Submitted: 18 July, 2024 Accepted: 12 September, 2024 Published: 30 September, 2025 DOI:10.22514/jomh.2025.115

JOMH
Journal of Men's Health

ORIGINAL RESEARCH

Entity embeddings, attention mechanisms and ordinal regression for predicting life satisfaction: a study of young south Korean male workers

Haewon Byeon[1,2,*]

[1]Workcare Digital Health Lab, Department of Employment Service Policy, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea
[2]Department of Convergence, Korea University of Technology and Education, 31253 Cheonan, Republic of Korea

***Correspondence**
bhwpuma@naver.com
(Haewon Byeon)

Abstract

Background: Young male workers in South Korea face unique challenges that can impact their subjective life satisfaction. This study aimed to develop and evaluate a novel machine learning model to predict their life satisfaction. **Methods**: We integrated Entity Embeddings with Attention mechanisms into an ordinal logistic regression framework. Entity Embeddings transformed categorical variables into dense vectors, while the Attention mechanism prioritized the most influential features. Model performance was compared against Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN) using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). **Results**: The Entity Embeddings with Attention model demonstrated superior predictive accuracy compared to all baseline models across all evaluation metrics. Ordinal logistic regression facilitated model interpretability, revealing key predictors of life satisfaction, including monthly salary, job satisfaction, company size, weekly working hours, and educational background. **Conclusions**: This study provides valuable insights for policymakers and employers to enhance the well-being of young male workers in South Korea. The proposed model offers a robust and interpretable approach for predicting subjective life satisfaction, enabling targeted interventions to address the specific needs and challenges of this demographic.

Keywords

Entity embeddings; Attention mechanism; Ordinal regression; Subjective life satisfaction; Predictive modeling

1. Introduction

Life satisfaction refers to an individual's overall assessment of their life as a whole, encompassing various domains such as physical health, mental well-being, relationships, and personal achievements [1]. Subjective life satisfaction is a critical indicator of individual well-being and quality of life, often measured on a 5-point Likert scale. For young male workers, understanding the factors that influence their life satisfaction is essential for developing effective policies and interventions aimed at improving their overall well-being. Previous research has extensively explored various factors influencing life satisfaction and turnover, leveraging traditional regression models and machine learning techniques. Studies have shown that factors such as job security, salary, work-life balance, and organizational culture significantly impact life satisfaction and turnover intentions. For instance, Doede (2017) [1] investigated race as a predictor of life satisfaction and turnover among nurses, while Lazzari *et al.* [2] (2022) examined employee turnover intentions using various predictive models. These studies have primarily utilized logistic regression, factorization machines, and ensemble methods like Random Forest and Gradient Boosting to model the relationships between different variables and life satisfaction or turnover [3, 4].

While regression models, such as logistic regression and linear regression, are widely used due to their interpretability and simplicity, they face significant limitations when dealing with high-dimensional data, especially when the data includes categorical variables with many levels. These models often require extensive feature engineering to handle interactions between variables and to transform categorical variables into numerical formats. Moreover, regression models may struggle to capture complex, non-linear relationships inherent in high-dimensional datasets, leading to suboptimal predictive performance [5, 6].

Entity Embeddings with Attention mechanisms offer a powerful alternative to traditional regression models in handling high-dimensional data with numerous categorical variables. Entity Embeddings transform categorical variables into dense, continuous vectors, enabling the model to learn meaningful representations of these features. This approach reduces the need for extensive manual feature engineering and allows the model to capture intricate relationships between variables. Also, The Attention mechanism further enhances the model's capability by dynamically focusing on the most relevant parts of the input data. This is particularly advantageous when dealing with large datasets with many features, as it allows the model to weigh the importance of different features and interactions, leading to more accurate and interpretable predictions [7, 8]. The combined use of Entity Embeddings and Attention mechanisms can significantly improve the model's ability to predict subjective life satisfaction by capturing both the global structure and local nuances of the data [9–12].

This study aimed to predict the subjective life satisfaction of young male workers in South Korea using Entity Embeddings with Attention mechanisms. Additionally, to enhance the interpretability of the developed model, ordinal logistic regression analysis was applied to understand the relationship between predictors and the ordinal outcome of subjective life satisfaction. The primary objectives were to develop a predictive model that leveraged Entity Embeddings and Attention mechanisms to handle high-dimensional, categorical data effectively, to compare the performance of this model against established reference models, including Random Forest, Gradient Boosting, and K-Nearest Neighbors, and to conduct a variable importance analysis to identify key predictors of subjective life satisfaction among young male workers. The remainder of this paper is organized as follows: Section II reviews related works. Section III describes the details of the proposed model. Section IV presents the experimental analysis. Finally, Section V concludes the paper.

