

1 Leading Data Science Journals: The Top 20th Percentile

1.1 Methodology for Identification and Ranking

The identification of the top 20th percentile of journals relevant to Data Science involved a multi-step process designed to capture the breadth of the field while applying a consistent quality threshold:

1. **Data Retrieval:** Lists of Scopus-indexed journals, along with their 2024 SJR scores, were conceptually retrieved for each of the four selected primary subject categories (Computer Science: Artificial Intelligence; Computer Science: Information Systems; Mathematics: Statistics and Probability; Decision Sciences) using the SCImago Journal & Country Rank database platform.
2. **List Combination:** The individual lists were merged into a single master list containing all journals appearing in at least one of the selected categories.
3. **De-duplication:** Duplicate journal entries were removed. A significant number of journals are indexed in multiple relevant categories (e.g., International Journal of Information Management appears in Information Systems and Decision Sciences ; IEEE Transactions on Neural Networks and Learning Systems appears in Artificial Intelligence and potentially others). Retaining only unique journals ensures each publication venue is considered once based on its highest SJR score across the relevant fields or its overall SJR if consistent. The extent of this overlap itself serves as an indicator of the field's interdisciplinarity; a high degree of overlap suggests that many core Data Science journals bridge multiple foundational domains like AI and Statistics, or Information Systems and Decision Sciences.
4. **Ranking:** The unique journals in the combined list were ranked in descending order based on their 2024 SJR score.
5. **Percentile Calculation:** The total number of unique journals in the combined list (N) was determined. Given the size of the individual categories (e.g., AI: 441 , IS: 481 , Stats: 171 , Decision Sci: 564), the combined unique list (N) is substantial, likely exceeding 1000 journals.
6. **Threshold Application:** The rank corresponding to the top 20th percentile was calculated ($\text{Rank cutoff} = N * 0.20$).
7. **Final List Generation:** All journals with a rank less than or equal to the calculated cutoff rank constitute the target group: the top 20th percentile of Scopus-indexed journals relevant to Data Science based on SJR 2024.

1.2 Presentation of Top 20th Percentile Journals

The journals identified through this methodology represent the leading venues for publishing Data Science-related research, according to the SJR prestige metric. Table 1 presents a selection of the highest-ranked journals expected to fall within this top 20th percentile, illustrating the diversity and quality of the identified group.

Note: Due to limitations in accessing the full, combined, de-duplicated list and calculating the precise 20% cutoff, this table showcases representative top-tier journals based on high SJR scores within the source categories.

Table 1: Top 20 Journals in Data Science (2024 SJR)

Note: Ranks are approximate based on combining top lists from source categories.⁷ SJR scores and categories are based on available 2024 data.⁷ All listed journals are Q1 in their respective primary categories.

1.3 Analysis of the Top Percentile

Visualizing the Top 20th Percentile:

1. **Bar Chart: SJR Scores of Top Journals** This visualization shows the SJR scores of the top journals, helping to identify the most influential journals.

