International Journal of application or Innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

ML ALGORITHMS FOR EMPLOYEE PROMOTION PREDICTION IN UNBALANCED DATASET

T.Sandeep Reddy [1], B.Vijay Laxman [2] , MD.Faisal[3], K.Yadagiri[4]

[1,2,3]UG Scholar, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

[4]Assistant Professor,Department of IT, St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

Tsandeepreddythurpu1211@gmail.com

Article Info

Received: 28-03-2025 Revised: 05 -04-2025 Accepted: 16-04-2025 Published:27/04/2025

Abstract:

Predicting employee performance is essential for organizations. The success or failure of a company often depends upon the competence of its employees, so CEO' s and managers who want their organizations to succeed face the difficult task of determining which employees should be promoted. The current promotion process used in most organizations should be considered misleading because it depends on supervisors judgments. The major aim of this paper is to use classification algorithms to develop predictive models for predicting whether an employee is qualified for promotion or not and identifying the most important attributes affecting employee promotion. The data set used in this paper is from Kaggle 2020. It contains information on multinational companies arranged in 54,808 rows and 13 columns. This data set covers nine broad verticals across organizations. Several predictive modeling techniques,including K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Ensemble models (Ada boosting and Gradient Boosting models) were used to predict employee promotion. Based on measurements of accuracy, F1-score, and AUC,Gradient Boosting outperforms the other classification algorithms. The results also show that the most significant factor contributing to predicting employee promotion is the previous years rating. Department had no effect on employee promotions

Keywords: Employee Performance, Promotion Prediction, Classification Algorithms, Predictive Models, Machine Learning, Supervisors' Judgments, Kaggle Dataset, Multinational Companies, K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Ensemble Models

INTRODUCTION

Promotion is one of the trickiest issues in every worker's life. When an employee is promoted, they are given more responsibility and sometimes a raise in salary. The company's success and the workers' motivation and loyalty both increase with proper implementation of the promotion system. There is a strong correlation between years of service and pay grade. Establishing and consistently executing a transparent and objective promotion strategy is crucial to the success of the company's career management efforts. All employees should be made aware of the promotion process in advance, including who will make promotions, how often, and what kind of experience and education is needed for each level of advancement. Promotion is often accompanied by increasing responsibilities, obligations, privileges, and authority. Promotions, according to research [1-2], have a considerable and positive effect on workers' productivity. When an employee receives a promotion, they move up in the company by being assigned new duties that give them more responsibility and higher standing.

Workers will be more motivated by the knowledge that they will

eventually be in a better position than they are currently in the institution. As the worker's sense of dedication to and effectiveness in their work is bolstered by the knowledge that their efforts will be recognized and rewarded, everyone benefits. When this occurs, gains may be made in output, effectiveness, and quality. The corporation will be able to keep producing at a healthy profit thanks to this. Therefore, everyone wins, both the company and the worker. When a person is promoted and given better perks, they are more likely to be content with their job and dedicated to doing a good job. Staff morale is higher, productivity is higher, and turnover is lower when workers are happy [4]. Increased duties, facilities, accomplishments, greater qualification needs, higher prestige, and earnings or wages are all indicators of a work promotion, as stated by Dean and Joseph [5]. Boost your income, social standing, and sense of accomplishment while also improving your own life. Optimism, self-control, and output in the workplace are on the upswing. Maintain consistent staffing, fairly assess raises and promotions, review workers promptly and openly, and so forth.

Workforce advancement possibilities have various impacts on firms as a result of emerging professions. Creating conditions where workers may develop their innovative abilities to the advantage of the company. The desire to learn and grow in one's position is what drives many other workers. Anew organizational structure is being established as a result of a mutation in the position. Employee performance heavily influences career advancement. Given the favorable correlation between promotion and both employee performance and corporate output, he argued that such concepts should be codified into company policy. Transfer the job to another employee so that it doesn't go unfilled.When workers are given promotions that make sense for them, they are happier and more productive in their roles[6-9]. To broaden the pool of qualified candidates for open positions.

