Asia Pacific Journal of Information Systems
Vol. 34 No. 1 (March 2024), 209-225

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2024.34.1.209

A Prediction of Work-life Balance Using Machine Learning

Youngkeun Choi*

Associate Professor, Sangmyung University Seoul, Korea

ABSTRACT

This research aims to use machine learning technology in human resource management to predict employees' work-life balance. The study utilized a dataset from IBM Watson Analytics in the IBM Community for the machine learning analysis. Multinomial dependent variables concerning workers' work-life balance were examined, categorized into continuous and categorical types using the Generalized Linear Model. The complexity of assessing variable roles and their varied impact based on the type of model used was highlighted. The study's outcomes are academically and practically relevant, showcasing how machine learning can offer further understanding of psychological variables like work-life balance through analyzing employee profiles.

Keywords: Human Resource Management, Work-life Balance, Machine Learning, Data Science

Ⅰ. Introduction

Organizations incur significant costs from recruitment, selection, induction, and training of employees, aiming to align them with the organization's objectives (Raheef, 2019). The departure of a key employee represents a substantial loss, not only financially but also in terms of knowledge, information, experience, and skills, which are vital for organizational growth and may benefit competitors when employees leave. Thus, employee retention is crucial for organizational success. Attrition can result from various factors, with dissatisfaction being a common reason for employees seeking better opportunities elsewhere. This issue is closely linked to work-life balance, which involves effectively balancing family and work commitments to achieve satisfaction and happiness (Balven et al., 2018). Work-life balance comprises three key elements: the balance of time between work and family, the psychological engagement with work and non-work roles, and the satisfaction derived from both work and family roles. Individuals experiencing difficulties in meeting their responsibilities due to work or facing work-related

*Corresponding Author. E-mail: penking1@smu.ac.kr

challenges due to their personal responsibilities are said to experience work-life balance issues.

Human resources possess vast amounts of data (De Mauro et al., 2018), including built-in data such as employee information, participation scores, and performance records. However, much of this data is lost or underutilized after its initial use, preventing organizations from extracting valuable insights, conducting detailed analysis, and offering new opportunities and benefits. In today's world, even seemingly minor details like an employee's name and address or their purchases can hold significant importance. Therefore, data is a crucial element for organizational success. The sheer volume, variety, and velocity of this data require new approaches to big data analytics, including various analysis and storage methods. It is essential to accurately analyze this vast amount of big data and extract relevant information. The HR department has begun leveraging data analytics to identify top performers, improve retention rates, and enhance overall employee satisfaction. As technology and the Internet evolve, the amount of available information continues to grow exponentially. With increasing storage capabilities and data classification methods, organizations can now store and analyze massive amounts of data to extract value. It is crucial for organizations to effectively utilize their vast stores of data due to the low cost of storage and the potential insights it can provide.

The term "Big Data" has gained popularity in various data analytics efforts (Gandomi and Haider, 2015). Initially, it emerged to describe a technological revolution enabling the collection of massive amounts of data (Jacobs, 2009). Over time, its usage has evolved across different domains to represent various aspects of analysis, depending on the context in which it is used. Nowadays, it encompasses both technical and commercial aspects of data collection activities, representing data processing capabilities and characteristics (Nunan and Di Dominico, 2017). Mayer-Schönberger and Cukier (2013) view big data as a set of new features that enable the collection and immediate analysis of vast amounts of information (Kitchin, 2014). Similarly, Boyd and Crawford (2012) argue that big data is not solely about the size of the data but also about the ability to search, aggregate, and cross-reference large datasets. The development of computer and Internet-based technologies has significantly increased the amount and availability of data worldwide, enabling the capture of large-scale data more easily than ever before. This abundance of data allows for the development of insights through new information processing or decision-making tools, utilizing analytical formulas and rule-development possibilities (data-processing algorithms) to solve problems. More recently, the proliferation of intelligent machine learning algorithms in computer science has provided a powerful quantitative method for deriving insights from industrial data.

Statistical data analysis and machine learning differ in several key aspects, as highlighted by Garg and Tai (2013). First, their approach differs: statistical data analysis focuses on understanding and modeling data distributions and relationships using statistical techniques, while machine learning emphasizes learning patterns from data and building predictive models using algorithms. Second, their purpose differs: statistical data analysis aims to characterize data through statistical inference, such as hypothesis testing and estimation, while machine learning aims to create predictions for new data through tasks like prediction and classification. Third, their data processing methods differ: statistical data analysis primarily involves data preprocessing considering variable distribution and handling outliers and missing values, whereas machine learning transforms input data into a model through preprocessing and feature extraction. Fourth, their algorithm selection differs: statistical data analysis primarily uses statistical techniques like t-tests and regression analysis, while machine learning employs various algorithms such as linear regression, decision trees, and neural networks for training and prediction. Finally, their automation and scaling capabilities differ: machine learning can handle large amounts of data and complex patterns with automated model training, while statistical data analysis often relies on small datasets and explicit assumptions. Supervised machine learning methods have been applied in diverse fields like biology, medical science, transportation, and political science, and researchers have explored machine learning approaches to enhance human resources management outcomes in response to advances in information technology.

