Team Contributions

Project Title: Final Project (Social Media vs. Productivity)

Class Section: CS-301-011

Team Members: *Nona Harris, Melany Chaparro Mendez, Kaylee Hernandez*

I. Overview:

For our algorithms, we used the Regression Training models and ensured to follow the professor's instructions on pre-processing to clear out any inconsistencies using pre-processing and engineering featuring. We were already noticing issues of inconsistency visualizations, so we were hoping the regression model can present us with better circumstances.

We chose Decision Tree Regressor to extract non-linear relationships, while handling numerical and categorical features. As well as, Support Vector Regression with the RBF kernel allows us to also focus on non-linear regression but the models seemed to both underperform in our expectations. We used residual plots to visually assess how well the regression models captured actual productivity values. Allowing us to find out if the model was making accurate predictions. This allowed us to confirm our model is showing signs of underfitting and high bias. We believe our model corresponded to the dataset oversimplifying the complexity of human productivity by only measuring measurable behaviors.

II. Individual Contributions:

Name: *Nona Harris*

- Key Contributions:
  - Worked on the Canva Presentation theme and completing the Exploratory Data Analysis & Visualizations
  - Completed the Box Plot, Bar Chart, and Histogram Visualizations; Helped with decision tree figure

Name: *Melany Chaparro*

- Key Contributions:
  - Line plot, Correlation Matrix using HeatMap, Regression Model evaluation and comparison(SVR and Decision Tree), preprocessing check, engineering featuring
  - Filled in Canva slides for visualizations and training model evaluations

Name: *Kaylee Hernandez*

- Key Contributions:
    - Filled in Canva slides for visualizations/analysis
        - Did code for key features statistics in slides
    - Completed scatter plots
    - Completed facet grids

III. Challenges Encountered

- Visualizations had little to no correlation between variables
- The metrics to evaluate the performances of our algorithms had an overall low fit to our model.
- There were issues finding an algorithm that fit well with our model.
- Pre-processing was a stage that was not considered from the start, so we had an odd group of data due to outliers and null values.