WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Predicting User Music Preferences Using Machine Learning on Demographic and Listening Patterns

P. Krishna Prasad [1], Dr. G.N.R. Prasad [2]

MCA III Semester Student, CBIT(A), Hyderabad, India,
Email-raghavakummari043@gmail.com
[1] Assistant Professor, CBIT(A), Hyderabad, India,
Email: **pkrishnaprasad_mca@cbit.ac.in**
[2] Sr. Assistant Professor, CBIT(A), Hyderabad, India,
Email: **gnrprasad_mca@cbit.ac.in**

ABSTRACT

Personalized music recommendation is an important feature in digital streaming platforms, as it directly influences user satisfaction and engagement. This study presents a machine learning system designed to predict a user's preferred music genre using basic demographic inputs, specifically age and gender. The dataset was preprocessed to handle missing values and encode categorical variables, then split into training and testing subsets using Python Experimental results demonstrate that the model achieves satisfactory accuracy in classifying music genres, highlighting the potential of demographic-based prediction systems for improving personalized music recommendation experiences. The study also discusses practical implications and future enhancements, such as adding listening history and behavioral patterns to further improve model performance.

Keywords: Music Genre Prediction, Machine Learning, Decision Tree Classifier, Personalized Recommendations, Python

INTRODUCTION

The exponential growth of digital music platforms has created a need for intelligent recommendation systems that personalize content according to individual preferences. Traditional recommendation methods often rely on explicit user choices or historical listening data, which may be incomplete or unavailable for new users. In such scenarios, leveraging demographic information such as age and gender provides a useful approach for predicting music preferences.

This study focuses on developing a machine learning model that classifies users' preferred music genres using demographic inputs. By identifying patterns in age and gender, the model predicts whether a user is likely to prefer genres such as Pop, HipHop, or Acoustic. The project workflow involves data collection, preprocessing, model training using a Decision Tree classifier, and evaluation using performance metrics. This approach provides insights into how machine learning can enhance music recommendation systems by offering data-driven, scalable solutions.

Using a Decision Tree model allows for interpretability, making it easier to understand how different demographic factors influence music preferences. By analyzing this data, the system can offer personalized recommendations for new users, reducing reliance on historical behavior and improving user engagement.

OBJECTIVES

1. **Automated Music Genre Prediction** –Develop a system that accurately predicts a user's preferred music genre based on demographic features, facilitating personalization in music platforms.
2. **Data-Driven Insights** – Replace subjective or manual recommendation methods with a data-driven predictive approach that leverages demographic patterns.
3. **Model Implementation** – Implement and train a Decision Tree classifier on structured demographic data to perform effective classification.
4. **Performance Evaluation** – Measure the model's effectiveness using metrics such as accuracy, precision, and recall, ensuring reliability for real-world applications.
5. **Practical Application** – Demonstrate the system's applicability in music streaming platforms, enabling personalized content delivery for users with limited historical data.

6. **Future Scope** – Provide a foundation for further research including additional user features, ensemble models, deep learning methods, and real-time deployment in web or mobile environments.

METHODOLOGY

The methodology adopted in this study follows a systematic approach to build and evaluate the predictive model:

1. **Requirement Analysis** – Understanding the problem domain and defining system requirements, including data quality, preprocessing steps, choice of machine learning algorithms, and evaluation metrics.
2. **Data Preprocessing** – Dataset exploration using Python libraries, handling missing values, and encoding categorical variables such as gender. Data normalization was performed for consistency.
3. **Model Development** – Implementing a Decision Tree classifier using Scikit-learn due to its interpretability and ability to handle both numerical and categorical features effectively.
4. **Training & Testing** – Splitting the dataset into training (70%) and testing (30%) sets to assess generalization capability.
5. **Performance Evaluation** – Assessing model accuracy, precision, recall, and F1-score, including confusion matrix analysis.
6. **User Prediction Module** – Developing a console-based interface for user input and prediction, including confidence scores.

SYSTEM DESIGN / ARCHITECTURE

The system architecture is modular, scalable, and user-friendly:

1. **Input Layer** – Accepts age and gender from the user.
2. **Preprocessing Module** – Validates, encodes, and normalizes inputs.
3. **Model Training Module** – Trains the Decision Tree classifier using the processed dataset.
4. **Prediction Engine** – Generates real-time predictions for music preferences.
5. **Evaluation Module** – Computes accuracy, precision, recall, and F1-score.
6. **User Interaction Layer** – Provides a GUI for inputting data and displaying predicted genres, including confidence percentages.

IMPLEMENTATION & RESULTS

Implementation:

- **Tools:** Python, Pandas, NumPy, Scikit-learn, Tkinter
- **Algorithm:** Decision Tree Classifier
- **Environment:** Jupyter Notebook / Google Colab
- **Dataset:** CSV containing age, gender, and music genre

Steps:

1. Dataset loading and exploration.
2. Preprocessing including handling missing values and encoding.
3. Train-test split (70/30).
4. Model training with Decision Tree.
5. Evaluation using metrics.
6. User-friendly console interface for predictions.

