




The primary business problem is to predict the likelihood that individuals will receive the seasonal flu vaccines in the **United States** based on their backgrounds, opinions, and health-related behaviors. By understanding the factors that influence vaccine uptake, we can help the public health organizations and healthcare providers to design more effective vaccination campaigns and help policy makers allocate resources more efficiently. This will ultimately improve vaccination rates thus reducing the spread of these viruses.

## Important Features used

- This was based on the important features selected by our best performing model **Decision Trees Model**  [Important Features](#)

## *Specific Questions to Address*

1. *Which factors most influence the decision to get vaccinated?*

- Identifying key demographic, behavioral, or opinion-based variables that are strong predictors of vaccine uptake.

2. *Can we accurately predict who is likely to get vaccinated?*

- Building a predictive model to classify individuals as likely or unlikely to get vaccinated.

3. *How can public health interventions be targeted more effectively?*

- Using the model's insights to inform public health strategies, such as targeted communication or incentives for groups less likely to get vaccinated.

## *Use Case or Stakeholders to our model*

This analysis could be used by:

- **Public Health Departments:** To identify segments of the population that are less likely to get vaccinated and develop targeted interventions.

- **Healthcare Providers:** Can use the model to identify patients who are less likely to get vaccinated and engage in targeted outreach.
- **Policy Makers:** To allocate resources and plan vaccination drives more effectively based on predicted uptake.

## ***Outcome***

The expected outcome is a **predictive model** that helps in identifying the likelihood of individuals receiving the vaccine, which can then be used to tailor public health campaigns and interventions, ultimately aiming to increase vaccination rates and reduce the spread of illness.

## ***Potential Impact***

- **Improved Vaccination Rates:** By identifying the determinants of vaccine uptake, targeted interventions can be designed to increase vaccination rates, leading to better public health outcomes.
- **Resource Optimization:** Health departments can allocate resources more effectively by focusing efforts on groups identified as less likely to get vaccinated.
- **Policy Development:** Data-driven policies can be developed to address barriers to vaccination, such as misinformation or access issues.

## **Methods used to clean data**

---

- Imputer
- One Hot Encoding
- Log Transformation to handle skewness

## **Models used to train our features**

---

- Logistic Regression
- Decision Trees Classifiers

## Evaluation

---

- ROC Curve
- Confusion Matrix

## Conclusion

---

By leveraging Logistic Regression for feature selection and using Random Forest for the final predictive modeling, I have developed a streamlined and focused model that is based on the most relevant features. The evaluation of the Random Forest model indicates how effectively these features contribute to accurate predictions. This approach balances both feature selection with model evaluation, ensuring that the final model is both efficient and effective.

## Recommendations

---

### To All Our Stakeholders

-- Using the features provided by our models, new policies can be made, resources allocated, and plans made to drive vaccination based on them. -- Stakeholders are now able to use the features to come up with incentives for the groups that are less likely to be vaccinated thus curbing the spread of the virus. -- Public health departments can now use this features to their advantage in developing target interventions in specific populations whereby the feature depicts that they are lesss likely to get the vaccine