Machine Learning for KPI Development in Public Administration

Simona Fioretto[a], Elio Masciari[b] and Enea Vincenzo Napolitano[c]
Department of Electrical and Information Technology Engineering, University of Naples Federico II, Naples, Italy
fi

Keywords: Public Administration, Key Performance Indicators, Variable Importance, Machine Learning.

Abstract: Efficient and effective service delivery to citizens in Public Administrations (PA) requires the use of key performance indicators (KPIs) for performance evaluation and measurement. This paper proposes an innovative framework for constructing KPIs in performance evaluation systems using Random Forest and variable importance analysis. Our approach aims to identify the variables that have a strong impact on the performance of PAs. This identification enables a deeper understanding of the factors that are critical for organizational performance. By analyzing the importance of variables and consulting domain experts, relevant KPIs can be developed. This ensures improvement strategies focus on critical aspects linked to performance. The framework provides a continuous monitoring flow for KPIs and a set of phases for adapting KPIs in response to changing administrative dynamics. The objective of this study is to enhance the performance of PAs by applying machine learning techniques to achieve a more agile and results-oriented PAs.

1 INTRODUCTION

The success of an organisation depends on its ability to meet internal and external objectives. This involves the alignment of the mission and strategy of the organisation with the needs of its customers. In fact, once the needs of the customers are identified, they must be translated into organisational goals driving mission and strategy of the organization. Then, to evaluate the achievement of these goals, organisations require an objective measurement system. In fact, the ability of an organization of performing activities by pursuing efficiency and efficacy, is a measure of its performance results. Therefore, measuring performance is a complex and structured system. In fact, performance is a multifaceted phenomenon that requires integrated and simultaneous analysis of several indicators. Individual indicators often capture only a portion of the complexity of the organization, which instead is influenced by many variables. The identification of Key Performance Indicators (KPIs) aligned with the objectives of the organization, is crucial for assessing the performance in the organization and identify areas for improvement (Banu, 2018). These indicators use quantitative metrics to summarise information about specific phenomena of interest to stakeholders (Jahangirian et al., 2017). Indeed, to evaluate whether a process adheres to policies, meets deadlines, or is able to respect a fixed budget, it may be necessary to use a combination of multiple indicators. In the context of Public Administration (PA), the implementation of a proper performance measurement system can be fundamental to assure high quality services to citizens. However, the definition of KPIs in the PA sector is not as simple as it can be in private companies. In fact, the PAs significantly differ from the dynamics mechanism which are typical of the private sector. This is due to PAs characteristics. In fact, PAs differ from each other for the offered services, and for offices characteristics, such as the number of citizens served, the number of employees, and the level of office digitization (Kerzner, 2019). For instance, in justice sector or in education, only simple and measurable indicators are needed such as the required average time to resolve a legal case or graduation rates(Amato et al., 2023). However, these simple measures, which are called in the following macro-KPI, may not fully capture the quality level or fairness of the services provided. Additionally, PAs face challenges with bureaucracy and resistance to change. Administrative procedures can oppose to the adoption or modification of KPIs, even when they are no longer effective. The definition and interpretation of KPIs can also be heavily influenced by political environment, with changes in administration sometimes resulting in

[a] https://orcid.org/0009-0006-8700-8188
[b] https://orcid.org/0000-0002-1778-5321
[c] https://orcid.org/0000-0002-6384-9891

Fioretto, S., Masciari, E. and Napolitano, E.
Machine Learning for KPI Development in Public Administration.
DOI: 10.5220/0012820300003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 522-527
ISBN: 978-989-758-707-8; ISSN: 2184-285X

a complete restructuring of objectives and evaluation metrics. Accountability is another critical dimension in PA. Unlike the private sector, where accountability is primarily focused on financial results, in PAs the accountability is towards citizens. This is translated into the need of higher levels of transparency and communication, with resulting understandable KPIs by the public. One of the main issues in KPI definition, is the imprecise definition of objectives to which parameters should be aligned. If the objectives are unclear, the chosen parameters may not be relevant, resulting in the collection of meaningless data that do not provide insights into performance levels. In addition, markets and operating environments evolve quickly, and parameters previously defined may no longer be appropriate for the current necessities. However, the use of digital technologies in the context of PA can properly support the objective definition and monitoring of KPIs based on data. By leveraging the information contained in operational data, it is possible to define objectively KPIs, guaranteeing a more suitable performance management system. In this paper we aim to introduce a framework, which starting from macro-KPI, leverages data for identifying the specific micro-KPI. Micro-KPIs investigate and measure the variables results leading to the results of the macro-KPIs. In particular, can be leveraged the power of Machine Learning to select the most influential features, which can be used to properly define micro-KPIs which contribute to simpler macro-KPIs.

