Reimagining the Future of Work:
Predicting Job Satisfaction of the Changing Workforce

Michael Sam, Shuang Geng, Kavie Yu

May 2, 2023

Abstract: The US workforce today faces pressing challenges such as decreasing participation and high turnover despite the substantial unmet talent demand in the job market. Much evidence has pointed to people's dissatisfaction with jobs underlying these trends. What constitutes a good job? While individuals may have different definitions, there are potentially universal criteria pointing to workplace satisfaction and overall happiness in life. Our project uses the National Study of the Changing Workforce 2016 survey, a nationally representative survey of U.S. workers that includes diverse information on essential dimensions of work and life. Using the decision tree algorithm, we identify important contributing factors to a productive and happy workforce for the future with a prediction accuracy of 0.60. Our research findings will offer valuable insights to address the most critical challenges in today's talent development, retention, and engagement at both policy and organizational levels.

Table of Contents

Problem Motivation

The US workforce is facing challenges such as unmet talent demand, decreasing participation rates, and high turnover rates due to a variety of reasons, such as the need for upskilling, concerns of workplace safety, and care responsibilities hindering work outside of the home (CNBC, 2021; Forbes, 2021; Tankersley, 2021). To understand what constitutes a good job and contributes to a productive and happy workforce for the future, this project uses the National Study of the Changing Workforce 2016 survey data. It applies ML algorithms to identify important contributing factors to workplace satisfaction. The study's results will provide perspectives for tackling the primary obstacles in talent growth, retention, and commitment in policy and organizational dimensions.

Description of Data

The National Study of the Changing Workforce survey was conducted online and via phone interviews in 2015. It focused on work and personal lives, physical and mental well-being, healthcare, and general attitudes about today's workplace. The study sample was selected from the AmeriSpeak panel and SSI (Survey Sampling International) 's non-probability opt-in panel. Recruited respondents were those at least 18 years old who worked for an employer or owned a business. The sample was selected based on age, race and ethnicity, gender, and education.

Description of Data Cleaning of Feature Engineering

Hypothesis Formation:

Our research project is inspired by the job evaluation criteria developed by Two Sigma Impact. Two Sigma develops the Good Job Score to measure job quality. This “good jobs” framework and survey tool tested over 10,000 employees from 1,000 companies cross-industry. The framework is based on four core dimensions: leadership, purpose, growth, and fairness. This assessment tool is standardized and scalable by meeting several critical criteria, including face, construct, internal, and external validity. Figure 1 shows the 12 questions that form the basis of the Good Job Score assessment. (Two Sigma Impact, 2019).

Furthermore, we want to examine the common perception that a good job is a high-paying job. The monetary lever has been the center for traditional management studies and heavily used for workplace incentivization (Judge et al., 2010). But to what degrees can it lead to satisfaction and engagement? We aim to address this question by comparing our analysis to a baseline model, which is a simple multiclass logistic regression of income on job satisfaction.

Figure 1: Good Job Score Metric

Data Cleaning Workflow:

1. Apply exclusion criteria as the first step of data cleaning

Predictor variables missing more than 20% of data, that have no dictionary (to map legend to meaning), or correlated with other predictor variables due to transformation were excluded. For predictor variables that are perfectly correlated (since it is a transformation of another variable), the variable with continuous values is always kept (instead of the categorical grouping of the continuous variable).

2. Further selected predictor variables based on our hypothesis grounded in the Two Sigma Good Job Framework

The original data set ("NSCW2016_July2018_v2019.csv") consists of 1,516 observations and 417 variables. 65 predictor variables relevant to the four dimensions of the Two Sigma Good Job Framework were selected as features.

Overall, the following features were used[1]:

[1] For the variable dictionary, see the 2016 National Study of the Changing Workforce Questionnaire documentation.