One approach for organizations to address this challenge is by utilizing machine-learning technology to predict employees' work-life balance. This allows leaders and HR personnel to proactively take measures or plan succession strategies for retention. However, traditional machine-learning technology used for this purpose often fails to consider data noise present in many HR Information Systems (HRIS). Many organizations have not prioritized investments in efficient HRIS solutions that accurately capture employee data throughout their tenure. A key issue is the limited understanding of the benefits and costs associated with such investments, making it challenging to measure the return on investment in HRIS (Jahan, 2014). This data noise weakens the generalizability of these algorithms.

To address these concerns, a generalized linear model (GLM) algorithm is employed to predict employees' work-life balance. In the HR field, measuring employees' psychological dependent variables is challenging due to the variance in psychological standards among individuals. GLM is advantageous in transforming and analyzing the shape of the dependent variable, especially when psychological variables measured in multinomial categories need to be converted into other types of variables. This study analyzes the values of multinomial categorical dependent variables related to workers' work-life balance by categorizing them into continuous and categorical types.

In typical scenarios, machine-learning technology can develop algorithms based on employee attributes relevant to the job performance of the current workforce. However, there are challenges with using such algorithms, particularly regarding gender bias. While there may be a causal relationship between traits like gender and work-life balance, algorithms that promote one gender over another may be unreliable. This is because job performance itself can be influenced by biased indicators, the composition of the current workforce, and historical data (e.g., if few women were previously employed). This paper discusses key machine learning algorithms used to address employees' work-life balance issues, providing a new contribution by exploring the application of machine learning. The preprocessing and modeling methodology outlined in this paper can serve as a guide for readers to follow the steps taken in this study and apply similar procedures to identify the causes of various other HR issues. Consequently, this paper offers a straightforward and immediate way to select potential employees, providing a unique benefit to HR departments.

Ⅱ. Literature Review on Work-Life Balance

Bachmann (2000) and Schwartz (1994) discovered that family-friendly work environments, including flexible working hours and telecommuting, significantly influence individual workers' preferences and job choices. Hyman et al. (2003) conducted empirical research in the UK, revealing that the intrusion of work demands into personal life, such as working on weekends, is associated with increased reports of stress and emotional exhaustion among employees. Employees also acknowledged that this intrusion negatively impacted their health. Ferguson et al. (2012) found that work-family balance plays a mediating role in facilitating social support, which contributes to both job and family satisfaction.

Dundas (2008) defined work-life balance as the effective management of balancing paid work with other important activities such as family, community involvement, voluntary work, personal development, and leisure. Khallash and Kruse (2012) suggested that work-life balance involves separating work from private life, with balance achieved through an equal division between the two. Balven et al. (2018) described work-life balance as organizational support for aspects of employees' personal lives, including flexible working hours, dependent care, and family/personal leave, requiring alignment between organizational and individual factors.

Yawalkar and Sonawane (2017) studied the relationship between demographic variables and work-life balance among police personnel in the Jalgaon district police department. They found that poor work-life balance leads to workplace mistakes, increased errors, and negative effects on health. Richert-Kazmierska and Stankiewicz (2016) investigated the impact of age on workers' assessment of work-life balance, revealing a statistically significant finding that older employees are more likely to maintain work-life balance.

Imna and Hassan (2015) conducted a study on various independent variables, including career and development, training and development, performance appraisal, reward and compensation, and health and safety, to measure human resource management practices and their impact on employee retention. Syed (2015) highlighted the specific implications of work-life balance for individuals facing multiple disadvantages due to intersecting identities, recommending various management policies to address these issues. Kamau et al. (2013) discussed stress and work-life balance, emphasizing the importance of individuals having control over when, where, and how they work. The authors identified time and stress as two key variables and emphasized the consequences of work-life imbalance on daily life, as well as the role of organizational policies in addressing this imbalance. Santhana and Gopinath (2013) highlighted the challenges faced in balancing work demands and family responsibilities, noting a high correlation between these challenges and the need for work-life balance. They also proposed specific HRM interventions to enhance work-life balance. Sudha and Karthikeyan (2014) identified various aspects such as career progress, work stress, career aspirations, work-family conflict, and family-work conflict in child care in the context of work-life balance. They outlined the challenges and issues faced by women employees in achieving work-life balance. Asiedu-Appiah et al. (2013) emphasized the importance of work-life balance in improving employee performance at work and home, noting gender differences in work-life balance needs. They found that women generally displayed a greater need for work-life balance compared to men. Cetin et al. (2013) surveyed 90 accounting and finance professionals in Turkey, focusing on the impact of mentoring on organizational commitment, job satisfaction, and work-life balance. Their study showed that mentoring, including career development, role modeling, and social support, significantly affected these variables. McCarthy et al. (2010) highlighted the importance of organizational support in maintaining a good work-life balance, noting that organizations that neglect this balance may exploit their employees' hard work and effort.

