AJIT-e Academic Journal of Information Technology
10.5824/ajite.2024.03.001.x

2024 . Summer - Yaz . Volume – Cilt: 15 . Issue – Sayı: 3

Analyzing & Predicting Black Friday Sales of a Retailer

Aşkın DEMİRAĞ, *Yeditepe University, Faculty of Computer and Information Sciences, Department of Management Information Systems, Associate Prof. Dr., ademirag@yeditepe.edu.tr, 0000-0001-7868-0438*

Fatma Dilan ÖNCEVARLIK YILDIZ, *Yeditepe University, Graduate School of Social Sciences, Department of Management Information Systems, M.A., dilanfatma.oncevarlik@std.yeditepe.edu.tr, 0000-0002-9886-2340*

ABSTRACT

In this project, a retail store's Black Friday sales from prior years were examined, and sales projections for the next years were made using this information. A few demographic details and consumer basket amounts are among the data examined for this research. Correlations were found using the data that was examined for the study. Stores can identify their Black Friday sales strategy and use this knowledge to increase sales by evaluating the relationships between demographic data about their clients and their shopping behaviors. The artificial intelligence-supported forecasting model created as part of the research was trained using the same data, and sales projections for the upcoming years were computed. The consistency rate of the system was evaluated by comparing the sales data of the same retail store to the system that was trained using past sales data.

Keywords **:** Black Friday Sales, Machine Learning Based Predictions,Sales Prediction

Analyzing & Predicting Black Friday Sales of a Retailer

ÖZ

Bu projede bir perakende mağazasının geçmiş yıllardaki Black Friday satışları analiz edilmiş ve bu verilerden hareketle gelecek yıllara ait satışları tahmin edilmiştir.
Proje kapsamında analiz edilen veriler, müşterilerin bazı demografik bilgilerini ve sepet tutarlarını içermektedir. Proje kapsamında analiz edilen veriler ile koreasyonlar ortaya çıkarıldı. Müşterilerin demografik bilgileri ve satın alma alışkanlıkları arasında bulunan korelasyonların analizi ile black friday dönemi mağazaların satış stratejileri belirlenebilir ve daha çok satış yapmak için bu bilgiler kullanılabilir.
Aynı veriler kullanılarak proje kapsamında geliştirilen yapay zeka destekli tahmin modeli eğitilmiş ve gelecek yıllara ilişkin satış beklentileri hesaplanmıştır. Daha önceki satış verileri ve demografik verilerle eğitilen sisteme aynı perakende mağazanın mevcut olan ve eğitimde kullanılmayan satış verileri sorularak sistemin tutarlılık oranı test edilmiştir.

Araştırma Makalesi / Research Article
Date Received: 1.07.2024
Publishing Date: 31.08.2024

AJIT-e Academic Journal of Information Technology
2024 Summer/Yaz – Cilt/Vol: 15 - Sayı/Issue: 3
10.5824/ajite.2024.03.001.x

Anahtar Kelimeler : Black Friday Satışları, Makine Öğrenmesi Temelli Tahminleme, Satış Tahmini

INTRODUCTION

The things a person purchase can be used to explain shopping. However, there is more to why individuals spend money than merely buying stuff. Almost everything a person does, including shopping, is done for pleasure. It fits into the daily quest for happiness that people have. (Murray, 2014) Retail and e-commerce companies, which are aware of this urge of people, are in pursuit of making the most sales by offering the best campaign to their customers, especially on Black Friday, the most known shopping day of the year. America's favorite shopping day of the year, Black Friday, saw over 100 million people flock to stores and e-commerce platforms in 2016 in pursuit of what were supposedly the best prices of the year.

Since the debut of the modern Macy's Thanksgiving Day Parade in 1924, the Friday after Thanksgiving has been considered the unofficial start of a hectic holiday shopping season. In the 1960s, Philadelphia police who were handling big consumer crowds coined the name "Black Friday" (in the retail sense). But a lot of people also think about the phrase from an accounting standpoint. "Black" refers to the shift in store colors from "red" to "black," during the manual accounting record-writing era when red ink represented a loss and black ink a profit. (BlackFriday.com, 2020)

Figure 0 Origin of the name Black Friday (BlackFriday.com, 2020)

In the modern world, every movement of a person means data. The products s/he looks at, the items s/he buys and the social media posts s/he likes actually inform the systems that can analyze this data about that person. Thanks to this advanced technology, companies may prefer to get to know their customers and offer special services instead of advertising campaigns whose results cannot be followed efficiently. Companies that make big campaigns, especially during Black Friday, can make predictions about the next year using the purchasing data they already have. Based on these estimates, it can determine its new campaigns and target audience. This study examined the sales of a retail market during the Black Friday

period. In addition, the effects of factors such as gender on sales were analyzed, and forecasts were made for the upcoming Black Friday sales.

1.1 Motivation and Goals

This study's inspiration and main objective stem from an investigation of the complex interactions that exist between demographic characteristics and consumer purchasing patterns during the hugely important Black Friday shopping event. Finding empirical information that clarifies whether a noticeable relationship exists between these two crucial aspects of customer behavior is the main goal. In addition, this project aims to innovate by utilizing artificial intelligence (AI) to create a complex predictive model that can anticipate future sales trajectory in relation to Black Friday.

Businesses and marketers may use this information to modify their tactics more effectively and efficiently, which can prove to be extremely lucrative. Knowing these relationships can help with marketing campaign optimization, which can ultimately save expenses while improving customer service. The results of this study are expected to play a significant role in directing future investigations and shaping business procedures.

A key component of this research project is the creation of an AI model. This model will be a cutting-edge instrument that can forecast future sales trends in addition to analyzing historical data. It is anticipated that the model's predictive powers will offer businesses and marketers insightful information that will empower them to make data-driven decisions and predict consumer behavior. This proactive strategy could transform how companies get ready for and profit from the Black Friday phenomenon, increasing their competitiveness in the retail market.

