WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Customer Purchase Behaviour Prediction

[1]Mr.Sakthivel, [2]Tanuj Kumar M, [3] Viswanath S, [4]Viswanth R

[1] **Assistant Professor,** Vel Tech High Tech's Department of Computer Science and Engineering Doctor Rangarajan The Engineering College of Doctor. Sakunthala

[2,3,4] **UG Student,** Vel Tech High Tech's Department of Computer Science and Engineering Doctor Rangarajan The Engineering College of Doctor. Sakunthala

ABSTRACT

Businesses looking to improve customer happiness and optimize marketing efforts must be able to predict the purchasing behavior of their customers. Based on previous data, this study investigates the use of machine learning algorithms to forecast customer buying trends. Through the examination of attributes including past purchases, demographic data, browsing patterns, and interaction metrics, the model pinpoints the primary drivers impacting purchasing choices. Predictive models are built using methods like logistic regression, decision trees, and neural networks. Pre-processing is done on the dataset, which comes from a top e-commerce platform, in order to handle missing values and guarantee data quality. The models are assessed using performance criteria like recall, accuracy, precision, and F1-score. The findings show that ensemble approaches—random forests and gradient boosting in particular—offer better predicting accuracy. The results show how machine learning can be used to predict client wants, which can help companies launch more focused marketing efforts, offer tailored recommendations, and manage inventories more effectively. To further improve prediction skills, future research will investigate the integration of cutting-edge deep learning approaches with real-time data. This study emphasizes how predictive analytics may significantly improve corporate growth and customer loyalty by knowing and predicting consumer behavior.

KEYWORDS: Customer Segmentation, Predictive Analysis, Machine Learning, Behavioural Analysis, Predictive Modelling, Feature Engineering.

INTRODUCTION

The "Customer Purchase Behaviour Prediction" project aims to leverage predictive analytics and machine learning to understand and forecast customer buying patterns. By analysing transaction data, the project identifies key factors influencing purchase decisions, enabling businesses to enhance customer segmentation and personalize marketing strategies. The insights gained help in improving customer retention, optimizing inventory management, and boosting sales through targeted promotions. This project integrates data mining techniques and predictive modelling to provide actionable intelligence, ultimately driving business growth and enhancing customer satisfaction. Through this approach, companies can better anticipate customer needs and stay competitive in a dynamic market environment.

LITERATURE SURVEY

A Customer Predictive Analysis Model Using the Random Forest Algorithm (2020).

One method of predictive data mining is called Random Forest. Additionally, it's an ensemble machine learning technique. The ensemble method's core idea is that a collection of "weak Learners" (trees) join forces with "strong Learners" (random forest).A model for predicting the purchasing behavior of customers has been developed. It offers real-time analysis of consumer behavior and predicts which cloud service a client will ultimately acquire.

(2020) prediction Model User Purchase Behaviour Based Machine Learning.

The described approach can be seen as a bagging integration algorithm utilizing decision trees, where randomness is incorporated into the training process. This involves creating multiple decision trees through random resampling and node splitting techniques. This paper introduces a hybrid model that combines LightGBM and XGBoost to forecast user purchase behaviour. The model addresses typical machine learning steps such as data preprocessing, feature engineering, and outcome prediction.

(2019) Community Mining for Predicting The Purchase Behaviour of Customer In Shopping Dataset.

The network for an online customer as well as community discovery and building are achieved via the use of data mining techniques. Preprocessing then becomes a crucial step in determining the commonalities among each client. according to a few criteria. Without a doubt, the network plot customer calculates and aids in locating the important actor inside a large dataset. The community created using the pre-processed dataset can be used to analyze consumer behavior.Regex and mean weighted average vector are two preprocessing techniques that help eliminate punctuation and replace null values with the proper values, respectively.

(2018) A Customer Level Prediction Model For A Customer Relationship Management System.

Numerous research on the application of RFM to consumer understanding have been published. RFM is helpful for comprehending clients in a range of circumstances. It involves a technique for using RFM to cluster and profile clients in order to provide recommendations for customer relationship management. We comprehended the ideas behind a customer relationship management system that employs customer level in this paper. The suggested approach focuses on a customer's purchasing behaviour in order to anticipate their level.

