Predicting Customer's Choice Using Big Data Analytics

Ganesh Naikwade, Prof Anil Ambore
Computer science
REVA UNIVERSITY, Bangalore Karnataka

Abstract- Big data is large volume data gather from different sources like online transaction, clicking on internet sites, social media, search history, online payment etc. When this large volume data filtered and visualized can be very useful for firms to understand trend of products and improve business strategies. Determining the choice of customers is very difficult task but it is very important for service oriented businesses. Research on determining consumer choice is continues process and choice of customer varies as per culture and geographic location. Many methods for this purpose have been developed like data mining, machine learning and statistics. Many e-commerce companies use Big Data Analytics for predicting the product customer will buy on getting his profile form social media site or analyzing his tweets. After buying any product which product he can buy this can be predicted. In this paper data is extracted from social media site and visualized to understand the trend.
On knowing the trend firms focuses more on product which are in trend. It encourages them to focus more on product which can be bought more and increase their sale. Data mining has been successfully applied by many ecommerce firms.

Keywords- Big data analytics, e-commerce, trend, Data mining, customer, sentiment.

I. INTRODUCTION

Now a day's internet is growing very fast from computer to mobile devices, Data is growing also very fast. Internet has made life very easy. It is very easy to access data on internet. As internet connection are increasing, According to Internet Live Stats which collects data from different sources 40 % of world population has internet connection and as wealth of middle class are growing, they want to do everything on same time, and for them ecommerce sites are the best option or buying online. This gives firms to data which they analyze as per their need. People want to do more activities at the same time, to improve their quality of life. Style of customer to buy the product has changed compare to decades ago. For instance if somebody wanted to buy a book he had to go into more book stores to know that book exists and if that exists he had to compare the price in order to decide the price of book. Nowadays customer can very easily choose their product without going to store and easily find out pros and cons of product and compare the price and buy the product by just pressing the button as a result saving time and energy. As the process of buying and selling the product is changed competition among firms are increased.
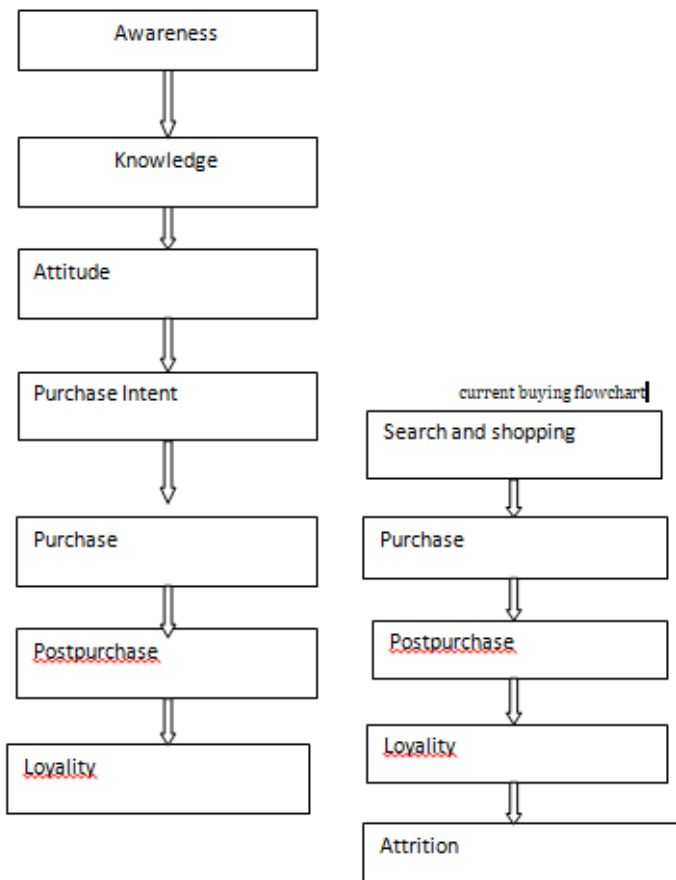
Customer satisfaction is main goal of firms. There are many studies and researches are conducted and still continue to understand the choice of customer. Number of studies indicated that customer behavior can be predicted. For instance positive remarks by customer on product encourage firms to invest in that product and positive review encourage customers to buy that product. [1] Preference of product not changed but steps of buying the product changed even paying the bill at physical store is changed. People talk about product on social media sites and buy accordingly. Advertisement is given by companies over sites which visited by user in large number. Data from all these sites are gathering and analyzed by firms and used to increase their growth. Data mining has become a mature market after crush of dot.com bubble. Customer analytics is very important tool to get information of customer behavior. Unfortunately for many organizations the access to data and necessary analytics tools are limited. Customer analytics allow companies to understand trend and behavior as a result it improves their operation. It is very expensive to make new customer than to retain existing ones.

