

A Web Application for Customer Segment based Black Friday Sales Forecast

Tanisha G. Patil

*Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
tanisha.gpatil15@gmail.com*

Varsha Venkatesh

*Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
venkyvarsha@gmail.com*

Vanitha R. Patil

*Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
punnu.rp@gmail.com*

Kanchan.A.Purohit

*Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
kanchanpurohit@bit-bangalore.edu.in*

Shivangi Karwa

*Computer Science and Engineering
Bangalore Institute of Technology
Bengaluru, India
karwashivangi1999@gmail.com*

Abstract—The study shows how factors internal to the sales forecasting process can introduce bias. Black Friday is upon us, and retailers everywhere stock up on inventory at dramatic discounts. The products aren't accessible to some clients due to poor arrangements, despite good sales. Managing a small staff and keeping potential customers in mind is more of a concern for shopkeepers. Despite some attempts at effective inventory management and sales forecast, the results have not been favorable. A prediction model has proven effective in managing this issue. By analyzing the previous spending patterns of clients, a precise and efficient algorithm can predict future spending. A variety of data analysis strategies, including regressors and classifiers, are employed to evaluate specificity and accuracy. Then, different algorithms are used to implement machine learning techniques. An extensive report of sales for Black Friday 2019 has been compiled. To calculate the volume, we took into account various factors, including age group, occupation, product category, city, etc. A store that wishes to predict sales could do so from this information. We have deployed our Flask-based user interface after determining which model is the best fit.

Keywords—Black Friday, Cyber Monday, Sales Forecast, Customers, Customer Segmentation, Retailers, Data Analysis, Machine Learning, Flask

I. INTRODUCTION

A shopping holiday known as Black Friday is observed on the fourth Thursday of the month following Thanksgiving Day in the United States. The Black Friday sales data set was acquired from kaggle. Statistical overview indicates that the data set encompasses 537577 rows and 12 indicators of predictability. Also included in this data is geographic data of existing customers. Age, gender, occupation, city category, time spent in the city, and marital status are some of the profile data available for retail employees who worked at least once on Black Friday.

Additionally, it includes information about the products purchased, such as their IDs and categories. In order to predict the amount of the purchase based on the profiles of the customers and the various product categories, we have built machine learning models. The information will enable us to predict Black Friday shopping capital.

II. MOTIVATION AND BACKGROUND STUDY

The popularity of Black Friday has consistently been the highest during the holiday shopping season. Traditionally, this is the day of the year when most people go shopping in the United States (US).

As the holiday shopping season begins the Friday after Thanksgiving in the United States, Black Friday marks the beginning of the holiday shopping season. As retailers point out, this is one of the busiest shopping days of the year.

Sales forecasting can take a variety of forms, but there are two basic approaches: Data-driven and manual processes. In either case, significant time is necessary to develop forecasts. A system like the one proposed can overcome this problem. Predictions can be obtained instantly.

Rodoula H. Tsotsou. [1] There were 540 consumer responses to the online survey. Within two days after Black Friday, Facebook and LinkedIn rolled out a questionnaire with several open and closed questions regarding purchasing intentions, knowledge, and behavior of consumers. Since this Black Friday in Greece was considered for the first time, shopping trends may be inconclusive, which means that the results of the survey might not be as accurate as they ought to be. Consumers' preferences for brands and products differ based on categories, but the survey did not reveal which brands they prefer.

Esther Swilley, Ronald E. Goldsmith.[2]. 225 American study participants were asked about their mindsets and actions for these two shopping occasions. This study modified several constructs from literature available. One point represents strongly disagreed with each question, while seven represents strongly agreed with each question. There was only one category for which moderation was allowed, namely gender. There may have been other variables that would have provided more insight. There was a limitation to the geographical region of the sample.

