# International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

# A Machine Learning-Based Gender Prediction System Using Consumer Shopping Behavior

**REVANTH AREPALLI**

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

**K. Venkatesh**

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

## ABSTRACT

In recent years, the rapid growth of e-commerce platforms and digital transactions has generated vast amounts of consumer data, enabling businesses to gain deeper insights into customer behavior. This project presents a machine learning-based system designed to predict gender based on shopping behavior patterns. The system utilizes features such as purchase amount, number of items, preferred product category, time of purchase, discount usage, return history, device usage, loyalty membership, payment method, and shipping preferences to build a predictive model. The proposed solution integrates a trained machine learning model with a user-friendly graphical interface developed using Python's Tkinter library. The model is trained on historical shopping datasets and saved using joblib for efficient reuse. The application allows users to input behavioral parameters, which are then processed and fed into the model to predict gender as either male or female. The system aims to assist businesses in personalizing marketing strategies, improving customer targeting, and enhancing user experience. By understanding customer demographics through behavioral data, companies can tailor recommendations, advertisements, and promotions more effectively. The model leverages classification techniques to identify patterns and correlations within the data, ensuring accurate predictions.

The implementation focuses on usability and accessibility by providing a scrollable interface to accommodate multiple input fields. Error handling mechanisms ensure robustness by validating user inputs and preventing incorrect data entry. The system demonstrates how machine learning can be effectively integrated into real-world applications to derive meaningful insights from consumer data. Overall, this project highlights the significance of data-driven decision-making in modern business environments. It showcases the potential of predictive analytics in understanding customer behavior and emphasizes the role of machine learning in transforming raw data into actionable intelligence. Future improvements may include incorporating deep learning models, expanding datasets, and enhancing prediction accuracy.

**KEYWORDS:** Gender Prediction, Machine Learning, Consumer Behavior, Classification, Shopping Patterns, Data Analytics, Predictive Modeling, User Profiling

## I. INTRODUCTION

With the exponential growth of online shopping and digital transactions, consumer behavior analysis has become a crucial aspect of modern business strategies. Companies continuously seek innovative ways to understand their customers better in order to deliver personalized experiences and maximize revenue. One such approach involves predicting demographic attributes, such as gender, based on purchasing patterns. Gender prediction using shopping behavior is an emerging application of machine learning that leverages user interaction data to infer demographic characteristics. Traditional methods of collecting demographic information often rely on surveys or manual input, which may be inaccurate or incomplete. In contrast, machine learning models can analyze behavioral data to uncover hidden patterns and make predictions with higher efficiency and scalability. This project focuses on developing a gender prediction system using machine learning techniques. The system is designed to analyze various behavioral features, including purchase amount, number of items purchased, preferred product categories, time of purchase, and payment methods. These features provide valuable insights into consumer habits and preferences, which can be used to classify users into different gender categories.

The application is implemented using Python, with Tkinter used to create an interactive graphical user interface. The trained model is loaded using joblib, allowing for quick and efficient predictions. The interface is designed to be user-friendly, enabling users to input data easily and receive instant predictions. The importance of this system lies in its practical applications. Businesses can use gender prediction models to enhance targeted marketing campaigns, improve recommendation systems, and optimize product offerings. For example, understanding whether a user is more likely to be male or female can help tailor advertisements and product suggestions, leading to increased customer engagement and satisfaction. Furthermore, this project demonstrates the integration of machine learning models into real-time applications. It highlights the end-to-end process of data preprocessing, model training, deployment, and user interaction. By bridging the gap between theoretical concepts and practical implementation, the project provides a comprehensive understanding of applied machine learning. In conclusion, gender prediction based on shopping behavior represents a powerful tool for businesses seeking to leverage data for strategic decision-making. This project showcases how machine learning can transform consumer data into valuable insights, ultimately contributing to improved business performance and customer experience.

## II. LITERATURE SURVEY (WITH EXISTING METHODS)

Several research studies have explored the use of machine learning techniques for predicting demographic attributes based on user behavior. These studies highlight the growing importance of data-driven approaches in understanding consumer patterns and preferences. One of the early approaches to gender prediction involved the use of statistical methods and rule-based systems. These methods relied on predefined assumptions about user behavior, such as associating certain product categories with specific genders. However, these approaches were limited in their ability to handle complex and dynamic data. With the advancement of machine learning, more sophisticated models have been developed to analyze consumer behavior. Techniques such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM) have been widely used for classification tasks. These models are capable of identifying patterns in large datasets and making accurate predictions based on multiple features. Recent studies have also explored the use of deep learning models, such as Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), for gender prediction. These models can capture complex relationships in data and achieve higher accuracy compared to traditional methods. However, they require large datasets and significant computational resources.