The paper is structured as follows: Section 2 introduces the theoretical concepts fundamental to our study. We discuss the nature and importance of KPIs and outline the characteristics of the chosen ML model. Section 3 reviews existing literature on the use of KPIs in PA and the construction of KPIs through machine learning techniques. Finally, in Section 4, we propose an innovative framework for constructing KPIs based on the use of machine learning. This framework aims to enhance the accuracy and relevance of KPIs used in PA by utilising the potential of Random Forest to analyse and interpret large volumes of data. The approach demonstrated in this text provides practical and meaningful insights for the evaluation and optimisation of PA processes.

2 THEORETICAL BACKGROUND

This section provides the theoretical background required throughout the paper, offering basic concepts of Key Performance Indicators (KPIs) and Random Forest. The theoretical background is essential to understand the innovations proposed in our research framework, which will be outlined in the following sections.

2.1 Key Performance Indicators

Organizations continuously set goals in order to achieve better results in terms of efficiency and efficacy. These goals are both a translation of the mission and the strategy of the organization. They need to be objectively monitored, in order to understand the status of their achievement. In fact, by monitoring their activities, organisations can determine whether or not they have achieved their objectives (Domínguez et al., 2019). The evaluation of goals achievement can be done by defining objective metrics, known as KPI. KPIs are a collection of crucial measures, both financial and non-financial, that are utilised to convert objectives into tangible measures. In details, the authors in (Domínguez et al., 2019) demonstrate that KPIs can provide organisations with reliable information to establish the basis for implementing growth strategies. KPIs can provide a way to see whether the strategic plan being adopted is working, serving as a tool to drive desired behaviours, and that their use can increase and improve operational efficiency, productivity and profitability. By establishing a set of KPIs, an organization can evaluate whether it has reached its goals (Velimirović et al., 2011). The relationship between the success of the organization and KPIs is evident, as they are closely linked to goals achievement.

2.2 Importance Factor for Random Forest

The Random Forest (RF) method (Parmar et al., 2019) is an ensemble learning technique for classification, regression, and other tasks. It constructs multiple decision trees during the training phase and outputs the mode of the classes (classification) or the mean of the predictions (regression) of the individual trees. The model's robustness is enhanced by its ability to withstand variation without significantly increasing bias, thanks to its natural ensemble. One significant contribution of RF is its ability to assess the importance of variables, known as feature importance, in the predictive model. This is typically calculated in two ways (Strobl et al., 2008):

1. **Importance Based on Decreasing Impurity:** is a method used to measure the importance of a variable in decision trees. It calculates how much the Gini index or entropy decreases due to the splits made on that variable. This method aggregates the total decrease in impurity attributable

to each variable across all forest trees, normally weighted by the number of observations passing through those splits.

2. **Importance of Allowed Variance:** This text evaluates the impact of a variable by mixing its values across observations in the test dataset. If there is a significant decrease in model performance after permutation, it indicates a high importance of the mixed variable. This is because its direct alteration deteriorates the model's ability to make accurate predictions.

These methods for evaluating variable importance are essential not only for optimizing RF models but also for providing insights into the characteristics that have the greatest impact on the target variable, thereby offering guidance for understanding and interpreting performance to define KPIs.

3 RELATED WORK

This section of the paper explores the literature and research related to identifying and developing KPIs in PA, as well as the use of Machine Learning techniques to improve these processes. The review is divided into two parts, reflecting the two main aspects of the research focus.

3.1 The Identification of Key Performance Indicators in Public Administration

The selection of an appropriate KPI must take into account several factors. Many studies address the definition and selection of KPIs in organizations, such as the proposal in (Parmenter, 2015), which suggests considering a variety of factors when selecting KPIs. Organizational characteristics must be taken into account, such as the identification of the appropriate KPI based on the Critical Success Factors (CSFs) of the organization. Secondly, when dealing specifically with PA, it is necessary to make further assumptions and link KPIs to the Balanced Scorecard (BSC) perspectives. In their work, the authors in (Parmenter, 2012) propose specific techniques for supporting PAs in identifying and selecting KPIs, providing a comprehensive methodology. In fact, while private organization are focused on profit and so on budget optimization, PAs are non-profit organizations. This led to a different perception of KPIs, as in PAs they need to measure variables related to the effectiveness of the organisation in providing high quality services and the efficiency of the organisation in optimising resources, which is not driven by profit optimisation.

