Name: Sirisha Gathpa

Campus ID: KG86292

Project Proposal: Deep Dive into COVID-19's Impact on Hospital Systems

Title:
Deep Dive: COVID-19's Pervasive Impact on Hospital Staffing and Patient Admissions Across States

Introduction:

The project begins with an introduction to COVID-19, its global impact, and its challenges to healthcare systems. It emphasizes the need for effective data analysis to understand and mitigate these challenges. The introduction sets the stage for the following detailed analysis, focusing on two key aspects: predicting staffing shortages in healthcare facilities and forecasting hospitalization rates. The introduction frames the project as a vital step towards strategic healthcare planning in response to the pandemic.

Business Problem:

How has the COVID-19 pandemic reshaped the dynamics of hospital staffing and patient admissions, and what insights can I extract to strengthen healthcare infrastructure in preparation for future waves or pandemics?

Data Sources:
Our analysis will lean on a comprehensive dataset that includes hospital-specific data from various states. The dataset highlights:

- Staffing shortages
- Inpatient bed utilization
- Patient admissions (divided by age)
- ICU occupancy rates
- Availability of therapeutic supplies ...among other pertinent metrics.

Modeling Approach:

- **Classification:** This will be used to group hospitals based on the intensity of staffing shortages and to identify primary factors leading to these shortages.
- **Regression:** A regression approach will be employed to anticipate future patterns in hospital admissions, ICU demands, and bed occupancies, thus offering predictions on potential healthcare system strains.

Project Approach:

Our approach aligns with the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. This cyclical method initiates with:

- Business understanding
- Data Exploration
- Data Preparation
- Modeling
- Evaluation
- Deployment

The study used extensive COVID-19 data, including 71,129 entries across 135 columns. The data was analyzed using both classification and regression models. The classification models aimed to predict critical staffing shortages, while the regression models focused on forecasting COVID-19 hospitalization rates.

Significant Assumptions/Constraints:

- **Data Integrity:** I operate under the presumption that our dataset is both current and reflects authentic conditions.
- **Influence of COVID-19:** It is assumed that, within the timeline of our dataset, hospital operations and challenges were mainly a consequence of the COVID-19 pandemic, with minimal impact from other external factors.
- **Projection Limitations:** The trends extracted from this data are poised for short-term forecasts. For extended projections, recalibration may be necessary.

Data Preprocessing:

The data, consisting of 71,129 entries across 135 columns, was cleaned and preprocessed. This involved handling missing values, normalizing data, and selecting relevant model features. Effective preprocessing is vital for ensuring the accuracy of the model's predictions.

Classifiers Used:

The classification models applied various algorithms, though the exact types aren't specified in the provided materials. Generally, such models could include logistic regression, decision trees, or random forests, each with unique strengths in handling binary or multi-class classification problems.

Regression Models:

The regression analysis likely employed models such as linear regression, ridge regression, or support vector regression. These models excel in predicting continuous outcomes, like hospitalization rates, by finding relationships between independent variables and a continuous dependent variable.

Findings:

Critical Staffing Shortages Classification Model:

The classification model focused on predicting critical staffing shortages in healthcare facilities. Although the specific algorithms used aren't detailed, such models typically employ logistic regression or decision trees. The model achieved an accuracy of 58.5% with a macro-averaged F1-score of 0.100,

indicating challenges in accurately predicting staffing shortages due to the complexity of the factors involved.

COVID-19 Hospitalization (Random Forest) Regression Model:

The regression model, likely using a Random Forest algorithm, aimed to predict hospitalization rates. It demonstrated high precision (92.5%) and recall (93.3%), with an F1-score of approximately 92.9%, indicating strong predictive capabilities.

Correlation Matrix Heatmap:

The correlation matrix heatmap provided insights into the relationships between different variables. This visualization helped in understanding which factors were most strongly associated with hospitalization rates and staffing shortages.

Data Visualizations and Scatter Plot:

Various data visualizations, including scatter plots, were used to explore the data further. These visualizations aided in identifying trends, outliers, and patterns, crucial for interpreting the data and understanding the dynamics of COVID-19's impact on healthcare systems.

Insights Gained and Future Scope:

- The regression model's effectiveness in predicting hospitalizations can inform resource allocation and preparedness in healthcare settings.
- The classification model's lower accuracy underscores the need for more nuanced approaches to predict staffing shortages.
- Future work includes refining these models and exploring additional data sources for improved predictions, vital for strategic planning in healthcare to manage future pandemic scenarios.
- This comprehensive summary provides a deeper understanding of the methodologies, findings, and implications of the project, emphasizing its importance in healthcare strategy and pandemic management.

Conclusion:

The conclusion drawn from the "COVID-19 Data Analysis" project emphasizes the potential of data-driven models in healthcare planning during pandemics. It highlights the effectiveness of the regression model in predicting hospitalization rates, which can guide resource allocation and preparedness. However, it also acknowledges the need for improvement in the classification model for predicting staffing shortages. The project underscores the importance of refining these models and exploring additional data for better predictions, reinforcing their value in strategic healthcare planning and response to future pandemic scenarios.