The University of Memphis
Computer Science
COMP 7/8150-Fundamentals of Data Science
Project Proposal: **Health Insurance Premium at PSI**
Vinay Badam, Nikhil Banka, Data Science, Fall 2022

PROJECT GOAL AND SCOPE

The goal of this project is to predict the health insurance premiums of customers of PSI six months in advance. The solution will be one python program that predicts the premium with an average relative error of at most 20%.

BACKGROUND/LIT REVIEW

The articles [2] and [3] worked on the insurance prediction problem using a linear regression model based on the dataset [4]. The data available from the dataset [4] contains the required data of insurance firms. However, this resulted in a higher-than-expected relative error or mean squared error for the data. This dataset [4] contains data about the insurer, their dependents, their medical expenses over a year, etc. These resources and others will be used to complete the project successfully.

CASE STUDY

This study provides a way to predict insurance premiums using several features of individuals such as age, sex, region, physical/family condition, expenses, and location. The concepts presented in [2] will be used to predict the average relative error of insurance premiums.

TAKE-HOME DELIVERABLE

The deliverables for this project will be one python program that predicts the premium with an average relative error of at most 20%.

REFERENCES

[1] Centers for Medicare Services (CMS)
https://data.cms.gov/search?keywords=HEALTH%20INSURANCE&sort=Relevancy

[2] https://www.kaggle.com/code/mariapushkareva/medical-insurance-cost-with-linear-regression

[3] https://www.kaggle.com/datasets/hhs/health-insurance-marketplace

[4] https://www.kaggle.com/datasets/hhs/health-insurance-marketplace