

ALY 6020:

PREDCTIVE ANALYTICS

**Week 3: Logistic Regression Analysis: Subscription Behavior Prediction Models**

Submitted To:

Prof. Chinthaka Pathum Dinesh Herath Gedara, Faculty Lecturer

Submitted By: Abhilash Dikshit

Academic Term: Winter 2024

Northeastern University, Vancouver, BC, Canada

Master of Professional Studies in Analytics

January 27, 2024

**Title: Subscription Behavior Prediction Models**

1. **Objective:**

The goal of this analysis is to predict subscription behavior using logistic regression and support vector machine (SVM) models based on a dataset provided by a magazine company. The dataset contains information about customers, their demographics, and past interactions with the company.

1. **Data Overview:**

The dataset includes features such as customer demographics (age, income, education, marital status), past campaign acceptances, complaints, and various spending behavior metrics. The target variable, "Response," indicates whether a customer accepted the offer in the last campaign.

A screenshot of a data table

Description automatically generated

1. **Data Preprocessing:**

Handled missing values using mean imputation.

Selected relevant features, including 'Year\_Birth', 'Kidhome', 'Teenhome', 'Income', 'MntWines', 'NumWebVisitsMonth', and 'Recency'.

Converted categorical variables to numerical representations.

Split the data into training and testing sets.

1. **Modeling:**

Two models were trained and evaluated: Logistic Regression and Support Vector Machine (SVM).

A screenshot of a computer

Description automatically generated

|  |  |
| --- | --- |
| **Logistic Regression:**  Accuracy: 0.8482  Precision: 0.5263  Recall: 0.1449 | **SVM:**  Accuracy: 0.8594  Precision: 0.6364  Recall: 0.2029 |

**Performance Comparison:**

Accuracy: Both models demonstrated similar overall accuracy, with SVM slightly outperforming Logistic Regression.

Precision: SVM exhibited higher precision, indicating a better ability to correctly identify true positives among predicted positives.

Recall: SVM had a higher recall, suggesting a better ability to capture positive cases among all actual positive cases.

1. **Recommendation:**

If precision and recall are equally important, or if there is a need for balance, the SVM model is recommended due to its higher precision and recall.

If precision is a critical factor, SVM may be preferred for its better performance in minimizing false positives.

Logistic Regression may be preferred for its interpretability, as it provides coefficients for each feature, aiding in understanding the impact of individual variables.

**Next Steps:**

Consider further hyperparameter tuning for both models.

Explore additional feature engineering to enhance model performance.

Evaluate models on a larger dataset if available for more robust conclusions.

1. **Conclusion:**

The SVM model, with its slightly superior accuracy, precision, and recall, appears to be a promising choice for predicting subscription behavior. However, the final decision should be based on the specific business goals, interpretability needs, and the importance of precision and recall in the context of the magazine company's objectives.