

## Project Proposal

This project aims to build a recommendation system for health insurance plans using data contained in five CSV files [Benefits\_Cost\_Sharing.csv, Rate.csv, Network.csv, Plan\_Attribute.csv, Business Rule] with approximately 270 parameters in total. These CSV files contain information about different insurance plans, the benefits included in those plans, cost of the plans, the networks of providers that offer these plans, and the attributes of these plans. The objective of the recommendation system is to suggest health insurance plans based on the value of the parameter provided by the user. Some examples of the parameters are health conditions, budget, location, marital status, dependencies, etc.

The dataset used our project is termed as Health Insurance Marketplace dataset. The dataset has 5 tables which we find the most relevant. The first is Benefit Cost and Sharing [Parameters: benefit name, business year, PlanID, IssuerID, LimitQty, LimitUnit]. In the benefit cost sharing, we combine benefits based on common plan ID. Limit Qty and Limit Unit provides an overview of how many visits/amounts is included in insurance. Second table is Rate [Parameters: Age, IssuerId, PlanID, StateCode, Tobacco, Primary Subscriber, one dependent, Primary subscriber 2 dependents] which would help us find out the rate of the insurance plan. The third table is Plan Attributes [ Parameter: IssuerId, PlanID, Source, Plan Start, Plan expiration Date, Guaranteed Rate, NetworkId, CSRVariation Time etc..] which provides information about different aspects of the plan. Fourth table is the Business Rules [Parameters: IssuerID, Source Name, StandardComponent, Two Parent Family Max dependent rule, Single Parent Family max dependent rule] which helps us to filter out best suited plan for the user.

We aim to address few questions which will help in creation of recommender system and analyse it. Firstly, what are the plans prevalent in the area where the user lives. How would the situation of a person impact the cost of the plan? We aim to draw a comparison between the two algorithms. Since we do not have any pre-existing demographic data, Content based filtering will be used to create a recommendation system. The other algorithm that we'll be using is Decision Tree. Content-based filtering and decision tree algorithm are both commonly used in recommendation systems to provide personalized recommendations to users. Content-based filtering is a technique that recommends items to a user based on the features of items that the user has previously shown interest in. On the other hand, a decision tree algorithm is a machine learning technique that can be used to build a model for decision-making. In the context of recommendation systems, a decision tree algorithm can be used to predict which items a user will be interested in based on the attributes of the items and the user's preferences.