Beyond Rankings: A K-means Approach to Evaluating Research Universities in Emerging Higher Education Systems

TARIK KUCUKDENIZ[1], MELIKE SOLAK ALTUNTAS[2*],
CANAN HAZAL AKARSU[3], EROL INCE[4]

[1]Department of Industrial Engineering
Istanbul University-Cerrahpasa
Avcilar Campus, 34320
TÜRKİYE

[2]Corporate Data Management Office
Istanbul University-Cerrahpasa
Avcilar Campus, 34320
TÜRKİYE

[3] Corporate Data Management Office
Istanbul University-Cerrahpasa
Avcilar Campus, 34320
TÜRKİYE

[4]Department of Chemical Engineering
Istanbul University-Cerrahpasa
Avcilar Campus, 34320
TÜRKİYE

Abstract: - This study introduces a novel approach to evaluating research universities in developing countries, using Türkiye as a case study within the broader context of global higher education trends. By combining the national University Ranking by Academic Performance (URAP-TR) metrics with K-means clustering analysis, we address the limitations of international ranking systems in assessing institutions outside the Global North. Our comparative analysis of 23 Turkish research universities, implemented using Python and scikit-learn, resulted in three distinct clusters that reflect diverse patterns of institutional development. This clustering approach allows for a nuanced comparison of university performance within Turkey's higher education landscape, while also connecting to global debates on university rankings and performance metrics. A focused examination of Istanbul University-Cerrahpasa illustrates how this method can inform targeted improvement strategies, offering insights applicable to institutions in similar contexts worldwide. By moving beyond traditional rankings, this approach facilitates data-driven decision-making in higher education policy and institutional strategy.

Key-Words: - University ranking, K-means clustering, URAP-TR, Research performance, Higher education policy, Türkiye.

Received: March 11, 2024. Revised: August 19, 2024. Accepted: December 14, 2024. Published: March 27, 2025.

1 Introduction

University ranking systems have become crucial tools for evaluating and comparing academic and research performance of higher education institutions worldwide. These systems, utilizing various metrics such as publication output, citation impact, and international collaboration, provide stakeholders with valuable insights into institutional strengths and areas for improvement. While international ranking systems like QS, THE, and ARWU offer comprehensive global comparisons, they often favor institutions in developed countries, leaving a significant gap in the evaluation of universities in developing nations. This limitation has led to the emergence of national ranking systems, tailored to specific country contexts and needs. In Türkiye, the University Ranking by Academic Performance (URAP-TR) system was developed in 2009 to address this gap, providing a more nuanced evaluation of Turkish universities. However, despite the progress made by URAP-TR, there remains a need for more sophisticated analytical approaches to fully leverage the multidimensional data provided by such ranking systems. [1].

The application of clustering techniques, particularly the K-means algorithm, offers a promising way for addressing this gap. By grouping universities with similar performance characteristics, clustering analysis can reveal patterns and insights that may not be apparent from traditional ranking methodologies alone. This approach is especially valuable for research universities, which play a critical role in advancing scientific knowledge and driving innovation. Despite the potential of clustering techniques in university performance analysis, there is a lack of comprehensive studies applying these methods to national ranking data, particularly in the context of developing countries. This gap in the literature presents an opportunity to enhance our understanding of university performance dynamics and provide more targeted strategies for improvement. The present study aims to address this gap by conducting a comparative analysis of research universities in Türkiye using the K-means clustering algorithm applied to URAP-TR metrics. By focusing on Istanbul University-Cerrahpaşa as a case study, this research seeks to demonstrate how clustering analysis can provide a more nuanced understanding of a university's performance relative to its peers, identify specific areas for improvement to advance to a higher performance cluster, and offer insights that can inform strategic decision-making in higher education management. This approach not only contributes to the methodological advancement of university performance analysis but also offers practical implications for university administrators and policymakers in developing countries. By combining the strengths of national ranking systems with data analysis techniques, this study aims to provide a more comprehensive and actionable framework for university performance evaluation and improvement.

This research addresses a critical gap in higher education assessment by developing a context-sensitive analytical framework. While global ranking systems such as THE, QS, and ARWU provide standardized comparisons, they systematically privilege institutions in developed nations through metrics that reflect Western academic traditions and resource advantages. Our clustering methodology counterbalances this structural inequity by analyzing universities within their national higher education ecosystem, acknowledging distinctive historical trajectories, funding mechanisms, and societal missions. The application of K-means analysis to URAP-TR metrics offers a transferable model for other emerging higher education systems—from Southeast Asia to Latin America and Africa—where universities face similar challenges of global recognition while serving distinct local priorities. The identification of performance clusters delivers actionable intelligence for multiple stakeholders: institutional leaders can benchmark against appropriate peer institutions rather than unattainable global exemplars; national policymakers can implement targeted interventions reflecting the differentiated needs of each university cluster; and funding agencies can allocate resources more strategically to maximize system-wide advancement. This approach transcends the reductionist nature of conventional rankings, challenging the 'one-size-fits-all' evaluation paradigm dominating global higher education discourse. By demonstrating how sophisticated data science techniques can extract insights from performance metrics, this study contributes to the emerging critical literature on university evaluation methodologies, advocating for contextually-grounded assessment frameworks that serve as genuine catalysts for institutional development rather than mere competitive instruments.