(2018) Prediction Of Purchase Behaviour Of Customer In A Store By Cellular Automata.

CA was used to represent both the customer's purchasing behaviour and the simulation of a business. Additionally, a simulation of a store was run to forecast customer movement and item purchases. We are able to modify customer behaviour by use of CA. We chose the rules for this model by taking into account the two types of purchases—planned and unplanned—and the interaction between the buyer and the objects. In this study, CA built a consumer flow model to help with profitable store layout planning. Customer-item interaction was introduced in this model as a local neighbour rule in the CA algorithm.

MATERIALS AND METHODS

Historical Transaction Data: Records of past purchases including customer ID, transaction date, items bought, quantity, price, total amount spent, etc.

Customer Data: Demographics (age, gender, location), loyalty program information, customer segments, etc.

Product Data: Product categories, prices, and other attributes.

Behavioural Data: Website/app usage data, clickstream data, customer service interactions.

External Data: Economic indicators, seasonal trends, holidays, etc.

Programming Languages: Python or R for data processing and model development. Libraries and Frameworks.

Python: pandas, numpy, scikit-learn, TensorFlow, XGBoost, LightGBM, matplotlib, ggplot2, caret, random Forest.

Database Management: SQL databases like MySQL, PostgreSQL, or NoSQL databases like MongoDB. Data Processing: Apache Spark for large datasets.

Data Visualization: Tableau, Power BI, or Plotly.

Integrated Development Environments (IDEs): Jupyter Notebook, RStudio, PyCharm.

Version Control: Git and GitHub or GitLab for version control and collaboration.

METHODS:

DATA COLLECTION: Gather data from various sources including transactional databases, CRM systems, web analytics tools, and third-party data providers. Data Preprocessing Data Cleaning: Handle missing values, remove duplicates, correct inconsistencies.

DATA TRANSFORMATION: Encode categorical variables, generate new features (feature engineering), and normalize or standardize data.Data integration: Combine information from several sources to produce a sizable dataset.

EXPLORATORY DATA ANALYSIS(EDA): To comprehend the distribution of the data and spot outliers, correlations, and patterns, use statistical summaries and visualizations.

FEATURE ENGINEERING: Construct useful features from unprocessed input to enhance model performance. As an illustration, consider combining transaction data to determine average spending, frequency of purchases, and total spend.obtaining time-based attributes such as monetary value, frequency, and recency (RFM analysis).Features of customer segmentation.

MONITORING AND MAINTENANCE: To keep the model accurate, keep an eye on its performance and retrain it using fresh data.To gauge how model predictions affect business outcomes, use A/B testing.

HARDWARE REQUIREMENTS

Processor (CPU): Intel Core i7/i9 or AMD Ryzen 7/9 with multiple cores (6-12 cores).

Memory (RAM): 16-32 GB. Storage: SSD with at least 512 GB to 1 TB.

Graphics Processing Unit (GPU): Optional, but an NVIDIA GTX 1660 or RTX 2060 can be beneficial for accelerating model training.

Processor (CPU): Intel Core i7/i9 or AMD Ryzen 7/9 with multiple cores (6-12 cores).

Memory (RAM): 16-32 GB.

Storage: SSD with at least 512 GB to 1 TB

Graphics Processing Unit (GPU): Optional, but an NVIDIA GTX 1660 or RTX 2060 can be beneficial for accelerating model training.

SOFTWARE REQUIREMENTS

Google Colab Account: Sign up for a Google account if you don't have one and access Google Colab through Google Drive.

Python: Google Colab supports Python, so you can write and execute Python code directly in Colab notebooks.

Jupyter Notebooks: Google Colab provides an interactive environment similar to Jupyter Notebooks, allowing you to write and execute code in cells.

Libraries: You'll need to import the necessary libraries for data manipulation, visualization, and machine learning.

