

SALDA-ML: Machine Learning Based System Design to Predict Salary Increase

Yasin Görmez^{1,*}, Halil Arslan², Suat Sarı³, Mücahit Daniş³

¹ Sivas Cumhuriyet University, Management Information Systems, Sivas, Turkey

² Sivas Cumhuriyet University, Computer Engineering, Sivas, Turkey

³ Detaysoft, İstanbul, Turkey

Abstract

Number of employees are increases with growing in companies. Firms basically make salary increases for their employees in order not to lose their talents and moreover to increase them. Although there is not much problem in how to in-crease the salary in small organizations, this process should be carried out carefully in terms of many parameters in large organizations and should not result in negativities that may disrupt employee motivation. For companies with a large number of employees, creating a model in which the market conditions are determined correctly and all economic parameters are taken into account reveals the need for a process that needs to be worked on for months. In this context, a machine learning-based salary increase prediction system was designed with the study. Specific attributes were determined and a specific scale was developed for performance score for this study.

Keywords: Salary Increase, Machine Learning, Work Flow, Personnel Performance Measurement, System Design, Software Develop

1. Introduction

There is an increase in the human resources with the growth of companies. Personnel management and keeping personnel motivation high are very important for these companies to execute their business process properly. According to the research, one of the important factors affecting the high motivation is that the personnel think they are getting the salary they deserve [1]. Salaries in companies in Turkey are usually determined annually. Determining the most appropriate salary increase every year is one of the important issues for these companies. Methods such as collective agreement, opinion of the person manager, and increase according to the rate determined by the governments due to inflation can be preferred in salary increase. However, none of these methods is sufficient on its own to determine proper salary for employees. According to the research many factors, which also include gender, effect the salary increase [2]. One of the most important problems of collective bargaining or inflationary salary increase methods is that the personnel who do more work and those who do less work cannot be distinguished. It has been observed that, personnel with a high capacity in the companies that applied these methods for a long time lost their motivation over time. Therefore, they started to do less work or changed jobs [3]. In addition, inflation and living conditions in a country may differ for each region. When all these cases are examined, it is clear that salary increase is a problem that does not have a linear solution depending on many factors.

Within the scope of the study, it is aimed to design a system based on machine learning that can predict personnel salary increase. During the system design phase, the experience of our company, which has been doing business with holding-level institutions around the world for more than 20 years, was utilized. In this context, taking into account our previous studies, it is aimed to use the inflation rate in the country, the inflation rate in the region, regional information, personnel level, exchange rate information, company policy, business area multiplier and performance score as attributes for the machine learning model. Data from consulting firms such as Korn Ferry (Hay Group) that collected market analysis data as a result of surveys were used to create these features [4]. The performance score is an indicator that represents the efficiency of the personnel during the year. A novel scale was developed to calculate this score that specific to this study. This performance measure will be calculated using data that are collected from enterprise resource planning (ERP) databases. In this context, a table will be designed in the ERP database that is compatible with personnel work tracking systems. Table fields are explained in detail in the following parts of the study. In addition, the system salary increase rate also differs on the basis of the sector [5]. Therefore, this system will be developed using micro-service architecture to make it modularize it. In this way, different machine learning model can be developed for different companies. If a new company wants to use the system, a company-specific model will be trained using the historical data of that company. Virtualization will be used to provide all of these attributes. In addition to all of these, system will be improved with correction by taking opinion of the experts.

*Corresponding author

E-mail address: yasingormez@cumhuriyet.edu.tr

2. Literature Review

Various studies have been conducted in the literature on the effects of salary on individuals. Kathawala et al. conducted a study on the preference for increased salary or job security for automotive industry workers. As a result of the research, it has been determined that white-collar personnel prefer salary increase rather than job security and personnel who prefer salary increase are less satisfied with their salary [6]. Mohan and Muthuswamy showed that people tend to leave their jobs in organizations with insufficient salary increase or no promotional payments [7]. Indrasari et al. analyzed the relationship between salary satisfaction, job satisfaction and organizational commitment. They concluded that job satisfaction has a positive effect on organizational commitment, while job satisfaction, salary satisfaction and organizational commitment have a negative effect on quitting [8].

