The 14th International Conference on Computing, Communication and Networking Technologies (ICCCNT)

IIT DELHI, July 6th - 8th, 2023

Exploring the Factors behind COVID-19 Surge: Predictive Modeling and Analysis

Presented By: Tushar Sharma

OUTLINE

- Introduction
- Goals of Research
- Dataset
- Methodology
- Results
- Discussions / Future Work
- Conclusion
- References

INTRODUCTION

- The global outbreak of SARS-CoV-2 has strained healthcare systems and caused significant disruptions worldwide.
- COVID-19 is primarily transmitted through respiratory droplets and close contact with infected individuals.
- The high transmissibility of COVID-19 has led to community spread, causing severe illness, high mortality, and disruptions to societal and economic activities.
- Sufficient healthcare system capacity is crucial for managing and providing adequate care to COVID-19 patients.
- Delhi NCR is one of the most affected regions in India, with densely populated areas contributing to community transmission and increased infection risk.

INTRODUCTION (Continued)

- The research contributes to understanding the factors behind the surge in COVID-19 cases in the Delhi NCR region.
- Predictive modeling techniques can help understand the relationships between parameters, the number of cases, and the spreading patterns of COVID-19.
- Exploratory data analysis techniques are employed to gain insights and identify potential relationships within the dataset.
- The findings can help government officials and medical experts in creating practical plans to control the spread of the virus and lessen its impact on people's health.

GOALS OF RESEARCH

- Understand the relationships between various parameters and the number of COVID-19 cases.
- Analyze the spreading patterns of COVID-19 and identify key factors driving the surge in cases.
- To develop accurate predictive models for forecasting COVID-19 cases in the region.
- Support public health officials, policymakers, and medical experts in creating practical plans to limit the spread of the virus and mitigate its impact on the population.
- To contribute to the existing knowledge and understanding of the complex dynamics underlying the COVID-19 pandemic.

DATASET

- The dataset used in this research spans from January 2020 to December 2022.
- 12 Parameters Included: Temperature, Humidity, Air quality index, Senior citizen population, Gender diversity, Labour/hazardous work percentage, Prior Medical and Surgical history, Cardiovascular and Gastrointestinal diseases, Smoking and alcohol habits, and the Athletic individual's percentage.
- To ensure the accuracy and reliability of the data, information was collected from various government official sources and recognized international organizations.
- Questionnaire-like approach was utilized to gather specific information for parameters not available from official sources, enhancing the dataset's comprehensiveness.

METHODOLOGY

- The dataset was preprocessed to handle outliers, missing data, and other irregularities and employing data augmentation techniques to ensure accurate and reliable predictions.

- Exploratory Data Analysis (EDA) was performed to discover key elements and trends in the dataset, such as using plots and visualizations, scale of data, seasonal decomposition, Standardization, and Principal Component Analysis.

- The testing data was carefully selected over a period of 120 days to identify time-related trends and patterns, resulting in a reliable and applicable evaluation methodology.

- Multiple machine learning techniques such as time series analysis, decision trees, and regression models have been evaluated for their performance and suitability to accurately predict and analyze the factors driving COVID-19 cases.

METHODOLOGY (Continued)

- The Light Gradient Boosted Machine (LightGBM) model was selected as the optimal option due to its advanced boosting techniques, high performance in handling large datasets, and powerful predictive capabilities.
- LightGBM model parameters were fine-tuned, including boosting_type, num_leaves, max_depth, learning_rate, n_estimators, and random_state.
- The LightGBM model is being used to accurately predict COVID-19 cases and gain useful insights into potential contributing factors.

