Jata MacCabe

Senior Consultant - Data Scientist

5 years of experience in Software Development and Data Science

Profile

Jata MacCabe is a seasoned Data Scientist and Senior Consultant with over five years of experience blending software development and data science expertise. With a Master's degree in Computational Linguistics, Jata combines creativity and analytical precision to tackle complex challenges and deliver innovative, data-driven solutions. Known for proficiency in Natural Language Processing (NLP), machine learning, and full-stack development, Jata has a proven track record of leading impactful projects across diverse industries, including IT consulting, academia, and non-profit organizations.

Jata has demonstrated her Data Science leadership by spearheading the development of cutting-edge language models with reduced hallucinations and enhanced human-like communication. She has partnered with marketing, finance, and operations teams to integrate predictive analytics and visualization into actionable business strategies. Her research work has focused on computational linguistics, focusing on neural sub-word modeling for under-resourced languages.

Her data expertise is complimented by her Full-Stack Development experience where she designed and implemented end-to-end web-based solutions, enhancing user experiences and accessibility. Her technical expertise includes Python, SQL/NoSQL, Vue.js, React, and cloud platforms such as Azure and Google Cloud as well as Data Science Tools and Frameworks like LangChain, RAG, Jupyter, and Pandas for robust data science and visualization applications.

CGI Experience

Senior Consultant - Data Scientist (11/24 to Present)

* Gathering and cleaning data and conducting exploratory data analysis.
* Developing models to forecast trends and behaviors using machine-learning techniques.
* Creating visual representations of data insights.
* Collaborating with marketing, finance, and operations teams to integrate data insights into business strategies.
* **Technology:** LangChain, RAG, Jupyter, and Pandas

Health Insurance Client, Full Stack Software Developer (02/22 to 10/22)

* Key full-stack developer for all levels of the project, including database development, API creation, and front-end development
* Mentored junior and co-op team members, building strong relationships and guiding them to develop their skills and reach their professional goals
* Developed new features and functionalities to improve the user experience and increase the portal’s accessibility
* **Technology: SQL/NoSQL, Vue.js, React, and cloud platforms such as Azure and Google Cloud**

Other Experience

Rocketbrew, Data Scientist (08/23 to Present)

* Led development of core language model
* Improved content accuracy to near-zero hallucinations, using programmatic and RAG techniques
* Created new techniques to produce more human-like communication
* Highly proficient in prompt engineering and NLP techniques for customized model interaction

Faculty of French, Italian and Hispanic Studies, UBC, Graduate Academic Assistant (05/23 to 01/24)

* Conducted statistical analysis to identify patterns in large volumes of textual data
* Developed web scraping methods to produce unsupervised data-collection

Legal Hackers Ontario, Future of Law Lab (UofT), Data Scientist – Captone Internship (05/23 to 06/23)

* Primary developer of text-mining methods for >40 thousand unstructured web documents with domain-specific jargon
* Led the development and management within the data science team

Community Forests International, Junior Software Developer – Data Analytics Intern (04/18 to 04/19)

* Contributed to the design and development of a web-based dashboard and a data collection and analysis system
* Upon completion of my term as a Data Analytics Intern, rehired by the company as a Junior Software Developer

Publications:

* Glossy Bytes: Neural Glossing using Sub-word Encoding, 07/23  
  Association for Computational Linguistics
* Published in Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology
* Developed neural sub-word modelling based approaches to interlinear glossing for seven under-resourced languages

Volunteer Experience:

* Mentor & Invited Panel Alumna, 09/23 - Present  
  UBC MDS-CL Graduate Program, Vancouver, BC
* Published in Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology
* Developed neural sub-word modelling based approaches to interlinear glossing for seven under-resourced languages
* Workshop Facilitator & Mentor, 09/23   
  Vancouver Datajam, Vancouver, BC
* Conducted an interactive workshop on prompt engineering, LLMS, and other advanced NLP techniques
* Mentored winning team on NLP best practices and code-specific advice
* Charity Representative, Diversity & Inclusion Representative, 09/21 – 05/23  
  UNB Computer Science Association, Fredericton, NB
* Elected member of UNB’s Computer Science Association, planned and ran events for Computer Science students
* Advocated for under-represented groups in the faculty and organized charity events for the community

Education

* Master of Data Science: Computational Linguistics, University of British Columbia, 2023
* Honours Bachelor of Computer Science: Information Systems, Business Minor, University of New Brunswick, 2021

|  |  |
| --- | --- |
| Industry experience   * Information Technology * Academia and Research * Non-Profit and Environmental   Technical specializations   * Natural Language Processing (NLP) * Machine Learning (ML) * Data Analysis and Statistical Analysis * Web Scraping and Data Collection * Full-Stack Software Development   Areas of expertise   * Developing and Leading AI-Based Projects * Building and Optimizing Language Models * Prompt Engineering * Text Mining and Analysis * Data Visualization and Dashboard Development   Languages   * English * French | Environments   * Cloud Platforms: Azure, Google Cloud * Web Development: Full-stack development environments * Data Science and Analytics: Machine learning and NLP environments * Academic Research: University research settings   Tools & software   * Programming Languages: Python, C, Java, SQL/NoSQL * Machine Learning and NLP: LLMs, LangChain, RAG, ML/NLP techniques * Web Development: Vue/Vue3, HTML & CSS, Spring, Selenium, React * Cloud Platforms: Azure, Google Cloud * Data Science Tools: Compound Systems, Jupyter, Pandas, NumPy |

Skills summary

| Skill | Number of years | Skill level\* |
| --- | --- | --- |
| Technical skills | | |
| Python | 5 | 4 |
| SQL/NoSQL | 5 | 4 |
| C | 5 | 3 |
| Java | 5 | 3 |
| Vue/Vue3 | 2 | 3 |
| HTML & CSS | 5 | 4 |
| Spring | 2 | 3 |
| Selenium | 2 | 3 |
| React | 2 | 3 |
| LLMs | 2 | 4 |
| LangChain | 2 | 4 |
| RAG | 2 | 4 |
| ML/NLP | 5 | 4 |
| Compound Systems | 2 | 3 |
| Application knowledge | | |
| Azure | 2 | 3 |
| Google Cloud | 2 | 3 |
| Jupyter | 5 | 4 |
| Pandas | 5 | 4 |
| NumPy | 5 | 4 |
| IT disciplines | | |
| Data Analysis and Statistical Analysis | 5 | 4 |
| Machine Learning and NLP | 5 | 4 |
| Web Scraping and Data Collection | 5 | 4 |
| Full-Stack Software Development | 5 | 4 |
| Data Visualization | 2 | 3 |
| Industry knowledge | | |
| IT Consulting | 5 | 4 |
| Academia and Research | 1 | 3 |
| Non-Profit and Environmental | 1 | 3 |
| Other relevant skills | | |
| Prompt Engineering | 2 | 4 |
| Text Mining and Analysis | 5 | 4 |
| Project Management | 2 | 3 |
| Mentorship | 4 | 3 |
| Communication | 5 | 4 |

\*Skill Level: 1 = Beginner, 2 = Experienced, 3 = Advanced, 4 = Expert