Sam Loyd

Predicting Burnout

August 2021

Executive Summary

The purpose of this project was to analyze and prepare a dataset to determine if a model could be used to adequately predict burnout within a unique corporate environment. It also sought to determine if the predictive power was improved by asking a dimension such as the level of mental fatigue of each employee through a survey. Burnout adversely affects the mental and physical health of an individual. It also reduces productivity which can affect corporate profitability. This work sought to apply mitigations by staff in human relations targeted at those with the highest level of burnout. Levels of burnout were on the rise before COVID, and the pandemic has only exacerbated the problem. With all that in mind, a dataset was acquired from an open-source community popular in education and data science. It contained nine dimensions. Most of these dimensions could be obtained without querying the employee on current mental state, but one dimension called mental fatigue score was included and was likely the result of a survey. This dataset was analyzed, cleansed, and used to train a machine learning model to predict the level of burnout an employee experienced. An automation library was used to reduce coding efforts, and the dataset was split into a training and validation set before being passed to the model. The training set was further split by the library into separate training and testing sets. The model was tuned after selection. Ultimately, the final model proved successful at explaining the variance between the actual values and those predicted by the model. The secondary goal of determining the importance of the queried dimension indicated a significant boost to the model's predictive power by asking the mental fatigue score from each employee. It was twenty percent better at doing so with that dimension included. With that in mind, it is recommended to perform such a

survey, whenever possible and ethical, to increase the likelihood of success when applying mitigations to improve the burnout rate.

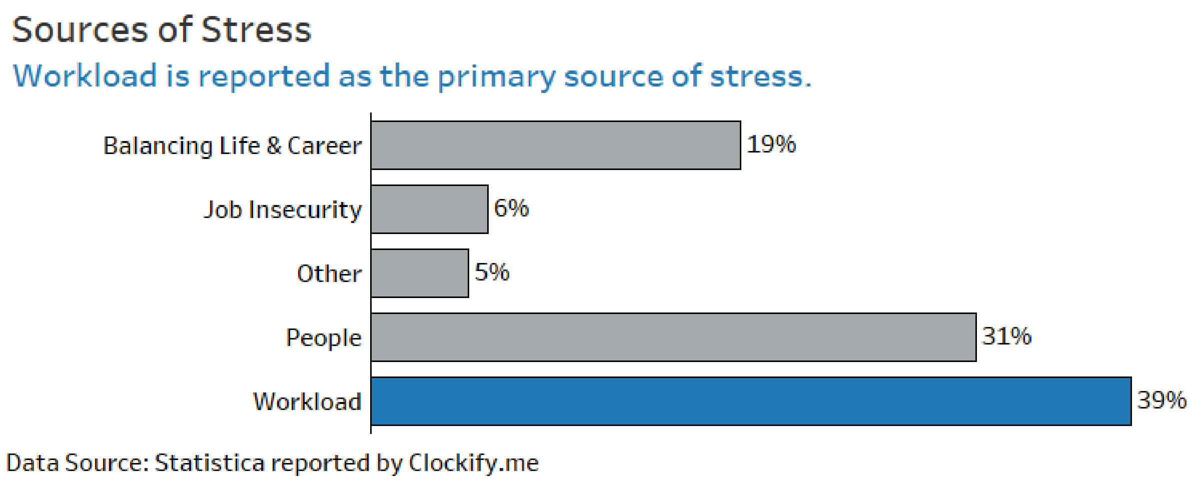
Abstract

Harvard Business Review reports that in 2019, 190 billion dollars was spent on stress induced by the workplace and that it accounted for 120,000 annual mortalities (2019). This stress often leads to employee burnout. With those staggering figures in mind, this work sought to minimize the impact of corporate burnout by targeting mitigations at highly stressed individuals by using a corporate burnout score to measure the degree. A secondary goal of determining the need of querying each employee as to the state of their mental fatigue was also desired. A comparison should be drawn between the two methods to decide if any improvements would be justified given the likely cost and possible intrusion of privacy and other ethical considerations.

