

K-Nearest Neighbor Algorithm for Data-Driven IT Governance: A Case Study of Project Outcome Prediction

Agung Suharyanto¹, Kraugusteeliana Kraugusteeliana², Yuniningsih Yuniningsih³, Danu Eko Agustinova⁴, Ari Ramdani⁵, Robbi Rahim^{6*}

¹Universitas Medan Area, Medan, Indonesia

²Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Indonesia

³Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia

⁴Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

⁵Sekolah Tinggi Ilmu Administrasi YPPT Tasikmalaya, Tasikmalaya, Indonesia

⁶Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia

agungsuaharyanto@staff.uma.ac.id, kraugusteeliana@upnvj.ac.id, yuniningsih@upnjatim.ac.id, danu_eko@uny.ac.id, dosenmuda1983@gmail.com, usurobbi85@zoho.com (Corresponding Author)

Abstract. This study examines the application of the k-Nearest Neighbor (k-NN) algorithm for predicting IT project outcomes to support data-driven decision-making in IT governance. The algorithm was applied to a dataset of historical projects from a mid-sized technology company. Despite limitations like sensitivity to parameter tuning, the simplicity and interpretability of k-NN demonstrate its potential as an IT governance decision tool. However, the single case study design restricts generalizability. Further research should explore ensemble approaches to improve robustness, compare k-NN with other methods, and assess its effectiveness across diverse organizational contexts.

Keywords: Artificial Intelligence, Enhanced IT, k-Nearest Neighbors (k-NN)

1. Introduction

In the era of digitalization, Information Technology (IT) has emerged as the fundamental pillar supporting a wide range of operations in businesses across diverse industries (Krajewski et al., 2007; OECD, 2021; Rosda et al., 2023). The role of IT is no longer confined to just providing technical support; it has become an integral part of strategic planning and decision-making processes. This transformation has made IT governance a critical aspect of business management, ensuring the optimal utilization of IT resources in alignment with the organization's strategies and objectives (Lisaria Putri et al., 2023; Muna et al., 2023). IT governance is not a standalone concept; it is deeply intertwined with the organization's leadership, organizational structures, and processes. It involves the establishment of a framework that guides the management and use of IT resources to ensure that they not only support but also enhance the organization's strategies and objectives. This includes setting the direction for IT, making decisions about IT investments, and managing IT performance. In essence, IT governance provides a roadmap for how IT should be used to drive business value and achieve organizational goals (Supardi et al., 2023; Triyono & Nuariyani, 2019; Umar, 2011; Yanwastika Ariyana & D. Abdillah, 2023).

Despite its importance, effective IT governance can be a complex task. Challenges can arise from rapidly changing technology landscapes, cybersecurity threats, regulatory compliance, budget constraints, and the need to align IT initiatives with business goals. These challenges necessitate innovative solutions that can enhance the effectiveness of IT governance.

The significance of IT governance in modern enterprises is of utmost importance, given its pivotal role in guaranteeing operational efficiency, safeguarding data security, and fostering strategic alignment between IT and business objectives. Based on a survey published by Gartner, it has been observed that a significant majority of firms, specifically 80%, that establish a harmonious relationship between their IT governance and business objectives, are able to attain tangible results in terms of operational efficiency and the expansion of their business. Nevertheless, the augmentation in the scope of responsibilities is accompanied by a distinct array of difficulties. The dynamic nature of technical advancements, in conjunction with the growing regulatory requirements such as GDPR and CCPA, presents a challenge for governance frameworks to maintain up-to-date relevance. The issue of cybersecurity is of significant importance, as evidenced by the findings of the FBI's Internet Crime Report. The report highlights a substantial 69% surge in cybercrime complaints between 2019 and 2021, underscoring the imperative for the implementation of strong IT governance measures to effectively manage and reduce associated risks. Furthermore, the proliferation of remote work settings and the adoption of cloud computing have broadened the scope of IT administration, hence intensifying the complexity of this undertaking. Therefore, in light of the current organizational landscape, the significance of IT governance has become increasingly paramount. However, it is important to acknowledge that this field is confronted with unprecedented obstacles, necessitating the development of adaptable and all-encompassing approaches.

A cutting-edge solution that has gained prominence in recent times is the utilization of Artificial Intelligence (AI). In particular, the k-Nearest Neighbors (k-NN) algorithm stands out as a remarkable tool. This algorithm is a form of instance-based learning, also referred to as lazy learning, where the function is approximated locally, and all computations are postponed until the function evaluation is required. The k-NN algorithm can be applied to both classification and regression predictive problems, although it is more commonly employed in classification problems within the realm of data mining (Adhitya et al., 2015; Liantoni, 2016; Lu et al., 2021; Siburian et al., 2019). The significance of the k-NN AI algorithm in the context of IT governance is underscored by its capacity to analyze vast quantities of data and make predictions or decisions based on that data. This ability to handle large datasets and provide valuable insights makes it an invaluable tool for IT governance, where data-driven decision-making is crucial.

