1 Defining the Landscape of Data Science Journals

1.1 Background and Rationale

Data Science has emerged as a pivotal, highly interdisciplinary field, integrating techniques and theories from statistics, computer science, information science, and domain-specific knowledge to extract insights and understanding from data. Its rapid growth and profound impact across academia, industry, and society underscore the critical need for researchers, institutions, and funding bodies to navigate its evolving publication landscape effectively. Identifying the most influential and prestigious journals associated with Data Science is crucial for disseminating cutting-edge research, evaluating academic output, making strategic submission decisions, and understanding the intellectual currents shaping the field.

This report aims to address this need by providing a comprehensive analysis of the top-tier Scopus-indexed journals relevant to the broad domain of Data Science. The analysis focuses on objective metrics, specifically utilizing the SCImago Journal Rank (SJR) for 2024 to identify the leading 20th percentile of journals. Furthermore, it categorizes the publishers of these elite journals to illuminate the structure of the publishing ecosystem. Finally, for the absolute top 20 journals within this percentile, the report investigates reported publication timelines, offering practical information regarding the speed of dissemination in these high-impact venues. The intended audience includes academic researchers seeking publication outlets, research managers assessing field dynamics, librarians involved in collection development, and university administrators making strategic decisions related to research evaluation and support.

1.2 Operationalizing "Data Science" within Scopus

A primary challenge in analyzing the Data Science publication landscape is that "Data Science" itself is not designated as a discrete subject category within major indexing databases like Scopus. Its inherent interdisciplinarity means relevant research is published across a spectrum of established fields. Discussions and resources often highlight key journals but defining a comprehensive list requires a systematic approach.

Therefore, to operationalize "Data Science" for this analysis, relevant Scopus subject categories were selected based on their strong alignment with the core components and common application areas of the field. These primary categories are:

1.2.1 Artificial Intelligence (AI)

Encompassing machine learning, pattern recognition, robotics, computational linguistics, and related areas fundamental to modern data analysis. Journal scopes often explicitly mention AI's broad aspects and applications.

1.2.2 Computer Science: Information Systems

Covering the management, processing, retrieval, and strategic use of information, crucial for data-intensive applications and organizational decision-making. The interdisciplinary nature, especially with IT transformation, is noted in journal scopes.

1.2.3 Mathematics: Statistics and Probability

Providing the theoretical and methodological foundations for data analysis, modeling, inference, and uncertainty quantification. Journals focus on developing new statistical methods motivated by real-life

problems.

1.2.4 Decision Sciences

Including areas like Management Information Systems, Information Systems and Management, Operations Research, and Statistics, Probability and Uncertainty, which focus on using data and models to support decision-making processes.

1.2.5 Computer Science: Theory and Methods

Covering the theoretical underpinnings of computer science, including algorithms, computational complexity, and formal methods, which are essential for understanding data processing and analysis.

1.2.6 Computer Science: Software Engineering

Focusing on the design, development, and maintenance of software systems, including programming languages, software architecture, and software engineering methodologies.

1.3 Primary Ranking Metric: SCImago Journal Rank (SJR)

To rank journals within this defined scope, the SCImago Journal Rank (SJR) indicator was selected as the primary metric. SJR is a prestige metric based on the concept that "all citations are not created equal." It weights citations based on the SJR of the citing journal, effectively transferring prestige, and considers a three-year citation window. This approach aims to reflect journal influence more effectively than simple citation counts. SJR is widely used in academic assessment and provides a standardized way to compare journals across different fields, although field-specific citation patterns should always be considered.

This report utilizes the latest available SJR data, corresponding to the 2024 values derived from Scopus data indexed as of March 2025. Associated with SJR are quartile rankings (Q1, Q2, Q3, Q4), which divide the journals within a specific subject category into four equal groups based on their SJR scores. Q1 represents the top 25% of journals in that category. While this report focuses on the top 20th percentile across the combined categories, quartile information provides additional context for individual journal standing within their primary fields.