Singh (2013) examined role stress theory in his paper titled "Work-Life Balance: A Literature Review," focusing on the negative aspects of the work-family interaction. Recently, there has been a shift towards investigating the positive interaction between work and family roles. Hanif and Shao (2013) linked HR generic strategies with retention, highlighting factors such as succession planning, employer branding, motivation, and effective development policies. Kumari and Devi (2013) addressed work-life balance as one of the most challenging issues for women employees in the 21st century, considering the roles they play at home and the spillover of personal life into work. Lakshmi and Gopinath (2013) emphasized that educational institutions should address work-life balance issues among their staff, particularly women, by implementing policies to support teaching staff in managing their work-life balance. Bee, Baskar, and Vimala (2013) noted that family-work conflict and work-family conflict negatively impact work-life balance, with several variables influencing the level of conflict. Sajjad and Nas (2013) analyzed the impact of organizational commitment and support on employee retention and commitment. Gupta and Kumar (2012) researched staff retention measures, effectiveness of exit interviews, competitiveness of reward programs, and workplace environment conduciveness, highlighting these as factors contributing to workforce turnover. Human resource management practices are now considered key factors in organizational success. Thriveni and Rama (2012) found that women are more dependent and have greater responsibilities, with demographic factors influencing women's work-life balance.

Ⅲ. Methodology

3.1. Dataset

The dataset utilized in this paper for experimentation pertained to work-life balance and is accessible at IBM Watson Analytics in the IBM Community (https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset). IBM Watson Analytics in IBM Community offers several advantages due to its diverse datasets. Firstly, it provides datasets from various fields, making it easier for users to locate and utilize data in their desired field. Secondly, it offers large datasets, enabling more accurate analysis results by providing the necessary data for analysis. Thirdly, it offers high-quality datasets, increasing the reliability of analysis results and preventing incorrect conclusions by ensuring data accuracy. Lastly, it provides pre-cleaned datasets, allowing users to perform analyses more quickly by reducing data cleaning time. The IBM dataset used in this study was collected from 1470 individuals. Work-life balance was set as the outcome variable in this dataset, with the remaining 34 variables set as causal variables influencing the prediction of work-life balance. The minimum tolerance of 0.612 and maximum variance inflation factor (VIF) of 1.634 indicate that the statistical significance of the data analysis was not compromised by multicollinearity. Upon calculating the VIF for each variable, all VIF values were found to be less than 2, which did not exceed the permissible range

<Table 1> The Measurements of Variables

Variables		Measurement
Independent Variables	Age	Continuous Integer
	Attrition	Binomial (True or False)
	BusinessTravel	Polynomial (Travel_Rarely 71%, Travel_Frequently 19%, Other 10%)
	DailyRate	Continuous Integer
	Department	Polynomial (Research & Development 65%, Sales 30%, Other 4%)
	DistanceFromHome	Continuous Integer
	Education	Continuous Integer (1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'
	EducationField	Polynomial (Life Sciences 41%, Medical 32%, Other 27%)
	EmployeeCount	Continuous Integer
	EmployeeNumber	Continuous Integer
	EnvironmentSatisfaction	Continuous Integer
	Gender	Binomial (Male 60%, Female 40%)
	HourlyRate	Continuous Integer
	JobInvolvement	Integer (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
	JobLevel	Continuous Integer
	JobRole	Polynomial (Sales Executive 22%, Research Scientist 20%, Other 58%)
	JobSatisfaction	Continuous Integer (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
	MartialStatus	Polynomial (Married 46%, Single 32%, Other 22%)
	MonthlyIncome	Continuous Integer
	MonthlyRate	Continuous Integer
	NumCompniesWorked	Continuous Integer
	Over18	Binomial (True or False)
	OverTime	Binomial (True or False)
	PercentSalarHike	Continuous Integer
	PerformanceRating	Continuous Integer (1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding')
	RelationshipSatisfaction	Continuous Integer (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
	StandardHours	Continuous Integer

<Table 1> The Measurements of Variables (Cont.)