Artificial Intelligence (AI) signifies a monumental transformation in the realm of technology (Lalithendra Nadh & Syam Prasad, 2018). AI encompasses a vast area of research, incorporating a multitude of theories, methodologies, and technologies. Machine Learning uses neural networks, genetic algorithms, and support vector machines, among others. Reasoning—logic, limits, decision theory, and game theory—is another important topic (Kaushik et al., 2021; Melanie, 1996; Nguyen et al., 2019). Robotics—manipulation, locomotion, and perception—is essential to AI. AI is set to bring about a revolutionary impact across all sectors, making it a focal point of interest for various stakeholders. It is altering the business landscape, from automating mundane tasks to providing intelligent insights that drive decision-making. In essence, AI is reshaping the way businesses operate, offering unprecedented opportunities for innovation and growth.

There is a growing body of research that explores the application of k-NN and other AI techniques in many aspects. For instance, a study by Nicola Altini et al. (Altini et al., 2021) uses deep learning and classical machine learning to segment and identify vertebrae. The suggested system contains two phases: a binary completely automated spine segmentation using a 3D convolutional neural network, and a semi-automated vertebrae centroids location using classic machine learning algorithms. Another research by Jeff Calder (Calder & García Trillos, 2022) increases spectral convergence for graph-based approximations of weighted Laplace-Beltrami operators from random input. The method can analyze many graph constructions, including ϵ -graphs and k-NN graphs. Also research by Quy (Quy & Kim, 2021) presents a gas pipeline leak detection method employing a k-NN classifier and hybrid features from acoustic emission (AE) signals. For real-time leak detection, the algorithm is incorporated in an MCU.

The objective of this paper is to explore the revolutionary potential of k-Nearest Neighbors (k-NN) algorithms in strengthening IT governance frameworks. The essay examines the practical implementation of k-NN algorithms in an operational IT governance context through the presentation of a detailed case study on IT projects. The evaluation of the outcomes of this implementation focuses on the optimization of resource allocation and decision-making processes. The findings are then extrapolated to provide significant insights for other businesses. The primary aim is to provide a practical example of how artificial intelligence, particularly k-NN algorithms, can be strategically utilized to address persistent issues in IT governance, hence enhancing its overall effectiveness.

2. Literature Review

The instance-based k-Nearest Neighbors (k-NN) algorithm can predict classification and regression (Basha et al., 2019). k-nearest neighbors use a set number of training samples that are closest to a new point to predict its label. The user can employ a constant number of samples (k-NN) or adjust the number based on point density (radius-based neighbor learning). The most used metric system for measuring distance is Euclidean.

An example of a non-parametric method that is used for classification and regression is the k-Nearest Neighbors (k-NN) algorithm. Because it does not make any explicit assumptions about the functional form of h , it is able to avoid the risks associated with incorrectly modeling the underlying distribution of the data because it is non-parametric. Consider the following scenario: the data exhibit a significantly non-Gaussian distribution, yet the learning model expects a Gaussian form. If this occurs, the consequences could be extremely severe (Kakoly et al., 2023; Sidik & Sen, 2019; Sultana et al., 2022).

k-nearest neighbor categorization produces class membership. An object's k nearest neighbors will vote on its classification, and it will be placed in the category with the most votes. If k equals one, the object is simply classified as its closest neighbor. k-NN regressions determine an object's property value. The closest k neighbors' values are averaged to get this value. Instance-based learning, or "lazy learning" approximates the function locally and delays computation until the function is assessed.

The term "lazy" in the context of lazy learning pertains to the algorithm's practice of postponing computing tasks until a specific query is initiated. This stands in stark contrast to algorithms such as decision trees or neural networks that exhibit "eager learning," wherein the majority of computing tasks are carried out during the initial training phase. The dynamic computation capabilities of lazy learning enable adaptable and context-specific model modifications. However, this approach might incur significant computational costs and exhibit sluggish performance when used to extensive datasets.

The effectiveness of k-NN is greatly influenced by the distance measurement(Adeniyi et al., 2016). The most common type of distance measurement is Euclidean distance, which is suitable for continuous variables. For categorical variables, Hamming distance can be used. In some cases, the variables may also need to be standardized before calculating distance.