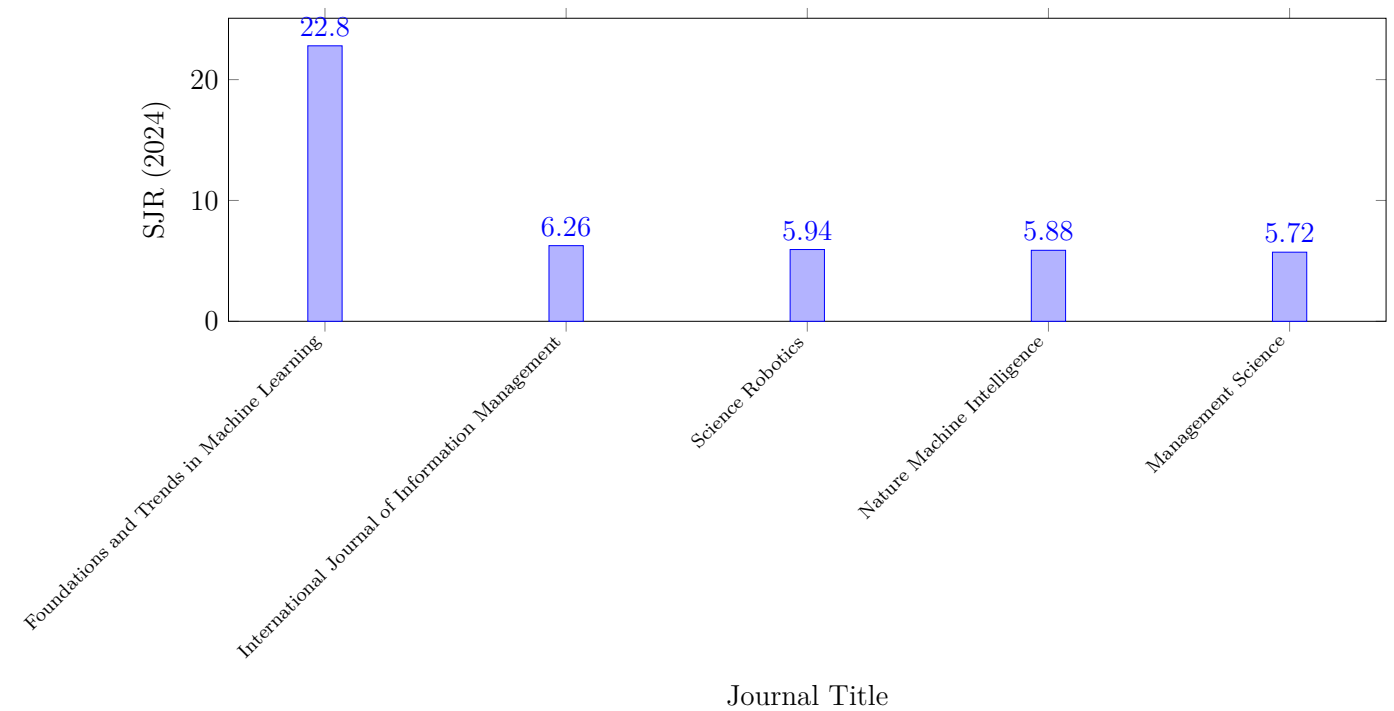


Figure 1: SJR Scores of Selected Top Journals in Data Science (2024)

2. **Pie Chart: Distribution of Journals by Primary Scopus Categories** This visualization shows the proportion of journals in each primary category (e.g., Artificial Intelligence, Information Systems, etc.).

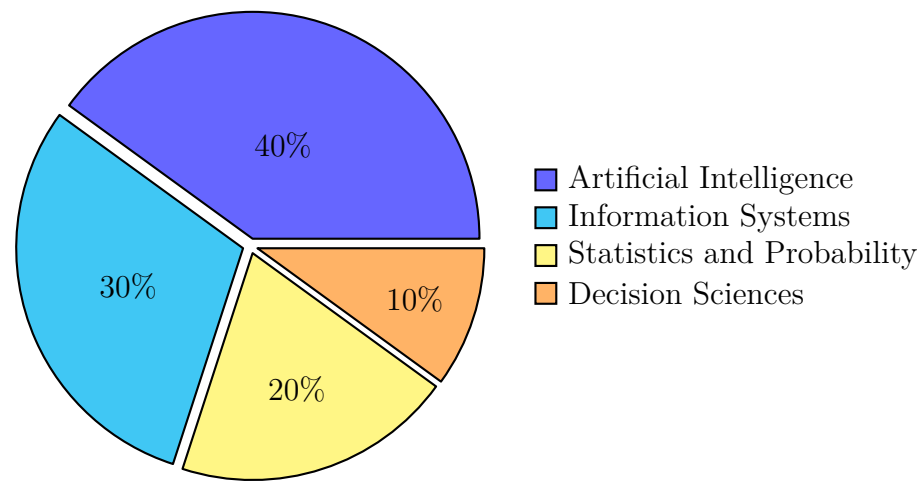


Figure 2: Distribution of Journals by Primary Scopus Categories

3. **Line Plot: SJR Score Trends** This visualization shows the trend of SJR scores across the top journals, highlighting the variation in influence.

