THEME 3:

FUTURE - READY WORKFORCE

Rishika Mehta, Krystal Pek,
Bernard Iskandar, Kriti Raja

TABLE OF *Contents*

01 Dataset Selection

02 ML models used

03 Our Solution

How do working conditions affect employees’ stress levels?

UNDERSTANDING

The Data

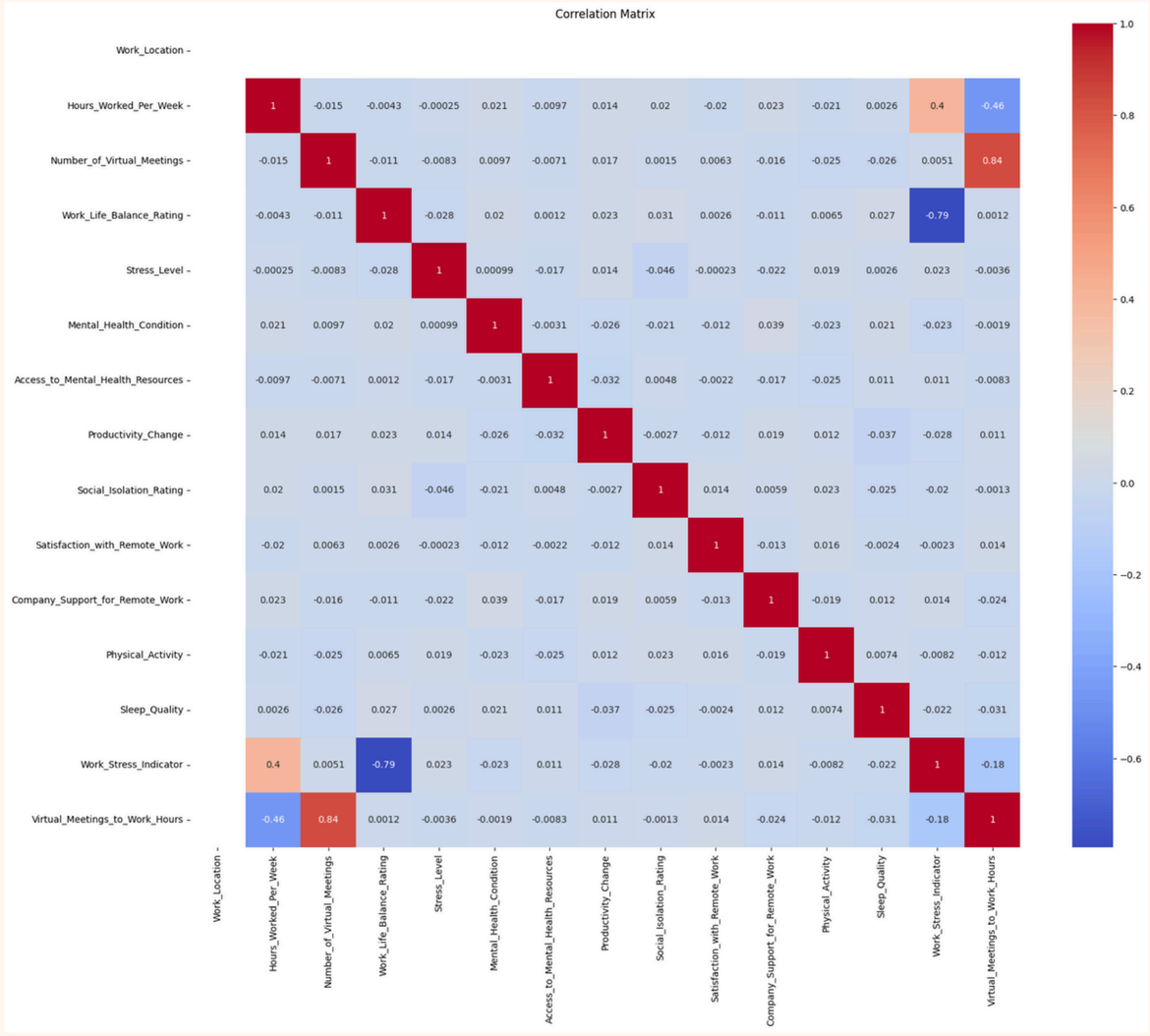
Work_Location	Hours_Worked_Per_Week	Number_of_Virtual_Meetings	Work_Life_Balance_Rating	Stress_Level	Physical_Activity	Sleep_Quality
Hybrid	47	7	2	Medium	Weekly	Good
Remote	52	4	1	Medium	Weekly	Good
Hybrid	46	11	5	Medium	NaN	Poor
Onsite	32	8	4	High	NaN	Poor
Onsite	35	12	2	High	Weekly	Average