In addition to these studies, several machine learning models developed to predict salary increase. Khongchai and Songmuang compared k-nearest neighbor (k-nn), naïve Bayes (NB), decision trees (DT), artificial neural networks (ANN) and support vector machines (SVM) to measure of their performance on estimating salary of students after graduation. According to results 84.69%, 43.63%, 73.96%, 38.08%, 43.71% accuracies were obtained with k-nn, NB, J48 DT, ANN and SVM respectively [9]. Wang et al. developed a machine learning system using bi-directional long short-term memories (bi-LSTM) and convolutional neural networks (CNN) to predict salary on dataset that generated using job search engine and they obtained better results than the literature [10]. Martin et al. analyzed machine learning methods to predict salary of employees that works in the field of information technology in Spain and they showed that the best accuracy, which is 84%, was obtained with decision tree algorithm [11]. Bansal et al. compared simple linear regression and multiple linear regression to predict personnel salary and they showed that multiple linear regression was obtained better accuracy results [12]. Das et al designed a system to estimate personnel salary after specific time and they shared the visual results of that system [13]. Li et al. predict the salary of employees using support vector regression, nearest geometric center, linear regression, logistic regression, k nearest regression and random forest regression on dataset that generated using job posting in England and they obtained 0.184% mean absolute error [14]. Zhang and Cheng obtained 88.10% accuracy for salary range estimation with k-nn for Java back-end developer using knowledge in basic Java programming, database principle, Java web score, system programming and Linux course and education information [15]. Mobasshera et al. developed a system using adaptive network-based fuzzy inference to predict salary increase and they made estimates very close to the real increase [16]. Viroon-luecha and Kaewkiriya obtained 0.774 root mean square error with deep learning approach for estimation of monthly salary using job posting data in Thailand [17].

3. Proposed System

In this study, a system, which can work integrated with ERP system, was designed to predict salary increase that uses inflation rate in the country, inflation rate in the region, region information, personnel level, exchange rate information, company policy, business multiplier and performance score a feature. Firstly, a performance score should be calculated for this system. In order to calculate score, a job tracking table must first be created. The table can be created in the existing business tracking systems of the institutions or can be integrated into the existing business tracking systems when necessary. This table contains columns for job identifier number, estimated job duration (man-day), job value ratio, job assignment date, and job closure date. The job value ratio represents a real number between zero and one determined by the manager according to the difficulty of the assigned job. If this value is close to one, it means that the task is more difficult. The score calculation for each job assigned to the personnel will be calculated using the formula in equation 1, and the total score of the personnel will be calculated using the formula in equation 2.

$$ts_{ij} = \frac{\{(estimated\ work\ time\ | \ present\ day - job\ assignment\ date)\} \times job\ value\ ratio}{total\ workers\ in\ assignment} \quad (1)$$

$$s_j = \sum_{i=1}^N ts_{ij} \quad (2)$$

In equation 1, while calculating $(estimated\ work\ time\ | \ present\ day - job\ assignment)$ part, the value will be *estimated work time* if job is finished, and if not, the value will be *present day - job assignment*. For example if *estimated work time* is 2 months and the starting date of job is first of January, $(estimated\ work\ time\ | \ present\ day - job\ assignment)$ part will be 30 where ts_{ij} calculated in first of February and it will be 60 where ts_{ij} calculated after first of March. Later on, calculated performance score will be used as a feature for proposed system. In the first example, if performance score calculated after March, job value ratio is 0.7 (difficulty level of the job) and 4 employees assigned for this job, ts_{ij} will be $(60 \times 0.7) / 4 = 10.5$. This means a performance score for this job calculated for each employee in the job. Performance score of employees will be computed as sum of each performance score calculated from their tasks. Figure 1

shows the architecture of proposed system.