Briana Milavec. [3] The purpose of consumer research is to determine how consumers interact with their surroundings, which includes not only the environment but also employees

and other consumers. GAM (General Aggression Model) makes it possible to predict Black Friday misbehavior, provided that misbehavior by consumers is categorized as aggression. GAM was the key idea behind this study. For the study, a convenience sample was used. A majority of survey respondents shopped at least once on Black Friday, and all interviewed Jasmin Kwon and Thomas M Brinthaup [4] Authors examined some of the Seasonal Shopping Events (SSEs) from both the perspective of retailers and consumers. In line with the criteria, the examples of events known for their popularity and volume have the potential (Expanding or already doing so) beyond their respective cultures. All four of these events usually occur at the end of the calendar year. (November-December/early January). This paper examines the novelty of seasonal shopping's cultural aspects. A thorough understanding of the origins of such events is lacking. Over the past 100 years, SSE has evolved significantly. An SSE tends to stem from a cultural event that is tied to a national or religious holiday. In contrast, as they develop and become popular, SSEs depart from their origin culture. Gina Castle Bell, Melinda R. Weathers, Sally O. Hastings, Emily B. Peterson. [5] An interview of 65 candidates was done. The data was transcribed and analyzed. The data was analyzed using several coding schemes such as coding initially, incident-by-incidence, focused coding, and axial coding. Data was classified according to themes found in them. The data was then segmented based on theory and theme through thematic analysis. The results were then verified by a member check and a respondent check. The inputs for this survey were derived from only one North Virginian city. It is a survey of brick-and-mortar store customers only, not those of online retailers. Marko Bohaneca, Mirjana Kljajić Borštnarb, Marko Robnik- Šikonja [6] EXPLAIN and IME (Input Method Editor) were utilized as examples. Among its many characteristics are expressive power, algorithmic complexity, and portability. Both of these models provide interpretation and comparison of data.

It may be slow to use this method interactively in a discussion session when dealing with large datasets. In addition due to the overly optimistic nature of these forecasts, further investigation is needed.

Sharron J. Lennon, Minjeong Kim, Jaeha Lee, Kim,K. P. Johnson[7] A pretest was conducted to evaluate the authenticity of the emotion items and determine the affiliation between emotional reactions and three scenarios. 99 volunteers from the same pool of classes as in the main study were considered. A hypothesis test used three variables (Deals on Black Friday, retailer capabilities, and customer service expectations). This article does not take into account vibrant culture. The process has another drawback, as consumer behavior is not incorporated beyond evaluating the Black Friday sales experience.

Shiwani Joshi, Lavi Samuel Mr. Shrey Harsh Baderiya, Prof. Pramila M. Chawan .[8] A purchase decision can be classified into three types based on the system. Various levels of involvement, the price of the purchase, the purchase frequency, and time spent contemplating purchase was considered. This research used item set mining. The item sequences are frequently examined. Due to the large number of complex datasets generated each year, scanning takes a considerable amount of time.

Goutam Majumder. [9] The Random Forest algorithm is used to forecast the purchasing activity for particular demographics of consumers in specific cities. This algorithm builds the decision trees and uses sampling and voting to find the best solution. Only the random forest model is examined. XG Boost serves as a descriptive model for more accurate continuous nature forecasts.

Shubham Banthia, Swapnil Joshi, Pooja Darda. [10] This paper provides a comprehensive analysis of the sales during Black Friday. In this case, the study focuses on the sales generated by online campaigns during Black Friday in India. Researchers surveyed 100 young Indians for data collection. Teenagers were the only members of the group. Kovács András, Tamás Sikos T. [11] A cross-tabulation, cluster analysis, and descriptive methods are used in this study. Analyses of the results of the online survey concerning Black Friday, particularly consumer awareness of Black Friday, relationships between purchase and non-purchase patterns, as well as factors affecting purchasing decisions, such as expenditure, shopping inclinations, and opinions concerning Black Friday. There is too small a sample size to be representative. Those with higher education and a higher level of digital literacy, largely consisting of millennial city dwellers form the majority of participants.

Jose Ramon Saura, Ana Reyes-Menendez, Pedro Palos-Sánchez. [12] This study used an improved three-step method. In the first step, using the Latent Dirichlet Allocation Model (LDA), topics associated with Black Friday were identified. After that, a sentiment analysis (SA) was carried out in order to determine perceptions of the identified topics and overs by Twitter accounts owned by companies. Lastly, a text mining process was used to gain insights from the data. Results of this study were affected by social and economic developments of the time, adding to the intrusiveness of the narrative.

