

Review

Not peer-reviewed version

---

# A Systematic Literature Review on the Evolution of Skyline Query on Uncertain Database: Trends and Insights

---

[H. M. Ikram Kays](#)<sup>\*</sup>, Raini Hassan , Dini Oktarina Dwi Handayani

Posted Date: 23 December 2025

doi: 10.20944/preprints202512.2082.v1

Keywords: distributed systems; probabilistic databases; query optimization; skyline queries; uncertain data



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# A Systematic Literature Review on the Evolution of Skyline Query on Uncertain Database: Trends and Insights

H. M. Ikram Kays, Raini Hassan and Dini Oktarina Dwi Handayani

Department of Computer Science, Kulliyah of Information and Communication Technology  
International Islamic University Malaysia, Kuala Lumpur, Malaysia

\* Correspondence: ikramkays@gmail.com

## Abstract

Skyline Query, one of the profound tools that holds up tremendously when it comes to extracting valuable insights has witnessed multiple significant evolutions both in application domain and problem complexity over the years. In this SLR (Structural Literature Review), this study has tried to investigate the trends, evolutions of the application domain, and problem complexity from as early as 2008 until now. The authors divided the timespan into three major periods and analyzed 28 Scopus-indexed papers which this study chose using the PRISMA methodology. When looking at insights on application domain evolution, in the early years fundamental algorithmic research was taking place and it gradually shifted towards more specialized applications such as smart cities, IoT, and distributed computing. As the domains evolved, the complexity of the problems also spiked as a need to handle higher dimensionality in data, larger volume, and increased uncertainty became apparent. This paper provides impactful insights into how skyline query research domains have changed and tries to highlight future directions for addressing newer and more complex data management challenges.

**Keywords:** distributed systems; probabilistic databases; query optimization; skyline queries; uncertain data

---

## I. Introduction

Skyline queries, an essential method of extracting useful insights from high-dimensional and uncertain data, have attracted significant focus in the field of data management research. In the last decade, the skyline query field has grown rapidly, fueled by the ever-growing demands of managing data in real-time requirements and technological advances.

The understanding of how the applications of research on skyline queries have developed as well as how the nature of the issues changed will be crucial in determining future research direction and dealing with new challenges to managing and processing data.

This paper will analyze the Skyline Query research from the year 2008 until 2024. This study has focused on studying the evolution of the application domains as well as the problem complexity. By conducting a thorough study of 28 papers that span three distinct periods of time this study aims to illuminate the most important driving factors and trends.

The papers are classified by key domains of application as well as the types of uncertainties they address the dimensionality of data as well as volume and the main problems identified by the authors. This paper not only reveals the shift from basic algorithmic enhancements to advanced applications but also highlights the increasing difficulty of the problems that are addressed. In shedding light on these developments our goal is to provide a greater understanding of the changing characteristics and help future researchers to create guidelines on future directions for research in skyline query over uncertain databases.

## II. Methodology

### A. Research Questions

- How have the application domains of skyline query on uncertain databases evolved over the years?
- How has the complexity of problems addressed in skyline queries on uncertain databases changed over the years?

### B. Search Strategy

For this paper, the authors developed a search strategy to identify relevant literature. This search strategy was retrieved from the SCOPUS database using the advanced search strategy,

(TITLE-ABS-KEY (skyline OR "skyline queries" OR "Skyline query" ) AND TITLE-ABS-KEY ( "uncertain" OR "graph database" ) ) AND ( LIMIT-TO ( SUBJAREA, "COMP" ) ) AND ( LIMIT-TO ( DOCTYPE, "ar" ) ) AND ( LIMIT-TO ( EXACTKEYWORD, "Skyline Query" ) OR LIMIT-TO ( EXACTKEYWORD, "Query Processing" ) OR LIMIT-TO ( EXACTKEYWORD, "Uncertain Data" ) )

### C. Selection Criteria

The Selection Criteria were based on the PRISMA statement. The Search mainly focused on mapping existing literature on Skyline queries in the fields of computer science, information technology, and engineering. The search was then narrowed to the subject area of computer science.

