The Harms of Alcohol on Society and Health: A Machine Learning Approach

---

Submitted By:

1. Bhargav yoga

NEU ID: 002793012

Email ID: yoga.b@northeastern.edu

Percentage Contribution: 50%

2. Sara Aminoshari

NEU ID: 002820310

Email ID: Amonishari.s@northeastern.edu

Percentage Contribution: 50%

Course: IE7945 53267 Master's Project SEC 05 Summer Full 2024 [XCR-2-OOO]
Submission Deadline: June 1, 2024

Section 1: Introduction to the Project

Project Goals and Objectives

The primary goal of this project is to analyze the harms of alcohol on society and health using machine learning and data analysis techniques. The specific objectives are:

1. To identify patterns and trends in alcohol consumption and its health impacts.
2. To develop predictive models that can forecast alcohol consumption rates and associated health risks.
3. To provide insights that can inform public health policies and interventions aimed at reducing the negative impacts of alcohol consumption.

Problem Statement

Alcohol consumption is a major public health concern, contributing to a range of social and health issues, including addiction, chronic diseases, accidents, and economic costs. Understanding and predicting alcohol consumption patterns and their health impacts is crucial for developing effective public health strategies. This project seeks to leverage machine learning techniques to analyze alcohol-related data and provide actionable insights for policymakers and healthcare providers.

Expected Outcomes and Impacts

- **Predictive Models**: Development of robust machine learning models that can predict alcohol consumption rates and associated health risks.
- **Data Insights**: Identification of key patterns and trends in alcohol consumption and its health impacts.
- **Policy Recommendations**: Evidence-based recommendations for public health policies aimed at mitigating the harms of alcohol.
- **Public Awareness**: Increased awareness of the societal and health impacts of alcohol consumption through data-driven findings.

Section 2: Annotated Bibliography

1. **Artificial intelligence and machine learning**
    - **Authors**: Ashton Pike, Barlas Benkli, Sayyed Omar Gilani, Salman Hirani
    - **Link**: Artificial intelligence and machine learning
    - **Citation**: A. Pike, B. Benkli, S. O. Gilani, and S. Hirani, "Artificial intelligence and machine learning," Frontiers in Artificial Intelligence, vol. 3, pp. 1-10, 2020. [Online]. Available: Frontiers.
    - **Summary**: Explores AI and ML applications in healthcare, emphasizing predictive models for health risks, including alcohol-related conditions. Highlights the potential of these technologies to enhance preventive measures and personalized treatment plans.
    - **Relevance**: Provides foundational knowledge on using AI and ML in healthcare, which is crucial for developing models to predict and analyze the health impacts of alcohol consumption.
2. **A Survey of Deep Learning for Electronic Health Records**
    - **Authors**: Jiabao Xu, Xuefeng Xi, Jie Chen, Victor S. Sheng, Jieming Ma, Zhiming Cui
    - **Link**: A Survey of Deep Learning for Electronic Health Records
    - **Citation**: Xu, Jiabao & Xi, Xuefeng & Chen, Jie & Sheng, Victor & Ma, Jieming & Cui, Zhiming. (2022). A Survey of Deep Learning for Electronic Health Records. Applied Sciences. 12. 11709. 10.3390/app122211709.
    - **Summary**: Reviews deep learning applications in electronic health records, focusing on methods for analyzing large health datasets. Relevant for identifying patterns and predicting outcomes related to alcohol consumption.
    - **Relevance**: Offers insights into deep learning techniques that can be applied to health data, aiding in the analysis and prediction of alcohol-related health impacts.
3. **Alcohol consumption rate prediction using machine learning algorithms**
    - **Authors**: Advait Singh; Vinayak Singh; Mahendra Kumar Gourisaria; Ashish Sharma
    - **Link**: Alcohol consumption rate prediction using machine learning algorithms
    - **Citation:**A. Singh, V. Singh, M. K. Gourisaria and A. Sharma, "Alcohol Consumption Rate Prediction using Machine Learning Algorithms," 2022 OITS International Conference on Information Technology (OCIT), Bhubaneswar, India, 2022, pp. 85-90, doi: 10.1109/OCIT56763.2022.00026.

