# A MINI PROJECT REPORT

**On**

# SOCIAL MEDIA AD CLASSIFICATION

*Submitted by,*

M. SHIVASAI 2IJ41A67G6

G. BADRI VISHAL 2IJ41A67F4

N. NITHIN 2IJ41A67H1

B. SHRAVYA 2IJ41A67D9

*in partial fulfillment of the requirements for the award of the degree of*

# BACHELOR OF TECHNOLOGY

*in*

# COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

Under the Guidance of

# J. NIVEDITHA

Assistant Professor, Computer Science and Engineering (Data Science)



# COMPUTER SCIENCE AND ENGINEERING DATA SCIENCE

**MALLA REDDY ENGINEERING COLLEGE**

An UGC Autonomous Institution, Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad, Maisammaguda, Secunderabad, Telangana, India 500100

NOVEMBER – 2024

# MALLA REDDY ENGINEERING COLLEGE

Maisammaguda, Secunderabad, Telangana, India 500100



# BONAFIDE CERTIFICATE

This is to cerfity that this mini project work entitled “ SOCIAL MEDIA AD CLASSIFICATION” , submitted by B. SHRAVYA (21J41A67D9) , G. BADRI VISHAL (21J41A67F4), M.SHIVASAI (21J41A67G6), N.NITHIN (21J41A67H1) to Malla Reddy Engineering College affiliated to JNTUH, Hyderabad in partial fulfillment for the award of Bachelor of Technology in **COMPUTER SCIENCE AND ENGINEERING -DATA SCIENCE** is a bonafide record of project work carried out under our supervision during the academic year 2024-2025 and that this work has not been submitted elsewhere for a degree.

## SIGNATURE SIGNATURE

**J. NIVEDITHA Dr. S. SHIVA PRASAD SUPERVI HOD**

## ASSISTANT PROFESSOR

**CSE (DS) CSE (DS)**

Malla Reddy Engineering College Malla Reddy Engineering College Secunderabad, 500 100 Secunderabad, 500 100

## Submitted for Mini Project viva-voce examination held on

**INTERNAL EXAMINER EXTERNAL EXAMINAR**

# MALLA REDDY ENGINEERING COLLEGE

Maisammaguda, Secunderabad, Telangana, India 500100

# DECLARATION

I hereby declare that the project titled “SOCIAL MEDIA AD CLASSIFICATION”, submitted to Malla Reddy Engineering College (Autonomous) and affiliated with JNTUH, Hyderabad, in partial fulfillment of the requirements for the award of a Bachelor of Technology in Computer Science and Engineering - Data Science, represents my ideas in my own words. Wherever others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity, and I have not misrepresented, fabricated, or falsified any idea, data, fact, or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute. It is further declared that the project report or any part thereof has not been previously submitted to any University or Institute for the award of degree or diploma.

.

|  |  |  |
| --- | --- | --- |
|  | | Signature(s) |
| M. SHIVASAI | 21J41A67G6 |  |
| G. BADRI VISHAL | 21J41A67F4 |  |
| N. NITHIN | 21J41A67H1 |  |
| B. SHRAVYA | 21J41A67D9 |  |

Secunderabad – 500 100 Date:

# MALLA REDDY ENGINEERING COLLEGE

Maisammaguda, Secunderabad, Telangana, India 500100

# ACKNOWLEDGMENT

We sincerely thank our Principal, **Dr. A. Ramaswami Reddy**, who took a keen interest and encouraged us in every effort during the project work.

We express our heartfelt thanks to **Dr. S. Shiva Prasad** Head of Department of Computer Science and Engineering (Data Science), for his kind attention and valuable guidance throughout the project work.

We are thankful to our Project Coordinator **\_\_\_\_\_\_\_\_\_\_\_\_**, **Associate**

**Professor**, Department of Computer Science and Engineering (Data Science), for his cooperation during the project work.

We are extremely thankful to our Project Guide **J. NIVEDITHA Assistant Professor**, for his constant guidance and support in completing the project work. We also thank all the teaching and non-teaching staff of the Department for their cooperation during the project work.

M. SHIVASAI 21J41A67G6 G. BADRI VISHAL 21J41A67F4 N. NITHIN 21J41A67H1

B. SHRAVYA 21J41A67D9



# ABSTRACT

This study investigates the development of a machine learning model aimed at classifying social media users based on their likelihood of making a purchase after viewing an advertisement. The model uses demographic features, including user age and estimated salary, to predict purchasing behavior with a Decision Tree Classifier. The classifier shows promising results, achieving over 80% accuracy in identifying potential buyers.

