Predicting Antibiotic Resistance in Salmonella enteritidis

●●●

Rebecca Garlena

Project Introduction and Goal

- Antibiotic resistance poses a significant public health threat, complicating the treatment of bacterial infections

- Salmonella enteritidis is a major cause of foodborne illness globally, exhibiting growing resistance to antibiotics

- This project aims to develop a predictive model to analyze and forecast antibiotic resistance trends in Salmonella enteritidis

Salmonella

Dataset Overview

- The data used in this study is sourced from the CDC National Antimicrobial Resistance Monitoring System (NARMS)

- Key features in the dataset include:

  - Specimen ID: Unique identifier for each isolate

  - Antibiotic Conclusion Columns: Indicating susceptibility (S), intermediate (I), or resistance (R)

  - Year and Region: Collection year and geographic region.

 - Age Group: Demographic information of the individual

Data Wrangling

Data cleaning and preparation involved:

- Examination of duplicates to ensure unique isolates

- Removal of irrelevant columns and those with excessive missing data

- Imputation of missing values using the most common entries

- Filtering the dataset to focus on data from 2003 to 2023

- The AR Level for each specimen was calculated by summing all counts of resistant (R) values across the antibiotic conclusion columns.

EDA: Sample Collection Trends

- Variations in sample collection by year and region suggest shifts in infection rates or surveillance efforts

- The dominance of stool specimens reflects clinical practices for testing gastrointestinal pathogens

- Higher incidence of infections in younger populations highlights the need for targeted public health interventions
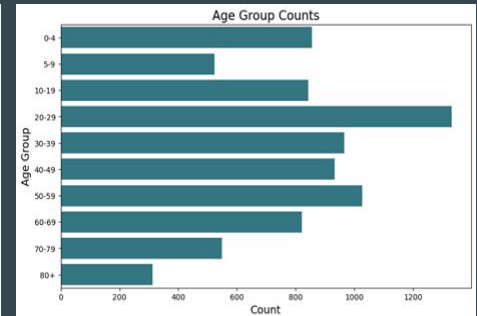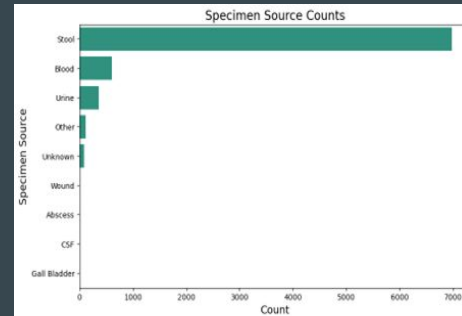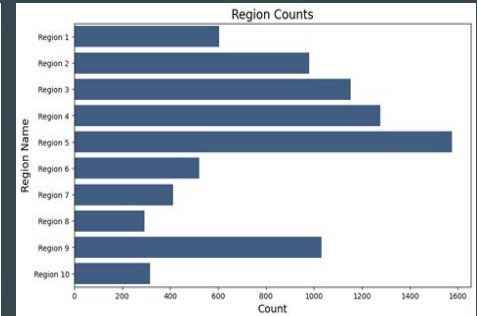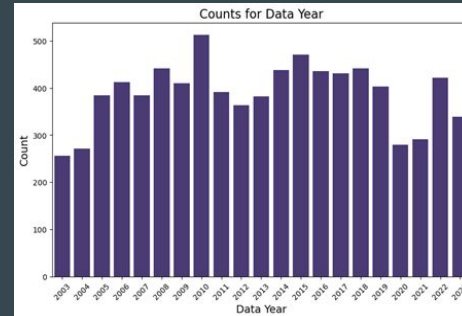
EDA: Antibiotic Susceptibility and Resistance

- Prevalence of susceptible strains underscores antibiotic effectiveness

- Emerging resistance noted for Ampicillin (AMP), Fosfomycin (FIS), Nalidixic Acid (NAL), Streptomycin (STR), and Tetracycline (TET)