Introduction

Three fourths of corporate workers report feeling burned out at some point during their careers and almost the same percentage report a negative effect on their mental health (Morrison, 2021). Spring Health states that the pandemic has increased this problem with 76 percent actively feeling increased stress from the pandemic. (2020). This is backed up by a survey from Robert Half which highlighted that 44 percent reported more stress than at the same point a year before the pandemic began (Mayer, 2021). Burnout can come from many sources and is often reported as work related stress. Figure 1 shows a surveyed report of sources of work-related stress (Kojic, 2019).

Figure 1



With all this in mind, a Kaggle dataset was acquired with dimensions suitable for creating a machine learning model to predict burnout. The dataset, which was also previously used in a competition, was provided by Paras Varshney. It contained nine dimensions with *burnout rate* as the desired target dimension. There was an *employee id, date of joining* or hire date*, gender, company type, WFH setup available* or work from home option*, designation, resource allocation, and mental fatigue score* dimension in addition to the target dimension. *Julian date* was also created from hire date as well. These were then all cleansed for model preparation.

Methods

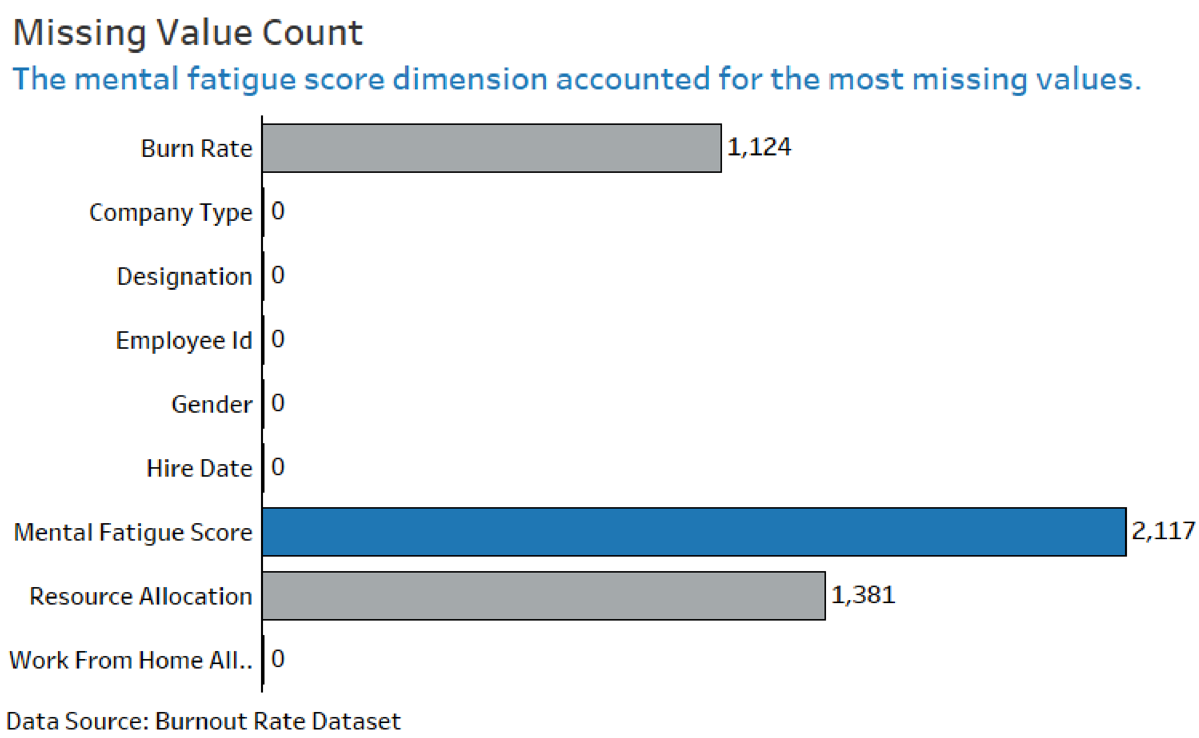
The CRISP-DM methodology was implemented for this project (Siegel, 2016). As such, six iterative stages were followed.

