A SINGLE PAGE APPLICATION TO PREDICT BLACK FRIDAY SALES

Shivangi Karwa[1], Tanisha G Patil[2], Vanitha R Patil[3], Varsha V[4], Prof. Kanchan A Purohit[5]

Department of Computer Science and Technology, Bangalore Institute of Technology

Abstract--- This paper deals with the problem of internal bias and manual-driven process during the sales forecasting. The Black Friday sales are pervaded, all the retailers are packed, & many items are priced for sale at a significantly reduced price. The sales are good that it is hard for clients to obtain the items even with a decent arrangement.

The shop proprietors are more troubled with handling a constrained staff and focusing on prospective customers. While some approaches have been used, they have not been well received.

The prediction model is a system that demonstrated promise in managing this problem. We will use the prediction model to build an exact & efficient algorithm to analyze previous clients spending and forecast their future spending based on similar characteristics.

We utilize different data analysis strategies counting repressors & classifiers to actualize predictions and assess the effectiveness of predictions based on their specificity and accuracy. These machine learning techniques are executed using various algorithms.

Here we present a detailed and visualized analysis and report of 2019 Black Friday sales figures. We considered various parameters such as age group, occupation, product category, city etc, to assess the volume. From this, any store that wants to predict the number of sales could do so.

After finding out the best-fit model, we will deploy it on our user interface developed using Flask.

Keywords: Black Friday, Retailer, Sales, Data Analysis, Machine Learning, Flask

I. INTRODUCTION

Black Friday is celebrated as a shopping holiday on the fourth Thursday following Thanksgiving Day in the United States.

The dataset on Black Friday has been extracted from Kaggle. Summary statistics show that the data set contains 537577 rows and 12 predictor variables. Summary statistics show that the data set accommodates 537577 rows and 12 predictor variables. Additionally, this data accomadates demographic information about customers who have shopped in the past. Among the demographic statistics available are age, gender, occupation, city category, length of time spent in the city, and marital status.

Furthermore, it contains information about the purchased products, such as their IDs and product categories. Using this data, we have created machine learning models that can predict the purchase amount based on the demographics of the customer and the categories of the products. We will use the information to predict the amount spent by individuals during the Black Friday sale festivities.

I. LITERATURE SURVEY

Rodoula H. Tsiotsou. [1] The author conducted research online by running a survey and collected data from 540 consumers. The questionnaire was posted on Facebook and LinkedIn two days after Black Friday with both open and closed questions considering consumer knowledge and behavior, purchase intentions and demographics. Considering it is Greece's first Black Friday, the shopping trends are not very specific, so the results of the survey may not be accurate. It is not clear what brands and what items shoppers prefer in

what categories. Esther Swilley, Ronald E. Goldsmith.[2] The study examined attitudes and behaviors of shoppers for these two shopping occasions based on data collected from 225 U.S. consumers. All constructs in this study were developed and modified from existing literature. They were assessed using a seven-point rating scale, from 1=strongly disagree to 7=strongly agree. Moderation was used only with regard to gender. Other variables would provide useful information. The sample was restricted to a certain area geographically. Briana Milavec. [3] Research focuses on the interaction between consumers and the environment, which includes not only concrete environmental stimuli, but also employees and fellow consumers. GAM (General Aggression Model) can predict Black Friday misbehavior if consumer misbehavior is consumer misbehavior is considered a form of aggression.

In creating the theoretical underpinnings for this research, GAM was influential. Participants for this study are convenience samples. All respondents to the survey shopped at least once on Black Friday, and all interviewed Jasmin Kwon, Thomas M Brinthaupt. [4] Theauthors examine a number of Seasonal Shopping Events (SSE) from the viewpoints of both retailers and consumers.

According to the criteria, the chosen examples of events with a large volume and popularity that have the potential (or have already expanded) beyond their respective cultures. These four SSEs typically occur at the end of the calendar year. The authors examine a number of Seasonal Shopping Events (SSE) from the viewpoints of both retailers and consumers. According to the criteria, the chosen examples of events with a large volume and popularity that have the potential (or have already expanded) beyond their respective cultures.

These four SSEs typically occur at the end of the calendar year (November-December/early January).

