Forecasting and Monitoring Hospitalization Trends from Respiratory Viruses in Kenya: A Time Series Analysis by County

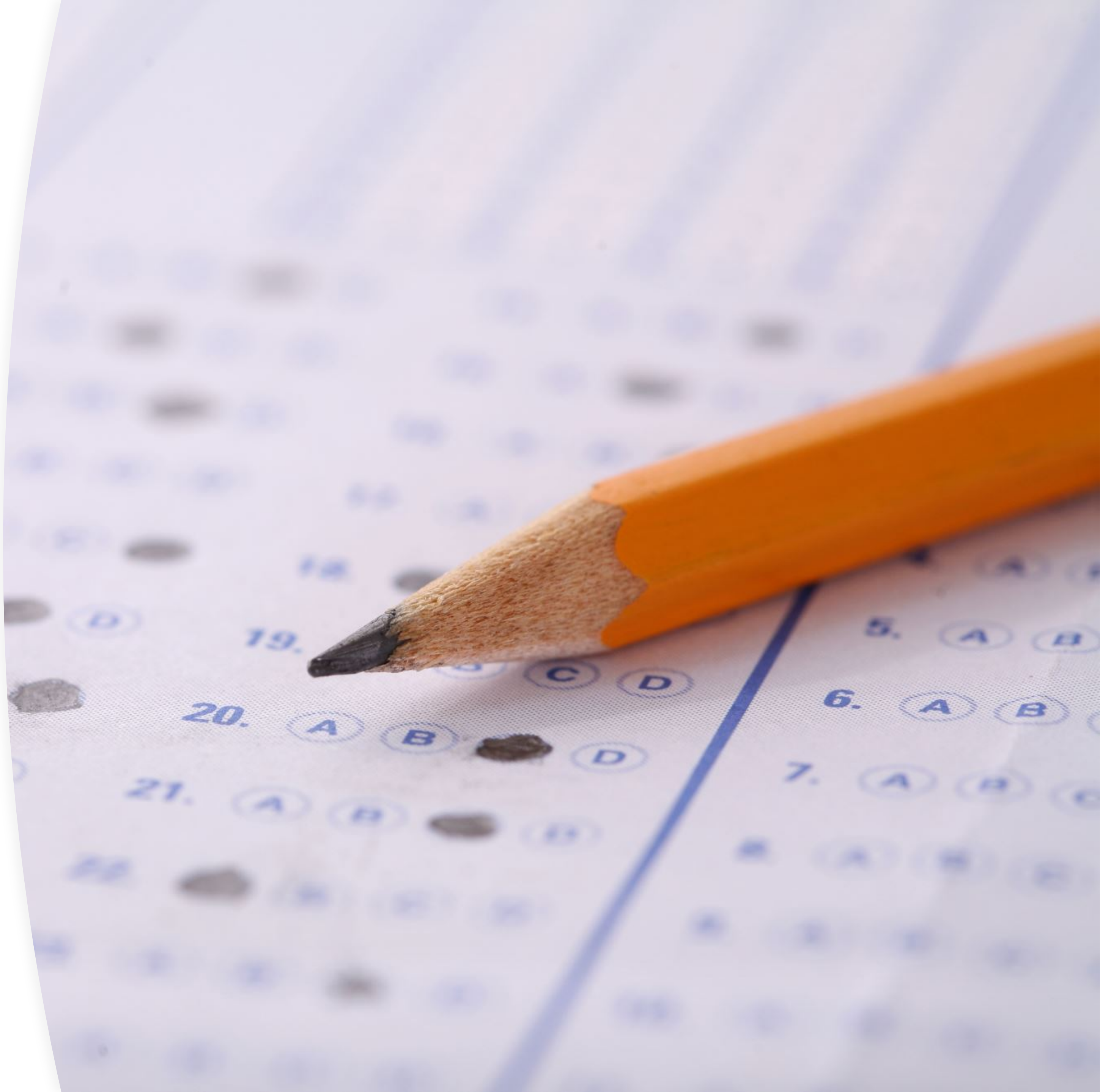
Group 10

Business Understanding

The project aims to address Kenya’s lack of real-time health surveillance for respiratory viruses by demonstrating a data-driven forecasting approach using U.S. hospitalization data. By building predictive models, the project showcases how Kenya could anticipate surges, allocate resources effectively, and improve its public health response once a local data infrastructure is established.