The value of k that is selected for use in k-nearest neighbors is greatly dependent on the data being analyzed. In general, a bigger k reduces the impacts of noise, but it also blurs the classification boundaries. The ideal choice of the value k is extremely dependent on the data. In general, a bigger k reduces the impacts of noise, but it also blurs the classification boundaries.(Beeche et al., 2022; Karimi & Yousefi, 2012)

The k-nearest neighbors approach is easy to develop, forgiving of noisy training data, and effective when used to huge amounts of training data. However, it is necessary for it to decide the value of the parameter K, which stands for the number of nearest neighbors. Additionally, the calculation cost is large since it is necessary for it to compute the distance between each instance and all the training samples.

Table 1. k-NN and Artificial Intelligence Research Summary

No	Title	Summary
1	How AI Can be Used for Governance of Messaging Services: A Study on Spam Classification Leveraging Multi-Channel Convolutional Neural Network(Waja et al., 2023)	The research goal is to develop a model for the detection of offensive content or spam in text messages, social media, and emails using MCC-NN
2	Intelligent Learning based Opinion Mining Model for Governmental Decision Making(Sharma & Shekhar, 2020)	This research is to develop a socio-affective framework equipped with opinion mining to gain insight into how government programs and initiatives are perceived by citizens and relevant stakeholder groups. The framework aims to answer questions about citizens' perception of new initiatives, the benefits and pitfalls of programs, the possibility of replicating positive attributes to other programs, the need to address negative feedback, and the attention given to the concerns of the general mass when formulating administrative decisions. The research also aims to analyze the findings and observations based on statistical measures and provide recommendations for refining strategies and improving communication and community outreach programs.
3	Flood Damage Analysis Using Machine Learning Techniques(Snehil & Goel, 2020)	Based on limited data about damages and rainfall in India, this study will evaluate flood risk analysis using data mining algorithms like logistic regression and decision tree.
4	Machlne learning for nutrient recovery in	This research goal is to explore and promote the

	the smart city circular economy – A review(Soo et al., 2023)	implementation of Circular Economy (CE) practices within smart cities, specifically focusing on nutrient recovery and its end-to-end implementation using machine learning approach
5	Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda(Nishant et al., 2020)	This research discusses the potential of artificial intelligence (AI) in addressing major societal problems, particularly in the field of sustainability. It emphasizes the need for advanced and innovative solutions to combat the degradation of the natural environment and the climate crisis, also aims to promote groundbreaking research and practical solutions for using AI to achieve environmental sustainability.
6	The application of PROMETHEE multi-criteria decision aid in financial decision making: Case of distress prediction models evaluation(Mousavi & Lin, 2020)	This article discusses the application of PROMETHEE multi-criteria decision aid in financial decision making, specifically in the evaluation of distress prediction models. It focuses on the use of corporate governance indicators and machine learning in predicting corporate distress, as well as the issue of conflicting rankings in the evaluation process. The article provides valuable insights and presents a comprehensive analysis of the subject matter.
7	Assessing the Effectiveness of Masking and Encryption in Safeguarding the Identity of Social Media Publishers from Advanced Metadata Analysis(Khader & Karam, 2023)	Masking and encryption are tested to shield social media producers from advanced metadata analysis. The research helps understand social media privacy security. The essay examines how masking and encryption protect social media producers from advanced metadata monitoring. It examines how machine learning algorithms might mine social media data, potentially compromising anonymous publishers' privacy. Twitter is cited as a platform where machine learning may identify anonymous users/publishers. URL shortening, pseudonymization, obfuscation, and cryptographic algorithms are tested to safeguard social media producers' identities and privacy. These methods disguise or alter metadata to prevent enemies from identifying social media publishers and accessing sensitive data.
8	A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes(Gupta et al., 2021)	This research focuses on the use of different machine learning and deep learning-based models for fraud detection in healthcare systems. The study aims to explore various models and their effectiveness in detecting fraudulent activities, with a specific focus on the financial consequences of fraud in Universal Health Coverage Schemes, such as investigation costs, revenue losses, and reputational risk.
9	Sentiment Analysis of User Feedback in e-Learning Environment(Omar et al., 2020)	This document discusses the importance of sentiment analysis (SA) in the context of user-generated content, specifically in e-learning environments. SA is valuable because it allows for the extraction of useful insights from large

		<p>volumes of unstructured data found in social media and micro-blog websites. However, conducting SA on user-generated content can be challenging due to its informal nature, which introduces additional variables and properties that need to be evaluated. The author emphasizes the need for evaluating these informal texts and highlights the relevance of SA in understanding user feedback in e-learning environments.</p>
--	--	---

The k-Nearest Neighbors (k-NN) algorithms have been applied in domains beyond typical machine learning tasks, demonstrating its relevance in areas such as project management and risk analysis. The adoption of this algorithm across many disciplines provides support for its potential to enhance IT governance system.