DATA ANALYSIS

Understanding and identifying the important variables needed to build the model

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 15 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Work_Location                   5000 non-null   int64  
 1   Hours_Worked_Per_Week           5000 non-null   int64  
 2   Number_of_Virtual_Meetings      5000 non-null   int64  
 3   Work_Life_Balance_Rating        5000 non-null   int64  
 4   Stress_Level                    5000 non-null   int64  
 5   Mental_Health_Condition         5000 non-null   int64  
 6   Access_to_Mental_Health_Resources  5000 non-null   int64  
 7   Productivity_Change             5000 non-null   int64  
 8   Social_Isolation_Rating         5000 non-null   int64  
 9   Satisfaction_with_Remote_Work   5000 non-null   int64  
 10  Company_Support_for_Remote_Work  5000 non-null   int64  
 11  Physical_Activity               5000 non-null   int64  
 12  Sleep_Quality                   5000 non-null   int64  
 13  Work_Stress_Indicator           5000 non-null   float64
 14  Virtual_Meetings_to_Work_Hours  5000 non-null   float64
dtypes: float64(2), int64(13)
```



ML models / data analysis

Random Forest to determine feature importance

	Feature	Importance
1	Hours_Worked_Per_Week	0.341580
5	Social_Isolation_Rating	0.127975
2	Work_Life_Balance_Rating	0.112315
0	Work_Location	0.085844
8	Sleep_Quality	0.081028
6	Satisfaction_with_Remote_Work	0.077960
3	Mental_Health_Condition	0.076690
4	Access_to_Mental_Health_Resources	0.050987
7	Physical_Activity	0.045621

Neural Network Model