The title was restricted to "skyline", "skyline queries" and "uncertain" or "graph database" so that authors could filter out publications that match our SLR targets. Also, the document type was set to articles only. And a keyword filter was set that limits the search to documents that have either "Skyline Query", "Query Processing" or "Uncertain Data" as exact keywords.

**Table I.** Inclusion and exclusion criteria.

| No | Inclusion  | Exclusion                                     |
|----|--|---|
| 1  | Articles retrieved from the SCOPUS database using the specified search strategy.                               |   |
| 2  | Articles related to Skyline queries in the fields of computer science, information technology, and engineering |   |
| 3  | Articles specifically focused on computer science.   |   |
| 4  | Articles categorized as articles only  | Articles categorized as other document types  |
| 5  | Articles with exact keywords "Skyline Query" or "Uncertain Data"   | Articles without the exact keywords specified |

### D. Result

A total of 54 articles were found in Scopus, and after doing a full-text assessment, 28 journal papers were selected for this systematic literature review.

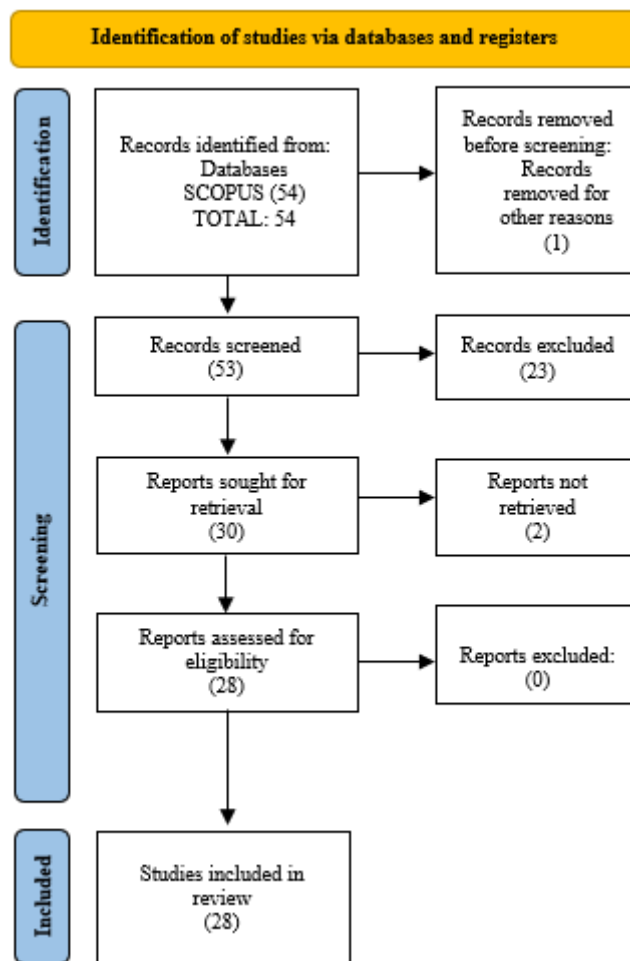


Figure 1. PRISMA flow diagram of the paper selection process.

### III. Results and Discussion

#### A. Evolution of Application Domains in Skyline Query Research (RQ1)

The domains of skyline query research have significantly changed in the past, demonstrating the ever-changing requirements for data management as well as the growing complexity of applications in real life. At the beginning it was focused primarily on the fundamental aspects, dealing with issues that were posed by the uncertainty of information. As the field grew more mature, researchers began exploring deeper and more complex applications.

The initial period defined for this paper ran from 2008 until 2012, the study primarily was focused on methodological and theoretical advancements. For example, earlier research on Probabilistic Group Nearest Neighbor and Efficient Evaluation of Probabilistic Advanced Spatial Queries laid the foundations by looking at the uncertainties of group queries as well as advanced spatial queries [9,16]. These studies played a crucial role in developing the foundational concepts of handling uncertain data within database systems. The domains of application at the time were usually wide and did not have a high degree of specificity. They were more focused on enhancing the efficiency of queries and tackling the uncertainty of data in general-purpose databases. Further research has been done to improve the performance of existing algorithms, resulting in more and more refined and efficient algorithms [6,21].

As research progressed and the field grew, application domains became increasingly diverse and specific. In 2010, researchers began to cover specialization in applications. Reverse Skyline Search and Continuous Monitoring of Skylines over Uncertain Data are two examples of this trend [3,20].

