Analysis of Workplace Burnout

MIS 381

Team Members: Fernando Chapa, Chang Li, Monica Martinez, Biyun Yuan

Table of Contents

Project Introduction

Overview:

- Investigate factors that lead to higher burnout rates for employees based on the workplace settings and requirements provided by their employer

Goal:

- Use predictive analysis to identify burnt out employees and determine ways for employers to decrease burnout rate and retain employees over time

Exploratory Data Analysis (EDA)

Data Cleaning

- 22,752 to 18,590 values
- Removed “Employee ID” variable
- Re-named “Designation” to “Job Level” and “Resource Allocation” to “Hours Worked”

Stratified Train-test Split

- 70/30 train:test split
- Check class imbalance

```
# Create Stratified and Test Samples
y=dataset.pop('Burn Rate').to_frame()
X=dataset
```

```
# Using a 70"30 split in train:test
X_train,X_test,y_train,y_test=train_test_split(X,y,stratify=y,test_size=0.3)
```

```
# Checking the number of rows and columns
X_train.shape,y_train.shape,X_test.shape,y_test.shape
```

```
((13013, 6), (13013, 1), (5577, 6), (5577, 1))
```

Exploratory Data Analysis (EDA)

Six Predictors:

- Gender (male or female)
- Company Type (service or product)
- Work-from-home Setup Availability (yes or no)
- Job Level (intern, entry-level, associate, supervisor, or executive)
- Hours Worked (20 to 70 hours per week)
- Mental Fatigue Score (“not at all” to “extremely”)

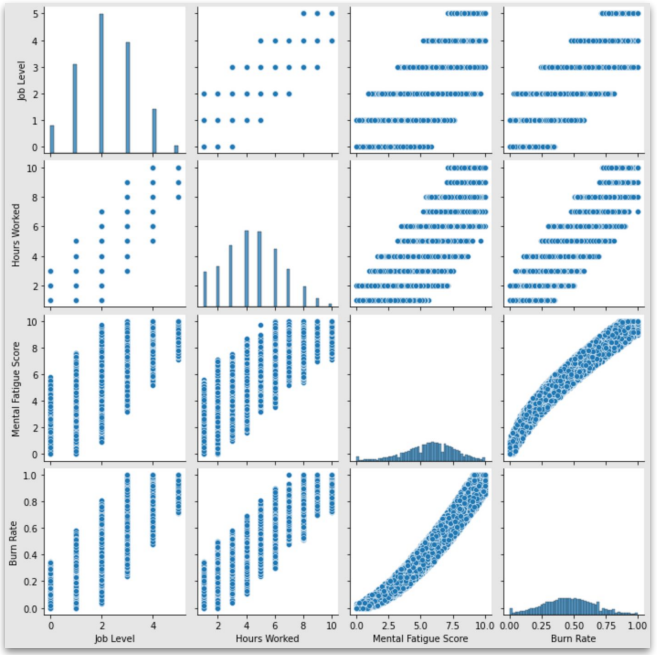
Predicted Variable:

- Employee Burnout Rate (0.0 to 1.0)

	Gender	Company Type	WFH Setup Available	Job Level	Hours Worked	Mental Fatigue Score	Burn Rate
0	Female	Service	No	2	3	3.8	0.16
1	Male	Service	Yes	1	2	5.0	0.36
2	Male	Service	Yes	1	1	2.6	0.20
3	Female	Service	No	3	7	6.9	0.52
4	Male	Product	Yes	2	4	3.6	0.29
...	...	...	...	...	...	...	...
18585	Female	Product	Yes	1	3	6.0	0.48
18586	Male	Product	No	3	7	6.2	0.54
18587	Female	Product	Yes	3	6	6.7	0.59
18588	Female	Service	No	2	5	5.9	0.52
18589	Male	Product	No	3	6	7.8	0.61

18590 rows × 7 columns

Exploratory Data Analysis (EDA)

