World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Targeted customer classification using machine learning techniques

Mrs. Vijayajyothi Chiluka, Pravalika Bandi *, Pranathi Enumula, Laxmi Sowjanya Korvi and Varun Teja Seelam

Department of CSE (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

World Journal of Advanced Research and Reviews, 2025, 25(02), 174-183

Publication history: Received on 23 December 2024; revised on 01 February 2025; accepted on 03 February 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.2.0334

Abstract

Targeted Customer Classification Using Machine Learning Techniques aims to optimize marketing strategies through data-driven customer segmentation and personalization. Traditional models, often limited by simplistic attributes like income, fail to capture the nuanced behaviors and preferences of diverse customer bases, resulting in static clusters and generalized offers that do not adapt to evolving consumer trends. To address these challenges, this project proposes an advanced model leveraging machine learning for dynamic and precise customer classification. The system employs KMeans clustering to segment customers into distinct groups based on demographic attributes such as age and spending patterns. Following this, a Random Forest classifier is used to predict and classify new customers into these predefined clusters. By tailoring personalized offers aligned with each cluster's characteristics, the system enhances marketing efficiency, improves customer engagement, and drives sales. Implemented using Python-based frameworks and integrated with a user-friendly interface, the model supports seamless data input and real-time predictions. This approach empowers businesses to gain actionable insights and make more informed decisions. Future enhancements include integrating real-time data processing, expanding feature sets, and exploring deep learning techniques to improve segmentation and prediction accuracy further.

Keywords: Customer Segmentation; Machine Learning; KMeans Clustering; Random Forest Classifier; Personalized Marketing

1. Introduction

In today's competitive business environment, understanding customer behavior is crucial for personalized marketing and effective decision-making. Targeted Customer Classification using Machine Learning enables businesses to segment customers based on their purchasing patterns, demographics, and behavioral attributes. This approach helps in identifying high-value customers, optimizing marketing strategies, and enhancing customer engagement.

Traditional customer segmentation methods rely on predefined rules, which may not capture complex relationships in the data. Machine Learning (ML) techniques, such as clustering and classification, provide a data-driven approach to segment customers dynamically. Clustering algorithms, like K-Means, group customers with similar characteristics, while classification models, such as Random Forest, predict segment membership for new customers based on input features.

This study applies K-Means clustering to categorize customers into distinct groups based on age and spending score, followed by Random Forest classification to predict customer segments for new inputs. The proposed model enables businesses to classify potential customers into targeted groups and provide personalized offers, improving customer retention and revenue generation.The paper discusses the dataset selection, feature engineering, model selection, and evaluation metrics used in the classification process. Experimental results demonstrate the effectiveness of the

* Corresponding author: Pravalika Bandi

proposed model in accurately segmenting customers, making it a valuable tool for businesses looking to enhance their marketing strategies.

Furthermore, the integration of machine learning in customer classification enhances decision-making by leveraging data-driven insights rather than subjective assumptions. By analyzing key features such as age and spending score, businesses can create tailored marketing campaigns, optimize product recommendations, and improve customer satisfaction. The implementation of this approach using Flask ensures user-friendly interaction, allowing businesses to input customer details and receive instant segment predictions. This research highlights the significance of ML-driven customer classification in modern business strategies, showcasing its potential to revolutionize targeted marketing and customer relationship management.

2. Related Work

The literature on customer classification and segmentation using Machine Learning (ML) highlights the growing importance of data-driven marketing strategies. Foundational studies in customer segmentation, such as those by Wedel and Kamakura in *Market Segmentation: Conceptual and Methodological Foundations*, establish the significance of clustering techniques in consumer analytics. These studies emphasize the role of data mining in identifying purchasing patterns and behavioral trends, which form the basis for modern classification models.

Building on these principles, recent research by Bhatia and Sinha explores advanced segmentation methodologies, including K-Means clustering and hierarchical clustering, to analyze large-scale consumer datasets. Their work underscores the effectiveness of unsupervised learning techniques in forming meaningful customer groups based on purchasing behaviors and demographic attributes. The introduction of classification models, such as Decision Trees and Random Forest, has further enhanced customer prediction capabilities, allowing businesses to tailor personalized marketing strategies effectively.