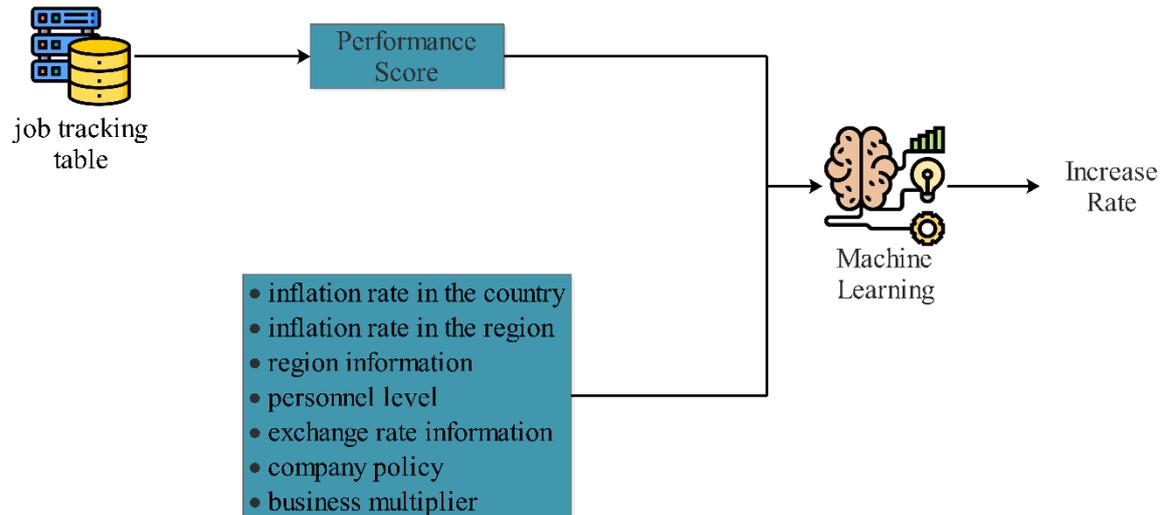


Figure 1. Proposed System Architecture

In figure 1, performance score represents value that calculated using equation 1 and equation 2. Remaining features will be taken from the database that integrated with ERP systems. Each attributes in this databases determined according to analysis and research [4]. Each field can be expressed in detail as:

- Inflation rate in the country: Inflation rate that covers the entire country, announced monthly or annually.
- Inflation rate in the region: Inflation rate in the province where the institution is located. This information was needed because the inflation in the region may differ from the inflation in the country.
- Region information: Information on the province where the institution is located. It is known that basic needs expenditures (food and beverage, transportation, rent, etc.) differ on a provincial basis and these expenditures have a direct related with salary. Therefore, this information is needed.
- Personnel Level: The position of the personnel is one of the most important factors affecting the salary increase. This effect becomes much more important especially in level changes. Considering the mentioned factors, it was concluded that this parameter value should be used in the model.
- Exchange rate information: Exchange rate information at the time of salary increase. Since exchange rate information is one of the important factors (especially for countries with a current account deficit) that determine the price of basic needs or secondary needs, it is considered to be in the model.
- Company policy: It represents a multiplier between 0 and 1 that the company has created for its own policies. For example, the company may want to raise more or less than inflation this year. If this multiplier is close to 1, it means that there will be more raises.
- Business multiplier: It represents the value between 0 and 1, which indicates the coefficient of the work area (information processing, human resources, etc.) in which the personnel works. If this number is close to 1, it represents a further increase in that area.

Considering the structure of the system and the number of parameters in the system, it is seen that there is a problem that cannot be solved linearly. Therefore, it is thought that, machine learning models, which were obtained accurate results in many field such as sentiment analysis [18] – [20], object recognition from image [21] – [23] and bioinformatics [24] – [26], can be used to predict salary increase with the least error. In this context, it's aimed to use methods based on machine learning in the last stage of the proposed system. It is possible for different machine learning models to yield more successful results according to the data structure of each institution and the number of data collected. Before installation for the intuitions, linear regression, artificial neural networks, random forest regression and deep learning approaches will be analyzed to determine most suitable models for that intuition. Then, a final model will be trained using all the data in the institution and this model will be used as the decision maker of the system.