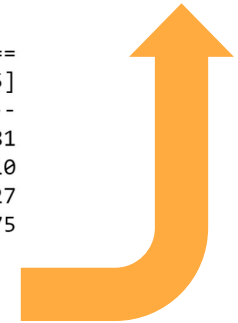


Multi Linear Regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              Burn_Rate   R-squared:                       0.920
Model:                            OLS   Adj. R-squared:                  0.920
Method:                 Least Squares   F-statistic:                 7.160e+04
Date:                Thu, 04 Aug 2022   Prob (F-statistic):               0.00
Time:                        15:54:36   Log-Likelihood:                 27262.
No. Observations:               18590   AIC:                        -5.452e+04
Df Residuals:                   18586   BIC:                        -5.448e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                 -0.0835      0.002    -54.753      0.000      -0.087      -0.081
WFH_Setup_Available       -0.0119      0.001    -13.852      0.000      -0.014      -0.010
Resource_Allocation        0.0263      0.000     78.357      0.000       0.026       0.027
Mental_Fatigue_Score       0.0740      0.000    207.783      0.000       0.073       0.075
==============================================================================
Omnibus:                      194.854   Durbin-Watson:                   1.979
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              134.814
Skew:                           0.088   Prob(JB):                     5.32e-30
Kurtosis:                       2.622   Cond. No.                         30.9
==============================================================================
```

