

Predict H1N1 and Seasonal Flu Vaccines

From backgrounds, opinions, and health behaviors - Multi target classification case

Predict H1N1 and Seasonal Flu Vaccines

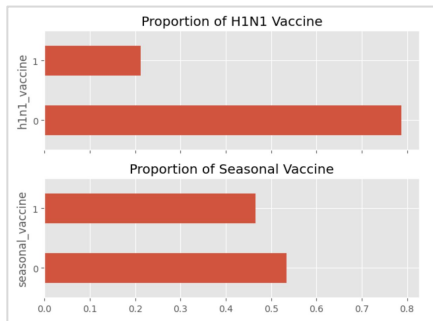
Predecir si las personas se vacunaron contra la gripe H1N1 y la Gripe Estacional usando la información que compartieron sobre sus antecedentes, opiniones y conductas de salud.

<https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/>

Datos claves del dataset

1. Contexto: Data Science for social good
2. Dataset público
3. Origen, United States National Center for Health Statistics
4. 26707 rows, 36 cols
5. Variables numéricas: 24
6. Variables categóricas: 12
7. Variables target: 2

2 Variables target H1N1 / Seasonal



La mitad de la población recibió la vacuna de gripe estacional (clases balanceadas), pero solo un 20% la de H1N1 (clases moderadamente desbalanceadas).

EDA y calidad del dato

- Número final de variables: 49
- Label/One-Hot Encoder para convertir las variables categóricas en numéricas.
- Eliminadas variables con datos irrelevantes o baja correlación con las variables target.

Hyperparameters tuning with GridSearchCV

→ Scoring: **roc_auc**

→ Algoritmos evaluados:

1. Logistic Regression
2. Random Forest
3. SVM (Support Vector Machines)
4. Naive Bayes
5. Decision Tree
6. K-Neighbors

Training Results

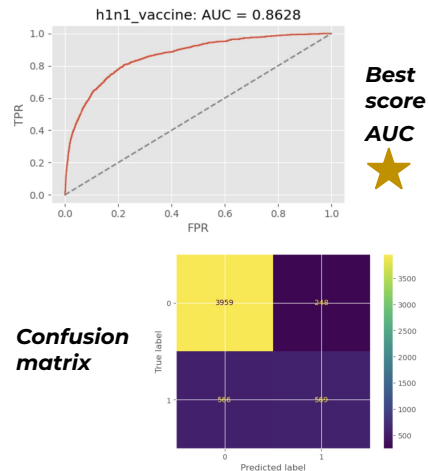
Manejo desbalance con `class_weight`

Algoritmos TOP 3

1. **H1N1 / Seasonal: SVM** ★
2. **H1N1 / Seasonal:** Random Forest
3. **H1N1:** Logistic Regression / **Seasonal:** K-Neighbors

HIGHEST SCORES		
SPLIT	H1N1	SEASONAL
80-20%	SVM balanced ★ AUC = 0.862754	SVM balanced AUC = 0.855847
70-30%	SVM balanced AUC = 0.858395	SVM balanced ★ AUC = 0.858624

* SVM balanced = Ejecución con `class_weight='balanced'`.



Conclusiones

- EDA es muy importante para lograr un dataset óptimo para el training.
- Mejor algoritmo, **SVM(C=1, gamma=0.01, kernel=RBF)**, por su versatilidad y robustez ante la presencia de outliers.
- El desbalance fue tratado con `class_weight=balanced`, mejorando el score logrado con SVM.