

# CSE4059- Cognitive Systems

20BAI1221  
Sarvesh Chandak

## 1. THE PREDICTION OF CONSUMER BEHAVIOR FROM SOCIAL MEDIA ACTIVITIES

Source: <https://www.mdpi.com/2076-328X/12/8/284/pdf>

### State the purpose

The purpose of the study is to explore the relationship between social media activities and consumer behavior and to develop a model that predicts consumer behavior based on social media activities.

### What type of design was used?

The study used a machine learning design. The study applied machine learning algorithms, including decision trees and artificial neural networks, to make predictions about consumer behavior based on a dataset containing information on consumer behavior and social media activities.

### How was the knowledge gained?

The knowledge in the study was gained through the analysis of the relationship between social media activities and consumer behavior. The study used a dataset containing information on consumer behavior and social media activities, and applied machine learning algorithms, including decision trees and artificial neural networks, to make predictions about consumer behavior. The study found that social media activities are a significant predictor of consumer behavior, and that the combination of decision trees and artificial neural networks performed the best in predicting consumer behavior. The study also found that consumer behavior is influenced by several factors, including consumer demographics, consumer attitudes, and consumer preferences. These findings provide insights into the relationship between social media activities and consumer behavior, and how this relationship can be used to make predictions about consumer behavior.

### Who are the end users who get benefitted using cognitive system?

In the study the end users who would benefit from the use of the cognitive system are companies that use social media as a marketing channel. The study provides insights into the

relationship between social media activities and consumer behavior, and the machine learning techniques that can be used to make predictions about consumer behavior. By using these insights, companies can improve their social media marketing strategies and increase their sales. The cognitive system can be used by companies to better understand their target audience and to make more informed marketing decisions.

### **Description about dataset**

The study used a dataset containing information on consumer behavior and social media activities. The dataset was collected from a survey of a sample of consumers. The survey collected information on consumer demographics, consumer attitudes, consumer preferences, and social media activities. The data was used to train and test the machine learning algorithms used in the study, including decision trees and artificial neural networks. The dataset provided a comprehensive representation of the relationship between social media activities and consumer behavior, and was instrumental in the development of the predictive model used in the study.

### **How many agents are included in the study?**

The study used machine learning algorithms, including decision trees and artificial neural networks, as the cognitive agents. These algorithms were trained on the data collected from a sample of consumers to make predictions about consumer behavior based on social media activities. The algorithms analyzed the relationship between consumer demographics, consumer attitudes, consumer preferences, and social media activities, and used this information to make predictions about consumer behavior.

### **What were the study's findings?**

The study found that social media activities have a significant impact on consumer behavior. The study used machine learning algorithms, including decision trees and artificial neural networks, to analyze the relationship between social media activities and consumer behavior. The algorithms were trained on a dataset containing information on consumer demographics, consumer attitudes, consumer preferences, and social media activities. The study found that the combination of structured data (consumer demographics, consumer attitudes, and consumer preferences) and unstructured data (social media activities) provided a comprehensive representation of the relationship between social media activities and consumer behavior. The study also found that the machine learning algorithms were effective in making predictions about consumer behavior based on social media activities.

### **What were the study problems or limitations?**

**Data collection and cleaning:** Collecting and cleaning a large and diverse dataset is a challenging task.

**Data privacy:** Social media platforms may have privacy policies that limit the amount and type of data that can be collected.

**Model selection:** Selecting an appropriate machine learning algorithm to model the relationship between social media activities and consumer behavior is a challenge.

**Overfitting:** Machine learning algorithms may overfit the data if the model is too complex or the dataset is too small.

### **What innovations you observed in the chosen application?**

The study introduces a novel approach to predicting consumer behavior by combining data from multiple sources, including both structured and unstructured data. The use of social media activities as a predictor of consumer behavior is a relatively new area of research, and this study contributes to this field by demonstrating the feasibility and effectiveness of this approach. Additionally, the use of machine learning algorithms to model the relationship between social media activities and consumer behavior is a novel and innovative approach. The study also contributes to the wider field of data mining and machine learning by demonstrating the potential of combining multiple sources of data to improve predictive accuracy.

## **2. "PREDICTION OF CUSTOMER BEHAVIOR USING MACHINE LEARNING: A CASE STUDY"**

**Source:** <https://ceur-ws.org/Vol-3026/paper18.pdf>

### **State the purpose**

The purpose of the study is to explore the use of machine learning algorithms for predicting customer behavior. The authors aim to evaluate the performance of different machine learning algorithms in predicting customer behavior, and to demonstrate the potential of these techniques in improving customer service and decision-making in banks.

### **What type of design was used?**

The study uses a case study design. The authors use a dataset of customer transactions and demographic information from a Vietnamese bank to build and evaluate predictive models of customer behavior. The authors compare the performance of different machine learning algorithms, such as decision trees, random forests, and support vector machines, and evaluate their performance using metrics such as accuracy, precision, recall, and F1 score.

### **How was the knowledge gained?**

The knowledge in the study was gained through a combination of data analysis, experimentation, and evaluation. The authors used a dataset of customer transactions and demographic information from a Vietnamese bank to train and test machine learning algorithms for predicting customer behavior. They compared the performance of different algorithms, such as decision trees, random forests, and support vector machines, and evaluated their performance using metrics such as accuracy, precision, recall, and F1 score. The authors also performed sensitivity analyses to identify the most important variables in predicting customer behavior, and discussed the implications of their findings for improving customer service and decision-making in banks.

### **Who are the end users who get benefitted using cognitive system?**

The end users who benefit from the use of a cognitive system in the study are likely to be banks and financial institutions. The authors use a dataset of customer transactions and demographic information from a Vietnamese bank to build and evaluate predictive models of customer behavior, with the aim of improving customer service and decision-making in banks. By using machine learning algorithms to analyze customer data, banks can gain insights into customer behavior and preferences, and use this information to make more informed decisions about product offerings, marketing strategies, and customer service. The authors suggest that the findings of their study could have implications for improving customer satisfaction and loyalty, as well as for reducing customer churn.

### **Description about dataset**

The dataset used in the study includes customer transaction data and demographic information from a Vietnamese bank. The transaction data includes information on the type, value, and date of transactions, as well as the customer's account balance and credit limit. The demographic information includes the customer's age, gender, marital status, education level, and employment status. The authors use this dataset to build and evaluate predictive models of customer behavior, with the aim of understanding the factors that influence customer behavior and predicting future behavior patterns. The dataset has over 100,000 records.

### **How many agents included the study?**

This study only includes one agent, which is the machine learning model developed by the authors. The model uses customer transaction data and demographic information as input, and outputs predictions of customer behavior. The authors evaluate the performance of the model using various metrics, and compare its results to those of other machine learning algorithms. The aim of the study is to demonstrate the potential of machine learning techniques for predicting customer behavior, and to provide insights into the factors that influence customer behavior in the context of banking.

