were earlier predictable. The issues in return were related to two cases either from the customer side or the seller side. The two cases are discussed below,

Case I: The study was made on the customer return reasons as stated in several e-commerce sites. Accordingly, the major reasons for return from the customer side were noted to be the order placed by accident, the better price being available on other sites, the product is not useful for the intended purpose, and they don't want the product anymore, size issues either too large or too small, the quality issues on the product, the color issues.

Case II: The study was focused on the seller's sides issues like the product being delivered too late, damaged product received, poor packing quality, missing parts or missing accessories, dispatching different product other than ordered, defective product delivered, defect emerged after product usage, the product is different from website description, missing paperwork, warranty or manual, recycled and old item was sent.

The e-commerce firm should focus on the reason for the issue and develop prediction systems to predict the returns in advance so that they can be reduced to some extent using some mechanisms also the weightage score needs to be calculated to find the maximum contributor playing part in the return process to be either the vendor, the logistics, or the customers.

3.2 The description of the data

For our research, we use the dataset obtained from the URL [12]. The final sample operational dataset after preprocessing, cleaning, and dropping unwanted values is shown in Table 1. It has five columns representing the id of the product in alphanumeric value which is unique for each product type, the recommendation state of the product represented as a Boolean value, the reviews rating represented as integers ranging from 1 to 5, the reviews title describing a summary on reviews and finally the review text which has the detailed description about the feedback of the product.

Table 1

A Sample Operational Dataset

	id	reviews.doRecommend	reviews.rating	reviews.text	reviews.title
0	AVqVGZNvQMlgsOJE6eUY	False	3	I thought it would be as big as small paper bu...	Too small
1	AVqVGZNvQMlgsOJE6eUY	True	5	This kindle is light and easy to use especiall...	Great light reader. Easy to use at the beach
2	AVqVGZNvQMlgsOJE6eUY	True	4	Didnt know how much i'd use a kindle so went f...	Great for the price
3	AVqVGZNvQMlgsOJE6eUY	True	5	I am 100 happy with my purchase. I caught it o...	A Great Buy
4	AVqVGZNvQMlgsOJE6eUY	True	5	Solid entry level Kindle. Great for kids. Gift...	Solid entry-level Kindle. Great for kids
...	...	...	...	...	...
4995	AVqkIdZiv8e3D1O-leaJ	True	5	This is a great tablet for the price. Amazon i...	Good product
4996	AVqkIdZiv8e3D1O-leaJ	True	5	This tablet is the perfect size and so easy to...	Great Tablet
4997	AVqkIdZiv8e3D1O-leaJ	True	4	Purchased this for my son. Has room to upgrade...	Great for kids or smaller needs
4998	AVqkIdZiv8e3D1O-leaJ	True	5	I had some thoughts about getting this for a 5...	Very sturdy for a 5 year old
4999	AVqkIdZiv8e3D1O-leaJ	True	5	this is a steal, have 8 gb model as well.This ...	great little tablet

3.3 Variables declaration

The final output value is defined as Δ_{RPS} Δ_{RPS}, where RPS stands for Return Prediction Score and Δ_{RPS} Δ_{RPS} is the sum of the total return prediction scores obtained from each entity. The entities used to calculate the prediction score are obtained through the product feedback score denoted as $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$, the vendors return rate score of individual product denoted as $\gamma_{vendor\_return\_rate}$ $\gamma_{vendor\_return\_rate}$ and the scores of customers return behavior stated as $\beta_{customer\_return\_behaviour\_score}$ $\beta_{customer\_return\_behaviour\_score}$. The final value is obtained by summing and taking the average value of the above three scores.

Table 2
Definition of Variables

CATEGORY	VARIABLE NAME	DEFINITION
FEEDBACK EFFECT	$\alpha_{Product\_feedback\_score}$	AVERAGE OF RECOMMENDATION, RATINGS & POLARITY
CUSTOMER BEHAVIOR EFFECT	$\beta_{customer\_return\_behaviour\_score}$	DIVIDEND OF THE NUMBER OF PRODUCTS RETURNED TO THE NUMBER OF PRODUCTS PURCHASED BY THE CUSTOMER
VENDOR RETURN EFFECT	$\gamma_{vendor\_return\_rate}$	DIVIDEND OF THE NUMBER OF PRODUCTS RETURNED TO VENDOR TO THE NUMBER OF PRODUCTS SOLD BY THE VENDOR FOR A PARTICULAR PRODUCT
FINAL RESPONSE VARIABLE	Δ_{RPS}	AVERAGE OF THE FEEDBACK SCORE, CUSTOMER RETURN BEHAVIOR, VENDOR RETURN RATE

6. ERPSO ILLUSTRATION

The **E**-commerce based **R**eturn **P**rediction using **S**entiment analysis and **O**ptimization techniques (ERPSO) will work as an intelligent software agent model to predict the product returns in advance and alert the e-commerce firm for taking preventive actions through checking the maximum contributor responsible for the cause of return. The product return can be predicted daily and measures can be implemented to stop the unnecessary losses.

