

# BLACK FRIDAY SALES PREDECITION USING MACHINELEARNING

K.SUPARNA <sup>1</sup>, YEKULA JASWANTHI<sup>2</sup>

<sup>1</sup> Assistant Professor MSc (CS), DEPT, Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**

Email id:- [suparnakalidindi@gmail.com](mailto:suparnakalidindi@gmail.com)

<sup>2</sup>PG Student of MSc Computer Science, Dantuluri Narayana Raju College,

**Bhimavaram, Andhra Pradesh**

Email id :- [yjaswanthi1999@gmail.com](mailto:yjaswanthi1999@gmail.com)

## ABSTRACT

Understanding the purchase behavior of various customers (dependent variable) against different products using their demographic information (IS features where most of the features are self-explanatory). This dataset consist of null values, redundant and unstructured data. Machine learning is the most common applications in the domain retail industry. This concept helps to develop a predictor that has a distinct commercial value to the shop owners as it will help with their inventory management, financial planning, advertising and marketing. This entire process of developing a model includes preprocessing, modelling, training testing and evaluating. Hence, frameworks will be developed to automate few of this process and its complexity will be reduced.

## 1 INTRODUCTION

“Black Friday” is the name given to the shopping day after thanksgiving. This day was actually considered as “Black Friday” on the grounds that the number of customers made auto collisions and some of the time even violence. Police begat the saying to depict the disorder encompassing the congestion of pedestrian and auto traffic in downtown shopping regions. In retail industry, the number of sales play an important part that decide the loss a profit for the company. Predicting the sales accurately gives the efficient industry management. Black Friday is like a carnival sale in the USA. In this day huge sale occurs in a very less price for the products which are much demanded. To incur the sales, a prediction model is made to hover on the type of product which is sold in maximum numbers. A customer’s behavior is to be analyzed in order to predict the amount of purchase to be done by him/her on a particular day. In this paper, we will predict the sales of a company on "Black Friday" .

## 2. LITERATURE SURVEY AND RELATED WORK

Empirical assessment

The wide-variety of real-time software systems, including telecontrol/telepresence systems, robotic systems, and mission planning systems, can entail dynamic code synthesis based on runtime mission-specific requirements and operating conditions. This necessitates the need for dynamic dependability assessment to ensure that these systems perform as specified and not fail in catastrophic ways. One approach in achieving this is to dynamically assess the modules in the synthesized code using software defect prediction techniques. Statistical models; such as stepwise multi-linear regression models and multivariate models, and machine learning approaches, such as artificial neural networks, instance-based reasoning, Bayesian-belief networks, decision trees, and rule inductions have been investigated for predicting software quality. However, there is still no consensus about the best predictor model for software defects. In this paper; we evaluate different predictor models on four different real-time software defect data sets. The results show that a combination of IR and instance-based learning along with the consistency-based subset evaluation technique provides a relatively better consistency in accuracy prediction compared to other models. The results also show that "size" and "complexity" metrics are not sufficient for accurately predicting real-time software defects.

#### A comparative study

The purpose of this paper is to compare the accuracy of various linear and nonlinear models for forecasting aggregate retail sales. Because of the strong seasonal fluctuations observed in the retail sales, several traditional seasonal forecasting methods such as the time series approach and the regression approach with seasonal dummy variables and trigonometric functions are employed. The nonlinear versions of these methods are implemented via neural networks that are generalized nonlinear functional approximators. Issues of seasonal time series modeling such as deseasonalization are also investigated. Using multiple cross-validation samples, we find that the nonlinear models are able to outperform their linear counterparts in out-of-sample forecasting, and prior seasonal adjustment of the data can significantly improve forecasting performance of the neural network model. The overall best model is the neural network built on deseasonalized time series data. While seasonal dummy variables can be useful in developing effective regression models for predicting retail sales, the performance of dummy regression models may not be robust. Furthermore, trigonometric models are not useful in aggregate retail sales forecasting.