RESULTS

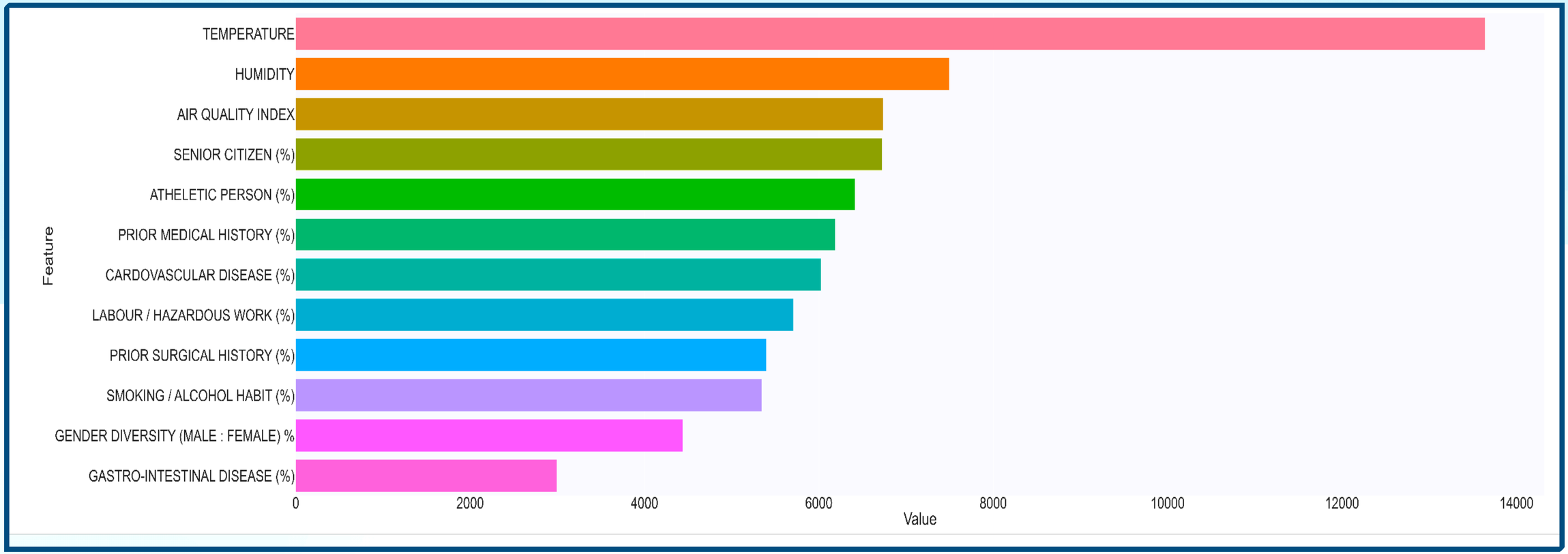
- The analysis reveals significant factors contributing to the surge in COVID-19 cases in the Delhi-NCR region. Temperature shows a substantial effect, Previous medical history, climate variables, and labor/hazardous work percentage are also identified as important determinants.

- Demographic analysis highlights the higher susceptibility of elderly individuals to the virus.

- Use of Tobacco, Smoking, Alcohol consumptions, and a lack of physical activity directly affect the spread of COVID-19.