In summary, the goal of this research project is to make a substantial contribution to our understanding of the relationship between consumer demographics and Black Friday purchase behaviors. It also seeks to equip companies with a state-of-the-art AI model that predicts future sales patterns. By fulfilling these goals, this work contributes to the body of knowledge in academia and provides real value to companies trying to successfully negotiate the challenging landscape of Black Friday sales.

1.2 Related Works

The technique of forecasting future consumer demand over a specified period using historical data and other information is known as demand forecasting. (Hand, 2024) Demand forecasting can have different usage and advantages for different business models. This approach can help companies that sell fast-moving consumer goods to create a strategy to keep their products in their warehouses for a short time, as well as help companies with larger products such as cars, furniture, and houses to make predictions about the future and prepare according to this forecast. Demand forecasting, from warehouse management to determine a

10.5824/ajite.2024.03.001.x

sales strategy at the end of the day, is an approach that will benefit businesses. E-commerce sites, on the other hand, can learn the user load that will periodically visit their sites by demand estimation and improves their infrastructure and the service they will provide according to these estimates.

In A prediction model for automobile sales in Turkey using deep neural networks study, the authors used the experts to determine the indicators such as the number of automobile sales, time (month), dollar exchange rate, gross domestic product, consumer confidence index, and consumer price index while predicting automobile sales. (Kayapınar Kaya & Yıldırım, 2020)

DNNs (Deep Neural Networks) are a type of neural network that has numerous layers between the input and output layers. When compared to shallow structures, the ability to generalize is said to improve with network depth. (LeCun, Bengio , & Hinton , 2015) Due to their classification and prediction capability, the DNN model was preferred by authors Kayapınar Kaya & Yıldırım.

Researchers intend to generate predictions with historical sales data using the time series model in the study Research on sales information prediction system of e-commerce enterprises based on the time series model. (Liu, Liu, Xu , & Zhang, 2020) In their research, they use a time series model to create a sales forecasting and analysis model for products with common features, and they forecast the sales inventory of a certain type of product from a quantitative standpoint. Their tests revealed that the time series model has a positive impact on sales forecasting.

DEFINITION OF THE SELECTED PROBLEM

As mentioned in the Introduction, Black Friday is a shopping event that takes place once a year, where brands offer discounts for consumers. Stores that provide online and offline campaigns provide their consumers with a shopping event only once a year. Meanwhile, stores sell a lot of products in a short time and melt their stocks. This event, which takes place in November every year, has become a tradition that consumers look forward to and in some cases plan their shopping accordingly. For the stores, this event, which takes place near the end of the year, can be seen as a good opportunity for them to enter the new year with new stocks.

In this study, a retail sales data of 2017, 2018, 2019, 2020 and 2021 are included and forecasts are made for next year. In addition to the data used in the study, the information about the Black Friday period of these years in general was examined.

According to an article published in PracticalEcommerce in 2018, Black Friday in 2018 broke the record by generating more revenue than any Black Friday that took place until that year:

Consumers embraced 2018 online holiday shopping with gusto starting on Thanksgiving Day and culminating on Cyber Monday, which became the highest U.S. ecommerce sales day in history with $7.9 billion in revenue. Thanksgiving Day, Black Friday, and Cyber Monday all exhibited healthy year-over-year sales growth. The most popular products were televisions, computers and other electronics, and toys. (Kaplan, 2018)

The year 2018 set a record, garnered a lot of attention, and generated a lot of study. This interest has meant that there is more data for companies to study and prepare for the upcoming shopping events, as many research companies have approached this issue from different angles. The product groups that consumers preferred the most were revealed as a result of the examination Google conducted in online sales channels in 2018. According to the results, consumers preferred the most to buy books, and household goods was placed at the bottom of the ranking. (Eberhard, 2018)

In 2019, the biggest shopping day of the year was measured in terms of time spent in the store, not Black Friday, but Christmas Eve. According to the same study, although users' purchases broke a record on Black Friday in 2019, Christmas Eve surpassed Black Friday in terms of spending time in the store. In the research, it is interpreted that Black Friday customers tend to shop online, while users who shop on Christmas Eve cannot wait for shipping time because they shop for last-minute gifts.

In 2020, the inclination for online delivery increased and took over as the primary method of purchasing for that season. Thanksgiving Day shopping was all but eliminated by the coronavirus since most stores chose to stay closed for the first time in decade. (BlackFriday.com, 2020)

In this study, the sales of a retail store were examined and predictions were made about the sales of the future period with more than one machine learning method. In addition to these estimations, inferences were made about the shared data. Research was conducted to choose the most accurate method to be used during the study. The research steps are explained.

The data used in the study was taken from kaggle.com, which includes five years of sales data from a retail store. The sales data for 2017, 2018, 2019, 2020, and 2021 includes gender, elite membership status, age, and country information in an anonymous way. This data set, which has approximately 136,950 rows of data in total, contains the total amount of expenditure made by customers during shopping. In this project, the whole year is examined to predict the Black Friday week sales.

10.5824/ajite.2024.03.001.x

RESEARCH METHODOLOGY

When the data set quality is examined for this study, the topics to be considered are as follows;

1. Null value
2. Incorrect data that can be accepted as Outlier
3. Input method that does not conform to the standard of the data less than 1% in total.

As a result of the quality of the data, there was no visible decrease in the amount of data that can be used in the study. The total amount of data that can be used has increased. In this way, while the machine learning methods suitable for the system were tried, the data was sufficient for training and testing. The unused 1% was removed from the data for study. The remaining data was made available to train the systems.

Machine learning methods were preferred in the study. The reason why machine learning methods are preferred is to keep the algorithm simple and understandable compared to deep learning methods. Because deep learning is a method used to capture complex relationships found in larger datasets, it was not an appropriate approach to the existing dataset.