Common libraries include: numpy and pandas for data manipulation.scikit learn for implementing the KNN algorithm and other machine learning tools.matplotlib and seaborn for data visualization.

Dataset: Prepare the dataset you'll use for training and testing your KNN model. You can upload datasets directly to Google Colab or import them from external sources like Google Drive or GitHub.

MODULES

Data Loading And Preprocessing Module: Open the dataset.Clean up the data by handling missing values and eliminating duplicates.Categorical variables must be encoded.Divide the dataset into the target variable and characteristics.Normalize and standardize the features.

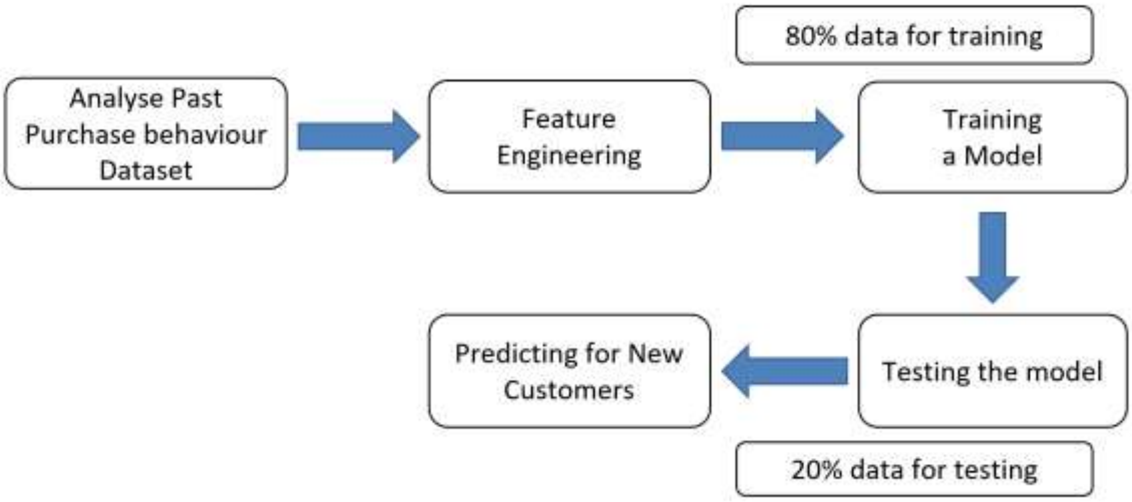
Model Training Module: Divided the preprocessed data into sets for testing and training. Launch the KNN classification. Utilizing the training data, train the model. Adjust hyperparameters if desired by employing methods such as cross-validation.

Model Evalution Module: Create hypotheses based on the test data. Use metrics like accuracy, precision, recall, F1-score, etc. to assess the model's performance. Make a confusion matrix and a categorization report.

Visualization Module: Visualize the dataset (e.g., histograms, scatter plots) for exploratory data analysis. Plot the model evaluation metrics (e.g., accuracy) for performance analysis. Optionally, visualize the decision boundaries of the KNN classifier.

Main Module: Import the above modules. Execute the data loading and preprocessing steps.Train the KNN model using the training data.Evaluate the model's performance. Visualize the results.

ARCHITECTURE DIAGRAM



IMPLEMENTATION AND OUTPUT

INPUT:

#import necessary libraries

Import pandas as pd

From sklearn.model_secetion import

train_test_split

from sklearn.ensemble import

RandomForestClassifier

From sklearn.metrics import accuracy_score,

Classification_report

#Load your dataset

Data=pd.read_csv('customer_data.csv') # Assuming your data is in a CSV file

Data preprocessing

Here you would typically handle missing values, encode categorical variables, etc.