Within the field of project management, k-NN algorithms have been utilized to predict project outcomes by analyzing past data. For example, the k-Nearest Neighbors (k-NN) algorithm has been employed in the domain of project management to forecast project completion durations and budget deviations. This is achieved by comparing existing projects with historical projects that exhibit comparable characteristics within the dataset. The algorithm's ability to identify the 'k' most similar historical projects enables it to generate predictions that are more localized and context-sensitive. This, in turn, assists project managers in making well-informed decisions and implementing necessary revisions. This is especially advantageous in agile settings characterized by high levels of dynamism and the need for frequent evaluation.

The k-Nearest Neighbors (k-NN) algorithm has been utilized in various domains of risk analysis, including credit scoring, fraud detection, and operational risk assessment. In financial institutions, the implementation of k-NN algorithms enables the identification of potentially high-risk loan applicants by a comparative analysis of their profiles with those of previous applicants who have either defaulted or successfully repaid their loans. Likewise, within the realm of cybersecurity, the k-Nearest Neighbors (k-NN) algorithm can be employed to identify aberrant activities by juxtaposing network traffic patterns or system behavior with past records. The algorithm's capacity to make localized decisions renders it proficient in detecting outliers or anomalies, thereby limiting potential dangers.

The utilization of the k-Nearest Neighbors (k-NN) algorithm has been observed in the field of operations and supply chain management for the purpose of enhancing resource allocation. By doing a comparative analysis of present resource use patterns with historical data, the system is capable of making predictions regarding future resource requirements. This feature is especially advantageous in the context of IT governance since it enables the allocation of resources to be flexible and responsive in order to effectively address the evolving requirements of the organization.

3. Method

The case study focuses on a mid-sized technology company, which operates in a highly competitive market and relies heavily on its IT infrastructure for various business operations. IT governance challenges include managing IT risks, aligning IT initiatives with business goals, and optimizing the allocation of IT resources.

Media Digital Indonesia is small company that exist in Indonesia that focus on Software and Hardware developer for Regencies and Cities in North Sumatra, this company cooperation with government office in Province North Sumatra to provide software and hardware to support government activities.

A. Data Collection

Data collecting began implementation. Data on past IT records projects, risks, resource allocation, and results. All of this will be trained using feature selection and data splitting techniques.

B. Data Preprocessing

For the k-NN algorithm, the data was preprocessed. This required handling missing values, normalizing numerical data to guarantee that all features contribute equally to the distance calculation, and encoding categorical variables.

C. Model Training

Preprocessed data trained k-NN. To maximize model performance, the number of neighbors (k) was tuned during training. Cross-validation was used to identify the optimal model performance k value.

D. Model Testing

The trained model was tested on a new data set (test set). To test the model's ability to anticipate fresh data, this was done.

E. Model Integration

Once the model was trained and tested, it was integrated into IT governance structure. The model was used to make predictions and decisions in various areas of IT governance, such as project management, risk management, and resource allocation. General k-NN process can be described as in figure 1 below:

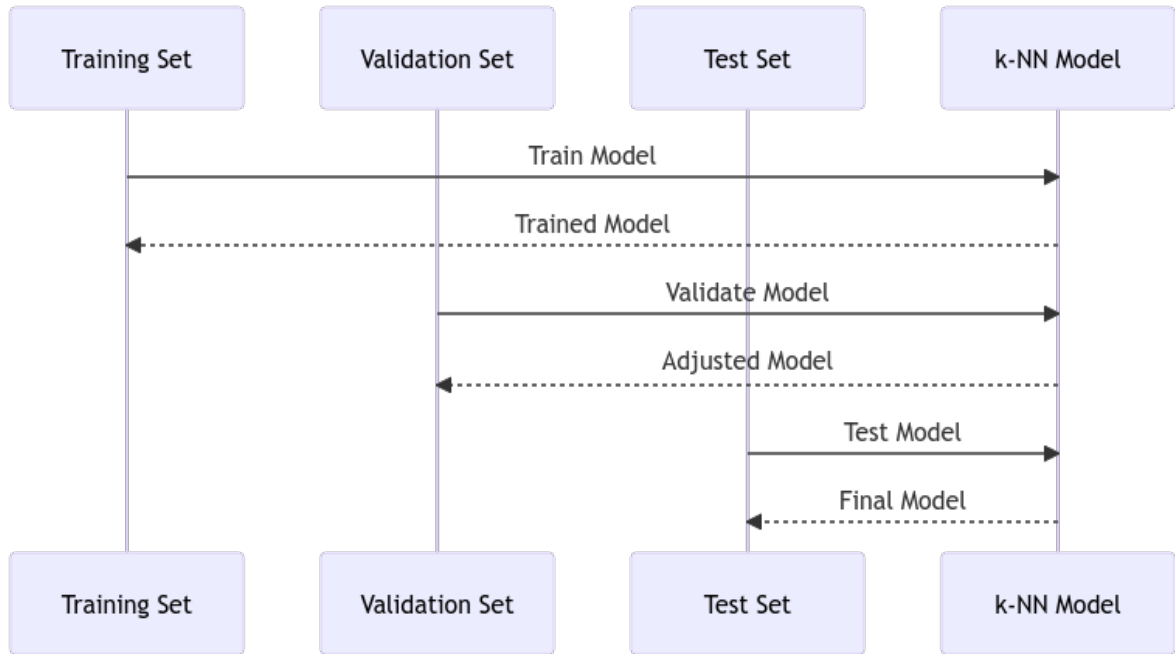


Fig 1. k-NN process

4. Result and Discussion

k-Nearest Neighbors (k-NN) algorithm can be applied to the data table to make predictions about the outcome of future IT governance projects, see table 1 for simple dataset.

Table 2. Dataset Project

Project ID	Project Budget	Project Duration	Team Size	Technology Used	Risk Level	Resources Allocated	Project Outcome
1	50000	6 months	5	Java	High	30000	Success
2	75000	9 months	7	Python	Medium	45000	Failure
3	60000	12 months	6	C++	Low	35000	Success
4	80000	8 months	8	Java	High	50000	Failure

5	70000	10 months	7	Python	Medium	40000	Success
---	-------	-----------	---	--------	--------	-------	---------

Description table 1:

- Project ID: a unique identifier for each project.
- Project Budget: total budget for the project.
- Project Duration: total duration of the project.
- Team Size: number of team members involved in the project.
- Technology Used: main technology or programming language used in the project.
- Risk Level: level of risk associated with the project, categorized as High, Medium, or Low.
- Allocated Resources: Project resources.
- Outcome: project success or failure.

This data will be processed with the k-NN algorithm to make predictions about future projects based on their budget, duration, team size, technology used, risk level, and resources allocated.

Before using k-NN, data must be preprocessed. This stage converts raw data into an algorithm-friendly format. Preprocessing table data:

a. k-NN algorithm requires numerical input. Thus, "Technology Used" and "Risk Level" must be numerical variables. A one-hot encoding method allows this. "Technology Used" can be split into "Uses Java", "Uses Python", and "Uses C++". These variables would be 1 if the technology was employed and 0 otherwise. "Risk Level" can also be split into "High Risk," "Medium Risk," and "Low Risk" variables.

b. Normalizing: k-NN algorithm calculates the distance between data points, so it's important that all numerical variables are on the same scale. Otherwise, variables with larger values (like "Project Budget") could dominate the distance calculation. To prevent this, numerical variables can be normalized so that they all have a range between 0 and 1. This can be done using a technique called min-max normalization.

After these preprocessing steps, data would be processed using k-NN algorithm. The transformed data would look something like this:

Table 3. k- NN Data Processing

Project ID	Project Budget (Normalized)	Project Duration (Normalized)	Team Size (Normalized)	Uses Java	Uses Python	Uses C++	High Risk	Medium Risk	Low Risk	Resources Allocated (Normalized)	Project Outcome
1	0.5	0.3	0.4	1	0	0	1	0	0	0.6	Success
2	0.75	0.45	0.55	0	1	0	0	1	0	0.9	Failure
3	0.6	0.6	0.5	0	0	1	0	0	1	0.7	Success
4	0.8	0.4	0.6	1	0	0	1	0	0	1.0	Failure
5	0.7	0.5	0.55	0	1	0	0	1	0	0.8	Success

0: The value frequently denotes the minimal attainable value for a specific attribute subsequent to the process of normalization. Within the framework of our IT project case study, a budget or resource allocation of zero would indicate the minimal value when compared to all other projects under consideration.

1: On the contrary, this value signifies the uppermost attainable value for a specific attribute following the process of normalization. In the context of our case study, a budget or resource allocation value of 1 signifies the highest level among all the projects under consideration.

Within the range of 0 and 1: Each value falling within the interval of 0 and 1 denotes a relative location in relation to the minimum and maximum values associated with the given feature. An

instance of a normalized budget value equal to 0.5 indicates that the budget falls precisely at the midpoint between the minimum and highest budgets within the dataset.

The process of normalization holds particular significance in the context of algorithms such as k-NN, wherein distance measures are employed to generate predictions. When a particular feature has a significantly larger size in comparison to other features, it can exert a dominant influence on distance computations, hence potentially resulting in inaccurate predictions. By applying a normalization technique that scales all features to a range of 0 to 1, each feature is afforded an equitable chance to impact the final decision.

Instance-based learning, or lazy learning, uses the k-Nearest Neighbors (k-NN) technique to approximate the function locally and defer computation until function evaluation. Among machine learning algorithms, the k-NN algorithm is basic.