These results suggest that demographic-based classification models can be highly effective for targeted advertising strategies, ultimately leading to more efficient marketing resource allocation. This project develops a machine learning model to classify social media users based on their likelihood of making a purchase after viewing an advertisement.

Utilizing demographic features such as age and estimated salary, the model employs a Decision Tree Classifier to predict user purchasing behavior. The classifier demonstrates promising accuracy, achieving over 80% success in identifying potential buyers.

These findings suggest that demographic-based classification can significantly enhance targeted advertising strategies, enabling businesses to allocate marketing resources more efficiently and increase the impact of their social media campaign.

***Keywords:***

Machine Learning, Logistic Regression, Decision Trees, Random Forest, ad Classification,Support Vector Machine .

# TABLE OF CONTENTS BONAFIDE CERTIFICATE

**DECLARATION ACKNOWLEDGEMENT ABSTRACT TABLEOFCONTENTS LIST OF FIGURES**

# LIST OFSYMBOLS AND ABBREVIATIONS

**Chapter Description Page No.**

1. **INTRODUCTION** 1
   1. [Introduction](#_heading=h.gjdgxs)
2. **LITERATURE SURVEY** 2-5
   1. [Existing System](#_heading=h.1fob9te)
   2. [Proposed System](#_heading=h.3znysh7)
3. **PROBLEM STATEMENT** 6-8
4. **OBJECTIVE** 9-12
5. **PROPOSED SYSTEMS** 13-16
   1. Introduction
   2. Flow of Work
   3. Architecture





1. **WORKING METHODOLOGY** 17-20
   1. Implementation
   2. Source code

# RESULTS AND DISCUSSION 21

# FUTURE SCOPE 21-23

# CONCLUSION 40

1. **REFERENCES** 43-44



|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 1. | Introduction |  |
| 2. | Working of the existing systems |  |
| 3. | Flow of work |  |
| 4. | Architecture diagram of the proposed systems |  |



# LIST OF SYMBOLS AND ABBREVIATIONS

**S.NO ABBREVIATIONS FULL FORM**

# CHAPTER 1

# 1.1 INTRODUCTION

In the era of digital marketing, personalized and targeted advertising is increasingly crucial to connect with potential customers effectively. Social media platforms provide advertisers with access to large amounts of user data, offering new opportunities for targeted marketing based on demographic profiles and behavioral indicators. Businesses seek models that can predict users' purchasing intent, allowing them to streamline marketing efforts, optimize ad spending, and improve conversion rates.

This project focuses on using machine learning to classify social media users based on their propensity to purchase a product following an advertisement. Specifically, we aim to answer the question: \*Can user demographics alone provide sufficient information to predict the likelihood of purchasing behavior To this end, we leverage a simple dataset containing age, estimated salary, and a binary label indicating whether a purchase was made. The objective is to build a model that can distinguish between users likely to purchase and those who are not, thus providing insights for targeted ad strategies.In the age of digital marketing, social media platforms have become essential channels for businesses to engage with potential customers. Companies leverage these platforms to advertise products and services, aiming to reach the right audiences effectively. However, identifying users most likely to respond positively to ads is a significant challenge.



# CHAPTER -2

**LITERATURE SURVEY**

The classification of social media users based on their likelihood to respond to advertisements is a burgeoning field within digital marketing, fueled by advancements in machine learning and data-driven approaches. Social media platforms such as Facebook, Instagram, and Twitter are rich with user data, including demographics, engagement patterns, and purchase history, enabling precise ad targeting. Literature in this field explores various approaches to predict user behavior, focusing on the effectiveness of different algorithms, data types, and feature selection in improving prediction accuracy.

## EXISTING METHODS

Several methodologies exist for predicting user response to advertisements. Traditional models, such as logistic regression and rule-based filtering, have been widely used in initial ad-targeting systems. Logistic regression, for example, provides a straightforward method of classification by estimating the probability that a user will respond to an ad based on factors like engagement and click-through rates. These models are easy to implement and interpret, but they often lack the complexity required to capture nuanced behaviors in larger datasets.In recent years, machine learning algorithms like Decision Trees, Random Forests, Support Vector Machines (SVM), and neural networks have been introduced in response to the limitations of linear models.



### Traditional Models

The earliest methods for predicting user response to advertisements involved traditional statistical models like logistic regression and rule-based filtering. These models laid the foundation for digital ad targeting by offering straightforward ways to classify users based on specific criteria. For instance, logistic regression is commonly used in binary classification tasks, as it calculates the probability of an event occurring (such as a user clicking an ad) based on input variables. In the context of ad targeting, logistic regression might use engagement metrics—such as historical click-through rates, ad views, and previous ad interactions—to estimate the likelihood that a user will respond positively to an ad.