3.2 Machine Learning to Develop KPIs

The application of Machine Learning (ML) techniques to the development and optimisation of KPIs gained significant attention in various fields, as it is shown by several recent studies. Each study uses different ML techniques and data sources to identify and predict KPIs that meet specific industry needs.

Using Google Analytics and ML techniques, Ahmed et al.(2017) in (Ahmed et al., 2017) attempted to establish a set of standard rules that must be employed to identify the best KPIs for an e-commerce business website. This study highlights the potential of ML to enhance the analytical capabilities of standard business tools and provide a structured approach to KPI development.

Fanaei et al. (2018) in (Fanaei et al., 2018) explored the application of various ML techniques to qualitatively predict overall project KPIs at critical project stages. They used methods such as artificial neural networks (ANN) and neuro-fuzzy techniques, integrating fuzzy C-means (FCM) and subtractive clustering to predict project KPIs. This comparative approach illustrates the versatility and robustness of ML in dealing with complex, diverse datasets typically found in project management.

Micu et al.(2019) in (Micu et al., 2019) used ML to analyse over a thousand e-commerce websites, with the aim of identifying KPIs capable of determining the success of the companies under consideration. Their study highlights the scalability of ML techniques in processing large datasets, and their utility in extracting meaningful insights across numerous domains.

El Haddad et al (2021) in (El Mazgualdi et al., 2021) presented the use of different ML algorithms under different configurations to predict Overall Equipment Effectiveness (OEE) and its application. This research demonstrates the adaptability of ML algorithms in industrial environment and their potential to improve manufacturing efficiency through accurate KPI measurement.

Tavakolirad et al.(2023) in (Tavakolirad et al., 2023) introduced an innovative approach in ML techniques to identify effective indicators and improve understanding of the relationships between them. By integrating supervised and unsupervised models, they analysed customers that directly impact on the goals of the enterprise. Their novel approach also leverages clustering algorithms to analyse high-risk customers, demonstrating the innovative application of

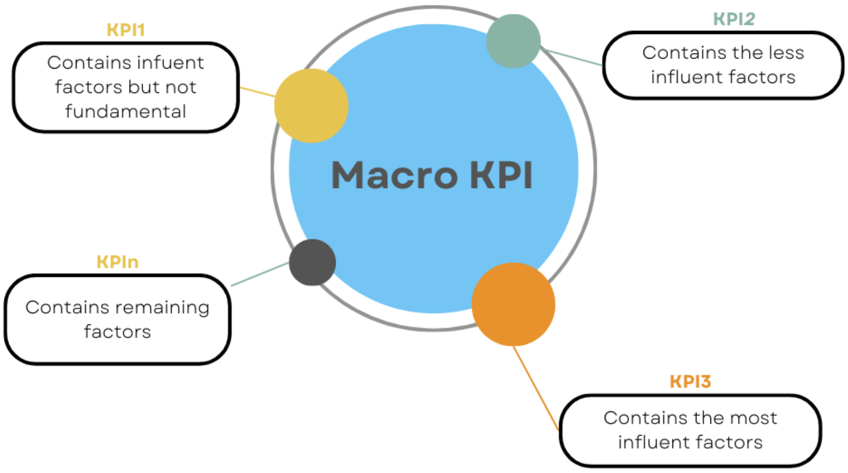

Figure 1: KPI details.

ML in customer segmentation and risk management.

These studies demonstrate the broad applicability and transforming potential of ML in KPI development across industries ranging from e-commerce to project management and manufacturing.

4 FRAMEWORK PROPOSAL

In this section we propose a structured framework for performance evaluation in PAs. The proposal involves precise goal-setting, data analysis, and ML techniques. The framework is divided into several key phases, each of them built upon the insights gained from the previous stages. The ultimate goal is to realize an environment supported by stakeholder engagement and continuous improvement.