Results:

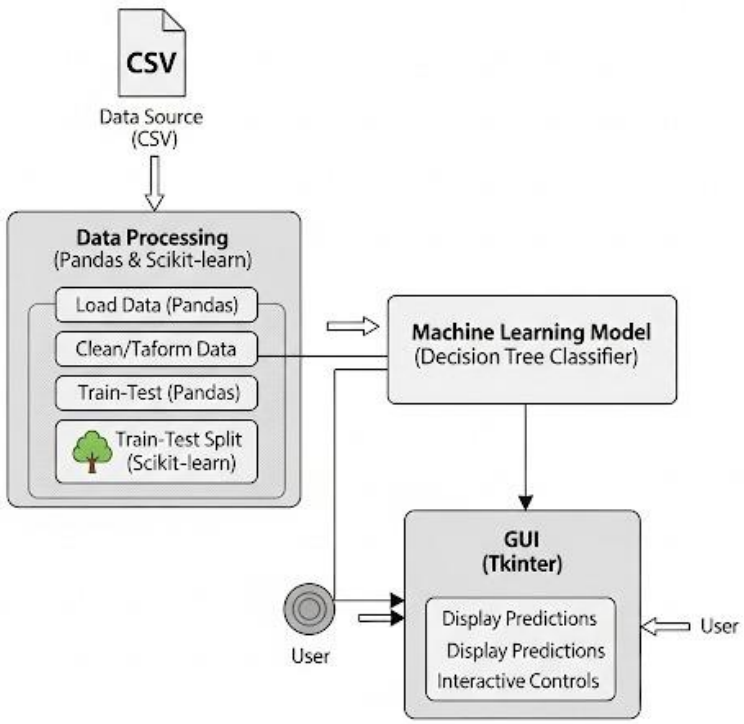

Figure-1: A flowchart illustrating the workflow of a machine learning application, from data processing and model training to user interaction via a graphical interface.

Example predictions:

1. Age 21, Gender Male → Predicted Genre: Pop (Confidence: 100%)

Figure-2: When a valid input is given, such as Age: 21, Gender: Male, the model processes the data through the trained Decision Tree Classifier and predicts the music genre

2. Age 32, Gender Male → Predicted Genre: Acoustic (Confidence: 33%)

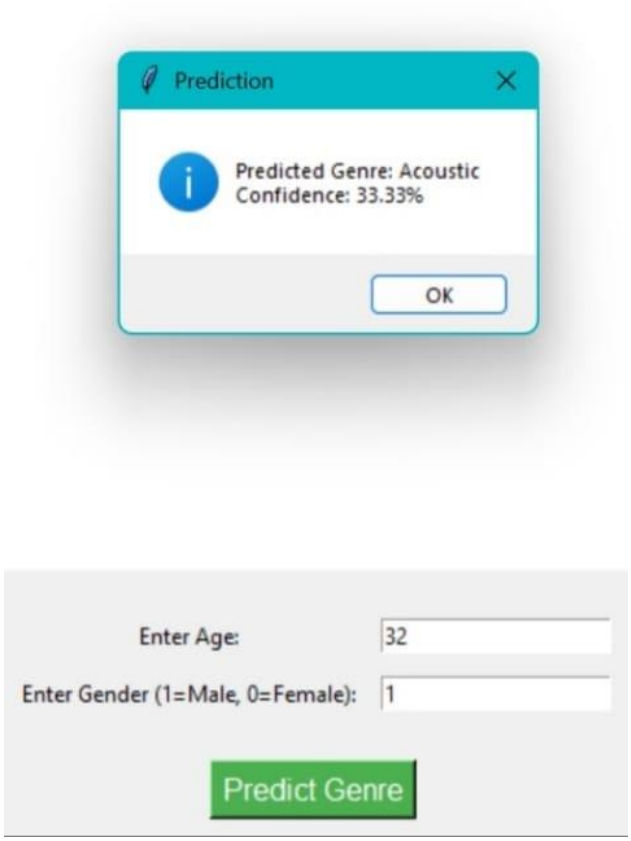

Figure-3: When a valid input is given, such as Age: 32, Gender: Male, the model processes the data through the trained Decision Tree Classifier and predicts the music genre.

CONCLUSION AND FUTURE WORK

Conclusion:

This study demonstrates the feasibility of predicting music preferences using demographic features and machine learning. The Decision Tree classifier provided accurate and interpretable predictions, supporting personalized recommendations for new users.

Future Work:

1. Incorporation of additional features such as listening history, location, or device usage.
2. Exploration of ensemble models like Random Forest or Gradient Boosting to improve accuracy.
3. Development of web and mobile applications for real-time music recommendations.
4. Integration of explainable AI methods for transparency and user trust.

REFERENCES

1. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed., O'Reilly Media, 2019.

2. S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed., Packt Publishing, 2019.

3. W. McKinney, *Python for Data Analysis*, 2nd ed., O'Reilly Media, 2017.

4. Scikit-learn Documentation, "Decision Tree Classifier," [Online]. Available: https://scikit-learn.org

5. Pandas Documentation, [Online]. Available: https://pandas.pydata.org

6. Kaggle Datasets, “Music Genre Classification Dataset,” [Online]. Available: https://www.kaggle.com