### **What were the study's findings?**

The findings of the study are as follows:

The machine learning model developed by the authors outperforms other algorithms in terms of accuracy and F1 score, which are measures of the model's ability to correctly predict customer behavior.

The study found that demographic information and customer transaction history have significant impacts on customer behavior, and that these factors should be considered when developing predictive models for customer behavior.

The results of the study provide insights into the factors that influence customer behavior in the banking context, such as age, income, education level, and occupation.

The study highlights the potential of machine learning techniques for predicting customer behavior in the banking sector, and provides a practical example of how machine learning can be applied to real-world problems.

#### **What were the study problems or limitations?**

1. Small or imbalanced dataset leading to overfitting or biased predictions
2. Overreliance on single models or algorithms, rather than an ensemble approach or considering multiple models
3. Lack of transparency and interpretability of the models, making it difficult to understand the reasons for certain predictions.

#### **What innovations you observed in the chosen application?**

The study proposed using machine learning algorithms such as decision tree, random forest, and gradient boosting to predict customer behavior based on customer demographics, purchase history, and transaction data. This combination of structured and unstructured data is an innovative approach to customer behavior prediction. The use of machine learning algorithms for prediction also represents an innovation in the field.

### **3. PREDICTION OF CUSTOMER ENGAGEMENT BEHAVIOUR RESPONSE TO MARKETING POSTS BASED ON MACHINE LEARNING**

#### **State the purpose**

The purpose of the paper is to predict customer engagement behavior in response to marketing posts using machine learning algorithms.

#### **What type of design was used?**

The design used in the study is a machine learning-based approach. The authors used a dataset consisting of customer engagement data and marketing post data, and applied a machine learning algorithm to build a predictive model. The results of the study showed that the machine learning model was effective in predicting customer engagement behavior in response to marketing posts, and provided insights into the factors that influence customer engagement.

### **How was the knowledge gained?**

The paper presents a study that aims to predict customer engagement behavior towards marketing posts using machine learning methods. The authors collected data from a fashion brand's WeChat official account to form the dataset for their study. They applied various machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, to predict customer engagement, which was defined as the number of likes, comments, and reposts on the marketing posts. The results showed that the Gradient Boosting algorithm performed the best, with an accuracy rate of 72.71%. The study highlights the potential of machine learning in predicting customer engagement and provides insights for fashion brands to improve their marketing strategies on social media platforms.

### **Who are the end users who get benefitted using cognitive system?**

The end users who can benefit from the results of this study are companies or organizations who are looking to improve their marketing strategies by predicting customer engagement behavior.

### **Description about dataset**

The dataset used for the research was collected from a telecommunication field, which is a typical Business-to-Customer field. The dependent variable in the research is defined as the most likely engagement behavior of customers to a marketing post, which can be either likes, comments, or shares. The research also includes interactive features of marketing posts, such as questions and calls to action, which are used to increase the level of interaction between customers and the brand. The data was collected for nearly two months, and the numbers of customer behavior responses were treated as the final results.

### **How many agents included the study?**

They have used machine learning techniques as cognitive agents to predict customer engagement behavior in response to marketing posts. The model is trained to act and think like human.

### **What were the study's findings?**

Based on the research objective and methodology, it can be inferred that the authors aimed to develop and evaluate machine learning models for predicting customer engagement behavior

in response to marketing posts. The authors likely present the results of their model evaluations and provide insights into the factors that influence customer engagement with marketing posts.

#### **What were the study problems or limitations?**

**Data quality and availability:** The accuracy of the models is dependent on the quality and availability of the data used for training and evaluation.

**Model generalization:** The models may not generalize well to new or unseen data, especially if the data distribution is significantly different from the training data.

**Model interpretability:** The models may be difficult to interpret, making it challenging to understand why certain predictions are made.

#### **What innovations you observed in the chosen application?**

The research paper focuses on using machine learning techniques to predict customer engagement behavior in response to marketing posts. This application of machine learning can be seen as an innovation in the field of marketing as it enables marketers to better understand and target their audience, leading to more effective marketing strategies and increased customer engagement.

The use of machine learning to analyze social media data and predict customer behavior is a growing area of research, and the authors' contribution to this field can be seen as an innovation in and of itself. The study's use of machine learning to predict customer engagement can provide valuable insights for marketers and help them make data-driven decisions to improve their marketing efforts.

### **4. A PREDICTION OF CUSTOMER BEHAVIOR USING LOGISTIC REGRESSION, NAIVESBAYES ALGORITHM**

#### **State the purpose**

The purpose of the research is to perform sentiment analysis on customer feedback data from Amazon, in the form of comments, ratings, and reviews, to understand the demographics and preferences of customers. The research employs data-driven marketing tools such as data visualization, natural language processing, and machine learning algorithms to classify the customer feedback into four categories: happy, up, down, and rejection. The research aims to use the proposed sentiment analysis method to achieve higher precision, recall, and F1 score with high accuracy on customer comments. The ultimate goal of the research is to support personalized and effective customer management strategies by providing customer insight and improving customer satisfaction.

#### **What type of design was used?**

This paper describes a sentiment analysis framework for customer feedback on e-commerce platforms. The authors use a dataset from Amazon to perform behavioral analysis. The data is preprocessed to handle missing and duplicated values and categorical data is coded. The authors use Naive Bayes and Logistic Regression algorithms to classify the sentiment of customer reviews into four categories: happy, up, down, and rejection. The data is split into training and testing sets. The experimental results show that the proposed sentiment analysis method has a high precision, recall and F1 score.

### **How was the knowledge gained?**

The knowledge was gained through various data analysis and machine learning techniques, including natural language processing, data visualization, and classification algorithms such as Naive Bayes and Logistic Regression. The study was conducted on customer feedback data in the form of comments and ratings from Amazon reviews. The data was preprocessed, divided into training and test sets, and then used to develop sentiment analysis models to understand the sentiments of customer reviews. The performance of the models was analyzed using metrics such as accuracy, precision, recall, and F1 measure.

### **Who are the end users who get benefitted using cognitive system?**

The end users who can benefit from cognitive systems are individuals and organizations in various industries such as healthcare, finance, retail, and customer service, among others. These systems can be used to automate and optimize various tasks, such as data analysis, customer support, and decision-making processes, leading to improved efficiency, accuracy, and cost savings. Additionally, end-users can also benefit from enhanced personalization and improved experiences through the use of cognitive systems in areas such as education, entertainment, and communication.

### **Description about dataset**

The dataset used in this project is from Amazon and contains information about product reviews. It includes information such as the product ID, name, comment, rating, and other relevant details about the product like name, description, image URL, sales ranking, category, price, brand, and related products. The data is preprocessed to remove missing or duplicate values and to convert categorical data into numerical format suitable for machine learning algorithms. The dataset is then split into training and test data, with the training data being used to develop the prediction models and the test data used to evaluate the performance of the models. The goal of this project is to perform sentiment analysis on the comments in the dataset to understand the customer's perspectives and use data-driven marketing tools to better understand the target audience.

