**Project Overview**

As remote work becomes the new norm, it's essential to understand its impact on employees' mental well-being and productivity. What are the effects of remote work on employee satisfaction and productivity? What factors influence productivity and employee satisfaction with respect to work location? Are employees working fully remote, hybrid or onsite most likely to be satisfied and perform at a higher productivity level? This project seeks to answer the above questions.

We used a data set obtained from Kaggle that contained data across various industries and different regions of the world. The data provided valuable insights into how factors such as the number of hours worked, number of virtual meetings, mental health condition and job role affect employee satisfaction and productivity.

This analysis could help businesses, HR professionals, and researchers assess the growing influence of remote work on productivity and well-being and make appropriate decisions so that it would benefit both employees and companies in the long run.

Group Roles:

Data Clean up – Luke

Data Preprocessing – Bhagi and Luke

Data Analysis/Tableau/graphs etc. – Jay and Maher

Building the ML Model – Neural Network Deep Learning Model – Bhagi

Building the ML Model – Logistic Regression and Random Forest – Luke

Testing the Models – Bhagi and Luke

Creating the Documentation – Bhagi, Luke, Jay, Maher

Creating the Presentation - Jay, Maher, Luke, Bhagi

Data Set

Obtained from Kaggle: Remote work and mental health.

<https://www.kaggle.com/datasets/waqi786/remote-work-and-mental-health/data>

The original data set consisted of the following fourteen columns of which productivity and satisfaction were our target variables and all others were the feature variables.

Employee ID

Age

Gender

Job\_Role

Work\_Location

Hours Worked Per Week

Number of Virtual Meetings

Mental Health Condition

Access to Mental Health Resources

Productivity

Satisfaction

Physical Activity

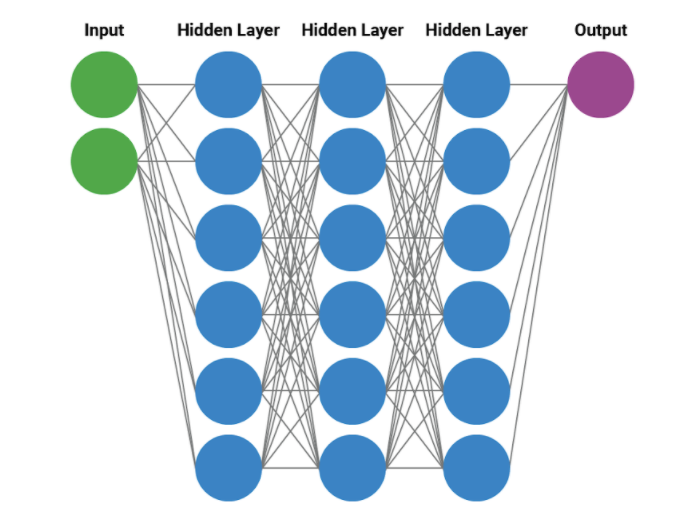
Sleep Quality

Region

Machine Learning Models

Our goal was to create a model with a high accuracy score. We explored the neural network deep learning model and as the accuracy was not optimal for this data set, we also explored two other models which were a Logistic Regression Mode and a Decision Tree Model.

Neural Network Model: Deep Learning Model



NN Deep Learning Modeling Process:

* We split the data set, scaled and standardize each feature.
* Fitted the model to our training data
* Transformed the data
* Created a sequential model
* Compiled and ran Model
* Created accuracy report

NN Model Definition

We ran several iterations our NN model by changing parameters to find the optimal setting with the highest accuracy:

Our Optimal Model Definition:

Input Layer: 33 Neurons

Hidden Layer 1: 40 Neurons

Hidden Layer 2: 40 Neurons

Hidden Layer 3: 20 Neurons

Output Layer: 1

Epochs: 30

* The best accuracy we could obtain was 0.52 (52% )with a loss of 1.12 which is not optimal.
* We tried to improve accuracy by increasing the number of neurons per hidden layer, Increasing the number of hidden layers, changing the number of epochs and none of these methods seem to improve accuracy

**Possible Reasons for low accuracy:**

1)Over Fitting: We believe the data adjusted too much to fit the training data and did not generalize well when fitting and predicting test data

2)Another possible cause could be the uniformity of data specifically the non categorical data in which the unique values were distributed evenly among the population and hence lack of variability in the dispersion.

3)To further improve accuracy we evaluated two additional models to see if the accuracy would improve with the same data,. The models we built and evaluated were  a logistic regressions model and a Decision Tree Model.