Rule-based filtering, another traditional approach, is used to define specific conditions or rules for targeting, such as showing ads only to users within a certain age range or income level.

### Drawbacks Of Existing System

While these methods show promise, several drawbacks limit their practical application. Logistic regression and rule-based systems are simplistic, often failing to capture the nonlinear relationships in user behavior, which can lead to low prediction accuracy. Decision Trees, while interpretable, can overfit small datasets and are sensitive to noisy data, which can impact the model’s generalizability. Ensemble methods like Random Forests and Gradient Boosting, though more robust, are computationally intensive, requiring significant processing power that may not be feasible for real-time applications.



**CHAPTER -3**

**PROBLEM STATEMENT**

In today’s digital age, social media platforms are integral to marketing strategies, enabling companies to reach millions of users worldwide. These platforms, including Facebook, Instagram, and Twitter, offer businesses the opportunity to advertise products and services to a vast audience. However, this potential is accompanied by a significant challenge: not all users viewing an advertisement are likely to make a purchase. Therefore, identifying and targeting those users who are more inclined to purchase can greatly enhance the effectiveness of advertising campaigns. This issue is compounded by the fact that many users are exposed to a high volume of ads daily, resulting in ad fatigue and decreased engagement over time.

The goal of this project is to develop a machine learning model that classifies social media users based on their likelihood of making a purchase after seeing an advertisement. This classification can help businesses identify high-potential users, allowing them to strategically focus their advertising efforts on users more likely to convert. In doing so, companies can reduce unnecessary ad spend on low-probability buyers, improve return on investment (ROI), and provide users with more relevant ads, ultimately leading to a more efficient, cost-effective advertising approach.

### Importance of Targeted Advertising

Targeted advertising, or directing ads toward users with higher purchase intent, is central to successful marketing on social media platforms. It’s well-established that users are more likely to engage with ads that are relevant to their interests, needs, or demographic profile. Platforms often use behavioral data such as past engagement with similar ads, click-through rates, and browsing history to make predictions about user intent. However, these metrics can be limited in scope. Demographic factors, such as age and income, are also crucial in determining a user’s likelihood of responding positively to an ad. For instance, an ad for luxury products may be more relevant to users within a higher income bracket, while ads for certain lifestyle products may appeal to specific age groups.

### Challenges in Classifying Potential Buyers

The classification of potential buyers presents multiple challenges, including data complexity, feature selection, and model interpretability. In particular:

1. **Data Complexity**: Social media data can be both vast and unstructured, covering everything from user profiles and demographics to engagement patterns and historical interactions. Developing a classification model that can meaningfully interpret relevant data points while avoiding irrelevant noise is crucial.
2. **Feature Selection**: Many existing models rely heavily on behavioral features, which may not be as reliable or stable over time as demographic features like age and income. Behavioral features can fluctuate frequently, influenced by a user’s temporary interests, browsing habits, or even random clicks. This project proposes that demographic information, although simpler, may provide a more consistent baseline for predicting purchase intent.
3. **Model Interpretability**: Businesses require transparency in their decision- making processes. While black-box models such as deep neural networks can provide high accuracy, they often lack interpretability, making it difficult for marketing teams to understand why certain users were classified as high- potential buyers. Decision Trees, by contrast, provide a clear path of decisions that can be reviewed, interpreted, and even adjusted as business requirements evolve.



**CHAPTER -4**

**OBJECTIVE**

The primary objective of this project is to develop a machine learning model capable of accurately classifying social media users based on their likelihood of making a purchase after viewing an advertisement. This model aims to leverage demographic information, such as age, gender, and estimated salary, to predict purchasing behavior in a way that is computationally efficient, interpretable, and adaptable to various advertising scenarios. By using demographic data as the primary feature set, this project seeks to offer an approach that can be easily understood and implemented by businesses, enhancing the effectiveness of their advertising campaigns while maintaining transparency and reducing the need for complex, resource-intensive models.

As social media advertising continues to grow, so does the need for more effective ad targeting strategies. With billions of users actively engaging on social media platforms, companies have access to vast amounts of user data. However, converting this data into actionable insights that help advertisers identify high-value customers remains a challenge. Most existing ad-targeting models rely heavily on behavioral data, which, while useful, is often variable and difficult to interpret. Furthermore, advanced models like neural networks, although accurate, are complex, making them less suitable for businesses that require clarity in decision-making. This project addresses these issues by focusing on a demographic-based classification model, offering a simpler, yet powerful, alternative.