2. Related work

Traditional regression models, such as logistic regression and linear regression, have been widely used to predict various outcomes, including life satisfaction and turnover intentions. Logistic regression is particularly popular for binary classification tasks due to its interpretability and simplicity. For instance, Doede (2017) [1] used logistic regression to study race as a predictor of life satisfaction and turnover among nurses. Similarly, linear regression has been employed to model the relationship between continuous dependent variables and multiple independent variables. However, these models face significant limitations when dealing with high-dimensional data, especially with numerous categorical variables. They often require extensive feature engineering to transform categorical variables into numerical formats and to create interaction terms, which can be labor-intensive and prone to human error. Additionally, traditional regression models may struggle to capture complex, non-linear relationships, leading to suboptimal predictive performance [2].

Ensemble learning methods, such as Random Forest and Gradient Boosting, have been introduced to overcome some of the limitations of traditional regression models. Random Forest, an ensemble method that constructs multiple decision trees, aggregates their predictions to improve accuracy and robustness. This method has been widely applied in various domains, including employee turnover prediction. For example, Zhao *et al.* [4] (2019) utilized Random Forest to predict employee turnover with machine learning, demonstrating its effectiveness in handling large datasets with multiple features. Gradient Boosting, another powerful ensemble method, builds an ensemble of weak learners, typically decision trees, to iteratively minimize prediction error. This method has been shown to outperform traditional regression models in many predictive tasks, including life satisfaction and turnover prediction. Skelton *et al.* [6] (2020) used Gradient Boosting to predict manufacturing employee turnover intentions, highlighting its ability to capture complex relationships between variables.

Deep learning has revolutionized the field of predictive modeling by providing advanced capabilities for capturing complex, non-linear relationships in high-dimensional data. Neural network-based approaches, such as Deep Neural Networks (DNNs), have been successfully applied in various domains, including image recognition, natural language processing, and recommendation systems. However, traditional DNNs are typically suited for dense, numerical inputs and may not be directly applicable to high-dimensional, sparse datasets commonly found in Click-Through Rate (CTR) prediction and employee satisfaction studies.

To address this issue, Factorization Machine-supported Neural Networks (FNN) have been proposed. FNN combines an embedding layer pretrained by Factorization Machines (FM) to convert sparse features into a low-dimensional dense space with a DNN component to capture high-order interactions. This approach has shown promise in handling high-dimensional categorical data and improving predictive performance [13].

D. Entity Embeddings and Attention Mechanisms Entity Embeddings with Attention mechanisms offer a novel approach to handling high-dimensional data with numerous categorical variables. Entity Embeddings transform categorical variables into dense, continuous vectors, enabling the model to learn meaningful representations of these features. This reduces the need for extensive manual feature engineering and allows the model to capture intricate relationships between variables. The Attention mechanism further enhances the model's capability by dynamically focusing on the most relevant parts of the input data. This is particularly advantageous when dealing with large datasets with many features, as it allows the model to weigh the importance of different features and interactions, leading to more accurate and interpretable predictions [14].

Recent studies have demonstrated the effectiveness of Entity Embeddings and Attention mechanisms in various applications. For instance, Rezaei and Houmansadr (2021) [8] utilized Entity Embeddings with Attention to improve the accuracy of network flow fingerprinting, while Umuroglu *et al.* [10] (2017) applied these techniques to enhance the performance of binarized neural network inference. These studies highlight the potential of Entity Embeddings and Attention mechanisms to address the limitations of traditional regression models and improve predictive performance in high-dimensional, sparse datasets.

Entity Embeddings with Attention mechanisms offer a novel approach to handling high-dimensional data with numerous categorical variables. Entity Embeddings transform categorical variables into dense, continuous vectors, enabling the model to learn meaningful representations of these features. This reduces the need for extensive manual feature engineering and allows the model to capture intricate relationships between variables. The Attention mechanism further enhances the model's capability by dynamically focusing on the most relevant parts of the input data. This is particularly advantageous when dealing with large datasets with many features, as it allows the model to weigh the importance of different features and interactions, leading to more accurate and interpretable predictions.