The analysis of this research paper focuses on the novelty of the cultural aspect of seasonal shopping. A thorough understanding of events' origins has not been established. The SSE phenomenon has evolved significantly over the past 100 years. In particular, SSEs tend to begin as cultural events tied to national or religious holidays. However, as they mature and become more successful, SSEs evolve away from the culture of originGina Castle Bell, Melinda R. Weathers, Sally O. Hastings, Emily B. Peterson. [5] Sixty-five candidates were interviewed from a

queue of people waiting. A transcription and analysis of the data followed. For data analysis, several coding schemes were used, including coding initially, coding incident by incident, and focused coding and axial coding. Categorization was based on themes found in the data. Thematic analysis was used to segment the data into theory-based and theme-based categories.

A member check and a respondent check verified the results. Only one North Virginian city is used as an input. This survey is only conducted with customers who shop at brick-and-mortar stores and not online retailers. Marko Bohaneca, MirjanaKljaji´cBor˘stnarb, Marko Robnik- Sikonja [6] To illustrate the work, EXPLAIN is utilized as well as IME (Input Method Editor) It exhibits properties that include expressive power, transparency, and algorithmic complexity and Portability. Interpretation and comparison of data are done by these two models.

For large datasets, this method might be slow and require recalculation to be used interactively in a discussion session. A weakness of these forecasts is that they are overly optimistic, necessitating additional investigation.

SharronJ.Lennon,MinjeongKim,JaehaLee,Kim K. P. Johnson. [7] In order to determine the content validity of emotion items used in this study and determine the association between emotional reactions and three scenarios, a pretest was conducted. Volunteers (n = 99) were solicited from classes and were drawn from the same pool as in the main study. Hypotheses testing utilized three expectation variables (expectations of Black Friday deals, expectations of retailer capability, and expectations of poor customer service).The paper fails to include culture vitality. Furthermore, there is another disadvantage to the process, Consumer behavior beyond evaluating black Friday sales experience is not embodied.

Mr. Shrey Harsh Baderiya, Prof. Pramila M. Chawan. [8] The system distinguishes between three types of purchasing decisions. Different levels of involvement, covering different values and frequency of purchase and the amount of time spent reviewing the purchase decision is considered. An item set mining approach was used in this study. Frequent

mining of item sets is plagued by two major problems.

To begin with, the data must be scanned repeatedly each time a search is conducted. Another reason is that complex datasets are generated, the amount of time required to scan is substantial. Goutam Majumder. [9] Using Random Forest algorithm as a working approach, estimation of the purchase amount for certain age groups of customers in certain cities is carried out. With this algorithm, the decision trees are built, based on sampling and voting, the best solution is selected. This paper discusses only one model, the random forest model. A model like XG Boost can serve as a descriptive model for more precise continuous nature predictions. Shubham Banthia, Swapnil Joshi, Pooja Darda. [10] Detailed analysis of sales during Black Friday, as well as the data source and analysis of data have been incorporated into this paper. In this case, the study is confined to the sales resulting from online promotions for Black Friday in India. A sample of 100 young Indians was used to collect data. The 100 Indians' age group is restricted to teens only. KovácsAndrás, TamásSikos T. [11] This research makes use of descriptive methods, cross-tabulations, and cluster analysis. Examination of the results of the online questionnaire research on Black Friday, specifically, to examine the extent to which consumers are aware of Black Friday, purchase/non-purchase relationships and non-purchase trends, primary determinants of purchases including spending, shop and product preferences, and opinions pertaining to Black Friday. Sample size is not representative. Young urban residents make up the majority of respondents. Overrepresented are those with higher education and digital literacy above the average. Jose Ramon Saura, Ana Reyes-Menendez, Pedro Palos- Sanchez. [12] An improved three-step method was used in this study. As a start, using the Latent Dirichlet Allocation Model (LDA), topics were identified in the sample associated with Black Friday. Following that, a sentiment analysis (SA) using Python was carried out in order to determine perceptions of the identified topics and overs by Twitter accounts owned by companies. Finally, we have a data-text mining process called text mining to gain insights. During this study, the socioeconomic events of the period impacted the outcomes, thus contributing to the narrative noise. Menuka Maharjan. [13] Unlike other frameworks, this one can handle different data sets efficiently and easily, analyzes consumer classification rules to produce results.