In summary, this study addresses a significant gap in the literature by applying K-means clustering analysis to national university ranking data in the context of a developing country. By doing so, it aims to provide a more nuanced, context-specific, and actionable approach to university performance evaluation, with implications for institutional

strategy, national policy, and the broader field of higher education research.

2 Literature Review

The application of data mining and machine learning techniques to analyze academic performance in higher education has gained significant traction in recent years. This review examines the current state of research in this field, with a particular focus on clustering algorithms and their application to university ranking data.

Educational Data Mining (EDM) has emerged as a powerful tool for extracting valuable insights from the vast amount of data available within higher education institutions. [2] highlight the potential of EDM approaches in enhancing our understanding of academic performance. Building on this, [3] demonstrate how machine learning techniques, including deep neural networks and clustering algorithms, can improve the prediction of student academic performance by considering various factors such as attendance and class participation.

In the context of performance prediction, several studies have explored different methodologies. [4] employed Decision Tree Classification (DTC) models to forecast student outcomes accurately. [5] investigated machine learning methods for developing decision support systems to predict students' grades, while [6] demonstrated the effectiveness of the Naive Bayes classifier algorithm in evaluating student academic performance. The application of machine learning in higher education extends beyond student performance prediction. [7] proposed frameworks integrating performance management and machine learning algorithms to predict markers for student success, faculty productivity, and institutional efficiency. [8] explored multi-category prediction models using Support Vector Machines (SVM) and Random Forest to forecast students' academic achievements. Researchers have also emphasized the importance of leveraging machine learning and deep learning algorithms for performance analysis in higher education computing institutions [9]. [10] and [11] highlight how EDM approaches can guide educational processes and refine learning strategies based on collected datasets.

The critique of international ranking systems is essential, as these often fail to accurately reflect the diverse contexts of universities in developing nations. For instance, the work of Lee et al. highlights how global rankings can lead to strategic misalignments in institutional priorities, particularly in non-Western contexts, where local needs may be overshadowed by the pursuit of status in international rankings [12]. Moreover, the use of clustering analysis, as proposed in the study, can provide a more nuanced understanding of institutional performance. The application of K-means clustering to categorize universities based on performance metrics is supported by the literature on comparative education, which emphasizes the importance of context-specific evaluations. For example, the research by Altinyelken et al. discusses critical debates surrounding Turkish higher education, providing a contextual framework that could inform the clustering analysis in our study [13]. This approach allows for a more tailored evaluation that recognizes the unique trajectories of universities in Türkiye, as opposed to relying solely on global benchmarks.

Several studies provide valuable insights into the application of K-means clustering in analyzing university performance. [14] utilized K-means clustering to analyze key university leadership factors based on international rankings, focusing on the Top 50 universities according to the QS ranking. This approach allowed for the identification of patterns and groupings within the university rankings data. [15] elucidate the fundamental principles of K-means clustering, emphasizing its objective of partitioning observations into clusters based on their proximity to cluster means. This foundational understanding serves as a basis for applying K-means to university rankings data, where the goal is to group universities based on similarities in ranking criteria. [16] introduced the concept of supervised clustering of label ranking data, showcasing the versatility of the K-means algorithm in handling diverse data types. This approach can be particularly useful when dealing with the complex nature of university ranking data. [17] explored the application of various clustering algorithms, including K-means, GMM, Agglomerative, and Fuzzy C-Means, to unveil university groupings based on academic rankings. This comparative approach provides a comprehensive understanding of the similarities and distinctions among universities in terms of their ranking performance. [18] delved into the application of K-means clustering in university libraries, highlighting the iterative process of selecting initial clustering centers, assigning data points to clusters based on similarity, and refining cluster centers until convergence. This iterative nature of K-means clustering aligns well with the dynamic nature of university rankings data.

The literature reveals a growing trend in applying advanced analytical techniques, particularly clustering algorithms, to understand and evaluate academic performance in higher education. However, there remains a gap in the comprehensive application of these techniques to national ranking

systems in developing countries. Our study aims to address this gap by applying K-means clustering to the URAP-TR ranking system in Türkiye, providing a novel approach to understanding and improving the performance of research universities in a developing nation context.

3 Methodology

The methodology employed in this study combines quantitative data analysis with advanced clustering techniques to provide a comprehensive evaluation of research university performance in Türkiye. This section outlines the data sources, metrics, and analytical approaches used to conduct our comparative analysis. We begin by detailing the URAP-TR ranking system and its metrics, which form the foundation of our dataset. Following this, we describe the K-means clustering algorithm and its application to our data, explaining how this method allows us to group universities with similar performance characteristics. We then present the status of research universities based on these metrics, setting the stage for our focused analysis of Istanbul University-Cerrahpaşa (IUC). Throughout this section, we emphasize the rationale behind our methodological choices and how they address the research gaps identified in the introduction. By providing a clear and detailed account of our methodology, we aim to ensure the reproducibility of our study and facilitate future research in this area.

3.1 URAP-TR Metrics

University Ranking by Academic Performance (URAP) Research Laboratory has been ranking universities in Türkiye according to their academic performance since 2009. For a university to be ranked in URAP, it must be among the top 3000 universities with the highest score according to the indicators of the ranking methodology. The indicators used in the general ranking of universities are as shown in **Table 1** for the years 2023-2024.