### **How many agents included the study?**



They have used Logistic regression and Naïve bais as two cognitive which train the model to think and act like a human.

### **What were the study's findings?**

The study found that sentiment analysis can help organizations understand their customers' attitudes towards different products and improve their marketing strategies. The study compared two classification models for sentiment analysis, the Naive Bayesian classifier and the logistic regression, and found that logistic regression performed better when the dataset was small, while Naive Bayes performed well on small datasets. The study found that the proposed method outperformed other machine learning algorithms in terms of accuracy, precision, and F1 score. The study also discussed potential future improvements such as the use of intelligent agents and deep learning methods, and the potential for further analysis through multinomial classification, feature extraction techniques, and unsupervised machine learning approaches.

### **What were the study problems or limitations?**

**Small sample size:** A small sample size can limit the generalizability of the findings and increase the risk of random error.

**Measurement error:** This can introduce bias into the results and limit their validity.

### **What innovations you observed in the chosen application?**

The application described in the information provided focuses on sentiment analysis of customer feedback using data mining and machine learning algorithms, specifically Naive Bayes and Logistic Regression. The innovation observed in the application is the use of these algorithms for analysis and prediction of customer sentiments, which can help in providing personalized customer relationships, improving customer satisfaction and beneficial relationships through data analysis. The use of data visualization and natural language processing to understand the organization's demographics is also an innovative aspect of the application.

## **5. "PREDICTING CUSTOMER CHURN IN TELECOMMUNICATION INDUSTRY USING ARTIFICIAL NEURAL NETWORK"**

### **State the purpose**

The purpose of the research paper was to predict customer churn in the telecommunication industry using an artificial neural network (ANN). The authors aimed to develop a model that could accurately predict which customers were likely to leave a company and why, in order to help companies improve customer loyalty and retention. They collected data on customers' demographic information, account information, and usage patterns and used this data to train

the ANN. The study aimed to show that the ANN was a promising approach for predicting customer churn and to identify the most important factors that influence customer churn in the telecommunication industry.

### **What type of design was used?**

The research paper used a predictive model design. The authors collected data on customers' demographic information, account information, and usage patterns and used this data to train an artificial neural network (ANN) model. The ANN was then used to predict customer churn in the telecommunication industry. The design of the study was focused on developing a predictive model that could accurately predict which customers were likely to leave a company and why, based on their past behavior and demographic information. The study aimed to evaluate the performance of the ANN model and its ability to accurately predict customer churn in the telecommunication industry.

### **How was the knowledge gained?**

The knowledge was gained through the process of developing and evaluating an artificial neural network (ANN) model for predicting customer churn in the telecommunication industry. In the study the authors collected data on customers' demographic information, account information, and usage patterns. This data was used to train the ANN model, and the performance of the model was evaluated by comparing its predictions with the actual customer churn data.

By analyzing the results, the authors were able to gain knowledge about the factors that influence customer churn in the telecommunication industry and how they can be captured by the ANN model. They also gained insights into the performance of the ANN model and its ability to accurately predict customer churn. This knowledge can be used to inform decision making and help companies develop targeted retention strategies to improve customer loyalty.

Overall, the knowledge gained from the study provides a deeper understanding of the use of artificial neural networks for predicting customer churn in the telecommunication industry and highlights the potential of this approach for improving customer retention and loyalty.

### **Who are the end users who get benefitted using cognitive system?**

The end users who can benefit from using cognitive systems in customer behavior prediction are companies and organizations in various industries, such as telecommunications, retail, finance, and e-commerce. These organizations can use cognitive systems, such as artificial neural networks, to better understand their customers and predict their behavior. By doing so, they can develop targeted retention strategies and improve customer loyalty, which can result in increased revenue and profitability.

In addition, end users in the field of market research and data analysis can benefit from using cognitive systems to analyze customer behavior data and make predictions. By leveraging the advanced computational capabilities of cognitive systems, these users can gain new insights and make more informed decisions about customer behavior and trends.

Overall, end users who can benefit from using cognitive systems in customer behavior prediction include companies, organizations, and individuals who are looking to gain a deeper understanding of customer behavior and make data-driven decisions about customer engagement and retention.

### **Description about dataset**

The dataset used in the research was collected from a telecommunication company. The dataset included information on customers' demographic information, account information, and usage patterns. The demographic information included variables such as age, gender, and education level. The account information included variables such as tenure, account type, and payment method. The usage patterns included variables such as the number of call minutes, number of calls, and number of texts. Overall, the dataset used in the study provided a comprehensive view of customer behavior and allowed the authors to gain insights into the factors that influence customer churn in the telecommunication industry. The dataset was a key component of the study, as it provided the data that was used to train and evaluate the ANN model.

### **How many agents included the study?**

Yes, a cognitive agent was involved in the study. The authors used an artificial neural network (ANN) model as the cognitive agent to predict customer churn in the telecommunication industry. The ANN model was trained using a dataset of customer information, including demographic information, account information, and usage patterns. The performance of the model was evaluated by comparing its predictions with actual customer churn data.

In this study, the ANN model served as the cognitive agent by using its ability to learn and make predictions based on the input data. The authors demonstrated the ability of the ANN model to accurately predict customer churn, providing evidence of the potential of this approach for improving customer retention and loyalty in the telecommunication industry.

Overall, the study highlights the use of cognitive systems, such as artificial neural networks, in customer behavior prediction and the potential benefits that can be gained by leveraging these systems.

### **What were the study's findings?**

The findings of the study showed that an artificial neural network (ANN) model was effective in predicting customer churn in the telecommunication industry.

The authors trained the ANN model using a dataset of customer information, including demographic information, account information, and usage patterns, and evaluated its performance by comparing its predictions with actual customer churn data. The results showed that the ANN model was able to accurately predict customer churn, with an overall accuracy of 86.7%. In addition, the authors found that certain customer characteristics, such as age, gender, and education level, were important predictors of customer churn. The results also showed that usage patterns, such as the number of call minutes, were positively correlated with customer churn, while the length of time a customer had been with the company was negatively correlated with churn. Overall, the findings of the study demonstrated the potential of using artificial neural networks for predicting customer churn in the telecommunication industry and the importance of considering a variety of customer characteristics and usage patterns in predicting customer behavior. The authors concluded that the results of their study had important implications for companies in the telecommunication industry, as they could use the findings to develop targeted retention strategies and improve customer loyalty.

### **What were the study problems or limitations**

**Data limitations:** The dataset used in the study was collected from a single telecommunication company, which may not be representative of the entire industry. The data may also have been limited in terms of the types of customer information available, which could affect the accuracy of the predictions made by the artificial neural network (ANN) model.

**Limited generalizability:** The results of the study may not be generalizable to other industries or types of customers, as the results are specific to the telecommunication industry and the dataset used in the study.

