Chronic Conditions Prediction Methodology: Epidemiological Analysis in a Brazilian Self-Managed Health Provider

Poster number

[1]Antonio C. Neto, [1]Weverton V.S. Rosa, [1]Raylayne F.B. Bernardo, [1]Carlos W.G. de Barros, [1]Bruna M. dos Santos, [1]Wanderley A. Shiguti
([1]Weverton Vieira da Silva Rosa)
[1]Caixa de Assistência dos Funcionários do Banco do Brasil (CASSI), Brasília, Federal District, Brazil

The study **implemented a predictive methodology** to estimate the prevalence of *Diabetes Mellitus*, *Hypertension*, and *Dyslipidemia* using **supervised Machine Learning** and **Logistic Regression**. The findings show a **heterogeneous distribution** of these chronic conditions, highlighting significant differences by **age**, **gender**, and **geographic location**. Beneficiaries were categorized with probabilities of **85%** for Diabetes, **70%** for Hypertension, and **50%** for Dyslipidemia. The analysis provides **valuable insights** for targeted **interventions**, supporting the development of **public policies** and **prevention programs**.

BACKGROUND

Chronic diseases like Diabetes Mellitus, Hypertension, and Dyslipidemia are major public health concerns due to their prevalence and impact. Effective management of these conditions is often hampered by incomplete clinical records and inconsistent data collection. To address these issues, this study aimed to develop a predictive methodology using supervised Machine Learning and Logistic Regression to estimate the prevalence of these chronic conditions among beneficiaries of a self-managed health provider in Brazil. By analyzing five years of administrative data, the study identified patterns in healthcare utilization and provided insights for personalized health management, targeted interventions, and the formulation of data-driven public policies and prevention programs.

METHODS

The study employed a predictive model using supervised Machine Learning and Logistic Regression to estimate chronic disease prevalence among beneficiaries based on healthcare utilization data from 2016 to 2020. The data included variables like age, gender, consultations, emergency visits, and hospital admissions. The dataset was divided into training (80%) and testing (20%) sets, and the model was evaluated for specificity and accuracy. Beneficiaries were categorized based on predicted probabilities (85% for Diabetes, 70% for Hypertension, 50% for Dyslipidemia). This approach enabled early identification of chronic conditions, offering a valuable tool for targeted healthcare interventions and policy planning.

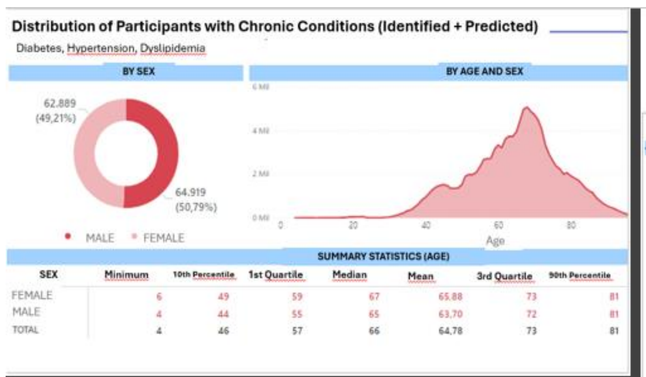
RESULTS

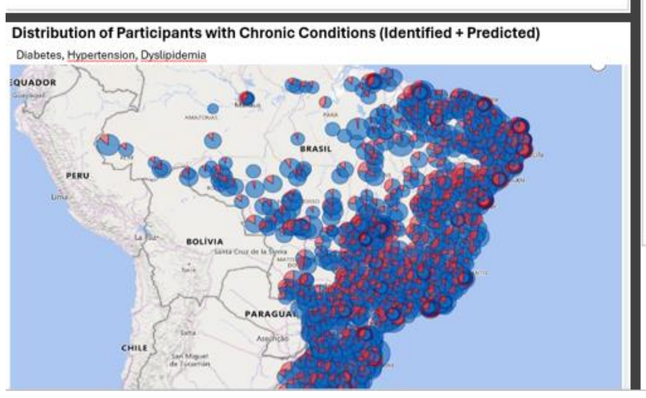
The study found a heterogeneous distribution of chronic conditions—Diabetes Mellitus, Hypertension, and Dyslipidemia—among the beneficiaries of a self-managed health provider in Brazil. The predictive model developed using supervised Machine Learning and Logistic Regression identified 83,977 individuals as hypertensive, with an additional 35,054 likely to have the condition, representing 23.87% of the active population. For Dyslipidemia, 46,683 individuals were identified, and 69,882 more were predicted to be at risk, totaling 23.6% of the population. Regarding Diabetes, 26,097 individuals were identified, with 17,386 additional likely cases, affecting 8.7% of active beneficiaries.