Variables		Measurement
Independent Variables	StockOptionLevel	Continuous Integer
	TotalWorkingYears	Continuous Integer
	TrainingTimesLastYear	Continuous Integer
	YearsAtCompany	Continuous Integer
	YearsinCurrentRole	Continuous Integer
	YearsSinceLastPromotion	Continuous Integer
	YearsWithCurrManager	Continuous Integer
Dependant Variable	WorkLifeBalance	Continuous Integer (1 'Bad' 2 'Good' 3 'Better' 4 'Best')

of 10 (Chatterjee et al., 2000). The details of the 35 attributes are shown in <Table 1>.

3.2. Generalized Linear Model

The Generalized Linear Model (GLM) is a widely used tool, even when the dependent variable does not adhere to a normal distribution (Calabrese and Elkink, 2016). GLM is applicable when the dependent variable follows diverse distributions, including binomial, Poisson, and gamma distributions. This versatility is essential in predictive modeling, as datasets often encompass variables that conform to different distributions. For instance, in a binomial distribution scenario, consider a model predicting the likelihood of a person testing positive in a medical examination. Here, the dependent variable is binary, assuming either a negative (0) or positive (1) value. The logistic regression model exemplifies GLM in such instances. Similarly, in a Poisson distribution context, imagine a model forecasting the daily number of traffic accidents in a city. Here, the dependent variable is an integer, taking discrete values. The Poisson regression model represents GLM in this scenario. Lastly, in a gamma distribution context, think of a model predicting the production cost of a company's manufactured product. Here, the dependent variable is continuous, assuming positive values. The gamma regression model serves as an example of GLM in this case. Therefore, GLM is suitable for addressing dependent variables adhering to various distributions, enhancing the accuracy of predictive models. Given its wide application in data analysis, understanding and utilizing GLM is crucial.

This study seeks to predict employees' work-life balance levels. GLMs are versatile in handling various types of dependent variables, including multinomial categorical variables. GLM's capability to transform and analyze the dependent variable's shape is particularly beneficial. This study demonstrates these strengths by analyzing the work-life balance of workers, dividing it into continuous and categorical types. By measuring work-life balance levels in a continuous form, this study can compare these levels among employees. Additionally, by categorically measuring work-life balance levels, individual employees' balance can be managed as a grade. Thus, GLM, adept at handling both discrete and categorical dependent variable values, is ideal for this analysis and prediction.

3.3. Preprocessing and Data Mining Models

In this study, we aim to analyze the factors influencing work-life balance. Job satisfaction is rated on a scale of 1 to 4. The goal is to assess whether GLM can address two types of problems: numerical prediction and binomial classification. To achieve this, the original numerical dependent variable in the dataset was converted into a binomial category. We utilized statistical and data mining techniques to develop decision prediction models. Data mining techniques help discover patterns or relationships in data and predict or classify behavior by creating a model based on available data.

When separating the learning dataset and the test dataset for machine learning, the test dataset must meet specific criteria. First, both datasets must be in the same format. Second, the test dataset should not be part of the training dataset. Third, the datasets must be consistent in terms of data. However, creating a test dataset that meets all these requirements is challenging. To address this, various validation frameworks using a single dataset have been developed in data mining. In this study, we utilized the Split Validation operator in RapidMiner. This operator divides the input dataset into a training dataset and a test dataset for performance evaluation. We chose relative segmentation among the operator's segmentation method parameters and used 70% of the input data as the learning dataset.

Ⅳ. Results

4.1. Linear Regression Model

In linear regression analysis, the model is represented as a function. <Table 2> displays the intercept, coefficient, and standard coefficient obtained from the regression analysis. The regression coefficient indicates how each explanatory variable influences the intercept of the dependent variable. When using different units of measurement for explanatory variables, it becomes challenging to explain how a one-unit increase in these variables affects the dependent variable. To address this, the standard coefficient is calculated by standardizing the variables and then estimating the regression model. The standard coefficients enable comparison of the effects of each explanatory variable on the dependent variable. In <Table 2>, variables such as Attrition.No, BusinessTravel. Travel_Frequently, Department.Human Resources, Education, EducationField.Medical, EducationField.Other, EducationField. Technical Degree, EmployeeNumber, EnvironmentSatisfaction, Gender.Male, HourlyRate, JobLevel, JobRole.Laboratory Technician, JobRole. Research Director, JobRole.Sales Executive, JobRole.Sales Representative, MaritalStatus. Single, MonthlyRate,

<Table 2> The Results of the Linear Regression Model

Attribute	Coefficient	Std. Coefficient
Age	-0.003	-0.003
Attrition.No	0.075	0.075
Attrition.Yes	-0.075	-0.075
BusinessTravel.Non-Travel	0	0
BusinessTravel.Travel_Frequently	0.025	0.025
BusinessTravel.Travel_Rarely	-0.014	-0.014
DailyRate	-0.000	-0.030
Department.Human Resources	0.311	0.311
Department.Research & Development	-0.037	-0.037

<Table 2> The Results of the Linear Regression Model (Cont.)