Recent studies have demonstrated the effectiveness of Entity Embeddings and Attention mechanisms in various applications. For instance, Rezaei and Houmansadr (2021) [8] utilized Entity Embeddings with Attention to improve the accuracy of network flow fingerprinting, while Umuroglu *et al.* [10] (2017) applied these techniques to enhance the performance of binarized neural network inference. These studies highlight the potential of Entity Embeddings and Attention mechanisms to address the limitations of traditional regression models and improve predictive performance in high-dimensional, sparse datasets.

In summary, while traditional regression models and ensemble learning methods have been widely used for predicting life satisfaction and turnover, they face significant limitations when dealing with high-dimensional data with numerous categorical variables. Neural network-based approaches, particularly those leveraging Entity Embeddings and Attention mechanisms, offer a powerful alternative by effectively handling high-dimensional data and capturing complex relationships between variables. This study aims to build on these advancements by applying Entity Embeddings with Attention mechanisms to predict subjective life satisfaction among young male workers in South Korea, comparing its performance against established reference models such as Random Forest, Gradient Boosting, and K-Nearest Neighbors.

3. Proposed method

3.1 Proposed model: entity embeddings with attention

The primary model employed in this study is based on Entity Embeddings with Attention mechanisms. This approach is well-suited for handling high-dimensional data with numerous categorical variables.

1. Entity Embeddings: Categorical variables are transformed into dense vectors using Entity Embeddings. This step reduces the need for extensive manual feature engineering and allows the model to capture intricate relationships between variables. Formally, let (x_i) be a categorical variable with (k) categories. The Entity Embedding layer maps (x_i) to a dense vector ($e_i \in R^d$), where (d) is the embedding dimension.

$$[e_i = Embedding(x_i)]$$

2. Attention Mechanism: The Attention mechanism dynamically focuses on the most relevant parts of the input data. This is achieved by assigning attention weights to each feature, allowing the model to weigh the importance of different features and interactions. Given a set of feature embeddings ($e_1, e_2, \ldots, e_n$), the Attention mechanism computes a context vector (c) as follows:

$$\left[\alpha_i = \frac{\exp(u^{Ttanh}(Wei + b))}{\sum j = 1^n \exp(u^{Ttanh}(We_j + b))}\right]$$

$$\left[c = \sum_{i=1}^{n} \alpha_i e_i\right]$$

Here, (W) and (u) are learnable parameters, (α_i) are the attention weights, and (c) is the context vector that summarizes the important features.

The architecture of the proposed model consists of the following components:

- Input Layer: Takes the preprocessed data as input.
- Embedding Layer: Converts categorical variables into dense vectors using Entity Embeddings.
- Attention Layer: Applies attention weights to focus on relevant features.
- Dense Layers: Captures high-order interactions between variables through multiple fully connected layers.
- Output Layer: Predicts the subjective life satisfaction on a 5-point Likert scale.

The final prediction ($\hat{y}$) is computed as:

$$[\hat{y} = \sigma\left(W_o h + b_o\right)]$$

Where (h) is the output of the final dense layer, (W_o) and (b_o) are the weights and bias of the output layer, and (σ) is the sigmoid activation function.

3.2 Variable importance analysis

To identify key predictors of subjective life satisfaction, a variable importance analysis was conducted. For Random Forest and Gradient Boosting models, the importance of each variable was assessed based on how much each variable reduces the impurity (*e.g.*, Gini impurity or entropy) in the decision trees. For the Entity Embeddings with Attention model, attention weights were analyzed to determine which features the model focuses on the most. This analysis provides insights into the factors that significantly influence the subjective life satisfaction of young male workers.

3.3 Regression analysis

In addition to the advanced machine learning models, ordinal logistic regression analysis was conducted to understand the relationship between predictors and the ordinal outcome of subjective life satisfaction. Ordinal regression is suitable for modeling outcomes that have a natural order but unknown distances between categories, such as the 5-point Likert scale used in this study. This analysis enhances the interpretability of the model by providing odds ratios for each predictor, indicating how changes in predictor variables affect the likelihood of higher subjective life satisfaction.

The ordinal regression model is formulated as follows:

$$\left[\log\left(\frac{P\left(Y\leq j\right)}{P\left(Y>j\right)}\right)=\alpha_j-X\beta\right]$$

Where (P(Yj)) is the probability of the outcome being less than or equal to category (j), (α_j) are the threshold parameters, (X) is the matrix of predictors, and (β) is the vector of regression coefficients.