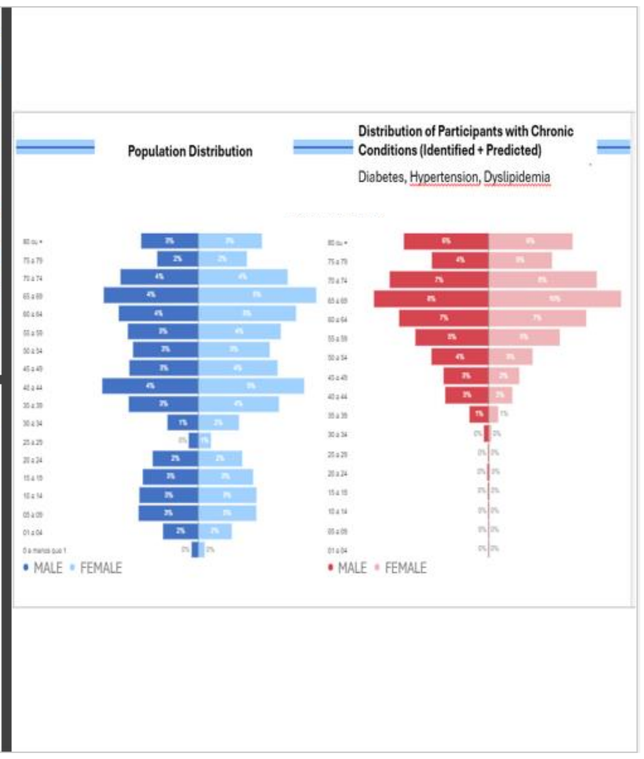
To provide a visual and dynamic representation of these findings, a Business Intelligence (BI) dashboard was developed to allow the health provider to monitor these data on a monthly basis. The figures presented in this section reflect demonstrating the effectiveness of the predictive model and the distribution patterns identified. These images offer insights into the demographic breakdown of the population, the prevalence of chronic conditions by age, gender, and geographic location, and the utilization of healthcare services, such as consultation rates, emergency visits, and hospitalizations.

For example, the dashboard's figures show the stratification of the population by the use of healthcare services, highlighting that the highest prevalence of chronic conditions is found in groups with more frequent service use. Additional graphs detail the distribution of chronic conditions across different age and gender groups, providing a clearer picture of the population's health status.

RESULTS CONTINUED







CONCLUSIONS

The study's findings demonstrate the effectiveness of using predictive modeling with Machine Learning and Logistic Regression to identify and predict the prevalence of chronic conditions like Diabetes Mellitus, Hypertension, and Dyslipidemia among beneficiaries of a self-managed health provider in Brazil. The results revealed a heterogeneous distribution of these conditions, varying by age, gender, and geographic location, and highlighted the importance of leveraging healthcare utilization data to identify high-risk groups. This approach enables healthcare providers to allocate resources more effectively and design targeted interventions and personalized care plans tailored to the specific needs of different population segments.

These findings are crucial for advancing data-driven approaches in healthcare management. By regularly updating data through a Business Intelligence (BI) dashboard, health providers can monitor trends in chronic disease prevalence and healthcare usage, enabling more proactive and informed decision-making. This strategy supports the development of more efficient public health policies and prevention programs, enhancing the sustainability of healthcare systems and improving patient outcomes by focusing on early identification and effective management of chronic diseases.

ADDITIONAL KEY INFORMATION

Other Key Information:
This study utilized a Business Intelligence (BI) dashboard to allow for real-time monitoring and management of chronic disease prevalence and healthcare service utilization, enabling data-driven decision-making for tailored public health strategies.

Additional Resources:
Further details on the methodology, data, and predictive models used in this study are available upon request. Researchers and healthcare providers interested in applying similar approaches can contact the authors for guidance and collaboration opportunities.

Author Contact Information:
For more information, please contact **Weverton Vieira da Silva Rosa** at **weverton.rosa@cassi.com.br** or reach out to any of the co-authors at the Caixa de Assistência dos Funcionários do Banco do Brasil (CASSI), Brasília, Federal District, Brazil at gestaopopulacional@cassi.com.br.

Funding Source:
This research was funded by the **Caixa de Assistência dos Funcionários do Banco do Brasil (CASSI)**, which provided the necessary data and resources for the completion of this study.

Conflicts of Interest:
The authors declare no conflicts of interest related to this study. All findings and conclusions are based on independent analysis and research conducted by the authors.

Acknowledgements:
The authors wish to acknowledge the support and contributions of the data management and analytics team at CASSI, as well as the healthcare professionals who provided valuable insights during the research process.

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024