Advances in Consumer Research
ISSN (Print): 0098-9258
ISSN(Online): 3079-1766

Vol. 2, Issue 4 (2025) https://acr-journal.com/

# Customer Recommendation Prediction for a Retail Furniture Store Using Machine Learning Techniques

**Lakshmi Devasena C[1]**

[1]ICFAI Business School Hyderabad, The ICFAI Foundation for Higher Education (IFHE) University, Hyderabad, India
Email ID: ldevasena@ibsindia.org

**Cite this paper as:** Lakshmi Devasena C, (2025) Customer Recommendation Prediction for a Retail Furniture Store Using Machine Learning Techniques. *Advances in Consumer Research*, 2 (4), 332-345

**KEYWORDS**

*Customer Recommendation prediction, Artificial Neural Network, Logistic Regression, Linear Discriminant analysis, Recommender Systems, Machine Learning*

**ABSTRACT**

As customer recommendation by word-of-mouth can attract more customers, this research deals with Customer recommendation prediction for a furniture store like IKEA. This research compares the machine learning models for Recommendation, namely, Artificial Neural Network (ANN), Logistic Regression (LR), and Linear Discriminant Analysis (LDA). The research is based on customers' feedback about the first store in Hyderabad, India. In this research work, IKEA's service is explored, and the models created are successful based on the predictor variables used in the study as the model's inputs. The abovementioned techniques were used to predict the customer recommendation of IKEA to their neighbours, family, and friends. It is evident from the classification outcomes of different models that ANN outperforms the other techniques, namely, the Logistic Regression model and LDA. As a managerial implication, this model could be used by any furniture or multinational retail stores to predict their customer recommendation status to improve their market and profit. Customer recommendation prediction in national/multinational stores, especially retail furniture stores, does not exist in the literature. This research aims to fill the gap in the literature to predict the customer recommendation status of retail furniture stores or any other multinational stores

## 1. INTRODUCTION

Any information technology tool or recommendation system analyzes the critical firm's factors, improving the customer service experience. Also, retaining the customer is most important for the organization's success. Any top retail branding company looks to sustain itself for a longer time by ensuring factors like Cost, Time to market, Profits, Quality, time, Flexibility, Agility, Innovation, and so on. Customer satisfaction plays a prominent role in retailing – a burgeoning sector with significant impacts on potentials and growing opportunities in fast-growing economies like India [1]. There are various factors that customers look at when opting for a retail store, like product quality, brand awareness, delivery and service, Location, Expectation matching, In-store convenience, and so on [2]. Hence, the retail firm must ensure such factors to improve customer satisfaction. Today, in the era of a competitive world, there is intense market competition, capital shortages, and dynamic retail sources. Hence, they require attention to analyze customer expectations accurately. Customer satisfaction leads to their retention and recommendation of the store and the products used by their neighbors, friends, and relatives. Evidence shows that customer reviews are essential in making purchase decisions, mainly in the e-business and offline business.

This research work has taken IKEA store, Hyderabad as an instrument for collecting the data and analyzed the data for predicting the customer recommendation. IKEA, one of the largest furniture retail stores globally, wanted to reach 50 billion euros in total trade by 2020, with an annual growth of 10 percent [3]. In order to expand its market and multiply its profits, it needs to closely monitor customer recommendations about IKEA to others. Thereby, there is a need for research to analyze whether IKEA should expand its retail services throughout India using predictive modeling. Researchers apply Neural networking in various aspects for model prediction, data mining, and classification in the areas where traditional statistical methods are used. The objective of this work is to develop a prototype to envisage whether the customer will recommend IKEA to others or not. This paper will appraise the best foretelling model for recommending furniture stores like IKEA based on customer satisfaction and service experiences. The objective is to compare and differentiate the extrapolative powers of ANN, Logistic Regression, and LDA to predict customer recommendation. The data is collected from the feedback responses of the customers who visited IKEA in Hyderabad, and the models are analyzed through SPSS.

### 1.1 Aim and Objectives

This research involved predicting customer recommendation by developing various models (like ANN, LR, and LDA) and comparing them to see the outperforming model for the same. This can be attained over and done with the succeeding research objectives:

1) To develop a Logistic Regression predictive model proficient in envisaging the perception of customers towards an international / national retail furniture store (whether to recommend the store (eg. IKEA) to others or not)
2) To develop an LDA model consummate to forecast the perception of customers towards an international / national retail furniture store
3) To build the ANN model to identify the perception of customers towards an international / national retail furniture store

To compare and differentiate the foretelling powers of all three models and come up with the best among all

## 2. LITERATURE REVIEW

Past studies of IKEA's entry engrossed its pecuniary properties in the retail industry. Retailing as an industry has changed a lot in India regarding managing the customer's service experience, meeting the demands of different segments of customers, dealing with sufficient inventory, and ensuring the growing needs and opportunities [4]. With the effect of globalization, Indian consumers are significantly looking to adopt the Western lifestyle and their way of shopping cultures, which leads to IKEA's easy success as a new entrant in this country [5]. However, it is necessary to study the consumers' underlying perception to make it successful. For defining the entry plan to India, the company took over three years to understand the consumer's cultural roots, tastes, and buying behaviours [3].