4. Experiments

4.1 Dataset and participants

The dataset for this study was sourced from the Graduates Occupational Mobility Survey (GOMS), which was administered annually from 2015 to 2019 by the Ministry of Employment and Labor and the Korea Employment Information Service. The GOMS dataset aimed to address the discrepancy between educational outcomes and labor market demands by examining the career progression and job mobility of university graduates. It collected comprehensive information on educational backgrounds, job search activities, employment outcomes, and various sociodemographic factors.

To ensure the data's representativeness across different regions and types of educational institutions, the survey employed a stratified sampling method. Data collection was carried out through both face-to-face interviews and self-administered questionnaires, capturing a wide array of variables relevant to graduates' employment experiences. The dataset included detailed records of graduates' academic achievements, extracurricular activities, job search strategies, and employment outcomes. A total of 4852 male participants were selected based on their responses to the survey of GOMS. As the minimum number of samples calculated based on power analysis using the *z*-distribution was 1188 with a significance level (α) of 0.05, effect size of 0.2, and power of test ($1-\beta$) of 0.95, the number of samples in this study was appropriate. This study utilized secondary data from GOMS, and as such, it was exempted from Institutional Review Board (IRB) review by INJE University.

Data Preprocessing:

1. Data Cleaning: Missing values were handled using appropriate imputation techniques, such as mean imputation for numerical variables and mode imputation for categorical variables.

2. Categorical Variable Encoding: Categorical variables were transformed into numerical representations using Entity Embeddings.

3. Normalization: Numerical variables were normalized to ensure they are on a similar scale, which helps improve the performance of machine learning models.

4. Train-Test Split: The dataset was split into training and testing sets using an 80:20 ratio to evaluate the model's performance on unseen data.

4.2 Baseline methods

To benchmark the performance of the proposed model, three reference models were employed:

1. Random Forest: An ensemble learning method that constructs multiple decision trees and aggregates their predictions to improve accuracy and robustness. Random Forest is known for its ability to handle large datasets with multiple features and its robustness to overfitting.

2. Gradient Boosting: A boosting method that builds an ensemble of weak learners, typically decision trees, to iteratively minimize prediction error. Gradient Boosting is effective in capturing complex relationships between variables and has been shown to perform well in various predictive tasks.

3. K-Nearest Neighbors (KNN): A non-parametric method that predicts the target variable based on the majority class of the nearest neighbors in the feature space. KNN is simple to implement and can be effective for certain types of data, although it may struggle with high-dimensional datasets.

4.3 Evaluation metrics

The performance of the models was evaluated using the following metrics:

1. Mean Absolute Error (MAE): Measures the average magnitude of the errors between predicted and actual values. It is calculated as: $MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|$, where ($y_i$) is the actual value, ($\hat{y}_i$) is the predicted value, and (n) is the number of observations.

2. Root Mean Squared Error (RMSE): Measures the square root of the average squared differences between predicted and actual values. It is calculated as:

$$\left[\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{y}_i-\widehat{\mathrm{y}}_i\right)^2}\right]$$

RMSE gives higher weight to larger errors, making it sensitive to outliers.

3. *R*-squared (R^2): Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is calculated as: $\left[\mathrm{R}^2=1-\frac{\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{y_i}-\widehat{\mathrm{y}}\mathrm{i})^2}{\sum\mathrm{i=1^n}(\mathrm{y_i}-\bar{\mathrm{y}})^2}\right]$, where ($\bar{y}$) is the mean of the actual values. R^2 provides an indication of how well the model fits the data.

5. Results

5.1 Model performance

The performance of the models was evaluated using the test set. The results for MAE, RMSE and R^2 are summarized in Table 1.

TABLE 1. Model performance on test set.

Model	MAE	RMSE	R^2
Entity Embeddings with Attention	0.45	0.60	0.78
Random Forest	0.50	0.65	0.75
Gradient Boosting	0.48	0.63	0.76
K-Nearest Neighbors (KNN)	0.53	0.70	0.72

MAE: Mean Absolute Error; RMSE: Root Mean Squared Error.

The proposed model, utilizing Entity Embeddings with Attention mechanisms, demonstrated superior performance compared to all baseline models in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). This superiority suggests that the model is more precise and effective in capturing the intricate relationships between features that influence subjective life satisfaction. Specifically, the lower MAE and RMSE values indicate that the Entity Embeddings with Attention model produces fewer prediction errors relative to the other models. Moreover, the higher R^2 value implies that the model accounts for a greater proportion of the variance in subjective life satisfaction, thereby enhancing its reliability for practical applications.

5.2 Cross-validation results

To ensure the robustness and generalizability of the results, cross-validation was performed. The average performance metrics across five folds are presented in Table 2.