#### Sample research the analysis

Ample research is carried out on the analysis and prediction of sales using various techniques. There are many methods proposed to do so by various researchers. In this section, we will summarize a few of the machine learning approaches. C. M. Wu et al. [1] have proposed a prediction model to analyze the customer's past spending and predict the future spending of the customer. The dataset referred is Black Friday Sales Dataset from analyticsvidhya. They have machine learning models such as Linear Regression, MLK classifier, Deep learning model using Keras, Decision Tree, and Decision Tree with

bagging, and XGBoost. The performance evaluation measure Root Mean Squared Error (RMSE) is used to evaluate the models used. Simple problems like regression can be solved by the use of simple models like linear regression instead of complex neural network models. Odegua, Rising [2] have proposed a sales forecasting model. The machine learning models used for implementation are K-Nearest Neighbor, Random Forest, and Gradient Boosting. The dataset used for the experimentation is provided by Data Science Nigeria, as a part of competitions

performance evaluation

based on Machine Learning. The performance evaluation measures used are Mean Absolute Error (MAE). Random Forest outperformed the other algorithms with a MAE rate of 0.409178. Singh, K et al [3] have analyzed and visually represented the sales data provided in the complex dataset from which we ample clarity about how it works, which helps the investors and owners of an organization to analyze and visualize the sales data, which will outcome in the form of a proper decision and generate revenue. The data visualization is based on different parameters and dimensions. The result of which will enable the end-user to make better decisions, ability to predict future sales, increase the production dependencies on the demand, and also regional sales can be calculated. S. Yadav et al [4] have analyzed and compared the performance of K-Fold cross-validation and hold-out validation method. The result of the experimentations where k-fold cross-validation gives more accurate results. The accuracy results of K - Fold cross-validation were around 0.1 - 3% more accurate as compared to hold-out validation for the same set of algorithms. Purvika Bajaj et al. [5] have performed sales prediction based on a dataset collected from a grocery store. The algorithms used for experimentations are Linear Regression, K-Nearest Neighbors algorithm, XGBoost, and Random Forest. The result precision is based on Root Mean Squared Error (RMSE), Variance Score, Training, and Testing Accuracies. The Random Forest algorithm outperforms the other three algorithms with an accuracy of 93.53%. Ramasubbareddy S. et al. [6] have applied machine learning algorithms to predict sales. The dataset for the experimentation purpose is taken from Kaggle, named as Black Friday Sales Dataset. The algorithms used for the implementation of the system are linear regression, Ridge Regression, XGBoost, Decision Tree, Random Forest, and Rule-Based Decision Tree. Root Mean Squared Error is used as the performance evaluation measure. As per RMSE lower the RMSE value better the prediction. As a result, based on the RMSE rate Rule-Based DT outperforms other machine learning techniques with a RMSE rate of 2291. Aaditi Narkhede et al.[7] has applied machine learning algorithm in tracking sales at places like shopping center big mart to anticipate the demand of customers and handle the management of inventory accordingly the methods presented here are an effective method for data shaping and decisionmaking. New ways that can better identify consumer needs and calculate marketing plans which will improve sales. M.Sahaya Vennila et al. [8] have analyzed, preprocessed, and applied machine learning techniques to predict sales. The dataset used for the analysis and experimentation purpose is Black Friday Sales Dataset from Kaggle. The dataset is preprocessed. K - Fold method is used for the purpose of splitting the dataset into training and testing datasets. The prediction model is implemented using Linear Regression.

### 3 PROPOSED WORK AND ALGORITHM

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

Using this system, the customer will be able to predict how much they are likely to spend during shopping festivities. The system can play a vital role in financial planning and inventory management .It will help to boost the sale trends of ecommerce websites and other related companies. 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. The web-application will help retailers to give customers differential experiences and offerings which isa prominent way to increase profits. It will help to enable scenario data-driven decision making.

