Targeted Wellness Benefits Promotion

Maximizing the impact of benefits programs on member health

Problem Statement

- Health and Welfare benefits in the United States are complicated and confusing.
- Employees often underutilize their benefits, costing both employers and employees.
- A personalized approach to promoting benefits, driven by specific behaviors and correlated health outcomes, will help guide employees in navigating and using their benefits.
- Helping employees use the benefits available to them will both help employees and employers save money. It would also help employees better maintain their health and wellness.

Related Work

There are a number of studies that have been conducted using the BRFSS data:

Correlative Analysis

- A 2023 study published in the Journal of Public Health and Emergency used the 2020 BRFSS dataset to investigate the association between adverse childhood experiences (ACEs) and self-reported mental health conditions in adults [1]
- A 2019 study published in Preventing Chronic Disease used BRFSS data to assess the prevalence of subjective cognitive decline in adults aged 45 years or older in 49 states [2]

Machine Learning Algorithms

- A 2023 study published in Patterns used the 2021 BRFSS dataset to investigate machine learning algorithms and data augmentation techniques for predicting chronic kidney disease [3]

Data Source

Behavioral Risk Factor Surveillance System (BRFSS)

- The 2022 BRFSS data contains 445,132 records. Each record contains the question results for one survey performed.
- The data has 326 columns, one column per potential question asked during the survey. This includes
  - Behavioral questions such as “Have you smoked at least 100 cigarettes in your entire life?”, “At what kind of place did you get your last flu shot or vaccine?”
  - Health outcome questions such as “Has a doctor, nurse, or other health professional ever told you that you tested positive for COVID 19?”, “In general, how satisfied are you with your life?”

Proposed Work

- The CDC collects a large amount of data nationally, gathering behaviors and health outcomes that we will use in this study.
- This data can be used to find correlations and associations between behaviors and health outcomes.
- By mining these relationships, we can discover questions that we can pose to employees. These questions can help us tailor benefits recommendations to employees directly.
- In addition, we can drive employees to use the benefits they have available to them to positively impact their health and wellness.

Proposed Work

Specifically, I will employ three primary data mining techniques:

- Clustering
  I intend on using clustering to attempt to identify groups of survey results that are likely to lead to specific health outcomes.
- Association Rule Mining
  Using methods like FP-Growth and Association Rule Mining can help identify survey answers that most likely result in a specific health outcome.
- Classification
  I would like to build decision tree models that attempt to classify specific health outcomes. Decision trees will provide both a classification method, but also explainable parameters we can use to determine specific health outcomes.

Proposed Implementation

- Using the relationships we identify, we can build models to help promote specific benefits that best meet employees' health needs.
- Using a personalized approach to meet employees where they're at, increasing employee utilization of their benefits will have a positive impact on employee's health and wellbeing.
- In addition by providing this as a service, we can improve the return on investment for employers. In addition to benefiting employers, this can help distinguish our company in the market.

Evaluation

To evaluate each of the specific methods above:

- For clustering, I intend on using silhouette coefficients to analyze how well the clustering methods perform.
- For association rule mining, using support and confidence will provide methods of identifying the most relevant rules.
- For classification, accuracy, precision, recall, and F1 scores will be the standard evaluation methods. In addition, a confusion matrix will help evaluate model performance.
- To measure the overall success of the project, success will be determined by if we can accurately determine health outcomes and associated health and welfare benefits using behavioral questions. Implementation will also be important. Determining whether the methods can be reasonably implemented in a real-world setting is critical.

Timeline

The project will be time boxed to an 8 week period total. The project will be divided into the following phases:

- Phase 1 (1 week): Data acquisition, basic data ingestion, and beginning EDA
- Phase 2 (1 week): Extended EDA. Beginning clustering methods.
- Phase 3 (1 week): Beginning Association Rule Mining.
- Phase 4 (1 week): Beginning Classification.
- Phase 5 (2 weeks): Wrapping up all methods.
- Phase 6 (1 week): Conclusion of all work
- Phase 7 (1 week): Final report

Preliminary Results

Work completed so far:

- Clustering
  - Attempted Heirarchical Clustering across the entire dataset using Gower's Distance as a distance metric. This attempt failed due to complexity of task and limited compute resources.
  - To accommodate above, move to mini-batch K-Means using reduced data via MCA. Ran 150 different K-Means models with varying number of MCA components and Ks. Used silhouette score from each to find the most performant model.
  - Found that Clustering is likely an inappropriate method for this data, because the complexity of the data doesn't yield interpretable clusters.

Preliminary Results

Work completed so far:

- Frequent Patterns
  - Began running FP Growth and Associate Rule mining on full dataset. This approach was too memory intensive for resources available.
  - Moved to running FP Growth and association rules on smaller, more targeted subsets of data.
  - In progress: Beginning to identify subsets of data that are of interest and run FP Growth and Association Rule mining on those subsets.
  - On track for delivery in coming weeks.

Preliminary Results

Work completed so far:

- Classification
  - Not yet started.
  - On track for delivery in coming weeks.