Calculate the training set distance from a test point. Manhattan and Minkowski distances are also employed, but Euclidean distance is the most popular. This study estimated Euclidean distance between two places (p1, p2):

$$d(p1, p2) = \text{sqrt}((p1_x - p2_x)^2 + (p1_y - p2_y)^2)$$

where p1_x and p1_y are the coordinates (Project Budget and Resources Allocated) of the new project, and p2_x and p2_y are the coordinates of a project in the training set. Calculation distance between the new project and each project in the training set as can see below:

- a. Distance to Project 1:
d(new project, Project 1) = sqrt((0.65 - 0.5)^2 + (0.75 - 0.6)^2)
 - b. Distance to Project 2:
d(new project, Project 2) = sqrt((0.65 - 0.75)^2 + (0.75 - 0.9)^2)
 - c. Distance to Project 3:
d(new project, Project 3) = sqrt((0.65 - 0.6)^2 + (0.75 - 0.7)^2)
 - d. Distance to Project 4:
d(new project, Project 4) = sqrt((0.65 - 0.8)^2 + (0.75 - 1.0)^2)
 - e. Distance to Project 5:
d(new project, Project 5) = sqrt((0.65 - 0.7)^2 + (0.75 - 0.8)^2)
- result Euclidean distance from calculation above can be seen in table 3 below.

Table 4. Euclidean Distance Result

Project ID	Distance to New Project
1	0.18
2	0.26
3	0.07
4	0.36
5	0.07

These distances represent the Euclidean distance between the new project and each project in the training set. The smaller the distance, the more similar the project is to the new project. In the step of selecting the 'k' nearest neighbors, we choose the 'k' data points in the training set that are closest to the new data point. This is done based on the distances calculated in the previous step.

There isn't a specific formula for this step, as it's more of a sorting and selection operation. We sort the distances in ascending order and select the top 'k' points. Let's say we choose 'k' as 3. From the distances that calculated earlier (table 3). The 3 projects with the smallest distances to the new project are Project 3, Project 5, and Project 1. So, the 'k' nearest neighbors to the new project are:

Table 5. *k* new project

Nearest Neighbor	Distance to New Project
Project 3	0.07
Project 5	0.07
Project 1	0.18

These are the 3 projects that are most similar to the new project, according to the k-NN algorithm.

k-Nearest Neighbors (k-NN) algorithm was applied to predict the outcome of a new IT project based on a set of historical project data. The attributes considered were the normalized project budget and resources allocated. The k-NN algorithm predicted that the new project will be a success based on the outcomes of its three nearest neighbors in the training set.

The instance-based k-NN method does not learn a model. Machine learning is the reason. Instead, it memorizes training examples and applies "knowledge" during prediction. This works. k-NN's simplicity makes it useful in interpretable fields like information technology governance.

The non-parametric k-NN algorithm makes no assumptions about the data or distribution. IT projects vary widely and may not follow a distribution, making this trait useful. k-NN has some drawbacks. The distance metric and 'k' matter. This case study employed $k=3$ and the Euclidean distance metric, although other selections could yield different predictions. It's sensitive to varying scale, therefore we standardized them before applying the method.

k-NN also struggles with categorical variables. Before using the technique, we converted category data to numerical variables in this case study. This can be a drawback if the categorical variables have several levels or a complex relationship with the outcome. The k-NN algorithm can forecast IT project outcomes when combined with other approaches and expert judgment, despite its limitations. This case study uses the k-NN technique due to its simplicity and interpretability.

5. Conclusion

This study demonstrated the application of k-NN algorithm to predict IT project outcomes, supporting data-driven IT governance. While the case study reveals its potential, limitations like sensitivity to parameter tuning persist. Further research should explore ensemble approaches to improve robustness, comparative assessments with other methods, and applications in diverse organizational contexts. The findings have practical implications for incorporating k-NN to enhance IT governance decision-making.

Nevertheless, k-NN method does have a few drawbacks of its own. It is sensitive to the choice of 'k' and the distance measure that is employed, and it does not perform well with categorical variables. When utilizing the k-NN algorithm, it is necessary to take these constraints into consideration.

Looking to the future, there are many different directions that research could go in the future. To handle categorical variables in the k-NN algorithm in a variety of different ways is one idea that may be investigated. Another option is to study various distance measurements and the effect that these metrics have on the predictions made by the algorithm. To further improve the reliability and precision of the predictions, one may make use of ensemble approaches, which integrate the results of several different machine learning algorithms.

In conclusion, the k-NN algorithm is an effective instrument forecasting the results of information technology projects. Despite its restrictions, it is an excellent option for using in applications related to IT governance due to its understandability and ease of implementation. It is possible for future

research to further improve both its efficacy and its usefulness in this sector.