Table 1 URAP-TR Ranking Indicators and Their Descriptions for 2023-2024

No	Indicator	Objective	Source	Description
1	Number of Articles	Research	InCites	In 2022, the number of articles with a maximum of 1000 authors published in journals included in SCI, SSCI and AHCI scans and ranked in the first 75% (Q1, Q2, Q3) in terms of efficiency multiplier
2	Number of Articles per Academic Staff	Research	InCites and Turkish Council of Higher Education (YÖK)	Number of articles published in 2022 in journals included in SCI, SSCI and AHCI scans and in the first 75% in terms of efficiency multiplier with a maximum of 1000 authors / Number of Faculty Members in 2022-2023
3	Citation Count	Research	InCites	Number of citations to Total Scientific Documents with maximum 1000 authors received between 2018-2022 (All citations to Total Scientific Documents are included in the evaluation)
4	Number of Citations per Academic Staff	Research	InCites and YÖK	Number of citations to Total Scientific Documents with maximum 1000 authors received between 2018-2022 / Number of Faculty Members in 2022-2023
5	Total Number of Scientific Documents	Research	InCites	Total number of publications, papers, etc. made between 2018-2022
6	Total Number of Scientific Documents per Academic Staff	Research	InCites and YÖK	Total number of publications, papers, etc. made between 2018-2022 / Number of Faculty Members in 2022-2023
7	Number of PhD Graduates	Education and Research	YÖK	Number of PhD graduates for the academic year 2021-2022
8	PhD Student Ratio	Education and Research	YÖK	Number of doctoral students in the academic year 2022-2023 / total number of students in the same period
9	Number of Students per Academic Staff	Education	YÖK	Total number of students in 2022-2023 Academic Year / Number of Faculty Members in 2022-2023
10	Number of International Collaborative Papers	Research	InCites	Total number of articles with a maximum of 1000 authors between 2018-2022 with universities of other countries
11	Number of International Collaborative Articles per Academic Staff	Research	InCites	Total number of articles with a maximum of 1000 authors between 2018-2022 with universities of other countries / Number of Faculty Members in 2022-2023
12	Number of Internal Collaborative Papers	Research	InCites	Total number of articles with a maximum of 1000 authors in collaboration with universities in Türkiye between 2018-2022
13	Number of Internal Collaborative Articles per Academic Staff	Research	InCites	Total number of articles with a maximum of 1000 authors in collaboration with universities in Türkiye between 2018-2022 / Number of Faculty Members in 2022-2023
14	Number of projects received from TUBITAK	Project	Scientific And Technological Research Council Of Türkiye (TUBITAK)	Number of projects received from TUBITAK between 2017-2021.
15	Number of projects received from TUBITAK per academic staff	Project	TUBITAK	Number of projects received from TUBITAK between 2017-2021 / Number of faculty members in 2022-2023

3.2 Status Of Research Universities in Türkiye

Research universities play a pivotal role in Türkiye's higher education landscape and national development strategy. Introduced in 2017 by the Council of Higher Education (YÖK), the research university initiative aims to enhance the global competitiveness of Turkish universities and boost the country's research and innovation capacity. Initially, 10 state universities and 5 foundation (private) universities were designated as research universities,

with the list being periodically reviewed and updated based on performance criteria.

Research universities in Türkiye are characterized by their focus on advanced research, innovation, and high-quality graduate education. These institutions are expected to lead in scientific publications, secure international research funding, foster industry collaborations, and contribute significantly to Türkiye's knowledge economy. The designation as a research university comes with additional funding and autonomy, but also with heightened expectations for research output and impact.

To provide a comprehensive view of the status of research universities in Türkiye, we have analyzed their performance across various metrics derived from the URAP-TR ranking system. These metrics include Article Score, Citation Score, Scientific Document Score, Doctorate Score, Faculty Member/Student Score, International Collaboration Score, and TUBITAK Project Score (**Fig. 1**). When examining Article Scores, there is a wide range of performance among institutions. Universities like Middle East Technical University (METU) and Istanbul Technical University (ITU) consistently produce a high volume of articles, showcasing their strong research foundations and productive faculty members. However, the performance across universities is not uniform. Some newer research universities are experiencing rapid growth in their publication output, while others are struggling to keep pace. This variability underscores the need for targeted strategies to enhance research productivity across all designated research universities. Looking at Citation Scores, which indicate the impact and visibility of research, we see a pattern that differs somewhat from Article Scores. While there is generally a correlation between high article output and high citation rates, some universities stand out for having disproportionately high citation impacts compared to their publication volume. This suggests that these institutions may be prioritizing quality over quantity in their research output, possibly by strategically targeting high-impact journals or focusing on particularly influential research areas. The Scientific Document Score provides a broader view of research output beyond just articles. This metric reveals interesting patterns across institutions. Some universities that may not be leading in article production show strengths in other forms of scientific communication, such as conference proceedings or book chapters. This diversity in research output types highlights the varied research cultures and priorities that exist across different institutions. Analyzing the Doctorate Score shows significant variations in both the scale and potentially the quality of doctoral programs across research universities. Well-established institutions like Ankara University and Hacettepe University demonstrate strong performances in this area, likely due to their long-standing graduate programs and extensive faculty resources. However, it's noteworthy that some younger or smaller research universities are making significant progress in expanding their doctoral education capacities. The Faculty Member/Student Score reveals a critical area of diversity among research universities. Some institutions, particularly certain foundation (private) universities, have very favorable ratios of faculty to students. This potentially allows for more personalized instruction and research mentorship. In contrast, many public universities face challenges in this area, with higher numbers of students per faculty member. This disparity points to resource allocation issues that could impact both the quality of teaching and research productivity. One of the most striking disparities among Turkish research universities is revealed by the International Collaboration Score. Certain institutions, notably Koç University and Sabancı University, show exceptionally high levels of international collaboration. This may be attributed to factors such as their use of English as the medium of instruction, their international faculty, and a strategic focus on global partnerships. Many public universities, despite their research designation, show lower levels of international collaboration, indicating a potential area for targeted improvement. Finally, the TUBITAK Project Score offers insights into universities' success in securing national research funding. The picture here is also diverse, with some universities showing particular strength in this area. The high performance of certain private universities like Koç and Sabancı in securing TUBITAK projects is noteworthy and may reflect their focused research strategies and efficient project management structures. This comprehensive analysis highlights the complex nature of research performance in Turkish universities. It reveals areas of excellence, disparities that need addressing, and opportunities for improvement across the higher education sector.