ERPSO MODEL

Initialize

Customer X places an order for the product Y from the vendor Z

For each order of product Y

{

Read the Input variables for $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$, $\beta_{customer\_return\_behaviour\_score}$ $\beta_{customer\_return\_behaviour\_score}$, $\gamma_{vendo\ r\_return\_rate}$

Where $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$ $= (\sigma + \tau + \upsilon)/3$ $(\sigma + \tau + \upsilon)/3$;

σ = Score received through Recommendations on products

τ = Score obtained for product based on Ratings

υ = Score generated based on textual feedback

$\beta_{customer\_return\_behaviour\_score}$ $\beta_{customer\_return\_behaviour\_score}$ = Total number of times the product returned back to the total number of products purchased

$\gamma_{vendo\ r\_return\_rate}$ $\gamma_{vendo\ r\_return\_rate}$ = Count on the number of times product returned to origin to the total number of products sold

$\Delta_{RPS} = (\alpha_{Product\_feedback\_score} + \beta_{customer\_return\_behaviour\_score} + \gamma_{vendo\ r\_return\_rate})/3$

Store the resulting output of Δ_{RPS}

}

Check the type of Δ_{RPS}

{

Type I: Return chances are very less where $\Delta_{RPS} = 0$.

Type II: Return chances are very less where $\Delta_{RPS} \leq 0.25$ & > 0.

Type III: Return chances are moderate where $\Delta_{RPS} \leq 0.5$ & > 0.25.

Type IV: Return chances are high where $\Delta_{RPS} \leq 0.75$ & > 0.5.

Type V: Return chances are very high where $\Delta_{RPS} \leq 1$ & > 0.75.

}

End

The flow diagram of the ERPSO model is shown in the below diagram. Fig.1 represents the input parameter, the processing of variables, and finally the output obtained. The input is further subdivided as consumer, product and vendor segments. Then the model undergoes two processing phases I and II, where the count of each product delivered and returned is calculated. Then the final obtained score is used to calculate the prediction of product returns.

In Fig.2 the block diagram represents the in-depth process and techniques for generating the prediction scores. The dataset is fed as input to the system and the final output is the prediction score denoted as Δ_{RPS} Δ_{RPS}. The illustration of the ERPSO modules is discussed below. We segregate the process into three modules as Module 1, Module 2, and Module 3. The first module explains the mechanism of equating the product feedback score, the second module illustrates the process of calculating the score of customers' return behavior, and the third module illustrates the process of calculating the vendor's return rate.