Numeric:
"RHRCPSMU","QEB44H","QEN8B","QPD1","QWF19","QSS8","C_FAMINC","QSS4","QSS8A","QPW16I","QPW16H"

Categorical:
"SEX","QEB2","QEB17","QEB22","QEB30B","QEB40","QEB44A","QEB44A2","QWC52B","QSUP1","QSUP5","QBP22B","QPW22J","QPW22A","QOC1"

Likert-type variables:
"QEB51","QWC1","QWC3","QWC9","QWC11","QWC12A","QWC13","QWC13B","QWC13C","QWC13D","QWC13E","QWC13F","QWC18D","QWC18E","QWC18F","QWC26B","QWC26A","QWC26D","QWC27","QWC26G","QWC52","QWC59E","QWC59G","QSUP9","QSUP14","QSUP15","QSUP16","QSUP17","QSUP20","QSUP8B","QBP21","QWF28","QSP13","QWF9","QWF18","QSS11F","QSS11G","QPW16C","QOC1_7"

3. Dropping missing values in the response variable

QWC38 (overall job satisfaction) was used as the response variable, and 8 observations were dropped due to missing values in QWC38.

4. Grouping response categories due to unbalanced distribution of data

When answering QWC38, “All in all, how satisfied are you with your job?”, respondents were presented with the following response options and asked to select statements that applied to them: Very satisfied (37.67%), Somewhat satisfied (45.36%), Somewhat dissatisfied (13.26%), and Very dissatisfied (3.71%). Due to the low number of responses in the “Very dissatisfied” category, the last two categories were combined to achieve a more even distribution of responses.

5. Further preprocessing: imputation and treatment of variables.

For numeric variables, we imputed the missing values using the median, and for categorical variables, we used the mode. We also reviewed each variable and reversed the encoding such that a higher number means "more" or "agree", and a lower number means "less" or "disagree". We treated the likert-type variables as continuous variables. We performed One-Hot-Encoding for the categorical variables and finally, we splitted 70% for train data and 30% for test data.

Explanation of Reasonable Alternative Machine Learning Methods

Multi-Class Logistic Regression (with LASSO regularization)

Multi-class logistic regression is suitable for our study as the project aims to classify employees into one of several categories of how they feel about job satisfaction (very satisfied, somewhat satisfied, dissatisfied) based on many potential predictor variables. Additionally, multi-class regression is computationally efficient compared to other models. However, the model can be hard to interpret with many predictor variables. Therefore, LASSO regularization is employed to reduce the number of predictors; it can also help improve the accuracy by reducing overfitting.

Single Decision Tree

A single decision tree produces a sequence of if-then rules that an average person can easily interpret. It is a suitable model for predicting employee job satisfaction based on various workplace metrics such as career growth, job fairness, and workplace leadership.

It can capture linear and non-linear relationships between the predictor variable and overall job satisfaction. This is essential because the relationship between workplace/life metrics and job satisfaction could be more complex than linear associations. Furthermore, a single decision tree can handle both categorical and numeric (continuous) predictor variables, which is ideal for this dataset containing both types of variables.

Random Forest

Unlike a single decision tree, a Random Forest utilizes multiple decision trees built on different subsets of the training data and a random subset of predictors for each tree. It is one of the popular ways to improve trees' prediction power. In the case of a powerful predictor and many moderately strong predictors, a Random Forest can de-correlate bagged trees to gain a higher reduction in variance so the committees do not come up with similar predictions.

GBM Boosting

Another popular means of improving the prediction power of trees is Boosting. Unlike Random Forests, in Boosting, each tree is built without considering other trees. Trees are formed sequentially, using information from the previous trees. GBM (Gradient Boosting Machines) may be better than Random Forest in this problem scope as it handles imbalanced data better and can help improve the classification of the minority class (which in this case is the combined category for “somewhat dissatisfied” and “very dissatisfied”) (Sun et al., 2009). Additionally, when there are complex relationships between predictor variables and the response variable, GBM boosting may perform more accurately than Random Forest (Friedman, 2001).

XG Boosting

The significant advantage of XG Boosting over GBM Boosting is that it is an upgraded version of GBM that adds regularization to the algorithm (Chen, 2016). Regularization can prevent overfitting and improve the generalizability of the model. XG Boosting can be particularly useful in high-dimensional datasets that utilize many features relating to the workplace and life.