The necessity to effectively handle massive information and identify correlations that traditional analytic methods can miss led to the choice to use artificial intelligence techniques. The chosen artificial intelligence techniques were followed for preprocessing the data to fill in any missing values. Conventional techniques of data preparation, which require manual analysis, are sometimes laborious. Artificial intelligence streamlines the analytical process by not requiring the knowledge of subject matter experts, in contrast to traditional methodologies. But it's important to recognize that artificial intelligence has limitations. Because of built-in limitations, it might miss some subtleties. There are a number of benefits to using artificial intelligence (AI) instead of conventional or classical methods, especially when it comes to difficult problem-solving and large-scale data processing.

1. Managing Big Data: Compared to traditional approaches, AI is superior at processing and analyzing large datasets, which may be time-consuming or impracticable.
2. Quickness & Effectiveness: Quick analysis and forecasts are produced by AI algorithms since they run quickly. When working with real-time or time-sensitive data, this efficiency is essential since it enables businesses to act swiftly

and decisively. Large datasets may not scale well for traditional approaches like manual analysis, which are often slower.

3. Identification of Patterns: AI is very good at identifying intricate patterns and relationships in data, especially with machine learning models. This is especially helpful in situations where the relationships are too complex or not immediately obvious for conventional approaches to recognize and understand.
4. Automation: By automating monotonous jobs, AI systems can minimize the need for human intervention.
5. Forecasting: AI is skilled in forecasting and predicting the future with accuracy based on past data. By using historical patterns and trends as a guide, machine learning models can forecast future events and offer insightful information for making decisions.
6. Scalability: Growing datasets and rising complexity may be handled by AI systems with ease. While traditional approaches may encounter scalability issues, AI systems can maintain their ability to perform effective analysis as the volume of data increases.

3.1. Design & Implementation

The Black Friday time period is taken from an official calendar of the United States.

Table 1: Black Friday Dates

Year	Black Friday Date
2017	November 24
2018	November 23
2019	November 29
2020	November 27
2021	November 26
2022	November 25
2023	November 24

The data used in the study are the sales of world-wide stores of an American retail store shared on kaggle.com. The data includes approximately 136,950 rows of sales in total. Each sales data includes the gender of the customer, how many products they bought, whether the customer is an elite customer or not, country and age.

10.5824/ajite.2024.03.001.x

Country 1	Argentina
Country 2	Canada
Country 3	Estonia
Country 4	Japan
Country 5	Spain

After the data is received, the part that cannot be used should be removed. It can be understood that a quality data set is obtained, since this unusable data is below 1%. In the data that could not be used, there were cases where one or more of the above-mentioned fields (gender, sales amount, member status, country, age) were empty, the entry was not made in accordance with the data standard, there was an outlier that could not be used, for example, the age was entered as 125. The data in this situation were manually extracted before the study for the health of the study. The dataset detailed below was prepared by following certain steps for the study. In order to try machine learning methods, the data must be made in a way that the model can understand. The available data includes information about the sales that were made in that year individually. Since this information will not be meaningful for the computer in its raw form, the following calculations were made from the dataset.

Total number of purchases according to gender for every year:

Year / Label	# Male Shopping	# Female Shopping	# Unknown Gender Shopping
2017	2,353,560	1,185,502	912,151
2018	2,538,950	1,237,520	947,491
2019	2,389,368	1,203,646	929,158
2020	2,200,165	1,074,575	811,758
2021	2,581,211	1,299,634	1,003,634

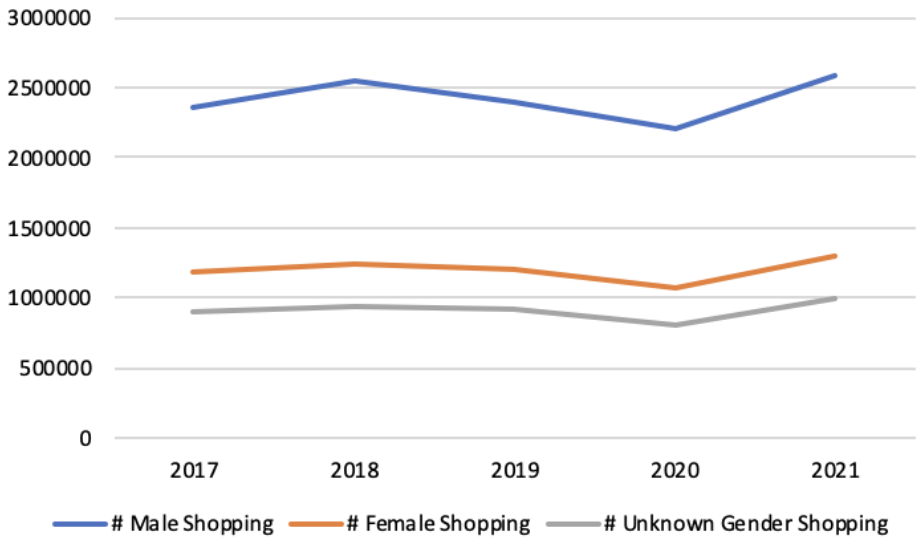

Figure 2: Sales Data according to genders for 5 years

The gender-segmented shopping participation trends from 2017 to 2021 are illustrated in the line chart. The blue line represents the male demographic, which shows a little increasing trend and may indicate a gradual increase in Black Friday sales activity. On the other hand, the orange line represents the female demographic, which stays flat over time, suggesting a steady participation rate. The group of unknown gender, represented by the gray line, exhibits the lowest level of interaction throughout time. It shows a modest increase in 2019 before leveling after that. According to the research, there may be differences in gender-based shopping habits during Black Friday sales. These differences could be a sign of underlying patterns in consumer behavior and could help anticipate future Black Friday sales trends.

