Are your Employees Burnt Out?

Classification Model and Analysis

Business Case

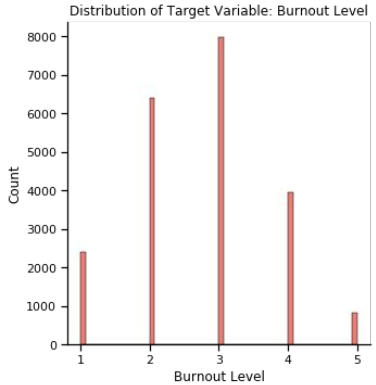
Due to the COVID 19 pandemic, Large Co. employees have been working from home since April. The company has noticed a decrease in productivity, and suspects it may be due to employee burnout. Large Co. would like to create a model that can predict employee burnout based on various factors, in order to identify and provide support to the employees most affected by the current situation.

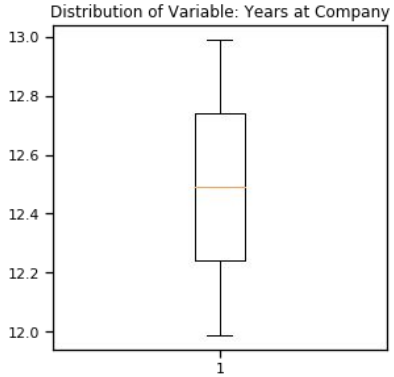
The Data

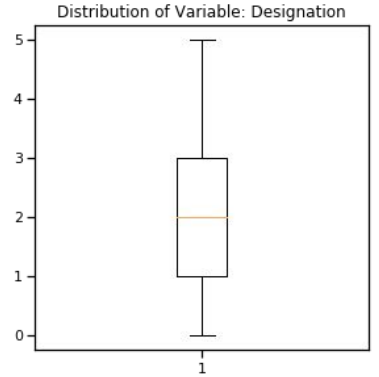
- Data was sourced from the Burn Rate Dataset from kaggle
- Labeled dataset included information on 22,750 employees and their respective burnout rates (split was 75/25 for training/testing).
- Predictors of burnout rates included: Years at Company, Gender, Company Type, Work From Home Setup Available, Designation, Resource Allocation, and Mental Fatigue Score
- Burn rates were measured on a scale of 1 - 5, with 1 being the lowest and 5 being the highest in terms of burnout

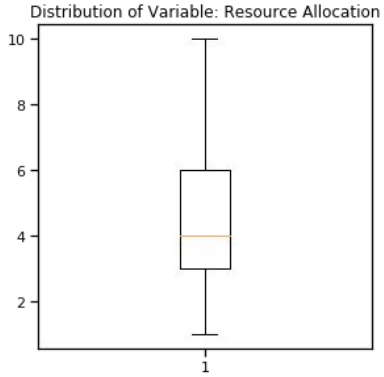
EDA

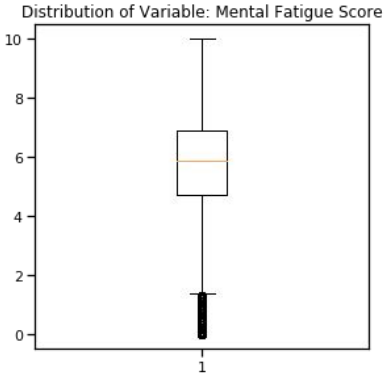
- Normally distributed target
- Lower lying outliers for mental fatigue score