Attribute	Coefficient	Std. Coefficient
Department.Sales	0	0
DistanceFromHome	-0.002	-0.015
Education	0.013	0.013
EducationField.Human Resources	-0.307	-0.307
EducationField.Life Sciences	-0.017	-0.017
EducationField.Marketing	-0.001	-0.001
EducationField.Medical	0.024	0.024
EducationField.Other	0.104	0.104
EducationField.Technical Degree	0.076	0.076
EmployeeNumber	0.000	0.001
EnvironmentSatisfaction	0.012	0.013
Gender.Female	-0.003	-0.003
Gender.Male	0.003	0.003
HourlyRate	0.000	0.003
JobInvolvement	-0.019	-0.013
JobLevel	0.074	0.081
JobRole.Healthcare Representative	-0.048	-0.048
JobRole.Human Resources	-0.004	-0.004
JobRole.Laboratory Technician	0.010	0.010
JobRole.Manager	-0.005	-0.005
JobRole.Manufacturing Director	0	0
JobRole.Research Director	0.095	0.095
JobRole.Research Scientist	-0.037	-0.037
JobRole.Sales Executive	0.019	0.019
JobRole.Sales Representative	0.157	0.157
JobSatisfaction	-0.017	-0.019
MaritalStatus.Divorced	-0.023	-0.023
MaritalStatus.Married	0	0
MaritalStatus.Single	0.052	0.052
MonthlyIncome	-0.000	-0.037
MonthlyRate	0.000	0.006
NumCompaniesWorked	0.002	0.006
OverTime.No	0.002	0.002
OverTime.Yes	-0.002	-0.002
PercentSalaryHike	-0.001	-0.005
PerformanceRating	0.024	0.009

<Table 2> The Results of the Linear Regression Model (Cont.)

Attribute	Coefficient	Std. Coefficient
RelationshipSatisfaction	0.014	0.015
StockOptionLevel	0.022	0.019
TotalWorkingYears	-0.003	-0.023
TrainingTimesLastYear	0.010	0.013
YearsAtCompany	-0.004	-0.024
YearsinCurrentRole	0.022	0.080
YearsSinceLastPromotion	-0.001	-0.003
YearsWithCurrManager	-0.012	-0.042
Intercept	2.674	2.688

NumCompaniesWorked, OverTime.No, PerformanceRating, RelationshipSatisfaction, StockOptionLevel, TrainingTimesLastYear, and YearsinCurrentRole are shown to positively impact WorkLifeBalance.

The model was created using Gaussian as the distribution function (family) and identity as the link function (link). Since cross-validation was used for validation, the results may vary for each subset. Performance indicators for the linear regression model are presented in <Table 3>.

4.2. Binomial Classification Model

In the original dataset, WorkLifeBalance is represented as numerical data. To create a binomial classification, a new variable called WorkLifeBalance2 is created, where ‘H’ is assigned if WorkLifeBalance is greater than or equal to 3, and ‘L’ if it is less than 3. The model for binomial classification is expressed as a function. <Table 4> displays the intercept, coefficient, and standard coefficient obtained from regression analysis. In <Table 4>, variables such as Attrition.No, BusinessTravel.Travel_Frequently, Department.Human Resources, EducationField.Life Sciences, EducationField.Other, EducationField.Technical Degree, Gender.Female, JobRole.Manager, JobRole.Research Director, JobRole.Sales Executive, JobRole.Sales Representative, MaritalStatus.Single, NumCompaniesWorkedtrue, OverTime.Yes, StockOptionLevel.false, TotalWorkingYears.false, TrainingTimesLastYear.true, YearsAtCompany.true, YearsinCurrentRole.false, YearsSinceLastPromotion.true, and YearsWithCurrManager.false are found to be associated with WorkLifeBalance levels greater than 3.