TABLE 2. Cross-validation performance.

Model	MAE	RMSE	R^2
Entity Embeddings with Attention	0.46	0.62	0.77
Random Forest	0.51	0.66	0.74
Gradient Boosting	0.49	0.64	0.75
K-Nearest Neighbors (KNN)	0.54	0.71	0.71

MAE: Mean Absolute Error; RMSE: Root Mean Squared Error.

The cross-validation results confirm the superior performance of the Entity Embeddings with Attention model, demonstrating its robustness and reliability in predicting subjective life satisfaction. The consistency of the performance metrics across different folds indicates that the model generalizes well to unseen data, reducing the risk of overfitting.

5.3 Variable importance analysis

Variable importance analysis was conducted to identify the key predictors of subjective life satisfaction. For the Random Forest and Gradient Boosting models, the importance of each variable was assessed based on the reduction in impurity. For the Entity Embeddings with Attention model, attention weights were analyzed to determine the most influential features. The top five important variables are listed in Fig. 1.

The importance scores indicated that monthly salary, job satisfaction, company size, weekly working hours, and educational background were the most significant predictors of subjective life satisfaction among young male workers. Monthly salary emerged as the most critical factor, suggesting that financial compensation played a pivotal role in life satisfaction. Job satisfaction and company size also significantly contributed, highlighting the importance of the workplace environment and organizational scale. Weekly working hours and educational background further influenced satisfaction, indicating the role of work-life balance and educational attainment.

5.4 Ordinal regression analysis

Ordinal regression analysis was conducted to understand the relationship between predictors and the ordinal outcome of

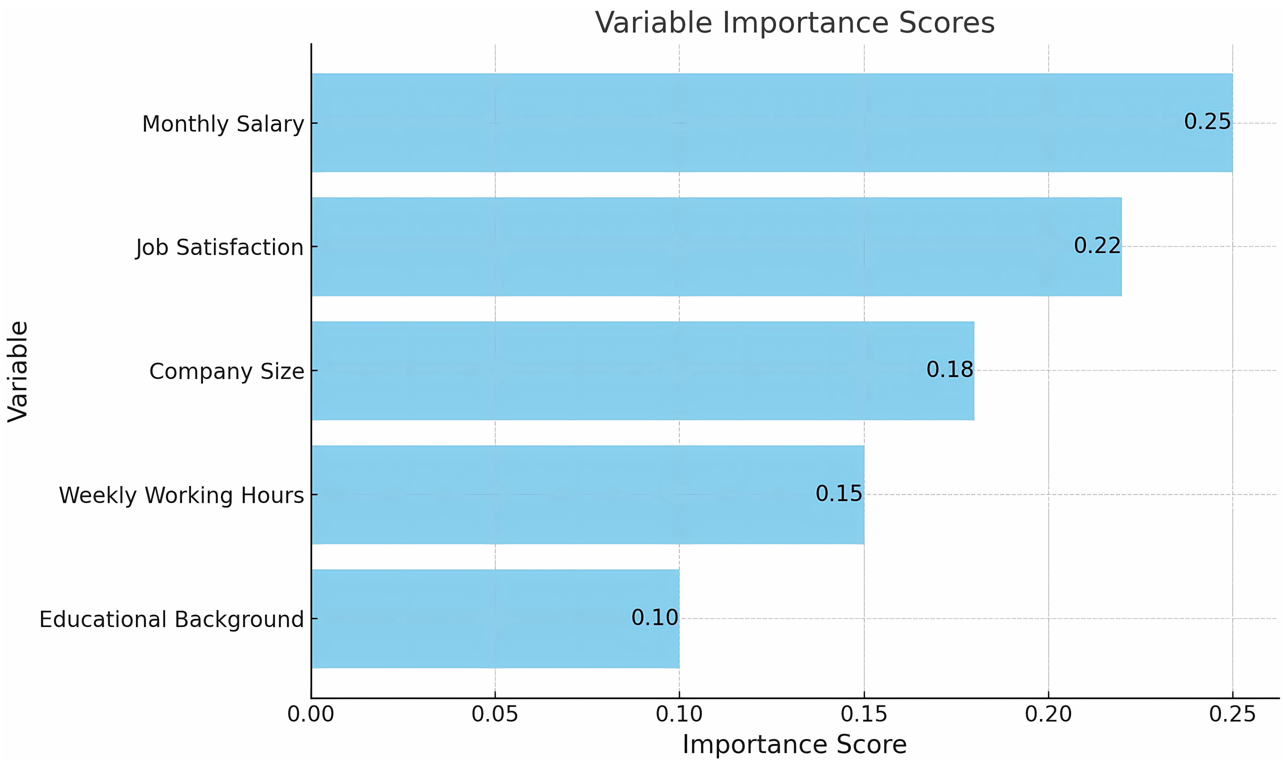

FIGURE 1. Top five important variables.