In addition to model selection, feature engineering plays a crucial role in improving prediction accuracy. Researchers have identified various behavioral features, such as purchase frequency, spending patterns, browsing history, and device usage, as important indicators of gender. Combining multiple features allows models to capture a more comprehensive view of user behavior. Another area of research focuses on privacy and ethical considerations in gender prediction systems. Since these models rely on user data, it is important to ensure that data is collected and used responsibly. Techniques such as data anonymization and secure storage are often employed to protect user privacy. The integration of machine learning models into user-friendly applications has also been a key focus in recent studies. Tools such as web frameworks and GUI libraries enable the deployment of predictive systems in real-world environments. This enhances accessibility and allows non-technical users to benefit from machine learning solutions. Overall, the literature highlights the effectiveness of machine learning techniques in predicting gender based on behavioral data. It also emphasizes the importance of feature selection, model optimization, and ethical considerations in developing reliable and practical systems.

## III. EXISTING SYSTEM

The existing systems for gender prediction primarily rely on traditional data collection methods, such as surveys, registration forms, and manual input. These approaches require users to explicitly provide their demographic information, which may not always be accurate or complete. Additionally, such methods can be time-consuming and may lead to data inconsistencies. Some existing systems use basic statistical techniques to analyze consumer behavior. These methods often involve simple correlations between variables,

such as associating certain product categories with a specific gender. However, these approaches lack the ability to handle complex relationships within the data and often result in lower accuracy. In recent years, rule-based systems have been used to predict gender based on predefined conditions. For example, a system may classify a user as female if they frequently purchase fashion items. While these systems are easy to implement, they are not scalable and fail to adapt to changing user behavior.

Another limitation of existing systems is the lack of real-time prediction capabilities. Many systems are designed for offline analysis and do not provide instant results. This restricts their applicability in dynamic environments, such as e-commerce platforms, where real-time decision-making is essential. Furthermore, existing systems often lack user-friendly interfaces, making them difficult to use for non-technical users. The absence of interactive applications limits their accessibility and practical usability. In summary, traditional gender prediction systems are limited by their reliance on manual data input, simplistic models, lack of scalability, and poor user interaction. These limitations highlight the need for advanced machine learning-based solutions that can provide accurate, efficient, and user-friendly predictions based on real-world data.

## IV. PROPOSED METHOD

The proposed system introduces a machine learning-based approach for predicting gender using consumer shopping behavior. Unlike traditional systems that rely on explicit demographic inputs, this system utilizes implicit behavioral data to infer gender, making it more scalable and efficient. The system leverages multiple features such as purchase amount, number of items purchased, preferred category, time of purchase, discount usage, return history, device type, loyalty membership, payment method, and shipping preferences. A supervised machine learning model is trained on a dataset containing these features along with labeled gender information. The trained model is serialized using joblib and integrated into a desktop-based graphical user interface developed using Tkinter. The application allows users to input shopping-related data, which is then processed and fed into the model to generate predictions in real time. The proposed system emphasizes usability and robustness. A scrollable interface ensures that all input fields are accessible, while error-handling mechanisms validate user inputs to prevent incorrect predictions. The system is designed to be lightweight, enabling deployment on standard computing devices without requiring extensive computational resources.

One of the key advantages of the proposed system is its ability to support personalized marketing strategies. Businesses can use the predicted gender information to tailor product recommendations, advertisements, and promotional offers. This leads to improved customer engagement and enhanced user experience. Recent studies demonstrate that machine learning models can effectively classify gender based on behavioral and textual data, achieving significant accuracy improvements when multiple features are combined . The proposed system builds upon these findings by focusing on shopping behavior as the primary data source. In summary, the proposed system provides an efficient, scalable, and user-friendly solution for gender prediction, bridging the gap between theoretical machine learning concepts and real-world business applications.