1. **Goals Identification:** the identification of the objectives of the organization is the initial phase of the framework. In this phase, the organization focuses on the identification and translation of the goals set by superior institutions. Once these goals have been identified, they define the specific organizational objectives. In this phase, it is essential to involve stakeholders to understand their expectations and performance measurement needs. They must be involved for contributing to the goal definition and must be informed about the mission and strategy of the organization. For instance, these objectives may refer to the reduction of response times or increasing citizen satisfaction.
2. **Macro KPI:** based on the goals defined in the previous step, this phase focuses on identifying macro measurements. By identifying macro-KPIs which measure goals achievement, results can be provided for both internal and external purposes. For example, processing time will be considered a macro-KPI for measuring response times. In the justice sector, the time taken to resolve a judgement process can be considered a macro-KPI that measures the goal response times reduction.
3. **Data:** the phase starts with precisely identification of required data, which demands a clear comprehension of the processes to be monitored within the PA. Once the KPIs are established, the next step is the identification of information systems containing the related data. PAs have various data collection systems, such as document archiving databases or human resources management systems, which are vital sources for acquiring the necessary data. The next step is the data collection phase, where all pertinent information from the identified systems is extracted. Data processing is the final step before analysis, which involves cleaning, pre-processing, and, if required data enrichment. These preparations are essential to facilitate the effective use of machine learning.
4. **Machine Learning:** in this phase it is applied the ML algorithm to the processed data. In particular, RF is effective in handling large volumes of data and identifying the most influential variables with precision.

   It realizes a forest of decision trees, which individually could be subject to over-fitting errors or biased interpretations. However, predictions of many trees are aggregated to obtain a final result which is generally more robust and reliable than single decision tree model result.

   In practical applications, the RF is trained using tabular data that includes input variables, which are specific indicators taken from the event logs of

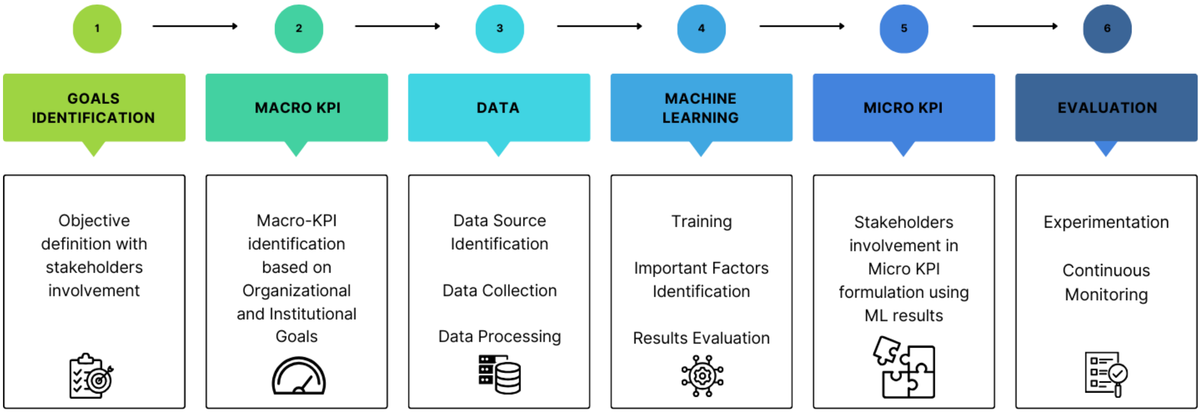

Figure 2: Framework for KPI identification.

information systems, and a target variable represented by the microscopic KPIs that one wishes to monitor and improve. During the training process, the algorithm analyses the correlation between the input variables and the target, identifying the variables that have the most significant impact on the performance measured by the KPIs.

RF is highly useful in quantifying the importance of each input variable in predicting the KPI. This enables administrators to identify the factors that truly influence results and direct resources towards interventions aimed at improving those aspects. Understanding the variables that play a key role in overall performance is essential for optimizing operations and increasing efficiency.

The information obtained from this process is crucial for developing Micro KPIs.

5. **Micro KPI:** thanks to the results of the ML algorithm, it is possible to identify the important factors that most contribute to the macro-KPIs identified. These results are then shared with stakeholders which contribute with knowledge domain to the confirmation of the importance of factors. Then, by leveraging the help of stakeholders, these identified factors are unified to make micro KPIs.

6. **Experimentation and Evaluation:** the last phase of the framework focuses on testing and evaluation of the proposed solutions. After identifying Micro-KPIs and implementing targeted interventions to improve Macro KPIs, it is crucial to test these changes in real scenarios within the organization.

   During the testing phase, interventions are applied on a small scale or under controlled conditions to monitor their effects and collect meaningful data on the effectiveness of the changes made.

