# InfectiShield ML Application



CONTENTS

**[InfectiShield ML Application 1](#_Toc31268)**

[Background 2](#_Toc9827)

[Proposed Solution - “InfectiShield” 3](#_Toc24211)

[Building the Solution 4](#_Toc9450)

[Limitation and Constraints 4](#_Toc21538)

[Conclusion 4](#_Toc4856)

## Background

The World Health Organization (WHO) has conducted numerous surveys that are crucial for promptly responding to spikes in mortality rates and potentially saving many lives. Frequently, deaths occur for reasons that remain unknown, prompting various speculations. However, research indicates that a significant increase in fatalities is often attributed to the rapid spread of viruses, often unidentified, posing a serious public health concern.

One such viral illness is 'Influenza’, or ‘the flu’ which can manifest with symptoms like fever, headaches, cough, cold, fatigue, among others. While some individuals recover within a week, others may face severe complications that can ultimately lead to fatality. To address these challenges, the development of a dependable application could prove invaluable. Such an application would empower individuals to identify their health issues promptly and seek appropriate medical attention.

Influenza can pose a significant risk during pregnancy, leading the American College of Obstetricians and Gynecologists (ACOG) and the Advisory Committee on Immunization Practices (ACIP) to advise that pregnant women should receive the influenza vaccine if they will be pregnant during the flu season, regardless of the trimester (1,2). During the 2009–10 flu season, pregnant women faced an elevated risk of severe illness and mortality due to infection with the influenza A (H1N1)pdm09 (pH1N1) pandemic virus. (see: [Influenza Vaccination Coverage Among Pregnant Women — 29 States and New York City, 2009–10 Season (cdc.gov)](https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6107a1.htm?s_cid=mm6107a1_e)

## Proposed Solution - “InfectiShield”

“InfectiShield” aspires to provide insights that illuminate the origins of health concerns and fatalities. Harnessing the potential of cutting-edge Big Data methodologies, the application utilizes extensive data to produce comprehensive health assessments.

The utilization of Big Data technology holds tremendous promise in predicting influenza outbreaks, enabling the efficient processing, seamless integration, and thorough analysis of extensive and varied datasets.

The outcome will be presented through an informative dashboard, featuring predictions derived from a meticulously trained dateset alongside charts, graphs and other visuals.

## Building the Solution - Architecture of the Application

### Downloading the Data

The Dataset used to build InfectiShield was downloaded from the [CDC website](https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6107a1.htm?s_cid=mm6107a1_e). ([link](https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6107a1.htm?s_cid=mm6107a1_e))

### Data Cleaning and Data Visualization

Originally, the dataset contains the number of cases of both Seasonal Influenza and the H1N1 Virus, along with the percentages of the Vaccination rate for both viruses in 29 states.

Data was cleaned and a total of 9 new columns was formed from the original dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| State | Seasonal n | Seasonal % | Longitude | Latitude | 2009 H1N1 n | 2009 H1N1 % | Longitude.1 | Latitude.1 |

Tableau was used primarily to visualize the Dataset: ([link](https://public.tableau.com/views/Viroshield/DashboardSeasonaland2009H1N1?:language=en-US&:display_count=n&:origin=viz_share_link ))

### Training the Data

InfectiShield uses the power of the Random Forest Classifier, which was chosen for its remarkable predictive accuracy and its inherent capability to adeptly manage geographical data. The code snippet below shows the essential steps in training the Random Forest Classifier using the dataset.

First, I encoded the risk labels and define the features (longitude and latitude) and the target variable (risk level). I employ K-Fold Cross Validation to ensure robust model evaluation. The Random Forest Classifier is instantiated with 1000 decision trees. After fitting the model, average cross-validation accuracy was calculated and displayed.

The Random Forest model is an apt choice due to its robustness in handling geographical data, which is pivotal for flu risk prediction. Its ensemble of decision trees allows for accurate risk assessment based on geographical information, making it a valuable asset to InfectiShield's functionality.