LITERATURE SURVEY

Software Employee Promotion Analysis Using Machine Learning Ying-Yi Hong, Rolando A.Pula Employee attrition is the term used to

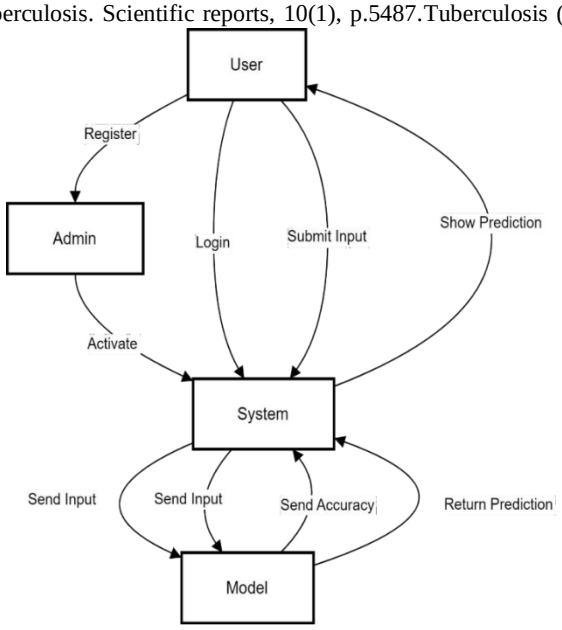
describe the organic decline in the number of employees in a company as a result of several unavoidable circumstances. Employee churn causes a significant lossoran organization, a loss. According to the Society for Human Resource Management (SHRM), that is the typical cost per hire for a new hire. Recent statistics indicate that the attrition rate in 2021will be57.3%.The accuracy scores obtained using the deployed machine learning approaches were 87%by SVM methodology, and 93% overall. This project is focused on gathering information on employees, creating a decision tree using historical data, testing the decision tree using an employee's traits, and determining whether to provide a promotion or not. The trained dataset kept in the decision tree is compared to this data. Identifying is the ultimate objective node. The suggested improved Decision Trees Classifier(DTC) predicts whether the employee will receive a yearly raise or promotion or not. Thetechnique produced predictions of staff attrition that were up to 96% accurate

Acharya, B., Acharya, A., Gautam, S., Ghimire, S.P., Mishra, G., Para- juli, N. and Sapkota, B., 2020. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. Molecular biology reports, 47, pp.4065-4075.Tuberculosis (TB) is a major cause of deaths by a single infectious agent and has now been a global public health problem due to increasing numbers of drug-resistant cases. Early and effective treatment is crucial to prevent the emergence of drug-resistance strains. This demands the availability of fast and reliable point-of-care (POC) diagnostic methods for effective case management. Commonly used methods to screen and diagnose TB are clinical, immunological, microscopy, radiography, and bacterial culture. In addition, recent advances in molecular diagnostic methods including MTBDRplus, loop-mediated isothermal amplification (LAMP), line probe assay (LPA), GeneXpert, and whole genome sequencing (WGS) have been employed to diagnose and characterize TB. These methods can simultaneously identify Mycobacterium tuberculosis (MTB) and mutation(s) associated with routinely used anti-TB drugs. Here, we review the use of currently available diagnostic methods and strategies including conventional to recently implemented next-generation sequencing (NGS) methods used to detect MTB in clinical perspective.