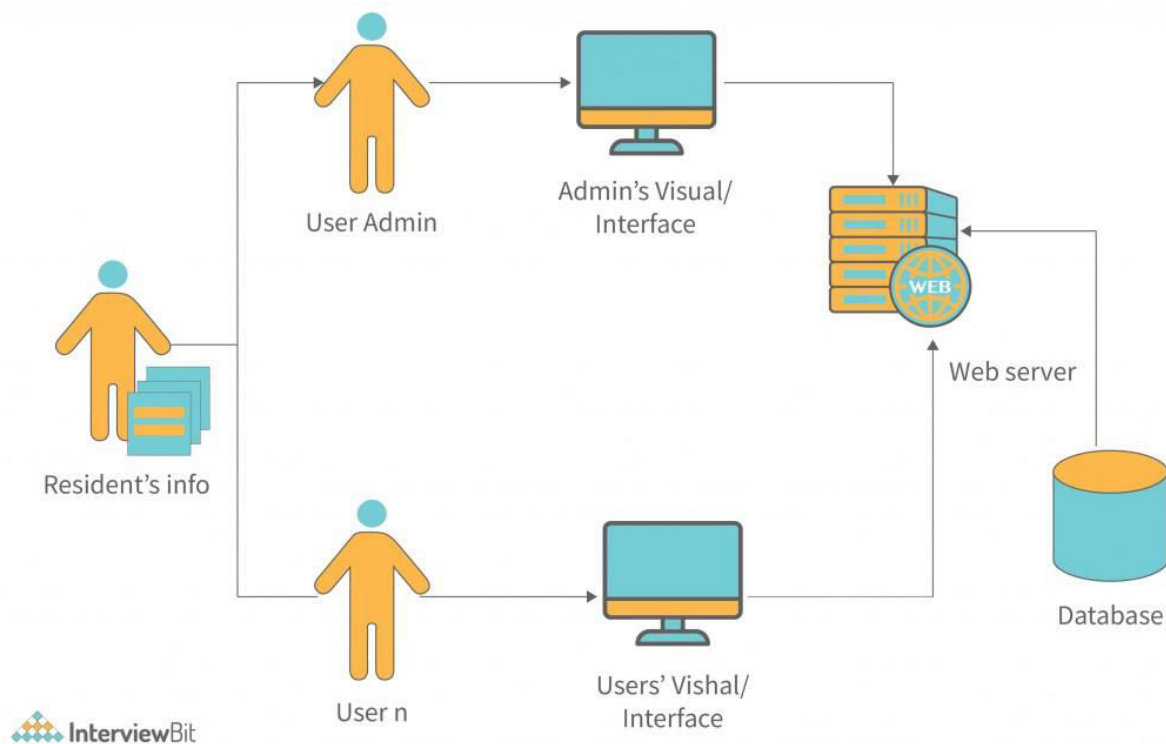


Fig-1: System Architecture

## 4 METHODOLOGIES

Building a campus placement prediction system involves several modules and steps. Here are some key modules to consider:

**Data Collection:** Gather historical data related to campus placements. This may include information about students, their qualifications, the companies that visited the campus, job offers made, and other relevant details.

**Data Preprocessing:** Clean and preprocess the data to handle missing values, outliers, and ensure data quality. This step is crucial for accurate predictions.

**Feature Engineering:** Create relevant features from the data that can help in making predictions. This may involve transforming and selecting the most informative features.

**Data Visualization:** Visualize the data to gain insights and understand patterns. Visualization can also help in feature selection and model evaluation.

**Model Selection:** Choose appropriate machine learning or deep learning algorithms for the prediction task. Common choices include logistic regression, decision trees, random forests, support vector machines, and neural networks.

**Model Training:** Train the selected model on the preprocessed data. This involves splitting the data into training and testing sets and fine-tuning hyperparameters.

**Model Evaluation:** Assess the model's performance using evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, depending on the nature of the problem (binary classification, multi-class classification, etc.).

