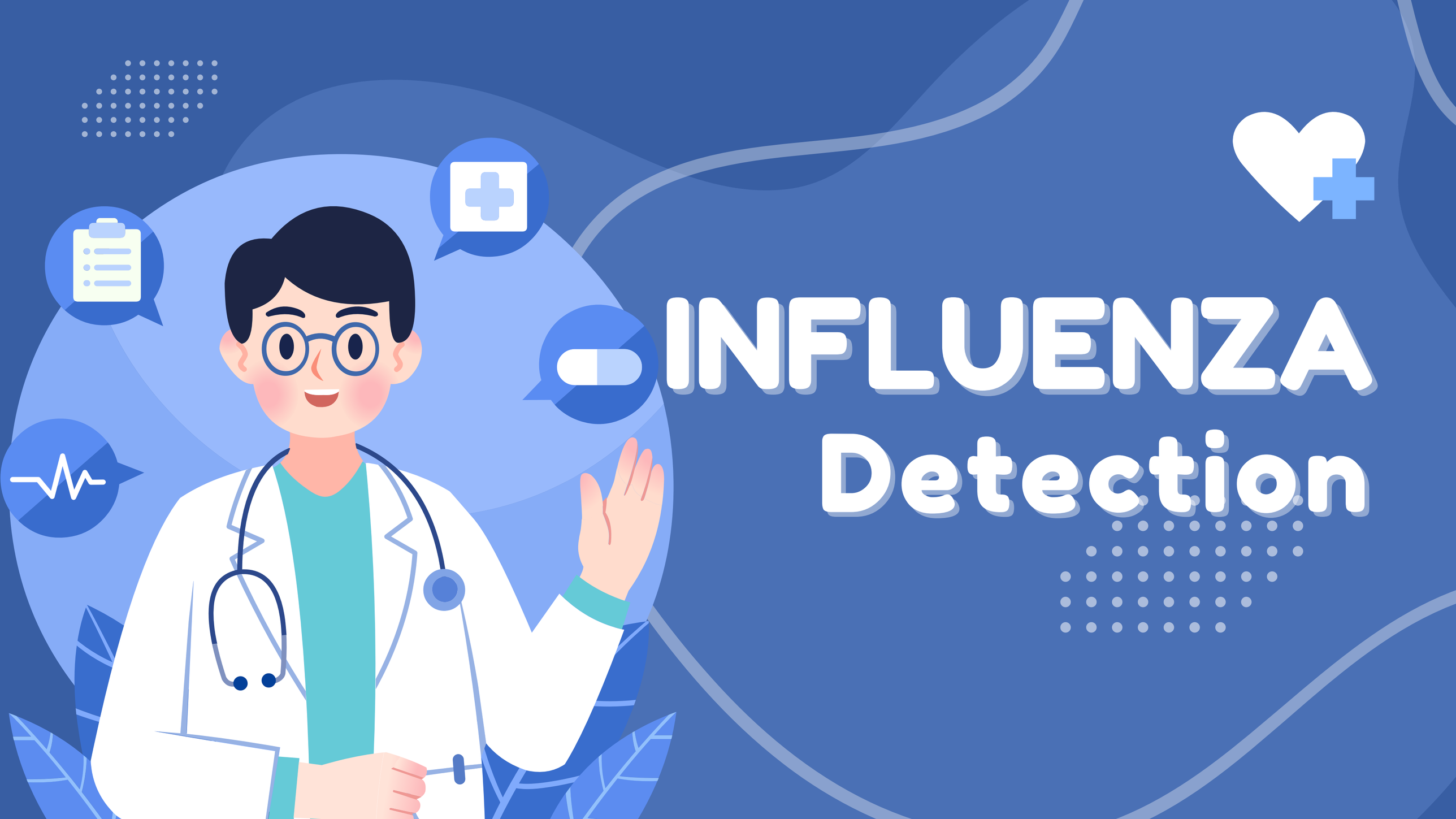

INFLUENZA
Detection

Problem

Health Organizations struggle with influenza detection.

Theie issue is understanding when to expect an influenza outbreak.

Countries are indecisive on how to accurately detect influenza outbreaks.

Goal

Analyze the data collected by the WHO from several countries.

Prepare a machine learning model that can detect influenza outbreaks.

Provide insights to health centers on when to dedicate resources and prevention strategies.

Our Data

The Dataset contained 10,835 values and 16 features.

Main variables:

- Country
- Surveillance site type
- Year-week/Week start date
- Specimen tested
- Influenza positive/negative
- Subtypes

Source: World Health Organization

Preprocessing steps

Null Values

Remove missing values in the dataset to remove noise.