- Business (Domain) Understanding
- Data Understanding
- Data Preparation
- Modeling

- Evaluation
- Deployment

In the context of this paper, the first four will be covered in the current methods section. The evaluation and deployment phases of CRISP-DM can be found in the results and conclusion sections. Domain knowledge was referenced in the introduction. The initial analysis indicated that most of the quantitative dimensions followed a normal distribution. There were 1,124 instances of missingness including in the target dimension. Since those could not be used for training or modeling, any rows without a burnout rate score were removed. Figure 2 shows the total counts of missing values in dimensions affected. Mental fatigue score proved to have the most missing values.

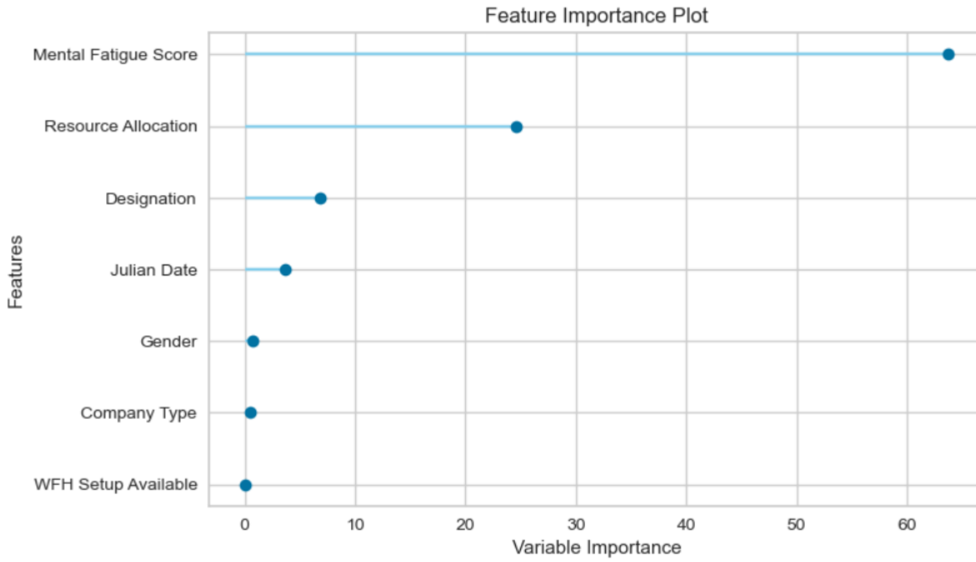
Figure 2



After removing the missing rows for the target dimension, the remaining dimensions affected included *mental fatigue score* and *resource allocation*. Missingness tended to be random between the dimensions affected (see Appendix A). As such, they were imputed or assigned a value using a nearest neighbor model predicted against the other dimensions. The *mental fatigue score* was affected by outliers on the lower end. These were replaced by assigning a value using a nearest neighbor model as well.

The data was then rescaled using a technique to reduce the impact a quantitative dimension with large range values would have over another dimension with smaller range values. Binary categories were converted to their numeric counterparts. The data was then split into a validation and training set. The training set was provided to an automation library where it was further split again into new training and test datasets. Models were compared and one was selected. The R-squared values, or explanation of variance from the predictions and the actual values, were in the upper 80 to low 90 percent ranges for the catboost and light gradient boosting machine (lightgbm) models predicting burn rate. Both methods use decisions trees where the former focuses on category and the latter uses splits based on best fit. The difference in the scores between the two was often minute. Despite catboost models tending to show insignificantly better scores, light gradient boosting machine models usually perform in around one-fifth the time. Given the markedly better processing time and close scores, light gradient boosting machine was selected for the final model. The model was tuned and validated. As indicated in the correlation analysis and top model feature selection, *mental fatigue score* proved to have the highest predictive power. Figure 3 illustrates the top feature results from the primary model.

Figure 3



The process was then repeated to address the secondary goal of creating a model without the *mental fatigue score* dimension. In this iteration, gradient boosting machine proved to have the highest explanation of variance between the predicted values from the model and the actual values in the testing and validation data sets. It was selected, tuned, and processed as well. This method uses multiple combined models often based on decision trees.

