

PRESENTATION

Predicting H1N1 and Seasonal Flu Vaccination

PROJECT OVERVIEW

The project analyzes what drives people to get vaccinated and also predict whether individuals receive the H1N1 and Seasonal flu vaccines using demographic, behavioural and opinion data.

The goal is to build a reliable multilabel classification with insight that can help guide future public health vaccination campaign.

BUSINESS UNDERSTANDING

This project aims to predict whether people received the H1N1 and Seasonal vaccine by uncovering how demographic, beliefs and behaviours affect vaccination decision.

The project helps public health agencies identify hesitant groups, design targeted campaigns and improve resource allocation.

The final model will support data driven strategies to boost vaccination rate and prevent future disease outbreak.

MODELLING

- Target Variable : H1N1_vaccines and seasonal_vaccine.
- Feature Selection : Used all the rest of 35 columns
- Used one-Hot Encoding to nominal categorical variables and ordinal Encoding for Ordinal categorical variables.
- Split the data and divided the preprocessed data into training 80% and Testing 20% using train_test_split, ensuring stratification to maintain the proportion of target classes in both sets.

Logistic Regression Modeling

Pipeline Construction : biult a pipeline using ColumnTransformer for preprocessing, standardScaler for numerical feature scaling and MultiOutputClassifier with LogisticRegression for classification task.

Training : Each model (H1N1 and Seasonal) was trained independently on its respective target variable.

Decision Tree Classifier Modeling

- Preprocessing : I used one-Hot encoding to all categorical variables including ordinal
- Model Initialization : DecisionTreeClassifier was initialized for both H1N1 and Seasonal flu.
- Training : Each Decision Tree model was trained separately on its corresponding target variable.

Evaluation

Accuracy : The proportion of correctly classified instances

ROC AUC Score : Measures the model's ability to distinguish between classes. A higher AUC indicates better discriminatory power

Precision : Of all instances predicted as positive, how many were actually positive?.

Recall : Of all actual positive, how many were correctly identified ?.

F1-Score : the harmonic mean of precision and recall, providing balance measure.

Confusion Matrix : A table showing counts of true positives, true negative, false positive and false negative

Logistic Regression Model Evaluation

- **H1N1 Vaccine Model:**
- **Interpretation** : while overall accuracy was good, the key focus was on recall for the vaccinated class (class 1)
- **Imbalance Handling** : The use of SMOTE successfully increased the Recall from 0.43 to 0.73 for the vaccinated class, demonstrating a much improved ability to identify individuals who received the H1N1 vaccines
- **ROC AUC** : High value of 0.839 indicating a strong class separability

Seasonal Vaccine Model

Interpretation : Exhibited balanced precision and recall for both vaccinated and non-vaccinate class

ROC AUC : high value of 0.854 confirms a strong performance

DecisionTree Classifier

- H1N1 Vaccine Model:
- Interpretation : Showed significant weakness in predicting H1N1 vaccination. Low recall (0.45) and precision (0.42) for the vaccinated class indicate it frequently missed vaccinated individuals and had many false positives.
- ROC AUC : (0.642) confirms its limited discriminatory power
- Seasonal Vaccine Model:
- Interpretation : performed better than the H1N1 Decision Tree, with more balanced precision and recall.
- ROC AUC : was moderate at (0.690)

RECOMMENDATION

- I would recommend leveraging the logistic Regression Models for predicting both H1N1 and Seasonal flu vaccination mainly because:
- **Superior Predictive Power:** Achieved a consistently higher ROC UAC Scores (H1N1 = 0.839, Seasonal = 0.854) compared to Decision Tree
- **Enhanced H1N1 Identification :** Critically, for H1N1 (the minority class), Logistic regression with class_weight='balanced' significantly boosted the recall for the vaccinated group to 0.73. This greatly improves the ability to find individuals who received the H1N1 vaccine.
- **Balanced Seasonal Flu Performance :** maintained strong and balanced performance across both vaccinated and non-vaccinated class for flu, ensuring reliable general prediction