# Extracting Features for the Model

le = LabelEncoder()

le.fit(df['Risk'])

X = df[['Longitude', 'Latitude']]

y = df['Risk']

# K-Fold Cross Validation and RF Model

kf = KFold(n\_splits=10, shuffle=True, random\_state=42)

rf = RandomForestClassifier(n\_estimators=1000)

scores = []

for train\_index, test\_index in kf.split(X):

X\_train, X\_test = X.iloc[train\_index], X.iloc[test\_index]

y\_train, y\_test = y.iloc[train\_index], y.iloc[test\_index]

# Fit the model

rf.fit(X\_train, y\_train)

score = rf.score(X\_test, y\_test)

scores.append(score)

Y\_pred = rf.predict(X\_test)

rf.fit(X, y)

print("Average CV Accuracy: {:.3f}".format(sum(scores)/len(scores)))

### Assigning Weights to Symptoms

In the quest to provide a comprehensive risk assessment, InfectiShield considers the presence or absence of various flu symptoms. Assigning weights to these symptoms is crucial to evaluate their significance in the risk prediction process. So after the prediction from the model, subjective weights were assigned to 12 different flu symptoms. (see Code Snippet below).

# Symptom weights

symptom\_weights = {

'Fever': 0.12, 'Cough': 0.12, 'Sore throat': 0.11, 'Body aches': 0.11,

'Headache': 0.08, 'Fatigue': 0.11, 'Vomiting': 0.04, 'Diarrhea': 0.05,

'Runny nose': 0.11, 'Eye pain': 0.01, 'Sneezing': 0.03, 'Dizziness': 0.01, 'Chills': 0.1

}

print("Select your symptoms:")

patient\_symptoms = []

for i in possible\_symptoms:

response = input(f"{i} (y/n)? ")

if response.lower() == 'y':

patient\_symptoms.append(i)

# Calculate symptom score

symptom\_score = 0

for i in patient\_symptoms:

if i in symptom\_weights:

symptom\_score += symptom\_weights[i]

# Print the prediction based on the symptom risk level

if symptom\_score >= 0.7:

print("Prediction: High Risk")

elif symptom\_score >= 0.4:

print("Prediction: Moderate Risk")

else:

print("Prediction: Low Risk")

### Final Risk Score Prediction

A risk matrix maps these two risk dimensions into a comprehensive risk level. See code snippet below showcasing the integration of these risk factors and efficiently combines the risk levels derived from the model and symptom-based analysis into a single, conclusive risk prediction. The risk matrix provides a structured approach to determine the overall risk level, ensuring that users receive a holistic assessment of their flu risk.

# Create final risk prediction with a risk matrix

risk\_matrix = {

(1, 1): 'Low',

(1, 2): 'Low',

(1, 3): 'Moderate',

(2, 1): 'Low',

(2, 2): 'Moderate',

(2, 3): 'High',

(3, 1): 'Moderate',

(3, 2): 'High',

(3, 3): 'High'

}

# Combine risks

combined\_risk = risk\_matrix[(ml\_risk, symptom\_risk)]

print(f"Final Risk Prediction: {combined\_risk}")

## 

## Limitation and Constraints of iNFECTISHIELD

While InfectiShield offers a robust tool for flu risk prediction, it is essential to acknowledge its limitations and constraints. This section outlines some of these considerations:

Data Quality: The accuracy of risk predictions heavily relies on the quality and reliability of the input data: geographical coordinates. Inaccurate or incomplete data may affect the precision of risk assessments. In adition, the data that was trained was data from 29 out of 50 states, iNFECTISHIELD’s effectiveness is contingent on the availability of geographical data. It may not provide accurate risk predictions in areas with limited data coverage

Symptom Subjectivity: Symptom weighting is based on subjective judgments and research findings. Individual experiences and interpretations of symptoms may vary, potentially leading to variations in risk predictions.

External Factors: iNFECTISHIELD primarily considers geographical and symptom-related factors. It did not account for other external variables that can influence flu risk, such as local healthcare infrastructure.

## Conclusion

InfectiShield represents a significant stride in leveraging machine learning and geographical data to provide users with insightful flu risk predictions. Using a combination of geographical information and symptom-based assessments, iNFECTISHIELD delivers a comprehensive evaluation of flu risk levels. However, it is vital to recognize its limitations and the potential impact of data quality and user compliance. InfectiShield stands as a valuable tool for individuals seeking to gauge their flu risk, offering a proactive approach to flu prevention and healthcare decision-making.