**INTERIM REPORT**

DATA301 Project

Exploratory Data Analysis:

Predicting the Risk of Hypertension Using NHANES Data

**Cheryl Chan**:

**Gale Bueno**: 300533945

**Nguyen Van**:

September 6, 2022

The aim of this interim report is to present and discuss the progress of the DATA301 project “Exploratory Data Analysis: Predicting Hypertension Using NHANES data”. The report will include the main objectives of the project and the exploratory analysis findings.

**Background and Data**

High blood pressure is a condition that affects over 103.3 million people in the US and 1.3 billion people globally (Ye et al., 2020). It can lead to various health conditions such as heart failure and stroke which is one of the most common leading causes of death in the United States. An estimated 17 million cardiovascular deaths globally are caused annually by factors such as hypertension  (López-Martínez et al., 2020). This sign indicates the necessity of having a thorough grasp of the numerous risk factors that might affect blood pressure since lifestyle decisions can impact both the prevalence and development of hypertension (Wang et al., 2015). ). Identifying the risk of hypertension early on can lead to early treatment and a lower likelihood that a patient will succumb to further ailments.

The utilization of machine learning to build models for disease categorisation is rapidly advancing. The pace at which data is processed and evaluated is accelerated by machine learning. With very slight deployment adjustments, predictive analytics algorithms can now train on even broader data sets and do more in-depth research on a variety of aspects (Elshawi et al., 2019). Several artificial neural network models have been developed to predict the risk of hypertension using a variety of data sources. Nevertheless, most of these models were developed based on smaller samples of data. Model accuracy can still be improved by representing a larger dataset.

The proposed objective of this project is to build a binary classification model which predicts whether an individual has hypertension or not using the National Health and Nutrition Examination Survey (NHANES) data. The NHANES is designed to evaluate the health and nutritional status of adults and children in the United States using interviews and physical examinations. Data from 2007- 2008 to 2017-2020 (pre-pandemic) will be used.

**NHANES DATA**

Table 1: A list of risk factors of hypertension across multiple studies.

|  |  |
| --- | --- |
| **Author** | **Risk Factors of Hypertension** |
| Marques et al. (2020) | Waist circumference, alcohol consumption, skin colour, smoking |
| Rodrigues et al. (2019) | Family history of hypertension, exercise, gender, weight, cholesterol/ lipid levels |
| Huang et al. (2019) | Age, resting heart rate, weight (overweight, obese etc.), lipid levels, uric acid levels, blood glucose levels/ diabetes, kidney function/kidney disease |
| López-Martínez et al. (2020) | Race, BMI, smoking status, gender, age, kidney disease, diabetes |

Reference

Elshawi, R., Al-Mallah, M. H., & Sakr, S. (2019). On the interpretability of machine learning-based model for predicting hypertension. *BMC medical informatics and decision making, 19*(1), 1-32.

Huang, Y., Deng, Z., Se, Z., Bai, Y., Yan, C., Zhan, Q., Zeng, Q., Ouyang, P., Dai, M., & Xu, D. (2019). Combined impact of risk factors on the subsequent development of hypertension. *Journal of Hypertension, 37*(4), 696-701.

López-Martínez, F., Núñez-Valdez, E. R., Crespo, R. G., & García-Díaz, V. (2020). An artificial neural network approach for predicting hypertension using NHANES data. *Scientific Reports, 10*(1), 1-14.

Marques, A. P., Szwarcwald, C. L., Pires, D. C., Rodrigues, J. M., Almeida, W. d. S. d., & Romero, D. (2020). Factors associated with arterial hypertension: a systematic review. *Ciência & Saúde Coletiva, 25*, 2271-2282.

Rodrigues, F., Coelho, P., & Mateus, S. (2019). Risk factors and arterial hypertension. *Journal of Human Sport and Exercise*, S1772-S1775.

Wang, A., An, N., Chen, G., Li, L., & Alterovitz, G. (2015). Predicting hypertension without measurement: A non-invasive, questionnaire-based approach. *Expert Systems with Applications, 42*(21), 7601-7609.

Ye, X., Zeng, Q. T., Facelli, J. C., Brixner, D. I., Conway, M., & Bray, B. E. (2020). Predicting optimal hypertension treatment pathways using recurrent neural networks. *International Journal of Medical Informatics, 139*, 104122.