Big data are proving to be valuable resource which is acquired from retail transaction data from credit card purchases and from digital footprints, called cookies. As the cost of big data is decreasing, this technology is more accessible to small business for predicting customer behavior. The study of customers choice prediction was started about 300 years ago by Nicholas Bernoulli. In this paper data extracted from social media sites is filtered and visualized and trend is identified.

II. LITERATURE REVIEW

Earlier customers buy the product on visiting different store and suggestion of friends and family. They visit many stores and compare price and quality of product that consume time and energy after satisfaction they buy the product. Nowadays customers have more option they can buy online as well as by visiting physical market. They don't have to depend only on suggestion of friend and family. Nowadays people read online review chat online on product. Search about price and quality of product and then take decision. Nowadays customers use more mobile for shopping they can track delivery of their product. Online customers have no need to visit different store to compare price and check quality of product. Nowadays customers are more aware having more knowledge about the product than earlier.

Traditional buying flowchart

Awareness

Knowledge

Attitude

Purchase Intent

Purchase

Postpurchase

Loyality

current buying flowchart

Search and shopping

Purchase

Postpurchase

Loyality

Attrition

Data mining has disadvantage also, [2] The Junzo Watada and Kozo in their paper "A Data mining approach to consumer behavior" improved defects in conventional methods by applying fuzzy clustering, principal component analysis and discriminate analysis.

[4] Tan, Blake, Saleh in their paper ''Social Network Sourced Big Data Analytics'' present various benefits of social media like identify and address customers point of view concerned to product, understand new requirement of customer, address competition among firms and obtain customer feedback for decision making. They presented stages of social media data analysis that are data collection, sentiment analysis of product in the market, visualization.

[5] B. Yildiz and B. Ergenc in their paper '' Comparison of Two Association Rule Mining Algorithms without Candidate Generation'' Association rule mining techniques find interesting correlations among sets of items in database. Aprori algorithm categorized under association rule mining, it has some drawback candidate generation phase very lengthy and time consuming. FP Growth and Matrix Apriori are two algorithm categorized under Association rule mining eliminate the problem of candidate generation.

Nowadays businesses are more data centric. Big data is used not only in IT but in every field. The first step is collecting data and storing it. Large data ware house are made for storing data from different sources. Data mining is done for extracting desired data from large data set.

Prasanna Mohan in his paper "Factor influencing customers brand preference" highlighted 6 factors influencing customers purchase decision namely Product reliability, monetary factor, trend, frequency of non-price promotions offered, trustworthiness and customer feeling towards brand. Organization must consider these point before increasing their product stock[3].

[6] S. Vijaylaxmi, V, Mohan, S. Suresh Raju in their paper "Mining of user's access behavior for frequent sequential pattern from web logs" explains in sequential pattern mining comes under association rule mining.

[7]Sharma in their paper "Changing consumer behavior-A challenge for sustainable business growth" discussed the factors effecting consumer behaviors.

Now products are made as per consumer choice than as per vendor's choice. Research also suggested that online promotion and advertising improves traffic as well as sales.

[10] Roselin and Hanupriya in their paper "Customer Behavior Analysis for Credit Card Proposers Based on Data Mining Techniques" explains strategy how to respond queries of every type of customers.

Several studies on consumer purchasing behavior had been used in real problem. Data mining techniques were expected to be more effective tool for analyzing consumer behavior. Therefore, it is important to select appropriate method to mine database. The Junzo Watada and Kozo Yamashiro in their paper "A Data mining approach to consumer behavior" tried to improve data mining analysis by applying several methods including fuzzy clustering, principal component analysis and discriminate analysis. Many defects included in the conventional methods are improved in this paper.Social Media Analytics enables organization to utilize social media for business growth. Various benefits of social media analytics are identify and address customer concerns to retain customer loyalty to products, understand new requirements and prospective customers, address competition and obtain customer feedback for proactive decision making. Various stages of social media analytics are data collection, measurement – market sentiment analysis of products, data analysis and visualization for better insights, innovative and Strategic business decisions.

III. METHODOLOGY

The paper for predicting customers choice is based on trend, by identifying trend in particular category of product, choice of customer which may be is identified. Data gather from social

media site is extracted and filtered using python language and that data is visualized using python language. Thus the first step is collecting data that is secondary research. Secondary data were collected and relevant literatures were referred for mining data from data source. On twitter application is made and access token and security key are used for authentication in program.

It has three necessary steps
- Creating twitter application
- Extracting data
- Visualizing data

Phase 1: Creating twitter application
For extracting data from twitter we made twitter application using twitter login id and password. On creating twitter application we get twitter access token and security token for authentication from twitter database and can extract twitter data.