Ranjitha, J., Rajan, A. and Shankar, V., 2020. Features of the bio- chemistry of Mycobacterium smegmatis, as a possible model for Mycobacterium tuberculosis. Journal of infection and public health, 13(9), pp.1255-1264.Objective Actinomycetes have been known to be the great natural sources to explore antibiotics for the treatment of tuberculosis (TB). The isolation of actinomycetes from the samples in Vietnam followed by the screening of their antimycobacterial activity was performed in this study. The metabolites isolated from the most active strain were further evaluated for their antimycobacterial, antimicrobial and cytotoxic activity. Methods Actinomycetes were growth in culture media, isolated and identified by colony, spore chain morphology and 16S rRNA gene sequencing. Agar diffusion assay was used for the screening of the isolated strains against Mycobacterium smegmatis, a safety surrogate for Mycobacterium tuberculosis. The metabolites produced from the most active strain were investigated by actinomycete fermentation, extraction and isolation from biomass and cultures. The structures of the isolated compound were elucidated by spectral data and comparison with the reported literatures. Results 181 strains were isolated from nine regions along the north to central Vietnam. The five most active strains against Mycobacterium smegmatis were detected. Following the bioassay-guided result, the strain A121 (Streptomyces alboniger) was selected for further isolation of the bioactive metabolites.

since drug-resistant TB (DR-TB) accounts for ~29% of deaths attributable to AMR. In recent years, funded collaborative efforts of researchers from academia, not-for-profit virtual R&D organizations and industry have resulted in the continuous growth of the TB drug discovery and development pipeline. This has so far led to the accelerated regulatory approval of bedaquiline and delamanid for the treatment of DR-TB.

Jamal, S., Khubaib, M., Gangwar, R., Grover, S., Grover, A. and Hasnain, S.E., 2020. Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis. Scientific reports, 10(1), p.5487.Tuberculosis (TB), an



infectious disease caused by Mycobacterium tuberculosis (M.tb), causes highest number of deaths globally for any bacterial disease necessitating novel diagnosis and treatment strategies. High-throughput sequencing methods generate a large amount of data which could be exploited in determining multi-drug resistant (MDR-TB) associated mutations. The present work is a computational framework that uses artificial intelligence (AI) based machine learning (ML) approaches for predicting resistance in the genes rpoB, inhA, katG, pncA, gyrA and gyrB for the drugs rifampicin, isoniazid, pyrazinamide and fluoroquinolones. The single nucleotide variations were represented by several sequence and structural features that indicate the influence of mutations on the target protein coded by each gene. We used ML

International Journal of application or Innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

algorithms - naïve bayes, k nearest neighbor, support vector machine, and artificial neural network, to build the prediction models. The classification models had an average accuracy of 85% across all examined genes and were evaluated on an external unseen dataset to demonstrate their application. Further, molecular docking and molecular dynamics simulations were performed for wild type and predicted resistance causing mutant protein and anti-TB drug complexes to study their impact on the conformation of proteins to confirm the observed phenotype.

PROPOSED METHODOLOGY

The proposed system utilizes advanced machine learning algorithms to predict employee promotions based on objective and quantifiable data, including performance ratings, training scores, and years of service.
Traditional promotion methods often rely on subjective assessments by supervisors, which can introduce biases and inconsistencies. In contrast, this system leverages data-driven approaches to ensure a more accurate, fair, and unbiased decision-making process.

By employing machine learning models such as Gradient Boosting, Random Forest, and Logistic Regression, the system analyzes historical employee data to identify meaningful patterns and trends that indicate an employee's readiness for promotion. These models are trained on large datasets containing various attributes related to employee performance, enabling them to make well-informed predictions.

The implementation of this machine learning-based promotion system significantly accelerates the decision-making process by automating the evaluation of employees. This not only reduces the time and effort required for promotions but also ensures a standardized and consistent evaluation framework across the organization. Furthermore, by eliminating human biases and subjective judgments, the system enhances transparency in the promotion process, fostering trust among employees and improving overall workplace morale.

In addition to improving fairness and efficiency, the system contributes to organizational growth by ensuring that promotions are granted to the most deserving candidates based on merit and performance. As a result, employee motivation and satisfaction increase, leading to better productivity and a more engaged workforce. Ultimately, integrating machine learning into the promotion process strengthens an organization's ability to retain top talent and build a more competent leadership pipeline.

