EdiGlobe

<u>Minor Project Report</u>

<u>Project title</u>: Predictive Modeling for Click-Through Rate Optimization at ConnectSphere Digital.

Submitted by: Rutu Nitin Kulkarni

BE in Electronics and Communication Engineering

KLS GIT, Belagavi.

Submitted to: EdiGlobe.

Internship on Machine Learning with Python.

INDEX:

1. Introduction

In the modern digital era, businesses heavily rely on online advertising to promote their products and services. However, with the vast number of users online, not all advertisements reach the right audience. Many ads are displayed to users who may have little to no interest in engaging with them, leading to poor campaign results and unnecessary expenditure.

ConnectSphere Digital, a digital marketing agency, faces this very challenge. The agency manages multiple campaigns simultaneously and must ensure that client budgets are used wisely. To achieve this, there is a need for a data-driven solution that can predict which users are more likely to engage with advertisements.

This project seeks to solve that challenge by developing a **predictive model** that uses historical user data to identify patterns in user behavior. The ultimate aim is to increase the **Click-Through Rate (CTR)**, which is a key measure of how effectively an ad generates engagement.

2.Problem Statement

At present, a significant portion of ConnectSphere Digital's advertising budget is being wasted on showing ads to audiences who are unlikely to click. For example, users with low online activity or who fall outside the intended demographic continue to receive ads, even though the probability of interaction is minimal.

This inefficient targeting strategy reduces the **Return on Ad Spend (ROAS)**, making campaigns less cost-effective and limiting client satisfaction. In the absence of a **systematic, data-driven method**, the agency's targeting remains broad and imprecise, which affects its overall competitiveness in the highly saturated digital marketing industry.

Thus, there is a clear need to build a solution that can accurately identify "likely clickers" versus "unlikely clickers," ensuring that advertising is focused only on high-value audiences.

3. **Project Goal**

The main goal of this project is to design, implement, and evaluate a **Logistic Regression model** capable of predicting the probability of a user clicking on an advertisement.

The model will classify users into two categories:

- **1 (Likely to Click)** – Users who show strong potential to interact with ads.
- **0 (Unlikely to Click)** – Users who are less likely to engage.

To achieve this, the model will rely on a dataset containing both **demographic factors** and **behavioral metrics**, including:

- Age of the user
- Area Income (average income in the user's geographical area)
- Daily Time Spent on Site
- Daily Internet Usage

By analyzing these features, the model will learn patterns that differentiate engaged users from disengaged ones. Performance will be validated using standard metrics such as **accuracy**, **precision**, and **recall**, ensuring the model is reliable before being applied in real campaigns.

4.Business Objective

From a business perspective, the integration of this model into ConnectSphere Digital's operations has several significant benefits:

- **Optimized Targeting:** Ads can be directed toward individuals with a higher probability of engagement, eliminating unnecessary spend.
- **Higher CTR:** By focusing on quality users, the overall Click-Through Rate across campaigns will rise.
- **Better ROAS:** Clients will achieve more conversions for the same budget, directly increasing return on investment.
- **Improved Client Trust:** Delivering better results ensures stronger client relationships and a competitive advantage in the digital marketing landscape.

In short, the business objective is not only to cut down wasted costs but also to **make advertising more effective, profitable, and client-friendly**.

5. Methodology

The approach to solving the problem involves several systematic steps:

Data Collection:

- o Historical user data is collected, including demographics and online behavior.
- o The dataset used for this project has been sourced from [dataset link].

Data Preprocessing:

- o Handling missing values to ensure data quality.
- o Feature scaling for numerical variables like "Time Spent on Site" and "Internet Usage."
- o Encoding categorical variables, if any, so they can be processed by the model.

Model Development:

- o Logistic Regression is chosen because it is simple, interpretable, and effective for binary classification problems.
- o The model learns patterns in user data to distinguish between likely and unlikely clickers.

Evaluation Metrics:

- o **Accuracy** – Measures how often the model's predictions are correct.
- o **Precision** – Focuses on how many of the predicted "clickers" are actually clickers.
- o **Recall** – Measures how well the model identifies all possible "clickers."

Deployment Possibility:

- o Once validated, the model can be integrated into campaign management systems.
- o Real-time predictions can then guide ad allocation dynamically.

6. **Outcomes**

By implementing this project, the following are outcomes:
- Clear identification of user groups with a **higher tendency to click on ads**.
- A measurable **increase in campaign CTR**, leading to more efficient performance.
- **Reduction of wasted ad spend**, allowing budgets to be allocated more strategically.
- Stronger positioning of ConnectSphere Digital in the competitive advertising industry, as a company that leverages data science to deliver superior results.

These outcomes will not only benefit the agency but also foster **trust and satisfaction among clients**, which is crucial for long-term business sustainability.