References

- Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, 12(1), 90–108. <https://doi.org/10.1016/J.ACI.2014.10.001>
- Adhitya, E. K., Satria, R., & Subagyo, H. (2015). Komparasi Metode Machine Learning dan Metode Non Machine Learning untuk Estimasi Usaha Perangkat Lunak. *IlmuKomputer.Com Journal of Software Engineering*, 1(2), 109–113.
- Altini, N., De Giosa, G., Fragasso, N., Coscia, C., Sibilano, E., Prencipe, B., Hussain, S. M., Brunetti, A., Buongiorno, D., Guerriero, A., Tatò, I. S., Brunetti, G., Triggiani, V., & Bevilacqua, V. (2021). Segmentation and Identification of Vertebrae in CT Scans Using CNN, k-Means Clustering and k-NN. *Informatics*, 8(2), 40. <https://doi.org/10.3390/informatics8020040>
- Basha, S. M., Bagyalakshmi, K., Ramesh, C., Rahim, R., Manikandan, R., & Kumar, A. (2019). Comparative study on performance of document classification using supervised machine learning algorithms: KNIME. *International Journal on Emerging Technologies*, 10(1).
- Beeche, C., Singh, J. P., Leader, J. K., Gezer, N. S., Oruwari, A. P., Dansingani, K. K., Chhablani, J., & Pu, J. (2022). Super U-Net: A modularized generalizable architecture. *Pattern Recognition*, 128, 108669. <https://doi.org/10.1016/J.PATCOG.2022.108669>
- Calder, J., & García Trillos, N. (2022). Improved spectral convergence rates for graph Laplacians on ϵ -graphs and k-NN graphs. *Applied and Computational Harmonic Analysis*, 60, 123–175. <https://doi.org/10.1016/j.acha.2022.02.004>
- Gupta, R. Y., Mudigonda, S. S., & Baruah, P. K. (2021). A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes. *International Journal of Engineering Trends and Technology*, 69(3), 96–102. <https://doi.org/10.14445/22315381/IJETT-V69I3P216>
- Kakoly, I. J., Hoque, M. R., & Hasan, N. (2023). Data-Driven Diabetes Risk Factor Prediction Using Machine Learning Algorithms with Feature Selection Technique. *Sustainability*, 15(6), 4930. <https://doi.org/10.3390/su15064930>
- Karimi, H., & Yousefi, F. (2012). Application of artificial neural network-genetic algorithm (ANN-GA) to correlation of density in nanofluids. *Fluid Phase Equilibria*, 336, 79–83. <https://doi.org/10.1016/j.fluid.2012.08.019>
- Kaushik, P., Sharma, A., & Bhathal, G. S. (2021). Revamping Supermarkets with AI and RSSI*. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*. <https://doi.org/10.1109/ICAIS50930.2021.9395814>
- Khader, M., & Karam, M. (2023). Assessing the Effectiveness of Masking and Encryption in Safeguarding the Identity of Social Media Publishers from Advanced Metadata Analysis. *Data*, 8(6), 105. <https://doi.org/10.3390/data8060105>
- Krajewski, L., Ritzman, L., & MK, M. (2007). Operations Management – Processes and Value Chain. *International Journal of Production Economics*. <https://doi.org/10.1016/j.ijpe.2008.07.023>
- Lalithendra Nadh, V., & Syam Prasad, G. (2018). Support vector machine in the anticipation of currency markets. *International Journal of Engineering & Technology*, 7(2–7), 66.

<https://doi.org/10.14419/ijet.v7i2.7.10262>

Liantoni, F. (2016). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *Jurnal ULTIMATICS*, 7(2), 98–104. <https://doi.org/10.31937/ti.v7i2.356>

Lisaria Putri, R., Nyoman Sri Werastuti, D., Dwi Astuti, E., Fatimah Khairunnisa, A., Wahyono, E., & Luh Apriani, N. (2023). Integrated Reporting: Corporate Strategy Towards Achieving Sustainable Development SDGs. *Apollo: Journal of Tourism and Business*, 1(2), 64–71. <https://doi.org/10.58905/APOLLO.V1I2.39>

Lu, J., Qian, W., Li, S., & Cui, R. (2021). Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery. *Applied Sciences*, 11(3), 919. <https://doi.org/10.3390/app11030919>

Melanie, M. (1996). An introduction to genetic algorithms. *Cambridge, Massachusetts London, England*, ..., 162. [https://doi.org/10.1016/S0898-1221\(96\)90227-8](https://doi.org/10.1016/S0898-1221(96)90227-8)