Phase 2: Extracting twitter data or data mining
Data is extracted from twitter in json format. First authentication is done with twitter and then data is extracted. Data is filtered according to our choice and filtered data is stored in storage.

Phase 3: Visualizing data
Extracted data visualized using python. Visualized data is easy to understand.

IV. ALGORITHM
The main objective of this method is to extracting data from twitter and visualizing that data. Twitter data are complete public and pullable. This is the reason we chooesen twitter for data analytics. Twitter's API allows to do complex queries like pulling every tweet concerned to particular topic. We can get tweet from certain location which is known as spatial data.
To use twitter's API, we have to create a developer account on the Twitter app. Once project created get from there consumer key, consumer secret, Access token and Access token secret.

4.1 Data extraction algorithm
```
Class listener (StreamListener):
Def on_data(self, data):
Print (data)
Return (True)
Def on_error(self, status):
 Print  (status)
Auth = OAuthHandler("consumer key", "consumer secret")
Auth.set_access_token("access token", "access secret")
Stream = Stream (auth, listener ())
Stream.filter(track = ["ACC", "BIRLA", "AMBUJA", "INDIA", "PRISM"])
```

VI. CONCLUSION
Analysis of customers choice and preferences enables firms to reduce the risk of loss and direct them to invest in right

This algorithm gives data on cement companies.

4.2 Visualizing data algorithm
```
tweets_data_path = "storage"
tweets_data = []
tweets_file = open(tweets_data_path, "r")
for line in tweets_file:
tweet=json.loads(line)
tweets_data.append(tweet)
tweets=pd.DataFrame()
tweets["text"]=map(lambda tweet:tweet['text'],tweets_data)
tweets["cement"]=map(lambda tweet:tweet['text'],tweets_data)
fig.ax=plt.subplots()
ax.tick_params(axis='x',labelsize=10)
ax.tick_params(axis='y',labelsize=10)
ax.set_xlabel('cement', fontsize=10)
plt.show()
```
On extracting data (twitter) and visualizing data of cement using python we get trend chart which states that top ACC Cement is sold more or twittes we analyze that if we sell ACC Cement more in our firm then sell of cement will increase.

V. RESULT
Below figure evaluates the data extracted from twitter on cement as it shows ACC cement term more used in twitter. It means trend of buying cement is ACC, firms in their store keep more item of ACC Cement and Prism cement than other then they no need to invest money in other product and their sale will increase and profit as well.we can apply this method on several product like shoes, washing machine and knowing the trend on social media site or data gather from other media and know the trend which in turn increase our growth.
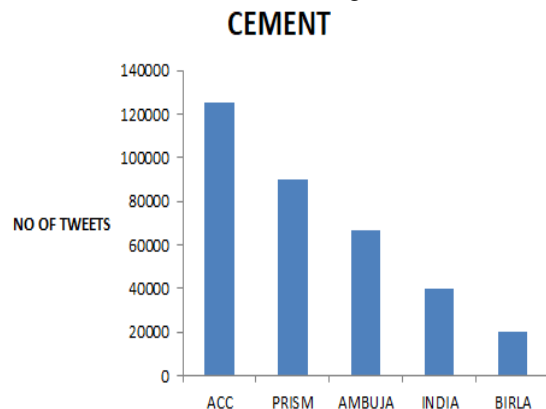

Fig.1:

direction. As a result improve their performance. As customer preferences changes and they have now more options, customer data analysis is important for customer attraction,

retention and development. This study focuses on analysis of data for determining trend and based on that predicting users choice. Data extracted from e-commerce sites and visualized using python.

VII. REFERENCES

[1]. CLEMONS, E.K. 'How Information changes consumer behavior'.

[2]. watada, j., &Yamashiro. A data mining approach to consumer behavior. In first International Conference on Innovative Computing, Information and Control-Volume I

[3]. Raj, M.P.M., Sasikumar, j., & Sriram, S.(2013). A Study on Customers Brand Preference in Suvs and Muvs: Effect of Marketing Mix Variables.

[4]. Tan, W., Blake, M. B., Saleh, I., & Dustdar, S. (2013). Social Network Sourced Big Data Analytics.

[5]. Yildiz, B., & Ergenc, B. (2010). Mining of user's access behavior for frequent sequential pattern from web logs.

[6]. Vijayalakshmi, S., Mohan, V., &Raja, S.S. (2010). Mining of user's access behavior for frequent sequential pattern from web logs.

[7]. Sharma, S., &Lal, k.(2012). Changing consumer behavior- A challenge for sustainable business growth.

[8]. Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior.