



0 stars

0 forks

1 watching

1 Branch

0 Tags

Activity

Public repository


Anniemarie001

updated notebook and presentation

8225095 · 1 minute ago

.ipynb_checkpoints	notebook in pdf format	6 minutes ago
Data	outliers and feature distributions	4 days ago
H1N1 VACCINE PRESENTATI...	updated notebook and presentation	1 minute ago
H1N1_IMG.jpg	README File	2 days ago
H1NI.jpg	README File	2 days ago
README.md	final notebook and README submis...	26 minutes ago
index - Jupyter Notebook.pdf	notebook in pdf format	6 minutes ago
index.ipynb	notebook in pdf format	6 minutes ago

README



Project Overview

A vaccine for the H1N1 flu virus became publicly available in October 2009. In late 2009 and early 2010, the United States conducted the National 2009 H1N1 Flu Survey. This phone survey asked respondents whether they had received the H1N1 and seasonal flu vaccines, in conjunction with questions about themselves. These additional questions covered their social, economic, and demographic background, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission. A better understanding of how these characteristics are associated with personal vaccination patterns can provide guidance for future public health efforts.

OBJECTIVES

Identify Key Predictors: Determine the most significant factors that influence whether an individual received the H1N1 vaccine.

Develop a Predictive Model: Build a binary classification model to accurately predict whether a survey respondent received the chosen vaccine.

Evaluate Model Performance: Assess the model's performance using appropriate metrics such as accuracy, precision, recall, AUC and F1 score to ensure its reliability in predicting vaccination behavior.

Provide Actionable Insights: Analyze the model's findings to provide public health authorities with actionable insights that can guide future vaccination campaigns and strategies, particularly in the context of managing public health responses to pandemics.

Dataset Description

The dataset was sourced from the [DrivenData Flu Shot Learning competition](#) and is composed of three primary files: `train.csv`, `test.csv`, and `labels.csv`.

Structure:

- **Train and Test Data:**
 - Both the `train` and `test` datasets contain 26,707 entries with 36 columns each.
 - The first column, `respondent_id`, is a unique identifier for each respondent.
 - The remaining 35 features encompass a range of demographic, behavioral, and health-related data points. These features are further described in the accompanying `feature_description.txt` file.
- **Labels Data:**
 - The `labels.csv` file includes two target variables:
 - a. **h1n1_vaccine:** Indicates whether the respondent received the H1N1 flu vaccine (0 = No, 1 = Yes).
 - b. **seasonal_vaccine:** Indicates whether the respondent received the seasonal flu vaccine (0 = No, 1 = Yes).

Target Variable:

- For this analysis, we focus on the `h1n1_vaccine` variable, aiming to model and predict whether respondents received the H1N1 vaccine based on the 35 features provided.

This dataset provides a rich source for analyzing factors that influence vaccine uptake, with a well-structured combination of features that support comprehensive modeling and prediction efforts.

Project Structure

The project is organized into several key directories and files for ease of navigation and understanding:

- **data/** : This directory contains the dataset files.
 - `training_set_features.csv` : The main dataset file with features data.
 - `test_set_features.csv` : The test dataset for testing the model performance on unseen data.
 - `feature_description.md` : Descriptions of the dataset columns.
- **notebooks/** : Jupyter Notebooks for analysis and modeling.
 - `index.ipynb` : The primary notebook containing the data analysis, visualization, and modeling steps.
- **src/** : Source code for custom functions and utilities used within the notebooks.
 - `data_preprocessing` : Functions for data cleaning and preparation.
 - `feature_engineering` : Functions for creating new features from the existing data.
 - `model_evaluation` : Utilities for evaluating model performance.
- **requirements.txt** : A list of Python packages required to run the project.
- **LICENSE** : The MIT License file.
- **README.md** : The project overview, setup instructions, and additional information.

Methodology

The project follows a structured data science process:

- **Data Collection and Inspection**: Gather and inspect the provided dataset.
- **Data Cleaning and Preparation**: Handle missing values, outliers, and incorrect data types.
- **Exploratory Data Analysis (EDA)**: Analyze the data to find patterns, relationships, and insights.
- **Modeling**: Build predictive models to predict vaccine uptake based on selected features.
- **Model Evaluation**: Assess the models' performance using appropriate metrics.
- **Interpretation**: Draw conclusions from the model results and provide recommendations.