- Intermediate resistance levels observed for Ciprofloxacin (CIP)
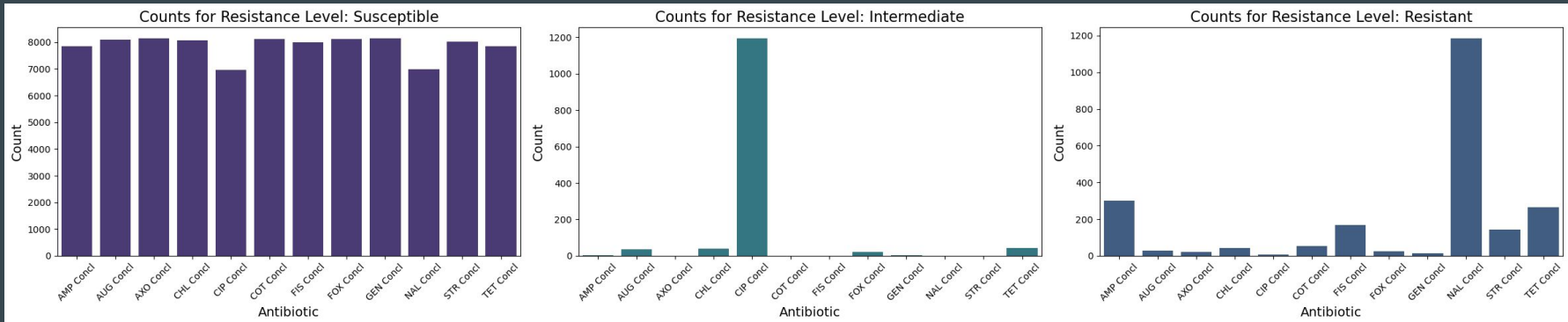
EDA: Susceptible vs. Resistant Counts Over Years

- The counts of susceptible isolates (AR Level 0) consistently exceed those of resistant isolates (AR Level 1+)

- However, a slight upward trend in resistant counts is observed from 2015 onward, raising concerns about increasing antibiotic resistance

- Concurrently, there is a downward trend in susceptible counts, indicating a potential decline in antibiotic effectiveness

Susceptible vs Resistant Counts Over Years
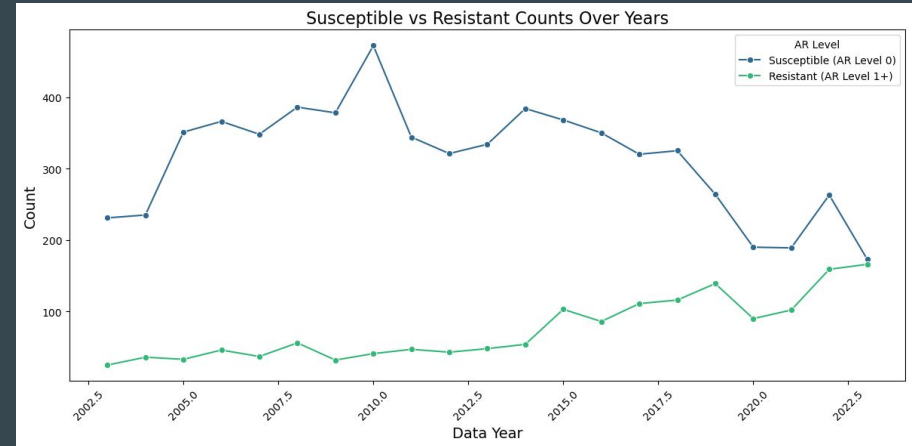
Modeling Approaches

The following modeling techniques were utilized:

- Random Forest Regressor: Trained to predict AR Levels using antibiotic conclusion features

- Exponential Smoothing: Employed to forecast future trends in antibiotic resistance

- Holt's Method: Similar to Exponential Smoothing, used for trend analysis

Random Forest Modeling Results

- Modeling results indicate strong predictive capabilities for the Random Forest model

- The Random Forest achieved a Mean Squared Error (MSE) of 0.0313 and an R-squared value of 0.9781

- Feature importance analysis revealed FIS Concl as a critical factor influencing predictions