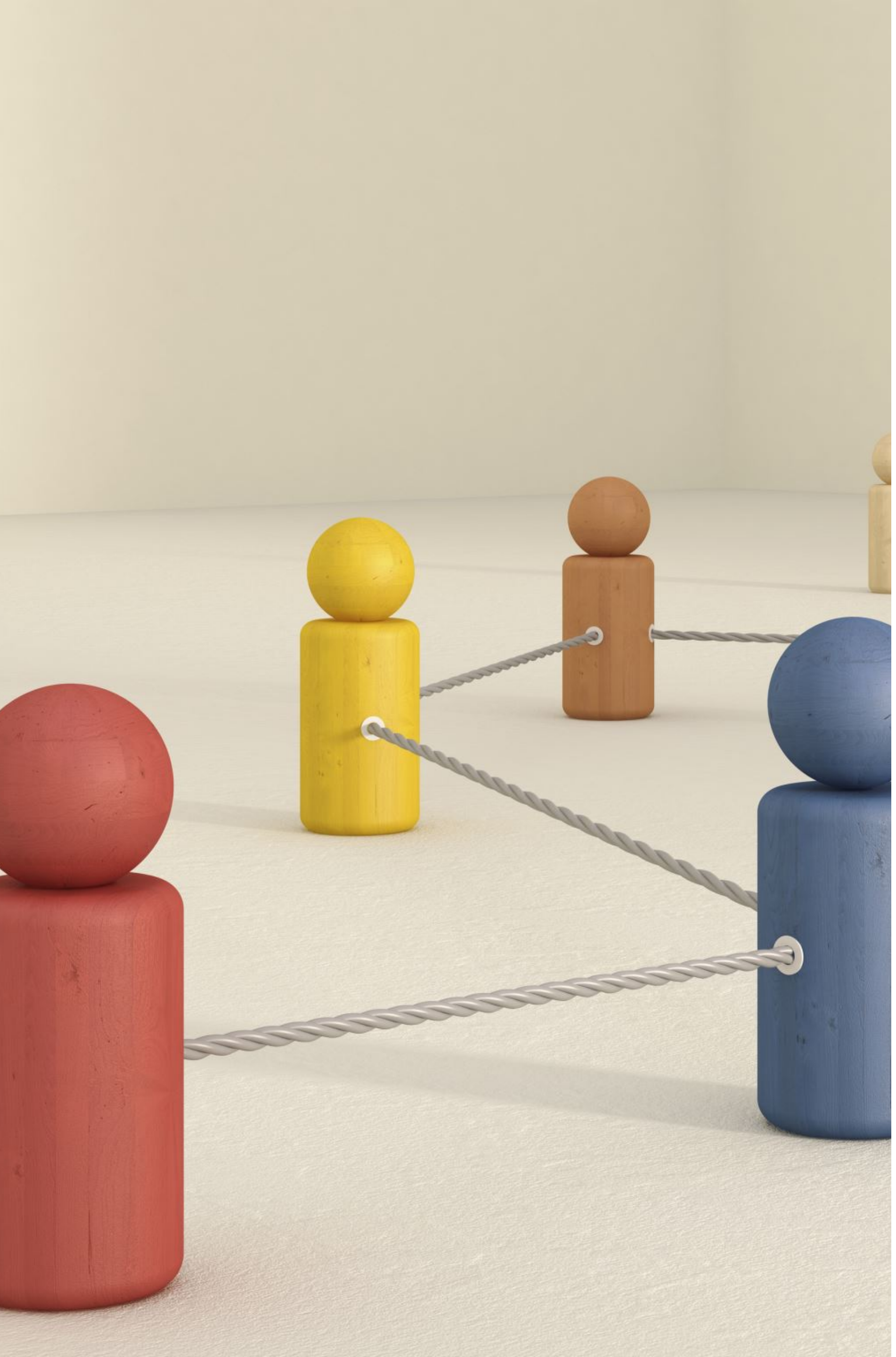
Problem Statement

Kenya currently lacks a real-time health surveillance system to track hospitalizations from respiratory viruses, making it hard to respond to surges. This project uses U.S. hospitalization data as a proxy to demonstrate how data-driven forecasting could work in Kenya. The goal is to guide future public health actions, resource planning, and preparedness once Kenya develops its own data infrastructure.