The varied performance across different metrics suggests that universities are adopting diverse strategies in their pursuit of research excellence. Some institutions focus on high-volume article production, others prioritize high-impact publications, while some excel in securing external funding or fostering international collaborations. This diversity of approaches enriches the research ecosystem but also presents valuable opportunities for cross-institutional learning and collaboration. These findings have important policy implications.

The observed imbalances and areas of excellence provide valuable insights for policymakers. There may be a need for more targeted support to help some research universities improve in specific areas, such as international collaboration or doctoral education. Moreover, the success of certain institutions in areas like TUBITAK project acquisition could inform best practices that could be shared and implemented across the sector.

Fig. 1 Comparative Performance of Turkish Research Universities Based on URAP-TR 2023-2024 Total Scores

3.3 K-means Clustering Algorithm

The K-means clustering algorithm is a vector quantization method derived from signal processing [19]. The K-Means method is an algorithm used to group a given data set into k number of clusters. This method assigns data points to clusters and collects data points around the centroid of each cluster. When an n-dimensional data set ($X = \{x_i | i = 1,2, \dots, n\}$) consisting of d-dimensional data points is divided into k clusters $\{c_j | j = 1,2, \dots, k\}$) , the error function of each cluster is defined as follows [20]:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The K-Means algorithm aims to minimise the sum of error squares for each set k:

$$\min J(C) = \sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

Initially, the K-Means algorithm randomly selects k centers from the dataset. The distances of each data point to the selected centers are calculated and each data point is assigned to the nearest center. When a data point is assigned to a new cluster, the cluster center is recalculated. This process continues iteratively until the membership of the clusters stabilizes [20].

4 Clustering of Research Universities According to Their URAP Performance

This study presents a K-Means clustering analysis to group research universities in Türkiye based on their academic and research performance metrics. The primary objective of this analysis is to identify universities with similar performance characteristics and create meaningful groups that can inform strategic decision-making processes in higher education [21]. By clustering universities with similar profiles, we aim to provide a robust foundation for identifying institutional strengths and areas for improvement, facilitating more efficient resource allocation, and developing tailored improvement strategies.

Our clustering analysis is grounded in the metrics provided by the University Ranking by Academic Performance (URAP) system. These metrics include article and citation scores, international collaboration rates, and the quality of doctoral programs. The URAP metrics were chosen for their comprehensive coverage of key aspects of university performance, providing a multidimensional view of each institution's research output, impact, and educational quality. This rich dataset allows for nuanced comparative analyses across the Turkish higher education landscape.

The K-Means clustering algorithm was implemented using Python, leveraging the scikit-learn library for its robust and efficient implementation of the algorithm. The process involved several key steps, beginning with data preprocessing. The URAP metrics for each university were normalized using z-score standardization to ensure that all features contributed equally to the clustering process, regardless of their original scale. We then carefully selected the most relevant URAP metrics for our analysis, focusing on those that best represent research output, impact, and quality. This included metrics such as article score, citation score,

international collaboration score, and doctoral program quality indicators.

To determine the optimal number of clusters (K), we employed the Elbow Method (**Fig. 2**). This involved running the K-Means algorithm for a range of K values (typically from 1 to 10) and calculating the Within-Cluster Sum of Squares (WCSS) for each K. The "elbow" in the resulting plot, where the rate of decrease in WCSS begins to level off, suggests the optimal number of clusters. Our analysis using the Elbow Method suggested that K=3 provides a good balance between cluster cohesion and separation. This choice was further supported by our objective to achieve a meaningful division of universities into categories that could be broadly interpreted as "good," "better," and "best" in terms of research performance. This three-tier categorization aligns well with common practices in performance evaluation and provides a clear, intuitive framework for understanding the landscape of research universities in Türkiye.