Additionally, Huang et al. examine the impact of feature selection and dimensionality reduction in customer segmentation. Their research highlights how Principal Component Analysis (PCA) and t-SNE improve clustering performance by eliminating redundant features, thus optimizing computational efficiency. Similarly, Singh et al. focus on supervised learning models, analyzing how classification algorithms, including Support Vector Machines (SVMs) and Gradient Boosting, improve segmentation accuracy in real-world applications. These studies validate the role of predictive analytics in customer classification, enabling precise targeting and enhanced decision-making.

Further advancements in ML-based customer classification integrate deep learning techniques and hybrid models. Researchers such as Chen and Liu investigate Convolutional Neural Networks (CNNs) and Deep Autoencoders for feature extraction, significantly improving segmentation quality in complex datasets. The application of transfer learning and ensemble methods, including XGBoost and LightGBM, has addressed challenges related to class imbalance and model interpretability. These developments contribute to the robustness and adaptability of targeted customer classification systems, paving the way for real-world applications in retail, finance, and e-commerce.

Together, these studies provide a comprehensive understanding of ML-based customer classification, offering insights into algorithm selection, implementation, and optimization. This research contributes to advancing intelligent marketing systems, enabling businesses to enhance customer engagement, optimize resource allocation, and drive data-driven decision-making with improved accuracy and real-time adaptability.

3. Existing System

Traditional customer classification relies on rule-based segmentation and statistical methods like RFM analysis, which lack adaptability to dynamic customer behavior. Machine Learning (ML) techniques, such as K-Means clustering and Decision Trees, have improved segmentation but still face challenges like rigid predefined rules and broad categorization.

Most existing systems process data in batches, limiting real-time classification and personalized targeting. Additionally, imbalanced datasets can bias predictions, reducing accuracy. While ML-based methods enhance segmentation, they still struggle with adaptability, real-time processing, and fine-grained personalization, requiring more advanced approaches for better results.

4. Proposed Model

The proposed Targeted Customer Classification system enhances segmentation by integrating K-Means clustering and Random Forest classification to improve accuracy and adaptability. Unlike traditional models, this approach dynamically groups customers based on age and spending score, forming precise customer segments.

With a real-time classification system powered by Flask, users can input customer details and receive instant segment predictions. The model addresses challenges like imbalanced data and broad categorization by ensuring well-distributed clusters and refining feature selection. By leveraging machine learning techniques, the system provides personalized marketing insights, improving customer engagement and business decision-making.

5. Methodology

The methodology for the Targeted Customer Classification system focuses on data collection, feature extraction, model building, and real-time classification. The process ensures that the system provides precise, adaptable, and personalized results for customer segmentation.

Below is a detailed explanation of each step involved:

Importing libraries

```
# for Data manipulation and analysis
import pandas as pd
import numpy as np

# for Data visualization
import matplotlib.pyplot as plt
import seaborn as sns

# for Machine learning algorithms and tools
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import OneHotEncoder
```

Figure 1 Libraries imported

5.1. Data Collection

Data collection forms the foundation of a robust classification model. The system utilizes customer data, such as age, spending score, and other demographic details, to create accurate customer profiles.

5.1.1. Dataset Selection:

The system uses datasets containing customer features, including age, spending score, and similar attributes to form distinct customer clusters.

5.1.2. Data Preprocessing:

Preprocessing includes handling missing values, normalizing numerical data, and encoding categorical variables. Feature scaling ensures uniformity across attributes for better model performance.

5.2. System Architecture

The Targeted Customer Classification system architecture illustrates the flow of data through various processing components, enabling personalized recommendations based on user input such as age and spending score.