Objectives

The project aims to monitor, model, and forecast respiratory virus-related hospitalizations across U.S. regions to aid in healthcare planning, outbreak response, and policy-making. Key objectives include detecting seasonal patterns, forecasting future trends, assessing the impact of interventions like vaccines and NPIs, and offering real-time dashboards for public health decision-makers.



Stakeholders

1. CDC and local health departments
2. Hospital administrators and planners
3. Epidemiologists and public health researchers
4. Policy-makers and emergency response teams

Data Understanding

- This project leverages two core datasets derived from publicly available U.S. health records (e.g., CDC and HHS):
- **Weekly COVID-19 hospital admissions** – used as the primary target variable.
- **State-level socioeconomic and demographic indicators** – providing contextual predictors.
- The hospitalization dataset is a time-series collection containing detailed records of average COVID-related admissions. It includes engineered features such as **lag values**, **rolling statistics**, and **aggregated state-level metrics**.

Data Understanding

Key insights from initial data exploration revealed:

- Clear **seasonal trends** and **admission spikes** during outbreak periods.
- **Regional variations** in admission patterns across states.
- Valuable **time-series decompositions** into trend, seasonal, and residual components.
- Rich **statistical summaries** that guided model selection and feature engineering strategies.

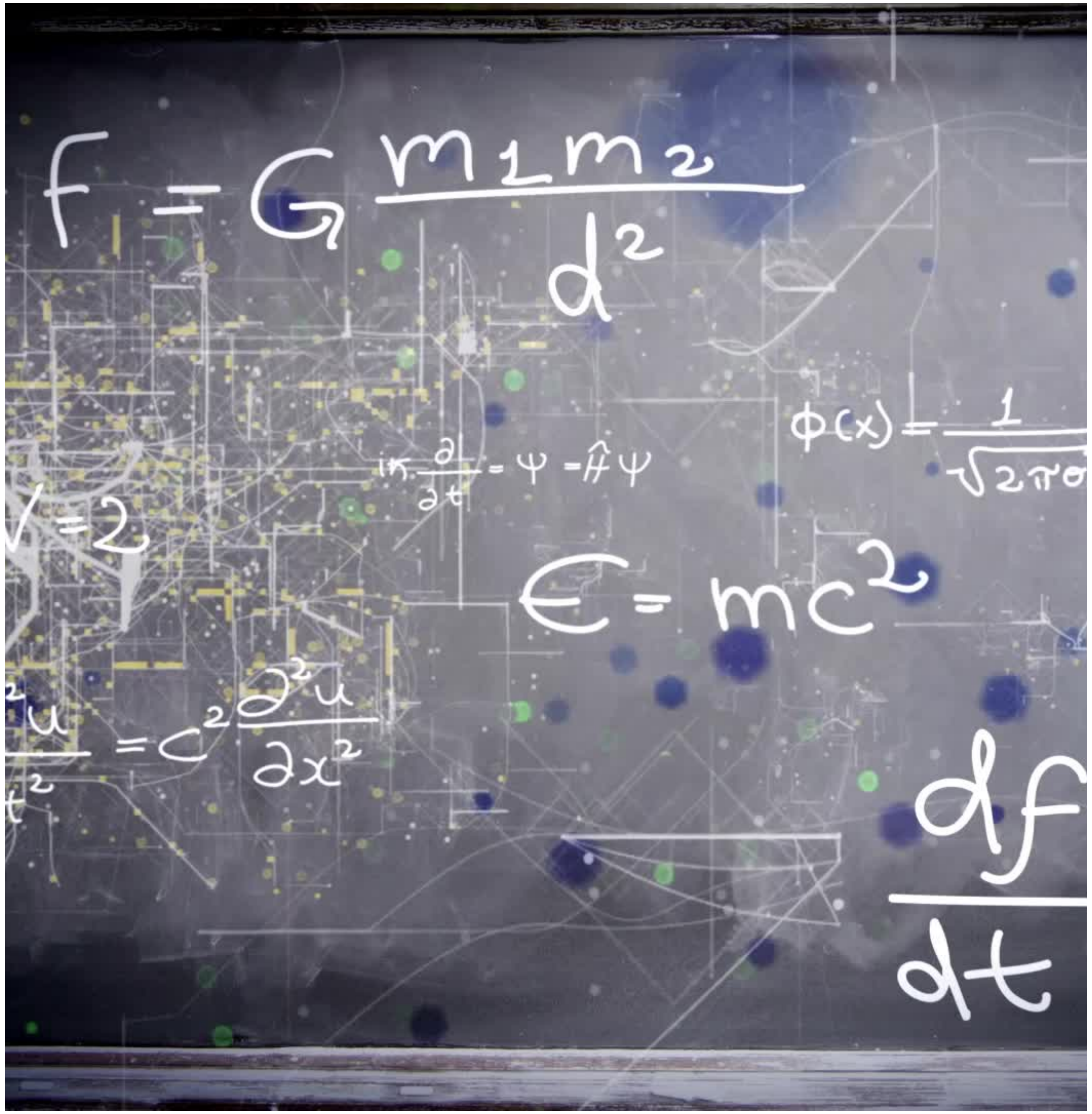
Data Preparation

Comprehensive preprocessing included:

- Log transformation of skewed features
- Lag features (1–8 weeks back)
- Rolling statistics (mean, std, min, max for 2, 4, 8 weeks)
- Percentage changes to capture weekly trends