Menuka Maharjan. [13] As opposed to other frameworks, this one has the ability for efficient and seamless processing of diverse data sets, analysis of consumer classification rules to generate insights.

With minimal confidence and support, the Apriori algorithm is used to examine the data. Interpretation of association rules should be done with care. This means that there are no causal links in these associations.

Rao, B. Ida Seraphim.[14] Visualizing the data is followed by developing a model to derive accurate predictions for the variable chosen. There are decision tree models and linear models that can be used. Despite the paper's central focus on sales, marketing strategies have not been examined. Samruddhi K, Dr. Ashok Kumar.[15] The random forest was determined to be the best-fitting model. As a result of this work, they considered two parameters in this algorithm, In particular, n_estimators, which has a value of 100. In addition, random_state allows us to randomly generate our bootstrapping samples when building trees whose value is set as one by default. There is no clear insight to be gained from this black-box approach. Due to the short testing window, it is difficult to find the full range of inputs. It is possible to run into unidentified inputs during the testing process.

III. SYSTEM DESCRIPTION

A. Architecture

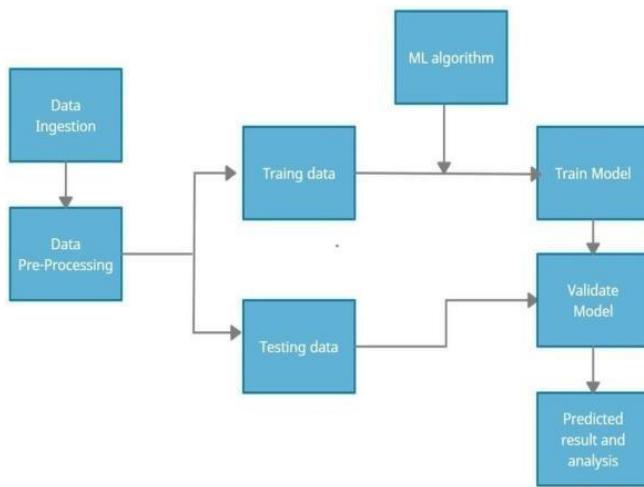


Fig. 1. System Architecture

System architecture is a collection of all the modules of our application. Input, output and processing of the input to get the desired output is shown.

The diagram in Fig. 1 indicates the next steps. Following data pre-processing, a training set is used to develop a model, and a test set is used to verify the model created.

Using five different ML algorithms, the most efficient model will be trained and further validated. Among the five different algorithms considered, the most efficient one is further trained and validated against the test set. The predictor User Interface displays the estimated purchase amount, which helps us to infer the spending capabilities of a particular segment of customers.

B. Activity Diagram

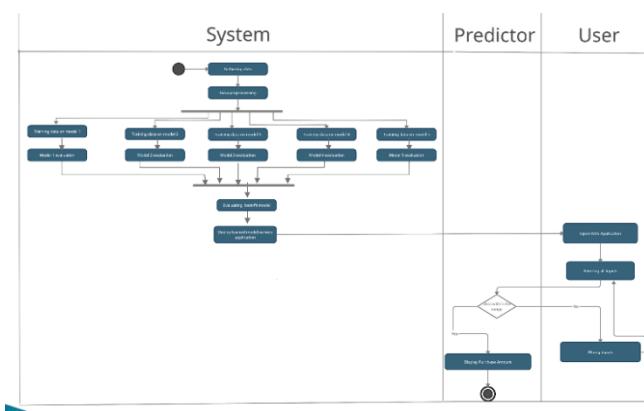


Fig. 2. Activity Diagram

Fig. 2 depicts the behavior between System, Predictor and User. The description of the Activity Diagram is as follows:

- Data should be gathered.
- Raw data will be converted into an understandable format using data preprocessing.

- The pre-processed data will be given to 5 different models for training for analysis, design, development, implementation, and evaluation.
- This will be further used for evaluating the model by comparing the values of evaluation metrics.
- A potential best-fit model will be deployed among all 5 models.