Justification of Final Model

We ran the Machine Learning models mentioned before for this multiclass classification task, and these are the results:

Model	**Train accuracy**	**Test Accuracy**
Multiclass Logistic Regression	0.7633	0.6504
Multiclass Logistic Regression with Lasso regularization	0.7462	0.6615
Single Decision Tree	0.7008	0.5996
Random Forest	1.0000	0.6814
GBM Boosting	0.7244	0.6593
XG Boosting	0.8438	0.6438
Baseline: Multiclass Logistic Regression of income alone	*0.4527*	*0.4558*

Our findings indicate that all models display a slightly higher training accuracy compared to their testing accuracy. Although not indicative of significant overfitting, this suggests that the models perform better on the training data than on unseen data or predictions. Notably, our analysis revealed that all ensemble methods demonstrated overfitting tendencies, with the RF model exhibiting the most pronounced overfitting behavior.

While the RF model demonstrated the highest accuracy on the testing set and would be an ideal choice for predictive purposes, our goal is to prioritize interpretability and avoid black-box models. Our project's primary objective is to comprehend the underlying factors that contribute to job happiness, rather than accurately predicting job happiness levels based on the survey questions. Therefore, we have decided to opt for the single decision tree model. Moreover, it has several advantages:

- It is easy to comprehend and interpret, making it particularly useful for creating organizational policies. Additionally, decision-makers and managers without a background in ML or statistics can readily comprehend the decision rules. Moreover, decision trees offer straightforward visualizations that show how decisions are made.
- The single decision tree model also handles mixed categorical and numerical data. Additionally, being a non-parametric model, we are not required to fix any assumptions about the underlying data distribution, further enhancing its flexibility and versatility.
- It is very easy to calculate the feature importance which reveals the most critical variables for making decisions. Additionally, the model can capture feature interactions, making it flexible for modeling complex relationships within the data.

Description of Final Model (Decision Tree)

As we explained before, a huge advantage of using a decision tree is that it is interpretable. We used a decision tree with max depth = 5 (as we can see in Figure 2, there are only 5 levels).

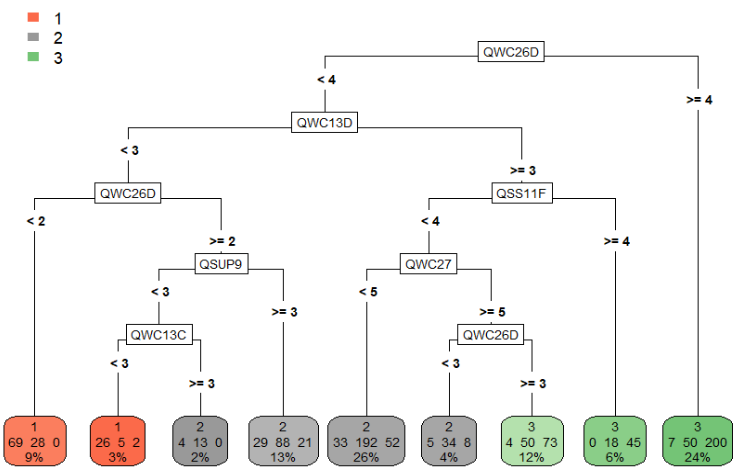

Figure 2: Decision Tree Plot

For mapping purposes, we present the legend of the variables in the model, which are, at the same time, the drivers of happiness at the job.

- **QWC26D**: Satisfied with opportunities for advancement
- **QWC13D**: Job lets me deal with feelings about work/life
- **QSUP9**: Super/manager supportive of work problems
- **QWC13C**: Job lets me create/sustain healthy relationships
- **QSS11F**: Satisfaction with earnings in main job
- **QWC27**: Feel a part of the group of people I work with

According to the model:

- A employee that is very satisfied with opportunities for advancement is very satisfied with his/her job
- An employee that is not fully satisfied with opportunities for advancement, but agrees that the job lets him/her deal with feelings about work/life and is very satisfied with earnings, is very satisfied with his/her job
- An employee that is is satisfied with opportunities for advancement, agrees that the job lets him/her deal with feelings about work/life despite not being fully satisfied with earnings in main job, is still very satisfied with his/her job
- An employee dissatisfied with opportunities for advancement and that disagrees that the job lets him/her deal with feelings about work/life, is dissatisfied with his/her job
- An employee not fully satisfied nor dissatisfied with opportunities for advancement, that disagrees that the job lets him/her deal with feelings about work/life, that has a relatively

unsupportive manager, and disagrees that the job lets him/her create healthy relationships, is not satisfied with his/her job.