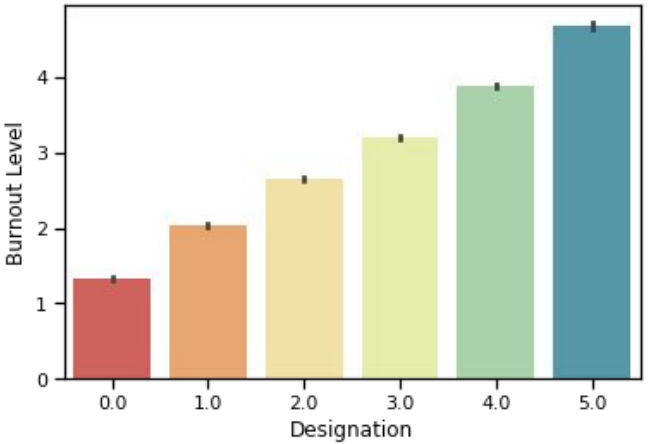


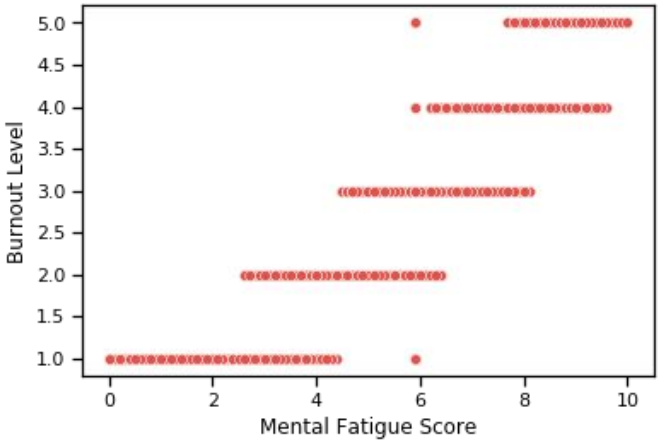


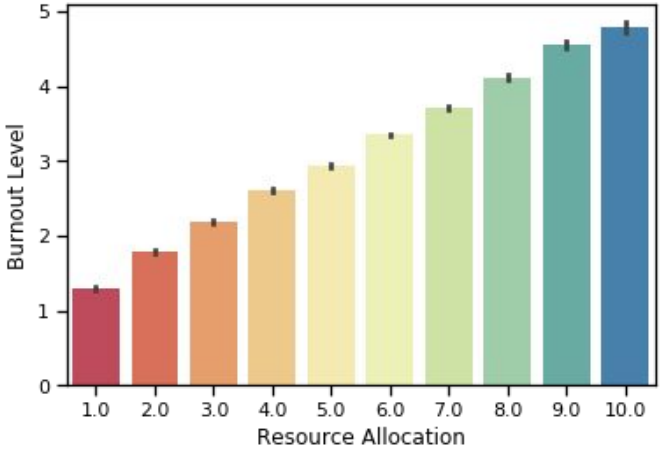


EDA

- Strong positive correlations between mental fatigue score and burnout
- Also strong correlations with designation, resource allocation and burnout



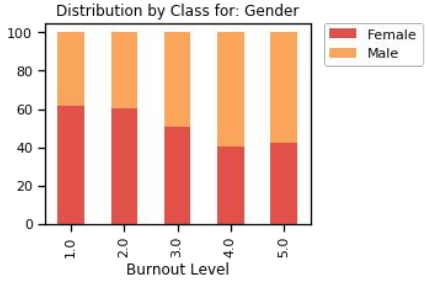


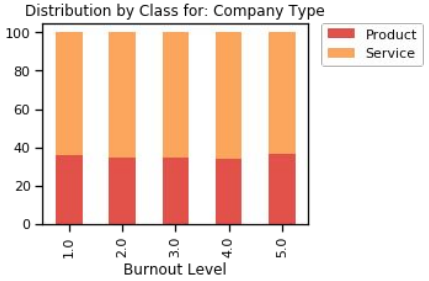


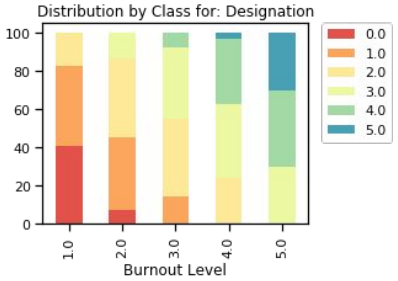
EDA

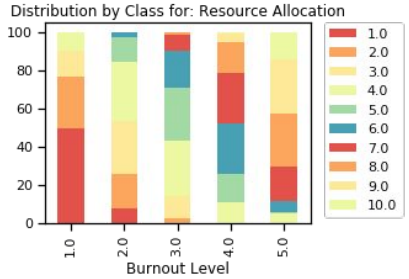
Of the highest burnout classes (level 4 and 5):