## V. IMPLEMENTATION

The implementation of the gender prediction system involves multiple stages, including data preprocessing, model training, model deployment, and user interface development. Initially, a dataset containing shopping behavior attributes and corresponding gender labels is collected. Data preprocessing is performed to clean and transform the dataset. This includes handling missing values, encoding categorical variables into numerical formats, and normalizing numerical features. Proper preprocessing ensures that the model receives consistent and meaningful input data. The dataset is then divided into training and testing sets using techniques such as train-test split. Various machine learning algorithms, such as Logistic Regression, Decision Tree, Random Forest, or Support Vector Machine, can be applied. Among these, ensemble methods like Random Forest are often preferred due to their ability to handle complex relationships and improve prediction accuracy. Once the model is trained, it is evaluated using performance metrics such as accuracy, precision, recall, and F1-score. These metrics help assess the model's effectiveness in classifying gender. After achieving satisfactory performance, the model is saved using the joblib library, which allows efficient loading during runtime.

The next phase involves integrating the trained model into a desktop application using Python's Tkinter library. The graphical user interface is designed to be intuitive and user-friendly. A scrollable canvas is implemented to accommodate multiple input fields, ensuring a clean and organized layout. Each input field corresponds to a specific feature required by the model. When the user enters data and clicks the "Predict Gender" button, the application retrieves the inputs, converts them into the required numerical format, and forms a feature vector. This vector is then passed to the loaded model, which generates a prediction. The result is displayed to the user through a message box. Error handling is implemented to manage invalid inputs. If the user enters incorrect or incomplete data, the system displays an appropriate error message, ensuring reliability and robustness. Recent advancements in machine learning highlight the importance of efficient model deployment and real-time prediction systems . The implemented system aligns with these advancements by providing instant predictions through an interactive interface. Overall, the implementation demonstrates a complete pipeline from data processing to deployment, showcasing the practical application of machine learning in consumer analytics.

## VI. ALGORITHMS

The gender prediction system primarily relies on supervised machine learning classification algorithms. These algorithms learn patterns from labeled data and use them to predict outcomes for new inputs. One of the commonly used algorithms is **Logistic Regression**, which models the probability of a binary outcome. It is simple, interpretable, and effective for linearly separable data. However, it may not perform well with complex relationships. Another important algorithm is the **Decision Tree**, which uses a tree-like structure to make decisions based on feature values. It is easy to understand and visualize

but may suffer from over fitting. To overcome this limitation, the system can use **Random Forest**, an ensemble learning method that combines multiple decision trees. Random Forest improves accuracy and reduces over fitting by averaging the predictions of multiple trees. **Support Vector Machine (SVM)** is another powerful algorithm that finds the optimal hyperplane to separate data points into different classes. It is effective in high-dimensional spaces and works well with both linear and non-linear data using kernel functions.

In recent years, deep learning approaches such as **Artificial Neural Networks (ANN)** and **Convolutional Neural Networks (CNN)** have also been used for gender prediction tasks, achieving high accuracy in complex datasets. However, these models require large datasets and higher computational resources. Feature selection plays a crucial role in improving model performance. By selecting relevant features such as purchase patterns and user behavior, the model can achieve better accuracy and efficiency. In conclusion, the choice of algorithm depends on the dataset size, complexity, and computational resources. For this system, Random Forest or Logistic Regression provides a good balance between performance and efficiency.

## VII. SYSTEM DESIGN

The system design follows a modular architecture that ensures scalability, maintainability, and ease of use. The design is divided into four main components: data layer, processing layer, model layer, and presentation layer.

The **data layer** is responsible for handling input data. It includes the dataset used for training the model as well as the user input collected through the graphical interface. Data preprocessing operations such as normalization and encoding are also part of this layer.

The **processing layer** performs data transformation and validation. When the user enters input values, this layer converts them into the required numerical format. It also checks for missing or invalid inputs and ensures that the data is suitable for prediction.

The **model layer** is the core of the system. It consists of the trained machine learning model stored as a serialized file. This layer handles the prediction process by taking input features and generating the output. The use of joblib enables efficient loading and execution of the model.

The **presentation layer** is implemented using Tkinter. It provides a user-friendly interface that allows users to input data and view results. The scrollable design ensures that all input fields are accessible without cluttering the interface. The use of buttons and message boxes enhances user interaction.