Feature Selection

Capture vital features into a defined variables for efficient modeling.

Encoded & Scaled

Convert categorical data into numerical form and equalizes numerical features

Train/Test

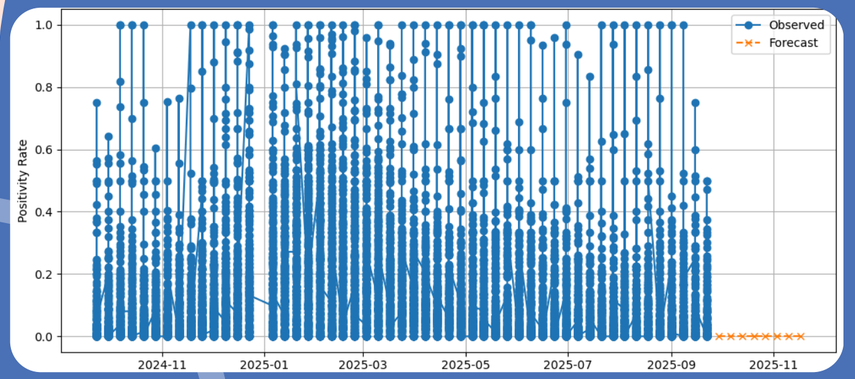
Split the data into training, testing, and validation sets

Our Models

Severity Forecasting Model

R^2 = 0.97

RMSE = 0.02

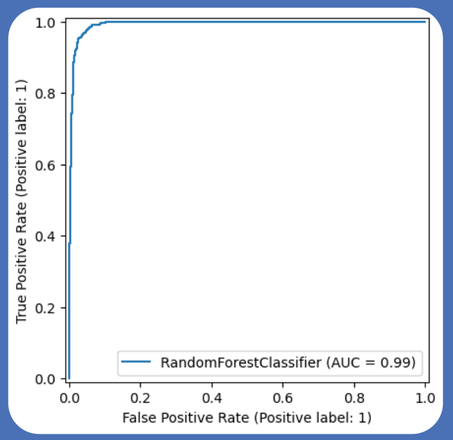


Outbreak Classification Model

Accuracy = 0.96

Precision = 0.93

Recall = 0.98



Our Findings

Forecasting Model

Estimate the future influenza positivity rate over upcoming weeks.

Classification Model

Determine whether an outbreak is likely

So What?

Predict

Detect patterns of growing influenza positivity in real time

Plan

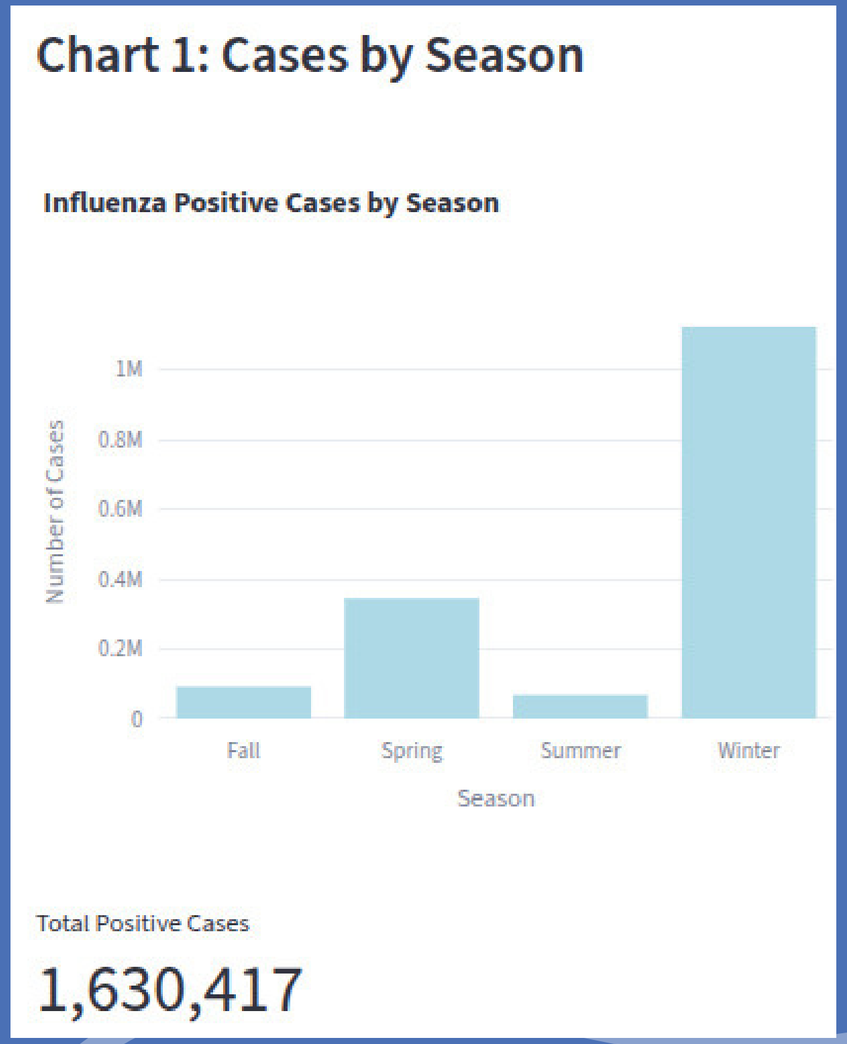
Prepare prevention strategies such as alerts, vaccines, hospital staffing...