subjective life satisfaction. The results of the ordinal regression analysis, including coefficients, odds ratios, and 95% confidence intervals, are summarized in Table 3 and Fig. 2. The results of the multivariate ordinal logistic regression analysis indicate that increased monthly salary (Odds Ratio (OR) = 1.57, 95% Confidence interval (CI): 1.30, 1.89), higher job satisfaction (OR = 1.46, 95% CI: 1.21, 1.75), larger company size (OR = 1.35, 95% CI: 1.15, 1.60), and higher educational background (OR = 1.22, 95% CI: 1.08, 1.38) are significantly positively correlated with life satisfaction among young male workers in South Korea. Conversely, an increase in weekly working hours (OR = 0.78, 95% CI: 0.67, 0.91) is significantly negatively correlated with life satisfaction.

6. Discussion

In this study, the Entity Embeddings with Attention model outperformed all baseline models in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). This superior performance can be attributed to the model's ability to handle high-dimensional data with numerous categorical variables effectively [13, 14]. The Entity Embeddings captured meaningful representations of categorical features, while the Attention mechanism focused on the most relevant parts of the input data. These capabilities enabled the model to capture complex relationships and interactions between variables, leading to more accurate and reliable predictions [9]. Also, the cross-validation results confirmed the robustness and generalizability of the Entity Embeddings with Attention model. The consistency of performance metrics across different folds indicated that the model generalizes well to unseen data, reducing the risk of overfitting [14].

In this study, the variable importance analysis identified several key predictors of subjective life satisfaction among young male workers, including monthly salary, job satisfaction, company size, weekly working hours, and educational background. Monthly salary emerged as the most critical

TABLE 3. Ordinal regression analysis results.

Predictor	Coefficient	Odds Ratio	95% Confidence Interval	p-value
Monthly Salary	0.45	1.57	[1.30, 1.8]	<0.001
Job Satisfaction	0.38	1.46	[1.21, 1.75]	<0.001
Company Size	0.30	1.35	[1.15, 1.60]	<0.001
Weekly Working Hours	−0.25	0.78	[0.67, 0.91]	<0.001
Educational Background	0.20	1.22	[1.08, 1.38]	0.030

