INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 07, July 2025, pp : 177-179
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

HOW TO FIND A PERFECT DATA SCIENTIST

Mr. K Sathish[1], G. Archana Yadav[2], Mohammed Afroz[3], S.Sai Divya[4]

[1]Professor Department of Computer Science Engineering Siddhartha Institute of Technology & Sciences Hyderabad, India.

[2,3,4]Department of Computer Science Engineering Siddhartha Institute of Technology & Sciences Hyderabad, India.

DOI: https://www.doi.org/10.58257/IJPREMS42738

ABSTRACT

The title of data scientist has been described as one of the sexiest jobs of the 21st century. Numerous efforts have been made to define the job of a data scientist in a qualitative manner by, for example, listing the job functions and required skill sets of data scientists. However, to the best of our knowledge, no attempt has been made to define the term data scientist in a scientific manner.

In this paper, we address this issue by using a data-driven approach to answer three questions:

What is a proper definition of the term data scientist from a market-demand perspective? Do self-described data scientists meet the market demand? And finally, how can companies efficiently recruit data scientists that match their openings? To answer these questions, we crawl two data sets for the supply and demand sides. For the former, we collect a set of data scientist user profiles from LinkedIn for the latter, we collect a set of data scientist job descriptions from Monster.

We first parse the set of data scientist job descriptions via natural language processing techniques and derive a scientific definition of the job of a data scientist via a clustering algorithm.

Second, we use the same approach to determine that, under the aforementioned definition, self-claimed data scientists on the market would meet the market demand with a high probability.

Finally, we introduce a distance-metric learning approach that can be used by companies to find data scientist candidates that match their openings. We achieve an average precision of 12.31%; i.e., one in ten candidates with matching qualifications would accept a given offer.

The application of this quantitative approach could significantly reduce the human resource costs incurred by companies in recruiting matching data scientists.

1. INTRODUCTION

The role of a data scientist has rapidly evolved over the past decade, combining skills from computer science, statistics, and business understanding. Organizations now seek data scientists who not only have technical expertise but also critical thinking abilities, communication skills, and domain knowledge. However, identifying candidates who meet all these dynamic requirements is a significant challenge. As businesses continue to depend on data-driven insights, the demand for multi-skilled professionals grows. This project explores the industry's definition of the "ideal" data scientist by systematically studying research articles, job listings, expert blogs, and reports. It aims to bridge the gap between what the market expects and what aspiring data scientists prepare for, offering a comprehensive framework for the perfect data scientist profile.

OVERVIEW

This mini project focuses on identifying the essential skills, qualifications, and attributes required to become an ideal data scientist. The methodology involves collecting and analyzing data from job portals, research articles, and industry reports to study current trends and employer expectations. Using the Knowledge, Skills, and Abilities (KSA) model, the project develops a framework that defines the characteristics of a perfect data scientist.
The results offer valuable insights for both recruiters seeking the right talent and students or professionals aiming to align their skills with industry demands.

2. PROPOSED SYSTEM

The proposed system aims to provide a holistic evaluation framework for data scientists, combining technical skills, business acumen, and soft skills. It goes beyond traditional hiring methods, which often focus only on technical abilities, by assessing candidates' ability to solve real-world business problems and communicate effectively.

- **Integrated Assessment Platform**: A unified online platform where candidates can take technical tests, complete business problem-solving scenarios, and participate in group assessments or interviews.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 07, July 2025, pp : 177-179
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

- **Real-World Data Projects**: Candidates will be asked to demonstrate their skills by working on real-world data problems, ideally sourced from current business cases or datasets relevant to the company's industry.
- **Soft-Skill Tests and Interviews**: In addition to technical interviews, soft-skill tests will be integrated into the process. These could include situational judgment tests, personality assessments, and peer-review tasks.
- **Continuous Feedback Loop**: Both candidates and hiring managers will receive feedback throughout the evaluation process, ensuring transparency and mutual understanding of the candidate's strengths and areas for improvement.

3. SYSTEM ARCHITECHURE

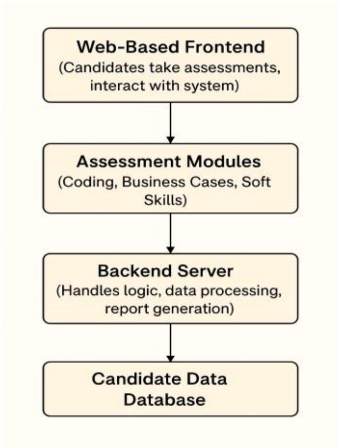

Figure 1: System Architecture Diagram

4. IMPLEMENTATION

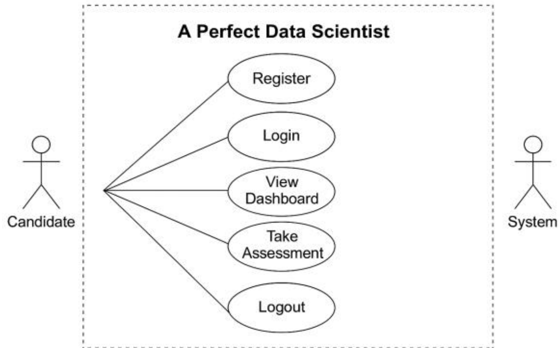

Figure 2: Use Case Diagram

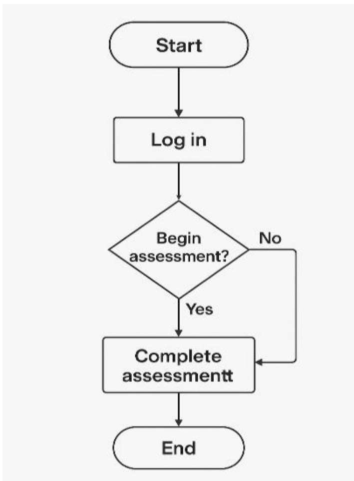

Figure 3: Activity Diagram

5. FUTURE WORK

Future iterations of the project could incorporate more advanced assessment tools, such as:

- **Automated skill assessment**: Integrating coding challenges or simulations for real-time evaluation.
- **Job market insights**: Providing real-time updates on the skills currently in demand through web scraping or API integration with job listing platforms.
- **Integration with AI**: Leveraging AI to predict the potential success of candidates based on past hiring data and industry trends.

6. CONCLUSION

The project "How to Find a Perfect Data Scientist" aims to provide a comprehensive understanding of the skills, qualifications, and attributes that define the ideal data scientist in today's competitive job market. Through careful analysis of industry trends, research papers, blogs, and expert opinions, this system facilitates the systematic evaluation of candidates for data science roles.

7. REFERENCES

[1] T. H. Davenport and D. J. Patil, "Data scientist: The sexiest job of the 21st century," Harvard Bus. Rev., vol. 90, no. 10, pp. 70–76, Oct. 2012.
[2] Hiretual. Accessed: Oct. 20, 2017. [Online]. Available: https://hiretual.com
[3] Why So Many Fake Data Scientists? Accessed: Oct. 20, 2017.[Online].Available:https://www.linkedin.com/pulse/why-so-many-fake-data-scientist-bernard-marr
[4] Being a Data Scientis. Accessed: Oct. 20, 2017. [Online]. Available: https://www.linkedin.com/pulse/being-data-scientist-carla-gentry
[5] Dhar, "Data science and prediction," Commun. ACM, vol. 56, no. 12, pp. 64–73, 2013.