The data is implementing in Apriori algorithm with minimal support and confidence. Interpreting association rules should be done with great care. As such, they are associations that are not necessarily causally related (compare correlations). Shiwani Joshi, LaviSamuel Rao, B. Ida Seraphim. [14] Visualizing the datais followed by building models to derive accurate predictions of the variable chosen. Models being considered include decision tree models and linear models. Sales are the pivotal component of this paper, even so, marketing strategies haven't been considered. Samruddhi K, Dr. Ashok Kumar. [15] Random forest was determined to be the best-fit model.

As a result of this work, they have considered two parameters for this algorithm, namely n_estimaters that has a value of 100.

Additionally, we have random_state, which allows us to randomize the bootstrapping of our samples when building the treesn whose value is taken as one. A black box approach has been employed, which doesn't divulge much insights.

In limited testing time, it is hard to detect all possible inputs. Probabilities of having unidentified paths in the testing process.

II.   EXISTING SYSTEM

Forecasting sales can be performed in a variety of ways, including both manually and through data-driven processes. Either way, extensive time is needed to create these forecasts.

A large percentage of sales organizations rely on spreadsheets and database software to monitor ongoing relationships with customers. Manual processes are less technology-intensive, but spreadsheets and database software are commonly used to manage the processes.

Sales reps make better looking forecasts when marketing or sales goals are high to gain more commissions. This leads to internal bias within the company when marketing and sales reps prepare forecasts.

IV. PROBLEM STATEMENT

There is a need to develop a system which takes little time to develop forecasts, that is technology inclined rather than manual

which can help the companies to prepare overall forecasts and avail them to surplus their previous profits.

V. PROPOSED SYSTEM

The system proposes a generic function, that takes the algorithm and data as input and builds a model. Itperforms cross-validation to find the best model which will help to improve overall performance.

There are **six modules** in the project are:
1. **Descriptive Analysis**
2. **Exploratory Data Analysis**
3. **Data Preprocessing and Visualization**
4. **Feature Engineering**
5. **Model building and Evaluation**
6. **Model Deployment**

1) Descriptive Analysis:

In this module, Categorical data are first described by counting the number of observations in each category, then expressing them as a percentage of the overall sample size. The dataset is first previewed followed by identifying the dimensions (rows, columns) of the dataset and the data types contained within it, such as target variables, unique elements within each attribute. A statistical summary of the dataset is then obtained. Null and unique values should be checked and the range of the target variable should be determined.

2) Exploratory Data Analysis:

First, some basic data exploration is done and then inference is made based on the assumptions. Univariate analysis is performed initially on the target variable along with numerical predictors and categorical predictors respectively. Then, bivariate analysis is conducted on numerical variables. Multivariate analysis is also performed for one's reference, and all the plots that we obtained are analysed. This module serves the purpose of taking a closer look at the data, including any irregularities, as well as correcting any inconsistencies for the next module, Data Pre-Processing.

3) Data Preprocessing and Visualization:

Categorical features are processed through label encoding, and different types of graphs are plotted accordingly. The missing values are filled in and the unnecessary columns are removed. For the pre-processed data, we identify the top columns which have the greatest influence on the Purchase amount.

4) Feature Engineering:

In this module, the features are scaled and transformed. Then, variables and interaction variables are derived. A function is created to count features. At last, the dataset is divided into test and training set, all the unnecessary columns are dropped and files are exported as modified versions.

5) Model building and Evaluation:

As many models are built, a generic function will be defined which takes the algorithmand data as input and makes the model, performs cross-validation and generatessubmission. The models include Ridge Regression, Decision Tree, Random Forrest, ADABoost, XG Boost.In order to evaluate the model, we will use two metrics: root mean square error (RMSE)and R squared score (r2 score). In statistics, the root of the square root of the variance of the errors is the RMSE.A model with a lower RMSE value is better. The R squared is a statistical measure of howclose the data are to the fitted regression line. Its value is between 0 and 1, and the higher the value the better fit the model is.