# of Sold Item According to Countries for every year:

	2017	2018	2019	2020	2021
Country / Label	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**
Country 1	444,806	369,035	301,174	244,235	323,994
Country 2	1,358,689	1,443,381	1,390,628	1,235,905	1,584,511
Country 3	617,405	722,914	707,842	678,162	850,996
Country 4	1,176,442	1,241618	1,230,347	1,155,435	1,209,944
Country 5	853,871	947,015	892,181	892,181	915,034

Based on a preliminary analysis of the data, a clear trend is apparent in Country 1, where sales of items consistently decline between 2017 and 2020, but noticeably climb in 2021. This volatility could be the result of COVID-19-related socioeconomic factors impacting purchase patterns, changes in customer preferences, market strategies, or economic variances.

10.5824/ajite.2024.03.001.x

Comparatively, Country 2 exhibits a general upward trend that peaks in 2021. The trends in Countries 3, 4, and 5 are different. Country 3 is gradually increasing, Country 4 is exhibiting a tiny fall that eventually stabilizes, and Country 5 is displaying a slight upward trend that fluctuates.

Looking at things more broadly, Country 2 routinely does better than the other nations in terms of volume sold. During the course of the five-year period from 2017 to 2021, the bar graph compares the annual sales volume across five nations, designated as Country 1 through Country 5. The horizontal axis shows the time sequence by year, and the vertical axis counts the quantity of products sold. The graph shows that, with a peak in 2021, Country 2 continuously reports the greatest sales volume. Sales in Country 1 show a downward trend through 2020, then a rebound in 2021. Country 5 has a consistent increase, while Countries 3 and 4 show a mixed trend with variations in sales volume.

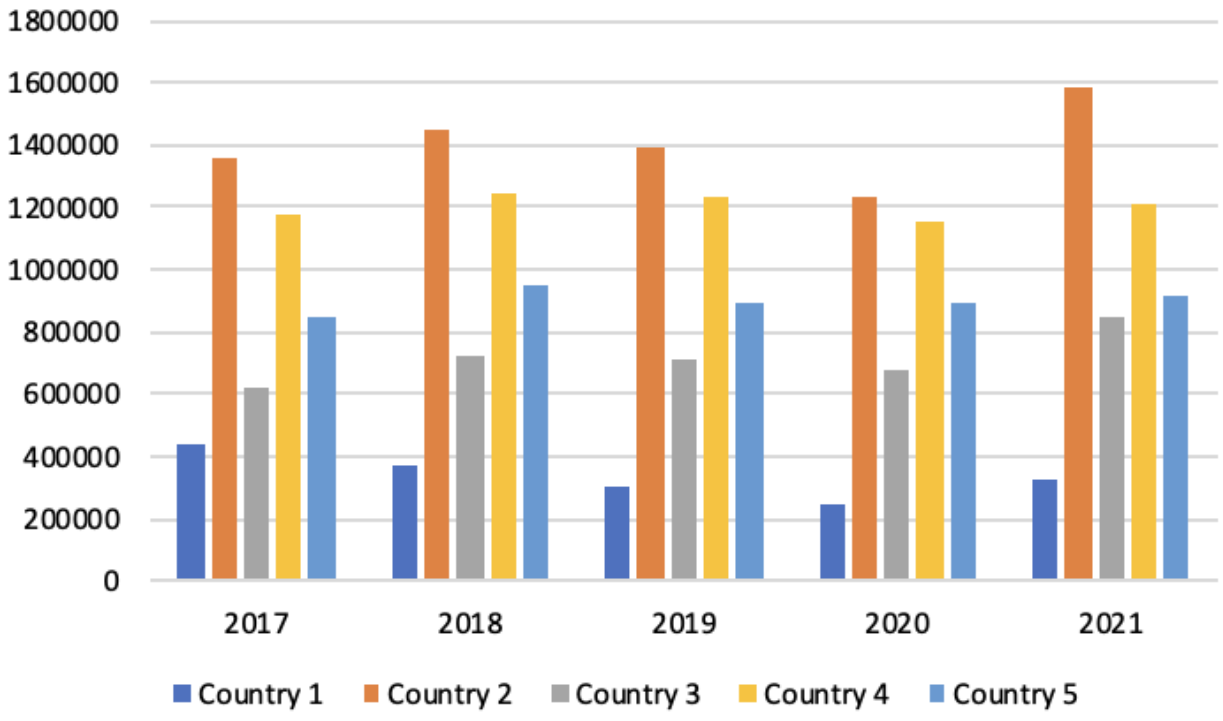

Figure 3: # of Sold Item According to Countries

# of Sold Item According to Age Group:

	2017	2018	2019	2020	2021
Age	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**
18-30	541,824	574,681	548,868	496,344	594,217
31-55	835,061	885,751	847,379	766,253	916,612
55<	3,074,328	3,263,531	3,125,925	2,823,901	3,373,650

Sales for the years 2017 through 2021 are broken down by age groupings in the statistics presented throughout the tables. A pattern appears showing that the age group of 55 and older routinely sells the most things, with quantities far exceeding those of the younger demographics. This pattern can indicate that seniors have more purchasing power or are generally more interested in in-person transactions. Significantly, sales in this age group peak in 2021, which may be a sign of rising consumer confidence or a strengthening economy following recessions like the COVID-19 pandemic. On the other hand, the sales volume for the 18–30 age group fluctuates, although in 2021 it shows an increasing tendency after declining in 2020. Over the course of five years, sales in the 31–55 age group have steadily increased, indicating a growing trend of on-site shopping, which may be attributed to middle-aged consumers' consistent income levels and household needs.

