Predicting Work-Life Balance Scores For Men

Hassan Ali, Jack Baxter

Slide 1:

Goal: To create an effective ML model that predicts a work life balance scores for men based on identified lifestyle variables of high influence

Why: Less than half of men who experience chronic stress and other mental health related issues speak up on the matter. By contributing to ML research for men's health we can advance the field of statistic-based ML to assist providers in recognizing the signs of a struggling patient.

Motive: With busy schedules balancing work and the MS AAI program, insights from the project can be applied to our own lives as adjustments can be made based on beneficial variables identified in work-life balance

Slide 2:

Data: Lifestyle_Wellbeing set from Kaggle

Variables: lifestyle_wellbeing(timestamp, FRUITS_VEGGIES, DAILY_STRESS, PLACES_VISITED, CORE_CIRCLE, SUPPORTING_OTHERS, SOCIAL_NETWORK, ACHIEVEMENT, DONATION, BMI_RANGE, TODO_COMPLETED, FLOW, DAILY_STEPS, LIVE_VISION, SLEEP_HOURS, LOST_VACATION, DAILY_SHOUTING, SUFFICIENT_INCOME, PERSONAL_AWARDS, TIME_FOR_PASSION, WEEKLY_MEDITATION, AGE)

- Most variables are int64 type indicating a survey rating by the subject
- Due to nature of the data a regression analysis will be necessary (cannot classify our target)
- Aim to choose a specific set of listed variables above that best represent WLB score
- N = 12,757 for ample analysis
- 23 variables total per subject tested

Slide 3: chapter one material

- Cleaning the dataset
- Kaggle dataset came cleaned with no missing values or rows needing to be dropped
- Preliminary data exploration of histograms, outliers, potential removal of outliers
- Box plot visualization, summary stats, etc

Slide 4: model selection, chapter 3, chapter 2

- Training and test sampling methods
- Random forest model for regression analysis
- Chosen for ability to prioritize variables of high correlation
- Test different values for K (number of trees in forest)
- Identify which K value produces the best precision, accuracy and recall.

Slide 5/6: model building, chapter 2

- With our identified K value

- Correlation heat map of logreg on each variable with WLB score
- Choose top variables for K value (ex, 3 best correlations)
- Use these variables in the final model, show final model

Slide 7: chapter three, chapter 4
- Conduct traditional hypothesis testing on the variables of high interest to test whether or not they actually had significant influence on the overall WLB score
- Show process, show null hypothesis and conclusion
- Illustrate the sample distributions, (normal, binomial, gaussian?)

Slide 8: findings & significance
- How accurate was our model on the test data
- Confucian matrix for extra emphasis on this
- What were the variables of most influence in our model?
- How significant?

Slide 9: future directions & our goal
- Restate our goal and how our findings helped us achieve the goal
- Include some future directions for our data: is there a variable of interest we should pay close attention to? Flow?
- Include any potential flaws in our design and the ethical statistic reporting guidelines