6) Model Deployment:

The best-fit model among all the 5 models is deployed to the User interface. Theuser will be able to able to enter the numerical and categorical values to predict purchaseamount. The predictor should able to validate data if it is within the given range anddisplay the purchase amount.

VI. SYSTEM FRAMEWORK

Figure 1. Use Case Diagram

Figure 1 shows the use case diagram where the system and user are the two actors.

Figure 2. Interface Design
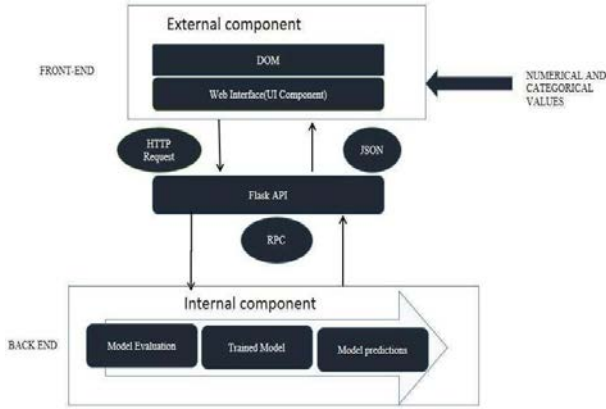
Theuser should be able to enter the numerical data on the user interface. After entering the valid values, the user will be able to view the predicted purchase amount.

Figure 2 provides an interface through which the user can communicate with a computer. There are 2 components-External component and Internal Component. An external component (Front-end) consists of DOM and Web interface. The internal component (Back-end) consists of the Training model, Model evaluation, and prediction.

- DOM-It stands for document object model. It provides an interface for HTML documents.
- JSON- JSON stands for JavaScript Object Notation. JSON is a lightweight format for storing and transporting data. It is used for transmitting data in web applications.
- RPC- It stands for Remote Procedure Call. It is used to request a service from a
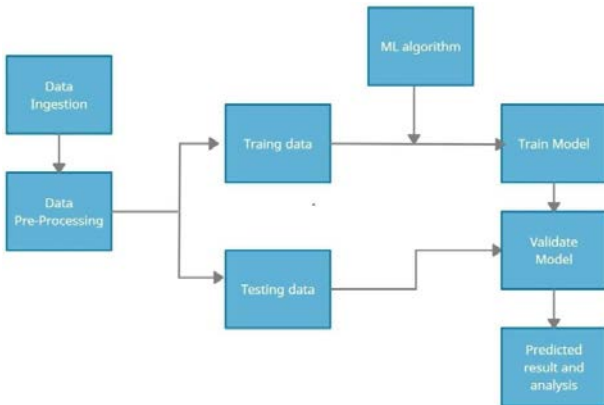
  program.
- Flask API- Flask API is a drop-in replacement for Flask that provides an implementation of browsable APIs.
- HTTP Request- Message sent by the client to initiate an action.

VII. SYSTEM ARCHITECTURE

Figure 3. Architecture Diagram

Figure 3 shows the architecture diagram where

after the completion of data pre-processing, a training set is implemented to build up a model, while a test (or validation) set is to validate the model built. . So, we use the training data to fit the model and testing data to test it. 5 different ML algorithms are applied and the most efficient model will be trained and further used for validation. The purchase amount will be displayed on the screen which gives us an estimate of how much an individual will spend.

VIII. RESULTS

The least RMSE value was found to be the XG Boost model making it the most efficient among all the models. The r2 score of XG Boost is the highest among all models and is comparatively close to one.
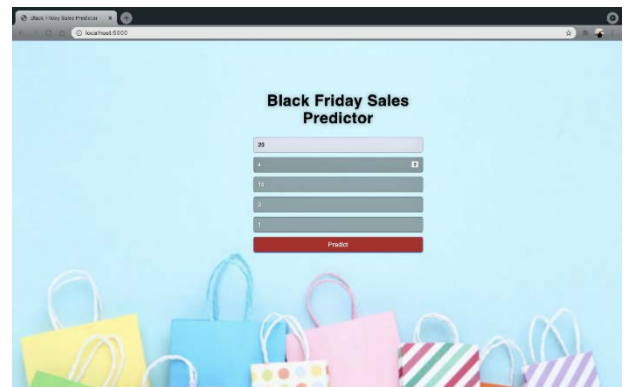