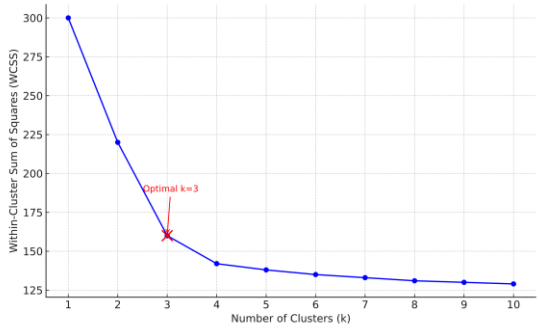

Fig. 2 Elbow Method Analysis for Determining Optimal Number of Clusters in K-Means Algorithm

With K=3, we ran the K-Means algorithm using scikit-learn's K-Means class. The algorithm was initialized using the 'k-means++' method to ensure more stable and optimal starting centroids. We set a high number of initializations (n_init=100) and allowed for a large number of iterations (max_iter=500) to ensure convergence to a global optimum. After running the algorithm, each university was assigned to one of the three clusters. The centroid of each cluster represents the average performance across all metrics for universities in that cluster.

The resulting clusters, as shown in **Table 2**, provide valuable insights into the stratification of research universities in Türkiye. Cluster 1, which we might label as "Best," typically includes the top-performing universities with high scores across most or all URAP metrics. These institutions often have a long-established research culture, significant resources, and a strong international presence. Cluster 2, or "Better," contains universities that show strong performance in many areas but may lag behind Cluster 1 in certain metrics. These institutions often have significant potential for growth and may be on a trajectory to join the top tier. Cluster 3, which we could term "Good," includes universities that, while still designated as research institutions, may face more challenges or have more areas for improvement compared to the other clusters. However, they still demonstrate significant research output and quality. It's important to note that these cluster designations are relative within the context of research universities in Türkiye and should not be interpreted as absolute quality judgments.

Table 2 K-Means Clustering Results for Turkish Research Universities Based on URAP-TR 2023-2024 Metrics

University	Cluster No.
Middle East Technical University	1
Istanbul Technical University	1
Sabancı University	1
Koç University	1
Ege University	2
Gazi University	2
Ankara University	2
Istanbul University-Cerrahpaşa	2
Hacettepe University	2
Istanbul University	2
Bursa Uludağ University	3
Izmir Institute of Technology	3
Karadeniz Technical University	3
Dokuz Eylül University	3
Çukurova University	3
Fırat University	3
Marmara University	3
Ihsan Doğramacı Bilkent University	3
Atatürk University	3
Yıldız Technical University	3
Erciyes University	3
Boğaziçi University	3
Gebze Technical University	3

For policymakers and education administrators, this analysis offers valuable insights into the current landscape of research universities in Türkiye. It highlights areas where targeted interventions or policy changes could have the most significant impact on improving the overall quality and competitiveness of Turkish higher education. Furthermore, this clustering approach provides a framework for universities to benchmark themselves

against peer institutions and set realistic goals for improvement. By understanding their position within their cluster and the characteristics of universities in higher-performing clusters, institutions can develop more focused and effective strategies for advancement.

5 Case Study: Performance Analysis and Strategic Recommendations for Istanbul University-Cerrahpaşa

Istanbul University-Cerrahpaşa (IUC) stands as a prominent institution among research universities in Türkiye, as evidenced by its performance in the URAP 2023-2024 ranking. This comprehensive analysis delves into IUC's current standing, its strengths and areas for improvement, and provides strategic recommendations for enhancing its position in the competitive landscape of Turkish higher education.

The URAP 2023-2024 ranking offers a comprehensive view of IUC's performance relative to other research universities in Türkiye. **Fig. 3** illustrates the total scores of research universities, with IUC highlighted to emphasize its position. This visualization clearly demonstrates that IUC ranks among the top research institutions in the country, reflecting its commitment to academic excellence and research impact.

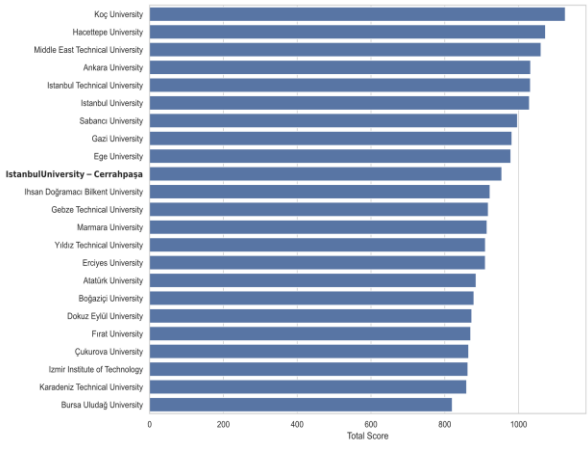

Fig. 3 Total Scores of Turkish Research Universities

A closer examination of IUC's performance across individual URAP metrics reveals a nuanced picture of its strengths and areas for potential improvement:

- Article Score: 6th place
- Citation Score: 9th place
- Scientific Document Score: 11th place
- Doctorate Score: 8th place
- Faculty Member / Student Score: 1st place
- International Collaboration Score: 11th place
- Internal Collaboration Score: 4th place

These rankings provide valuable insights into IUC's relative performance in key areas of academic and research activity. Notably, IUC excels in its faculty-to-student ratio, ranking first among all research universities in Türkiye. This suggests a strong commitment to providing quality education and personalized attention to students. The university also performs well in internal collaboration, indicating robust research partnerships within the country. However, there are areas where IUC has room for improvement, particularly in international collaboration and scientific document production. These areas represent opportunities for strategic focus and development.