Detailed Modeling Section

The modeling phase of this project involves several key steps, from feature selection to model evaluation, aimed at building a robust classification model to predict house prices. Here's a detailed breakdown:

1. Feature Selection

- Initial features were selected based on their expected impact on vaccination, as identified during the exploratory data analysis (EDA) phase.
- Correlation analysis was performed to identify highly correlated features that might cause multicollinearity.
- A combination of domain knowledge and statistical tests (e.g., ANOVA for categorical variables) helped refine the feature set.

2. Data Preprocessing

- Categorical variables were encoded using one-hot encoding to convert them into a format that could be provided to the model.

- Continuous variables were scaled to have a mean of 0 and a standard deviation of 1, ensuring that no variable would dominate the model due to its scale.

3. Model Building

- A linear regression model was chosen as the starting point due to its simplicity, interpretability, and the linear relationship observed between many predictors and the target variable during EDA.
- The model was implemented using the `Logistic Regression` and `Decision Trees` class from `scikit-learn`.

4. Model Training

- The dataset was split into training and testing sets, with 80% of the data used for training and 20% for testing, to evaluate the model's performance on unseen data.
- The model was trained using the training set, fitting it to predict the house prices based on the selected features.

5. Model Evaluation

- The model's performance was evaluated using several metrics, including accuracy, AUC score, precision, recall and F1 score to assess its accuracy and explanatory power.
- A comparison was made between the training and testing set performances to check for overfitting.

6. Model Interpretation

- The coefficients of the model were analyzed to understand the impact of each feature on the vaccine uptake, providing insights into the public health authorities.

7. Next Steps

- Based on the initial model's performance, further steps could include exploring more complex models, such as random forest or ensemble methods, and conducting feature engineering to uncover more nuanced relationships within the data.

This detailed approach to modeling ensures a thorough understanding of the factors influencing vaccine uptake, providing a solid foundation for making informed public health vaccination strategies.

Tools and Libraries Used

The project utilizes a range of tools and libraries for data analysis and machine learning. Here are the key technologies used:

- **Data Analysis and Manipulation:**

-  **PANDAS**
-  **NUMPY**

- **Visualization:**

-  **MATPLOTLIB**

- **Machine Learning:**

-  **SCIKIT LEARN**

How to Run the Project

Before you begin, ensure you have Python 3.8 or later installed on your machine. Follow these steps to set up and run the project:

1. Clone the Repository:

- Open your terminal and run:

```
git clone https://github.com/Anniemarie001/H1N1-AND-SEASONAL-FLU-VACCINES.git  
cd your-repository-directory
```



2. Install Dependencies:

- Ensure pip is up to date:

```
python -m pip install --upgrade pip
```



- Install project dependencies:

```
pip install -r requirements.txt
```



3. Launch Jupyter Notebook:

- Run the following command to open the project in Jupyter Notebook:

```
jupyter notebook index.ipynb
```



- Run the cells sequentially to reproduce the analysis and results.

Data Understanding

The initial phase of the project involves importing necessary libraries such as pandas, numpy, matplotlib, seaborn, and various sklearn modules for preprocessing, model selection, and evaluation. The dataset is then loaded for inspection, revealing it contains 26,707 rows and 36 columns, encompassing a wide range of different individual opinions.

Conclusion and Recommendations

This project aims to provide a thorough analysis of the H1N1 vaccine uptake to support public health officials and policymakers in making informed decisions regarding vaccination strategies and public outreach. Through detailed exploratory data analysis and model development, we have identified the key factors that significantly influence individuals' decisions to receive the H1N1 vaccine.

Recommendations

Targeted Interventions: Focus educational campaigns on older adults, especially those aged 65 and above, to increase vaccine uptake in this demographic.

Enhance Doctor Involvement: Encourage healthcare providers to play a more active role in recommending the vaccine, as their influence is crucial for increasing vaccination rates.

Address Misconceptions: Develop public health campaigns to dispel myths and boost confidence in the vaccine's effectiveness, emphasizing clear and evidence-based information.

Conduct Further Analysis: Explore additional health-related factors or interactions that might influence vaccination decisions, particularly for those with chronic medical conditions.

License



Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%