```
result.pvalues
```

```
Intercept                 0.000000e+00
WFH_Setup_Available       2.035339e-43
Resource_Allocation       0.000000e+00
Mental_Fatigue_Score      0.000000e+00
dtype: float64
```



p-values

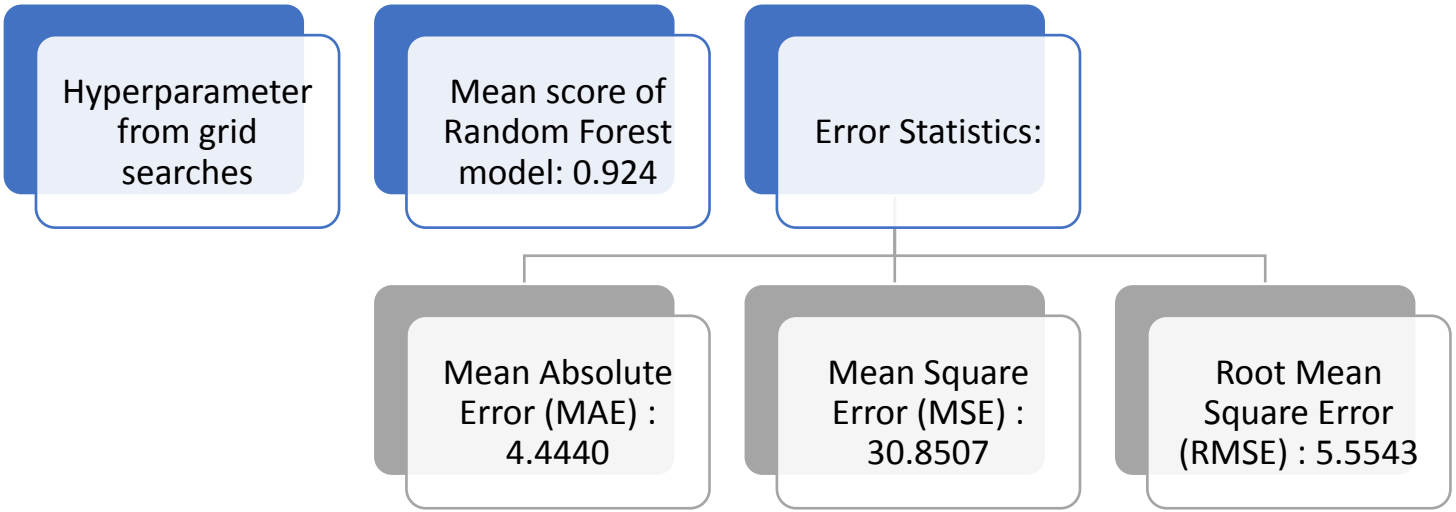
Multi Linear Regression

	Real_Values	Predicted_Values
0	0.26	0.313433
1	0.49	0.402237
2	0.49	0.474014
3	0.66	0.689059
4	0.49	0.536699
...	...	...
3713	0.49	0.492297
3714	0.63	0.588115
3715	0.25	0.258969
3716	0.92	0.854962
3717	0.63	0.673823

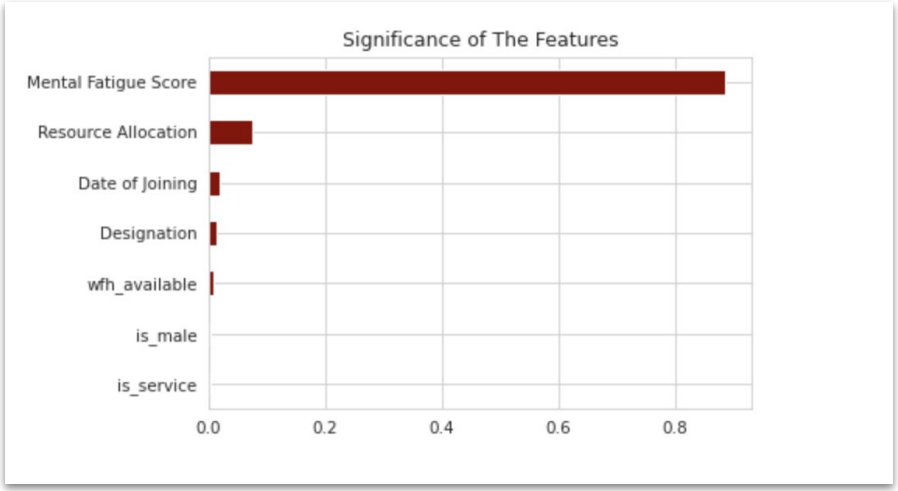
Calculating MSE

0.04632004267707683

Random Forest



Random Forest



Naive Bayes

Binned:

_binned_2	D_binned_<2	D_binned_>2	RA_binned_5	RA_binned_<5	RA_binned_>5	MFS_binned_(-0.001, 4.6]	MFS_binned_(4.6, 5.9]	MFS_binned_(5.9, 7.1]	MFS_binned_(7.1, 10.0]
1	0	0	0	1	0	1	0	0	0
0	1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	1	0	0	0
0	0	1	0	0	1	0	0	1	0
1	0	0	0	1	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...
0	1	0	0	1	0	0	0	1	0
0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	0
1	0	0	1	0	0	0	1	0	0
0	0	1	0	0	1	0	0	0	1

Gender	Company_Type	WFH_Setup_Available	Designation	Resource_Allocation	Mental_Fatigue_Score	Burn_Rate	Target	MFS_binned	D_binned
Female	Service	No	2	3	3.8	0.16	0.0	(-0.001, 4.6]	2
Male	Service	Yes	1	2	5.0	0.36	0.0	(4.6, 5.9]	<2
Male	Service	Yes	1	1	2.6	0.20	0.0	(-0.001, 4.6]	<2
Female	Service	No	3	7	6.9	0.52	1.0	(5.9, 7.1]	>2
Male	Product	Yes	2	4	3.6	0.29	0.0	(-0.001, 4.6]	2
...	...	...	...	...	...	...	...	...	...
Female	Product	Yes	1	3	6.0	0.48	0.0	(5.9, 7.1]	<2
Male	Product	No	3	7	6.2	0.54	1.0	(5.9, 7.1]	>2
Female	Product	Yes	3	6	6.7	0.59	1.0	(5.9, 7.1]	>2
Female	Service	No	2	5	5.9	0.52	1.0	(4.6, 5.9]	2
Male	Product	No	3	6	7.8	0.61	1.0	(7.1, 10.0]	>2

Concatenate new dummy columns into the old dataframe

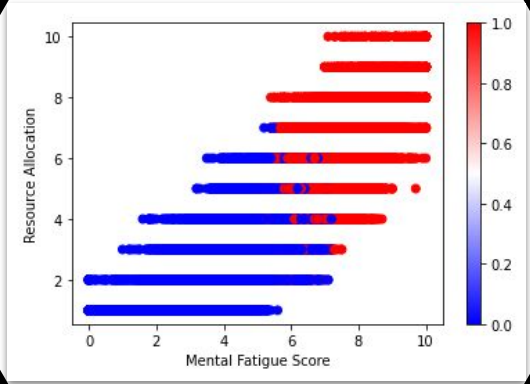
K-Nearest Neighbors

Design Matrices

	Mental_Fatigue_Score	Resource_Allocation	Burn_Rate	target
0	3.8	3	0.16	0.0
1	5.0	2	0.36	0.0
2	2.6	1	0.20	0.0
3	6.9	7	0.52	1.0
4	3.6	4	0.29	0.0
...	...	...	...	...
18585	6.0	3	0.48	0.0
18586	6.2	7	0.54	1.0
18587	6.7	6	0.59	1.0
18588	5.9	5	0.52	1.0
18589	7.8	6	0.61	1.0

18590 rows × 4 columns

Plot



Create Target: Burn Rate < 0.49 and > 0.49

- The combination of Mental Fatigue Score and Resource Allocation will influence the Burn Rate. The increase of the values of these two variables will cause the Burn Rate to be closer to 1

Solution and Insights

Most Influential Factors:

- Mental Fatigue Score and Hours Worked

Potential Solutions:

- Set a maximum hours worked per week
- Organize team events outside of the office
- Make more mental health resources available for employees
- Increase work-from-home availability


Thank You!
Q&A