A framework is developed from the outcomes of IKEA retail services to produce and succeed in resource configurations that empower, upkeep, and direct customers in service exchange and value co-creation [6]. Some studies analyze that consumers' attitudes are heightened by awareness, associations, perceived quality, and self-congruity, leading to purchase intentions [7] & [8], proposed that store environment and atmosphere also affect the customer's service experience in retail stores. They have concluded that factors like time spent in the store, frequent visits to the outlet, acceptance of retail among different age groups, merchandise purchased, locations of store, and discounts offered will impact customers' purchase intentions. Various sentiment analyses have been done based on the tweets around the opening of the IKEA store to understand customers' perceptions of IKEA [9]. Many models have been developed to analyze the customer experience using various techniques like the SERVQUAL model, Decision tree approach, Logistic Regression, Machine Learning, Multiple Regression, Naive Bayes, Neural Networking, etc. ([9],[10] & [11]).

Service quality in retail businesses is analyzed using the SERVQUAL model [12]. This model helps the retailer to assess where the service is weak and requires constant focus. [13], investigated the customer's point of view on shopping companion apps based on the service quality approach. This study showed how the various aspects expected by customers in shopping companion apps impact their purchasing in retail. [14], developed a predictive model to analyze customer gratification and fidelity in the Omni channel trading industry. This work guides practitioners in improving service quality to meet customer requirements.

Promotional gifts or offers are more effective than discounts the stores provide; promotional offers attract more customers despite their value [15]. The authors emphasized that promotional gifts will attract more buyers than the discount provided when promoting a product from a firm or store.

A study by [16] suggests that SERVQUAL models are used to foresee the overall consummation and emotional attitude of customers, and it is designed with three distinct regression models to ensure different types of fidelity. A study [17] used LR to select feasible features in clinical medicine to predict patients' health conditions. This Logistic regression model is used to classify heart attack cases in coronary heart disease. A study examined the accuracy of the LR model in analyzing the presence or absence of Pulmonary Tuberculosis and its risk-associated factors [18]. The ANN-based approach is used to prophesy customer churn in the telecom industry. This model helps predict the possible churning of customers and takes appropriate actions to prevent customer churn [19]. Five Machine learning techniques, like ANN, Decision tree learning, Support vector machines (SVM), Naive Bayes, and LR analysis, are used to analyze churn forecasts in the telecom industry [20].

Word-of-mouth (WOM) through social media networks like Instagram, WhatsApp, etc., reaches more people than other channels to reach customers. Recommendations of products or products to friends, relations, and neighbours will be more effective through electronic WOM [21]. So, the firm can keep the customer satisfied to have more recommendations.

However, it is necessary to compare the statistical predictive models to make better predictions in the retail industry. This research aims to develop the paramount model to predict the customer recommendation of a retail firm (like IKEA) by comparing ANN, Logistic regression, and LDA prediction models.

## 3. METHODS AND MATERIALS USED

In the following, the description of three methods used to identify the customer's perception of recommendations on retail stores is given.

### 3.1 Artificial Neural Networks (ANN)

An ANN resembles a biotic neural network in the human body [22]. The ANN model has an input layer, one or more hidden layers, and an output layer. The input neurons in the input layer acquire the inputs, and then the sum products of inputs and the random weight assigned are forwarded to the respective hidden layer neurons. Hidden layers neurons get the data from the input layer and use some activation functions like triangular, sigmoid, softmax, etc., to map them between the outputs in the range from zero to one. Then, the hidden layer neurons forward the outputs to the output layer, and sometimes activation functions in the output layer neurons further do calculations to produce the output [23].

X: Input Neurons W, Y: Weights H: Hidden Neurons. O: Outputs

The ANN output will be between 0 and 1 after backpropagation and after the number of iterations is over. The backpropagation is fine-tuning the weights of the connections till the output is in the anticipated range.

### 3.2 Logistic Regression

The LR model analyses the association between the output variable and one or more independent/latent variables. The dependent variable is categorical, i.e., dichotomous (Binary) [24]. In this study, the output variable is binary, where value ‘0’ takes the ‘not recommendable retail store,' and the value ‘1’ takes the ‘recommendable retail store.' It calculates the Recommendation probability concerning a set of values for the predictor variables. The model is given by

$Yj = ey / (1 + ey)$

Where $Y_j$ is the projected likelihood that the jth case in a category and 'y' is the ordered LR equation: $y = B0 + B1X1 + B2X2 + \ldots + BkXk$

The predictor value is given as follows,

$$p(x) = P(Yj = 1\,X = x) => E(YjX = x)$$

Odds values vary from zero to infinity, eliminating the difficulty of fitting a linear model to the highest asymptote. Considering the logarithm of the probabilities, the range of a dependent variable is -1 to 1. In an attempt to fit an LR model to the variable log odds, the model would be,

$$logit(p(x)) = \left(\log\left(\frac{p(x)}{1-p(x)}\right) = \beta 0 + \beta 1 \ldots\right)$$

The change in likelihood is not linear with linear changes in X. This means that the possibility of success or failure (Y = 1 or 0), given the independent variable (X), is a non-linear function. [17]

### 3.3 Linear Discriminant Analysis

LDA is a classification approach proposed in [24]. It is the widely used procedure for classifying data into predefined groups. The most naive type of LDA is a Linear Discriminant Function (LDF), which can discriminate between two groups and pass through the centroids (geometric centers) of the two groups [26].

