

# Lookalike Modeling

## Introduction

Lookalike modeling is a machine learning approach used primarily in marketing and recommendation systems. It identifies users or entities with similar characteristics to a given group, enabling targeted campaigns and personalized recommendations.

## Data Processing and Feature Engineering

- **Data Cleaning:** Duplicate records were removed, and missing values were handled appropriately.
- **Feature Engineering:** New features were derived using statistical aggregations, categorical encoding, and interaction terms.
- **Dimensionality Reduction:** Techniques such as PCA (Principal Component Analysis) were applied to remove redundant information and improve model performance.

## Model Selection and Evaluation

Several machine learning models were explored:

1.  
**Logistic Regression:** A baseline model to establish interpretability and performance benchmarks.
2.  
**Decision Trees and Random Forest:** Used to capture non-linear relationships between features.
3.  
**Gradient Boosting (XGBoost, LightGBM):** Advanced tree-based models were tested for optimal accuracy.
4.  
**Neural Networks:** A deep learning approach was considered for complex data structures.

## Results and Insights

- The best-performing model was identified using evaluation metrics like AUC-ROC, Precision, Recall, and F1-score.
- Feature importance analysis revealed the most influential variables affecting lookalike predictions.
- The final model provided high accuracy in predicting lookalike users, making it a valuable tool for marketing and user engagement strategies.

## Lookalike Modeling Notebook (KUMAR\_SAHIL\_Lookalike.ipynb)

This notebook deals with lookalike modeling, which is used in marketing or recommendation systems to identify users with similar behaviors. It likely includes feature engineering techniques to represent user profiles efficiently. The notebook explores supervised or unsupervised learning techniques such as logistic regression, decision trees, or similarity-based clustering. Feature selection and dimensionality reduction techniques (e.g., PCA) might be used to enhance model performance. The final results could include evaluation metrics like precision-recall or AUC-ROC curves, indicating the model's effectiveness in identifying lookalike users.