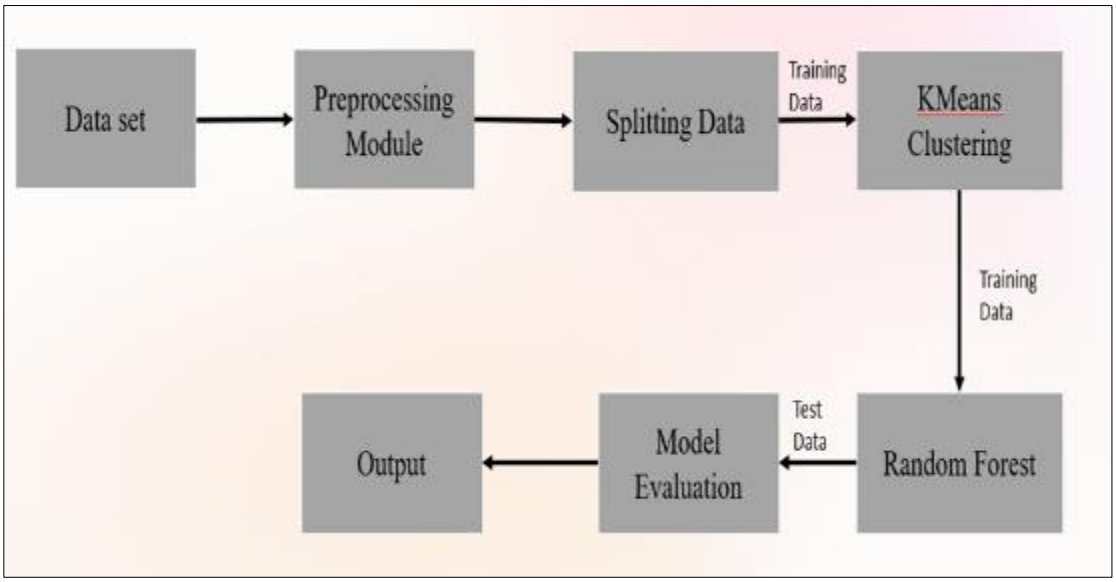

Figure 2 System Architecture

5.2.1. System Architecture Components

- **Input Module:**The system accepts user inputs (age and spending score) through a user interface built using Flask. This allows customers to input their details to receive personalized recommendations.
- **Customer Grouping Module:**The system processes customer data using an algorithm to group users based on similar features (age and spending score). The system assigns customers to one of the predefined segments, which helps in determining the type of offers they are eligible for.
- **Prediction Module:**This module uses Random Forest to evaluate the relationship between the user's features (age and spending score) and the resulting customer categories. The system determines which category the customer most closely aligns with based on the given inputs.
- **Recommendation Module:**Once the customer is assigned to a group, the system generates tailored offers based on predefined criteria for that group.
- **User Interface (UI):**The recommendations are displayed to the user in an easy-to-understand format via the UI.
- **Output:**After the prediction is made, the system shows the recommended offers in real-time through the UI.

5.3. Model Development

The Model Development phase focuses on building a robust model capable of grouping customers into meaningful segments and predicting personalized offers based on their behavior.

5.3.1. Data Preparation:

The dataset includes customer information such as age and spending score. Data preprocessing is performed to ensure consistency and quality.

