CS6220 Data Mining Technique Group Project

Diabetes Prediction and Prevention

Team member: Paul Cruz, Chih-Ming Sun, Haocheng Yang

***What’s the format of the final report?***

Although you are not expected to write a full-fledged research paper, this exercise helps you understand the structure of the components thereof. The project report might have the following seven main sections -

a. Abstract - This is a brief summary not exceeding 500 - 750 words describing the problem, the solution method, and the results.

b. Introduction – Approximately one to two pages long and contains (i) Statement of the problem you are trying to solve  
(ii) Why is it important to solve this problem?

(iii) Background information and bit of literature survey to present what’s already known about this problem.

c. Methodology - What’s your solution, how do you propose to solve the identified problem?

d. Code - Brief explanation of the code.

e. Results – here you just state the results as you see them. E.g., In a fictitious study you were studying a data collected over a period of 50 years of height of children aged 12 years. You can state the trends shown in the data, like in all countries the average height of 12-year-olds seem to be increasing by 3 mm every decade, except in countries C and D. In these two countries it seemed to decrease by 2 mm in the 1980s.

f. Discussion - interpretation of the results - what do the numbers really mean, do they make sense. E.g., In the fictitious study we used to illustrate the results section, Does it make sense that in countries C and D the average height decreased. Probably, there was a famine in Country C and Country D had an epidemic that affected the health of children in the 1980s.

g. Future Work - Is this study conclusive or does it lead to some future work? h. Conclusion - what conclusions can you draw.

i. References – Research papers, articles, and Internet resources referred to in the rest of the report.

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

<https://escholarship.org/uc/item/6rh07945>

Huang, R. (2021). Prediction of Pima Indians Diabetes with Machine Learning Algorithms. *UCLA*. ProQuest ID: Huang\_ucla\_0031N\_19508. Merritt ID: ark:/13030/m5md4q60. Retrieved from https://escholarship.org/uc/item/6rh07945