The following Equation symbolizes the LDF:

$LDF = a + b1x1 + b2x2 + \cdots .. + bkxk$,

where ‘a’ is a constant, and ‘b1’ to ‘bk’ are the regression coefficients for k predictor variables

#### *3.3.1 Wilks' Lambda Test for Significance of Canonical Correlation Hypothesis*

Wilks' lambda is used to measure the practicality.

Hypothesis for Canonical correlation are:

$H_0$: No linear association amid two variable sets

$H_a$: A linear association exists amid two variable sets

Test Statistic: $\lambda = |E|/(|E+V|)$, where E is error difference, V is variance due to a linear association, E+V is the total variance

Rule: Reject $H_0$ if $P < 0.05$; otherwise, accept default hypothesis $H_0$ at a 5% significance level.

### *3.3.2 Chi-Square Test*

Chi-Square Test hypotheses are

$H_0$: The variables are not dependent

$H_a$: The variables are dependent

Test Statistic:

$\chi 2 = \Sigma\Sigma\ ((Oij - Eij)\string^2)/Eij$, where Oij is the perceived value, and Eij is the anticipated value. Where summations are 'i' from 1 to 'r' and 'j' from 1 to 'c' ('r' and 'c' are constants).

Decision rule: Reject $H_0$ if $P < 0.05$; otherwise, accept $H_0$ at a 0.05% significance level.

### 3.4 Calculation of Predictive Powers

The foretelling powers of the three models are compared using the following parameters.

1. Accuracy: The number of observations predicted correctly (‘recommending the store’ and ‘Not recommending the store’) from the set of observations taken for analysis:

$A = (TP + TN) / (TP + FP + TN + FN)$

2. Specificity: Number of correctly classified negative results (‘Not Recommending the store’) obtained from the total negative results obtained: $Spec = TN / (FP + TN)$

3. Sensitivity: Number of correctly classified positive results (‘Recommending the store’) obtained from the total positive results obtained: $Sen = TP / (TP + FN)$

4. Precision: Number of correctly classified positive results (‘Recommending the store’) obtained from the actual number of favorable results obtained: $Pre = TP / (TP + FP)$

Where TP (True Positive): Recommendable and classified as recommendable

FP (False Positive): Recommendable but incorrectly classified as non-recommendable

TN (True Negative): Non-Recommendable but incorrectly classified as recommendable

FN (False): Non-Recommendable and classified as Non-recommendable

### 3.5 Materials Used

The dataset utilized in this research comprises 201 sample respondents from the shoppers of the Hyderabad IKEA store, with a Questionnaire of 16 Questions. The primary data is collected from respondents from all over India using the convenience snowball sampling method who are the purchasers of Hyderabad IKEA store (Exit Survey method). LR, LDA, and ANN are used to foresee customer recommendations. Of 201 responses, 168 answered ‘IKEA store as "Recommending IKEA store,' and 33 "Not Recommending IKEA store."

### *3.5.1 Selection of Predictor Variables*

It is vital to recognize which factors are essential and select them in the ultimate forecasting model since the factors decide the customer recommendation status to the store. The selection of independent variables should be based on theory support and statistical analysis. The customer service experience can be better measured with the selection of the best and most significant predictor variables of the model [27]. The selection of independent variables are based on the literature mentioned above to reach the accurate prediction.

To perform the customer recommendation prediction, using logistic regression, LDA, and ANN at IKEA, the following 12 factors are selected from the 16 questions (which are mentioned in Table 1 from 1 to 12).

**Table 1: Description of Data Set**

| S.No. | Question | Type of Variable | Scale | Variable |
|---|---|---|---|---|
| 1 | Do you know about the brand IKEA? | Input | Nominal | Brand Awareness |
| 2 | Have you ever visited the IKEA showroom? | Input | Nominal | Visiting store |
| 3 | Tell your age | Input | Scale | Customer Age |
| 4 | How was the experience at the store? | Input | Scale | Store Experience |
| 5 | How would you rate the product quality of IKEA? | Input | Scale | Product Quality |
| 6 | Does the brand do the pricing? | Input | Scale | Product Pricing |
| 7 | What is your family income or your income(if you are the only earning member)? | Input | Scale | Income Level of consumer |
| 8 | Comment on the store location | Input | Scale | Location of the store |
| 9 | Rate the service and delivery aspects of IKEA. | Input | Scale | Delivery and Service from IKEA |
| 10 | Did the products match your expectations? | Input | Scale | Expectation Matches |
| 11 | Did you like the idea of IKEA compelling you to visit their entire store? | Input | Scale | Store Facility Design |
| 12 | How frequently will you visit the store in the future? | Input | Scale | Visit frequency to the store |
| 13 | Will you recommend IKEA to others? | Output | Nominal | Recommendation Status |

**Table 2: ANN Model Summary**

**Model Summary**

| | | |
|---|---|---|
| Training | Cross Entropy Error | 59.415 |
| | Percent Incorrect Predictions | 18.2% |
| | Stopping Rule Used | 1 consecutive step(s) with no decrease in error[a] |
| | Training Time | 0:00:00.05 |
| Testing | Cross Entropy Error | 18.889 |
| | Percent Incorrect Predictions | 10.3% |

Dependent Variable: Recommendation_Status

a. Error computations are based on the testing sample.