By employing machine learning models such as Gradient Boosting, Random Forest, and Logistic Regression, the system analyzes historical employee data to identify meaningful patterns and trends that indicate an employee's readiness for promotion. These models are trained on large datasets containing various attributes related to employee performance, enabling them to make well-informed predictions

Figure 1: Proposed System Architecture

By employing machine learning models such as Gradient Boosting, Random Forest, and Logistic Regression, the system analyzes historical employee data to identify meaningful patterns and trends that indicate an employee's readiness for promotion. These models are trained on large datasets containing various attributes related to employee performance, enabling them to make well-informed predictions.

The implementation of this machine learning-based promotion system significantly accelerates the decision-making process by automating the evaluation of employees. This not only reduces the time and effort required for promotions but also ensures a standardized and consistent evaluation framework across the organization. Furthermore, by eliminating human biases and subjective judgments, the system enhances transparency in the promotion process, fostering trust among employees and improving overall workplace morale.

In addition to improving fairness and efficiency, the system contributes to organizational growth by ensuring that promotions are granted to the most deserving candidates based on merit and performance. As a result, employee motivation and satisfaction increase, leading to better productivity and a more engaged workforce. Ultimately, integrating machine learning into the promotion process strengthens an organization's ability to retain top talent and build a more competent leadership pipeline.

By employing machine learning models such as Gradient Boosting, Random Forest, and Logistic Regression, the system analyzes historical employee data to identify meaningful patterns and trends that indicate an employee's readiness for promotion. These models are trained on large datasets containing various attributes related to employee performance, enabling them to make well-informed predictions

EXPERIMENTAL ANALYSIS

International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Figure 2: Home Page

Figure 3: Registration Page

Figure 4: Admin Page

Fig 5 : User Login Page

International Journal of application or Innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Figure 6: User Home Page

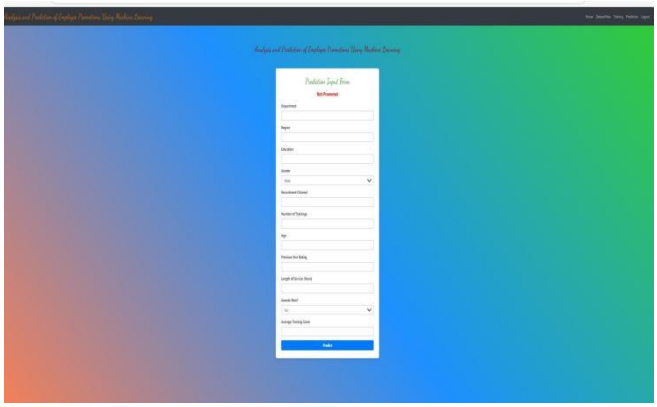


Figure 7: Final Prediction Page

CONCLUSION

This study has highlighted the significance of predicting employee promotions as a crucial aspect of human capital development. Ensuring that the right employees are promoted based on data-driven insights can lead to better workforce management, improved employee satisfaction, and increased organizational efficiency. To achieve this, we developed and evaluated three well-known machine learning algorithms for predicting whether an employee would receive a promotion. The goal was to create a reliable, automated decision- making system that minimizes bias and improves the accuracy of promotion processes within organizations.

To build and validate our predictive models, we utilized real-world company data, ensuring that the dataset accurately reflected employee characteristics, performance metrics, and other relevant factors. This dataset was carefully preprocessed and structured for training the machine learning models, allowing us to conduct a thorough analysis of how different factors influence promotion decisions.

In this quantitative study, we employed common supervised machine learning techniques, rigorously testing their effectiveness in predicting promotions. Each model was evaluated using key performance metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) to assess their predictive capabilities. Our results demonstrated that, while there were minor discrepancies between the models, all three machine learning algorithms effectively made accurate predictions regarding employee promotions.