Considering today's technological level and fast living conditions, the needs of institutions can change very quickly. Micro-service architecture was used to design the system considering this situation. Each module in the system will be virtualized and the modules will be able to communicate with each other via service

intermediary. In this way, it will be possible to add modules to the system in line with new requirements. In addition to all these, the developed machine learning model will be developed in structures that enable it to be developed with the online method, in line with expert opinions.

4. Conclusions

In the study, the relationship between the salary of the personnel and their motivation and productivity was analyzed with the literature review and salary prediction system in the literature was analyzed. As a result of these analysis, it has been concluded that the salary increase rate estimation can be made with machine learning methods. In this context, a salary increases estimation system based on machine learning, which can also be integrated into ERP systems, has been designed for institutions and organizations. The study specific features were determined and performance score were concatenated with these features. A study-specific scale was developed to calculate the performance score. In order to develop the designed system to meet future corporate needs, it is thought to use a microservice architecture. In addition, structures that allow the machine learning model to update itself in line with the needs have been designed.

In the future, it's aimed to install proposed system in to company from information technology sector for testing it. For this purpose, web services were developed according to features in this study to collect data from task tracking systems. In this way, the study-specific data will be collected. Therefore, contribution will be made the literature with this dataset. Machine learning model will be analyzed on this datasets and best accurate model will be determined. Then, the system will be installed in the relevant institution, and the operating performance of the system will be measured with real-time data, and system load tests will be carried out. For the mentioned reasons, it is thought that the system will be a study that will be open to development and will be the basis for machine learning-based systems.

Acknowledgment

This study is an output of studies conducted in DetaySoft research and development center. We appreciate their support.

Declaration of interest

It was presented as a summary at the ICAIAME 2021 conference.

References

- [1] D. Card, A. Mas, E. Moretti, and E. Saez, 'Inequality at Work: The Effect of Peer Salaries on Job Satisfaction', *Am. Econ. Rev.*, vol. 102, no. 6, pp. 2981–3003, Oct. 2012, doi: 10.1257/aer.102.6.2981.
- [2] R. A. Thacker, 'Gender, influence tactics, and job characteristics preferences: New insights into salary determination', *Sex Roles*, vol. 32, no. 9, pp. 617–638, May 1995, doi: 10.1007/BF01544215.
- [3] J. R. Hill and N. A. Jolly, 'Salary Distribution and Collective Bargaining Agreements: A Case Study of the NBA', *Ind. Relat. J. Econ. Soc.*, vol. 51, no. 2, pp. 342–363, 2012, doi: <https://doi.org/10.1111/j.1468-232X.2012.00680.x>.
- [4] 'Korn Ferry | Organizational Consulting'. <https://www.kornferry.com> (accessed Mar. 31, 2021).
- [5] K. S. Koong, L. C. Liu, and R. Fowler, 'Salaries of information technology managers; A trend analysis', p. 15, 2003.
- [6] Y. Kathawala, K. J. Moore, and D. Elmuti, 'Preference between Salary or Job Security Increase', *Int. J. Manpow.*, vol. 11, no. 7, pp. 25–31, Jan. 1990, doi: 10.1108/01437729010004174.
- [7] S. Mohan and P. R. Muthuswamy, 'A study on employee retention in BPO Sector with special reference to Coimbatore', *Int. J. Inf. Futur. Res.*, vol. 2, no. 6, 2015.
- [8] M. Indrasari, B. R. Purnomo, E. Yunus, E. Haryati, and A. R. Hashmi, 'The Effect of Salary Satisfaction, Work Satisfaction and Organizational Commitment To Work Intention', *J. Didask.*, vol. 1, no. 1, Art. no. 1, Nov. 2018, doi: 10.33856/didaskalia.v1i1.54.
- [9] P. Khongchai and P. Songmuang, 'Implement of salary prediction system to improve student motivation using data mining technique', in 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Nov. 2016, pp. 1–6. doi: 10.1109/KICSS.2016.7951419.
- [10] Z. Wang, S. Sugaya, and D. P. T. Nguyen, 'Salary Prediction using Bidirectional-GRU-CNN Model', p. 4, 2019.
- [11] I. Martín, A. Mariello, R. Battiti, and J. A. Hernández, 'Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study', *Int. J. Comput. Intell. Syst.*, vol. 11, no. 1, pp. 1192–1209, Jul. 2018, doi: 10.2991/ijcis.11.1.90.
- [12] U. Bansal, A. Narang, A. Sachdeva, I. Kashyap, and S. P. Panda, 'Empirical analysis of regression techniques by house price and salary prediction', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, p. 012110, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012110.
- [13] S. Das, R. Barik, and A. Mukherjee, 'Salary Prediction Using Regression Techniques', *Social Science Research*