   The evaluation phase analyses the collected data to determine whether the interventions have led to a concrete improvement in Micro and Macro KPIs. Based on the results obtained, the organisation may decide to extend interventions on a larger scale, make further changes, or possibly discontinue practices that did not bring the desired benefits. This phase is crucial to ensure that operational decisions are evidence-based and to promote continuous improvement within the organisation.

5 FUTURE WORK AND CONCLUSION

This paper presents a framework for constructing Key Performance Indicators in Public Administration scenarios. The framework leverages the RF algorithm to analyze variable importance and identify the most influential factors affecting public service performance. This provides a solid foundation for understanding the critical performance drivers. Additionally, with the integration of knowledge of domain experts, it is possible to develop relevant KPIs. This ensures that our contribution proposal is both theoretically grounded and practically focused. Finally, the resulting KPIs are continuously monitored and adapted, driving PA flexibility in response to changing conditions and ensuring consistent strategies. In addition, the implementation of real-time data analytics would enable instant updates to KPIs, reflecting the dynamic nature of PAs scenarios.

This work opens up several opportunities for future research. In future work, we plan to explore the application of several ML models to compare their results of the models, and therefore extend our hypothesis. Conducting comparative studies across different PA offices could help to generalize the application of our framework and identify universal best practices. Additionally, tracking the real-world impacts of KPIs adjustments could provide empirical evidence of the benefits of this data-driven approach. Furthermore, aligning public services with community needs could be achieved by prioritising user satisfaction when developing citizen-centring KPIs.

In conclusion, the application of ML techniques, particularly the application of RF and variable importance analysis, represents a step forward towards for a more agile and results-oriented PA. This study extends our understanding of key performance drivers and provides the basis for an effective and targeted performance evaluation system.

ACKNOWLEDGEMENTS

We acknowledge financial support from the project PNRR MUR project PE0000013-FAIR.

REFERENCES

Ahmed, H., Jilani, T. A., Haider, W., Abbasi, M. A., Nand, S., and Kamran, S. (2017). Establishing standard rules for choosing best kpis for an e-commerce business based on google analytics and machine learning technique. *International Journal of Advanced Computer Science and Applications*, 8(5).

Amato, F., Fioretto, S., Forgillo, E., Masciari, E., Mazzocca, N., Merola, S., and Napolitano, E. V. (2023). Evolving justice sector: An innovative proposal for introducing ai-based techniques in court offices. In *International Conference on Electronic Government and the Information Systems Perspective*, pages 75–88. Springer.

Banu, G. S. (2018). Measuring innovation using key performance indicators. *Procedia Manufacturing*, 22:906–911.

Domínguez, E., Pérez, B., Rubio, Á. L., and Zapata, M. A. (2019). A taxonomy for key performance indicators management. *Computer Standards & Interfaces*, 64:24–40.

El Mazgualdi, C., Masrour, T., El Hassani, I., and Khdoudi, A. (2021). Machine learning for kpis prediction: a case study of the overall equipment effectiveness within the automotive industry. *Soft Computing*, 25(4):2891–2909.

Fanaei, S. S., Moselhi, O., Alkass, S. T., and Zangenehmadar, Z. (2018). Application of machine learning in predicting key performance indicators for construction projects. *methods*, 5(9):1450–1457.

Jahangirian, M., Taylor, S. J., Young, T., and Robinson, S. (2017). Key performance indicators for successful simulation projects. *Journal of the Operational Research Society*, 68:747–765.

Kerzner, H. (2019). *Using the project management maturity model: strategic planning for project management*. John Wiley & Sons.

Micu, A., Geru, M., Capatina, A., Avram, C., Rusu, R., and Panait, A. A. (2019). Leveraging e-commerce performance through machine learning algorithms. *Ann. Dunarea Jos Univ. Galati*, 2:162–171.

Parmar, A., Katariya, R., and Patel, V. (2019). A review on random forest: An ensemble classifier. In *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*, pages 758–763. Springer.

Parmenter, D. (2012). *Key performance indicators for government and non profit agencies: Implementing winning KPIs*. John Wiley & Sons.

Parmenter, D. (2015). *Key performance indicators: developing, implementing, and using winning KPIs*. John Wiley & Sons.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, 9:1–11.

Tavakolirad, Z., Albadvi, A., and Akhondzadeh Noughabi, E. (2023). Key performance indicators analysis using machine learning techniques. *Available at SSRN 4520076*.

Velimirović, D., Velimirović, M., and Stanković, R. (2011). Role and importance of key performance indicators measurement. *Serbian Journal of Management*, 6(1):63–72.