- **Summary**: Discusses the use of various machine learning algorithms to predict alcohol consumption rates. Provides insights into how data-driven approaches can help in understanding and mitigating alcohol use disorders.
- **Relevance**: Directly relevant to the project's goal of predicting alcohol consumption rates, which can help in assessing the societal and health impacts of alcohol.

4. **Data Mining Techniques for Medical Data: A Review**
   - **Author**: Subhash Chandra Pandey
   - **Link**: Data Mining Techniques for Medical Data: A Review
   - **Citation:**Pandey, Dr. Subhash. (2016). Data Mining Techniques for Medical Data: A Review. 10.1109/SCOPES.2016.7955586.
   - **Summary**: Reviews data mining techniques applied to medical data, including the analysis of alcohol-related health impacts. Highlights methods for extracting meaningful patterns and improving healthcare outcomes.
   - **Relevance**: Essential for understanding how data mining can be used to extract insights from health data, supporting the analysis of alcohol's impact on health.
5. **Deep Learning for Health Informatics**
   - **Authors**: Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, Guang-Zhong Yang
   - **Link**: Deep Learning for Health Informatics
   - **Citation:**Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep Learning for Health Informatics. IEEE J Biomed Health Inform. 2017 Jan;21(1):4-21. doi: 10.1109/JBHI.2016.2636665. Epub 2016 Dec 29. PMID: 28055930.
   - **Summary**: Focuses on the application of deep learning in health informatics, discussing models and techniques applicable to alcohol-related health data analysis. Highlights the role of deep learning in improving diagnostic accuracy and patient care.
   - **Relevance**: Provides advanced techniques in deep learning that can enhance the predictive models for alcohol-related health outcomes.
6. **Deep Learning for Health Informatics: Recent Trends and Future Directions**
   - **Authors**: Siddharth Srivastava, Sumit Soman, Astha Rai, Praveen K Srivastava
   - **Link**: Deep Learning for Health Informatics: Recent Trends and Future Directions
   - **Citation:**Srivastava, Siddharth & Soman, Sumit & Rai, Astha & Srivastava, Praveen. (2017). Deep learning for health informatics: Recent trends and future directions. 1665-1670. 10.1109/ICACCI.2017.8126082.
   - **Summary**: Outlines recent advancements and future prospects in deep learning for health informatics. Discusses potential applications in studying the health impacts of alcohol consumption and developing predictive models.
   - **Relevance**: Discusses cutting-edge advancements in deep learning, providing valuable insights for applying these techniques to analyze and predict the impacts of alcohol on health.

Section 3: Dataset Overview and Repository Details

Dataset 1: Global Information System on Alcohol and Health

- **Source**: World Health Organization (WHO)
- **URL**: WHO Global Information System on Alcohol and Health
- **Description**: This dataset provides comprehensive data on alcohol consumption, health impacts, and policy measures at global, regional, and national levels. It includes various

indicators such as per capita alcohol consumption, prevalence of alcohol use disorders, and associated health outcomes.
- **Size**: Varies depending on the data subset (e.g., country-specific data, time series data)
- **Attributes**: Metrics such as age-standardized prevalence rates, alcohol consumption patterns, health consequences (e.g., liver cirrhosis, alcohol-attributable deaths), and policy responses.
- **Challenges**: Data may need preprocessing to handle missing values and integrate multiple sources. Analysis may require normalization and standardization techniques.

Dataset 2: Alcohol Data and Statistics from the CDC

- **Source**: Centers for Disease Control and Prevention (CDC)
- **URL**: CDC Alcohol Data and Statistics
- **Description**: This dataset includes data on alcohol consumption patterns, health outcomes, economic costs, and policy impacts in the United States. It provides valuable insights into the societal and health impacts of alcohol consumption.
- **Size**: Varies by specific dataset (e.g., national surveys, health statistics)
- **Attributes**: Alcohol consumption patterns, health outcomes (e.g., alcohol-related deaths, hospitalizations), economic costs, policy measures, and demographic information.
- **Challenges**: May require data cleaning and integration with other datasets for comprehensive analysis; data might have geographical or temporal limitations.

GITHUB:REPOSITORY