These are essential for the areas of environmental monitoring, media preference [7], financial analysis as well as other fields that require immediate data processing and decision-making.

The time period between 2013 and 2017 was characterized by a major shift toward more applied research which is characterized by a steady trend towards dealing with uncertainties in certain areas. As an example, researchers stressed the significance of skyline queries within intelligent transportation systems, highlighting the increasing interest in intelligent cities and applications [4]. Similarly, Trustworthy Answers for Top-K Queries as well as the GDPS approach tackled the problems of big data as well as distributed computing [12,26]. Researchers also highlighted the analysis of uncertain time series data in fields such as traffic flow management, meteorology, astronomy, remote sensing, market analysis, and object tracking [1,17]. This research reveals a rise in the use of skyline queries in the real world in which data volumes, speed, and diversity pose major problems.

In the most recent time frame which spans from 2018 through 2024, application domains have become more sophisticated and complex. Studies that include Top K probabilistic Skyline Queries, Spatial Skyline Queries for Smart Cities as well as probabilistic similarity search demonstrate the focus on tackling uncertainties in particular situations like smart cities, massive IoT settings, data integrations as well as sensor networks [14,23,25]. Skyline queries inclusion in these fields demonstrates their importance in the decision-making process, where information can be incomplete and rapidly evolving.

One of the most prominent developments in the last few years has been the growing importance of the use of parallel and distributed computing in order to handle the enormous amount of contemporary data. Parallel Computation using MapReduce and ProbSky approach illustrates how the field is evolving to the new challenges presented by the rise of data as well as the necessity to process data in real-time for distributed systems [15,18]. These studies are of particular importance when it comes to edge computing as well as IoT in which data is created at unprecedented levels and has to be efficiently processed.

Furthermore, research is also beginning to include factors that affect the quality of service as well as user preferences. There are studies that underscore the significance of not just processing data effectively but also ensuring the outcomes are useful and applicable for the end-user which is essential for applications like e-commerce, personalization services, as well as SaaS platforms [8,22,28].

In sum, the areas of skyline query research have shifted from general-purpose database management to highly specialized and contextually specific apps. This shift is indicative of wider developments in the field of data management and processing driven by the ever-growing magnitude and complexity of data across a variety of real-world situations. Research has gradually incorporated developments in distributed computation, real-time processing as well as user-centric aspects, which makes skyline queries essential to use in the modern-day data-driven application.

#### *B. Changes in Problem Complexity in Skyline Query Research (RQ2)*

The difficulty of the problems that are faced in Skyline Queries has increased significantly throughout the years. The change is evident in increasing attention towards handling more complex uncertain components in data. Researchers continue to develop new algorithms and methods to tackle these ever-growing complex issues, which reflect the evolving technological capabilities and demands of the applications that are real.

In the initial period, the complexity was to effectively process uncertain data within traditional databases. Research studies have focused on developing fundamental algorithms that can handle uncertain data in a streamlined manner [2,10]. The primary challenges faced this time were due to the computation overhead that came with uncertainty, and the necessity for optimizing query processing so that it could manage large-scale data efficaciously. The issues were daunting however, they were mostly restricted to improving performance in fairly controlled conditions.

When research advanced into the time period between 2013 and 2017 the difficulty of challenges that were addressed raised substantially. This was due to the requirement to manage increasingly dynamic and diverse data environments. Research on Continuous Probabilistic Skyline Queries for Uncertain Moving Objects in Road Networks and the GDPS approach brought new issues to data processing in real-time and distributed computation [4,12]. The research had to tackle the challenges of constantly changing the skyline value in accordance with new information and also ensuring that queries would scale effectively over many nodes of the distributed system. Another study of the SkyQUD algorithm reveals that it is a master at processing skyline query queries because of the efficient partitioning and effective pruning method but still, there were limitations in performance such as memory consumption and scalability [11]. The issues addressed in these studies were not only complex in computational terms, but they also demanded creative approaches to control information distribution as well as real-time updates.

In recent years research has been conducted to create algorithms that could process massive data sets in parallel and leverage tools such as MapReduce for distributing the workload of computation efficiently [15,25]. The difficulty here lies not only in the amount of data but also in making sure the distributed processing process is effective and that the payoff has been accurate and pertinent.