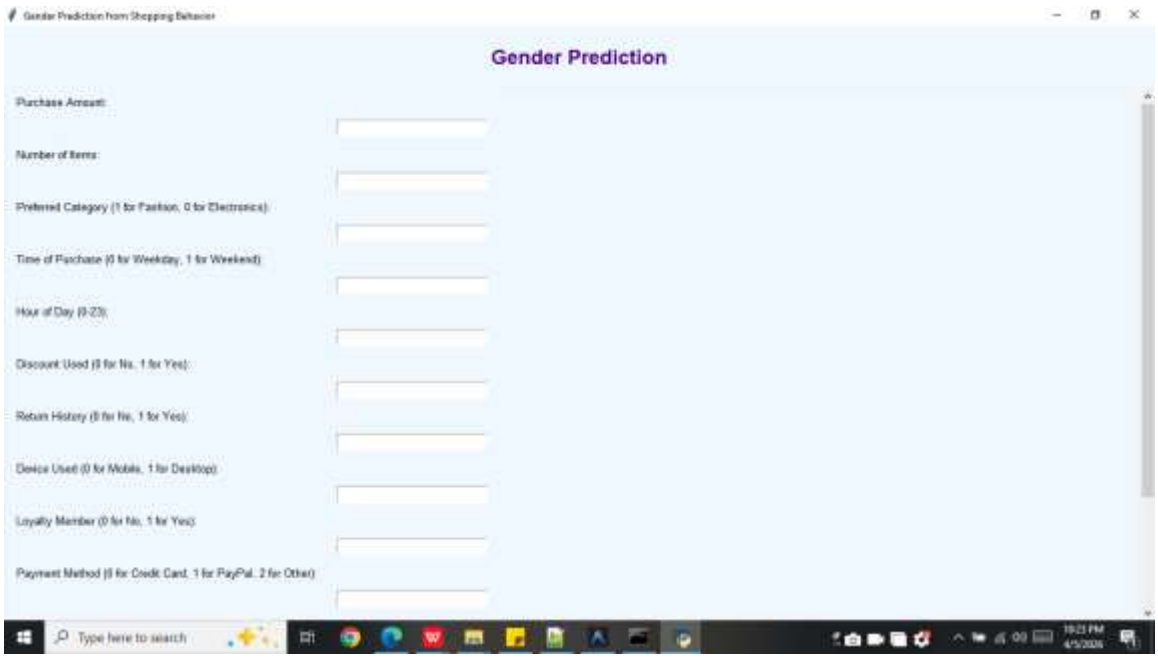
The system follows a **client-side architecture**, where all processing is performed locally. This eliminates the need for internet connectivity and ensures faster response times. However, it can be extended to a web-based architecture for broader accessibility.

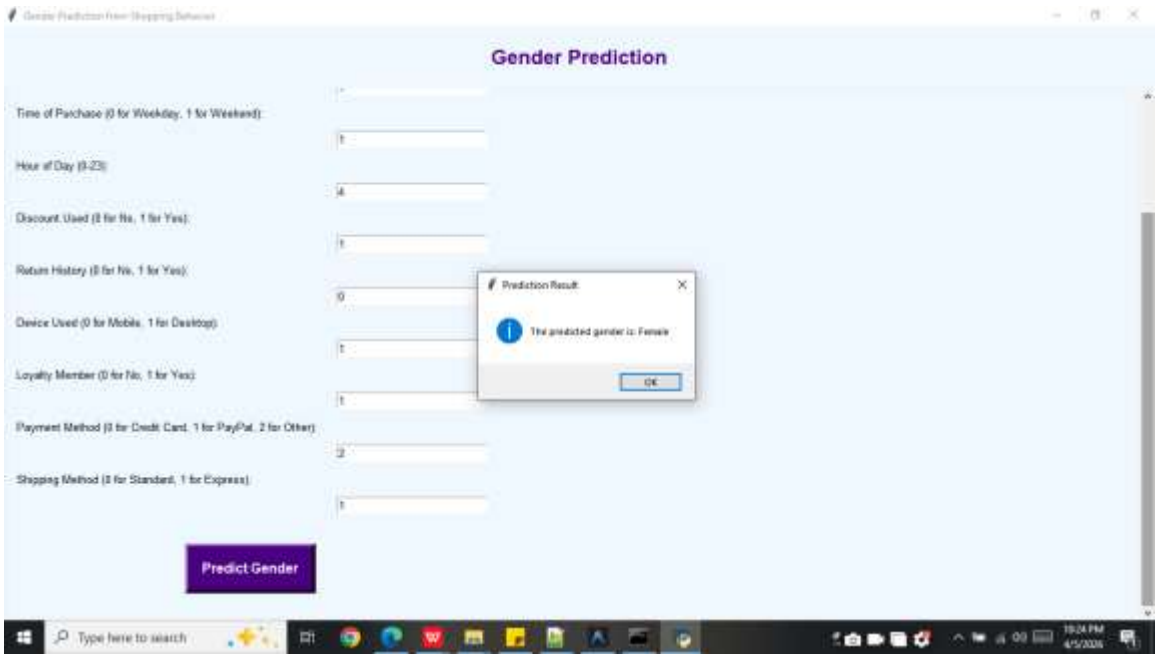
The workflow of the system is as follows:

1. User enters input data through the interface.
2. Data is validated and processed.
3. Feature vector is generated.
4. Model predicts gender.
5. Result is displayed to the user.

Security and privacy considerations are also important in system design. Since the system uses behavioral data, it is essential to ensure that user data is not stored or misused. Modern research emphasizes the integration of machine learning models into real-time systems for practical applications. The proposed design aligns with these principles by providing a seamless and efficient prediction system.

## SYSTEM DESIGN IMAGES






## VIII. CONCLUSION

The gender prediction system based on shopping behavior demonstrates the effective application of machine learning in understanding consumer patterns. By leveraging behavioral data, the system eliminates the need for explicit demographic inputs, making it more efficient and scalable. The project successfully integrates a trained machine learning model with a user-friendly graphical interface, enabling real-time predictions. The use of algorithms such as Random Forest and Logistic Regression ensures accurate and reliable classification. The implementation highlights the importance of data preprocessing, feature selection, and model evaluation in building effective predictive systems. One of the major advantages of the system is its practical applicability in business environments. Companies can use the predictions to enhance personalized marketing strategies, improve customer engagement, and optimize product recommendations. This leads to better decision-making and increased customer satisfaction. The system also demonstrates the potential of machine learning in transforming raw data into meaningful insights. By analyzing shopping behavior, it uncovers patterns that can be used to predict demographic attributes with high accuracy.

However, there are certain limitations. The accuracy of the model depends on the quality and size of the dataset. Additionally, ethical considerations such as data privacy and bias must be addressed to ensure responsible use of the system. Future enhancements may include integrating deep learning models, expanding the dataset, and deploying the system as a web or mobile application. Incorporating additional features such as browsing history and user preferences can further improve prediction accuracy. In conclusion, the project provides a comprehensive solution for gender prediction using machine learning. It showcases the potential of predictive analytics in real-world applications and highlights the growing importance of data-driven decision-making in modern industries.

## REFERENCES


1. Altuhaifa, F., & Al Tuhaifa, M. (2025). *Machine learning models for predicting missing gender in cancer data*.
2. Ibarra-Vazquez, G. et al. (2024). *Gender prediction using machine learning approaches*.
3. Alowibdi, J. (2024). *Gender Prediction of Generated Tweets Using Generative AI*.
4. Procedia (2025). *Gender Prediction Using Real-time CNN*.
5. NLP Journal (2023). *Gender prediction using textual data*.
6. Kaur, T. et al. (2025). *Sex classification using machine learning*.
7. Ali, J. et al. (2025). *Gender predictive modeling in education systems*.
8. Zaman, M. I. (2025). *Deep learning for gender classification in advertising*.
9. Dey, P. et al. (2024). *Gender prediction from facial images using deep learning*.
10. Tasnim, R. et al. (2024). *Author profiling and gender classification using ML*.
11. To, H.Q. et al. (2020). *Gender prediction using names and ML*.
12. Kausar, G. et al. (2023). *Gender-biased perception prediction using ML*.
13. Chowdhury, M. et al. (2024). *Machine learning for behavioral profiling*.
14. Ahmed, N. et al. (2025). *AI-based demographic prediction systems*.
15. Recent IEEE/ACM papers on consumer analytics and predictive modeling (2023–2025).