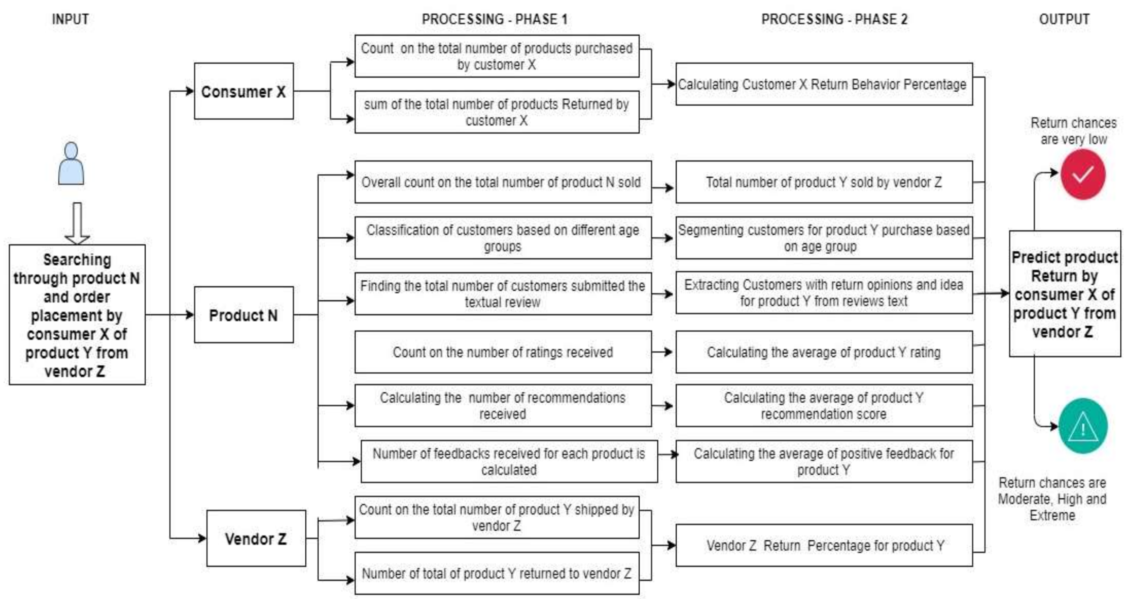

Fig.1 Overall Flow diagram

Module 1: The first module is designed to calculate the feedback score of a particular product based upon the product sales along with their reviews ratings, review text, review title, and recommendations obtained from the consumers which are represented as $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$.

Module 2: The second module is intended to find the consumers buying and return behavior based on their purchases and return history made with the particular e-commerce site and is denoted as $\beta_{customer\_return\_behaviour\_score}$ $\beta_{customer\_return\_behaviour\_score}$.

Module 3: The third module aims to calculate the return rate faced by the vendors based upon the total number of products dispatched from the vendor and the total number of products returned to the vendor which is stated as $\gamma_{vendor\_return\_rate}$ $\gamma_{vendor\_return\_rate}$. Comprising the values obtained from the above stated three modules, we find the **R**eturn **P**rediction **S**core (RPS) for any Customer X represented with their Customer Identification number ($X_{C\_ID}X_{C\_ID}$) while placing any particular product Y with a Product Identification number ($Y_{P\_ID}Y_{P\_ID}$) and dispatched from any particular Vendor Z, represented with the Vendor Identity ($Z_{V\_ID}Z_{V\_ID}$).

The resulting value of RPS is represented as Δ_{RPS} Δ_{RPS}, which is calculated by summing the values obtained from each module, and finally, the prediction conclusions are made accordingly.

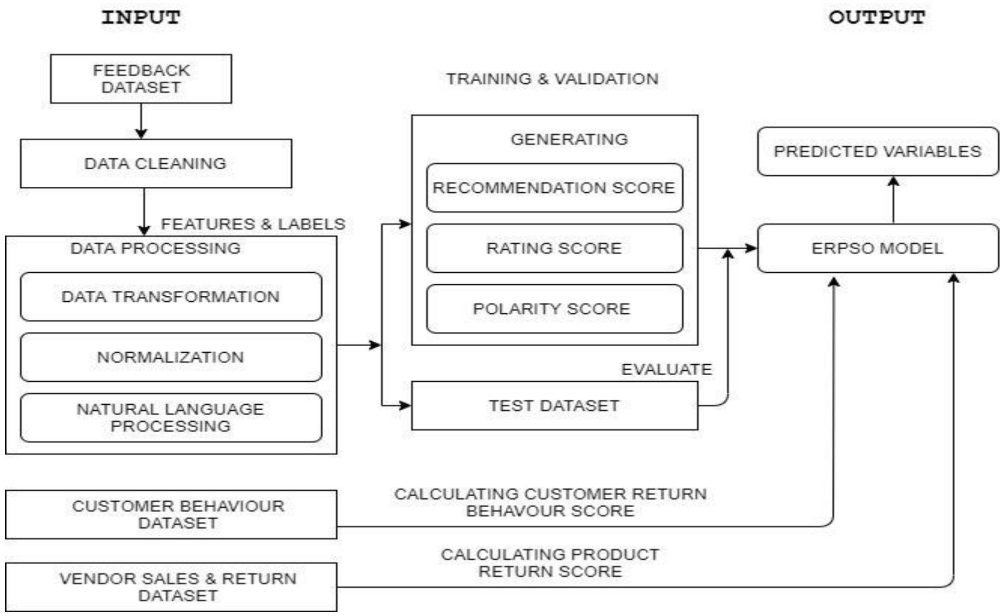

Fig.2 Block diagram of generating prediction output

Equation (4.1) represents the formula for calculating the value of Δ_{RPS} Δ_{RPS},

$$\Delta_{RPS} = (\alpha_{Product\_feedback\_score} + \beta_{customer\_return\_behaviour\_score} + \gamma_{vendor\_return\_rate}) / 3$$

$$\Delta_{RPS} = (\alpha_{Product\_feedback\_score} + \beta_{customer\_return\_behaviour\_score} + \gamma_{vendor\_return\_rate}) / 3 \quad (4.1)$$