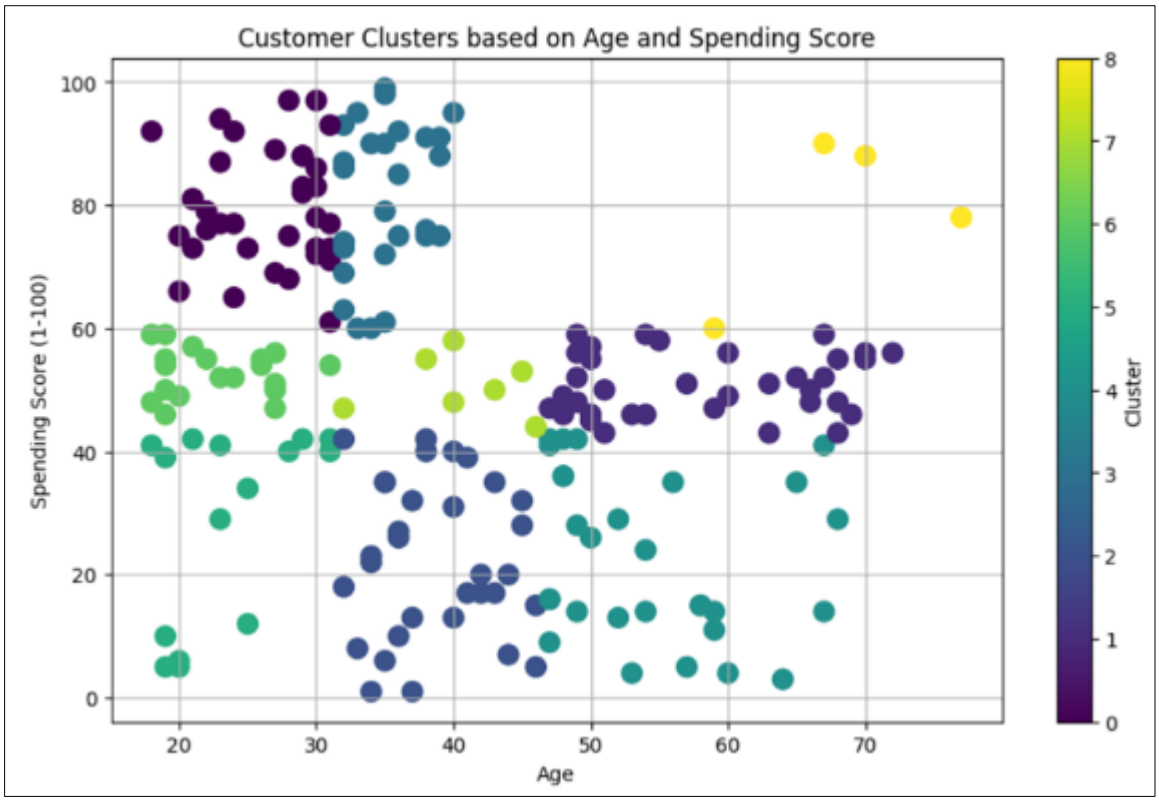

Figure 3 KMeans Clustering

5.3.2. Model Selection

K-Means Algorithm an unsupervised algorithm is used for grouping customers based on similarity in their age and spending score. Random Forest , after grouping customers into segments, the Random Forest model is applied to predict the customer's group based on new input data.

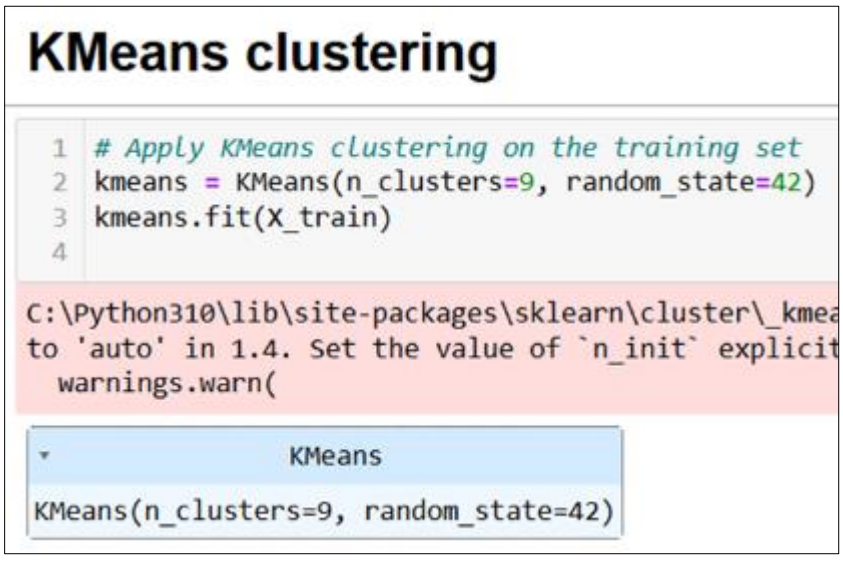

Figure 4 KMeans Clustering

5.3.3. Training:

The training dataset is split into subsets (training, validation, and test sets). The model is trained on the training set and evaluated using the validation set. Cross-validation ensures that the model generalizes well to new, unseen data. Metrics such as accuracy, precision, and recall are used for evaluation.

Figure 5 Model Training

5.3.4. Testing

After training, the model is evaluated on the test set to ensure its effectiveness in predicting new customer segments. Confusion Matrix and other evaluation metrics are used to assess how well the model is performing.

5.4. Real-Time Data Processing

This step ensures the system operates effectively in a real-time environment.

5.4.1. Input Capture

The user enters their age and spending score through a simple form in the Flask UI. This real-time input is processed immediately by the system.

