Short Term Disability Duration Prediction Using Machine Learning

Erik Prananta | Supervisor: Dr. Greg Doyle | MSc in Data Science

Introduction

- Short-term disability insurance is a type of voluntary insurance that replaces part or all of an employee's income in the event of a temporary disability.

- The charts currently used to predict the duration of the disability are inaccurate.

Research Question

- To examine which machine learning model is the most suitable to predict short term disability duration.

Literature Review

- Previous research using gradient boosting was used to predict the time required for sick employees to recover and return to work.
- The model shows an excellent binary classification, but falls off when asked to classify based on 3 groups categories. This is due to the lack of predictive factors and small database.
- The GBM model's main problem is outliers, since every classifier is obliged to fix the error of its predecessor and the lack of scaling ability.

Technology

Methodology

- Cleaning the dataset and identifying relevant predictive factors.
- Categorizing type of short term disability based on recovery length.
- Applying support vector machine, gradient boosting machine, k-means clustering and Naïve Bayes to the dataset.
- Fine tuning each model in order to improve accuracy.
- Comparing the models accuracy in order to determine the most suitable algorithm.

The Next Step

Implement the algorithm with the highest accuracy for testing in a real world environment and train the model to recognize a long term disability.

References

Obermeyer, Z. and Emanuel, E., 2016. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, 375(13), pp.-1216-1219.

Na, K. and Kim, E., 2019. A Machine Learning-Based Predictive Model of Return to Work After Sick Leave. Journal of Occupational and Environmental Medicine, 61(5), pp.e191–e199.

Natekin, A. and Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7.