**Table 3: Importance of Independent Variables**

**Independent Variable Importance**

| | Importance | Normalized Importance |
|---|---|---|
| Brand_Awareness | .027 | 17.3% |
| Visit_to_the_store | .085 | 53.7% |
| Age_of_the_Costomer | .159 | 100.0% |
| Experiance_at_the_Store | .085 | 53.4% |
| Product_Quality | .104 | 65.4% |
| Product_Pricing | .065 | 40.8% |
| Income_Level_of_Customer | .093 | 58.7% |
| Location_of_Store | .087 | 54.7% |
| Delivery_and_Service_from_IKEA | .141 | 89.0% |
| Expectation_Matches | .098 | 62.0% |
| Store_Facility_Design | .038 | 23.7% |
| Visit_Frequency_to_Store | .018 | 11.4% |

**Table 4: Predictive ability of ANN**

**Classification**

| | | Predicted | | |
|---|---|---|---|---|
| Sample | Observed | .0 | 1.0 | Percent Correct |
| Training | .0 | 2 | 24 | 7.7% |
| | 1.0 | 2 | 115 | 98.3% |
| | Overall Percent | 2.8% | 97.2% | 81.8% |
| Testing | .0 | 1 | 6 | 14.3% |
| | 1.0 | 0 | 51 | 100.0% |
| | Overall Percent | 1.7% | 98.3% | 89.7% |

Dependent Variable: Recommendation_Status

**Table 5: Walds statistics**

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | 1.627 | .190 | 73.054 | 1 | .000 | 5.091 |

**Table 6: Hosmer and Lemeshow Test**

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 170.508[a] | .044 | .074 |

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

**Hosmer and Lemeshow Test**

| Step | Chi-square | df | Sig. |
|---|---|---|---|
| 1 | 14.335 | 8 | .073 |

**Contingency Table for Hosmer and Lemeshow Test**

| | | Recommendation_Status = .0 | | Recommendation_Status = 1.0 | | |
|---|---|---|---|---|---|---|
| | | Observed | Expected | Observed | Expected | Total |
| Step 1 | 1 | 6 | 6.840 | 14 | 13.160 | 20 |
| | 2 | 4 | 4.528 | 16 | 15.472 | 20 |
| | 3 | 8 | 3.806 | 12 | 16.194 | 20 |
| | 4 | 3 | 3.467 | 17 | 16.533 | 20 |
| | 5 | 5 | 3.361 | 16 | 17.639 | 21 |
| | 6 | 0 | 2.930 | 20 | 17.070 | 20 |
| | 7 | 0 | 2.678 | 20 | 17.322 | 20 |
| | 8 | 3 | 2.383 | 17 | 17.617 | 20 |
| | 9 | 3 | 1.934 | 17 | 18.066 | 20 |
| | 10 | 1 | 1.072 | 19 | 18.928 | 20 |

**Table 7: Predictive Ability of Logistic Regression**

**Classification Table[a,b]**

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | Recommendation_Status | | Percentage Correct |
| | | | .0 | 1.0 | |
| Step 0 | Recommendation_Status | .0 | 0 | 33 | .0 |
| | | 1.0 | 0 | 168 | 100.0 |
| | Overall Percentage | | | | 83.6 |

a. Constant is included in the model.

b. The cut value is .500

**Table 8: Wilk's Lambda statistics**

**Wilks' Lambda**

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---|---|---|---|---|
| 1 | .957 | 8.417 | 12 | .752 |

**Table 9: Canonical Correlation-Eigen Values**

**Eigenvalues**

| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
|---|---|---|---|---|
| 1 | .045[a] | 100.0 | 100.0 | .207 |

a. First 1 canonical discriminant functions were used in the analysis.

**Table 10: Importance of Independent variables**

| Independent Variables | Structure Matrix | Standardized Discriminant Coefficients | Canonical Discriminant Coefficients |
|---|---|---|---|
| Brand Awareness | -0.300 | -0.328 | -2.343 |
| Visit to the store | 0.175 | 0.498 | 1.946 |
| Age of the customer | -.596 | -0.563 | -0.973 |
| Experience at store | 0.428 | 0.479 | 0.581 |
| Product Quality | 0.041 | 0.050 | 0.062 |
| Product Pricing | 0.087 | -0.138 | -0.181 |
| Income Level of consumer | -0.130 | -0.266 | -0.347 |
| Location of the store | 0.254 | 0.409 | 0.737 |
| Delivery and Service from IKEA | 0.154 | 0.433 | 0.267 |
| Expectation Matches | -0.186 | -0.263 | -0.497 |
| Store Facility Design | 0.112 | 0.158 | 0.178 |
| Visit frequency to the store | 0.097 | 0.130 | 0.154 |

| | | | |
|---|---|---|---|
| Constant | | | -0.689 |

**Table 11: Predictive ability of LD**

**Classification Results[a,c]**

| | | WILLYOURECOMMENDIKEATOOTHERS | Predicted Group Membership .0 | Predicted Group Membership 1.0 | Total |
|---|---|---|---|---|---|
| Original | Count | .0 | 20 | 13 | 33 |
| | | 1.0 | 54 | 114 | 168 |
| | % | .0 | 60.6 | 39.4 | 100.0 |
| | | 1.0 | 32.1 | 67.9 | 100.0 |
| Cross-validated[b] | Count | .0 | 11 | 22 | 33 |
| | | 1.0 | 64 | 104 | 168 |
| | % | .0 | 33.3 | 66.7 | 100.0 |
| | | 1.0 | 38.1 | 61.9 | 100.0 |

a. 66.7% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 57.2% of cross-validated grouped cases correctly classified.