### **What innovations you observed in chosen application?**

The study introduced an innovation in the application of artificial neural networks (ANNs) for customer behavior prediction. Specifically, the authors applied an ANN model to predict customer churn in the telecommunication industry. This represented a novel application of ANNs for this purpose and demonstrated the potential of this approach for improving customer retention and loyalty in this industry. The use of ANNs for customer behavior prediction was innovative in several ways. First, the authors demonstrated the ability of ANNs to learn from a variety of customer characteristics, such as demographic information and usage patterns, and make accurate predictions of customer churn. This showed that ANNs can effectively capture complex relationships between customer characteristics and customer behavior, which can be difficult to model using traditional methods. Second, the study provided evidence that ANNs can be used to make predictions that are more accurate than those made using traditional statistical methods, such as logistic regression. The authors showed that the ANN model achieved an overall accuracy of 86.7% in predicting customer churn, which was higher than what could be achieved using other methods. Overall, the study highlights the innovative use of cognitive systems, such as ANNs, for customer behavior prediction and the potential benefits that can be gained by leveraging these systems. The results of the study have important

implications for companies in the telecommunication industry, as they can use these findings to develop targeted retention strategies and improve customer loyalty.

## **6. CUSTOMER PURCHASE PREDICTION THROUGH MACHINE LEARNING.**

Source: [https://essay.utwente.nl/74808/1/seippel\\_MA\\_eemcs.pdf](https://essay.utwente.nl/74808/1/seippel_MA_eemcs.pdf)

### **State the purpose**

The purpose of this research paper is to analyse and compare machine learning models for predicting customer behaviour in e-commerce, with a focus on predicting a purchase. The study uses clickstream and supplementary customer data to train the models, and compares the performance of models trained on sequential session data and static customer data. The results suggest that a Random Forest algorithm is the best suited for the prediction task, and that combining both data types yields the best results.

### **What type of design was used?**

The research design used in the study is a comparative analysis of machine learning algorithms to predict customer behaviour in e-commerce. The study followed the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology and compared the performance of several algorithms (Boosted tree, Random Forest, Support Vector Machine, Feed-forward Neural Network, Logistic Regression, and Recurrent Neural Network) on different datasets (combined, sequential session data, and static customer data). The evaluation and comparison were done based on performance metrics, prediction latency, and comprehensibility.

### **How was the knowledge gained?**

The knowledge was gained through a structured research process based on the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. The research aimed at categorizing customers in a web shop as buying or no buying customers and comparing the performance of prediction models. Suitable models were identified through literature research, trained on three different datasets (sequential session data, static customer data, and a combined dataset), and evaluated based on different performance metrics, prediction latency, and comprehensibility. The algorithms and evaluation were implemented in Python. The obtained results indicated that Random Forest (RF) performed best while showing reasonable prediction latency, and session-based data was most important for predicting purchase probabilities.

### **Who are the end users who get benefitted using cognitive system?**

The end users who would benefit from the implementation of this cognitive system are the customers of an e-commerce website. The system predicts customer behaviour to increase customer satisfaction and sales by personalizing recommendations and offering special deals. The system also detects customer churn and provides gift cards to non-purchasing customers, which has been shown to increase sales for a German clothing retailer.

### **Description of the dataset:**

According to the information provided, three different datasets were used in this research:

- i) Sequential session data - data generated from the customer clickstream
- ii) Static customer data - data that doesn't change over time and provides information about the customer
- iii) Combined dataset - a combination of the above two datasets.

The datasets were used to train and evaluate machine learning algorithms for categorizing whether a web shop session will end in a purchase or not.

### **How many agents included in the study?**

The use of machine learning models such as boosted tree, random forest, support vector machine, feed-forward neural network, logistic regression, and recurrent neural networks have been mentioned as part of the methodology. These algorithms can be considered as cognitive agents that were used to categorize customers in a web shop as buying or non-buying customers.

### **How were the study's findings?**

The findings of the study showed that the Random Forest (RF) algorithm performed best in categorizing web shop sessions as buying or no buying sessions, with a reasonable prediction latency. A combined dataset of sequential session data and static customer data led to the best results. The study emphasized that session-based data generated from the customer clickstream is most important for predicting purchase probabilities and personal customer information is not necessary. The study also showed that Recurrent Neural Networks (RNN) trained on less engineered features performed well while requiring less time-consuming feature engineering, when detecting sequential patterns.

### **What were the study's problems or limitations?**

An RNN, representing the class of stateful machine learning models, is additionally trained on datasets, which required fewer feature engineering, to show if stateful models provide good results while reducing the need for feature engineering. After obtaining the results of the different models on all datasets, the models are tested for their robustness to different conditions. Examples of such conditions are the gender of the visitor or the device on which the web shop was visited. This analysis indicates how the models will perform under real conditions after deployment. Deployment is out of scope for this research and can only be regarded as an implication if a tested model outperforms the baseline model.

### **What innovation you observed in the chosen application?**

The innovation described in the text is the use of data mining techniques to analyse customer purchase behaviour in e-commerce, specifically the collection and analysis of personal information and clickstream data from web shop visitors. The goal is to improve customer satisfaction and gain a competitive advantage through increased personalization, efficiency, and engagement. The author is implementing and comparing different machine learning models, including Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Feed-forward Neural Networks (FNN), Higher-order Markov Chains (HMC), and Recurrent Neural Networks (RNN), trained on different datasets to predict the likelihood of a customer aborting a shopping process. The study also aims to explore the predictive power of different datasets, such as static customer data and dynamic clickstream data, through exploratory data analysis.

## **7. A COMPUTATIONAL MODEL TO PREDICT CONSUMER BEHAVIOUR DURING COVID-19 PANDEMIC**

**Source:** <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7643087/>

### **State the purpose:**

The purpose of this paper is to present a computational model to predict consumer behaviour during the COVID-19 pandemic in online shopping. The model uses machine learning techniques to extract implicit knowledge from the logs of online shopping sites, with the aim of improving the understanding of consumer behaviour and providing relevant insights for the retail industry, government, and supply chain management. The paper examines the performance of several individual classifiers and their ensembles using Bagging and Boosting,

and performs a correlation analysis to determine the most important factors influencing the volume of online purchases during the pandemic.

### **What type of design was used?**

In this paper, the authors use two approaches to predict consumer behaviour in e-commerce: a statistical approach and a machine learning approach. The statistical approach involves calculating the correlation between different features, while the machine learning approach involves training a predictive model to anticipate consumer behaviour in online shopping using data collected from DigiKala, a popular online shopping site in the middle east. The authors compare five machine learning classification methods (SVM, DT, SMO, ANN, and NB) and their ensembles using Bagging and Boosting, and evaluate the results using accuracy, precision, recall, and F-Measure criteria.

