**Engagement at Work**

MJ Shim

Department of Organizational Leadership, Policy, and Development

PSY8712: Data Science

Professor. Richard Landers

**Introduction**

This project examines work engagement among Korean employees using secondary data obtained from the Occupational Safety and Health Research Institute (OSHRI). The questionnaire originates from the European Working Conditions Surveys. Specifically, the data utilized in this project is derived from the 6th panel study conducted at the research institute. The 6th Korea Working Conditions Survey (KWCS) conducted between 2020 and 2021 collected responses from individuals aged 15 years and above. Although demographic information is not available in this panel data, it includes valuable data concerning job characteristics and workplace environments. Through the analysis of these variables, I aim to explore the factors that influence work engagement.

**Hypotheses**

Research Question: What factors influence work engagement?

Hypothesis 1 (H1): Workers who work remotely exhibit higher levels of work engagement compared to those working in an office setting.

Hypothesis 2 (H2): There exists a positive relationship between work engagement and an individual's sleep quality.

Hypothesis 3 (H3): When predicting work engagement, non-linear statistical models such as the Random Forest model or XGBoost model are superior to linear models like OLS regression model or Elastic Net model.

**Method**

**Open Science Materials**

The following link directs you to MyBinder, where readers can replicate my R environment and reproduce the analysis results based on the libraries and RStudio version utilized in this project: <https://mybinder.org/v2/gh/mjshim0220/psy8712-final/HEAD?urlpath=rstudio>

Furthermore, all materials used in this project are available on the GitHub repository linked below. The front page of this link contains a README file detailing how to utilize the git repository for this project: <https://github.com/mjshim0220/psy8712-final>

**Participants**

Out of the 50,538 participants in the 6th KWCS, 29,549 were utilized in this research after filtering out those who did not respond to the questions or lacked sufficient information to answer.

**Measures**

|  |  |  |
| --- | --- | --- |
| Variable | Item | Scale |
| Remote work | Where did you usually work last week | 1. Home 2. Designated place other than place of work or office |
| Contract Type | What type of your working hours is? | 1. Full-time work 2. Part-time work |
| Side Work | Do you have one, or more than one job or business? | 1. One job or business 2. Two job or business 3. More than two job or business |
| Engagement | At work, I feel full of energy | Likert 5-point scale (1:Always ~ 5:Never) |
| I am enthusiastic about my job |
| Time flies when I am working |
| Sleep Quality | Difficulty falling asleep | Likert 5-point scale (1:Daily ~ 5:Never) |
| Waking up repeatedly during the sleep |
| Waking up with a feeling of exhaustion and fatigue |
| Support from Colleague | Your colleagues or peers help and support you | Likert 5-point scale (1:Always ~ 5:Never) |
| Support from Manager | Your manager helps and support you |
| Perceived Justice | You are treated fairly at your workplace |

**Procedure**

Firstly, I conducted descriptive statistics using the variables employed in this project. Additionally, I created a histogram of the dependent variable (engagement level) to observe its distribution. Next, I performed a T-test to assess Hypothesis 1. Subsequently, I conducted correlation analysis and plotted the relationship between sleep quality and work engagement to test Hypothesis 2. Finally, I ran four analytical models to determine the most effective statistical approach for predicting work engagement.

**Analysis**

**Descriptive Statistics and Static Visualizations**

Before conducting the hypothesis testing, descriptive statistics were utilized to ascertain the mean and standard deviation of both dependent and independent variables. Furthermore, the distribution of the dependent variable, work engagement, was examined.

**Table 1.**

*Mean and Standard Deviation of Independent and Dependent Variables*

|  |  |  |
| --- | --- | --- |
|  | **Mean** | **SD** |
| Peer Support | 3.561 | 0.812 |
| Manager Support | 3.552 | 0.849 |
| Perceived Justice | 3.520 | 0.825 |
| Sleep Quality | 4.327 | 0.774 |
| Satisfaction | 2.090 | 0.492 |
| Engagement | 2.433 | 0.666 |

**Figure 1.**

*Histogram of the Work Engagement*

A graph of a bar graph

Description automatically generated

**Interactive Visualization**

Various types of work environments and conditions are present in the survey, including work mode (remote or office), work contracts, and side jobs. Among them, I selected these three variables to analyze the distribution of work engagement. To share the results, I used shinyspps.io (Link: <https://mjshim0220.shinyapps.io/shiny/>). This interactive approach allows readers to explore specific segments of the overall work engagement distribution. Two chart options are available to visualize the distribution of work engagement: histogram or box plot.

**Data Cleansing**

In this project, I excluded responses that included "8. No opinion" or "9. Refused" for any of the variables used in this study. Additionally, to create an engagement and sleep quality variable, I calculated the overall score of relevant items. Lastly, since engagement, sleep quality, support from colleagues, and support from managers were measured on a Likert 5-point scale, with 1 indicating "Always" and 5 indicating "Never," I conducted reverse coding for these four variables.

**Analysis**

Related to Hypothesis 1, I first created a box plot (see Figure 2) to visualize the distribution of engagement. Subsequently, I conducted a t-test to compare engagement levels between different work modes. The t-test yielded a result of , suggesting a significant difference in work engagement between remote and office workers. With the mean work engagement of remote workers at 3.444 and office workers at 3.567, on average, remote workers exhibit significantly lower work engagement than office workers, not supporting H1.

**Figure2.**

*Comparison of Work Engagement Between Remote and Office Workers*

A graph with a diagram

Description automatically generated with medium confidence

H2. To investigate the connection between sleep quality and work engagement, I visualized their relationship. As shown in Figure 3, the fitted line on the scatter plot suggests a negative association between sleep quality and engagement. Further correlation analysis using the Pearson method revealed a significant negative correlation between sleep quality and work engagement (). Consequently, this project does not support H2.

**Figure 3.**

*A graph with a blue line

Description automatically generatedThe Relationship Between Sleep Quality And Engagement*

Lastly, I compared the R-square values for predicting work engagement levels using peer support, manager's support, sleep quality, and perceived justice across four models, including a machine learning approach. The results (see Table 2) indicate that all four models exhibit poor fit for the data in predicting work engagement, with R-square values below Notably, the random forest model demonstrates the highest R-square value in both cross-validated and holdout sets. However, when considering the disparity with other models, the margin is not significantly greater. Additionally, given concerns about overfitting, the random forest model may not be the optimal choice in this scenario; instead, the OLS model or Elastic Net model might prove more effective. Therefore, H3, which posits that non-linear models are superior to linear models for predicting work engagement, is not supported.

**Table 2.**

*Comparison of R-Square Values Between Cross-Validated and Holdout Sets of Algorithms*

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **cv\_rsq** | **ho\_rsq** |
| OLS regression | 0.18 | 0.17 |
| Elastic net | 0.18 | 0.17 |
| Random forest | 0.182 | 0.177 |
| eXtreme Gradient Boosting | 0.167 | 0.166 |

**Reflection**

Throughout this course, I aim to deepen my understanding of data manipulation and develop a data analysis mindset. Despite having conducted projects and research, I often find myself investing considerable time in reproducing previous work. Learning about the Binder function and GitHub enables me to efficiently share my research and projects with others. Moreover, through this course, I havee gained the ability to independently create code, rather than relying on copying and pasting from online searches. Finally, exploring new techniques such as web scraping and machine learning has expanded the scope of research possibilities within my current areas of interest.