Figure 4. Black Friday sales predictor

Figure 4 shows the single page web-application which consists of five input fields and one predict button.The 5 different numerical and categorical fields are Age, Occupation code, Stay in current city years, Product Category 1, and Product Category 2. Theuserwill be able to fill the five input fields with the appropriate numerical values. All the values must be entered within the given range.
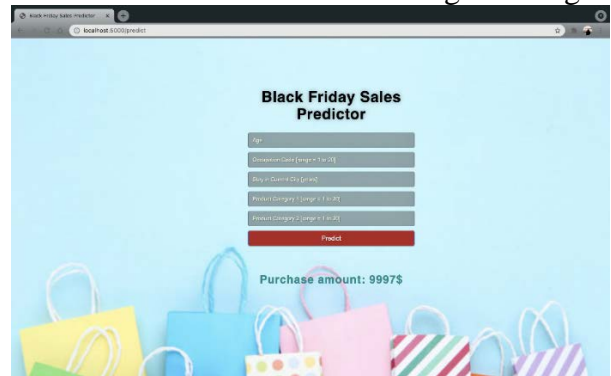

Figure 5. Predictor with purchase amount

Figure 5 shows the single-page web application displaying the purchase amount for the numerical values entered by the user.

IX. APPLICATIONS

1. Using this system, the customer will be able to predict how much they are likely to spend during shopping festivities.
2. The system can play a vital role in financial planning and inventory management.
3. It will help to boost the sale trends of e-commerce websites and other related companies.
4. The model can be applied in technology sector and beyond for digital marketing companies to expand their network and to obtain insights that can help improve their strategies.
5. The web-application will help retailers to give customers differential experiences and offerings which is a prominent way to increase profits.
6. It will help to enable scenario data-driven decision making.

X. CONCLUSION

In this paper, we addressed many challenging issues related to sales forecasting. Based on previous research, we have proposed a system which will help the customers as well as the retailers to prepare them in advance for the shopping festivities by making them aware of the predicted amount they might spend.

XI. FUTURE ENHANCEMENT

Future contributors of this project can make use of stronger gradient algorithms like Light GBM. More efficient black-box approaches like ANN can be incorporated. The predictor model can be further enhanced by retailers and customers belonging to different nations to suit one's needs.

XII. REFERENCES

[1] Tsiotsou, RodoulaH, "Black Friday: Attitudes, Behaviorand Intentions of Greek Consumers",2017.

[2] Swilley, Esther & Goldsmith, Ronald, "Black Friday and Cyber Monday: Understanding consumerintentionson two major shopping days", Journal of Retailing and Consumer Services,2013.

[3]Briana Milavec, "An Analys is of Consumer Misbehavior On Black Friday",2012.

[4] Jasmin Kwon Vu, Thomas M Brinthaupt "The Evolution of Seasonal Shopping Events: Global Perspectives",November19,2018.

[5] Castle Bell, Gina & Weathers, Melinda & Hastings, Sally & Peterson, Emily, "Investigating the Celebration of Black Friday as 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 forConsumers,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, November2019.

[10] Shubham Banthia, Swapnil Joshi, PoojaDarda "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-Sanchez, Pedro, "Are Black Friday Deals Worth It? Mining Twitter Users' Sentiment and Behavior Response", Journal of Open Innovation: Technology, Market, andComplexity,2019.

[13] Maharjan, Menuka, "Analysis of Consumer Data on Black Friday Sales Using Apriori Algorithm", SCITECHNepal,2019.

[14] Shiwani Joshi, Lavi Samuel Rao, B. Ida Seraphim," Customer Centric Sales Analysis and Prediction", IJEAT, ISSN: 2249 – 8958, Volume-8 Issue-4, April,2019.

[15] SamruddhiK ,Dr Ashok Kumar R, "Applying Different Machine Learning Techniques for Sales Forecasting", ISSN:1001-1749, Volume- 16, Issue-5, May2020.