### 5.RESULTSANDDISCUSSION SCREENSHOTS

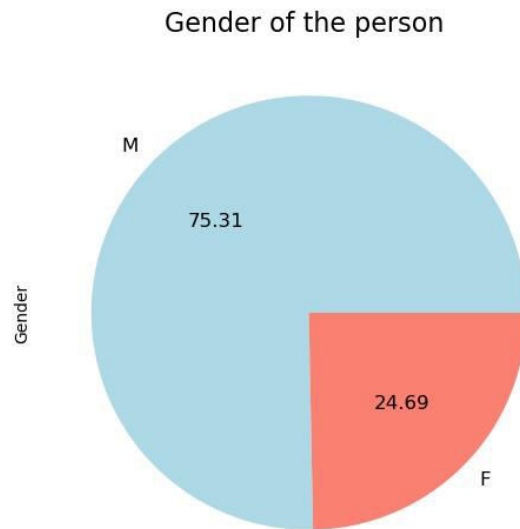


Fig 2:- Pie-chat represents number of males and females in dataset

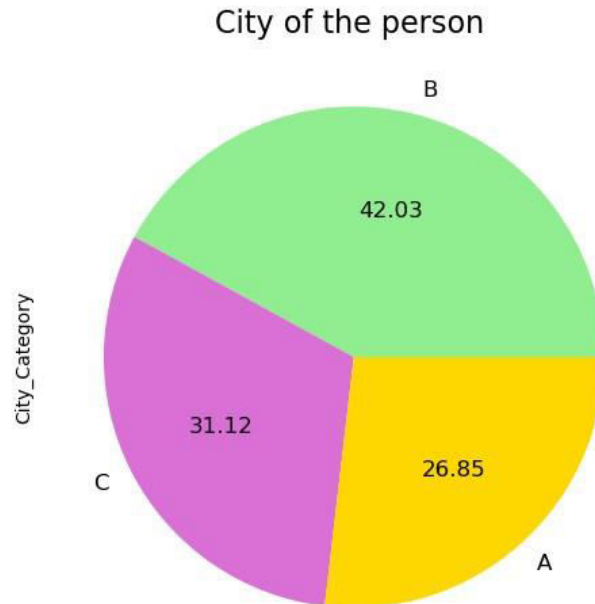


Fig 3:- Person city

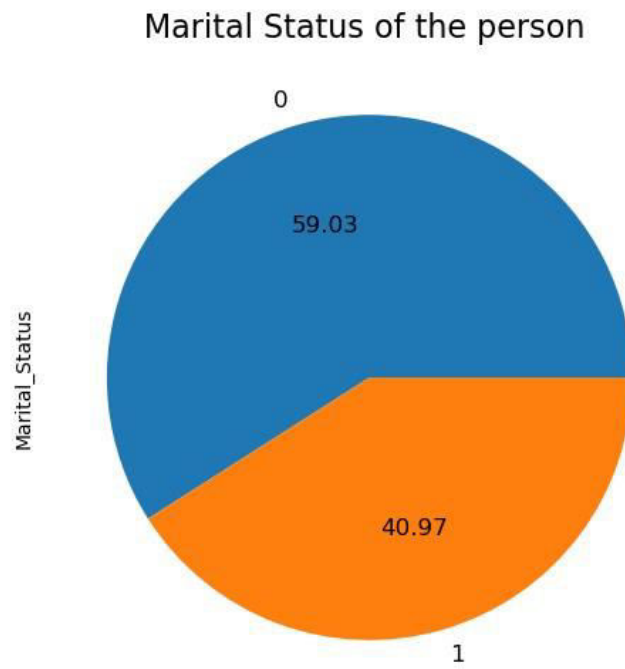


Fig 4 :- Marital status

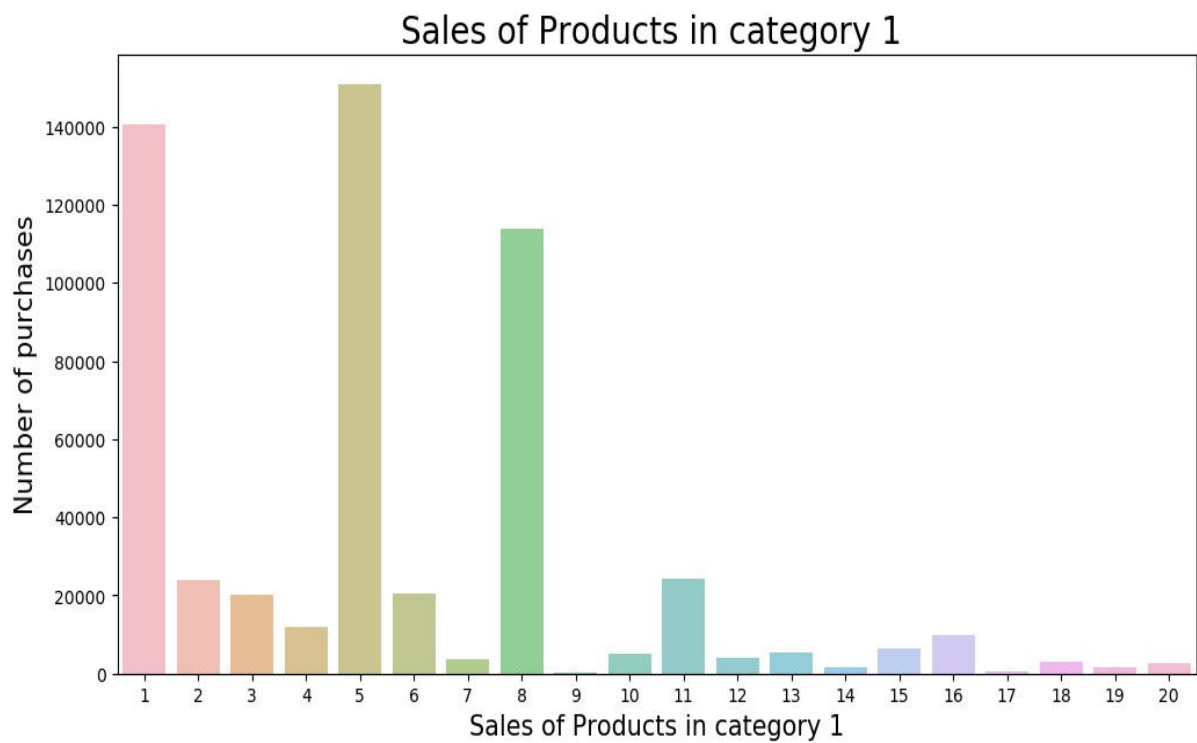


Fig 5 :- Sales of product in category 1

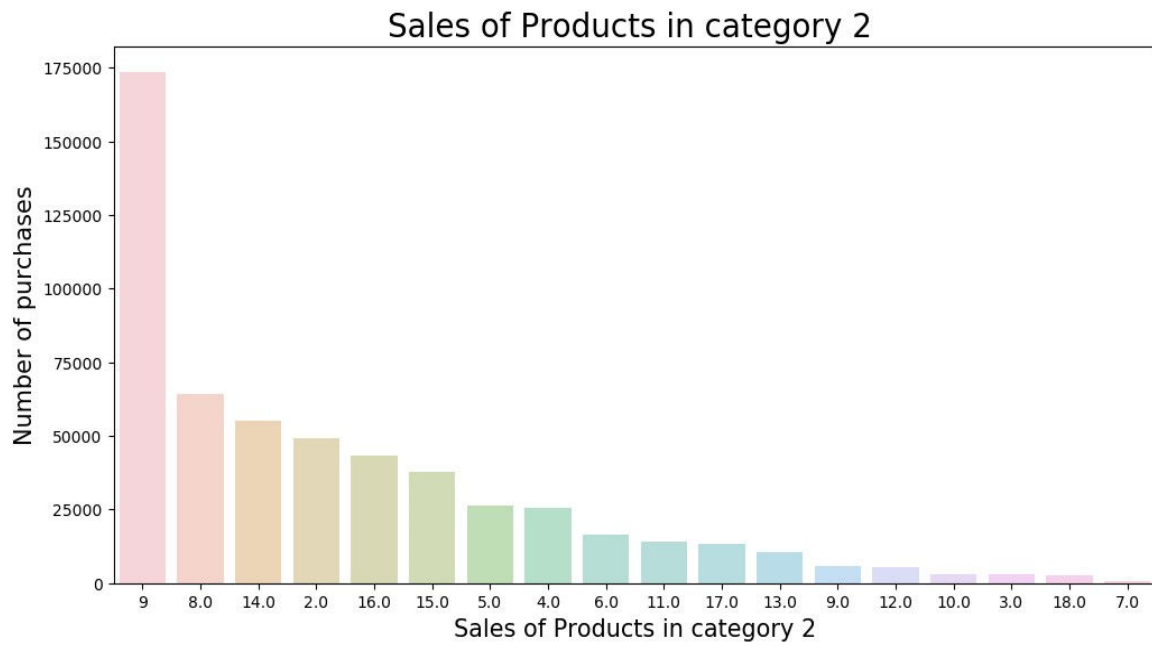


Fig 6 :- Sales of product in category 2

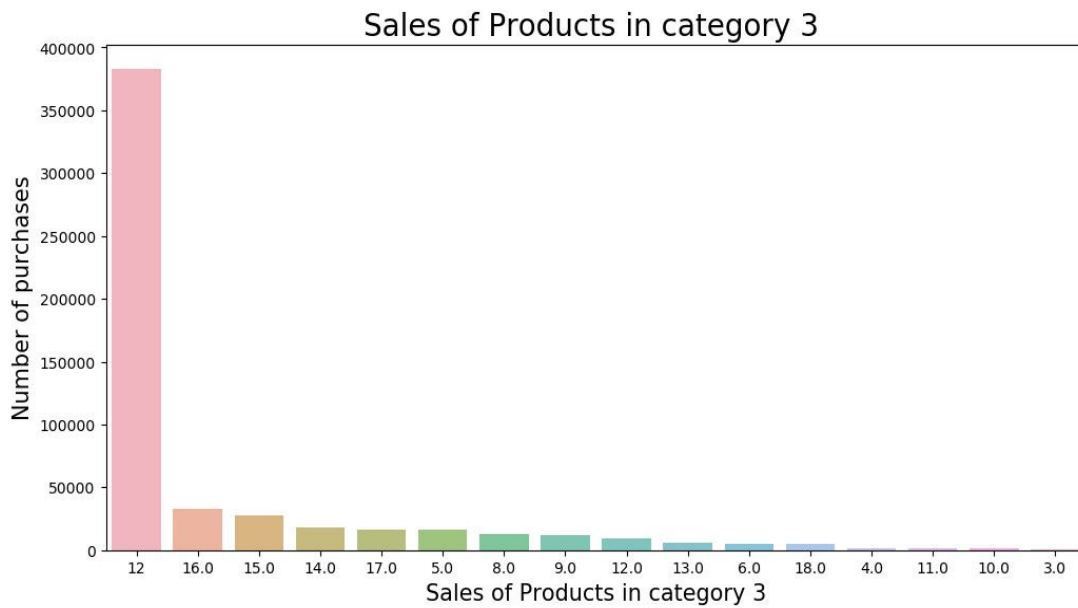


Fig 7 :- Sales of product in category 3