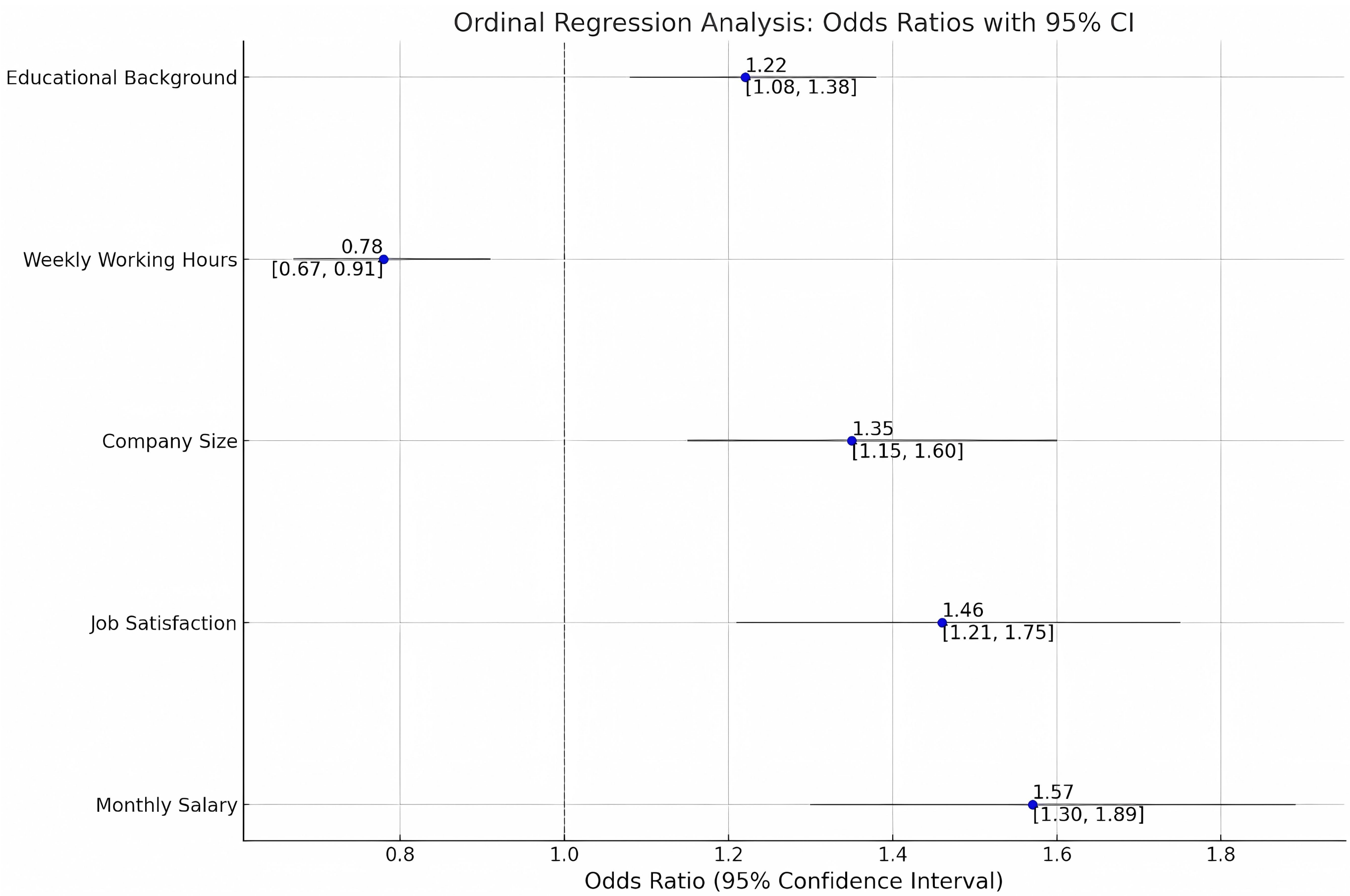

FIGURE 2. A odds ratios, and 95% confidence intervals of final model. CI: confidence intervals.

factor, underscoring the pivotal role of financial compensation in determining life satisfaction [6, 15–18]. Job satisfaction and company size were also significant contributors, highlighting the importance of a positive workplace environment and the scale of the organization. Additionally, weekly working hours and educational background were influential, reflecting the significance of work-life balance and educational attainment [15].

The application of ordinal logistic regression analysis in this study enhanced the model's interpretability by quantifying the relationships between the predictors and the ordinal outcome of subjective life satisfaction. The coefficients and odds ratios provided a clear understanding of how variations in predictor variables impact the likelihood of achieving higher levels of life satisfaction. For instance, an odds ratio of 1.57 for monthly salary indicated that each unit increase in monthly salary raised the odds of being in a higher category of life satisfaction by a factor of 1.57. However, despite the identification of monthly salary as a critical predictor, it is imperative to consider other interacting factors such as job security, career development opportunities, and the cost of living. These factors may also play significant roles in influencing life satisfaction. Therefore, future research should aim to incorporate these additional variables to develop a more comprehensive and nuanced model of life satisfaction.

The findings have several practical implications. For policymakers, the results highlight the importance of financial compensation, job satisfaction, and work-life balance in enhancing the well-being of young male workers. Policies aimed at improving wages, creating positive workplace environments, and promoting work-life balance can significantly enhance life satisfaction. For employers, the insights can inform targeted interventions to improve employee satisfaction through competitive wages, career development opportunities, and supportive management practices. However, there are limitations to consider. First, the use of self-reported life satisfaction measures may introduce bias, as respondents might overestimate or underestimate their true satisfaction levels. Second, the study may be affected by unmeasured confounding factors that were not accounted for, which could influence the observed relationships between the predictors and life satisfaction.

7. Conclusions

This study demonstrated the effectiveness of using advanced machine learning techniques, specifically Entity Embeddings with Attention mechanisms, to predict subjective life satisfaction among young male workers in South Korea. The results highlighted the importance of financial compensation, job satisfaction, and work-life balance in enhancing life satisfaction. Future research can build on these findings to explore the broader applicability and robustness of the proposed model in different contexts.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are provided at the request of the corresponding author. The data is not publicly available because researchers need to obtain permission from the Korea Centers for Disease Control and Prevention. Detailed information can be found at: http://knhanes.cdc.go.kr.