Presentation of Results

As previously mentioned, we have ultimately selected the single decision tree as our final model. It achieves a training accuracy of 0.7 and a testing accuracy of 0.6, meaning that it can predict job satisfaction levels correctly for 70% of the training dataset and 60% of unseen data. Despite the fact that the model is not perfect, it remains an invaluable tool for comprehending how organizations can enhance workplace happiness by implementing policies centered around professional growth and advancement, work-life balance, earnings and compensation, supportive leadership and colleagues, and healthy relationships. Next, we present the confusion matrix for the test dataset, as well as some classification metrics like precision, recall, and F1 Score.

Actual Predicted	Dissatisfied	Somewhat satisfied	Very satisfied
Dissatisfied	**29**	17	0
Somewhat satisfied	42	**119**	44
Very satisfied	8	70	**123**

- **Class "Dissatisfied"**: Precision=0.630, Recall=0.367, F1=0.464
- **Class "Somewhat satisfied"**: Precision=0.580, Recall=0.578, F1=0.579
- **Class "Very satisfied"**: Precision=0.612, Recall=0.737, F1=0.668
- **Weighted Metrics:** Precision=0.601, Recall=0.600, F1=0.592

As a reminder, precision measures how often the model correctly predicts the proportion of true positives among all predicted positives; recall measures the proportion of true positives among all actual positives, and F1 score is the harmonic mean of precision and recall which provides a balanced measure of both metrics.

Finally, this is the result of the feature importance, where we show the 8 most relevant variables:

Feature	Importance
QWC26D: Satisfied with opportunities for advancement	173.17
QWC27: Feel a part of the group of people I work with	150.42
QWC13D: Job lets me deal with feelings about work/life	109.09
QWC18E: Can be myself on my job	103.38
QWC13C: Job lets me create/sustain healthy relationships	71.41
QWC26B: I can trust what managers say	40.39
QWC13: Job lets me use my skills/abilities	34.77
QSS11F: Satisfaction with earnings in main job	22.80

Discussion of Results

The model exhibits higher accuracy on the training dataset compared to the test dataset, indicating better performance on known data than on unseen data. With an accuracy of 60% on the test dataset, the model can correctly predict the categories of 60% of new observations (meaning that its performance is considerably better than a random model). The similarity of the weighted precision, recall, and F1 scores is a positive sign, suggesting a balanced or nearly balanced dataset across the three classes. Breaking down these numbers, and analyzing the precision and recall scores for each class, we observe similar precision values across all three classes, which is also a positive sign. However, the recall for the "dissatisfied" class is significantly lower than that for the "very satisfied" class which suggests potential biases towards different classes regarding this metric.

This indicates that when the model predicts that an employee is dissatisfied, it is less likely to be correct than when it predicts that an employee is somewhat satisfied or very satisfied. In other words, the model has a harder time identifying dissatisfied employees, which could be a problem in addressing their concerns. On the other hand, the "Very satisfied" class has a high precision and recall which could lead to a bias in favor of satisfied employees.

As previously mentioned, the primary objective of the problem is to identify factors that contribute to employee happiness at work, rather than predicting individual happiness levels. While tuning the model to achieve a more balanced recall is a potential solution, it is important to address bias in the model by considering this analysis in the implementation of the policy. Given that it is easier to identify factors associated with happiness, more attention should be given to analyzing factors associated with unhappiness, which may be more challenging to predict accurately. By focusing on these factors, we may gain additional insights into the underlying causes of unhappiness at work and identify potential areas for improvement.

Contribution

Benchmarking to Hypothesis

An increasing amount of evidence has suggested that companies that focus on job quality are more successful on business performance (Bach, Kalloch, & Ton, 2019; Two Sigma Impact, 2023). The most systematic metric to measure job quality so far is the newly released Two Sigma Impact's Good Job Score, which consists of four dimensions of workplace excellence. Among the eight most important features of our final model, three features are connected to 3 out of the 4 dimensions: "Satisfied with opportunities for advancement" (QWC26D) reflects the growth aspect of a good job; "I can trust what managers say" (QWC26B) shows the importance of leadership support; "Satisfaction with earnings in main job" highlights the pivotal role of fairness in workplace.