```
Epoch 1/20
125/125 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.3326 - loss: 0.2853 - val_accuracy: 0.3260 - val_loss: -0.3737
Epoch 2/20
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - accuracy: 0.3274 - loss: 0.3841 - val_accuracy: 0.3260 - val_loss: -0.3849
Epoch 3/20
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - accuracy: 0.3259 - loss: -0.1950 - val_accuracy: 0.3260 - val_loss: -0.3970
Epoch 4/20
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - accuracy: 0.3344 - loss: -0.3647 - val_accuracy: 0.3260 - val_loss: -0.5128
Epoch 5/20
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step - accuracy: 0.3229 - loss: -0.5454 - val_accuracy: 0.3260 - val_loss: -0.6320
```

CHOSEN VARIABLES

- Physical_Activity
- Sleep_Quality
- Mental_Health_Condition

Neural Network Model

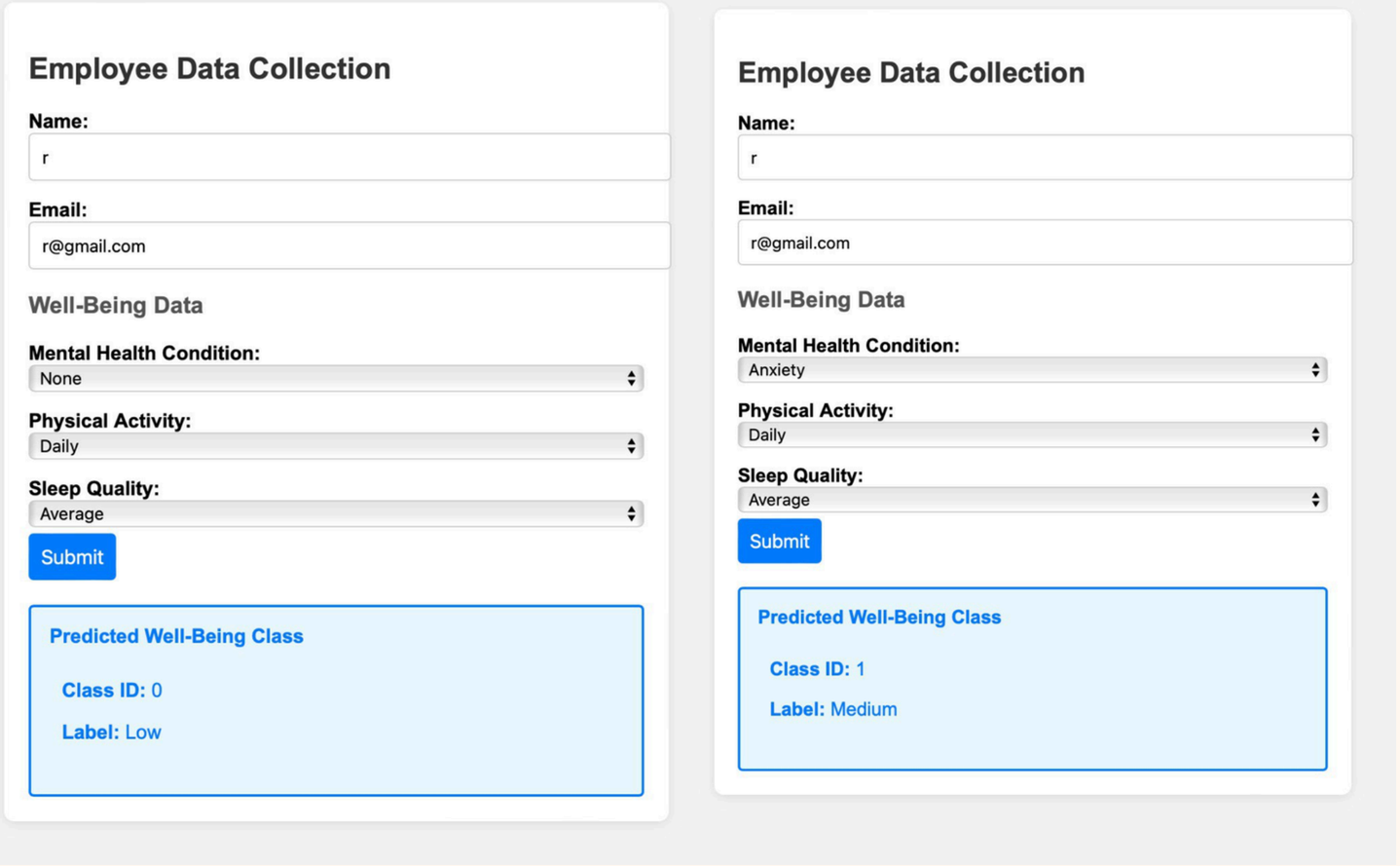
Why we chose this model

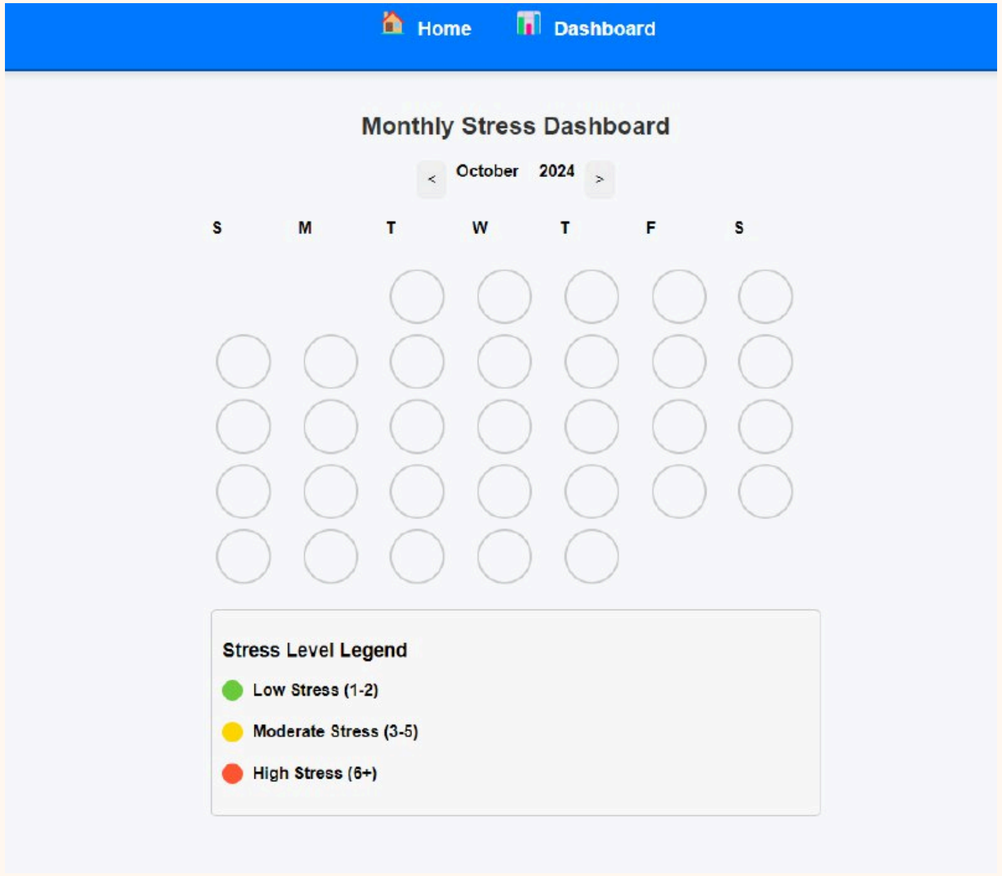
The problem involves predicting categorical outcomes (different levels of stress) based on several input features (mental health condition, physical activity, sleep quality). Therefore, it is a multiclass classification problem.

The chosen architecture (dense layers with ReLU activation) can effectively capture relationships in the input features.

The input features are categorical, and the code includes preprocessing steps such as Label Encoding and Standard Scaling. This is necessary to convert categorical variables into a numerical format that the neural network can understand.

OUR SOLUTION





THANK *You*