4.1 Equating product feedback score (Module 1)

The product feedback and the returns are interrelated to each other. Whenever a product is returned by the customer, they will state negative feedback on the product on the e-commerce site. The feedback is collected in different forms and stored in the dataset.

The machine learning algorithms process only on the numeric data, so all the textual data on the dataset have to undergo several transformations to attain the numeric shape. The product feedback score $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$ is calculated using the below entities as shown in Table 3.

4.1.1 *Recommendation: The recommendation data is either true or false. The Boolean value is converted into a numeric value. We consider the value as 1 for the true cases and value 0 for the false cases. So the final recommendation score, if the product is recommended, will be 1 or otherwise 0 denoted as σ.*

4.1.2 Review Rating: The rating for the reviews is obtained as numeric and the range of value is from 1 to 5. We convert the values in the range between 0 to 1 where 1 is taken as 0, 2 as 0.25, 3 as 0.50, 4 as 0.75, and 5 as 1. The reviews rating value is denoted by τ

4.1.3 Reviews Text: Here, the sentence gives even more deep insights when compared to the information obtained from other entities. NLP technique is applied to do opinion mining and the polarity score is recorded as a new column represented as υ.

Table 3
A Sample Dataset with recommendation, Rating and Polarity score

	id	reviews.doRecommend	reviews.rating	reviews.text	reviews.title	recommend	rating	polarity
0	AVqVGZNvQMlgsOJE6eUY	False	3	I thought it would be as big as small paper bu...	Too small	0	0.50	-0.108974
1	AVqVGZNvQMlgsOJE6eUY	True	5	This kindle is light and easy to use especiall...	Great light reader. Easy to use at the beach	1	1.00	0.277778
2	AVqVGZNvQMlgsOJE6eUY	True	4	Didnt know how much i'd use a kindle so went f...	Great for the price	1	0.75	0.165625
3	AVqVGZNvQMlgsOJE6eUY	True	5	I am 100 happy with my purchase. I caught it o...	A Great Buy	1	1.00	0.240497
4	AVqVGZNvQMlgsOJE6eUY	True	5	Solid entry level Kindle. Great for kids. Gift...	Solid entry-level Kindle. Great for kids	1	1.00	0.468750
...	...	...	...	...	...	...	...	...
4995	AVqkIdZiv8e3D1O-leaJ	True	5	This is a great tablet for the price. Amazon i...	Good product	1	1.00	0.750000
4996	AVqkIdZiv8e3D1O-leaJ	True	5	This tablet is the perfect size and so easy to...	Great Tablet	1	1.00	0.577083
4997	AVqkIdZiv8e3D1O-leaJ	True	4	Purchased this for my son. Has room to upgrade...	Great for kids or smaller needs	1	0.75	0.500000
4998	AVqkIdZiv8e3D1O-leaJ	True	5	I had some thoughts about getting this for a 5...	Very sturdy for a 5 year old	1	1.00	0.016667
4999	AVqkIdZiv8e3D1O-leaJ	True	5	this is a steal, have 8 gb model as well.This ...	great little tablet	1	1.00	0.500000

5000 rows × 8 columns

Fig.3 represents the graphical representation of the recommendations made by consumers for each product represented by individual id is shown below. The product recommendation score is used to analyze the recommendation score of each product given by the consumers.