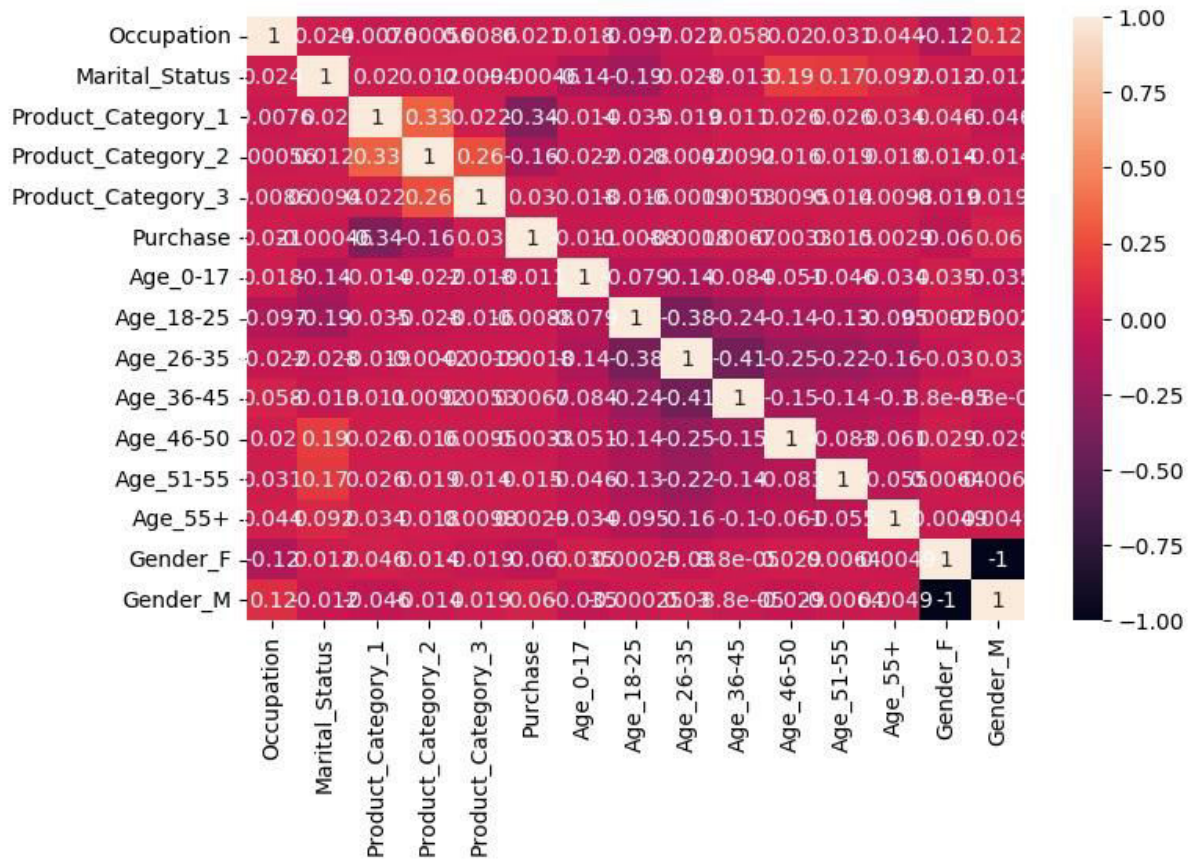


Fig 8 :- Detailed representation of Dataset

### 6.CONCLUSION

In this project, 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.

### 7.REFERENCES

1. Tsiotsou, RodoulaH, "Black Friday:Attitudes, Behaviorand Intentions of GreekConsumers",2017.
2. Swilley,Esther&Goldsmith,Ronald,“Black Friday and Cyber Monday: Understandingconsumerintentionsontwomajorshopping days”,JournalofRetailingandConsumerServices,2013.
3. Briana Milavec, “An Analys is of ConsumerMisbehaviorOn Black Friday”,2012.