Output:

```
Accuracy: 0.98
Precision: 0.9897959183673469
Recall: 0.97
ROC-AUC: 0.9902

Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.99      0.98       100
           1       0.99      0.97      0.98       100

    accuracy                           0.98       200
   macro avg       0.98      0.98      0.98       200
weighted avg       0.98      0.98      0.98       200
```
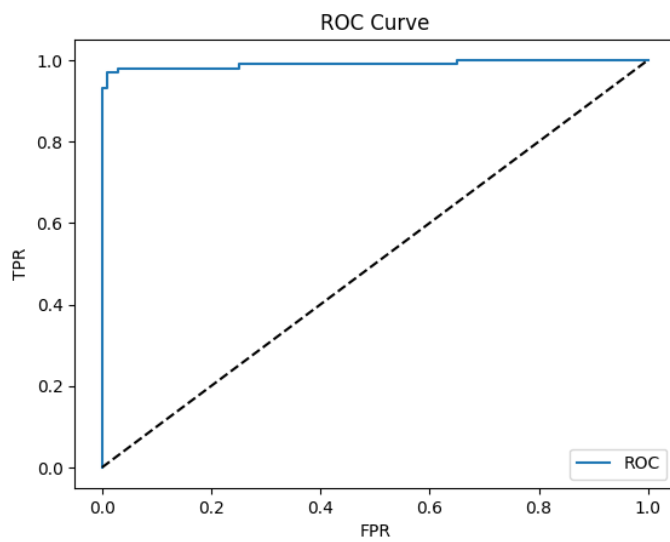
```
Final ROC-AUC: 0.9903
Final Report:
              precision    recall  f1-score   support

           0       0.97      0.99      0.98       100
           1       0.99      0.97      0.98       100

    accuracy                           0.98       200
   macro avg       0.98      0.98      0.98       200
weighted avg       0.98      0.98      0.98       200
```
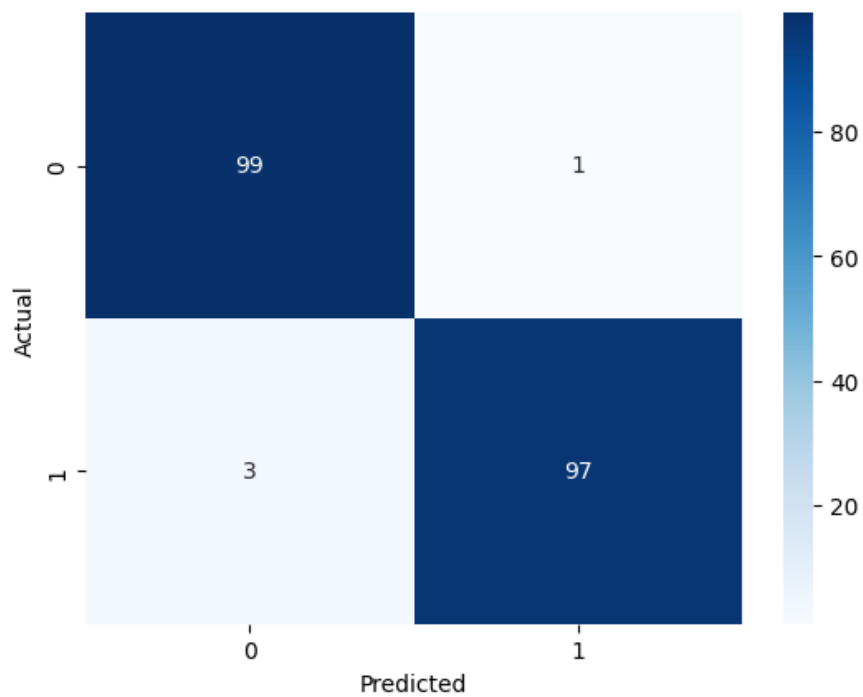
ROC Curve:

ROC Curve

Confusion Matrix:


```
Daily Time Spent on Site    -3.236110
Daily Internet Usage        -2.605517
Area Income                 -1.776209
Age                          1.439396
Male_0                       0.914304
Male_1                       0.292785
dtype: float64
```

7. **Conclusion**

This project demonstrates the power of predictive modeling in transforming online advertising strategies. By applying logistic regression to user data, ConnectSphere Digital gains the ability to distinguish between likely and unlikely clickers with greater accuracy.

Such a solution provides multiple business benefits: optimized spending, higher returns, and improved client relationships. Beyond immediate campaign improvements, this also sets the foundation for the company to adopt more advanced machine learning models in the future, further enhancing competitiveness.

In conclusion, predictive modeling is not just a technical exercise but a **strategic business enabler** that ensures smarter targeting, efficient budget utilization, and higher overall client satisfaction.

ProjectLink:
https://colab.research.google.com/drive/1FHxPBouDzC5KDgk3sV3Orphok9gLvEwv#scrollTo=P86eZfLi8qAm