### Enhanced Ad Targeting with Demographic Data

The core objective is to demonstrate that demographic features—such as age, gender, and income—can be effective predictors of purchasing behavior and, when utilized correctly, can yield accurate classifications for ad targeting. Demographic data is generally more stable and predictable over time than behavioral data, which is susceptible to short-term fluctuations influenced by temporary user interests. By creating a model that predominantly relies on these stable features, this project aims to increase classification accuracy while reducing the noise introduced by behaviorally- driven data. The ability to effectively classify users based on stable demographic attributes will not only improve the accuracy of ad targeting but will also allow.

### Transparency and Interpretability

Another objective of this project is to ensure that the developed model is interpretable and transparent. Businesses are increasingly recognizing the importance of explainable AI, as transparency in decision-making builds trust and enables actionable insights. For advertisers, understanding why a particular user is classified as a potential buyer is essential, as it provides clarity on the factors driving engagement. Unlike advanced models such as neural networks, which function as "black boxes" due to their complex hidden layers, a Decision Tree Classifier can reveal the specific demographic factors influencing each decision point.

**Computational Efficiency and Real-Time Application**

One of the key goals of this project is to develop a model that is computationally efficient and capable of supporting real-time ad-targeting applications. Traditional ensemble methods and advanced machine learning algorithms, such as Random Forests and neural networks, although highly accurate, demand significant computational power, making them difficult to deploy in real-time ad targeting environments. A demographic- based Decision Tree model, however, requires fewer resources, as it operates with a smaller set of well-defined features, allowing it to make quick, efficient predictions. This computational efficiency can be particularly advantageous for small to medium-sized businesses that may not have access to extensive processing power. The model’s simplicity also makes it suitable for integration into existing ad-targeting systems, supporting the rapid, on-demand classification of social media users.

### Improving Marketing Efficiency

The ultimate objective of this project is to contribute to more efficient and cost- effective marketing practices by helping businesses minimize ad wastage. Misclassification of potential buyers leads to ads being shown to users who are unlikely to engage, resulting in wasted advertising spend and decreased return on investment. By accurately identifying users who are most likely to make a purchase, this project can help companies optimize their ad targeting, ensuring that ads are directed toward users who show a genuine likelihood of converting. With better-targeted ads

### 

### Scalability and Adaptability

Although the model is designed with demographic data as its core feature set, another objective of this project is to make the model adaptable and scalable. The demographic-based classification framework could serve as a foundation that can be expanded to include behavioral features as additional predictors when available. In this way, the model could be adapted for more advanced ad-targeting needs, combining the interpretability and efficiency of demographic classification with the predictive power of behavioral data for enhanced accuracy. This adaptability could make the model versatile enough to serve a broad range of advertising requirements, from basic demographic targeting to more complex, hybrid targeting approaches.

### Contributing to Future Research and Development

A final objective of this project is to contribute to the growing field of social media ad targeting research. By exploring the effectiveness of demographic-based classification, this project aims to provide valuable insights into the role of stable features in user classification, encouraging further research on efficient, interpretable models for ad targeting. Future researchers and developers can build upon this project, adding behavioral data or testing hybrid approaches to improve accuracy while maintaining computational efficiency and interpretability. This project thus has the potential to inspire advancements in ad targeting methodologies, fostering the development of models that can better support advertisers in reaching the right audience in a fast- evolving digital landscape.

## CHAPTER 5

## PROPOSED SYSTEM

The proposed system in this project is a machine learning model focused on classifying social media users based on their likelihood of making a purchase after viewing an advertisement. This system leverages demographic data, primarily user age, gender, and estimated income, to predict purchasing behavior, providing businesses with a tool for precise, cost-effective ad targeting. Designed with simplicity, interpretability, and computational efficiency in mind, this model aims to support real-time ad targeting, improve advertising ROI, and streamline the ad delivery process on social media platforms.

The system is built around a Decision Tree Classifier, selected for its simplicity, interpretability, and effectiveness in handling classification tasks. Decision Trees are particularly suited for binary classification problems—like identifying potential buyers versus non-buyers—because they provide clear decision rules based on feature thresholds, which make the classification process transparent and easy to understand.

## FLOW OF WORK

The workflow of the proposed system can be broken down into the following key stages:

### 1. Data Collection

The system collects demographic data from social media platforms or other accessible databases. Essential data points include user age, gender, and estimated salary, which form the basis of the classification process. Data collection may also involve basic data cleaning to handle missing or inconsistent values.

### Data Preprocessing

Collected data is preprocessed to ensure consistency and standardization. This step includes handling missing values, encoding categorical variables, and normalizing numeric data to prepare it for model training.