- 60% were male
- 70% did not have a WFH setup available
- 78 % had a resource allocation of over 5
- 72 % had a designation of 3 or 4



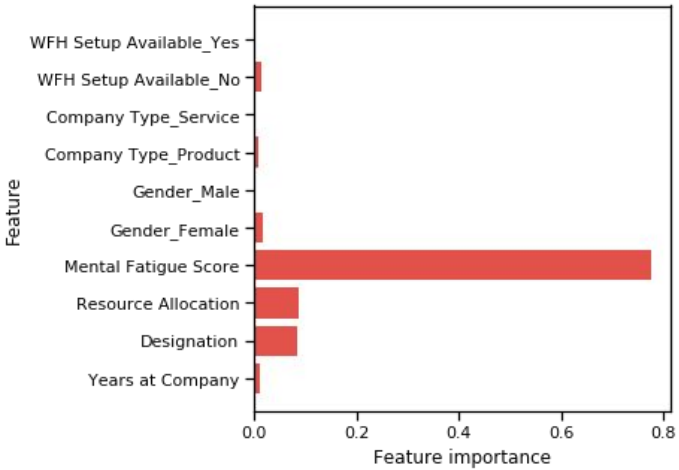






Modeling Process

- 9 vanilla models ran as baseline
- Best model was XGBoost with overall accuracy of 73% and recall of 85% for class 5
- Most important features were mental fatigue score, resource allocation and designation



	XGBoost Base				XGBoost Tuned			
	precision	recall	f1-score	support	precision	recall	f1-score	support
1.0	0.91	0.76	0.83	644.00	0.78	0.88	0.83	644.00
2.0	0.75	0.78	0.76	1551.00	0.74	0.75	0.75	1551.00
3.0	0.71	0.79	0.75	1979.00	0.76	0.67	0.71	1979.00
4.0	0.72	0.64	0.68	999.00	0.65	0.67	0.66	999.00
5.0	0.82	0.59	0.69	225.00	0.59	0.85	0.69	225.00
accuracy	0.75	0.75	0.75	0.75	0.73	0.73	0.73	0.73
macro avg	0.78	0.71	0.74	5398.00	0.70	0.76	0.73	5398.00
weighted avg	0.75	0.75	0.75	5398.00	0.73	0.73	0.73	5398.00

Predictions Preview

	Employee ID	Burnout Level (Preds)
0	fffe31003000390039003000	4.0
1	fffe31003003100370038000	2.0
2	fffe33003400380035003900	3.0
3	fffe3100370039003200	2.0
4	fffe32003600390036003700	3.0
...	...	...
12245	fffe3900310034003700	3.0
12246	fffe32003600330034003000	3.0
12247	fffe31003800340039003000	5.0
12248	fffe32003600380031003800	3.0
12249	fffe32003100390037003800	1.0

Next Steps and Recommendations

- Collect additional information on employees to use as predictors (ie performance evaluations, age, marriage/family status, etc.)
- Examine relationship between burnout, loss of productivity, and attrition
- Identify and reach out to managers of all employees who scored a 5 in terms of burnout
- Reach out to employees who scored a 4 and also had high mental fatigue score, high resource allocation and high designation

Limitations

- Labeled dataset only had employees that joined between 2008 and 2009 (No data on more recent employees)
- Limited features in dataset
- Blank (NaN) values were imputed using KNN Imputer, possibly skewing results

Thank You