The model was created using the binomial dis-

<Table 3> The Performance of Linear Regression Model

Performance Indicator	Measurement Value
root_mean_squared_error	0.716 +/- 0.045
absolute_error	0.548 +/- 0.031
relative_error	26.48% +/- 3.72%
squared_error	0.515 +/- 0.064
correlation	0.048 +/- 0.070

<Table 4> The Results of Binomial Classification Model

Attribute	Coefficient	Std. Coefficient
Attrition.No	0.296	0.296
Attrition.Yes	-0.247	-0.247
BusinessTravel.No-Travel	-0.087	-0.087
BusinessTravel.Travel_Frequently	0.069	0.069
BusinessTravel.Travel_Rarely	0	0
Department.Human Resources	1.429	1.429
Department.Research & Development	-0.181	-0.181
Department.Sales	-0.557	-0.557
EducationField.Human Resources	-0.520	-0.520
EducationField.Life Sciences	0.010	0.010
EducationField.Marketing	0	0
EducationField.Medical	-0.029	-0.029
EducationField.Other	0.142	0.142
EducationField.Technical Degree	0.312	0.312
Gender.Female	0.014	0.014
Gender.Male	-0.014	-0.014
JobRole.Healthcare Representative	-0.125	-0.125
JobRole.Human Resources	-0.782	-0.782
JobRole.Laboratory Techniciam	-0.033	-0.033
JobRole.Manager	0.026	0.026
JobRole.Manufacturing Director	0	0
JobRole.Research Director	0.214	0.214
JobRole.Research Scienctist	-0.314	-0.314
JobRole.Sale Executive	0.508	0.508
JobRole.Sale Representative	0.859	0.859
MaritalStatus.Divorced	-0.068	-0.068
MaritalStatus.Married	0	0
MaritalStatus.Single	0.017	0.017
NumCompaniesWorkedfalse	-0.095	-0.095
NumCompaniesWorkedtrue	0.104	0.104
OverTime.No	-0.008	-0.008
OverTime.Yes	0.008	0.008
StockOptionLevel.false	0.094	0.094
StockOptionLevel.true	-0.090	-0.090
TotalWorkingYears.false	0.068	0.068
TotalWorkingYears.true	-0.069	-0.069

<Table 4> The Results of Binomial Classification Model (Cont.)

Attribute	Coefficient	Std. Coefficient
TrainingTimesLastYear.false	-0.001	-0.001
TrainingTimesLastYear.true	0.001	0.001
YearsAtCompany.false	-0.130	-0.130
YearsAtCompany.true	0.139	0.139
YearsinCurrentRole.false	0.058	0.058
YearsCurrentRole.true	-0.058	-0.058
YearsSinceLastPromotion.false	-0.023	-0.023
YearsSinceLastPromotion.true	0.023	0.023
YearsWithCurrManager.false	0.108	0.108
YearsWithCurrManager.true	-0.107	-0.107
Intercept	0.842	0.842

<Table 5> The Performance of Linear Regression Model

Performance Indicator	Measurement Value
accuracy	70.88% +/- 0.63%
AUC	0.504 +/- 0.028
precision	71.11% +/- 0.43%
recall	99.52% +/- 0.68%
f_measure	82.95% +/-0.42%

tribution function (family) and the logit link function (link). <Table 5> displays the comprehensive performance indicators.

Ⅴ. Conclusions

5.1. Discussion

The main objective of this research paper is to assess the accuracy of existing models and develop a new model for predicting work-life balance. The study has two primary goals. Firstly, it aims to gain a better understanding of the variables' role in work-life balance prediction modeling. Secondly, it seeks to evaluate the predictive performance of the GLM, including both linear regression and binomial classification. The results indicate that determining the variables' role is complex, and their impact varies depending on the classification techniques used. GLM methods prioritize explanatory power, making it challenging to reach a unanimous conclusion about the most critical explanatory variables for work-life balance prediction. However, this research provides additional insights into employee profiles, suggesting that companies should use classification techniques to predict work-life balance.

5.2. Research Contributions and Practical Implications

This research contributes both theoretically and practically. It enhances the existing scholarly discussion by empirically examining the combined influence of variables on work-life balance modeling. While many studies have addressed work-life balance prediction, a universally applicable tool is lacking due to the complexities and interconnected nature of various factors. As a result, scholars often focus on a limited set of elements, overlooking the impacts of other variables, such as shifting employee demographics and privacy concerns. This study enriches the body of knowledge on work-life balance prediction by presenting a comprehensive model that encompasses the determinants of work-life balance prediction, including employee-related factors. Additionally, the methodology used in this paper provides a roadmap for readers to replicate the steps taken in this specific case study and apply the same approach to address other issues. The paper aims to develop an optimized predictive model for work-life balance, using a limited set of attributes that include employee-related factors. This is achieved by employing machine learning techniques such as Generalized Linear Models (GLM) and feature importance analysis to enhance accuracy. Using this methodology, the study reveals a discernible pattern in predicting work-life balance.