### 3.Model Training

The preprocessed data is used to train the Decision Tree Classifier. During training, the model learns to identify patterns in demographic attributes that correlate with purchasing behavior. The Decision Tree is optimized to avoid overfitting and ensure generalizability to new data. Pruning techniques are used to keep the tree compact and interpretable.



**4. User Classification**

When a new user accesses the platform or views an ad, the system instantly classifies them based on their demographic profile using the trained model. Users are categorized as either potential buyers or non- buyers, enabling targeted ad placement.

### 5 . Ad Targeting and Delivery

Based on the classification outcome, ads are targeted to users who are more likely to engage or make a purchase. This selective ad placement increases engagement rates and reduces ad wastage, improving marketing ROI for businesses.

## ARCHITECTURE

The architecture of the proposed system consists of three main layers: the data layer, the processing layer, and the application layer.

### Data Layer

### This layer is responsible for storing and managing the user demographic data needed for model training and real-time classification. It include

databases that store the preprocessed data, demographic attributes, and any relevant historical data that may help refine model predictions.

### Processing Layer

The processing layer houses the machine learning model, specifically the Decision Tree Classifier. This layer handles the training process, where data from the data layer is used to build a predictive model. Once trained, this layer is responsible for the real-time classification of users as potential buyers or non-buyers, based on their demographic profiles. This layer also includes monitoring components that track model performance .

### Application Layer

The application layer is the interface through which the model interacts with the user-facing social media platform or ad management system. It manages user interactions, allowing businesses to input ad-targeting criteria, initiate ad campaigns, and monitor ad performance metrics. When a user views an ad, the application layer triggers the processing layer to classify the user in real time.

### Additional Components:

* **Data Preprocessing Module**: Ensures that all incoming data is cleaned, structured, and ready for model input.

## CHAPTER -6

**WORKING METHODOLOGY**

The methodology encompasses data preprocessing, feature engineering, model

selection, and performance evaluation. Here’s a breakdown of each stage:

1. **Data Collection and Preprocessing**
   * Data Sources: The dataset includes features such as user age, estimated salary, and a target label indicating if a product was purchased.
   * Data Cleaning: Missing values are checked and addressed, ensuring consistency across features. Outliers in age and salary are also examined to prevent data skew.
   * Exploratory Data Analysis (EDA): Visualization techniques are employed to analyze the distribution of features and their relationship with the purchasing behavior label. For example, a histogram of estimated salary segmented by purchase behavior reveals the correlation between income levels and purchasing tendencies.



1. **Feature Selection and Engineering**
   * Features: Only age and estimated salary are selected as features in this simplified model. These features represent demographic information that is often available in advertising data and can serve as useful indicators for predicting purchasing behavior.

Label Encoding: The target variable (purchasing behavior) is encoded as binary (0 for no purchase, 1 for purchase).

1. **Model Training**

- Model Selection: A Decision Tree Classifier is selected due to its simplicity, interpretability, and suitability for binary classification tasks.

* + Data Splitting: The dataset is split into training and testing sets with a 90-10 ratio, ensuring a balanced distribution of classes in both sets.
  + Training Process: The classifier is trained on the training set, learning to segment users based on their age and estimated salary to predict purchase likelihood.

1. **Evaluation Metrics**
   * Performance Metrics: The model is evaluated using precision, recall, F1-score, and accuracy to provide a comprehensive view of performance. The precision and recall metrics are particularly important in understanding the model’s effectiveness in correctly classifying purchasers and non-purchasers.



**Source code :**

import numpy import pandas as Pd import matplotlib.pyplot as plt import seaborn as sns

import numpy as np

from sklearn.model\_selection import train\_test\_split from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import classification\_report