In addition, the inclusion of quality and user preference factors added another layer of difficulty. Studies such as Skyline Recommendation with Uncertain Preferences and Uncertain Big QoS Data-Driven Efficient SaaS Decision-Making Method dealt with the problem of integrating personal preferences from users into the search process [23,28]. It is not just about addressing all the uncertainties inherent in data, but also making sure that the end outcome is in line with the expectations of users and brings valuable insight. The algorithms used in these studies were required to be effective and versatile, able to adapt to changing requirements of users and the quality of data.

A greater focus on particular applications has also contributed to the complexity of the issues tackled. Research on Spatial Skyline Queries for Smart Cities and Distributed Indexing Schemes for Uncertain Edge-IoT Data, multi-criteria decision-making tackled the challenges in the real-world scenario and tried to come up with new solutions for dealing with uncertain data while connecting many data sources as well as process data in real-time for distributed systems [5,13,19,26]. This is a multi-faceted problem that includes not only computation aspects but also the managing and integration of various information sources.

In sum, the difficulty of issues addressed in skyline queries has grown due to the growing demands of today's data-driven environments. Initial research focused on the foundational algorithmic enhancements to manage complex probabilistic data effectively. The field grew more advanced, and the problems became more complicated, including the use of real-time processing and distributed computing and the interplay of the preferences of users with the quality of data. The latest research requires the need to come up with sophisticated algorithms that can handle such requirements. These developments reflect broad trends in the management of data and its processing due to the exponential expansion of data as well as the necessity for efficient flexible, user-centric, and scalable solutions.

### *C. Synthesis of Application Domains and Problem Complexity*

The growth of the application domains as well as the growing complexity of skyline query research are tightly interwoven. This can be seen in the research trajectory throughout the years, in which advances in one field often resulted in innovations in the next.

Early on, the wide and fundamental study of the topic allowed researchers to devise fundamental algorithms and methods to handle the probabilistic nature of data. The foundational research studies created the basis for dealing with larger and more complicated problems as areas of application grew. As an example, research to develop efficient algorithms to handle probabilistic group queries [2,3] opened the way for the development of more complicated applications such as real-time monitoring as well as distributed data processing in later decades [4,12]. Needing to manage real-time information in automated transportation systems [10] and distributed big data systems [12]

introduced new challenges that posed a problem for scalability and effectiveness. Efficiency of Optimized pruning methods and effective computational strategies, new strategies and the most advanced algorithms to process huge amounts of data in large-scale and distributed systems is crucial in real-time processing as well as scalable analysis to assure effective decision-making and optimization of performance [17].

The latest research has expanded limits further, merging user preferences with quality issues, and addressing situations in the real world, such as smart cities as well as IoT environments. The result is a new dimension of complexity and demands algorithms that are not just reliable but also flexible to the changing requirements of users and the quality of data [23,27]. A very short and brief overview of the key elements and focus of Skyline Query can be viewed in the table below.

**Table II.** Evolution of application domains and problem complexity in skyline query research (2008-2024).

| Year Range | Papers                        | Types of Uncertainties       | Data Dimensionality & Volume                   | Main Challenges   | Key Application Domains   |
|------------|-------------------------------|------------------------------|--|---|---|
| 2008-2012  | [2,3,6,9,16,20,24]            | Probabilistic, existential   | High-dimensional data, manageable volumes      | Efficient query processing, foundational algorithms               | General-purpose databases, foundational algorithmic improvements                      |
| 2013-2017  | [1,4,7,10,12,17,19,21,26,27]  | Real-time, distributed       | Dynamic, large volumes                         | Real-time processing, distributed systems                         | Intelligent transportation systems, smart cities, big data, distributed computing     |
| 2018-2024  | [5,8,11,13–15,18,22,23,25,28] | Incomplete, user preferences | Massive volumes, high-dimensional, distributed | Parallel computation, quality of service, user-centric algorithms | Smart cities, IoT environments, personalized services, SaaS platforms, edge computing |

## IV. Conclusion

In the end, the development of the application domains as well as the complexity that is growing with skyline queries represent the general trends in managing data and processing. Initial research focused on the foundational enhancements that laid the base for the development of more sophisticated issues in specific application domains. As technology has advanced the problems have become more complicated, including real-time computation as well as distributed computing and the consideration of users. The constant evolution of the field underscores the crucial role played by Skyline queries in today's information systems and has led to the development of new approaches in algorithms and strategies to address the increasing requirements in the real world.