**Table 12: Evaluation of Models Predictive powers**

| Techniques | Accuracy | Sensitivity | Specificity | Precision |
|---|---|---|---|---|
| LDA | 66.7% | 89.8% | 27% | 67.9% |
| Logistic Regression | 83.6% | 83.6% | 0% | 100% |
| ANN | 85.08% | 84.7% | 60% | 98.8% |

a. Brand Awareness
b. Visiting store
c. Customer Age
d. Store Experience
e. Product Quality
f. Product Pricing
g. Income level of consumer
h. Location of the store
i. Delivery and service from IKEA
j. Expectation matches
k. Store facility design
l. Visit frequency to the store

The data set contains Thirteen (13) variables (Table 1). Among them, three are categorical (nominal), and the rest are Numerical. Moreover, the data set has 12 predictor variables and one Output (Dependent) Variable.

## 4. DATA ANALYSIS USING DIFFERENT ML TECHNIQUES

### 4.1 ANN Architecture

An ANN model is developed using the multilayer perceptron approach. There are 12 independent variables; five neurons are in the hidden layer, and one output neuron is in the output layer for the output of binary value '0' or '1'. The analysis is done using SPSS software. The SPSS can mechanically choose the optimal design and construct the training neural network with one hidden layer. Setting the number of neurons in the hidden layer is obvious, and the automatic network selection approach decides the "optimal" number of neurons. The activation functions are used in the hidden layer (Hyperbolic Tangent) and the output layers (SoftMax). The Hyperbolic Tangent function is widely used in feed-forward networks to classify two classes. This function is much better than the logistic sigmoid function. Its range lies from -1 to 1. It provides an S- A shaped curve, as shown in Figure 2. The Hyperbolic Tangent (tanh) function has the benefit of mapping negative inputs to strongly negative and mapping the zero with near zero in the 'tanh' graph of the output [28]. The calculation of the 'tanh' function is as follows:

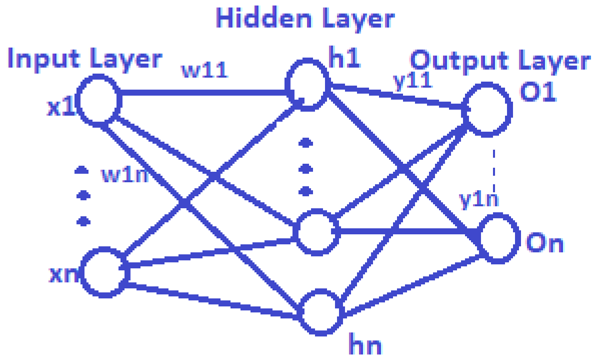


**Fig 1: ANN with one Hidden layer**

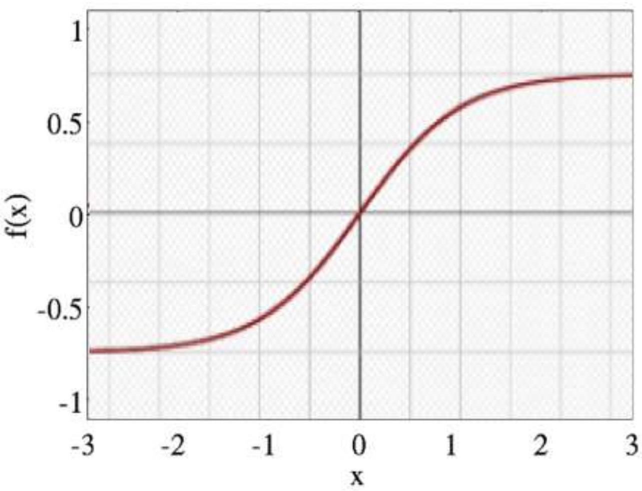


**Fig 2: Hyperbolic Tangent function**

$$\tanh(vi) = (e^{x} - e^{-x})/(e^{x} + e^{-x})$$

$\phi(vi) = U(vi)\tanh(vi)$, where U is the step function.

The softmax function is used to handle classification problems. It is non-linear and is used to handle even multiple classes. The softmax function would provide the outputs for each class between 0 and 1 and divide by the output sum. It is ideally used in the output layer of the classifier [28].

The Equation for the Softmax function is given by

$fi(x) = \frac{e^{xi}}{\Sigma e^{xj}}\ for\ i = 1, \ldots j$

Input variables consist of "Nominal" and "Continuous" variables. There are two nominal variables and ten continuous variables. In addition, 70% of the data is assigned to training the model, and 30% is allocated as test data to ensure model performance and avoid overfitting. Here, batch training, appropriate for 'smaller' datasets, is chosen to minimize the total error unswervingly. Besides that, the 'Scaled Conjugate Gradient' optimization algorithm supports the batch training method used in Training and testing [26].

#### *4.1.1 Model performance in ANN*

The outcomes of the ANN training are displayed in the subsequent model overview in Table 2, which describes the performance of the ANN model built. Since 'SoftMax' is used as the output activation function, Cross entropy error is exhibited. During Training, the network attempts to reduce this error function. Additionally, the percentage of improper estimates in the training data is comparable to 18.2%. Consequently, the proportion of accurate estimates is nearer to 81.8%, which is high.