Mousavi, M. M., & Lin, J. (2020). The application of PROMETHEE multi-criteria decision aid in financial decision making: Case of distress prediction models evaluation. *Expert Systems with Applications*, 159, 113438. <https://doi.org/10.1016/j.eswa.2020.113438>

Muna, A., Ausat, A., Al Bana, T., & Gadzali, S. S. (2023). Basic Capital of Creative Economy: The Role of Intellectual, Social, Cultural, and Institutional Capital. *Apollo: Journal of Tourism and Business*, 1(2), 42–54. <https://doi.org/10.58905/APOLLO.V1I2.21>

Nguyen, H., Bui, X. N., Tran, Q. H., & Mai, N. L. (2019). A new soft computing model for estimating and controlling blast-produced ground vibration based on Hierarchical K-means clustering and Cubist algorithms. *Applied Soft Computing Journal*, 77, 376–386. <https://doi.org/10.1016/j.asoc.2019.01.042>

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, 53, 102104. <https://doi.org/10.1016/j.ijinfomgt.2020.102104>

OECD. (2021). The Digital Transformation of SMEs. *OECD Publishing*, 1–4.

Omar, M. A., Makhtar, M., Ibrahim, M. F., & Aziz, A. A. (2020). Sentiment Analysis of User Feedback in e-Learning Environment. *International Journal of Engineering Trends and Technology*, 153–157. <https://doi.org/10.14445/22315381/CATI2P224>

Quy, T. B., & Kim, J.-M. (2021). Real-Time Leak Detection for a Gas Pipeline Using a k-NN Classifier and Hybrid AE Features. *Sensors*, 21(2), 367. <https://doi.org/10.3390/s21020367>

Rosda, R., Wali, M., & Imilda, I. (2023). Evaluation of the Successful Implementation of the SIMARDI Using the Technology Acceptance Model (TAM). *SAGA: Journal of Technology and Information System*, 1(2), 56–63. <https://doi.org/10.58905/SAGA.V1I2.67>

Sharma, A., & Shekhar, H. (2020). Intelligent Learning based Opinion Mining Model for Governmental Decision Making. *Procedia Computer Science*, 173, 216–224. <https://doi.org/10.1016/j.procs.2020.06.026>

Siburian, R. H. S., Karolina, R., Nguyen, P. T., Lydia, E. L., & Shankar, K. (2019). Leaf disease classification using advanced SVM algorithm. *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue), 712–718. <https://doi.org/10.35940/ijeat.F1138.0886S19>

Sidik, D. D., & Sen, T. W. (2019). Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan. *IT for Society*, 4(1), 21–27. <https://doi.org/10.33021/itfs.v4i1.1180>

Snehil, & Goel, R. (2020). Flood Damage Analysis Using Machine Learning Techniques. *Procedia Computer Science*, 173, 78–85. <https://doi.org/10.1016/j.procs.2020.06.011>

Soo, A., Wang, L., Wang, C., & Shon, H. K. (2023). Machine learning for nutrient recovery in the smart city circular economy – A review. *Process Safety and Environmental Protection*, 173, 529–557. <https://doi.org/10.1016/j.psep.2023.02.065>

Sultana, Z., Ferdousi, A., Tasnim, F., & Nahar, L. (2022). An Improved K-Nearest Neighbor Algorithm for Pattern Classification. *International Journal of Advanced Computer Science and Applications*, 13(8), 760–767. <https://doi.org/10.14569/IJACSA.2022.0130887>

Supardi, S., Eko Agustinova, D., Tri, S., Putro, J., & Septiantoko, R. (2023). E-Services: Implementation of Digital-Based Public Services in The 4.0 Era. *Athena: Journal of Social, Culture and Society*, 1(3), 87–92. <https://doi.org/10.58905/ATHENA.V1I3.40>

Triyono, A., & Nuariyani, N. W. (2019). Information and communication technology (ICT) and women empowerment in Indonesia. *Humanities and Social Sciences Reviews*, 7(3), 255–260. <https://doi.org/10.18510/HSSR.2019.7339>

Umar, H. (2011). Government Financial Management Strategy for Preventing Corruption in Indonesia. *The South East Asian Journal of Management*, 5(2), 19–36.

Waja, G., Patil, G., Mehta, C., & Patil, S. (2023). How AI Can be Used for Governance of Messaging Services: A Study on Spam Classification Leveraging Multi-Channel Convolutional Neural Network. *International Journal of Information Management Data Insights*, 3(1), 100147. <https://doi.org/10.1016/j.jjime.2022.100147>

Yanwastika Ariyana, R., & D. Abdillah, V. (2023). Network Design Using Load Balancing Equal Cost Multi Path (ECMP) Method At Lexima Consultant Office. *SAGA: Journal of Technology and Information System*, 1(2), 31–36. <https://doi.org/10.58905/saga.v1i2.81>