To ascend to the next level of excellence, represented by Cluster 1 in our analysis, IUC must set ambitious yet achievable performance targets. Our analysis has identified specific improvement percentages required for IUC to reach the average performance of Cluster 1 universities as shown in **Table 3**.

Table 3 Performance Improvement Targets for Istanbul University-Cerrahpaşa to Reach Cluster 1 Average

TUBITAK Project Score	**+%72**
International Collaboration Score	+%23
Citation Score	+%15
Scientific Document Score	+%12
Article Score	+% 9
PhD Score	%0 (already at par with Cluster 1 average)
Internal Collaboration Score	%0 (already at par with Cluster 1 average)
Academic Staff / Student Score	%0 (already at par with Cluster 1 average)
Total Score	%10

Overall, IUC needs to improve its total score by 10% to reach the average of Cluster 1 universities. This target, while challenging, is within reach given IUC's strong foundation and demonstrated commitment to excellence. Based on the detailed analysis of IUC's performance and the identified targets for improvement, we propose a comprehensive set of strategic recommendations to advance the university's standing. These recommendations span several key areas of academic and research activity.

To enhance research project acquisition, IUC should develop a dedicated task force focused on identifying and pursuing TUBITAK project opportunities. This effort should be supported by providing grant writing workshops and support for faculty members, helping them to craft more competitive proposals. Additionally, fostering interdisciplinary research teams can increase the competitiveness of project proposals, leveraging diverse expertise to address complex research challenges. Boosting international collaboration is another crucial area for improvement. IUC should work towards establishing new partnerships with high-ranking international universities, creating opportunities for joint research and academic exchange. Encouraging faculty participation in international conferences and research exchanges can help build these connections. Creating incentives for joint publications with international collaborators can further strengthen these relationships and increase the university's global research impact. Improving citation impact is essential for enhancing IUC's research profile. This can be achieved by implementing a strategic publication plan that focuses on high-impact journals in relevant fields. Providing resources and training on effective research dissemination techniques can help faculty members increase the visibility and impact of their work. Recognizing and rewarding highly cited researchers can also encourage the production of impactful research. To increase scientific document production, IUC should set departmental targets for research output, providing a clear goal for faculty members. Offering writing support and editing services for faculty and graduate students can help improve the quality and quantity of publications. Organizing regular research symposiums can foster a culture of active research and publication, providing opportunities for internal collaboration and idea exchange. Enhancing article output requires supporting faculty at all career stages. Implementing a mentoring program that pairs junior faculty with experienced researchers can help new academics develop their research skills and networks. Providing teaching load reductions for faculty actively engaged in high-quality research can give them more time to focus on their scholarly work. Establishing a university-wide research day to showcase ongoing projects can encourage collaboration and inspire new research initiatives.

While focusing on areas for improvement, it's crucial that IUC maintains its current strengths. The university should continue to support its strong faculty-to-student ratio through strategic hiring and resource allocation. Further developing internal collaboration networks can help maintain the university's strong position in this area, fostering a vibrant research community within the institution.

Revitalizing PhD programs is another key area for development. IUC should conduct a comprehensive review of existing PhD programs to ensure they align with current research trends and demands. Strengthening mentoring programs for doctoral students can improve completion rates and research quality. Increasing research funding opportunities for PhD candidates can attract top talent and support innovative research. Enhancing the quality of PhD education through innovative curriculum design and research-intensive coursework can prepare graduates for successful careers in academia and industry.

To ensure the success of these strategic initiatives, IUC should implement a robust system for implementation and monitoring. This includes establishing a dedicated committee to oversee the implementation of these recommendations, ensuring that progress is tracked, and adjustments are made as needed. Developing key performance indicators (KPIs) for each area of improvement can provide clear metrics for success. Regular progress reviews will allow the university to adjust strategies as needed, responding to changes in the academic landscape and internal capabilities.

6 Discussion

The application of K-means clustering to the URAP-TR metrics has provided valuable insights into the landscape of research universities in Türkiye, offering a more nuanced understanding than traditional ranking systems alone. This approach has allowed us to identify distinct groups of universities with similar performance characteristics, facilitating more targeted strategies for improvement and policy development.

Our analysis revealed three distinct clusters among Turkish research universities, which we have broadly categorized as "Best," "Better," and "Good." This stratification highlights the diversity within the Turkish higher education system and underscores the need for tailored approaches to university development and evaluation.

The "Best" cluster, comprising institutions like Middle East Technical University and Istanbul Technical University, demonstrates consistently high performance across multiple metrics. These universities are characterized by their strong research cultures, significant research outputs, and high levels of international collaboration. Their success in securing TUBITAK projects and their high citation impacts suggest that they have effectively aligned their research strategies with national priorities and global academic trends.