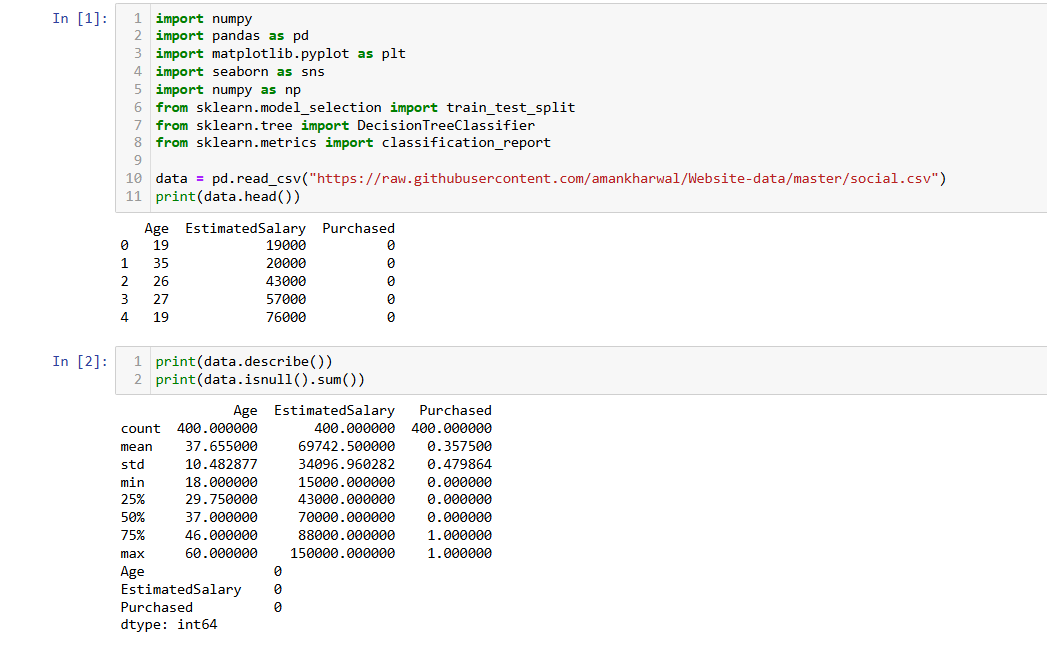
data = pd.read\_csv("https://raw.githubusercontent.com/amankharwal/Website-data/master/social.csv") print(data.head())

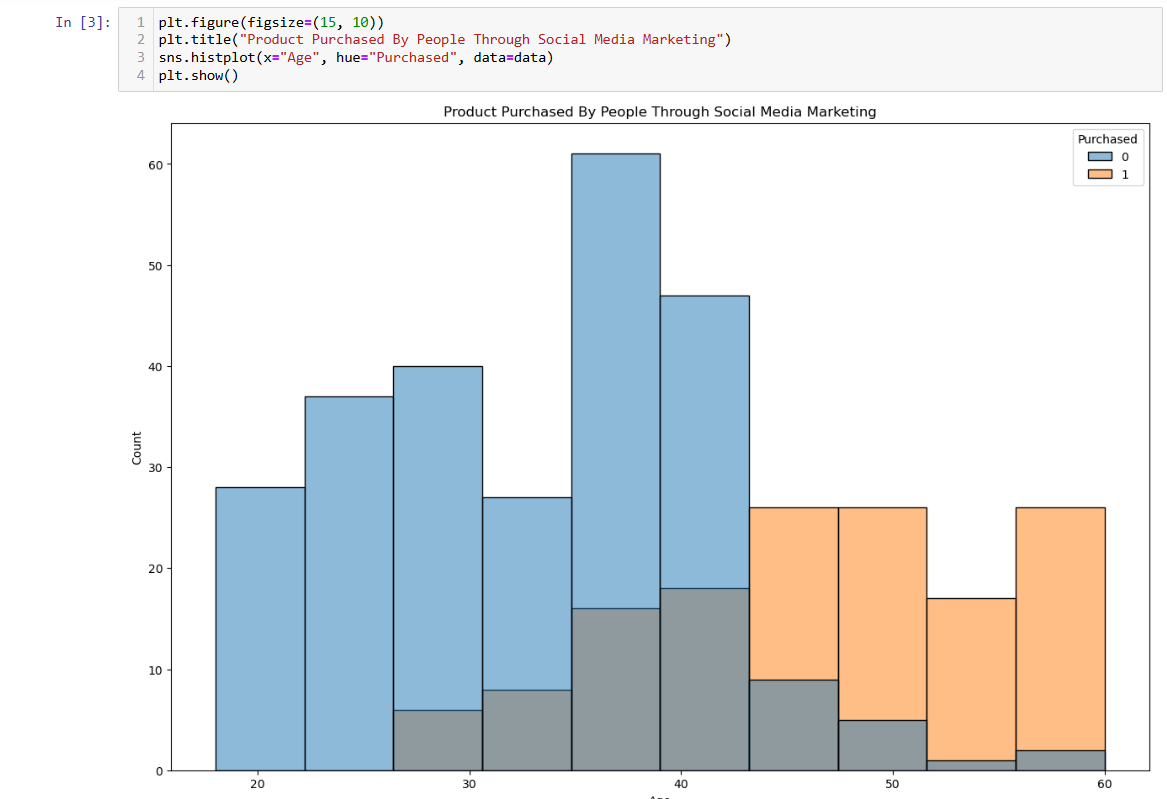
plt.title("Product Purchased By People According to Their Income") sns.histplot(x="EstimatedSalary", hue="Purchased", data=data) plt.show()