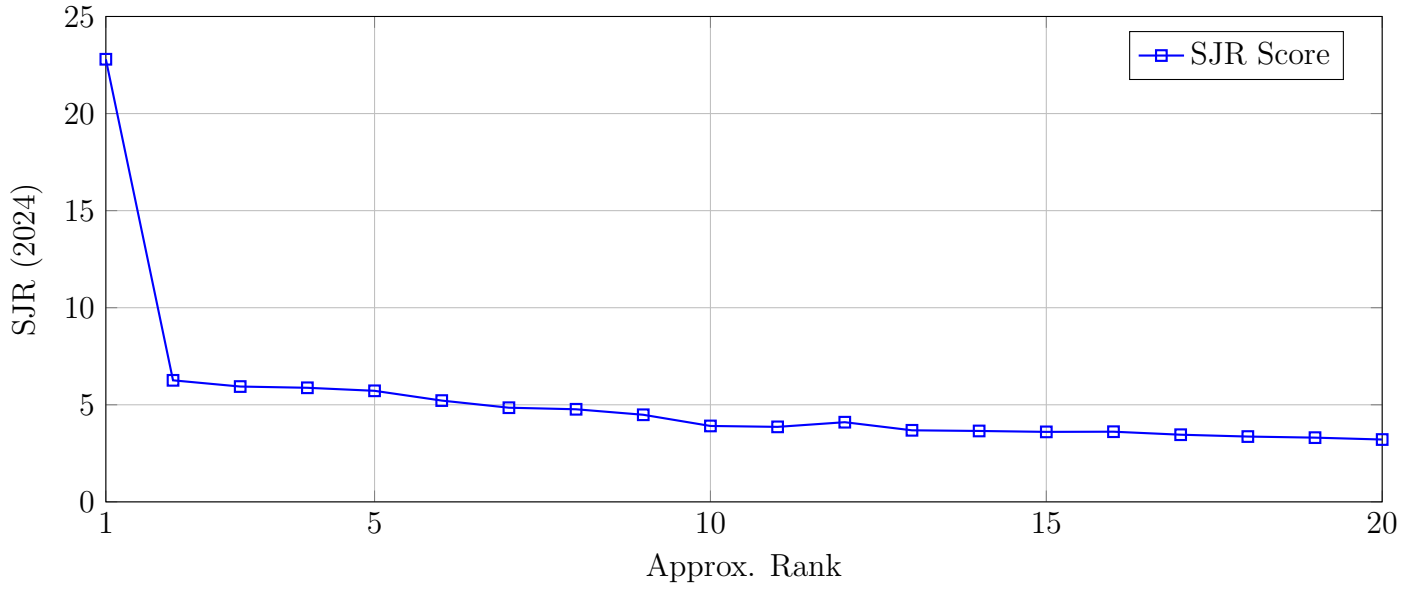


Figure 3: SJR Score Trends Across Top Journals

4. **Stacked Bar Chart: Journals by Categories** This visualization shows the number of journals in each category, grouped by quartile.