In the same way, for the testing data, the proportion of incorrect estimates is closer to 10.3%, which means the proportion of accurate estimates is closer to 89.7%. The mean relative error would be presented for the scale measurement level output variable. Table 2 shows the mean percentage of incorrect estimates for the nominal type output variable.

#### *4.1.2 Predictor Variables Importance*

ANN model presents various insights about the data analyzed. The importance of predictor variables, which the ANN model determines, is presented in Table 3. The calculations have been done based on the training and testing data used. The network's model-predicted value changes by how well an independent variable's value affects the network's relevance. There are two ways of expressing how important something is: a percentage of how important it is divided by how important it is. Table 3 shows that Customer Age is most important here, followed by 'Delivery and Service from IKEA.' From the literature, it could be evident that any company's 'Delivery and Service' leads to its growth and survival, which is projected from the data analyzed. 'Product Quality' is the next important variable, followed by 'Expectation Matches'. So, IKEA needs to identify the customer choices and keep the products related to their expectation. The 'Income level of the customer' and the 'experience at the store' are also more critical in recommending the store.

#### *4.1.3 Models Predictive Ability*

Predictive models based on artificial neural networks have high classification accuracy, achieved during the model formulation phase. Table 4 shows that from the training set of 117 recommendable responses, the neural network model classifies 115 respondents as 'Recommend the Store,' showing a classification accuracy of 98.3%. Also, the same ANN model classified 2 out of 26 respondents as 'not recommend the store' with a classification accuracy of 7.7%. Overall, the model can generate 81.8% accuracy as a whole. On the other hand, in the testing data, the 'Recommend the Store' category observation data are correctly predicted with 100% classification accuracy, and the 'Not-recommend the store' category is predicted with 14.3% accuracy. Overall, the model provides 89.7% accuracy for testing data for both 'Recommend the Store' and 'Not Recommend the Store.' Considering both Training and testing observations, the model produced **85.08%** accuracy for 'Recommend the Store' and 'Not Recommend the Store.'

### 4. 2 Logistic Regression Analysis

Coefficients in a logistic regression equation are computed using maximum-likelihood estimation. Finding coefficients that match the dependent variable's breakdown is the goal of this approach. The likelihood value, analogous to the error sum of squares value in multiple regressions, provides a complete valuation of the model's fit. A low likelihood value indicates that a model is well-suited to the data.

#### 4.2.1 Hypothesis Testing

Wald statistic test (Table 5) is used to identify which beta coefficients significantly contribute to the model among the logistic regression coefficients.

Hence, the hypothesis becomes:

$H_0$: $\beta_i = 0$

$H_a$: $\beta_i \neq 0$

However, the assumptions based on chi-square analyses are more powerful and reliable than Wald's test.

#### *4.2.2 Goodness of fit of the model*

Testing the LR model's Goodness of fit is appropriate to reveal whether it delivers an outstanding fit to the data used [18]. The Hosmer Lemeshow test splits the observations (i.e., the cases recommendable to the IKEA store) into each of ten equal groups based on projected likelihoods. Then, it calculates a chi-square from perceived and estimated occurrences. The model's Goodness of fit is verified with the Hosmer Lemeshow statistic. For this research, the computed Hosmer Lemeshow statistic is shown in Table 6.

#### *4.2.3 Models Predictive ability using LR*

From Table 6, it is observed that the Chi-square value for the Hosmer-Lemeshow Test statistic (i.e., 14.335) is less than the value obtained from the $X^2$ distribution table with 8 degrees of freedom and = 0.05 (i.e., 15.507), $H_0$ is accepted, and it is concluded that there is no difference between the perceived and model-driven or fitted values. This advocates that the model predictions are consistent with the observations at α= 0.05. Furthermore, the Hosmer-Lemeshow Test statistic obtained is C = 14.335, and the equivalent p-value received with 8 degrees of freedom is 0.073, which indicates that the model seems pretty well-fitting.

Table 7 shows that the LR model classified 168 respondents as Recommendable out of 168 recommended respondents. Hence, it provides 100% accuracy for the 'Recommend the store' category. Furthermore, the same LR model classified 0 out of 33 respondents as 'Not recommend the store.' Hence, it provides 0% accuracy for the 'Not Recommend the store' category. Overall, the LR model provides **83.6%** accuracy for both 'Recommend the Store' and 'Not Recommend the Store.

### 4.3 Linear Discriminant Analysis

Assumptions such as normality, linearity, and homoscedasticity must be satisfied for discriminant analysis. However, if met, the approach is generally resistant to violating these assumptions [29]. This study assumes that all necessary presumptions are true to apply the discriminant analysis's predictive capability to categorize the applicants.

#### *4.3.1 Measurement of model performance using LDA*

An improvement of at least 25% above the chance accuracy rate must be improved for a discriminant model to be considered "Useful." Wilks' lambda quantifies the model's utility. A smaller p-value suggests the discriminant function is more effective than randomness in separating the groups.

Here, Predictors (i.e., independent variables like brand awareness, visit to the store, customer age, and experience at the shop) are insignificantly discriminating the groups in the Wilks Lambda test with a likelihood of 0.752 (which is greater than the significance level of (0.05)). The percentage of overall variability that cannot be explained, or 95.7 percent, is shown in Table 8.

