Project Proposal Team 2 Project 4

Nick Mincucci

George Achie

Ian Cook

**Title**

Healthcare Cost

**Project Description**

The objective of this project is to create a predictive model utilizing the dataset of over 1300 entries containing information on users’ demographics, geographical location, and healthcare charges. The model should work to predict attributing factors to expected healthcare costs.

**Research Question**

* Which region of the United States has higher healthcare costs?
* The effect of the number of dependents on healthcare costs.

**Data Set to be Used**

US Health Insurance Dataset

<https://www.kaggle.com/datasets/teertha/ushealthinsurancedataset>

The dataset to be used contains 1338 rows of insured data, where the Insurance charges are given against the following attributes of the insured: Age, Sex, BMI, Number of Children, Smoker and Region.

**Breakdown of Tasks**

Data Cleaning/Filtering

* Remove duplicates
* Handle missing values
* Standardized data

Data Initialization/Training/Evaluated

* Python script to initialize, train, and evaluate the model
* Utilize SQL or Spark to retrieve the data
* Predictive measure of 75% classification accuracy is met

Data Model Optimization

* Iterative changes utilized to improve the model and create change to the performance
* Documentation of the model and optimization/evaluation process is included

Visualization Development

* Develop appropriate visualizations to show the results of the data and modeling
* Utilize multiple techniques such as Matplotlib and Tableau

Presentation

* Design and implement a 15 minute presentation displaying the visuals and model in an informative fashion. Allowing the audience to both understand the model while seeing appropriate visualizations.

**Ethical Considerations**

Throughout the project, ethical considerations will be considered, particularly concerning data privacy and representation. All personal data will be anonymized, and the visualizations will aim to present data accurately without misleading interpretations.