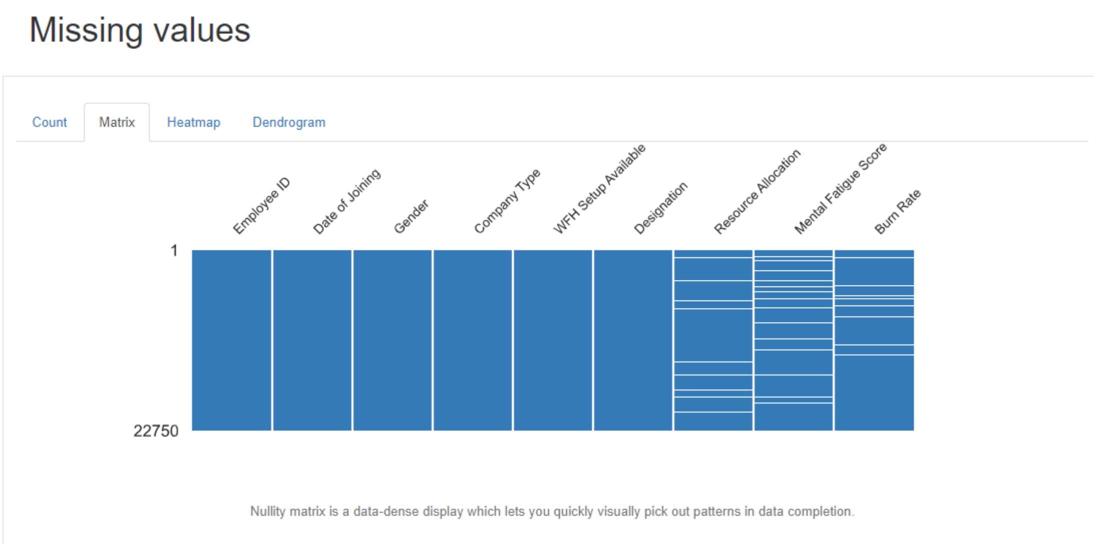
Results

The final primary model selected provided a slightly more than 90 percent goodness of fit score between the actual validation values and those predicted by the model. This is known as an R-squared value and is used to measure how well the model explains that variance. The secondary model which dropped the *mental fatigue score* dimension provided a slightly over 71 percent R-squared value. This drop was consistent with how many models listed the mental fatigue score as the top feature. Despite the drop in predictive power, the metrics prove the secondary model is usable and significant for this purpose.

Conclusions

Both models could be used to predict burnout rate in the corporate setting that the dataset was acquired from. In areas where privacy is of utmost concern or ethical constraints limit surveying for a mental fatigue dimension, it is recommended that the secondary model that did not include the *mental fatigue score* dimension be used. This comes with an understanding that these limitations would reduce model effectiveness. In an environment without those limitations, the original model with the *mental fatigue score* dimension is recommended. This comes with the extra burden of surveying the employee for that information. This should be done with the utmost care for privacy, security, and ethics. After which, mitigations such as those surveyed by Mental Health America could be offered to highly burned-out employees. These methods include meditation, courses in nutrition, exercise sessions, desktop yoga and mental wellness seminars which were all positively received by employees who participated in a recent survey (Morrison, 2021).

Appendix A



References

Kojic, M. (2019, June 7). Career burnout and its effect on health. Retrieved from https://clockify.me/blog/productivity/career-burnout

Mayer, K. (2021, June 1). Burnout is continuing to rise; is HR doing enough? Retrieved from https://hrexecutive.com/burnout-is-continuing-to-rise-is-hr-doing-enough/

Moss, J. (2019, December 11). Burnout Is About Your Workplace, Not Your People. Retrieved from https://hbr.org/2019/12/burnout-is-about-your-workplace-not-your-people

Morrison, C. (2021, May 4). 16 Employee Burnout Statistics You Can't Ignore. Retrieved from https://everyonesocial.com/blog/employee-burnout-statistics/

Siegel, E. (2016). Predictive Analytics. Hoboken, NJ: Wiley.