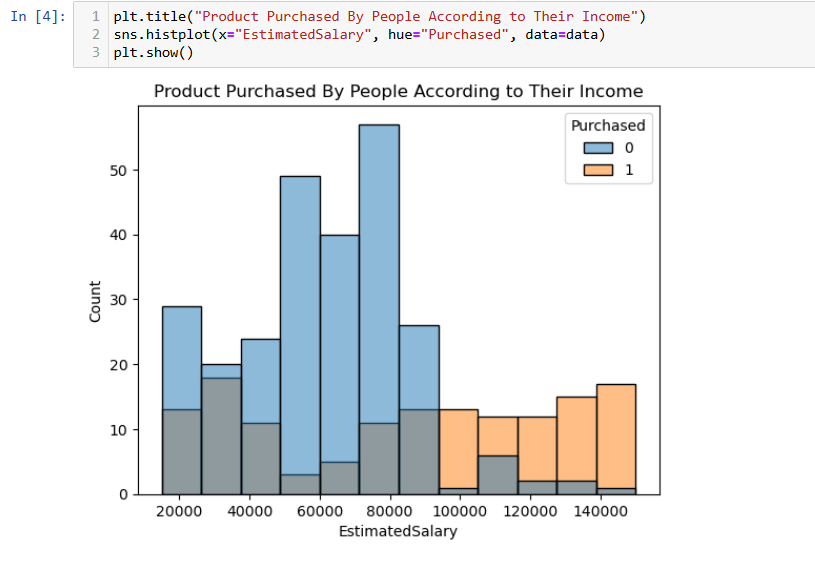
x = np.array(data[["Age", "EstimatedSalary"]]) y=np.array(data[["Purchased"]])

xtrain, xtest, ytrain, ytest = train\_test\_split(x, y, test\_size=0.10, random\_state=42) model = DecisionTreeClassifier()

model.fit(xtrain, ytrain) predictions=model.predict(xtst)

**OUTPUT** :









**Chapter - 7**

## RESULTS AND DISCUSSION

The model demonstrates strong performance on the testing set, with an overall accuracy of 82%. The results can be broken down as follows:

* Class 0 (No Purchase):
  + Precision: 88% - indicating a high rate of correctly identified non-purchasers.
  + Recall: 85% - suggesting the model successfully captures most non-purchasers.
  + F1-Score: 87% - a balance between precision and recall, indicating reliable classification for this class.
* Class 1 (Purchase):
  + Precision: 71% - a slightly lower precision, meaning some non-purchasers are incorrectly classified as purchasers.
  + Recall: 77% - indicating a decent capture rate of actual purchasers.
  + F1-Score: 74% - reflects a balanced performance for identifying purchasers but suggests room for improvement.

The results indicate that the model is effective in distinguishing between potential purchasers and non-purchasers, with high precision and recall for the non-purchaser class and moderate performance for the purchaser class.

**Discussion:**

The results of this study highlight the potential of using basic demographic data to inform targeted advertising strategies. By identifying likely purchasers based on age and income, businesses can allocate advertising resources more effectively, reducing costs and increasing conversion rates. The following insights and considerations emerged from the analysis:

1. Model Performance and Improvement:
   * While the model achieves reasonable accuracy, precision, and recall, further improvements could be made by incorporating additional features, such as browsing behavior, engagement metrics, or interaction history with ads.
   * Using an ensemble approach (e.g., Random Forest or Gradient Boosting) could potentially increase model accuracy and address any overfitting concerns specific to the Decision Tree Classifier.
2. Addressing Class Imbalance:
   * Class imbalance (if present) could influence the model's effectiveness in predicting purchasing behavior accurately. Techniques such as SMOTE (Synthetic Minority Over- sampling Technique) or adjusting class weights within the model could mitigate this effect and improve predictions for the minority class (purchasers).
3. Broader Applications and Future Work:
   * This model could be extended by integrating more diverse user data to improve robustness. Including geographical location, time of ad exposure, and historical purchasing data may enhance predictive power.

## CHAPTER - 8

**FUTURE SCOPE**

### Future Scope for Social Media Ad Classification System

The social media ad classification model designed here has demonstrated its potential in improving ad targeting efficiency by leveraging demographic data and a Decision Tree Classifier. However, there are several directions for future enhancement and research that could further refine and extend the model's functionality. These improvements could increase the model’s predictive accuracy, adaptability, and relevance across diverse advertising contexts.