### **How was the knowledge gained?**

The knowledge was gained in the paper by conducting a comparative study between two approaches, statistical approach and machine learning approach, to predict consumer behaviour in online shopping. The data was collected from the online purchase at DigiKala site, pre-processed to reduce the implementation time, and the correlation between different features was calculated and analysed. Five machine learning classification methods, SVM, DT, SMO, ANN, and Naive Bayes, and two ensemble methods, Boosting and Bagging, were compared and evaluated based on accuracy, precision, recall, and F-Measure criteria calculated from confusion matrix values. The best practice with the highest accuracy was identified.

### **Who are the end users who get benefitted using cognitive system?**

The end users who get benefitted using the cognitive system described in the paper are e-commerce companies. The system aims to predict consumer behaviour in online shopping, so that e-commerce companies can increase consumer purchase rates, increase sales and consumer satisfaction, and gain a competitive advantage.

### **Description about dataset:**

The dataset used in this paper was collected from online purchases at DigiKala, a large and successful online shopping site in the Middle East. The dataset contains 11 features, with the 11th feature being the impact of COVID-19 on consumer purchase volume. The data was pre-processed by normalizing the attributes and calculating the correlation between the features to determine the feature with the highest correlation with the "Effective" feature. The "Effective" feature accepts two values: 'yes' and 'no', which indicate the influence of COVID-19 on consumer purchase behaviour.



### **How many agents included the study?**

In the paper, five machine learning classification methods were used to predict consumer behaviour in online shopping: Support Vector Machine (SVM), Decision Tree (DT), Sequential Minimal Optimization (SMO), Artificial Neural Network (ANN), and Naive Bayes (NB). Additionally, ensemble meta-algorithms, Boosting and Bagging, were also applied with the above classifiers to improve the accuracy of the predictions.

### **What were the study's findings?**

This paper studies the impact of COVID-19 on consumer behaviour in online shopping. The authors propose a prediction model to anticipate consumer behaviour and test 5 classification models (DT, SVM, SMO, NB, and ANN) and ensembles of these models (Bagging and Boosting). The results showed that DT achieved the best accuracy (94.6%) followed by DT ensembles with Bagging (95.3%). Correlation analysis showed that Age and Diabetes were the most influential features in predicting consumer behaviour. The authors suggest future research could involve using meta-heuristic algorithms and different classifiers for higher accuracy, and using a dataset with more COVID-19 related features.

### **What were the study's problems or limitations?**

The study has limitations in terms of sample size and diversity, as well as the potential for selection bias in the data collected. Additionally, the findings are based on self-reported data and may not reflect the full extent of the experiences of the participants. These limitations should be considered when interpreting and generalizing the results of the study.

### **What innovation you observed in the chosen application?**

The given application appears to use cutting-edge technologies and innovative approaches. Some potential innovations observed include the use of machine learning algorithms, deep learning models, and natural language processing techniques, which allow the application to perform advanced tasks such as language translation, image recognition, and sentiment analysis. Additionally, the application may utilize cloud computing and big data technologies to store and process large amounts of data, allowing for real-time, accurate, and efficient results.

## **8. CUSTOMER PURCHASE BEHAVIOUR PREDICTION IN E-COMMERCE: A CONCEPTUAL FRAMEWORK AND RESEARCH AGENDA**

**Source:** <https://www.semanticscholar.org/paper/Customer-Purchase-Behavior-Prediction-in-A-and-Cirqueira-Hofer/743e180528c1223294d941f11abe6095bfbf222f>

### **State the purpose:**

The purpose of this paper is to present a systematic literature review of recent research dealing with customer purchase prediction in the E-commerce context. The review aims to provide a novel analytical framework, an analysis of the main tasks, predictive methodologies, and perspectives, and a research agenda for the field of purchase behaviour prediction online.

### **What type of design was used?**

The design used for analysis in this study is a systematic literature review. The authors followed systematic guidelines from Watson (2002) and Kitchenham, et al. (2009) and used a search query to collect comprehensive literature within the scope of purchase prediction in E-commerce. The search was performed in various scientific databases and covered papers from 2014 to 2019. The authors applied inclusion and exclusion criteria to retrieve papers focused on consumer purchase behaviour prediction in the E-commerce context and finally, a total of 63 papers were selected for extraction and mapping.

### **How was the knowledge gained?**

The knowledge was gained in this study through the systematic literature review of recent research on consumer purchase behaviour prediction in the E-commerce context. The authors collected literature from various scientific databases and applied inclusion and exclusion criteria to retrieve relevant papers. The final 63 selected papers were then analysed to provide a novel conceptual framework, an analysis of the main tasks, predictive methodologies, and perspectives, and a research agenda for the field of purchase behaviour prediction online. The authors gained insights into the current state of research and proposed future work opportunities in the field.

### **Who are the end users who get benefitted using cognitive system?**

The end users who benefit from a cognitive system, such as a predictive analytics system in E-commerce, are businesses and consumers. The businesses use the system to analyse consumer behaviour and predict their needs, allowing them to provide personalized services and improve their customer experience. The consumers benefit from a more personalized shopping experience, including recommendations tailored to their needs, which can make their shopping journey faster and more efficient. Additionally, by predicting consumer behaviour, businesses can also optimize their supply chain and improve the overall customer experience.

**Description about dataset:**

The data used in this literature review is related to customer behaviour in E-commerce platforms, specifically focused on the prediction of customer's next purchases. The research performed a systematic literature review from 2014 to 2019 in several scientific databases (Scopus, Web of Science, Science Direct, etc.) using a search query focused on consumer behaviour and purchase prediction. The search resulted in 9824 proposals, which were filtered to 429 papers, and then further reduced to 63 papers after applying exclusion criteria. The final data set consists of these 63 papers focused on purchase prediction in the E-commerce context.

**How many agents included the study?**

It provides a conceptual framework for analysis and a research agenda. The framework reveals three main tasks in consumer purchase behaviour prediction (i.e., prediction of buying sessions, purchase decisions, and customer intents) and provides eight applications enabled by each task. It also illustrates three perspectives on predictive methodologies. Based on the information provided in the paper, it can be concluded that multiple probabilistic and machine learning statistical models were applied to historical online customer data to predict the likelihood of consumer behavior patterns.

**What were the study's findings?**

The findings of this study show that there is a broad literature on consumer purchase prediction in E-commerce, but there is still a need for further investigation in specific directions. The study presents a novel conceptual framework for analysis and a research agenda for future work in the field. The framework maps the literature on datasets, predictive methods, and tasks with their applications. The study suggests that future research should focus on evaluating predictive methodologies using a benchmark dataset and evaluating performance in multi-task settings such as forecasting the next product, purchase time, or amount a customer will likely buy.

**What were the study's problems or limitations?**

This is a limitation, as the study is only a systematic literature review, and no new data or methods were analysed or tested. The study provides a review of existing literature in the field of consumer purchase prediction in e-commerce, but it does not provide empirical evidence or original findings on the problem. Therefore, the limitations of this study include the lack of empirical evidence and the fact that the results are only a synthesis of existing literature, rather than original research.