In practical terms, the application discussed in this paper helps businesses manage personnel records and speed up decision-making when the user's report is already on file. The paper introduces a prototype framework that businesses can use to make informed decisions about approving or denying employee work-life balance requests. Additionally, this study focuses solely on the managerial authority of the company, ensuring that the entire prediction process remains confidential and immune to external influence. Specific outcomes for a particular work-life balance inquiry can be shared with various departments within the company, allowing them to address the request appropriately and facilitate other formal processes across all divisions.

5.3. Limitations and Future Research Directions

Despite the academic contributions and practical implications of this study, it also highlights several limitations that warrant consideration. Firstly, as an experimental study proposing the application of machine learning to personnel management systems, it is constrained by the limitations inherent in using larger action datasets in practical machine learning models, such as those containing over a million different data points stored in electronic record systems. Secondly, the predictive performance of the study is limited due to insufficient data and the inclusion of variables that may distort the results. Specifically, in the results of binomial regression analysis, the Area Under the Curve (AUC) value was lower than other performance measures, which is unusual as AUC is typically close to other performance measures. Binomial regression analysis involves a division into two categories, which inherently introduces a risk of variance. To mitigate this risk, it is important not only to use a large dataset but also to carefully select independent variables that significantly impact the dependent variable of the binomial distribution. However, due to the small dataset limitations and the experimental nature of this study, independent variables were not rigorously screened. Therefore, future research should focus on improving analysis accuracy through the use of larger datasets and pre-processing of independent variables.

<References>

[1] Asiedu-Appiah, F., Dufie-Marfo, I., and Frempong, E. (2013). Work-life balance as a tool for stress management in selected banking institutions in Ghana. *Global Advanced Research Journal of Management and Business Studies, 2*(5), 291-311.

[2] Bachmann, K. (2000). *Work-life Balance. Are Employers Listening?* Ottawa: Conference Board of Canada.

[3] Balven, R., Fenters, V., Siegel, D. S., and Waldman, D. (2018). Academic entrepreneurship: The roles of organizational justice, championing, education, work-life balance, identity, and motivation. *Academy of Management Perspectives, 32*(1), 21-42. https://doi.org/10.5465/amp.2016.0127

[4] Bee, A. J., Baskar, K., and Vimala, V. (2013). Organisational culture on work-life balance among married women employees productivity. *Indian Streams Research Journal, 3*(1), 1-7. https://doi.org/10.4103/0253-7176.78508

[5] Boyd, D., and Crawford, K. (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. Information, *Communication & Society, 15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

[6] Calabrese, R., and Elkink, J. A. (2016). Estimating Binary Spatial Autoregressive Models for Rare Events. In *Spatial Econometrics: Qualitative and Limited Dependent Variables* (Advances in Econometrics, Vol. 37, pp. 145-166). Emerald Group Publishing Limited, Bingley. https://doi.org/10.1108/S0731-905320160000037012

[7] Cetin, A. T., Kizil, C., and Zengin, H. İ. (2013). Impact of mentoring on organizational commitment and job satisfaction of accounting-finance academicians employed in Turkey. *Emerging Markets Journal, 3*(2), 1-33. https://doi.org/10.5195/emaj.2013.40

[8] Chatterjee, S., Hadi, A. S., and Price, B. (2000) *Regression Analysis by Examples* (3rd ed.). Wiley VCH, New York.

[9] De Mauro, A., Greco, M., Grimaldi, M., and Ritala, P. (2018). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management, 54*(5), 807-817. https://doi.org/10.1016/j.ipm.2017.05.004

[10] Dundas, K. (2008). Work-Life Balance: There is no 'one size-fits-all' solution. In K O'Rourke (ed.), *Managing Matters* (pp. 00-00 누락). New South Wales: Southern Cross University.

[11] Ferguson, M., Carlson, D., Zivnuska, S., and Whitten, D. (2012). Support at work and home: The path to satisfaction through balance. *Journal of Vocational Behavior, 80*, 299-307. https://doi.org/10.1016/j.jvb.2012. 01.001

[12] Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management, 35*(2), 137-144. https://doi.org/10.1016/j.ijinfomgt .2014.10.007

[13] Garg, A., and Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control, 18*(4), 295-312. https://doi.org/10.1504/IJMIC.2013.053535

[14] Gupta, V., Kumar, S. (2012). Impact of performance appraisal justice on employee engagement: A study of Indian professionals. *Employee Relations, 35*(1), 61-78. https://doi.org/10.1108/01425451311279410

[15] Hanif, M. I., and Shao, Y. (2013). The role of talent management and HR generic strategies for talent retention. *African Journal of Business Management, 7*(29), 2827-2835. https://doi.org/10.5897/AJBM2012.1369

[16] Hyman, J., Baldry, C., Scholarios, D., and Bunzel, D. (2003). Work-life imbalance in the new service sector Economy. *British Journal of Industrial*