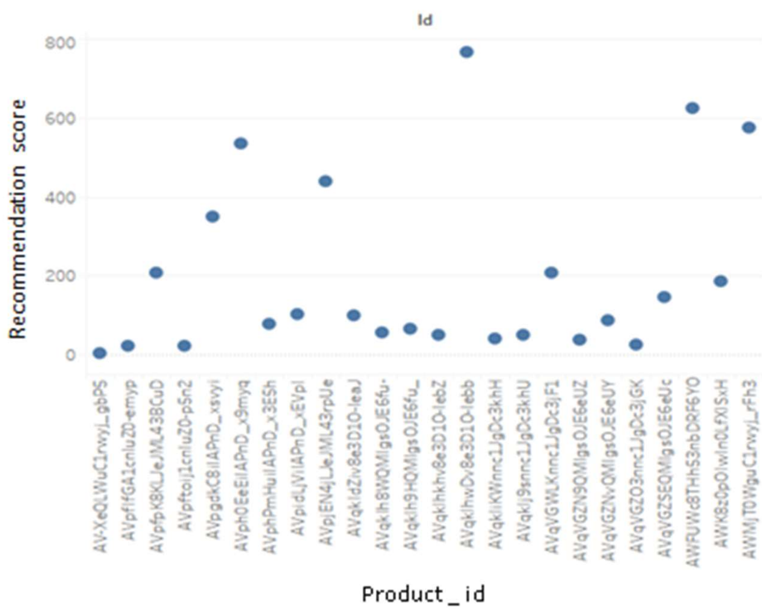

Fig.3 Product Recommendation Graph

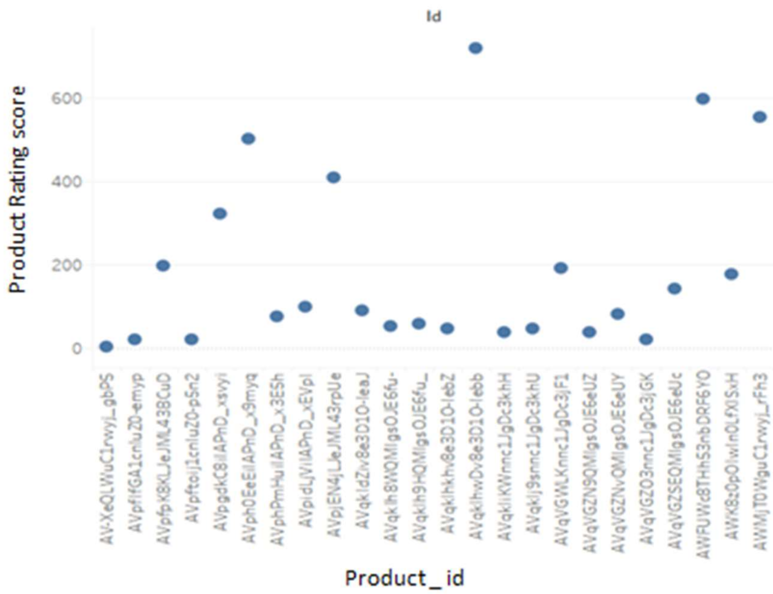

Fig.4 Product Rating Graph

The graph in Fig.4 represents the overall ratings received for individual products from the consumers who purchased the products. The ratings can be used to analyze the quality and other parameters of the product. The graphical representation in Fig.5 represents the polarity score obtained as the textual feedback from the consumers. Natural Language Processing Technique is used to mine the opinion from the texts. Then they are classified as positive, negative or neutral feedback. The feedback polarity score is calculated to find the opinion of customers upon a product.

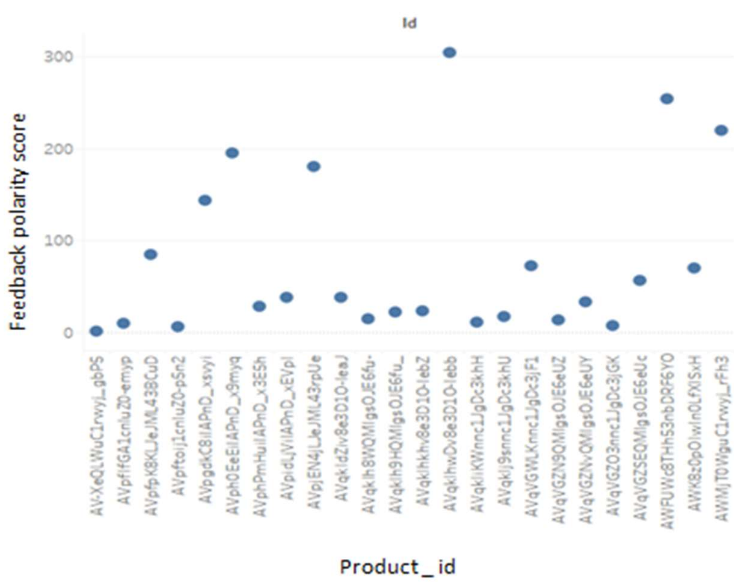

Fig.5 Polarity Analysis Graph

Finally, the product feedback score $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$ is calculated by adding all the values and taking the average of the summation obtained from the above-mentioned entities as shown in (4.1.1). The maximum score for $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$ will be 1 and the minimum score will be 0. So based upon the score value we can conclude that the feedback of the products fall under which range.

