Logo, company name

Description automatically generated

***Proposal for AI Application Development***

**Subject: Applied Artificial Intelligence and Machine Learning (ADMN5016-F22-101-3672-K)**

**For: Professor Joseph Santarcangelo**

**By: Pranjal Thakkar & Kevin Abraham**

**Date: 9th November 2022.**

***Proposal for AI Application Development for predicting insurance costs***

***Introduction:***

A health insurance provider can only be profitable if the premiums it collects exceed the amount it pays for the beneficiaries' medical care. However, even though some ailments are more common among particular populations, it is difficult to forecast how much a patient's medical expenses will be because the majority of the money comes from their unique conditions. The aim of this project is to forecast insurance premium costs with accuracy using information about individuals, such as age, BMI, smoking status, etc. Furthermore, we will establish what factor has the most impact on insurance rates. These projections could be used to produce actuarial tables that adjust annual premium prices upwards or downward in accordance with anticipated treatment expenses.

***Objective:***

* When properly applied, predictive analysis has offered health insurers the ability to improve the precision of their premium pricing, create specialized health insurance plans and services, and develop stronger relationships with their customers.
* As a result, the main goal of this project is to estimate insurance premium costs using the information collected from individuals, enabling insurance companies to make smart and trustworthy predictions.

***Model Development:***

* We want to develop a model that predicts the insurance premiums that beneficiaries should pay to insurance companies to preserve or boost earnings.
* We would utilize the Linear Regression approach and the Support Vector Machine method to create the model.

\*Statistical models will be chosen according to prediction accuracy.

***Assumptions for model development:***

1. We made the supposition that the insurance market in the United States of America is a monopolistic market for the sake of simplicity when developing this model.
2. The model is developed solely using data from the USA.

***Cost of developing the model:***

* The development, testing, and deployment of artificial intelligence application will cost anything between $20,000 and $35,000.

***Benefits of developing the application:***

1. This application can be used to generate actuarial tables which can help to adjust the cost of annual premiums depending on the anticipated medical cost.
2. The application would assist insurance providers in establishing premium costs based on previous data to generate profits.

***Risks associated with the application***

1. Generally, most medical costs are driven by the special health conditions of individuals. The insurance cost prediction model might have a higher error rate.
2. Since the model is based on the personal medical costs of individuals there is a significant amount of risk of data getting breached.
3. The limitation of the model is that since we don’t have any real-time health data of individuals it might affect the reliability of the predicted values.

***Market Size:***

* Ben Lewis claims that there are 900 insurance firms in the US, which is a sizable market in his article titled "Top 5 Health Insurance Companies in The U.S." So, it's safe to say that if the application is successful, it would be widely used.

***Description of the Dataset***

* The data set selected for this project is the U.S health insurance data set which is available on [Kaggle](https://www.kaggle.com/datasets/teertha/ushealthinsurancedataset) with 7 columns namely age, sex, BMI, children, smoker, region, and charges, 1338 rows, and 9366 available data points. There are no missing or undefined values in the dataset.

The description of the columns/features are as below.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Description** | **Data Type** |
| Age | Age of the beneficiary (primary) | Integer |
| Sex | Gender of the primary beneficiary (male/female) | Boolean |
| BMI | Body mass index, providing an understanding of the body (weight to height ratio) | Float |
| Children | the number of children/dependents covered by insurance | Integer |
| Smoker | Whether the primary beneficiary is a smoker or not. | Boolean |
| Region | Which region does the beneficiary belong to? i.e., the beneficiary’s residential area. | String |
| Charges | Individual medical costs billed by the health insurance. | Integer |