Monitor

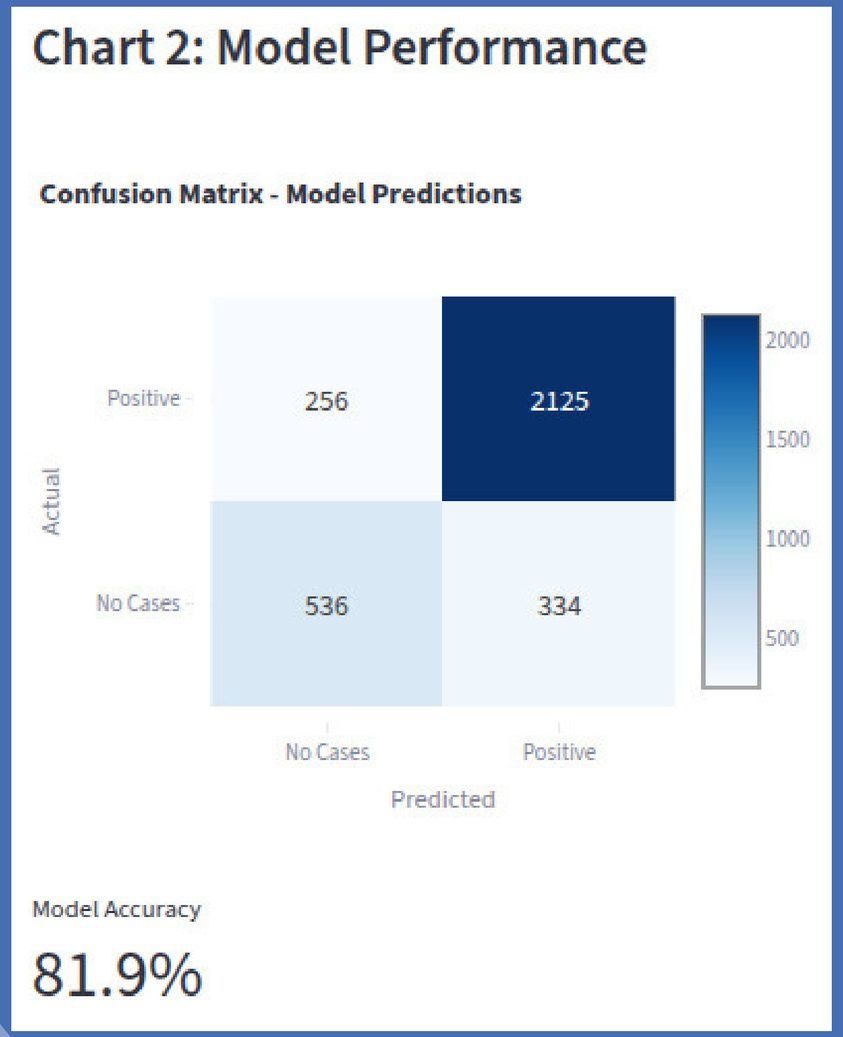
Evaluate the specimens while testing effective solutions

Dashboard

This bar chart displays the total number of confirmed influenza cases aggregated by season (Winter, Spring, Summer, Fall). The visualization reveals seasonal patterns in influenza transmission, with Winter typically showing the highest case count due to favorable conditions for viral spread. This analysis helps identify peak influenza activity periods and supports resource allocation and public health intervention planning.