4. JasminKwonVu,ThomasMBrinthaupt“TheEvolutionofSeasonalShopping.

5. Events:GlobalPerspectives”,November19,2018.
6. CastleBell,Gina&Weathers,Melinda&Hastings,Sally&Peterson,Emily,“Investigating theCelebrationofBlackFridayasCommunicationRitual”,JournalofCreativeCommunica tions,2014.
7. Bohanec,Marko,Borstnar,Mirjana,Robnik-Sikonja,Marko,“Explaining
8. machinelearningmodelsinsales predictions. ExpertSystemswith Applications”,2017.
9. Lennon, Sharron & Kim, Minjeong & Lee,Jaeha & Johnson, Kim,“ Consumer EmotionsonBlackFriday:AntecedentsandConsequence”,JournalofResearchforConsu mers,2018.
10. Mr.ShreyHarshBaderiya,Prof.PramilaM.Chawan,“CustomerbuyingPredictionUsing Machine-LearningTechniques:ASurvey”,Volume-05,Issue-10,October,2018.
11. GoutamMajumder,“AnalysisAndPrediction Of Consumer Behavior On BlackFriday Sales”,Journal of the Gujarat ResearchSociety, ISSN: 0374- 8588 Volume 21 Issue10s,November2019.
12. ShubhamBanthia,SwapnilJoshi,PoojaDarda“EffectofonlinemarketingonBlack Friday Sales in India”,Volumn-6,Issue-6,2019.
13. András,Kovács&SikosT.,Tamás.,“BlackFridayShoppingInHungary”,2019