5.4.2. Data Flow

Upon submission, the input data is processed in real-time, The Prediction Module receives the input and feeds it into the trained Random Forest model for immediate classification, The system quickly processes the data and determines the customer's assigned group (segmentation).

5.4.3. Output Processing

The Prediction Module generates the output (predicted group), which is passed to the Recommendation Module. Offer Recommendations are generated based on predefined criteria for the respective group.

5.5. Visualization

Delivering results in a user-friendly manner is crucial for system usability.

5.5.1. Real-Time Display

Developed a graphical user interface (GUI) using tool like Flask The UI includes fields where the user can enter their age and spending score. Upon submitting this information, the system will process the data and display the predicted group and recommended offers in real time.

6. Results and Discussion

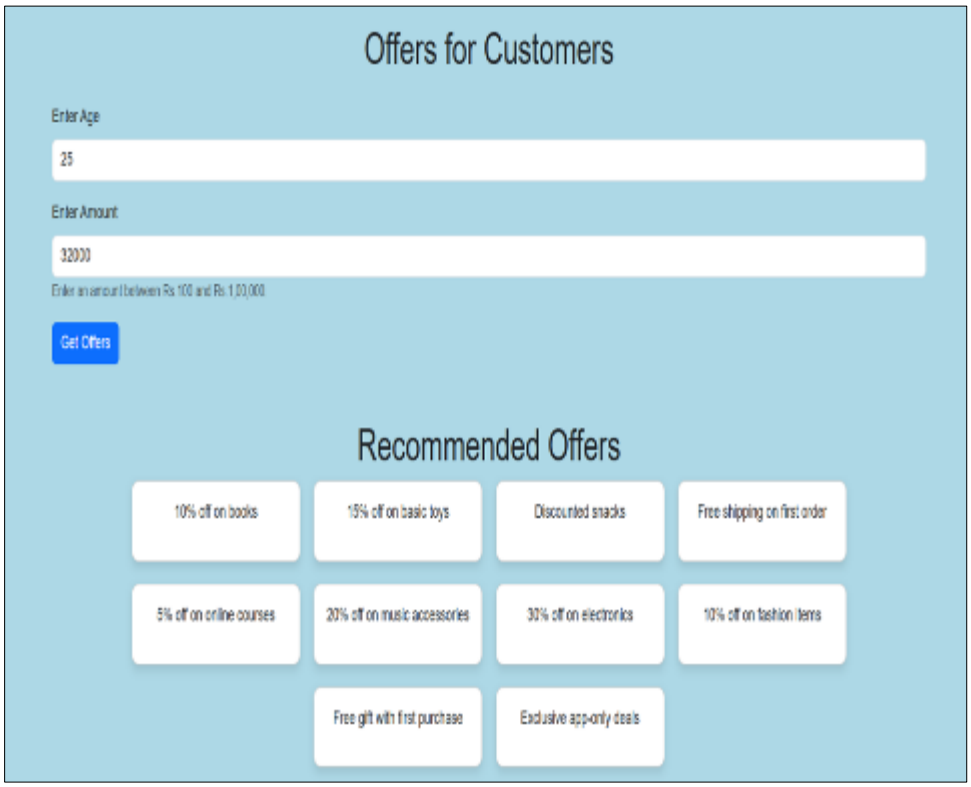

Figure 6 Young Age And Low Spender

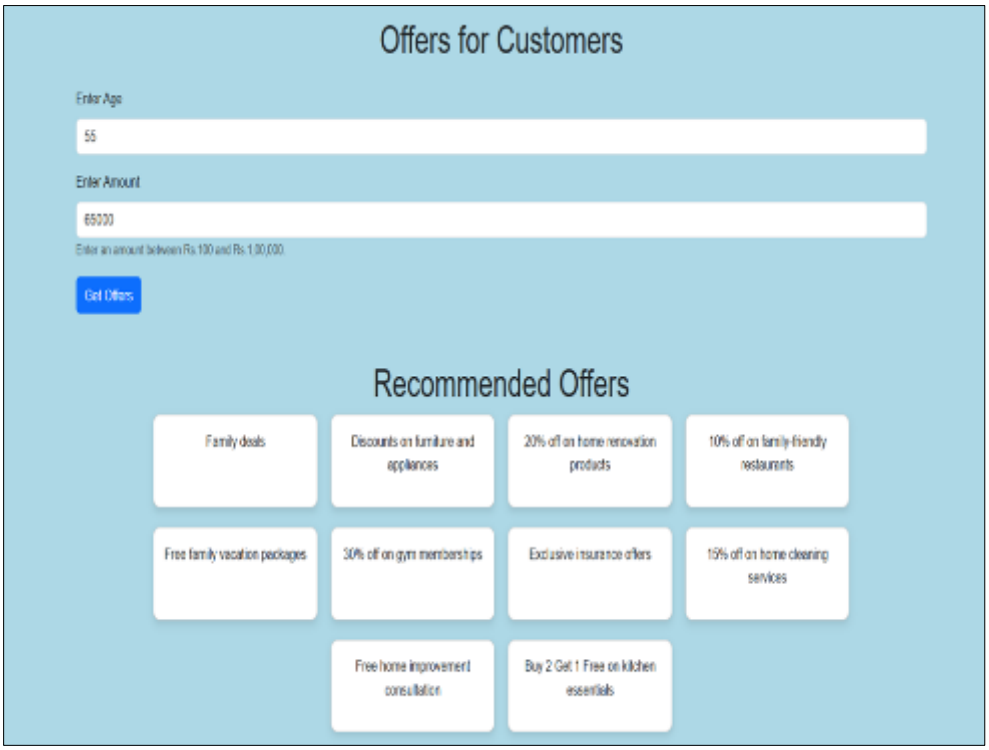

Figure 7 Middle Age and Medium Spender

