DATA MINING(Dr Uzma)

Report Design

M Ashir 19122144 Sec B

M Uzair 19122140 Sec B

Customer Conversion Classification Report

Objective:

The objective of this Data Mining Assignment is to develop a predictive model for customer conversion in the context of a Pakistani E-commerce platform. The goal is to leverage machine learning algorithms to predict whether a lead will convert into a customer based on various parameters. The ultimate aim is

to assist businesses in focusing their efforts on leads more likely to convert, thereby boosting overall customer conversion rates.

Introduction and Background of the Problem:

In the domain of E-commerce in Pakistan, understanding and predicting customer conversion are critical for optimizing marketing strategies and resource allocation. The major parameters identified for predicting customer conversion include lead source, time spent on the platform, pages viewed, emails sent, response time, follow-up emails, form submissions, CTR on product pages, social media engagement, and payment history. The successful prediction of customer conversion can significantly impact business growth.

Data Collection:

The dataset used for this analysis is retrieved from Kaggle, specifically tailored to the Pakistani audience for a Pakistani E-commerce platform. The dataset contains information on the identified parameters, providing a foundation for training a predictive model.

Data Preprocessing:

Before training the model, the dataset requires preprocessing. This involves tasks such as handling missing data, encoding categorical variables, scaling numerical features, and potentially addressing outliers. Additionally, exploration and analysis of the dataset will be conducted to gain insights into the distribution and characteristics of the data.

Modeling and Evaluation:

For the task of Customer Conversion Classification, two supervised learning algorithms have been chosen:

1. **Random Forest:**
   - Effective for both classification and regression tasks.
   - Handles non-linearity and interactions between features.
   - Robust and less prone to overfitting, even with a moderately sized dataset.
2. **Naive Bayes:**
   - Simple and computationally efficient.
   - Assumes independence between features, suitable for certain types of data distributions.

The modeling phase involves splitting the dataset into training and testing sets, training the models, and evaluating their performance using appropriate metrics such as accuracy, precision, recall, and F1-score.

Results:

The results of the model evaluation will provide insights into the effectiveness of the chosen algorithms in predicting customer conversion. The performance metrics will help in assessing the reliability and accuracy of the predictive models
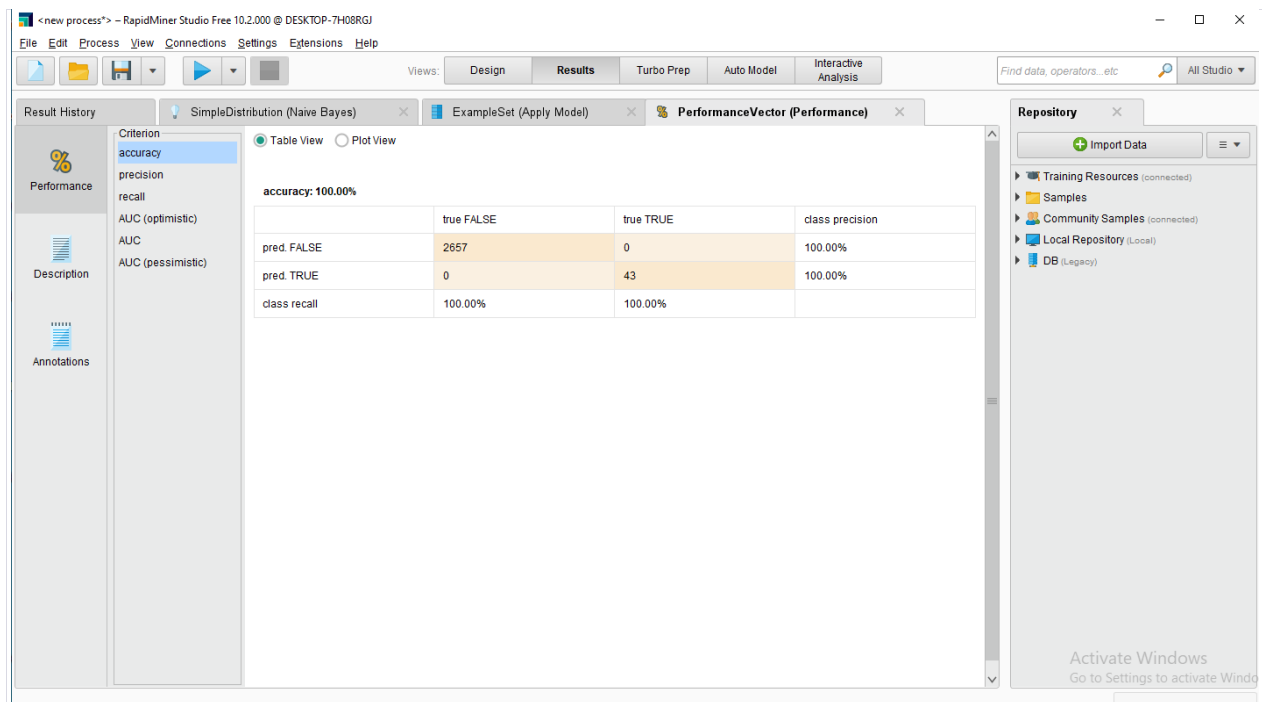
Conclusions:

In conclusion, the development of a predictive model for customer conversion in Pakistani E-commerce involves understanding and utilizing key parameters. The chosen algorithms, Random Forest and Naive Bayes, offer different advantages and are expected to provide valuable insights into lead conversion likelihood. The success of this model can empower businesses to strategically focus their efforts, ultimately enhancing customer conversion rates.