**Acknowledgments:** This research was supported by the Fundamental Research Grant Scheme (FRGS) with the Reference Code FRGS/1/2021/ICT01/UIAM/02/2 from the Ministry of Higher Education (MOHE) Malaysia.

## References

1. M. Nhut, J. Cao, and Z. He, "Answering skyline queries on probabilistic data using the dominance of probabilistic skyline tuples," *Information Sciences*, vol. 340–341, pp. 58–85, May 2016, doi: <https://doi.org/10.1016/j.ins.2016.01.016>.
2. M. J. Atallah, Y. Qi, and H. Yuan, "Asymptotically efficient algorithms for skyline probabilities of uncertain data," *ACM Transactions on Database Systems*, vol. 36, no. 2, May 2011, doi: <https://doi.org/10.1145/1966385.1966390>.
3. X. Ding, X. Lian, L. Chen, and H. Jin, "Continuous monitoring of skylines over uncertain data streams," *Information Sciences*, vol. 184, no. 1, pp. 196–214, Feb. 2012, doi: <https://doi.org/10.1016/j.ins.2011.09.007>.

4. S. Pan, Y. Dong, J. Cao, and K. Chen, "Continuous probabilistic skyline queries for uncertain moving objects in the road network," *International Journal of Distributed Sensor Networks*, vol. 2014, 2014, doi: <https://doi.org/10.1155/2014/365064>.
5. C. C. Lai, H. Y. Lin, and C. M. Liu, "Distributed indexing schemes for k-Dominant skyline analytics on uncertain edge-iot data," *IEEE Transactions on Emerging Topics in Computing*, 2023, doi: <https://doi.org/10.1109/TETC.2023.3326295>.
6. X. Ding and H. Jin, "Efficient and progressive algorithms for distributed skyline queries over uncertain data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 8, pp. 1448–1462, 2012, doi: <https://doi.org/10.1109/TKDE.2011.77>.
7. A. K. Pujari, V. R. Kagita, A. Garg, and V. Padmanabhan, "Efficient computation for probabilistic skyline over uncertain preferences," *Information Sciences*, vol. 324, pp. 146–162, Dec. 2015, doi: <https://doi.org/10.1016/j.ins.2015.06.041>.
8. N. Sukhwani, V. R. Kagita, V. Kumar, and S. Kumar Panda, "Efficient computation of top-k skyline objects in data set with uncertain preferences," *International Journal of Data Warehousing and Mining*, vol. 17, no. 3, pp. 68–80, Jul. 2021, doi: <https://doi.org/10.4018/IJDWM.2021070104>.
9. M. L. Yiu, N. Mamoulis, X. Dai, Y. Tao, and M. Vaitis, "Efficient evaluation of probabilistic advanced spatial queries on existentially uncertain data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 108–122, Jan. 2009, doi: <https://doi.org/10.1109/TKDE.2008.135>.
10. X. Lian and L. Chen, "Efficient processing of probabilistic group subspace skyline queries in uncertain databases," *Information Systems*, vol. 38, no. 3, pp. 265–285, 2013, doi: <https://doi.org/10.1016/j.is.2012.08.006>.
11. H. Mohd, H. Ibrahim, F. Sidi, R. Yaakob, and A. A. Alwan, "Efficient skyline computation on uncertain dimensions," *IEEE Access*, vol. 9, pp. 96975–96994, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3094547>.
12. X. Li, Y. Wang, X. Li, X. Wang, and J. Yu, "GDPS: An efficient approach for skyline queries over distributed uncertain data," *Big Data Research*, vol. 1, pp. 23–36, Aug. 2014, doi: <https://doi.org/10.1016/j.bdr.2014.07.003>.
13. Y. Zeng, G. Chen, K. Li, Y. Zhou, X. Zhou, and K. Li, "M-Skyline: Taking sunk cost and alternative recommendation in consideration for skyline query on uncertain data," *Knowledge-Based Systems*, vol. 163, pp. 204–213, Jan. 2019, doi: <https://doi.org/10.1016/j.knosys.2018.08.024>.
14. X. Miao, Y. Gao, L. Zhou, W. Wang, and Q. Li, "Optimizing quality for probabilistic skyline computation and probabilistic similarity search," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1741–1755, Sep. 2018, doi: <https://doi.org/10.1109/TKDE.2018.2805824>.
15. E. Gavagsaz, "Parallel computation of probabilistic skyline queries using MapReduce," *Journal of Supercomputing*, vol. 77, no. 1, pp. 418–444, Jan. 2021, doi: <https://doi.org/10.1007/s11227-020-03279-x>.
16. X. Lian and L. Chen, "Probabilistic group nearest neighbor queries in uncertain databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 6, pp. 809–824, Jun. 2008, doi: <https://doi.org/10.1109/TKDE.2008.41>.
17. G. He, L. Chen, C. Zeng, Q. Zheng, and G. Zhou, "Probabilistic skyline queries on uncertain time series," *Neurocomputing*, vol. 191, pp. 224–237, May 2016, doi: <https://doi.org/10.1016/j.neucom.2015.12.104>.
18. A. T. Kuo, H. Chen, L. Tang, W. S. Ku, and X. Qin, "ProbSky: Efficient computation of probabilistic skyline queries over distributed data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 5173–5186, May 2023, doi: <https://doi.org/10.1109/TKDE.2022.3151740>.
19. G. Xiao, K. Li, and K. Li, "Reporting l most influential objects in uncertain databases based on probabilistic reverse top-k queries," *Information Sciences*, vol. 405, pp. 207–226, Sep. 2017, doi: <https://doi.org/10.1016/j.ins.2017.04.028>.
20. X. Lian and L. Chen, "Reverse skyline search in uncertain databases," *ACM Transactions on Database Systems*, vol. 35, no. 1, Feb. 2010, doi: <https://doi.org/10.1145/1670243.1670246>.
21. S. Elmi, M. Anis, A. Hadjali, and B. B. Yaghlane, "Selecting skyline stars over uncertain databases: Semantics and refining methods in the evidence theory setting," *Applied Soft Computing Journal*, vol. 57, pp. 88–101, Aug. 2017, doi: <https://doi.org/10.1016/j.asoc.2017.03.025>.