A canonical correlation of 0.207 in Table 9 indicates that the model expounds 4.284% (i.e., .2072) of the disparity in the grouping variable, i.e., whether the IKEA store is recommendable.

#### *4.3.2 Independent Variables Importance*

One of the most vital tasks is identifying the essential predictor variables in the model development. Table 10 shows a high association between the discriminant model and the predictor variables. The structure matrix indicator shows each variable's associations with each distinguishing function. These structure matrix coefficients are comparable to factor loadings in factor analysis. The sign of the measure shows whether the particular variable positively or negatively impacts the output variable (recommendations). According to the LR model, Customer Age, Store Experience at the store, Brand Awareness, Store Location, and Expectation matches are of higher importance. Even product quality and pricing are less important than the variables mentioned above. Hence, customers do not care about pricing if their expectations match.

#### *4.3.3 LDA Models Predictive Ability*

A significant statistical model with exceptional prediction performance is developed during the development of the discriminant analysis model. Table 11 demonstrates that the discriminant model can categorize 114 of the 168 recommendable responses as "Recommendable Group." For the recommended group, it has a classification accuracy of 67.9 percent. It is possible to categorize 20 of 33 non-recommendable responses as belonging to the "Non-recommendable Group" using the same LDA model. The non-recommendable category is correctly classified with an accuracy rate of 60.6%. As a result, the model can produce a total group classification accuracy of 66.7%.

## 5. FINDINGS AND DISCUSSIONS

Based on the customer's degree of satisfaction, a multilayer perceptron neural algorithm of ANN is developed to estimate the likelihood that the consumer would recommend the IKEA outlet. The model results of Logistic Regression, LDA, and ANN are shown in Table 12 using different predictive measures. It is observed that Logistic regression classifies all responses as a reasonable group. This increases the Type 1 error and is unsuitable for predicting the correct Non-recommendable group. LDA has significantly less accuracy in prediction than the other two methods. The table also shows that ANN is outperforming the other two techniques.

## 6. IMPLICATIONS

### 6.1 Managerial Implications

The models developed in this research work are not only identical to IKEA; the same or similar model could be used for other furniture stores or multinational retail stores to predict customer recommendations and thus improve their market growth and profit further. Retail stores like IKEA could use the automated model specified here to improve their business. The organization's management should implement the research work proposed here to improve its customer base and enhance its marketplace and profit margin.

### 6.2 Academic Implications

This research concentrated on common factors from previous studies to predict customer recommendation status. However, the performance of the ANN model was revealed to be outperforming; other machine learning methods could be applied to study the data and explore more. This same research approach may be used for recommending products in a dynamic real world databases. This model can be extended further in the context of the global retail industry and building efficient retail prediction systems.

### 6.3 Practical implications

It is challenging in the business industry to get appropriate data to conduct research and conclude the general findings. Customer recommendations are the best way to improve the customer base in a business organization, and recommendations from existing customers will attract more loyal customers to the business. Therefore, business organizations must understand the users' views. The ANN model recommended in this research helps the retail stores appraise the significant factors to be considered while implementing the digitalized platform as well as to improve the customer base.

## 7. CONCLUSION

The customer base is significant for any shop/ store or business. As the customer base increases, the profit and the marketplace increase. Recommendations are more critical to the multinational/local retail furniture stores since they bring more customers and enhance the business to the top of the market. Mainly, word-of-mouth recommendations reach more people than advertisements. Cost-wise, it is also beneficial to attract more customers. This research explored how to identify the customer recommendation status by collecting the relevant data from a well-known multinational retail furniture store, like IKEA, and providing the appropriate model to predict the recommendation status of the customer. This research scrutinized the effect of factors like customer age, product quality, store location, expectation match, brand awareness, customer income, product pricing, and so on to predict whether the customer will recommend the store to their neighbours, relatives, and friends. ANN, LDA, and LR machine learning techniques are used in this research to compare the fitness of each model to predict the Recommendation of the retail furniture store and to compare the accuracy produced. This research also envisages the essential predictor variables to be included in the model developed. Hence, it indicates which factors should be considered/ focused on to improve the recommendation status. The models developed in this research reveal that ANN provides superior results over LDA and Logistic regression. From other literature works mentioned in this study, it can be concluded that ANN outperforms LDA and LR techniques, even though it may not be sufficient to draw a universal conclusion that ANN is superior to LDA and Logistic regression regarding predictive abilities, as this study only examines one dataset, namely IKEA customer experience data.

**Limitations and Future Research:**

This research has been done using the data collected from the IKEA store in Hyderabad. Similar research could be extended by collecting data from other retail stores, and the performance could be compared. Academic researchers can include more variables to predict customer recommendations to improve the performance of the proposed models. They could build models with other machine learning techniques to improve the performance

## REFERENCES


[1] Krishna Naik, C. N., Gantasala, S. B., & Prabhakar, G. v. (2010). Service Quality (Servqual) and its effect on customer satisfaction in retailing. European Journal of Social Sciences, Vol. 16, No. 2.