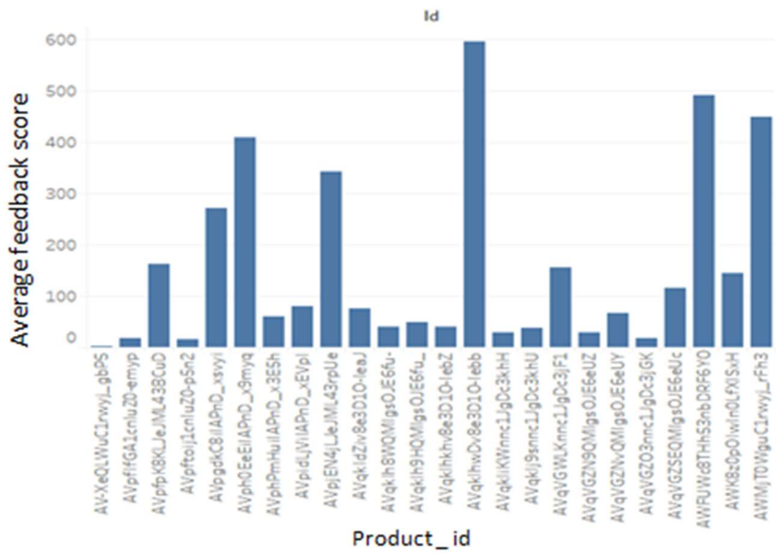

Fig.6 Average value of $\alpha\alpha$

This $\alpha_{Product\_feedback\_score}$ $\alpha_{Product\_feedback\_score}$ is found and lies as one of the parameters for Δ_{RPS}. Fig.6 represents the graph showing the average score of the three entities (i) Recommendation (σ) (ii) Rating (τ) and (iii) Polarity (υ).

$$\alpha_{Product\_feedback\_score} = \frac{(\sigma+\tau+\upsilon)}{3}$$

(4.1.1)

4.2 Calculating Customer Return Behavior score (Module 2)

Every customer is different. Returns are always considered an essential part to run a successful e-commerce business, but issues arise when some consumers use the opportunity and make return a serial habit. The behavior of a customer in making online purchases and their product return habit creates a greater impact affecting the revenue of the e-commerce firm. This data about returns need to be studied and recorded which can further be used in predicting the return rates in the future. The study can be focused to find,

- Count on the number of purchases made by the customer online
- Count on the number of returns made by the customer
- Customer demand on product exchange or cash return
- Feedbacks and ratings given by the product
- Customer product recommendation

All the above-studied data can be used to make a detailed note of the behavior of customers involved in online purchases. The concentration is made on the return aspect of the consumer. The return rate of the customer is calculated using the below two entities over a time period (t),

- The total number of products purchased online over a particular period $(TPP)_t$.
- The total number of products returned over a particular period $(TPR)_t$.

Here we consider the date as the period to calculate the return behavior of the consumers.
Table 4 represents a sample dataset created to illustrate the mechanism of calculating the behavior of the consumer in making the return process. Table 4 shows the sample data for a single consumer. The RTO column value is denoted as 1 if the customer makes a return else stated as 0. Likewise, the total number of products purchased by each customer is extracted from the dataset and summed together, which denotes the total number of products purchased, and also the total number of products returned is also calculated by summing the total number of returns.

Table 4
A Sample Dataset with customer purchase and return details

DATE	CUSTOMER_ID	PRODUCT_ID	RTO
12/05/2020	AF39WE23ST	Avqvgznvqmlgsoje6euy	1
14/05/2020	AF39WE23ST	Awmjt0wguc1rwyj_rfh3	1
22/05/2020	AF39WE23ST	Avqvgzseqmlgsoje6euc	0

The resultant value of $\beta_{customer\_return\_behaviour\_score}\beta_{customer\_return\_behaviour\_score}$ is calculated as follows as shown in equation (4.2.1),

$$\beta_{customer\_return\_behaviour\_score} = \frac{(TPR)_t}{(TPP)_t}\beta_{customer\_return\_behaviour\_score} = \frac{(TPR)_t}{(TPP)_t} \quad (4.2.1)$$

The output value is stored as the return behavior of the consumer and is the second entity in calculating the score of $\Delta_{RPS}\Delta_{RPS}$.