- Network, Rochester, NY, SSRN Scholarly Paper ID 3526707, Jan. 2020. doi: 10.2139/ssrn.3526707.
- [14] L. Li, X. Liu, and Y. Zhou, 'Prediction of Salary in UK', p. 5.
- [15] J. Zhang and J. Cheng, 'Study of Employment Salary Forecast using KNN Algorithm', Aug. 2019, pp. 166–170. doi: 10.2991/msbda-19.2019.26.
- [16] A. Mobasshera, K. Naher, T. M. Rezoan Tamal, and R. M. Rahman, 'Salary Increment Model Based on Fuzzy Logic', in *Artificial Intelligence and Algorithms in Intelligent Systems*, Cham, 2019, pp. 344–353. doi: 10.1007/978-3-319-91189-2_34.
- [17] P. Viroonluecha and T. Kaewkiriya, 'Salary Predictor System for Thailand Labour Workforce using Deep Learning', in *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, Sep. 2018, pp. 473–478. doi: 10.1109/ISCIT.2018.8587998.
- [18] R. Chiong, Z. Fan, Z. Hu, M. T. P. Adam, B. Lutz, and D. Neumann, 'A Sentiment Analysis-based Machine Learning Approach for Financial Market Prediction via News Disclosures', in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, New York, NY, USA, 2018, pp. 278–279. doi: 10.1145/3205651.3205682.
- [19] Ankit and N. Saleena, 'An Ensemble Classification System for Twitter Sentiment Analysis', *Procedia Comput. Sci.*, vol. 132, pp. 937–946, Jan. 2018, doi: 10.1016/j.procs.2018.05.109.
- [20] J. E. Dayhoff and J. M. DeLeo, 'Artificial neural networks', *Cancer*, vol. 91, no. 8, pp. 1615–1635, Apr. 2001, doi: 10.1002/1097-0142(20010415)91:8+<1615::AID-CNCR1175>3.0.CO;2-L.
- [21] G. Schaefer, 'ACO classification of thermogram symmetry features for breast cancer diagnosis', *Memetic Comput.*, vol. 6, no. 3, pp. 207–212, Sep. 2014, doi: 10.1007/s12293-014-0135-9.
- [22] N. F. Abubacker, A. Azman, S. Doraisamy, and M. A. A. Murad, 'An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification', *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3967–3980, Dec. 2017, doi: 10.1007/s00521-016-2290-z.
- [23] J. de N. Silva, A. O. de C. Filho, A. C. Silva, A. C. de Paiva, and M. Gattass, 'Automatic Detection of Masses in Mammograms Using Quality Threshold Clustering, Correlogram Function, and SVM', *J. Digit. Imaging*, vol. 28, no. 3, pp. 323–337, Jun. 2015, doi: 10.1007/s10278-014-9739-3.
- [24] O. Manor and E. Segal, 'Predicting Disease Risk Using Bootstrap Ranking and Classification Algorithms', *PLOS Comput. Biol.*, vol. 9, no. 8, p. e1003200, Aug. 2013, doi: 10.1371/journal.pcbi.1003200.
- [25] C. Kooperberg, M. LeBlanc, and V. Obenchain, 'Risk prediction using genome-wide association studies', *Genet. Epidemiol.*, vol. 34, no. 7, pp. 643–652, doi: 10.1002/gepi.20509.
- [26] M.-S. Cheung, M. L. Maguire, T. J. Stevens, and R. W. Broadhurst, 'DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure', *J. Magn. Reson.*, vol. 202, no. 2, pp. 223–233, Feb. 2010, doi: 10.1016/j.jmr.2009.11.008.