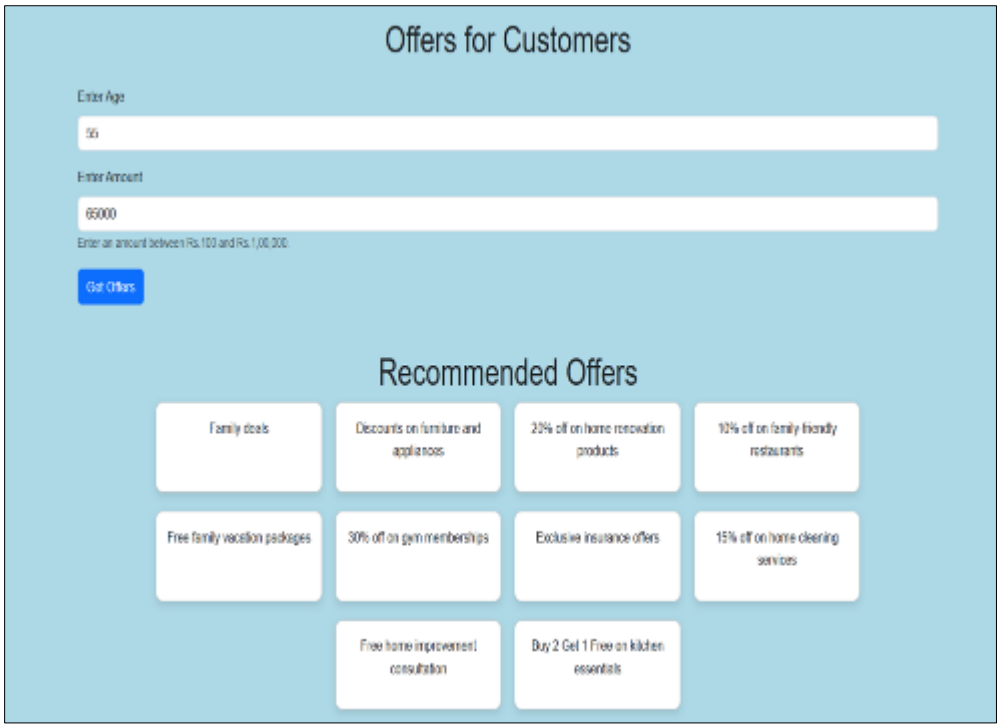

Figure 8 Old age and medium spender

7. Conclusion

In conclusion, the Targeted Customer Classification Using Machine Learning Algorithms project provides an effective solution for segmenting customers based on their age and spending score, enabling businesses to deliver personalized recommendations that enhance customer engagement. By leveraging KMeans clustering for segmentation and Random Forest for classification, the system accurately categorizes customers into meaningful groups, allowing businesses to target specific customer segments with tailored offers. The integration of a user-friendly Flask interface ensures smooth interaction making the system intuitive and efficient. This project highlights the potential of machine learning in optimizing marketing strategies, improving customer experiences, and driving business growth by enabling data-driven decision-making.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] CUSTOMER SEGMENTATION USING MACHINE LEARNING AMAN BANDUNI , Prof ILAVENDHAN A.School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P.

