Statistical Time Series Modelling for Predictive Epidemic Surveillance (STMES)

GROUP NO: 08

Submitted by-
1. Mrudula Deshmukh   (50605669)
2. Aishvarya Samuel Salvi   (50604254)
3. Siddhi Sunil Nalawade   (50613176)
4. Ragavi Muthukrishnan   (50604772)

Importance

The correct identification of infectious disease outbreaks remains a captious concern in public health. While current surveillance methods typically detect incidents after a large spread, complicated techniques to machine learning frequently build "black box" models that are not missing for decision makers. A method is required for both early detection and statistical interpretation.

Proposed Work

This work develops and assesses statistical time series modelling methods for earlier, more dependable outbreak identification which also focuses on Change-point ARIMA modelling.

1. Compare traditional ARIMA, Change-Point ARIMA, and Prophet models on historical outbreak data, measuring each method's predictive power and lead time advantages
2. Develop robust change-point detection algorithms to identify subtle statistical shifts that signal emerging outbreaks before they become obvious
3. Evaluate performance using sensitivity, specificity, and lead time metrics on diverse historical outbreaks
4. Test framework adaptability across multiple disease types and regional contexts
5. Create interpretable visualization tools that translate statistical insights into actionable information

Using these advanced time series methods, particularly Change-Point ARIMA, we intent to provide an approach that balances advanced statistical power with practical interpretability. This approach will assist public health professionals in detecting problematic patterns early while providing the statistical confidence required to justify immediate measures.

Link to dataset: https://www.kaggle.com/datasets/cdc/foodborne-diseases/

(Note: we are working in dataset as we are combining 2/3 different datasets.)

References

[1]Zhang, Z. (2022). Statistical Modeling of Daily Confirmed COVID-19 Cases and Deaths in the US and Europe.

[2]Gao, S., et al. (2025). Early detection of disease outbreaks and non-outbreaks using incidence data. PLoS Computational Biology, 21(2).

[3](2022). ARIMA MODEL IN PREDICTING OF COVID-19 EPIDEMIC FOR THE SADC.

[4](2023). Interrupted time series analysis using the ARIMA model of the incidence of notifiable communicable diseases. PMC10266306.

[5]Blonigan, P., Ray, J., & Safta, C. (2021). Forecasting Multi-Wave Epidemics Through Bayesian Inference.

[6](2024). A Comprehensive Statistical Analysis of COVID-19 Trends.

[7](2023). Accurately summarizing an outbreak using epidemiological models. PMC10523082.

[8]Sharma, A., et al. (2024). Predicting Disease Outbreaks Using Time Series Analysis and Classification