Feature Importance (Antibiotic Conclusions)
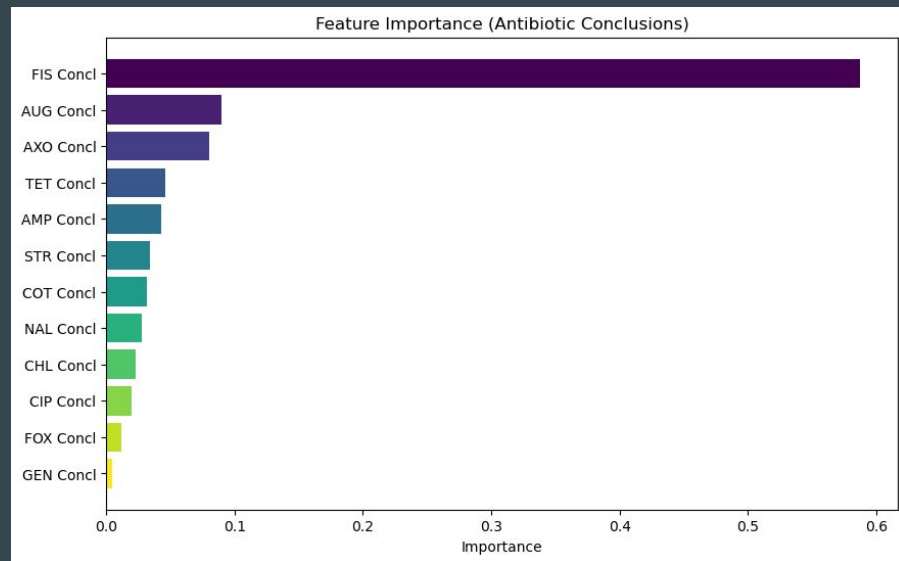
Exponential Smoothing Modeling Results

- The Exponential Smoothing model forecasts indicate an upward trend in antibiotic resistance levels

- This trend is projected to continue over the next five years, emphasizing the importance of ongoing monitoring and intervention efforts

AR Level Forecast for Next 5 Years (Exponential Smoothing)
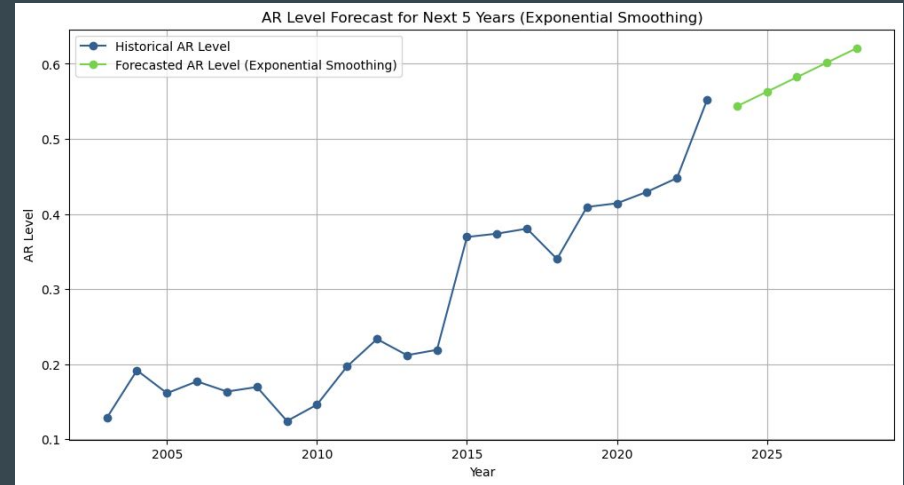
Model Metrics Overview

- The Random Forest model demonstrates strong performance with low MSE and MAE.

- Both Exponential Smoothing and Holt's models are effective for forecasting trends, but they have a higher MAPE

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Mean Absolute Percentage Error (MAPE)
Random Forest	0.0313	0.0384	0.1770	1.17%
Exponential Smoothing	0.0042	0.0538	0.0649	11.31%
Holt's Model	0.0042	0.0538	0.0649	11.31%

Conclusions

- Both models provide valuable insights into predicting antibiotic resistance levels

- The Random Forest model demonstrated strong predictive capabilities

- The Exponential Smoothing model provides valuable forecasts, indicating an upward trend in antibiotic resistance levels over the next five years

- This trend stresses the importance of continuous monitoring and proactive public health strategies to mitigate the rising threat of antibiotic resistance

Future Directions

Future research could focus on:

- Integrating additional data attributes, including geographical and demographic factors, as well as information on food outbreaks and salmonella-related hospitalizations.

- Continuous validation of models with new data to ensure adaptability to evolving resistance patterns

- Exploring advanced algorithms to enhance prediction accuracy and robustness

- Monitoring long-term trends beyond 2028 to inform and guide public health strategies effectively