22. V. R. Kagita, A. K. Pujari, V. Padmanabhan, and V. Kumar, "Skyline recommendation with uncertain preferences," *Pattern Recognition Letters*, vol. 125, pp. 446–452, Jul. 2019, doi: <https://doi.org/10.1016/j.patrec.2019.06.002>.
23. S. Elmi and J. K. Min, "Spatial skyline queries over incomplete data for smart cities," *Journal of Systems Architecture*, vol. 90, pp. 1–14, Oct. 2018, doi: <https://doi.org/10.1016/j.sysarc.2018.08.005>.
24. W. Zhang, X. Lin, Y. Zhang, M. A. Cheema, and Q. Zhang, "Stochastic skylines," *ACM Transactions on Database Systems*, vol. 37, no. 2, May 2012, doi: <https://doi.org/10.1145/2188349.2188356>.
25. Z. Yang, K. Li, X. Zhou, J. Mei, and Y. Gao, "Top k probabilistic skyline queries on uncertain data," *Neurocomputing*, vol. 317, pp. 1–14, Nov. 2018, doi: <https://doi.org/10.1016/j.neucom.2018.03.052>.
26. H. T. H. Nguyen and J. Cao, "Trustworthy answers for top-k queries on uncertain Big Data in decision making," *Information Sciences*, vol. 318, pp. 73–90, Oct. 2015, doi: <https://doi.org/10.1016/j.ins.2014.08.065>.
27. X. Liu, D. N. Yang, M. Ye, and W. C. Lee, "U-skyline: A new skyline query for uncertain databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 945–960, 2013, doi: <https://doi.org/10.1109/TKDE.2012.33>.
28. L. Zhang and J. Bai, "Uncertain big QoS data-driven efficient SaaS decision-making method," *IEEE Access*, vol. 12, pp. 11196–11216, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3355469>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.