The "Better" cluster, which includes Istanbul University-Cerrahpaşa, represents institutions with strong overall performance but with specific areas for improvement. These universities often excel in certain metrics but lag behind the top cluster in others. For instance, IUC's top ranking in faculty-to-student ratio demonstrates its commitment to quality education and personalized attention. However, its lower rankings in international collaboration and scientific document production indicate areas where strategic interventions could yield significant improvements.

The "Good" cluster encompasses universities that, while still designated as research institutions, face more significant challenges in competing with the top-tier universities. These institutions often have strengths in specific areas but require more comprehensive strategies to enhance their overall research performance and impact.

This clustering approach reveals several key insights about the Turkish research university landscape:

1. Diversity of Strengths: Each cluster, and indeed each university within the clusters, exhibits unique strengths. This diversity is a valuable asset for the Turkish higher education system, as it allows for a range of specializations and areas of excellence across the country.

2. Resource Allocation Disparities: The analysis highlights significant differences in resource allocation and utilization between public and private universities. Private institutions like Koç and Sabancı Universities consistently outperform in areas such as international collaboration and project acquisition, suggesting that their more flexible administrative structures and focused research strategies may offer lessons for public institutions.

3. International Collaboration Gap: A striking finding is the disparity in international collaboration scores, particularly between private and public universities. This suggests a critical area for improvement for many institutions, as international collaboration is increasingly crucial for enhancing research quality, visibility, and impact.

4. TUBITAK Project Success: The variation in TUBITAK project scores indicates differing levels of alignment with national research priorities or disparities in the capacity to secure competitive funding. Improving performance in this area could be a key lever for enhancing overall research output and impact.

5. Citation Impact Variations: The analysis reveals that some universities achieve high citation impacts relative to their publication volume, suggesting a focus on quality over quantity. This observation provides valuable insights for institutions looking to enhance their research impact.

The case study of Istanbul University-Cerrahpaşa offers a concrete example of how this clustering analysis can inform strategic planning at the institutional level. IUC's position in the "Better" cluster, combined with its specific performance metrics, provides a clear roadmap for improvement. The identified target improvements, such as a 72% increase in TUBITAK project score and a 23% increase in international collaboration score, offer quantifiable goals that can guide resource allocation and policy decisions.

The recommendations provided for IUC, such as developing a dedicated task force for TUBITAK projects, establishing new international partnerships, and implementing strategic publication plans, demonstrate how data-driven insights can be translated into actionable strategies. These recommendations are not only applicable to IUC but can serve as a template for other universities looking to enhance their research performance.

Moreover, this analysis has broader implications for higher education policy in Türkiye. The clear stratification of universities and the identification of common challenges across clusters can inform national-level interventions. For instance, the widespread need for improved international collaboration could prompt initiatives to facilitate global partnerships or provide funding for international research projects.

The study also highlights the potential limitations of relying solely on traditional ranking systems. While rankings provide valuable comparative data, the clustering approach offers a more nuanced understanding of university performance, accounting for the multifaceted nature of research output and impact. This suggests that policymakers and university administrators should consider adopting more sophisticated evaluation methods to complement existing ranking systems.

7 Conclusion

This study presents a novel approach to analyzing the performance of research universities in Türkiye, utilizing K-means clustering on URAP-TR metrics to provide a more nuanced and actionable understanding of the higher education landscape. By grouping universities with similar performance characteristics, we have identified distinct clusters that reflect the diversity and complexity of the Turkish research university sector. The analysis reveals both the strengths and challenges faced by Turkish research universities. While some institutions demonstrate world-class performance across multiple metrics, others show potential for significant improvement, particularly in areas such as international collaboration and research project acquisition.

The case study of Istanbul University-Cerrahpaşa exemplifies how this clustering approach can inform strategic planning at the institutional level. By identifying specific areas for improvement and setting quantifiable targets, IUC and similar institutions can develop focused strategies to enhance their research performance and impact.

This research makes significant contributions to the ongoing discussion about how universities are evaluated and improved. By applying sophisticated analytical methods to university performance data, our study reveals insights that traditional ranking systems often miss. This approach demonstrates the value of using advanced techniques to gain a deeper understanding of university performance. Our work also serves as a model that other countries can follow, especially those with developing higher education systems. It shows how nations can assess their research universities and develop strategic plans for improvement based on data-driven insights. This is particularly valuable for countries looking to enhance their global competitiveness in higher education. Furthermore, our research emphasizes the importance of taking a comprehensive approach to university performance evaluation. Rather than relying on a few simple metrics, we advocate for a more comprehensive assessment that considers a wide range of factors. This holistic approach allows for a more nuanced understanding of each institution's strengths and weaknesses, providing a fuller picture of their overall performance and potential.

Importantly, our study offers practical, actionable insights for both policymakers and university administrators. By providing a data-driven approach to resource allocation and strategic planning, we enable decision-makers to make more informed choices. This can lead to more effective policies and strategies for improving university performance, ultimately enhancing the quality of higher education and research output. In essence, our research not only advances the academic understanding of university performance evaluation but also provides practical tools and insights that can be applied to real-world scenarios. It bridges the gap between theoretical analysis and practical application, offering a pathway for substantial improvements in higher education systems.