**What innovation you observed in the chosen application?**

The study presents a novel conceptual framework for analysing the state-of-the-art in consumer purchase prediction in E-commerce. It suggests a research agenda for future work demands in this field. One possible innovation in this study could be to adopt a benchmark dataset and evaluate predictive methodologies in multi-task settings to better forecast customer behaviour. Another could be to investigate the construction of a framework for purchase prediction that considers the combination of the three tasks identified in the study.

## **9. CUSTOMERS' BEHAVIOUR PREDICTION USING ARTIFICIAL NEURAL NETWORK**

**Source:**

[https://www.iise.org/uploadedfiles/iie/community/technical\\_societies\\_and\\_divisions/sms/abstract\\_909.pdf](https://www.iise.org/uploadedfiles/iie/community/technical_societies_and_divisions/sms/abstract_909.pdf)

**State the purpose:**

The purpose of the paper is to predict customer restaurant preferences based on social media location check-ins and to analyse the accuracy of two machine learning algorithms (Artificial Neural Networks and Support Vector Machines) in predicting the customers' behaviour.

**What type of design was used?**

Traditional multilayer perceptron neural networks are trained by backpropagation to learn the relationship between the inputs and outputs. 67 neurons are used in the input layer, one per restaurant type, and 52 in the output layer. Different architectures for the configurations of hidden layers have been tried to reduce the predicting error.

**How was the knowledge gained?**

The study aims to improve customer behavior prediction by incorporating location-based information and social network influence. The data for the study is collected from Foursquare.com, a location-based social networking site. The data is preprocessed to calculate the Influence Factor, which measures the similarity between the user whose preference is predicted and that user's friends based on historical check-ins. The transience of a check-in location is also analyzed. The study uses Artificial Neural Network (ANN) and Support Vector Machine (SVM) to predict customer restaurant preferences based on the preprocessed data. The results are presented and conclusions and opportunities for future work are discussed.

**Who are the end users who get benefitted using cognitive system?**

The end users getting benefitted here by using the cognitive system, would be the restaurant customers.

**Description about dataset:**

It is unstructured data with text, time stamp and hyperlinks. 120,825 restaurant related check-ins have been collected with user ID, tweets, longitude, latitude, and time stamp.

**How many agents included the study?**

In the project, the following ML/cognitive agents were used for processing the data:

Artificial Neural Network (ANN)

Support Vector Machine (SVM)

These agents were utilized in the project to predict customer behavior based on social media data. Specifically, the ANN and SVM algorithms were used to predict customer restaurant preferences based on social media data. The results of the experiment were then analyzed and discussed in the paper.

**What were the study's findings?**

For each user, a SVM and an ANN are used to compare the predicting result. The accuracy (average percentage of correct prediction) is calculated based on 5-fold cross validation. From the chart, it is clear that the neural network provides a higher accuracy for most users. SVM with linear kernel function does not provide as high accuracy as other non-linear kernels. If the user feature space has a high number of dimensions which are linearly independent of each other, a linear kernel function is not able to separate the features.

**What were the study's problems or limitations?**

It has been seen that, as the social network evolves and increases, the ANN takes longer to train and gather information from the social network. Utilizing distributed processing techniques to speed up the training is one possibility for future work. To further improve the prediction accuracy and utilize more of the available information, a sentiment analysis based on the text of check-ins is another direction for future research. Currently, the user preferences are calculated based on the number of check-ins in the proposed model. From the text of check-ins, the opinions of users can be identified, which may present their preferences more accurately.

**What innovation you observed in the chosen application?**

In this paper, an ANN is utilized for customer behaviour prediction based on social media data. With 5-fold cross validation, the ANN provides 93.13% accuracy consistently, while support vector machine with a sigmoid function only gives 54.00% accuracy. This research proposes a novel way of using the information users receive from their social network. The high level of accuracy illustrates that this information plays an important role in the prediction model.

## **10. PREDICTING CUSTOMER BEHAVIOUR WITH COMBINATION OF STRUCTURED AND UNSTRUCTURED DATA**

**Source:** <https://iopscience.iop.org/article/10.1088/1742-6596/1299/1/012041>

### **State the purpose:**

The aim of this research is to address the problem of deciding which of the e-marketing mediums is most appropriate to target customer of a particular SME. Data was gathered by administering questionnaires and pre-processed based on structured and unstructured data sources.

### **What type of design was used?**

The J48 decision tree classification algorithm was used to mine the data, relevant predictions were made from the structured and unstructured data and the results were evaluated. Main steps for the entire process,

- i) Gather data for predicting customer behaviour
- ii) Pre-process the data based on structured unstructured basis
- iii) Predict from structured and unstructured data
- iv) Evaluate & deploy the model to predict on fresh data.

### **How was the knowledge gained?**

The survey instrument was employed to collect data by administering 400 questionnaires whereas 348 were gathered from the respondents. The questionnaire consists of two sections so as to bring out complete characteristics of the respondent demographic analysis.

### **Who are the end users who get benefitted using cognitive system?**

The end users who get benefitted would be the customers of SMEs in Nigeria.

### **Description about dataset:**

The questionnaires contained sections for both structured and unstructured questions. Structured questions provided information such as age, gender, marital status, occupation, highest academic qualification, average income per month, tribe, religion and internet subscription per month.

### **How many agents included the study?**

Data Gathering Agent: The survey instrument used to gather data by administering questionnaires to respondents.

Data Pre-processing Agent: The data pre-processing stage which consisted of two types, structured pre-processing and unstructured pre-processing.

Customer Behaviour Prediction Agent: J48 algorithm used to predict from structured and unstructured data.

Evaluation and Prediction Agent: The deployment of the model to make predictions on fresh data and evaluate its success rate.

### **What were the study's findings?**

Structured data predication model does better than unstructured data predication model but they both point towards the same direction. It was discovered that predicating from unstructured data expresses more of popular opinion, so decision can start from unstructured results and be fine tuned or validated with predicting from structured data.

### **What were the study's problems or limitations?**

There are several potential limitations of the study mentioned in the conclusion. Firstly, the conclusion only mentions the use of two marketing channels, Google and Facebook, as a basis for comparison in customer behavior prediction. It is possible that there are other important marketing channels that were not considered in the study and as such, the predictions made may not be comprehensive or representative of the customers' behavior as a whole. Secondly, the conclusion mentions using big data analysis to obtain an all-inclusive source of unstructured data. However, the use of big data analysis has its own limitations such as the potential for bias in data selection, difficulty in verifying data accuracy, and issues with data privacy. Additionally, unstructured data can also be difficult to clean and pre-process, which may impact the results of the analysis. Finally, the conclusion mentions that the study aims to turn a target customer into a lifetime customer and eventually a marketing customer, but it does not mention how this will be achieved or what metrics will be used to measure success.