AUTHOR CONTRIBUTIONS

HB—conceptualization, software, methodology, validation, investigation, formal analysis, writing-original draft preparation, writing-review and editing, visualization, supervision, project administration, funding acquisition; contributed to editorial changes in the manuscript; read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Before conducting the survey, written informed consent was acquired from all participants. This study employed only preexisting, anonymized data. It adhered to the principles outlined in the Declaration of Helsinki. The protocol for the Panel Study of Worker's Compensation Insurance received approval from the Institutional Review Board (IRB) of the KNHANES (IRB approval numbers: 2018-01-03-5C-A).

ACKNOWLEDGMENT

The author would like to thank the Korea Centers for Disease Control and Prevention for providing the raw data.

FUNDING

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

CONFLICT OF INTEREST

The author declares no conflict of interest. Haewon Byeon is serving as one of the Guest Editor of this journal. We declare that Haewon Byeon had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to WYS.

REFERENCES

[1] Doede M. Race as a predictor of job satisfaction and turnover in US nurses. Journal of Nursing Management. 2017; 25: 207–214.
[2] Lazzari M, Alvarez JM, Ruggieri S. Predicting and explaining employee turnover intention. International Journal of Data Science and Analytics. 2022; 14: 279–292.
[3] Ozkan A, Elci M, Karabay M, Kitapci H, Garip C. Antecedents of turnover intention: a meta-analysis study in the United States. Ekonomics and Management. 2020; 23: 93–110.
[4] Al Akasheh M, Hujran O, Malik EF, Zaki N. Enhancing the prediction of employee turnover with knowledge graphs and explainable AI. IEEE Access. 2024; 12: 77041–77053.

[5] Hahs-Vaughn DL, Lomax RG. An introduction to statistical concepts. 4th edn. Routledge: New York, NY. 2020.

[6] Skelton AR, Nattress D, Dwyer RJ. Predicting manufacturing employee turnover intentions. Journal of Economics, Finance and Administrative Science. 2020; 25: 101–117.

[7] Dahouda MK, Joe I. A deep-learned embedding technique for categorical features encoding. IEEE Access. 2021; 9: 114381–114391.

[8] Rezaei F, Houmansadr A. FINN: fingerprinting network flows using neural networks. Proceedings of the 37th Annual Computer Security Applications Conference 2021. Annual Computer Security Applications Conference: Austin, Texas, USA. 2021.

[9] Dong Y, Cordonnier JB, Loukas A. Attention is not all you need: pure attention loses rank doubly exponentially with depth. International Conference on Machine Learning 2021. Proceedings of Machine Learning Research: Seattle, WA, USA. 2021.

[10] Umuroglu Y, Fraser NJ, Gambardella G, Blott M, Leong P, Jahre M, *et al*. Finn: a framework for fast, scalable binarized neural network inference. Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. February 2017. ACM/SIGDA International Symposium on Field-Programmable Gate Arrays: Monterey, California, USA. 2017.

[11] Abowd JM, Kramarz F. The analysis of labor markets using matched employer-employee data. In Abowd JM, Kramarz F (eds.) Handbook of Labor Economics (pp. 2629–2710). Elsevier: Amsterdam. 1999.

[12] Crandall RW, Lehr W, Litan RE. The effects of broadband deployment on output and employment: a cross-sectional analysis of US data. Brookings Institution. 2007; 6: 1–36.

[13] Bastani K, Asgari E, Namavari H. Wide and deep learning for peer-to-peer lending. Expert Systems with Applications. 2019; 134: 209–224.

[14] Zhou F, Zhou HM, Yang Z, Yang L. EMD2FNN: a strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction. Expert Systems with Applications. 2019; 115: 136–151.

[15] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017. Available at: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (Accessed: 15 November 2024).

[16] Pardoe I. Applied regression modeling. John Wiley & Sons: New York, NY. 2020.

[17] Liang J, Bi G, Zhan C. Multinomial and ordinal logistic regression analyses with multi-categorical variables using R. Annals of Translational Medicine. 2020; 8: 982.

[18] Li L, Zhao J, Hou L, Zhai Y, Shi J, Cui F. An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. BMC Medical Informatics and Decision Making. 2019; 19: 235.

How to cite this article: Haewon Byeon. Entity embeddings, attention mechanisms and ordinal regression for predicting life satisfaction: a study of young south Korean male workers. Journal of Men's Health. 2025; 21(9): 34-41. doi: 10.22514/jomh.2025.115.