The predictor verifies that the data is within the range of validity. By using this predictor, a user can:

- Open the web application.
- A user can input the required fields to estimate the amount of purchase.

IV. RESULTS

According to the results, the most efficient model was the XG Boost model, which had the lowest RMSE. According to the r2 score, the XG Boost has the highest score of all the model and is fairly close to 1.



Fig. 3. Predictor User Interface

Fig. 4 shows the layout of a single-page web application comprising five entries and a button for predicting . This form contains five numerical and category fields: Age, Occupation code, Stay in current city years, Product Category 1, and Product Category 2.

A user can enter a numeric value in each of the five input fields. However, each value must fall within the specified range.



Fig. 4. Predictor with purchase amount

V. DISCUSSION

Most of the shoppers are between 18 and 25 years of age, and there are significantly more male shoppers than female shoppers. Greater number of shoppers come from the cities considered to be tier one cities in the United States.

Shoppers belonging to occupation code (4) spend the most during sale festivities. The results concerning Black Friday sales prediction of this study complement and extend the work of Gautam Majumder[9]. The random forest algorithm had been assessed as the best fit for sales forecasting. A random forest regressor was used for the prediction purpose. From the sample data, the algorithm creates decision trees, and then voting occurs, and finally voting determines the best solution. Overfitting is greatly reduced by the ensemble method as compared to a single decision tree.

In our study, we found that random forest is a good choice for sales forecasting; however, boosting algorithms are more effective as indicated by our evaluation metrics Root Mean Squared Error (RMSE) and R² (coefficient of determination) regression score.

This application is unique in that it allows users to understand the real-time implications of our analysis. The user can enter specific data for specific use cases and obtain predictions tailored to that use case. Users may also alter the parameters and draw insights from the prediction process.

The data which was contemplated in our project was taken from three different geographic locations across the United States (US). This data gives us a wide range of shopper activity across big, medium, and small cities across the US. This approach can be used for further research, applied to other North American countries that celebrate Black Friday like Canada. A comparative study between the shopping festivities in both these countries could be conducted and analyzed accordingly. The approach we used included black-box approaches like the Random-forest algorithm and a few gradient algorithms like Ada Boost and XG boost. A black-box approach such as ANN (Artificial Neural Network) can be incorporated. A gradient boosting approach such as Light GBM can also be used.

The evaluation metrics we used were RMSE (Root Mean Squared Error) and r²-square. In the future, other evaluation metrics like Mean Squared Error (MSE), Adjusted R can be used. Gradient boosting approaches like Light GBM can also be implemented.

The field of the retail industry is dramatically thriving. The retail sales are expected to boost to 13.5% from 9%. More and more shoppers are purchasing products via mobile devices, increasing e-commerce growth. Physical locations remain crucial, especially now that many e-commerce orders are fulfilled in stores. In today's retail landscape, companies are reimagining their offerings to meet consumers both online and in-store, creating a seamless consumer experience. Innovative thinking and smart investments by retailers are helping to drive this transformation. They improve the customer experience regardless of how they shop.

We have developed a platform to predict the purchase amount that the shoppers might spend based on different features that include age, city category, occupation code and product categories. These features should be inputted as categorical and numerical values and are constrained to a specified range. These 5 features were selected from the 12 features based on feature importance. The platform created to display purchase amount is a single page web application that was created using Flask API.

This study explores Black Friday as a consumption tradition in order to better understand the shopping, planning, and events that occur during this well-known American holiday. Consumers are only participating in 1/3 of Black Friday in-store sales, while the other half are switching to online shopping.

Several advancements and changes have been made in the approaches to shopping and even the way we shop traditionally has been influenced by many factors. As we live in a digital age, a lot of online shopping is taking place compared to in-store shopping. There was a fear that the sales would decrease because in-store shopping was canceled because of the recent pandemic. Since the advent of COVID-19, consumer shopping behavior has been transformed. People are now doing less shopping, spending less money, and preparing for the holidays differently.