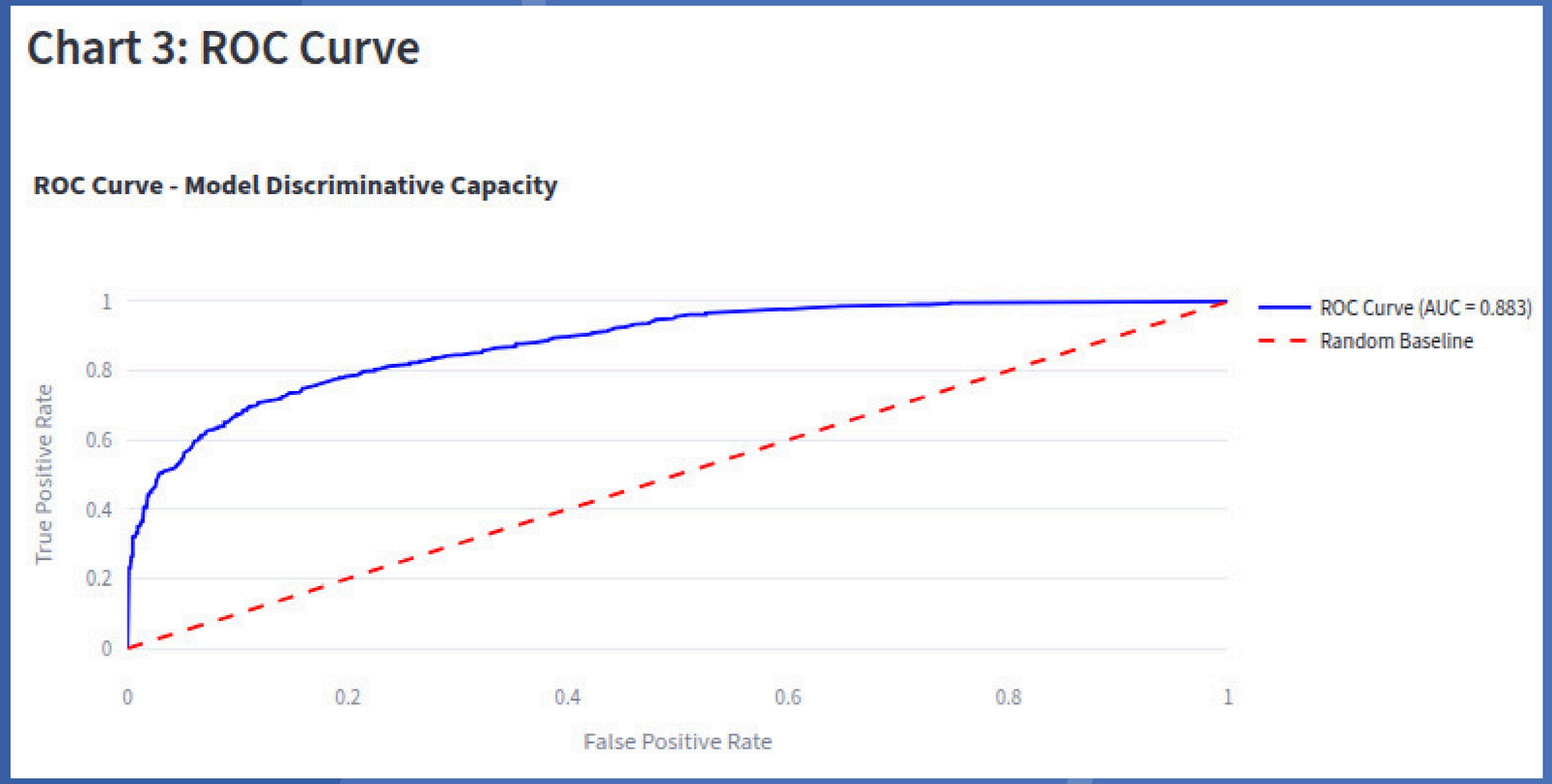


Dashboard



The heatmap visualization displays the confusion matrix of the Random Forest classification model. The matrix compares actual influenza cases (rows) against predicted cases (columns), showing four key outcomes: True Negatives (correctly identified non-cases), False Positives (false alarms), False Negatives (missed outbreaks), and True Positives (correctly detected cases). Each cell displays the count of observations, with color intensity indicating magnitude. This visualization is essential for understanding model accuracy and identifying systematic prediction errors

Dashboard



The Receiver Operating Characteristic (ROC) curve evaluates the model's ability to discriminate between positive and negative influenza cases across all classification thresholds. The blue line represents the model's performance, plotting True Positive Rate (sensitivity) against False Positive Rate (1-specificity). The red dashed diagonal line represents random chance performance. The Area Under the Curve (AUC) metric quantifies overall model discrimination, with values closer to 1.0 indicating excellent performance. This curve is fundamental for assessing predictive model quality in epidemiological surveillance systems.

THANK YOU