4.3 Calculating Vendor Return Rate (Module 3)

Vendors are the manufacturers or the source of suppliers for the e-commerce sites. They are tied up with e-commerce businesses to promote their sales. A vendor can make business with multiple e-commerce sites. The success of the vendor depends on the demand and quality of the product he supplies. The returned products always don't demand a refund, in some cases, they ask for replacement also. Even though the exchange costs unnecessary shipping charges, the return with demand on for exchange is omitted and only the refund issue is considered in this research work.

Table 5
A sample dataset with vendor dispatch and return details

DATE	VENDOR _ID	PRODUCT_ID	SALES	RTO
12/05/2020	VA59SD38YK	Avqvgznvqmlgsoje6euy	56	12
12/05/2020	VA59SD38YK	Awmjt0wguc1rwyj_rfh3	33	0
14/05/2020	VA59SD38YK	Avqvgzseqmlgsoje6euc	19	9

A sample dataset representing the product sales and return is shown in Table 5. The study can be focused in finding,

- Count on the number of products returned to the vendor
- Count of product returned for exchange
- Count of products returned demanding cash return
- Analyze the feedback statements and count the review ratings
- The condition of returned product
- Maximum limit of the return product
- investigation on the cause of product returns
- Improvements made based on the suggestion of analysis

All the above-studied data can be used to make a thorough note of the product sold through e-commerce. The concentration is made on the feedback on the product returns. The return rate of the product sold from a particular vendor is calculated using the below two entities over a particular period (t),

- The total number of products sold online over a particular period $(TPS)_t$.
- The total number of products returned to origin over a particular period $(TPO)_t$.

The product return rate of the vendor $\gamma_{vendor\_return\_rate}$ $\gamma_{vendor\_return\_rate}$ can be calculated concerning individual product_id as shown in equation (4.3.1),

$$\gamma_{vendor\_return\_rate} = \frac{(TPS)_t}{(TPO)_t} \gamma_{vendor\_return\_rate} = \frac{(TPS)_t}{(TPO)_t} \qquad (4.3.1)$$

4.4 Resulting Value of Δ_{RPS} Δ_{RPS}:

The resulting values from equations (2), (3), and (4) are substituted in equation (1) and the final Δ_{RPS} Δ_{RPS} value is obtained. The final score is categorized and checked with the conditional statements among the five types and the conclusions on return prediction are derived.

Checking the results with Conditional Hypothesis:

Input: Δ_{RPS}

Conditional check: If Δ_{RPS} = 0, then the chances of return are very less

Else If $\Delta_{RPS} \leq 0.25 \ \& > 0$, then the chances of return are less

Else If $\Delta_{RPS} \leq 0.5 \ \& > 0.25$, then the chances of return are moderate

Else If $\Delta_{RPS} \leq 0.75 \ \& > 0.5$, then the chances of return are high

Else If $\Delta_{RPS} \leq 1 \ \& > 0.75$, then the chances of return are very high

End If

Output: Δ_{RPS} Δ_{RPS} of the product is the Resultant score

of the product is the Resultant score

5 CONCLUSION AND FUTURE WORK

The e-commerce firm also sometimes lays as a reason for the return issues, where the consumer after receiving the product finds any comparatively lower prices for the same products in any other sites and it can even be due to discounts or offers they announce not being discussed in this research work. Machine learning techniques can be included when the system is implemented in real-time scenarios which will be considered in our future work. The prolonged research can be extended by implementing the system in a real-time scenario and checking with suitable machine-learning techniques for enhanced accuracy of the overall system.

6 REFERENCES

[1] H. Cui, S. Rajagopalan, and A. R. Ward, (2020), Predicting product return volume using machine learning methods, *Eur. J. Oper. Res.*, vol. 281, no. 3, pp. 612–627, 2020, doi: 10.1016/j.ejor.2019.05.046.

[2] Dimitrios Vlachos, Rommert Dekker, (2003), Return handling options and order quantities for single-period product , European Journal of Operational Research, Volume 151, Issue 1,2003, Pages 38-52, ISSN 0377-2217, https://doi.org/10.1016/S0377-2217(02)00596-9.

[3] Dheeraj Malhotra, Omprakash Rishi,(2021), An intelligent approach to the design of E-Commerce metasearch and ranking system using next-generation big data analytics", Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 2, Pages 183-194, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2018.02.015.