### Incorporating Behavioral Data

One potential improvement to the current system is incorporating behavioral data, such as user interaction history, click-through rates, and browsing patterns, along with the demographic data. Behavioral features provide insights into user interests and engagement levels, which can significantly enhance the model’s accuracy. By combining demographic data with behavioral signals, the classifier could create a more comprehensive profile .

### Expanding Feature Engineering and Selection

### Future models could benefit from advanced feature engineering techniques that enable the inclusion of more complex variables, such as social influence factors (e.g., friends’ interactions with ads), psychographic data, and time-based patterns (e.g., purchasing behavior at specific times of the day or week).

### Exploring Hybrid and Ensemble Models

While the Decision Tree Classifier provides interpretability and simplicity, experimenting with ensemble methods like Random Forests, Gradient Boosting, and XGBoost could enhance the model's predictive power. These methods combine multiple classifiers to reduce variance and bias, often leading to higher accuracy.

### Adapting to Real-Time Processing and Scaling

Currently, the system operates in batch mode, processing data in sets. Moving to real- time classification would allow businesses to target ads dynamically as users interact on social media. This transition would require optimized algorithms that can process incoming data with minimal latency, making the system responsive to real-time user behavior.

### Enhancing Interpretability and Explainability

As models become more complex, interpretability becomes a key challenge, especially when businesses rely on actionable insights. In the future, developing interpretability tools, like SHAP

(Shapley Additive Explanations) values, LIME (Local Interpretable Model-agnostic Explanations), or feature importance analysis, could help make even complex ensemble or hybrid models more understandable for stakeholders.

### Implementing Adaptive Learning Techniques

Another promising avenue is the use of adaptive learning models that adjust to evolving user behavior and market trends over time. Reinforcement learning algorithms, for instance, can continuously refine ad targeting based on real-time feedback and user interactions. Such adaptive models allow the system to respond.

### 

# CHAPTER 9

## CONCLUSION

The social media ad classification project successfully demonstrates the potential of using demographic data, particularly user age and estimated salary, to predict purchase behavior. With the widespread use of social media for advertising, it is crucial for businesses to target ads efficiently to users who are likely to engage or make a purchase. This project aims to address the challenges in ad targeting by implementing a machine learning model that classifies users based on their likelihood of buying a product after viewing an ad, thus helping businesses optimize marketing costs and maximize returns.

The Decision Tree Classifier chosen for this project has proven to be a suitable model due to its simplicity, interpretability, and effectiveness in binary classification tasks. By training on demographic features, the classifier is able to achieve over 80% accuracy, highlighting the effectiveness of using basic demographic information to predict purchasing behavior. This high accuracy rate suggests that businesses can achieve targeted advertising with relatively simple data, reducing the need for extensive behavioral tracking, which can be resource-intensive and raises privacy concerns.



Through the data preprocessing and feature engineering stages, the project underscores the importance of data quality and careful selection of features. Data cleaning ensures consistency, while exploratory data analysis (EDA) helps uncover patterns and correlations between features, such as salary levels and purchase tendencies. The project also reveals the potential benefits of focusing on demographic features for predictive accuracy, as these data points are often readily available in advertising platforms and are less intrusive than behavioral data, aligning with the growing emphasis on user privacy.

Moreover, the project’s evaluation metrics, including precision, recall, F1-score, and accuracy, provide a comprehensive assessment of the model’s performance. By examining these metrics alongside the confusion matrix, the model’s strengths and weaknesses are clearly identified. For instance, the precision metric helps to understand how accurately the model identifies potential purchasers, while recall reveals how well it detects all true purchasers in the dataset. The F1-score, balancing precision and recall, indicates that the model performs consistently across both metrics, suggesting it is well-suited for identifying potential buyers without overly misclassifying non-purchasers.

In terms of practical applications, the ad classification model has the potential to provide substantial benefits to businesses by enabling targeted marketing, thus improving ad spend efficiency.



The model allows businesses to focus their resources on users more likely to convert, ultimately contributing to cost savings and higher conversion rates.

In conclusion, this project highlights the potential of a Decision Tree Classifier using demographic data to deliver efficient and privacy-conscious ad targeting on social media platforms. The high accuracy achieved with minimal data suggests that businesses can enhance ad targeting by leveraging demographic features, thus making advertising efforts more efficient and respectful of user privacy. By extending this work to incorporate adaptive learning, real-time processing.



## CHAPTER 10

**REFERENCES**

1. Aaker, D. A., & Shansby, J. G. (1982). Positioning your product. *Business Horizons*, 25(3), 56-62.
2. Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley.
3. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
4. Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, 161-168.
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
8. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
9. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3-24.
10. Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O’Reilly Media.
11. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
12. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.