- Time encoding (week, month, quarter, cyclical features)
- Merge of temporal features with state-level aggregates
- Handling missing values and scaling for classification models

Modeling

Model	Type	Purpose
XGBoost Regressor	Machine Learning	Predict average hospital admissions (regression)
Random Forest Regressor	Machine Learning	Benchmark regression with interpretability
Prophet	Time Series Forecasting	Long-term trend and seasonality modeling with regressors
Logistic Regression	Classification	Predict “High” vs “Not High” admission weeks
SARIMA	Time Series Forecasting	Classical time series with autoregression + seasonality
Exponential Smoothing	Time Series Forecasting	Smoothing of trends and seasonality
Simple Moving Average	Baseline Model	Trend smoothing using fixed window sizes (SMA-3, SMA-6…)
LSTM (Keras)	Deep Learning	Sequence modeling using weekly features



Evaluation Metrics

01

Regression:
RMSE, MAE,
MAPE, R^2

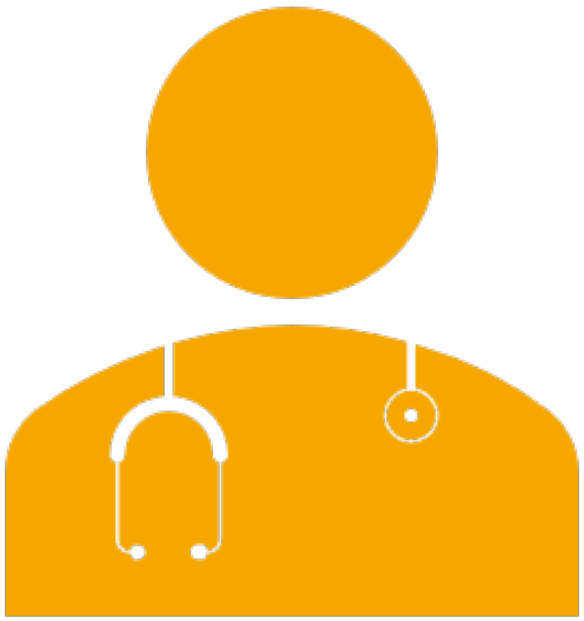
02

Classification:
Accuracy,
Confusion Matrix

03

Forecasting:
Visuals for actual vs predicted trends, prediction intervals

Conclusion & Recommendations



Conclusion

- Combining ML models (XGBoost, RF) with time series models (Prophet, SARIMA) enhanced robustness.
- Logistic Regression gave useful binary signals for critical weeks.
- LSTM holds promise but needs further tuning/data.
- Exponential smoothing and SMA were useful for interpretability and baseline comparisons.

Recommendations

- Kenya should prioritize developing a centralized, real-time health data collection system.
- Forecasting models should include lag, seasonal, and contextual features.
- Hybrid systems (ML + classical time series) improve both accuracy and explainability.
- Dashboards with alerts based on these models can significantly support the Ministry of Health and county health departments.

THANK YOU!