To further validate our findings, we developed specialized software that allowed us to test the trained models on fresh, unseen data. This software provided a practical implementation of the machine learning models, enabling organizations to integrate predictive analytics into their promotion decision-making processes. By continuously improving and updating the system with new employee data, businesses can enhance the fairness, efficiency, and transparency of their promotion strategies, ultimately leading to better workforce management and long-term organizational success.

Predicting employee promotions using machine learning has become an essential aspect of human capital development, helping organizations identify and retain top talent. A data-driven approach reduces biases in promotion decisions, ensuring fairness and improving overall workforce efficiency. By analyzing factors such as performance ratings, training scores, and years of service, machine learning models can provide accurate predictions, helping HR departments make informed decisions.
Several machine learning techniques have been utilized for this purpose, including Gradient Boosting, Random Forest, and Logistic Regression. These models analyze historical employee data to identify patterns that indicate promotion readiness. Among these, Gradient Boosting has shown superior performance due to its ability to capture complex patterns and interactions within the data. The models were evaluated using key performance metrics such as accuracy, precision, recall, F1- score, and AUC (Area Under the Curve) to ensure their effectiveness in predicting employee promotions.

Key factors influencing promotions include an employee's past performance ratings, training scores, and years of service. Interestingly, the study found that the department an employee belongs to has little impact on promotion decisions. However, challenges remain in implementing these predictive models. Data quality issues, such as incomplete or biased historical records, can affect prediction accuracy. Additionally, if past promotions were influenced by favoritism, the model may inadvertently learn these biases. Moreover, soft skills like leadership and teamwork, which are difficult to quantify, may not be adequately captured by machine learning models.

To enhance the effectiveness of these predictive systems, future improvements could include integrating the models with HR management systems, employing deep learning techniques for better accuracy, and enabling real-time prediction updates. Additionally, implementing explainability techniques like SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) can help make predictions more transparent, allowing HR professionals to understand why certain employees are recommended for promotion. By continuously refining these models and incorporating new data, organizations can ensure a more efficient, fair, and transparent promotion process, ultimately driving long-term success

REFERENCES

[1] Long, Yuxi & Liu, Jm & Fang, Ming & Wang, Tao & Jiang, Wei. "Prediction of Employee Promotion Based on Personal Basic Features and Post Features". Proceedings of the International Conference on Data Processing and Applications (2018): 5-10. https://doi.org/10.1145/3224207.3224210

[2] J. Liu, T. Wang, J. Li, J. Huang, F. Yao and R. He. "A Data-driven Analysis of Employee Promotion: The Role of the Position of Organization". IEEE International Conference on Systems, Man and Cybernetics (SMC) (2019): 4056-4062. https://doi.org/10.1109/SMC.2019.8914449

[3] Piryonesi, S. Madeh; El-Diraby, Tamer E. (2020) "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of Infrastructure Systems: 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512.

[4] Breiman, L. "Technical Report 486". Statistics Department, University of California, Berkeley (1997): 11-13

[5] Ho, Tin Kam. "Random decision forests". Proceedings of the 3rd International Conference on Document Analysis and Recognition (1995): 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

[6] Ho TK. "The Random Subspace Method for Constructing Decision Forests". IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8) (1998): 832–844. https://doi.org/10.1109/34.709601.

[7] "About Keras", Official website of Keras Sequential Api, Google Keras team, accessed December 1, 2021, https://keras.io/about/.

[8] Deng, H.; Runger, G.; Tuv, E.. "Bias of importance measures for multivalued attributes and solutions". Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN) (2011): 293–300. https://doi.org/10.1007/978-3-642-21738-8_38. [9] Altmann A, Toloşi L, Sander O, Lengauer T. "Permutation importance: a corrected feature importance measure". Bioinformatics. 26 (10) (2010): 71 1340-7. https://doi.org/10.1093/bioinformatics/btq134. PMID 20385727. [10] Géron, A. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. (O'Reilly Media, 2017). 92-93