### **What innovation you observed in the chosen application?**

The study of predicting customer behavior in regards to marketing channels is a crucial aspect of decision-making for small and medium-sized businesses. The models developed for this purpose form the foundation for informed marketing decisions and can be incorporated into decision support systems for marketing purposes. One of the main outcomes of this study is demonstrated by the result of structured prediction for PersonID "2". Here, the problem of determining the most suitable marketing channel is solved by weighing the options between Google and classification, revealing that the customer is not a fan of Facebook. The significance of this study lies in its potential to contribute towards achieving one of the main goals of Customer Relationship Management (CRM), which is to convert a target customer into a paying customer and eventually, a lifetime customer and marketing advocate. Furthermore, future plans involve incorporating big data analysis to gather a comprehensive source of unstructured data, adding even more value to the insights provided by the study.

## **11. CUSTOMER PURCHASING BEHAVIOUR PREDICTION USING MACHINE LEARNING CLASSIFICATION TECHNIQUES**

**Source:** [\(PDF\) Customer purchasing behavior prediction using machine learning classification techniques \(researchgate.net\)](#)

### **State the purpose**

The goal of the paper is to investigate how machine learning classification algorithms can be used to forecast client purchasing behaviour based on demographic data and previous purchase history.

### **What type of design was used?**

Predictive model design was used which involves building a model to predict an outcome based on input variables.

### **How was the knowledge gained?**

The data has been collected from wholesale customer's data from UCI-Irvin Machine Learning Data Repository

### **Who are the end users who get benefitted using cognitive system?**



The companies will get benefitted with such kind of research which focusses on customer purchasing behaviour. Using this information, the companies will be effectively able to target their customers, and focus their spendings on customer retention.

**Description about dataset:**

The dataset has 8 attributes, including channel as the target class. The attributes include fresh, milk, grocery, frozen, detergents\_paper, delicatessen, channel and region. The total number of instances is 440. No missing values are present in the dataset.

**How many agents included the study?**

Consumer characteristics from their pre-purchase behaviour was extracted. Using it, consumer purchase prediction model was constructed.

**What were the study's findings?**

The study found that a decision tree can have the same feature as its sub nodes. The number of nodes in the hidden layers vary with the problem. The accuracy of the model does not depend on the feature selection, so the best features that you have selected may or may not give the best accuracy. The highest accuracy was observed for the KnnSgd algorithm. When only the 2 best features are considered, the highest accuracy is obtained by the Adaboost algorithm. The highest accuracy obtained in 92.42%.

**What were the study's problems or limitations?**

The variety of the data can be changed to gain more accuracy on the prediction of personality. A combination of new hybrid algorithms and a new dataset with more instances may improve a model's accuracy. Data in the video format can also be used.

**What innovation you observed in the chosen application?**

The study used various algorithms as well as combinations of multiple algorithms such as SvmAda which consists of 2 base classifiers – linear support vector machine and AdaBoost, RfAda which is a combination of Random Forest classifier and AdaBoost along with KNN as meta model, KnnSgd with KNN and SGD, etc.

## **12. PREDICTING CUSTOMER BEHAVIOUR IN DIGITAL MARKET: A MACHINE LEARNING APPROACH**

**Source:** [\(PDF\) Predicting Consumer Behaviour in Digital Market: A Machine Learning Approach \(researchgate.net\)](#)

**State the purpose:**

This research paper aims to examine the use of machine learning algorithms to predict consumer behaviour in a digital market and to explore the potential benefits for businesses in terms of more informed marketing decisions and improved targeting of customers.

**What type of design was used?**

Predictive model design was used which involves building a model to predict an outcome based on input variables.

**How was the knowledge gained?**

The dataset is fetched from all the online customer transactions occurring between 10/12/2010 and 19/12/2011 for an international based and registered non-store online retail. It is accessible from UCI repository.

**Who are the end users who get benefitted using cognitive system?**

This study could benefit businesses and marketers looking to better understand and predict consumer behaviour in a digital market. By using ML algorithms to analyse data and make predictions, they may be able to make informed decisions about marketing strategies and target customers more effectively.

**Description about dataset:**

The dataset is a structured transactional dataset which contains all the online customer transactions occurring between 10/12/2010 and 19/12/2011 for an international based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers. The online retail purchase data has customer behaviour data with 8 attributes that have both continuous and symbolic attributes. The attributes are InvoiceNo, StockCode, Description, Quality, InvoiceDate, UnitPrice, CustomerId and Country. The dataset has 3 numeric and 8 nominal data type. All the data types need to be converted to uniform format for efficient analysis.

**How many agents included the study?**

Only 1 agent, i.e., the non-store online retail company is included as agent in the dataset.

**What were the study's findings?**

The paper was able to identify customer behaviour features patterns and mining association rules between frequent purchase behaviour on an online store. Optimum rule generation occurred at minimum support and confidence thresholds of 0.1 and 0.2. The paper discovered interesting frequent customer behaviour purchasing patterns that occurred in the online retail dataset and mined strong association rules.

**What were the study's problems or limitations?**

The limitation of this study is that the performance of the model created is greatly affected by the quality and dimensionality of the dataset used and nature of feature set.

**What innovation you observed in the chosen application?**

The innovative thing that I observed in this research is the use of apriori algorithm. The algorithm performs frequent item set mining and association rule learning over the database records. It is used to implement the association in rule mining. The algorithm needs a minimum support level and a dataset as input. It generates a list of all network feature itemsets with one item.

### **13. A NAÏVE BAYES STRATEGY FOR CLASSIFYING CUSTOMER SATISFACTION**

**Source:** [\(PDF\) A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services \(researchgate.net\)](#)

**State the purpose:**

The purpose of the study is to assess whether terms related to guest experience can be used to identify ways for hospitality services providers (such as hotels) to improve their offerings. It also aims at empirically identifying relevant features that classify customer satisfaction based on reviews from a social networking site, Yelp.

**What type of design was used?**

The research paper used a quantitative design which involves collecting and analysing numerical data. The study was conducted to empirically identify relevant features that classify customer satisfaction based on several reviews.

**How was the knowledge gained?**

The data was collected from the 9<sup>th</sup> Yelp Dataset Challenge to understand guests' preferences, filtering records whose category contained the term 'Hotels'.

### **Who are the end users who get benefitted using cognitive system?**

The primary beneficiaries of this research are hotel managers, as the model developed from it can help them understand guests' satisfaction better and provide insights into how hospitality services can be improved. Additionally, customers who stay in hotels will also benefit from these improvements to service quality.

### **Description about dataset:**

The dataset used in this research paper was 47,172 reviews of 33 Las Vegas hotels registered with Yelp - a social networking site. This data included customer ratings and comments on various aspects such as staff experience, professionalism, tangible and experiential factors, gambling-based attractions etc., which were then analysed using supervised machine learning approach to identify relevant features that classify customer satisfaction.

### **How many agents included the study?**

The study was conducted to empirically identify relevant features to classify customer satisfaction based on the reviews.

### **What were the study's findings?**

The findings of this research paper are that terms related to guest experience can be used to identify ways to enhance hospitality services. The model developed from the study was able to classify customer satisfaction with high precision and recall, while having a low computational cost compared to prior statistical results based on limited sample size. This provides insights into how hotels can improve their services by focusing on staff experience, professionalism, tangible, and experiential factors as well as gambling-based attractions.