While our results are largely consistent with Two Sigma's metric, new themes also emerge. Two features, "Feel a part of the group of people I work with" (QWC27) and "Job lets me create/sustain healthy relationships" (QWC13C) reveal the importance of community, highlighting the social facet of Workplace. In addition, three features, "Job lets me deal with feelings about work/life" (QWC13D), "Can be myself on my job" (QWC18E), and "Job lets me use my skills/abilities" (QWC13), points to the notion of self-worth. People don't want to be viewed only as a role for functional fulfillment but as unique individuals with distinct characters, skills, emotions, and ultimately humanity.

Comparing our findings to the baseline model that only considers the income factor, we've seen a 14% increase in prediction accuracy. This is consistent with modern management studies that develop larger focuses on other aspects at the workplace in worker's wellbeing and engagement. Overall, our research provides compelling evidence that increasing job satisfaction is a compound challenge and many factors play into it. Income is important but limiting, and other social and managerial layers should also be considered. The dominant framework for workplace satisfaction identifies four criteria (leadership, purpose, growth, and fairness), and we recommend the inclusion of two additional levers: community and self-worth. These two levers are arguably embedded within the existing ones but we believe are worth singling out.

Applying to Other Contexts

While applying our model to today's workplace, we can expect several changing forces that can affect its generalizability. One is the change in work mode post Covid that is expected to place a larger emphasis on hybrid collaboration, online community, and flexibility. The 2016 survey included questions on these factors but they have not been significant at that time. Another force to account for is the variability across industries and demographics which may entail very different work styles and expectations. Our model is generalizable across different contexts and can be adapted for more granular analysis.

Recommendation for Implementation

Our model can guide policymaking on both national and organizational levels about best strategies for talent management. Nationally, a growing community of flexible workers has emerged at an accelerated rate since Covid that leads to decrease in participation in traditional full-time work. The advancement in automation has also disrupted industries that are more labor-intensive and low-paid such as manufacturing and food services and face significant labor shortage. What are the best ways to incentivize participation, engagement, and retention? Our finding suggests that monetary incentives may not be the most effective (or objectively achievable) strategy. Employers should allocate more resources towards investment on the “soft” side, such as community building (establishment of communities of practice and learning groups), leadership development (increasing the mentoring role of managers), creation of clearer career advancement paths, and providing flexible policies and environments for employees to be seen and respected. Workplace needs to surpass the single purpose of business performances, but become a place for workers to strive professionally, socially, and personally.

References

Bach, K., Kalloch, S., & Ton, Z. (2019). The financial case for good retail jobs. Harvard Business Review, 26(June). https://hbr.org/2019/06/the-financial-case-for-good-retail-jobs

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

CNBC. (2021, August 9). The great resignation is here, and experts say 'the way we work has changed forever'. CNBC. https://www.cnbc.com/2021/08/09/the-great-resignation-is-here-and-experts-say-the-way-we-work-has-changed-forever.html

Forbes. (2021, August 27). The State Of The U.S. Job Market: What You Need To Know. Forbes. https://www.forbes.com/sites/advisor/2021/08/27/the-state-of-the-us-job-market-what-you-need-to-know/?sh=5b6f51c311af

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of statistics, 29(5), 1189-1232. https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-Function-Approximation--A-Gradient-Boosting-Machine/10.1214/aos/1013203451.full

Judge, T. A., Piccolo, R. F., Podsakoff, N. P., Shaw, J. C., & Rich, B. L. (2010). The relationship between pay and job satisfaction: A meta-analysis of the literature. *Journal of vocational behavior*, 77(2), 157-167.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence, 23(04), 687-719. https://www.worldscientific.com/doi/abs/10.1142/S0218001409007326

Tankersley, J. (2021, May 7). Why the U.S. is Running Out of Workers. The New York Times. https://www.nytimes.com/2021/05/07/business/economy/jobs-report.html

Two Sigma Impact. (2019, October 28). Introducing Two Sigma Impact's Good Job Score. Retrieved from https://www.twosigma.com/articles/introducing-two-sigma-impacts-good-job-score

Two Sigma Impact. (2023). Introducing the Good Job Score Assessment Tool. Retrieved from twosigma.com/wp-content/uploads/2023/04/Good-Job-Score-Paper-April-2023-1.pdf