# of Sold Item According to Membership:

	2017	2018	2019	2020	2021
Is Elite	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**	**# of Sold Item**
No	3,067,856	3,281,253	3,115,816	2,834,004	3,367,551
Yes	1,383,357	1,442,710	1,406,356	1,252,494	1,516,928

Sales data divided by membership status over a five-year period, from 2017 to 2021, is included in the tables. Two categories—'Elite' members and non-members ('No')—are used to dichotomize the data. The bulk of things sold annually are constantly accounted for by non-members, according to a review of the numbers. Nonetheless, in 2021, there is a discernible rise in the number of goods sold to "Elite" members, indicating a growing trend in sales for this demographic.

- Country 1: According to the data, Black Friday might not have a significant impact on sales in this nation.
- Country 2: The information suggests that Black Friday had little effect on sales here.

AJIT-e Academic Journal of Information Technology
2024 Summer/Yaz – Cilt/Vol: 15 - Sayı/Issue: 3
10.5824/ajite.2024.03.001.x

- Country 3: According to the data, Black Friday might have a longer-lasting impact here.
- Country 4: The data indicates that the impact of Black Friday was not felt right away but rather developed over time.
- Country 5: The data suggests that the COVID-19 outbreak may have had an impact on the delayed Black Friday impact.
- In every year, the number of male customers has outnumbered the number of female customers. This may be indicating that this retail store is mainly for household goods. Families may shop together.
- Elite members are also outnumbered in every year by non-elite members.
- Younger customer number are going higher in every year but +55 is the most common in every year.
- Countries have made consistent product sales every year compared to the previous year. There is nothing unexpected.
- When looking at the purchase by year graph, the increase and decrease in sales in all 3 categories (Male, Female, Un-known) coincide with the same time. It may have experienced a decrease due to 2020 Covid.

3.2. Model Design

The dataset contains the following columns; date of the sale, country, age group, the customer membership status, gender of the customer, number of sold items, and the year of the sale. Two distinct approaches have been chosen for this project to be applied to the dataset. Using the dataset, the first step is to generate a 2-dimensional array. This approach predicts future Black Friday sales based just on the year and the quantity of sold items. For this method, the first dataset is split into the years. After that, every year is split into weeks.

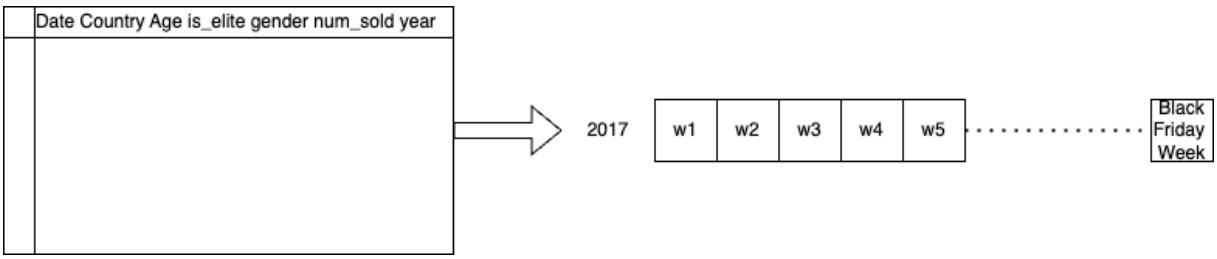

Figure 0: Representation of the data cell creation

This data cell contains the year, week number, and the total of the sold items.

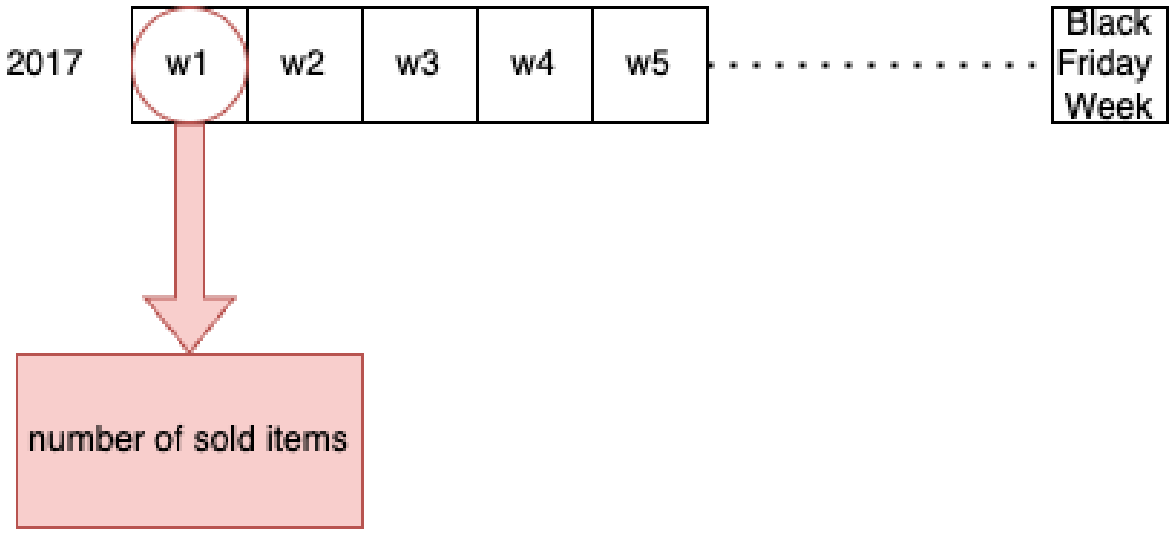

Figure 5: Representation of the data cell content

This method is repeatedly done until the whole dataset is split into data cells that contain the year, week number and the number of sold items of that year.