### **What were the study's problems or limitations?**

The main limitation of the study is that it was conducted using reviews from only 33 Las Vegas hotels registered with Yelp, which may not be representative of the entire hospitality industry. Also, since ML algorithms are based on statistical models and assumption about data distributions, there could be potential bias in the result due to these factors.

### **What innovation you observed in the chosen application?**

This paper provides a comprehensive assessment of how terms related to guest experience can be used to identify ways for hospitality services improvement. It is based on an empirical study

conducted with 47,172 reviews from 33 Las Vegas hotels registered in Yelp - a social networking site. The resulting model helps hotel managers understand guests' satisfaction and process vast amounts of review data using supervised machine learning approach. Compared to prior statistical results which were limited by sample size, this research offers more reliable and accurate insights into what factors influence customer satisfaction such as staff experience, professionalism etc., as well as gambling attractions offered at the hotel. Additionally, it provides insights into how hotels can enhance their services by focusing on staff experience, professionalism, tangible/experiential factors & gambling-based attractions etc.

#### **14. A PURCHASING PREDICTION MODEL CONSIDERING PRE-PURCHASE BEHAVIOURS**

**Source:** [A Purchasing Prediction Model Considering Pre-purchase Behaviors | Semantic Scholar](#)

##### **State the purpose:**

The purpose of the paper is to propose a purchasing prediction model that considers the purchasing behaviours of customers. The model aims to improve the accuracy of purchasing predictions by considering the various behaviours of customers, such as purchase frequency and purchase history.

##### **What type of design was used?**

The research uses a predictive model design. The authors aim to develop a purchasing prediction model that considers the behaviours of customers and evaluate its accuracy in comparison to existing models.

##### **How was the knowledge gained?**

The knowledge for this study was gained through the development and evaluation of the proposed purchasing prediction model. The authors gathered data on customer behaviour and used this information to build and test the model. They then compared the results of the proposed model with those of conventional methods to assess its accuracy. The results of this comparison were used to gain insights into the effectiveness of considering customer behaviours in purchasing predictions.

##### **Who are the end users who get benefitted using cognitive system?**

Businesses and organizations involved in e-commerce or retail sales will get benefitted from this research. The proposed purchasing prediction model can be used to help these entities

improve their predictions of customer purchasing behaviours, which can inform decisions related to inventory management, marketing, and other areas. By using the model, these organizations may be able to increase efficiency and revenue, and better meet the needs of their customers.

#### **Description about dataset:**

The data was collected from 01/07/2010 to 28/06/2011. A purchase prediction model was constructed using the conversion flag as the objective variable. Some of the attributes were age, sex, purchase amount, purchase quantity, session time, item browsing time, site access rate, preference diversity, etc.

#### **How many agents included the study?**

Consumer characteristics from their pre-purchase behaviour was extracted. Using it, consumer purchase prediction model was constructed.

#### **What were the study's findings?**

The proposed purchasing prediction model considers customer behaviours, such as purchase frequency and purchase history, is more accurate than conventional models. The results of the study indicate that the proposed model shows improved accuracy in comparison to existing methods. The authors conclude that considering customer behaviours in purchasing predictions is an effective way to improve the accuracy of these predictions.

#### **What were the study's problems or limitations?**

The data collected is not evenly distributed for consumer preference diversity, which has led to some random behaviour of the model. The study was not able to obtain information such as referral data for this survey, so it was not able to confirm customer migration. Also, the e-commerce website for which the data was collected did not have product reviews and product rating values, which reduced the effectiveness of the model.

#### **What innovation you observed in the chosen application?**

The innovative aspect of this research study is the proposed purchasing prediction model, which considers customer behaviours, such as purchase frequency and purchase history, in making predictions about customer purchasing behaviours. This is an innovative approach because conventional purchasing prediction models typically do not consider these types of customer behaviours. By including these behaviours in the model, the authors aim to improve the accuracy of purchasing predictions, which can benefit businesses and organizations involved in e-commerce or retail sales. The results of the study support the effectiveness of this

approach, as they indicate that the proposed model is more accurate than conventional methods.

## **15. IDENTIFYING THE FACTORS AFFECTING CUSTOMER PURCHASE INTENTION**

**Source:** [\[PDF\] Identifying the Factors Affecting Customer Purchase Intention | Semantic Scholar](#)

### **State the purpose:**

The purpose of this study is to examine the factors that influence customer purchase intention in the context of e-commerce. The study aims to identify the most significant factors affecting customer behaviour when making online purchases, and to provide insights for e-commerce businesses and marketers on how to enhance customer purchase intention. The research aimed to contribute to the existing literature in marketing and consumer behaviour by providing new insights on the determinants of customer purchase intention in the e-commerce context.

### **What type of design was used?**

This study follows the qualitative method. 100 self-earning and self-spending respondents were selected and asked to fill a questionnaire. Regression analysis was used to get the results.

### **How was the knowledge gained?**

100 self-earning and self-spending respondents were selected and asked to fill a questionnaire. Based on their responses, a relationship between various independent variables and customer purchase intention was discovered.

### **Who are the end users who get benefitted using cognitive system?**

E-commerce businesses, online retailers and marketers will benefit from this research. The study aims to provide insights into the factors that influence customer purchase intention in the e-commerce context, which can inform business and marketing strategies for improving customer satisfaction and increasing sales. By understanding the key determinants of customer purchase intention, e-commerce businesses can make data-driven decisions to enhance their customer experience and grow their business. The study may also benefit academic researchers by adding to the existing literature on customer behaviour and purchase intention in e-commerce.

### **Description about dataset:**

Based on the responses of the participants of the study, a relationship between various independent variables and customer purchase intention was discovered. The attributes used were celebrity endorsement, product packaging and customer knowledge along with their age, sex and demographic characteristics.

#### **How many agents included the study?**

100 respondents participated in the study. These self-earning and self-spending participants were asked to fill a questionnaire for the research.

#### **What were the study's findings?**

Data was found reliable to all the variables. Regression analysis was used. All the variables showed positive correlation with the independent variable. Celebrity endorsements and product packaging played an important role.

#### **What were the study's problems or limitations?**

The study considers that the participants will replicate the same behaviour in the real world, but there may be a difference in the response of the participants and their action in the real world while purchasing the product. The study has only considered the responses of 100 participants. The study has used regression algorithm, which may not cover all the aspects of the observation and can be unsuitable for the analysis. It would be better if the methodology also considered other algorithms. Very limited number of attributes were used, which did not consider the buying habits of customers.

#### **What innovation you observed in the chosen application?**

While many studies focus on the Western market, this study aims to identify purchasing behaviour patterns of customers in Asian markets. It studies the difference in purchasing behaviour based on age and economic conditions. The study has considered 4 hypothesis and used its methodology to prove all of them.