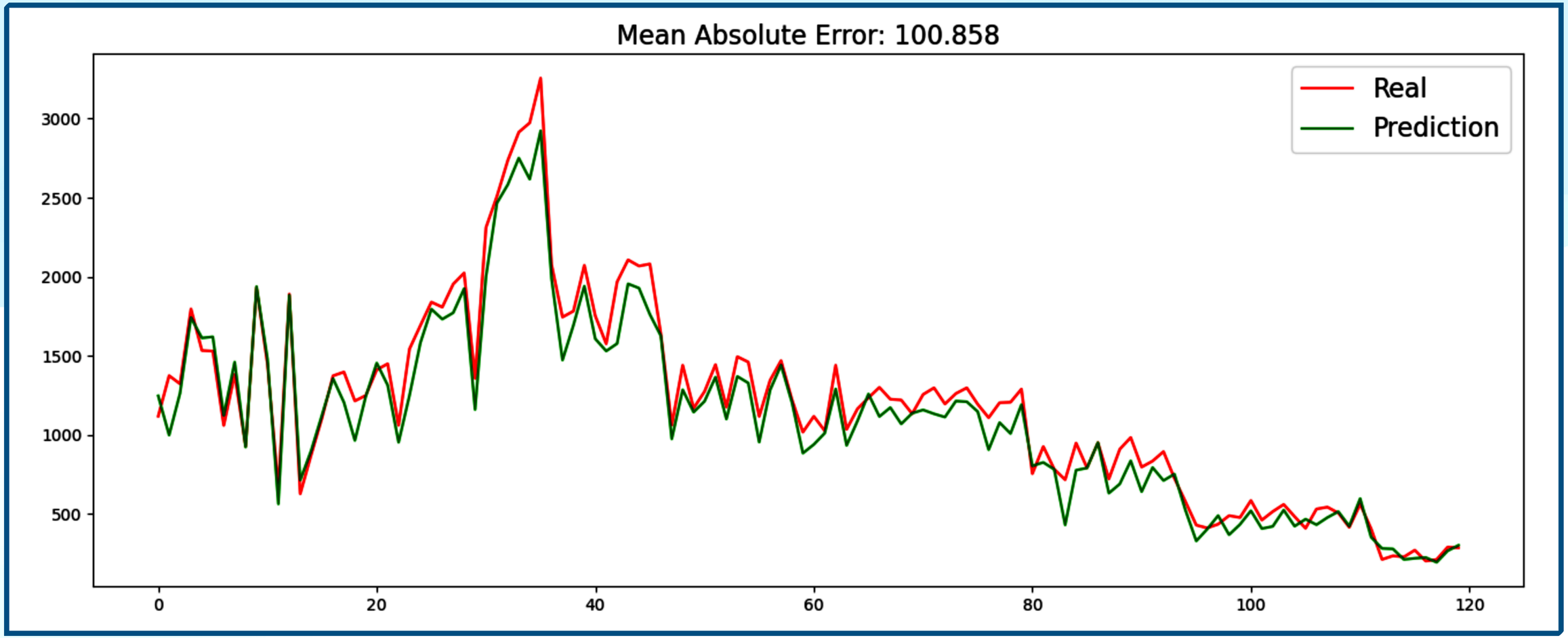
- The LightGBM model achieved an accuracy of 95.4%, with Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-Squared (R2) values, when applied to a testing dataset for predicting the number of cases.

RESULTS



LightGBM Model Features Importance

RESULTS



Real v/s Prediction

DISCUSSIONS / FUTURE WORK

- Integration of additional factors like vaccine rates, mobility data, and socioeconomic indicators for a more comprehensive understanding of the pandemic's dynamics.
- Evaluating the effectiveness of intervention strategies such as lockdown measures, mask requirements, and vaccine programs.
- Future work could involve expanding the analysis to other regions within India or globally to understand the impact of COVID-19 transmission patterns.
- Extending the prediction horizon to facilitate long-term planning can provide insights of the pandemic and facilitate resource allocation.

CONCLUSION

- This research paper successfully investigated the factors contributing to the surge in COVID-19 cases through predictive modeling and analysis.
- The study utilized a diverse set of variables and the Light Gradient Boosted Machine (LightGBM) model to accurately predict COVID-19 cases and capture complex relationships.
- The findings contribute to evidence-based decision-making, guiding targeted treatments, public health interventions, and resource allocation strategies.
- Future works can further enhance our understanding of the pandemic and facilitate proactive measures for control and prevention.
- By analyzing and exploring the factors influencing transmission, we can protect public health and mitigate the impact of the pandemic effectively.

REFERENCES

- World Health Organization. (2021). Coronavirus disease (COVID-19) pandemic. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019
- Government of India, Ministry of Health and Family Welfare. (2021). COVID-19 India. Retrieved from https://www.mohfw.gov.in/
- Centers for Disease Control and Prevention. (2021). COVID-19 Data Tracker. Retrieved from https://covid.cdc.gov/covid-data-tracker/
- LightGBM Documentation. (n.d.). Retrieved from https://lightgbm.readthedocs.io/
- Open Government Data (OGD) Platform India. Retrieved from https://data.gov.in/

THANK YOU