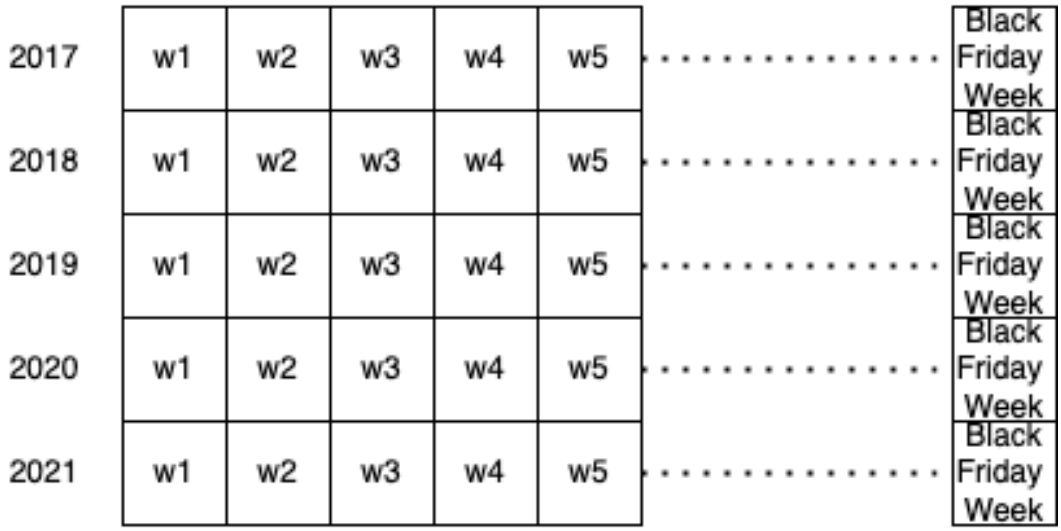

Figure 6: Representation of the whole dataset as data cells

In the second method, unlike the first one, 3-dimensional arrays are created. In this version, again data frame is created for each week of each year but, the data cell of that week contains the number of women/men who buy items, the number of elite members, the country, and the number of total items sold. To represent these data fields in one cell a 3-dimensional array is needed.

AJIT-e Academic Journal of Information Technology
2024 Summer/Yaz – Cilt/Vol: 15 - Sayı/Issue: 3
10.5824/ajite.2024.03.001.x

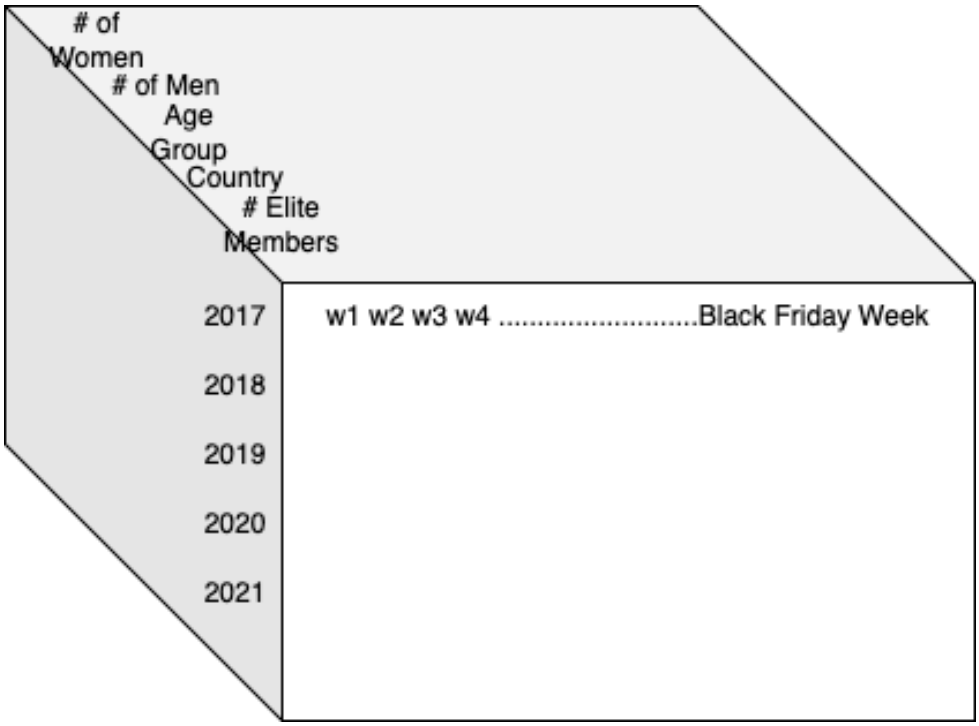

Figure 7: Representation of 3-dimensional data frame

In this data frame, all the sales data are connected to each other to use in the prediction of the Black Friday week total sales. In every data cell, all the columns that the data set contains are divided into that year's that week.