Relations, 41, 215-39.
[17] Imna, M., and Hassan, Z. (2015). Influence of human resource management practices on employee retention in Maldives retail industry. *International Journal of Accounting & Business Management, 3*(1), 50-80. https://doi.org/10.24924/ijabm/2015.04/v3.iss1/50.80
[18] Jacobs, A. (2009). Pathologies of Big Data. *Communications of the ACM, 52*(8), 36-44.
[19] Jahan, S. (2014). HRIS - a theoretical perspective. *Journal of Human Resource and Sustainability Studies, 2*, 33-39.
[20] Kamau, J. M., Muleke, V., Mukaya, S. O., and Wagoki, J. (2013). Work-life balance practices on employee job performance at Eco Bank Kenya. *European Journal of Business and Management, 5*(25), 179-185.
[21] Khallash, S., and Kruse, M. (2012). The future of work and work-life balance 2025. *Futures, 44*, 678-686. https://doi.org/10.1016/j.futures.2012.04.007
[22] Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1-12. https://doi.org/10.1177/2053951714528481
[23] Kumari, T. K., and Devi, R. V. (2013). Work-life balance of women employees: A challenge for the employee and the employer in the 21st century. *Pacific Business Review International, 6*(6), 33-40. https://doi.org/:10.1504/IJSOM.2018.10009105
[24] Lakshmi, K. S., and Gopinath, S. S. (2013). Work life balance of women employees – with reference to teaching faculties. *Abhinav International Monthly Referred Journal of Research in Management and Technology, 2*(3), 53-62.
[25] Mayer-Schönberger, V., and Cukier, K. (2013). Big Data: A Revolution that will Transform how we Live, Work, and Think. New York, New York: Houghton Mifflin Harcourt.
[26] McCarthy, A., Darcy, C., and Grady, G. (2010). Work-life balance policy & practice: Understanding line manager attitudes & behaviors. *Human Resource Management Review, 20*, 158-167. https://doi.org/10.1016/j.hrmr.2009.12.001
[27] Nunan, D., and Di Domenico, M. (2017). Big data: A normal accident waiting to happen?. *Journal of Business Ethics, 145*(3), 481-491. https://doi.org/10.1007/s10551-015-2904-x
[28] Raheef, O. O. (2019). Towards ensuring proper management of recruitment, selection and induction processes within business organizations. *Social Science and Humanities Journal (SSHJ)*, 1368-1378.
[29] Raheef, O. O. (2019). Towards Ensuring Proper Management of Recruitment, Selection and Induction Processes Within Business Organizations. *Social Science and Humanities Journal (SSHJ)*, 1368-1378.
[30] Richert-Kazmierska, A., and Stankiewicz, K. (2016). Work-life balance: Does age matter?. *Work, 55*(3), 1-10. https://doi.org/10.3233/WOR-162435
[31] Sajjad, M., and Nas, Z. (2013). Arising employees' retentions due to organizational commitment & support in pharmaceutical industry. *Management arts, 56*, 13407-13412.
[32] Santhana, L. K., and Gopinath, S. (2013). Work life balance of women employees-with reference to teaching faculties. *International Monthly Refereed Journal of Research in Management and Technology, 2*(3), 53-62.
[33] Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values?. *Journal of social issues, 50*(4), 19-45.
[34] Singh, S. (2013). Work-life balance: A literature review. *Global Journal of Commerce & Management Perspective, 2*(3), 84-91. https://doi.org/10.1108/XJM-10-2020-0186
[35] Sudha, J., and Karthikeyan. P. (2014). Work life balance of women employee a literature review. *International Journal of Management Research & Review, 4*(8), 797-804.
[36] Syed, J. (2015). Work-life balance. In Syed, J., & Özbilgin, M. (eds.), *Managing Diversity and Inclusion* (pp. 287-310). London: Sage.
[37] Thriveni, K. K., and Rama, D. V. (2012). Impact

of demographic variables on work-life balance of women employees (with special reference to Bangalore City). *International Journal of Advances in Management and Economics, 1*(6), 226-229.

[38] Yawalkar, V. V., and Sonawane, M. A. (2017). Impact of demographic variables on work-life balance of police personnel: With reference to Jalgaon police department. *International Journal of Science, Engineering and Management, 2*(9), 29-32.

◆ About the Authors ◆

Youngkeun Choi

Youngkeun Choi is an associate professor in the School of Management at Sangmyung University, Seoul Korea. Before joining Sangmyung University, he was working at Samsung corporation as a business developer, Atinum Investment and Neoplux as a venture capitalist, and POSCO research institute as a business data scientist.

Submitted: October 9, 2021; 1st Revision: April 4, 2023; 2nd Revision: August 6, 2023; Accepted: November 27, 2023