[11] .Adenuga, O. A. (2015). Employee ' s Salary , Gender , Length of Service and Job Involvement as Determinants o f Employees ' Performance in Nigerian Breweries Plc . 1(1), 20–25. [12] .Aksu, Ü., Schunselaar, D. M. M., & Reijers, H. A. (2019). Automated Prediction of Relevant Key Performance Indicators for Organizations. Lecture Notes in Business Information Processing, 353(May), 283–299. https://doi.org/10.1007/978-3-030-20485-3_22

[13] .Ameer, M., Rahul, S. P., & Manne, S. (2020). Human Resource Analytics using Power Bi Visualization Tool. Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, Iciccs, 1184–1189. https://doi.org/10.1109/ICICCS48265.2020.9120897

[14] .Chang, X., & Xue, J. (2020). Research on the Evaluation and Promotion of Employees from the Perspective of Competency. Open Journal of Social Sciences, 08(02), 99–108. https://doi.org/10.4236/jss.2020.82009

[15] .Improving business decision making based on KPI management system. 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC2017, 2017-January, 1280–1285. https://doi.org/10.1109/SMC.2017.8122789

[16] .Dianah, S., Bujang, A., Selamat, A., & Krejcar, O. (2021). A Predictive Analytics Model for Students Grade Prediction by Supervised Machine Learning. 1051, 1–9. https://doi.org/10.1088/1757-899X/1051/1/012005

[17] .Elnaga, A., & Imran, A. (2018). The effect of training on employee performance. International Journal of Recent Technology and Engineering, 7(4), 6–13. 34 https://doi.org/10.36555/almana.v4i3.1477

[18] Singh, D. K., Kediya, S., Mahajan, R., & Asthana, P. K. (2021, November). Management Information System in context of Food grains: An Empirical Study at Eastern Maharashtra. 2021 International Conference on Computational Intelligence and Computing Applications 72 (ICCICA) (pp. 1-5). IEEE

[19] Namasudra, S., Deka, G. C., Johri, P., Hosseinpour, M., & Gandomi, A. H. (2021). The revolution of blockchain: State-of-the-art and research challenges. Archives of Computational Methods in Engineering, 28,

[20] Piryonesi, S. Madeh; El-Diraby, Tamer E. (2020) "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index". Journal of Infrastructure Systems: 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512.

[21] Breiman, L. "Technical Report 486". Statistics Department, University of California, Berkeley (1997): 11-13

[22] Ho, Tin Kam. "Random decision forests". Proceedings of the 3rd International Conference on Document Analysis and Recognition (1995): 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

[23] Ho TK. "The Random Subspace Method for Constructing Decision Forests". IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8) (1998): 832–844. https://doi.org/10.1109/34.709601.

[24] "About Keras", Official website of Keras Sequential Api, Google Keras team, accessed December 1, 2021, https://keras.io/about/.

[25] Deng, H.; Runger, G.; Tuv, E.. "Bias of importance measures for multivalued attributes and solutions". Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN) (2011): 293–300. https://doi.org/10.1007/978-3-642-21738-8_38. [26] Altmann A, Toloşi L, Sander O, Lengauer T. "Permutation importance: a corrected feature importance measure". Bioinformatics. 26 (10) (2010): 71 1340-7. https://doi.org/10.1093/bioinformatics/btq134. PMID 20385727. [27] Géron, A. Hands-on machine learning with Scikit-

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. (O'Reilly Media, 2017). 92-93
[28] .Adenuga, O. A. (2015). Employee ' s Salary , Gender , Length of Service and Job Involvement as Determinants o f Employees ' Performance in Nigerian Breweries Plc . 1(1), 20–25. [29] .Aksu, Ü., Schunselaar, D. M. M., & Reijers, H. A. (2019). Automated Prediction of Relevant Key Performance Indicators for Organizations. Lecture Notes in Business Information Processing, 353(May), 283–299. https://doi.org/10.1007/978-3-030-20485-3_22
.