[2] Goswami, P., & Mishra, M. S. (2009). Would Indian consumers move from kirana stores to organized retailers when shopping for groceries? Asia Pacific J. of Marketing and Logistics, 21, No. 1. https://doi.org/10.1108/13555850910926281

[3] BS, R. (2013). IKEA's first India store not likely before the end of 2016. Business Standard. http://www.business-standard.com/article/companies/ikea-s-first-india-store-not-likely-before-the-end-of-2016-113101400773_1.html

[4] Srivastava, R. K. (2008). Changing retail scene in India. Int. J. of Retail and Distribution Management, Vol. 36, No. 9. https://doi.org/10.1108/09590550810890957

[5] McNamara, T., & Descubes, I. (2016). Can IKEA adapt its service experience to India? Emerald Emerging Markets Case Studies, Vol. 6, No. 1. https://doi.org/10.1108/EEMCS-10-2014-0245

[6] Edvardsson, B., & Enquist, B. (2011). The service excellence and innovation model: Lessons from IKEA and other service frontiers. Total Quality Mgt. and Business Excellence, Vol. 22, No. 5. https://doi.org/10.1080/14783363.2011.568242

[7] Das, G. (2014). Factors affecting Indian shoppers attitude and purchase intention: An empirical check. Journal of Retailing and Consumer Services, Vol. 21, No. 4. https://doi.org/10.1016/j.jretconser.2014.04.005

[8] Ghosh, P., Tripathi, V., & Kumar, A. (2010). Customer expectations of store attributes: A study of organized retail outlets in India. J. of Retail and Leisure Property, Vol. 9, No. 1. https://doi.org/10.1057/rlp.2009.27

[9] Li, Y., & Fleyeh, H. (2018). Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning. 2018 Int. Conf. on Computer and Applications, ICCA 2018. https://doi.org/10.1109/COMAPP.2018.8460277

[10] Chan, P., Finnegan, C., & Sternquist, B. (2011). Country and firm level factors in international retail expansion. European Journal of Marketing, Vol. 45, No.6. https://doi.org/10.1108/03090561111119985

[11] Enquist, B., Edvardsson, B., & Petros Sebhatu, S. (2007). Values-based service quality for sustainable business. Managing Service Quality, Vol. 17, No. 4. https://doi.org/10.1108/09604520710760535

[12] Dabholkar, P. A., Thorpe, D. I., & Rentz, J. O. (1996). A measure of service quality for retail stores: Scale development and validation. Journal of the Academy of Marketing Science, Vol. 24, No. 1. https://doi.org/10.1007/bf02893933

[13] Wulfert, T. (2019). Mobile App Service Quality Dimensions and Requirements for Mobile Shopping Companion Apps. Junior Management Science, Vol. 4, No. 3, pp. 339-391.

[14] Zhang, M., He, X., Qin, F., Fu, W., & He, Z. (2019). Service quality measurement for omni-channel retail: scale development and validation. Total Quality Management and Business Excellence, 30(sup1). https://doi.org/10.1080/14783363.2019.1665846

[15] Hudik, M., Karlíček, M., & Říha, D. (2023). Do consumers appreciate promotional gifts? Journal of Consumer Marketing, Vol. 40, No. 3, pp. 305-314.

[16] Baumann, C., Burton, S., Elliott, G., & Kehr, H. M. (2007). Prediction of attitude and behavioral intentions in retail banking. International Journal of Bank Marketing, Vol. 25, No. 2. https://doi.org/10.1108/02652320710728438

[17] Padmavathi, J. (2012). Logistic regression in feature selection in data mining. International Journal of Scientific & Engineering Research, Vol. 3, No. 8, pp. 1-4.

[18] Ogunsakin, R. R., & Adebayo, A. B. (2014). Performance of Logistic Regression in Tuberculosis Data. International Journal of Scientific and Research Publications, Vol. 4, No. 9, p. 1.

[19] Sharma, A., & Kumar Panigrahi, P. (2011). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications, Vol. 27, No. 11. https://doi.org/10.5120/3344-4605

[20] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55. https://doi.org/10.1016/j.simpat.2015.03.003

[21] Baykal, B., & Hesapci Karaca, O. (2022). Recommendation matters: how does your social capital engage you in eWOM? Journal of Consumer Marketing, Vol. 39, No. 7, pp. 691-707.

[22] Bishop, C. (2006). Participation. London: Whitechapel/Cambridge.

[23] Zakaryazad, A., & Duman, E. (2016). A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing. Neurocomputing, 175(PartA). https://doi.org/10.1016/j.neucom.2015.10.042

[24] Zhu, Y., Xie, C., Sun, B., Wang, G. J., & Yan, X. G. (2016). Predicting China's SME credit risk in supply chain financing by logistic regression, artificial neural network and hybrid models. Sustainability (Switzerland), Vol. 8, No. 5, 433. https://doi.org/10.3390/su8050433

[25] Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7, 179–188

[26] Alabi, M. A., Issa, S., & Afolayan, R. B. (2013). An application of artificial intelligent neural network and

discriminant analyses on credit scoring. Math Theory Model, Vol. 3, No. 11, pp. 20-28.

[27] Vesel, P., Orel, P. F., & Špende, M. (2012). Using multichannel marketing activities to build customer relationships: As pertaining to the consumer electronics retail sector. In Service Management: The New Paradigm in Retailing, Vol. 9781461415541. https://doi.org/10.1007/978-1-4614-1554-1_8

[28] Sagar Sharma. (2017). Activation functions in neural network. Towardsdatascience.com

[29] Meyers, L. S., Gamst, & A.J. Guarino, A. J. (2016). Applied Multivariate Research: Design and Interpretation (Third Edition). Sage Publication

❖❖❖❖❖