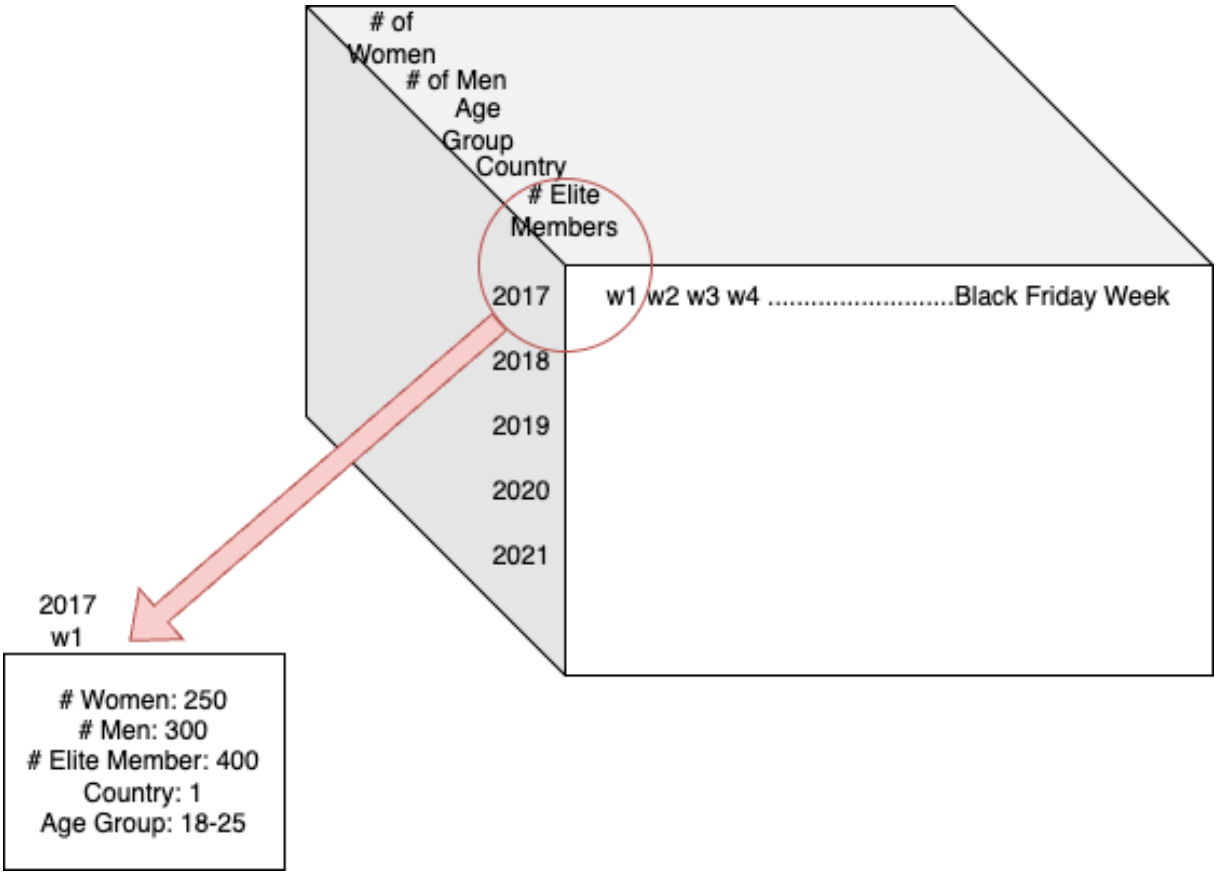

Figure 8: Representation of a data cell from the data frame

After the preparation part is done for the 2 methods, the training part begins. In the training part, different kinds of methods are used with different kinds of parameters. According to best practice, the following methods are used.

The objective was to identify trends and connections in the data and develop models that, using demographic information, could forecast sales for upcoming Black Friday events. Every one of these methods has advantages and disadvantages, and the way they work will depend on the particulars of the dataset and the issue that trying to be solved.

10.5824/ajite.2024.03.001.x

TESTS & RESULTS

Algorithm	Results (%)
logistic_regression	93,20425024
decision_tree	93,48748033
svm_kernel	92,31470579
light_gbm	88,84952433
XGBoost	92,47751064
LSTM	0,013393949

The error rates were compared after training the data with each model. The table of performance metrics for different machine learning algorithms used on a regression job is the data that is presented. The accuracy percentage in Mean Absolute Error is used to quantify the performance of the methods.

Using time series data, the methodology used includes an organized approach to model training and validation. Weekly sales information for every year under consideration was included in the dataset. Data from the first 45 weeks of each year, which captured the sales patterns leading up to Black Friday but excluded the week of Black Friday itself, were used for model training. By using this method, the model was able to learn from sales patterns, consumer behavior, and seasonal trends that preceded the event. The next objective given to the predictive model was to anticipate sales for the 46th week of the year, which is Black Friday week. The purpose of this prediction challenge was to evaluate the model's capacity to anticipate sales volume at a peak retail period by using patterns and trends from previous years' earlier parts of the year.

An assessment of the model's accuracy and predictive performance was carried out by contrasting the anticipated sales for the 46th week with the actual sales data for the same time. Notably, during the training phase, the model was never exposed to data from the years 2022 and 2023. This method simulated a real-world situation where future sales data would not be available at the time of prediction, allowing for a thorough assessment of the model's predictive ability on completely unseen data. A forward-testing set consisting of the years 2022 and 2023 was used to evaluate the model's generalization abilities and possible application in a real-world retail forecasting setting.

The table includes a deep learning model called LSTM (Long Short-Term Memory) in addition to more conventional machine learning models like logistic regression, decision trees, Support Vector Machine (SVM) with kernel, LightGBM, and XGBoost.

With scores of roughly 93.2% and 93.5%, respectively, the results demonstrate that the logistic regression and decision tree models performed similarly and had the greatest accuracy among the investigated methods. 92.3% accuracy is a little bit below the SVM with kernel. Both the gradient boosting models LightGBM and XGBoost perform worse than the previous three models, with XGBoost scoring 92.5% and LightGBM scoring 88.8%. But compared to the other models, the LSTM model has a significantly lower accuracy of about 0.013%, which is several orders of magnitude lower. An LSTM is often effective for sequential data and can attain high accuracy in various jobs; this kind of outcome is quite rare for an LSTM.

The comparison's findings for the conventional machine learning methods are very amazing, especially when it comes to the decision tree and logistic regression models, which show high accuracy rates above 93%. Given that logistic regression performs well in these situations, this degree of performance shows that the underlying patterns in the dataset are well-captured by these models, which may hint that the feature space is somewhat linearly separable.

Excellent accuracy is also demonstrated by the support vector machine (SVM) with a kernel function, indicating that the data has a complicated structure that makes use of the high-dimensional feature spaces in which SVMs are particularly effective. SVM can locate a hyperplane in the modified feature space that maximally divides the classes thanks to the kernel trick.

The gradient boosting machines, LightGBM and XGBoost, perform well despite having a slightly poorer accuracy than the previously stated models. These models are effective because they progressively construct an ensemble of weak learners (usually decision trees), with each new model aiming to rectify the mistakes of the previous ones. Significantly, LightGBM and XGBoost are both renowned for their efficacy and efficiency in managing big datasets and high-dimensional feature spaces; however, XGBoost is especially commended for its ability to prevent overfitting through regularization. Their slightly lower performance in this instance may reflect the features of the dataset or imply that more parameter tweaking may produce better outcomes.