[2] Blanchard, Tommy. Bhatnagar, Pranshu. Behera, Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities.

[3] Aazmaan Ahmed Khan. Tuijin Jishu/Journal of Propulsion Technology Customer Segmentation using Machine Learning Techniques R. Amutha1 y ISSN:1001-4055(2023)

[4] *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* by Aurélien Géron: A comprehensive guide to machine learning algorithms, including KMeans and Random Forests, with practical examples.

[5] *A survey of customer segmentation techniques"* by Anand et al. (2012).

[6] *A Study on Customer Segmentation and Its Application in Retail Marketing"* by Rahul M. Gawande and Dr. S. A. Dakhale (2019): Focuses on clustering techniques for customer segmentation in retail and marketing applications.

[7] *Customer Segmentation Using Clustering Techniques"* by G. K. Gupta and A. K. Singh (2018): Discusses the application of clustering techniques like KMeans for segmenting customers based on demographic and behavioral data.

[8] KUMAR, P. ASHOK, GSatish KUMAR, and SETTI NARESH KUMAR. "Improve the Capacity of Uniform Embedding for Efficient JPEG Steganography Based on DCT." (2015).

[9] Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In *XVIII International Conference on Data Science and Intelligent Analysis of Information*, pp. 13-26. Cham: Springer Nature Switzerland, 2023.

[10] Sunkavalli, Jayaprakash, B. Madhav Rao, M. Trinath Basu, Harish Dutt Sharma, P. Ashok Kumar, and Ketan Anand. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." In *2024 International Conference on Computing, Sciences and Communications (ICCSC)*, pp. 1-5. IEEE, 2024.

[11] Kumar, P. Ashok. "Event Based Time Series Sentiment Trend Analysis."

[12] PANDI, CHIRANJEEVI, THATIKONDA SUPRAJA, P. ASHOK KUMAR, and RALLA SURESH. "A SURVEY: RECOMMENDER SYSTEM FOR TRUSTWORTHY."

[13] Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Correction to: Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In *XVIII International Conference on Data Science and Intelligent Analysis of Information*, pp. C1-C1. Cham: Springer Nature Switzerland, 2023.

Author's short biography

Mrs. Vijayajyothi Chiluka

Mrs. Ch Vijayajyothi is an Assistant Professor with an M.Tech in Computer Science and Engineering and is pursuing a Ph.D. She has 12 years of professional experience in the field of computer science. Her research interests include Deep Learning and Analysis of Algorithms.

Pravalika Bandi

I am B Pravalika, an undergraduate student pursuing a Bachelor's degree in Computer Science and Engineering with a specialization in Data Science. My research interest lies in Machine Learning, where I explore data-driven techniques and their applications. As an aspiring data science student , I am keen on expanding my knowledge and gaining hands-on experience in the field.

Pranathi Enumula

E Pranathi is an undergraduate student pursuing a Bachelor's degree in Computer Science and Engineering with a specialization in Data Science. Her research interest lies in Artificial Intelligence, where she explores intelligent systems and their real-world applications. She is passionate about learning and applying AI techniques to solve complex problems.

Laxmi Sowjanya Korvi K Laxmi Sowjanya is an undergraduate student pursuing a Bachelor's degree in Computer Science and Engineering with a specialization in Data Science. Her research interest is in Data Science, where she explores data analysis techniques and their applications. She is eager to expand her understanding and contribute to advancements in the field.	
Varun Teja Seelam S Varun Teja is an undergraduate student pursuing a Bachelor's degree in Computer Science and Engineering with a specialization in Data Science. His research interest is in Data Science, focusing on extracting insights from data and developing data-driven solutions. He is keen on enhancing his knowledge and practical skills in this evolving field.