[4] Esmeli.R,Bader-EI-den,M.& Abdullahi H, (2020),Towards early intention prediction in online session based retailing systems.Electron Markets.https://doi.org/10.1007/s12525-020-00448-x

[5] Return to Sender: Re | Barclaycardtailers face a 'Phantom Economy' of £7bn each year as shopper returns continue to rise (home.barclaycard)

[6] Tolga Cenesizoglu, Allan Timmermann,(2012),Do return prediction models add economic value?,Journal of Banking & Finance,Volume 36, Issue 11,2012,Pages 2974-2987,ISSN 0378-4266, https://doi.org/10.1016/j.jbankfin.2012.06.008.
(https://www.sciencedirect.com/science/article/pii/S0378426612001604)

[7] Huibing Zhang, Junchao Dong,(2020), Prediction of repeat Customers on E-Commerce Platform Based Blockchain, Wireless Communications and Mobile Computing, vol.2020,Article ID8841437, 15 pages , 2020.https://doi.org/10.1155/2020/8841437

[8] Abdullah Al Imran, Md Nur Amin, (2020), Predicting the Return of Orders in the E-Tail Industry Accompanying with Model Interpretation", Procedia Computer Science, Volume 176, Pages 1170-1179, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.09.113

[9] Kranz, Johann & Urbanke, Patrick & Kolbe, Lutz. (2015). Predicting Product Returns in E-Commerce: The Contribution of Mahalanobis Feature Extraction.

[10] Baati, Karim, and Mouad Mohsil, (2020), Real-Time Prediction of Online Shoppers' Purchasing Intention Using Random Forest. Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part I vol. 583 43–51. 6 , doi: 10.1007/978-3-030-49161-1_4

[11] E-Commerce Product Return Rate – Statistics and Trends [Infographic] (2020), E-commerce Product Return Rate – Statistics and Trends [Infographic] (invespcro.com)

[12]https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products?select=Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products_May19.csv

[13] Stock, J. R., & Mulki, J. P. (2009). Product returns processing: An examination of practices of manufacturers, wholesalers/distributors, and retailers. Journal of Business Logistics, 30(1), 33–62.

[14] Arul Natarajan ,Panthagani vijayababu , M.Arsha ,Sasibhushana RaoPappu , Vidya Rajasekaran. (2023). Early Disease Diagnosis Using Multivariate Linear Regression. *Journal of Pharmaceutical Negative Results*, 1383–1390. https://doi.org/10.47750/pnr.2023.14.S02.168

[15] Hichem Rahab, Abdelhafid Zitouni, Mahieddine Djoudi,(2021),SANA: Sentiment analysis on newspapers comments in Algeria ,Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 7, Pages 899-907,ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2019.04.012.

[16] Grifs, S. E., Rao, S., Goldsby, T. J., & Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. Journal of Operations Management, 30(4), 282–294.

[17] Basavaraj N. Hiremath, Malini M. Patil, (2020), Enhancing Optimized Personalized Therapy in Clinical Decision Support System using Natural Language Processing ,Journal of King Saud University - Computer and Information Sciences,ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2020.03.006.

[18] S. K. Shivakumar and P. V. Suresh, (2014),Maximizing Knowledge Management return in e-commerce," 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 545-550, doi: 10.1109/IndiaCom.2014.6828018.

[19] Rajasekaran V., Priyadarshini R. (2021), An E-Commerce Prototype for Predicting the Product Return Phenomenon Using Optimization and Regression Techniques. In: Singh M., Tyagi V., Gupta P.K., Flusser J., Ören T., Sonawane V.R. (eds) Advances in Computing and Data Sciences. ICACDS 2021. Communications in Computer and Information Science, vol 1441. Springer, Cham. https://doi.org/10.1007/978-3-030-88244-0_22

[20] Priyadarshini, R. and L. Tamilselvan. (2014), Document clustering based on keyword frequency and concept matching technique in Hadoop. International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014 1367, ISSN 2229-5518

[21] H. Yang, J. Wang, M. He, and B. Kuang,(2010),Research of B2C E-Commerce Return Strategies Based on Return Price," 2010 International Conference on Management and Service Science, Wuhan, China, pp. 1-4, doi: 10.1109/ICMSS.2010.5576679.

[22] E-commerce Product Return Rate - Statistics and Trends 2020- Infographic (magnetoitsolutions.com)