All things considered, the effectiveness of these conventional machine learning techniques emphasizes how crucial it is to comprehend the properties of the dataset and select the appropriate model in accordance with them. It also highlights how important feature engineering, appropriate preprocessing, and hyperparameter optimization are to obtaining good forecast accuracy.

The following can be said about the project's outcome: critical low results were obtained since LSTM was unable to investigate enough of the data set due to its small size. On the other hand, traditional machine learning models produced good outcomes because, in contrast to deep learning, machine learning can operate with minimal amounts of data. It is

10.5824/ajite.2024.03.001.x

possible to conclude from the project's outcomes that machine learning models have practical applications in sales estimation. People can estimate sales to establish plans before there is any business knowledge. Businesses can gain from these technologies, but sales prediction and optimization are at important times, especially during Black Friday.

It is evident that there is interest in Black Friday, even though it is less than what was first anticipated in the study's data set. The similar approach can be attempted using data from the USA since Black Friday is a recognized shopping holiday there.

CONCLUSION

The project's final product is a comprehensive investigation of the complex relationships surrounding Black Friday sales and the significant influence of artificial intelligence (AI) on predictive analytics. This research has not only illuminated important relationships between demographic characteristics and consumer buying patterns during the Black Friday shopping frenzy, but it has also used artificial intelligence (AI) to predict future trends in sales.

One of the most important findings of this research is that Black Friday has a distinct effect on physical stores than it does on online retailers. Unlike what was anticipated, it turned out that the Black Friday seismic waves had a greater impact on e-commerce, changing the way consumers behaved and shifting sales to online channels. This discovery emphasizes how crucial it is to modify sales tactics to fit this changing environment, in which digital media becomes the focal point of customer interaction.

It is outside the purview of this investigation, but it is impossible to overestimate how disruptive the COVID-19 pandemic has been to sales trends. As a result of the pandemic, consumer preferences changed and the trend toward internet buying accelerated. This extraordinary worldwide incident served as a sobering reminder that, as they negotiate the shifting landscape of Black Friday sales, businesses must continue to be flexible and adaptive in the face of unanticipated setbacks.

This project has shown that a variety of machine learning models, such as Decision Trees, XG Boost, SVM with RBF Kernel, Linear Regression, and Light GBM, are effective at forecasting future sales trends. Equipped with past data and demographic insights, these models have the capacity to transform sales tactics. Businesses may improve customer experiences, optimize marketing efforts, and make data-driven choices by utilizing AI's predictive powers.

The most important factor affecting the margin of error in this study was the size of the data set. Big Data is needed as much as possible in order to do machine learning and

forecasting studies. In this study, a data set that cannot be classified as Big Data was used. For this reason, the error rates were calculated as acceptable and perfectible.

It is imperative to recognize that the margin of error was significantly influenced by the size of the dataset used in this investigation. Even though the dataset is large, it might not be considered "Big Data," but the computed error rates were still reasonable and might be improved. This supports the idea that businesses will likely see a considerable boost in the accuracy and predictive power of AI models as they gather larger datasets.

With the data set in the study, different systems prepared with the Black Friday sales of the past 5 years were trained and tested with a part that was separated from the same data set and was never shown to the system in training. Error rates were calculated by this method. The non-Big Data set included in this study obtained results with acceptable error rates. This study proves that large companies can predict Black Friday sales with their Big Data. Intelligent systems that continue learning and make instant predictions for each period can be realized not only for a certain period but also with continuously incoming data. Thus, companies that will integrate this work into themselves can have an accurate foresight for future periods, as well as make cost-saving plans. This can affect their growth plan, budget plan, and any other plan they wanted to forecast.

For future studies, an end-to-end system that can work not only with static data but also with live data can be designed. Thus, it can be tested how realistic such methodologies will help in the daily live data flow. The systems trained with Big Data can be integrated into a real sales application and can be retrained with each new incoming data and instantly generate individual forecasts for each customer who has not made a purchase yet at that moment. The potential for real-time personalized sales projections for consumers who haven't made a purchase is present when AI models trained on "Big Data" are included in useful sales apps. This rigorously studied research has the potential to significantly improve everyday business operations and customer experiences by pointing the way toward increased strategic advantage, flexibility, and foresight. Thus, the value these academically proven studies will add to daily life can be measured.

10.5824/ajite.2024.03.001.x

REFERENCES

BlackFriday.com. (2020). *Black Friday history and statistics*. https://blackfriday.com/news/black-friday-history

Eberhard, E. (2018, December). *Dashing to the stores: Last-minute holiday trends to watch*. Think with Google. https://www.thinkwithgoogle.com/consumer-insights/consumer-trends/last-minute-holiday-shopping-trends/

Hand, R. (2024, February 24). *What is demand forecasting? Importance and benefits of forecasting customer demand*. ShipBob. https://www.shipbob.com/blog/demand-forecasting/

Kaplan, M. (2018, November 28). *Sales report: 2018 Thanksgiving, Black Friday, Cyber Monday*. Practical Ecommerce. https://www.practicalecommerce.com/sales-report-2018-thanksgiving-black-friday-cyber-monday

Kayapınar Kaya, S., & Yıldırım, Ö. (2020). Prediction model for automobile sales in Turkey using deep neural networks. *Journal of Industrial Engineering*, 57-74.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 436–444.

Liu, J., Liu, C., Xu, Y., & Zhang, L. (2020). Research on sales information prediction system of e-commerce enterprises based on time series model. *Information Systems and e-Business Management*, 823–836.

Murray, P. N. (2014, June 11). *Why we really shop*. Psychology Today. https://www.psychologytoday.com/us/blog/inside-the-consumer-mind/201406/why-we-really-shop