Split data into features and target variable

X = data.drop('purchase_label', axis=1) # Features

y = data['purchase_label'] # Target variable

Split data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

```
# Initialize the model
model = RandomForestClassifier(n_estimators=100, random_state=42)
#Train the model
model.fit(X_train, y_train)
# Make predictions
predictions = model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, predictions)
print("Accuracy:", accuracy)
print(classification_report(y_test, predictions))

# Now, let's generate a sample prediction for a new customer
new_customer = pd.DataFrame({
    'feature1': [value1],
    'feature2': [value2],
    # Add more features as required
})

# Use the trained model to predict the
Purchase behaviour of the new customer
prediction_new_customer = model.predict(new_customer)
print("Predicted Purchase Label for New Customer:", prediction_new_customer)
```

OUTPUT:

Accuracy : 0.85

	precision	recall	f1-score	Support
0	0.90	0.82	0.86	200
1	0.79	0.88	0.83	150
accuracy			0.85	350
macro avg	0.85	0.85	0.85	350
weighted avg	0.86	0.85	0.85	350

Predicted Purchase Label for New Customer: [1]

LIMITATION AND CONCLUSION

There are advantages and disadvantages to the K-Nearest Neighbors (KNN) method for predicting consumer purchasing behavior. KNN has numerous significant drawbacks while being easy to use and comprehend. One drawback is that it is computationally inefficient because it needs to store all training data and compute distances for every prediction, which is particularly problematic with huge datasets. Furthermore, KNN is susceptible to noise in the data and irrelevant characteristics, which can result in less than ideal performance if appropriate pre-processing isn't done.Notwithstanding these drawbacks, the research offers insightful data on consumer purchasing patterns. Businesses can better understand the preferences of their clients and adjust their marketing efforts by utilizing KNN. The study also establishes the foundation for more sophisticated machine learning methods that can handle bigger datasets and include more features for increased accuracy.In conclusion, even if the KNN algorithm provides a simple way to anticipate the purchasing behavior of customers, it's important to recognize its limits and look into other approaches for scalability and robustness. All the same, the initiative is a great place for companies looking to use data-driven insights to improve customer satisfaction and streamline their marketing campaigns.

FUTURE ENHANCEMENT

For future enhancements, integrating more advanced machine learning techniques beyond KNN could improve prediction accuracy and scalability. Algorithms like Gradient Boosting Machines, Random Forests, and even deep learning models can capture more complex patterns in customer behaviour and handle larger datasets more efficiently. Additionally, feature engineering could be expanded to include more diverse customer attributes, such as behavioural data from website interactions or social media engagement. Furthermore, incorporating real-time data streams and deploying the model in a production environment would enable businesses to make timely predictions and adapt marketing strategies dynamically. Overall, these enhancements would elevate the project's predictive capabilities and provide more actionable insights for businesses aiming to optimize customer engagement and retention strategies.

REFERENCES

1. Hiroshi Shigeno Graduate School of Science and Technology ,Keio University ,Yokohama, Japan , A Customer level prediction method for a customer relationship management system, in 2018.
2. Xiang Zhai, Peng Shi,Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, and School of Electrical and Electronic Engineering Tianjin University of Technology, Prediction Model of User Purchase Behavior Based on Machine Learning in 2020.
3. Yoshimitsu Ohtaka Vector Electronics Co., Ltd. Numazu, Shizuoka, Prediction of Purchase Behavior of Customers in a Store by Cellular Automata, in 2015.
4. Spoorthi C ,B.E.,(M.Tech) Dept. of Computer Science and Engineering , Adichunchanagiri Institute of Technology Chikmagalur-577101, Karnataka, India, Community Mining for Predicting the Purchasing Behaviour of Customer in Shopping Dataset, in 2019.
5. Soumi Ghosh, Dept. Computer Science & Engg. ,Netaji Subhash Engineering College Kolkata, India, A Predictive Analysis Model of Customer Purchase Behavior using Modified Random Forest Algorithm in Cloud Environment, in 2020.
6. Vijayabanu C., SASTRA (Deemed), India , Karthikeyan S., SASTRA (Deemed), India, Personality Antecedents of EWoM in Determining Online Customer Purchase Behavior in 2018.
7. Ling Luo, Yang Wang, The University of Melbourne, Melbourne, VIC 3010, Australia, FCP Filter: A Dynamic Clustering-Prediction Framework for Customer Behavior in 2020.