The limitations are that firstly, the system is not customizable for different currencies. The amount of the purchase is forecast in dollars only. The second issue is that the data we have covers only limited geographical areas of the USA. With respect to three city categories. In addition, there is only one shopping event for the models.

The retail industry can benefit from the system by identifying an individual's likely spending behavior and financial planning as well as inventory management can be greatly impacted by this. This increases sales for websites and other businesses related to e-commerce. Also, this facilitates scenario-based decision making.

Digital marketing companies within and outside of the technology sector could apply a similar model to expand their reach and gather insights that can help improve their strategies. Retailers can use the web-application to provide their customers with unique experiences and offerings, increasing profits significantly. Retailers and customers from different countries can customize the predictor model to meet their needs.

VI. APPLICATIONS

1. Through this system, the customer can anticipate how much they will spend during a shopping festival.
2. The system can play a crucial role in asset management and fiscal planning.

3. E-commerce sites and associated businesses will be able to segment their customers and increase sales.
4. A model like this could expand the network of digital marketing firms and allow them to optimize their campaigns across various industries.
5. Merchants can offer their clients unique experiences and deals through the software, boosting profits.
6. By analyzing scenarios and data, it will lead to better decisions.

VII. CONCLUSION

Several challenging issues related to sales forecasting are addressed in the paper. Based on previous research, we have created a system for assisting customers and retailers to determine their potential spending before the holiday season begins

REFERENCES

- [1] Tsiotou, Rodoula H," Black Friday: Attitudes, Behavior and Intentions of Greek Consumers",2017.
- [2] Swilley, Esther & Goldsmith, Ronald, "Black Friday and Cyber Monday: Understanding consumer intentions on two major shopping days", Journal of Retailing and Consumer Services, 2013.
- [3] Briana Milavec,"An Analysis of Consumer Misbehavior On Black Friday",2012.
- [4] Jasmin Kwon Vu, Thomas M Brinthaup "The Evolution of Seasonal Shopping Events: Global Perspectives", November 19, 2018.
- [5] Castle Bell, Gina & Weathers, Melinda & Hastings, Sally & Peterson, Emily, "Investigating the Celebration of Black Friday as a Communication Ritual", Journal of Creative Communications,2014.
- [6] Bohanec, Marko, Borstnar, Mirjana, Robnik-Sikonja, Marko, "Explaining machine learning models in sales predictions. Expert Systems with Applications",2017.
- [7] Lennon, Sharron & Kim, Minjeong & Lee, Jaeha & Johnson, Kim, "Consumer Emotions on Black Friday: Antecedents and Consequence", Journal of Research for Consumers,2018.
- [8] Mr. Shrey Harsh Baderiya, Prof. Pramila M. Chawan, "Customer buying Prediction Using Machine-Learning Techniques: A Survey", Volume-05, Issue-10,October, 2018.
- [9] Goutam Majumder, "Analysis And Prediction Of Consumer Behavior On Black Friday Sales",Journal of the Gujarat Research Society, ISSN: 0374- 8588 Volume 21 Issue 10s, November 2019.
- [10] Shubham Banthia, Swapnil Joshi, Pooja Darda "Effect of online marketing on Black Friday Sales in India", Volumn-6,Issue-6,2019.
- [11] András, Kovács & Sikos T., Tamás., "Black Friday Shopping In Hungary", 2019.
- [12] Saura, José & Reyes-Menendez, Ana & Palos-Sánchez, Pedro, "Are Black Friday Deals Worth It? Mining Twitter Users' Sentiment and Behavior Response", Journal of Open Innovation: Technology, Market, and Complexity,2019.
- [13] Maharjan, Menuka, "Analysis of Consumer Data on Black Friday Sales Using Apriori Algorithm", SCITECH Nepal,2019.
- [14] Shiwan Joshi, Lavi Samuel Rao, B. Ida Seraphim," Customer Centric Sales Analysis and Prediction", IJEAT, ISSN: 2249 – 8958, Volume-8 Issue-4, April, 2019.
- [15] Samruddhi K , Dr. Ashok Kumar R, "Applying Different Machine Learning Techniques for Sales Forecasting", ISSN:1001-1749, Volume-16, Issue-5, May 2020.