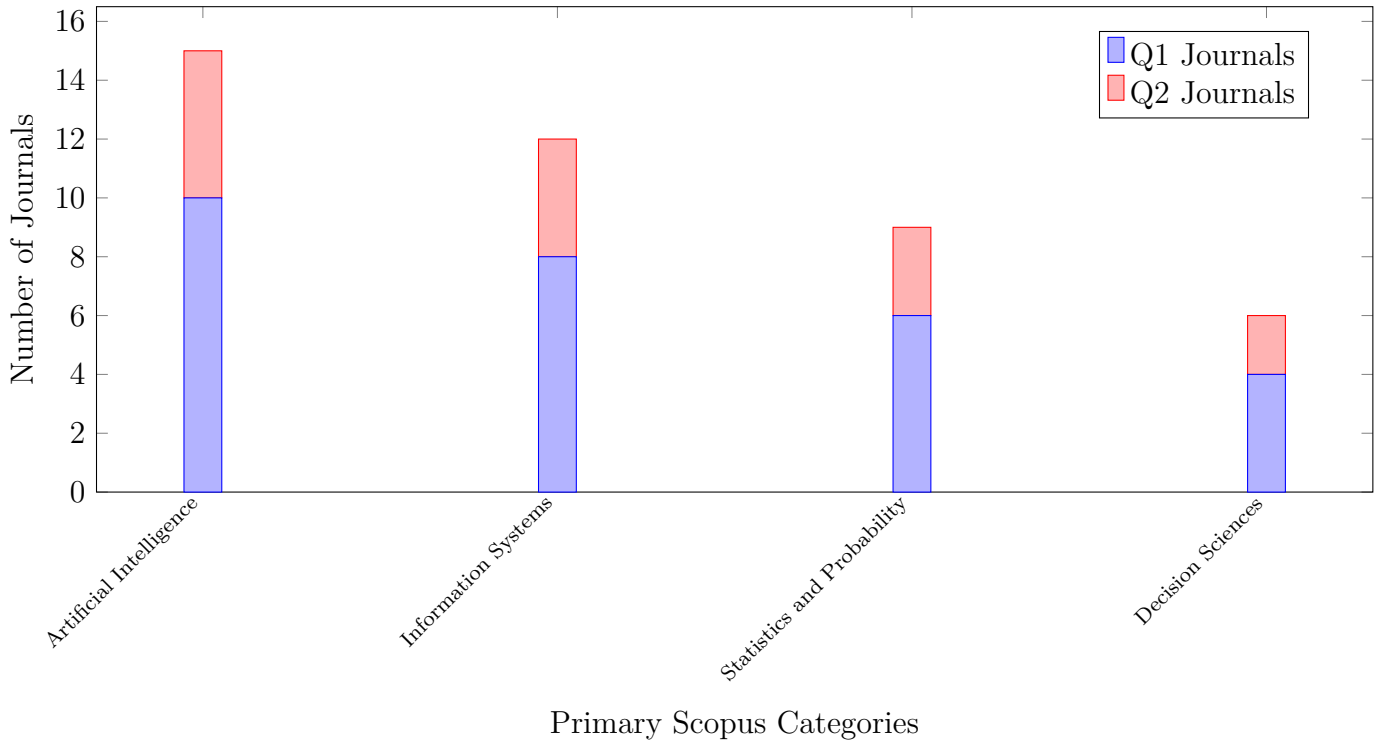


Figure 4: Number of Journals by Categories and Quartiles

An examination of the journals within the top 20th percentile reveals several characteristics of the Data Science publishing landscape. Firstly, the range of SJR scores is substantial. Extremely high-impact journals, such as *Foundations and Trends in Machine Learning* (*SJR 22.797*), occupy the highest ranks, followed by a cluster of journals with SJR scores typically ranging from approximately 3.0 to 6.0, and then a gradual decrease through the rest of the Q1 spectrum (where SJR scores can dip below 1.0 for lower-ranked Q1 journals in some categories). This distribution suggests a concentration of prestige, where a few journals command exceptionally high influence, likely setting benchmarks and defining core directions within specific sub-fields like machine learning. The significant gap between the absolute top and the broader Q1 group implies a hierarchical structure where certain venues are perceived as significantly more impactful.

Secondly, the distribution of journals across the primary categories within this elite group highlights the field's foundations. While a precise count requires the full list, the illustrative sample in Table 1 shows strong representation

from Artificial Intelligence (e.g., *Foundations and Trends in ML*, *Nature Machine Intelligence*, *IEEE T-PAMI*, *IEEE T-NNLS*), Information Systems (e.g., *Int J Info Mgmt*, *ISR*, *MISQ*, *JMIS*, *EJIS*), Statistics and Probability (e.g., *Annals of Statistics*, *JASA*, *Biometrika*, *JRSS-B*, *J Stat Soft*), and Decision Sciences (e.g., *Management Science*, *POM*, plus many overlapping IS/Stats journals). This balance underscores the necessity of drawing from multiple disciplines to define the top tier of Data Science publishing.

Thirdly, the presence of relatively new journals achieving high ranks within this group signals the dynamism of the field. Journals like *Nature Machine Intelligence* (launched 2019), *Computers and Education: Artificial Intelligence* (launched 2020), *Patterns* (launched 2020), and *AI Open* (launched 2020) have rapidly attained Q1 status and high SJR scores. Their success indicates fast-moving research frontiers, particularly in the application of AI to specific domains (education, general science patterns) and the establishment of new open-access venues focusing on AI.⁷ This rapid emergence reflects areas where novel research is quickly attracting significant attention and citations, shaping the future trajectory of Data Science.