

PREDICTING VACCINATION LIKELIHOOD

H1N1 AND SEASONAL FLU

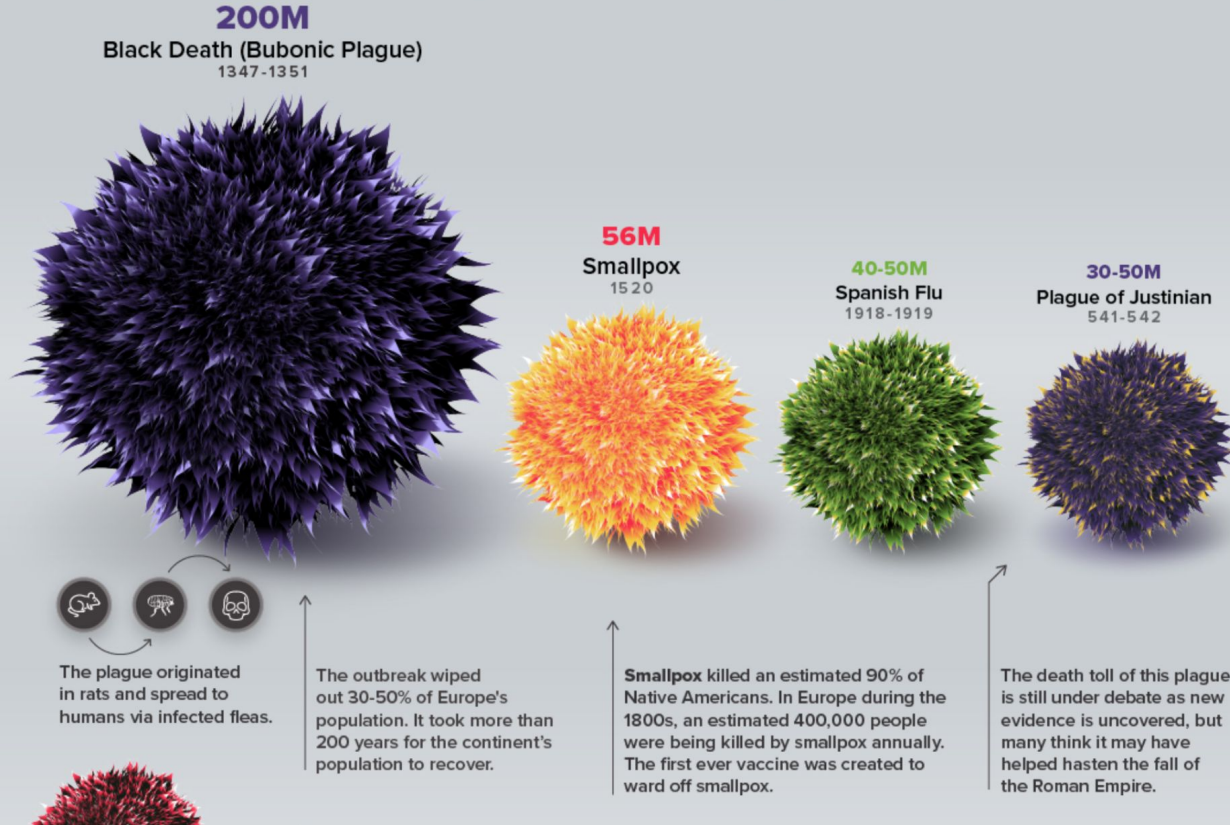


By: Chisum Lindauer and Greg Fatouras

“Prevention is better than a cure.” – Hippocrates

DEATH TOLL

[HIGHEST TO LOWEST]

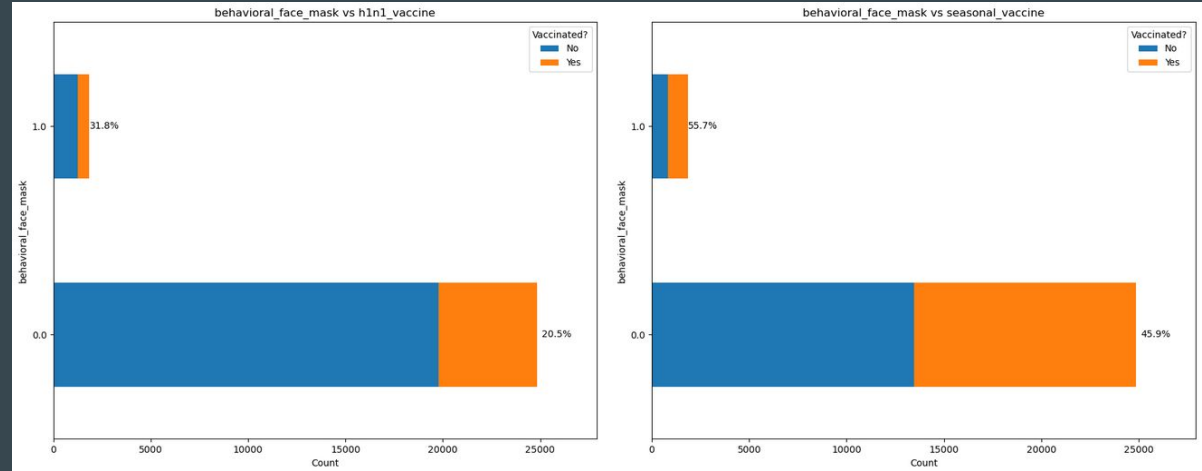


National 2009 H1N1 Flu Survey

- + 35 features
- + 26,707 entries
- + yes/no questions
- + categorical questions

Did They Get Vaccinated?

- + H1N1: The Swine Flu (2009)
- + Seasonal: Yearly Flu



Limitations

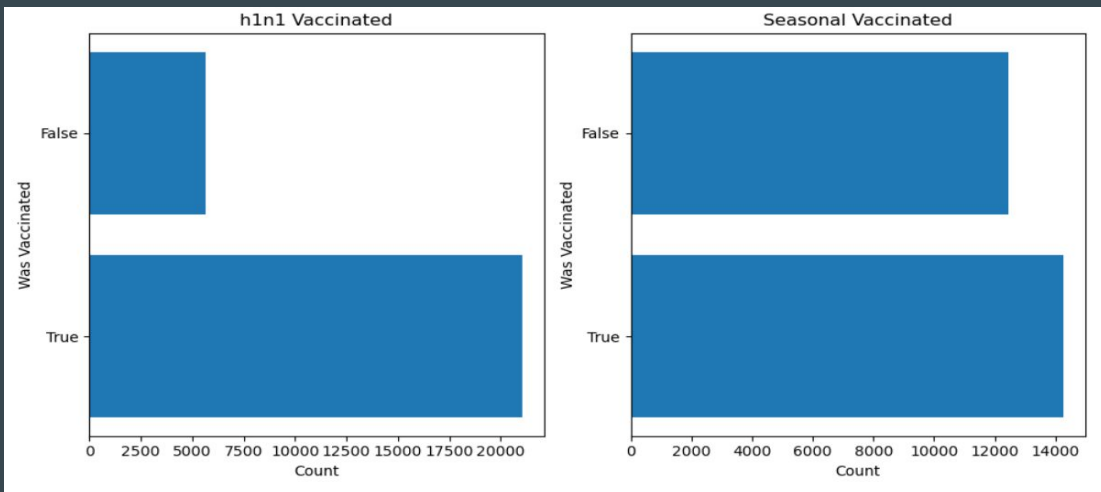
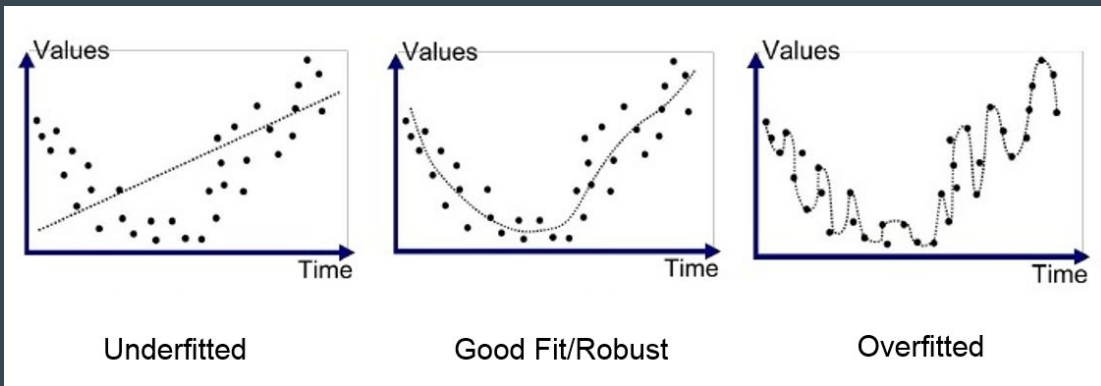
Overfitting - Cross Validation

Class Imbalance - SMOTE

Missing Data - Impute

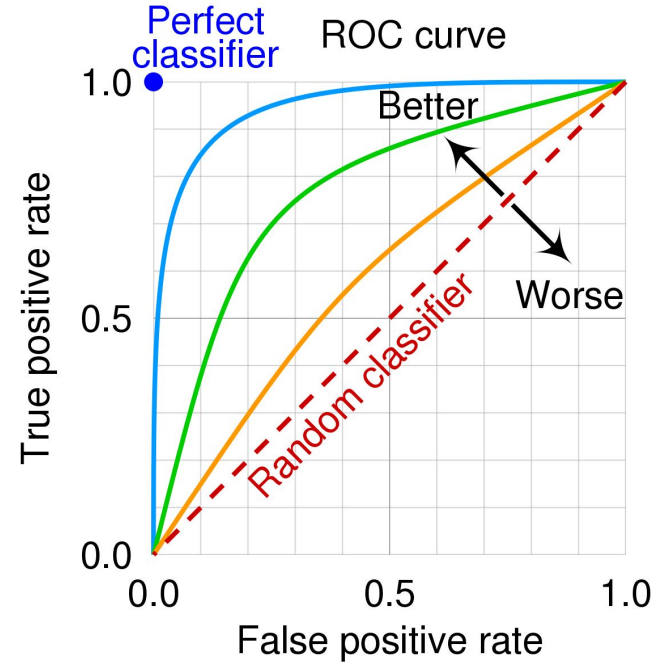
Dated Data - 2009

Population Source - US Bias

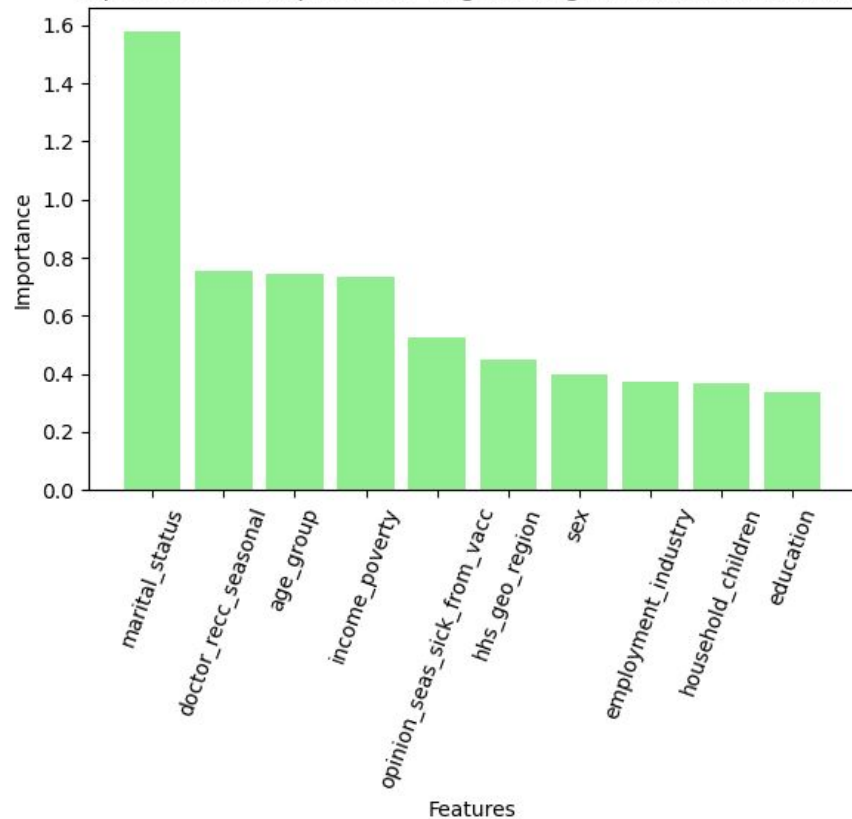


Predicting Vaccinations

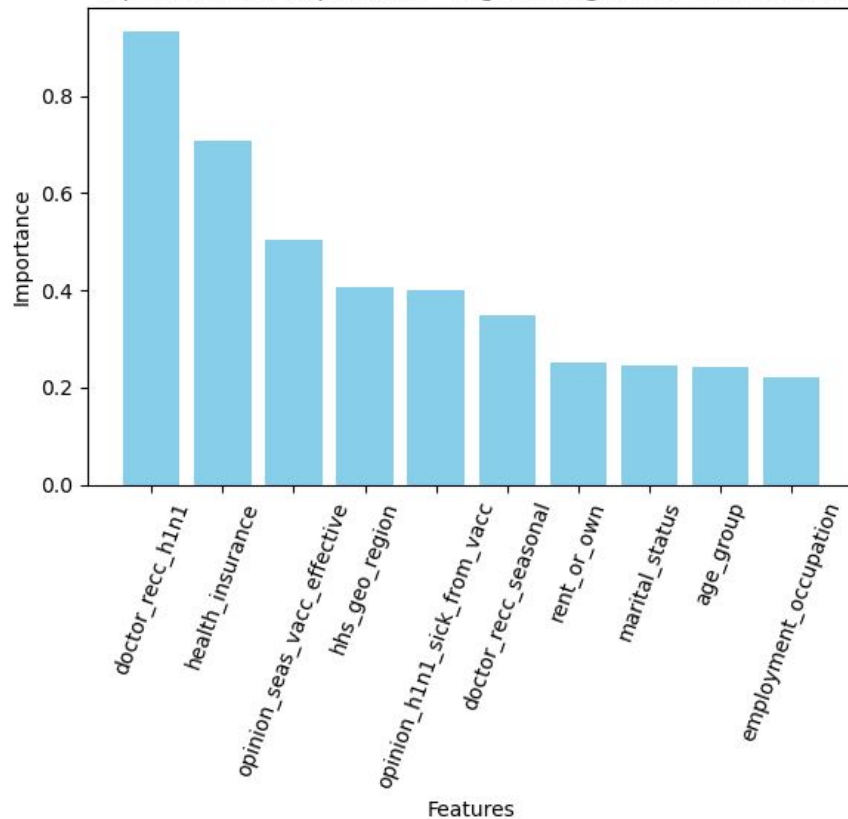
		True Class	
		Positive	Negative
Predicated Class	Positive	TP	FP
	Negative	FN	TN



Top 10 Feature Importance - Logistic Regression Seasonal Model



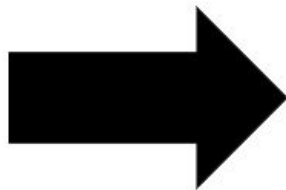
Top 10 Feature Importance - Logistic Regression H1N1 Model



Rationale

Training

Extract patterns from data



Evaluating

Use patterns to predict results



Preprocessing

Change Parameters to Preprocessing

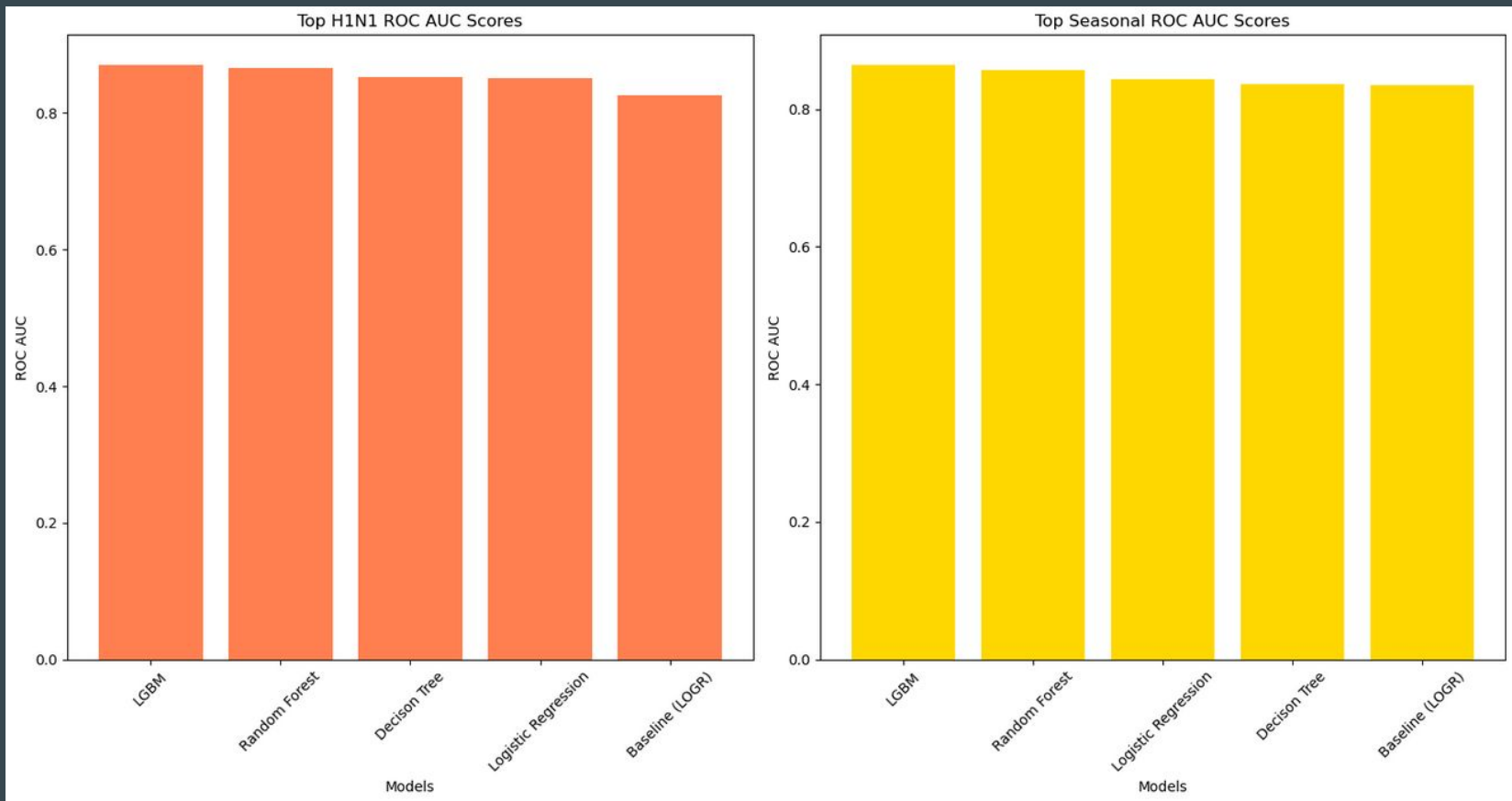
- + SMOTE
- + Impute Strategy
- + Scaler Type
- + Feature Selection
- + One Hot Encoding

Best Results?

- + SMOTE: Yes
- + Impute: -1 and None
- + Features: Top 87%
- + Scaler: Standard
- + OHE: Yes

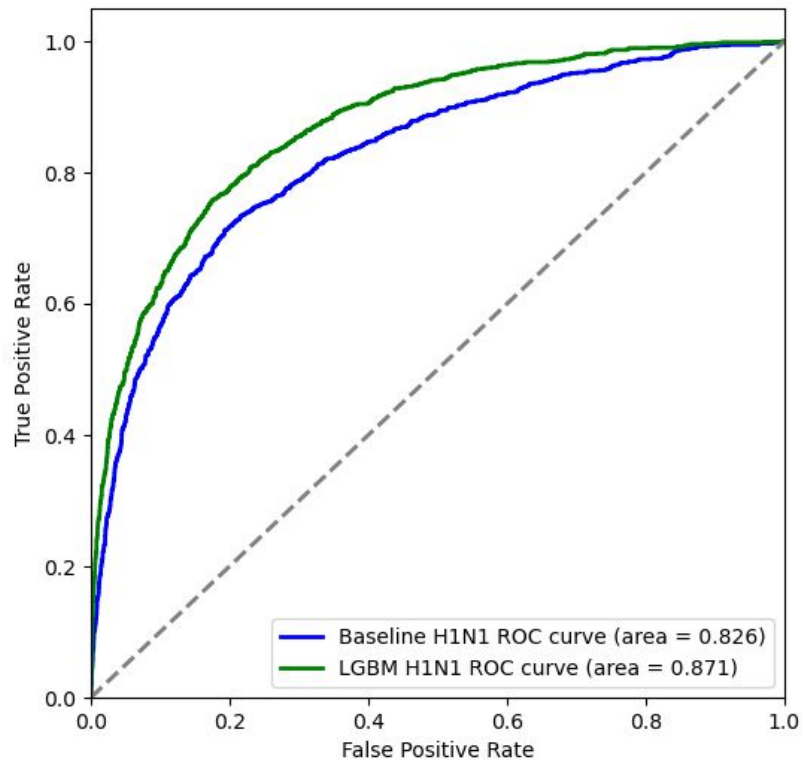


Predictive Models

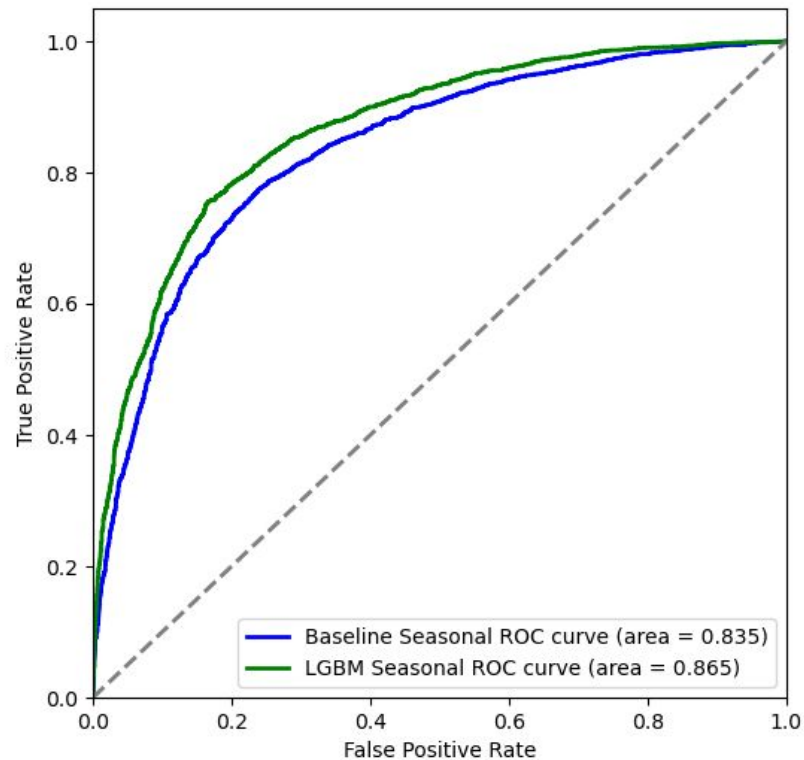


Results

ROC Curve - H1N1

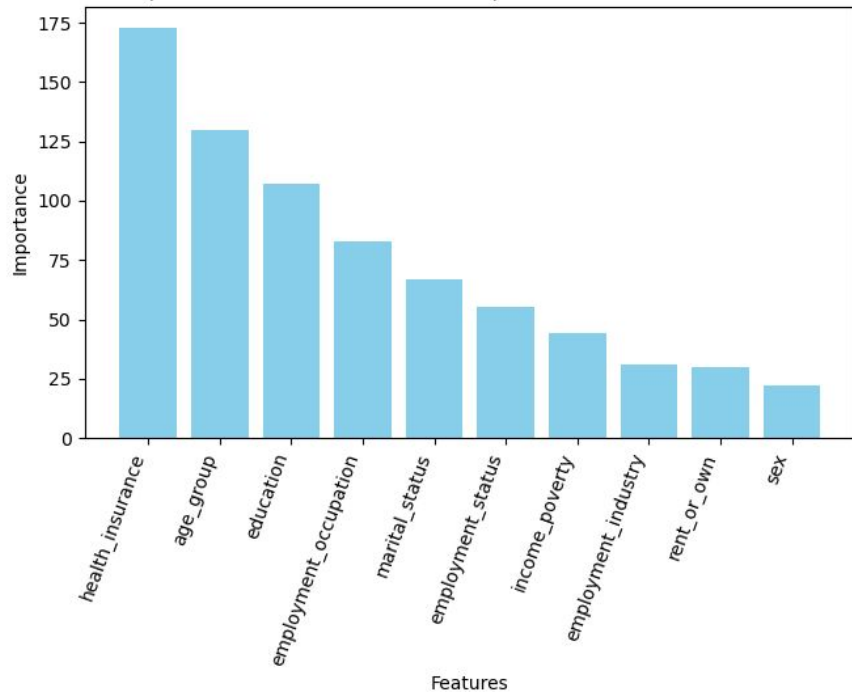


ROC Curve - Seasonal

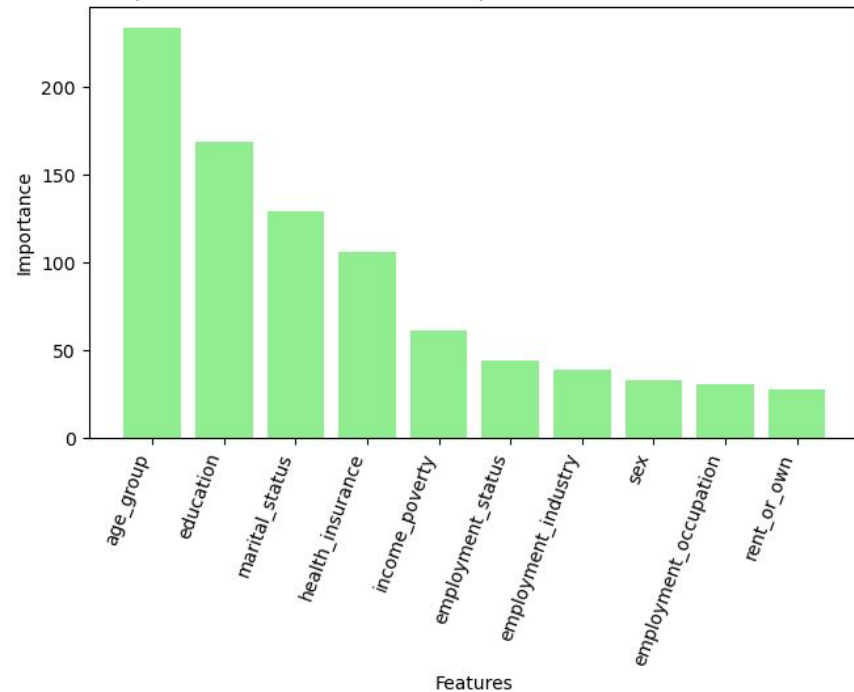


Socioeconomic Factors

Top 10 Socioeconomic Feature Importance - LGBM H1N1 Model



Top 10 Socioeconomic Feature Importance - LGBM Seasonal Model



Recommendations

Provide Free Vaccines To Those Without Health Insurance

Targeted Vaccination Drives To Low Vaccination Areas

Encourage Low-Risk Groups To Vaccinate

Educate and Inform Low Vaccination Groups

Train Doctors To Recommend H1N1 Effectively

Next Steps

Predict with historical and real-time data

Incorporate Survey Responses

Predict Outbreak Locations

Identify Unvaccinated Populations

Target at Risk Populations

Make a Recall Focused Model

Integrate Predictions from EHR

Visually Communicate Data With the Public

THANK YOU

Greg Fatouras:

<https://github.com/Gfatouras>

<https://www.linkedin.com/in/gfatouras/>

fatourasg@gmail.com

Chisum Lindauer

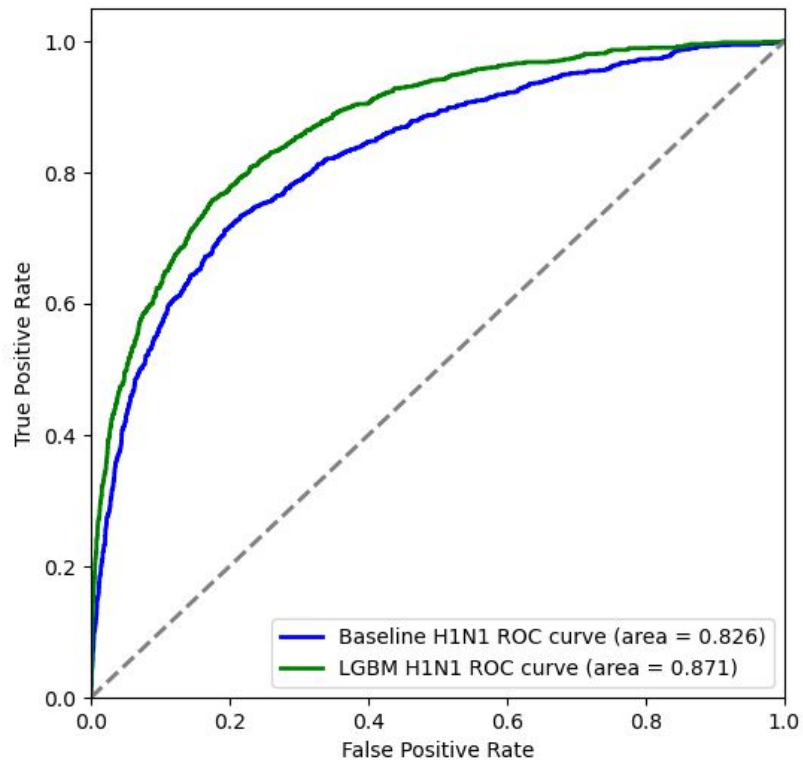
<https://www.linkedin.com/in/chisum-lindauer/>

<https://github.com/minerva-ds/>

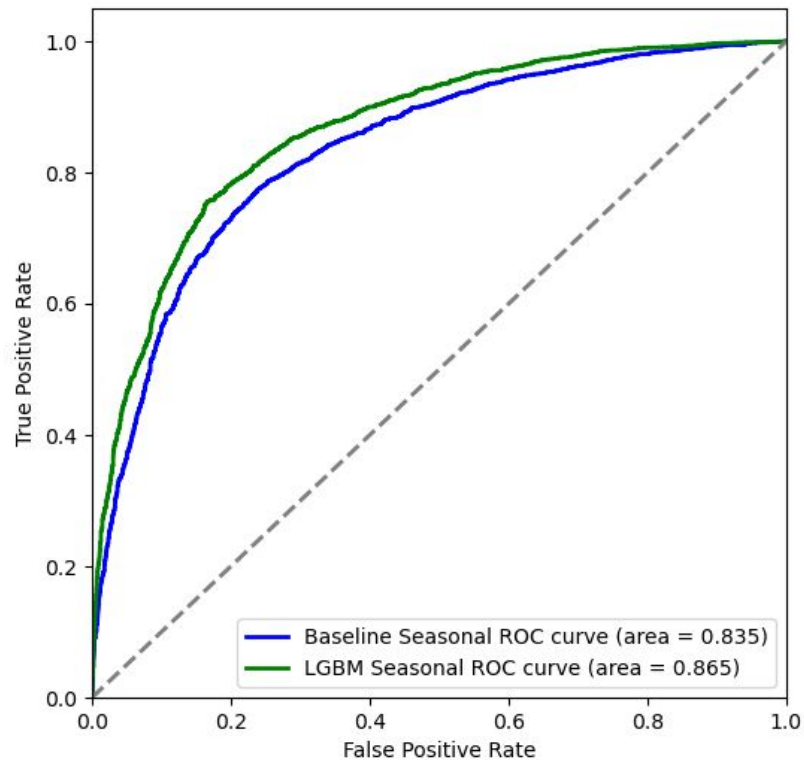
chisum@atrixtech.com

Baseline vs LGBM

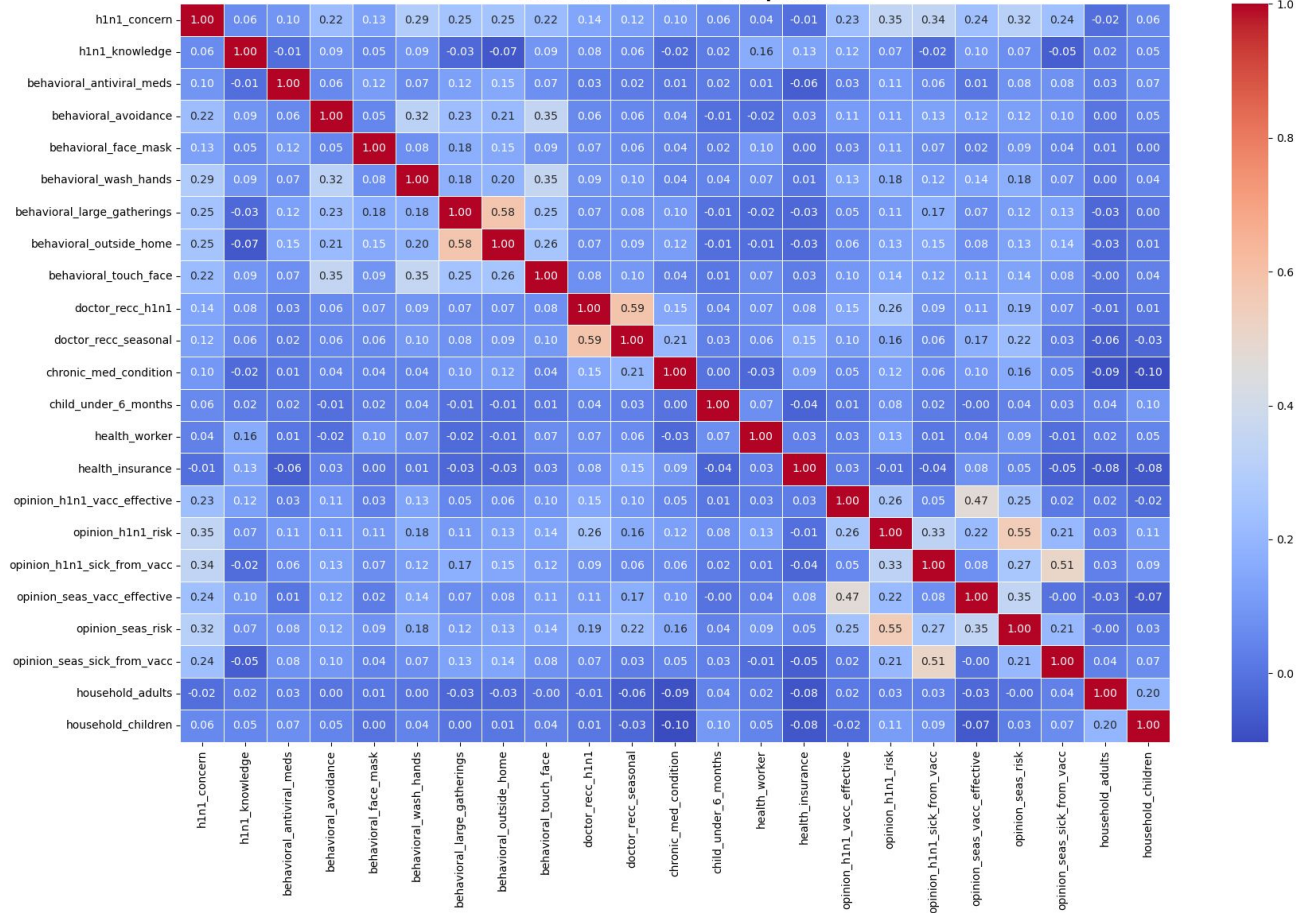
ROC Curve - H1N1



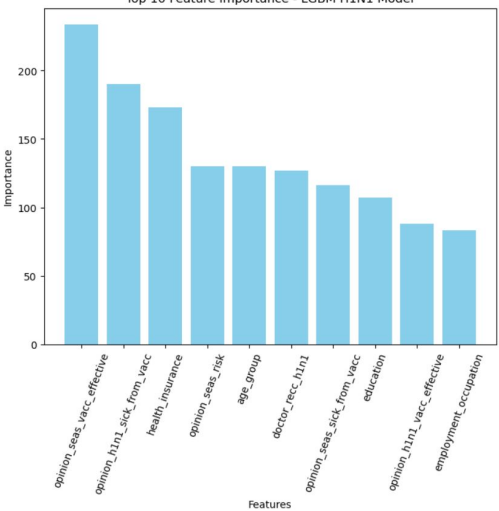
ROC Curve - Seasonal



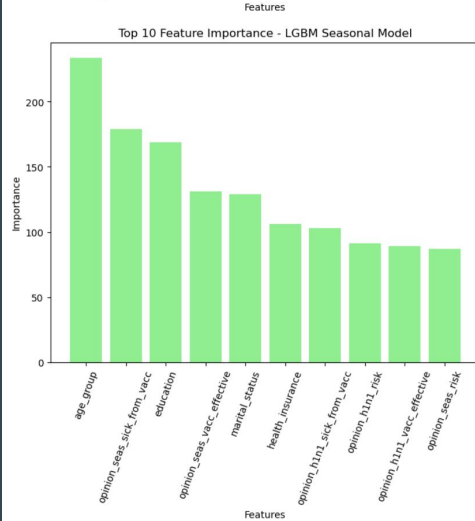
Correlation Matrix of Responses



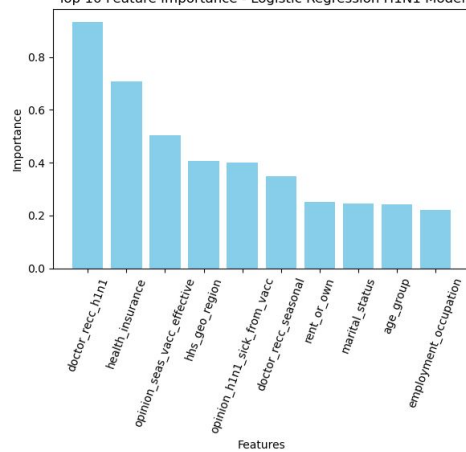
Top 10 Feature Importance - LGBM H1N1 Model



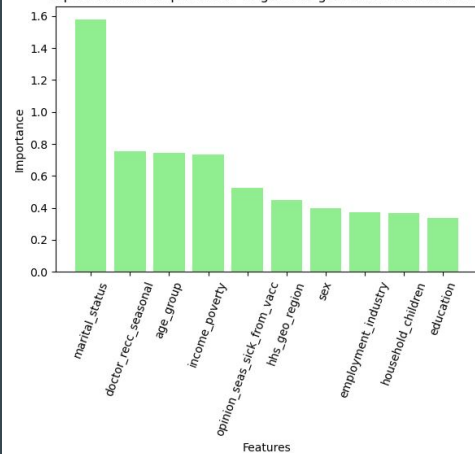
Top 10 Feature Importance - LGBM Seasonal Model



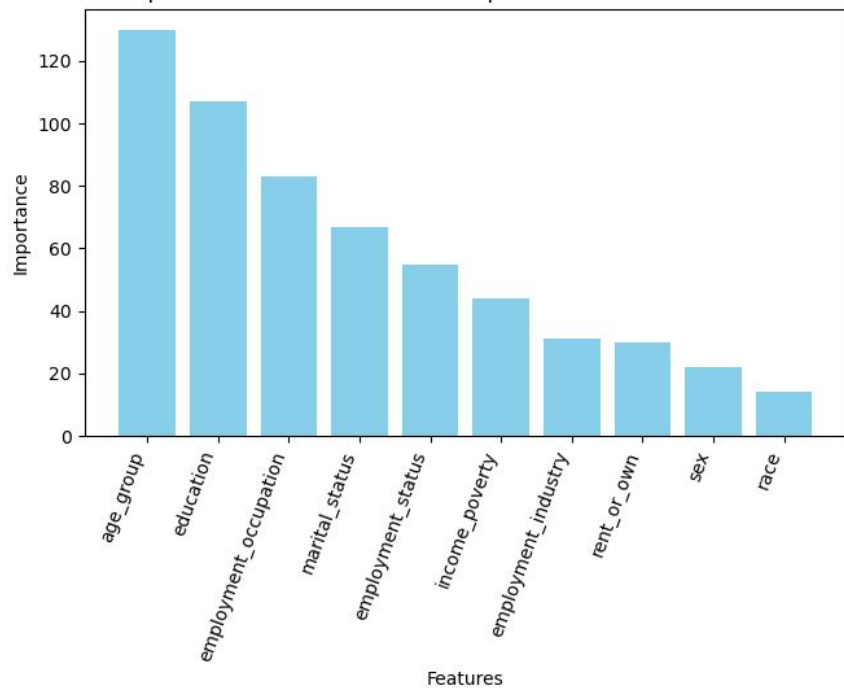
Top 10 Feature Importance - Logistic Regression H1N1 Model



Top 10 Feature Importance - Logistic Regression Seasonal Model



Top 10 Socioeconomic Feature Importance - LGBM H1N1 Model



Top 10 Socioeconomic Feature Importance - LGBM Seasonal Model