The findings of this study have significant implications for higher education policy in Türkiye. They suggest a need for targeted interventions to address common challenges across universities, such as enhancing international collaboration and improving success rates in securing research funding. Moreover, the clear stratification of universities indicates that a one-size-fits-all approach to higher education policy may be ineffective, and that policies should be tailored to the specific needs and potentials of different university clusters.

Future research could expand on this work by incorporating additional metrics, such as industry collaboration or graduate employment outcomes, to provide an even more comprehensive view of university performance. Longitudinal studies could also track the effectiveness of interventions over time, providing valuable data on the impact of strategic changes.

In conclusion, this study provides a robust framework for understanding and improving the performance of research universities in Türkiye. By combining the strengths of national ranking systems with advanced clustering techniques, it offers a powerful tool for strategic decision-making in higher education. As Türkiye continues to invest in its research capabilities and seeks to enhance its global competitiveness in higher education, approaches like the one presented in this study will be crucial in guiding effective policy and institutional strategies.

Acknowledgement:
Our study introduces a novel approach to evaluating research universities in developing countries, using Turkey as a case study within the broader context of international higher education trends. By combining national ranking metrics with K-means clustering analysis, we address the limitations of global ranking systems in assessing institutions outside the Global North. The manuscript presents:

1. A new methodological approach combining national metrics with advanced clustering techniques
2. Insights into the stratification of research universities in a developing country context
3. A case study demonstrating practical applications for institutional improvement
4. Implications for higher education policy and management in similar contexts worldwide

References:

[1] Alaşehir, O., Çakır, M. P., Acartürk, C., Baykal, N., & Akbulut, U., URAP-TR: a national ranking for Turkish universities based on academic performance, *Scientometrics*, Vol.101, 2014, pp.159-178.

[2] Trakunphutthirak, R., & Lee, V., Application of educational data mining approach for student academic performance prediction using progressive temporal data, Journal of Educational *Computing Research*, Vol.60, No.3, 2021, pp.742-776.

[3] Sikder, M., Hosen, M., Fatema, K., & Islam, M., Predicting students' performance in final examination using deep neural network, *Asian Journal of Research in Computer Science*, 2022, pp.218-227.

[4] LU, X., Modern education: advanced prediction techniques for student achievement data, *International Journal of Advanced Computer Science and Applications*, Vol.15, No.1, 2024.

[5] Kotsiantis, S., Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades, *Artificial Intelligence Review*, Vol.37, No.4, 2011, pp. 331-344.

[6] Umarova, Z., Use of the naive bayes classifier algorithm in machine learning for student performance prediction, *International Journal of Information and Education Technology*, Vol.14, No.1, 2024, pp.92-98.

[7] Osei-Asiamah, J., Towards a framework for performance management and machine learning in a higher education institution, *JIER*, Vol.4, No.2, 2024.

[8] Ahmed, A., Multi-category prediction of students' academic performance using machine learning: for students joining higher educational institution in Ethiopia, *Research Square*, 2023.

[9] Rahman, N., Exploring the research trend on student performance in education computing: a bibliometric analysis, *MyJICT*, 2022, pp.11-21.

[10] Kannan, R., Predicting student's soft skills based on socio-economical factors: an educational data mining approach, JOIV: *International Journal on Informatics Visualization*, Vol.7, No.3-2, 2023, pp.2040.

[11] Zhao, L., Quantitative analysis and prediction of academic performance of students using machine learning, *Sustainability*, Vol.15, No.16, 2023, pp.12531.

[12] Lee, J. J., Vance, H., Stensaker, B., & Ghosh, S., Global rankings at a local cost? the strategic pursuit of status and the third mission, *Comparative Education*, Vol.56, No.2, 2020, pp. 236-256.

[13] Altinyelken, H. K., Çayır, K., & Ağırdağ, O., Turkey at a crossroads: critical debates and issues in education, *Comparative Education*, Vol.51, No.4, 2015, pp. 473-483.

[14] Polyakov, M., Bilozubenko, V., Korneyev, M., & Nebaba, N., Analysis of key university leadership factors based on their international rankings (QS World University Rankings and Times Higher Education), *Problems and Perspectives in Management*, Vol.18, No.4, 2020, pp.142-152.

[15] Sundar, B., Devi, T., & Saravanan, N., Development of a data clustering algorithm for predicting heart, *International Journal of Computer Applications*, Vol.48, No.7, 2012, pp.8-13.

[16] Grbovic, M., Djuric, N., & Vucetic, S., Supervised clustering of label ranking data, 2012.

[17] Matlis, G., Unveiling university groupings: a clustering analysis for academic rankings, *Data*, Vol.9, No.5, 2024, pp.67.

[18] Wang, R., Tang, Y., Liu, G., & Li, Y., K-means clustering algorithm application in university libraries, 2011.

[19] Hu H., Liu J., et. al., An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis, *Pattern Recognition*, Vol.139, 2023.

[20] Ikotun A. M., et. al., K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data, *Information Sciences*, Vol. 622, 2023, pp. 178-210.

[21] Ozdemir, O., & Karakutuk, A. K., Comparison of fuzzy clustering algorithms on UI GreenMetric university rankings, *International Journal of Applied Mathematics, Computational Science and Systems Engineering*, Vol.6, 2024, pp.230-237.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.