



# Jay Kumar

✉ [Jay\\_tharwani1992@yahoo.com](mailto:Jay_tharwani1992@yahoo.com)  
🌐 <https://jaykumarr.github.io/cv/>  
📞 +1 902-488-1704  
LinkedIn | GScholar | Github

Extensive experience in data engineering for geospatial time-series data, text data, and weather data. Design, develop, and maintain scalable and efficient data pipelines using Python and Java. Optimize data workflows and processes. Adept at building proofs-of-concept, integrating mixed data sources, and driving machine learning model development to create impactful products.

## SKILLS

---

- **Programming Languages:** Python | SQL | Java
- **Machine Learning Libraries:** TensorFlow, PyTorch, Scikit-learn, Pandas, Numpy,
- **Data Science & Analytics:** Time-series forecasting, NLP (Unstructured data), Statistical Modeling, Geospatial data Analytics (structured data).
- Cloud/Big Data: Databricks, Docker
- Visualization: Tableau, Plotly, Geopandas, Matplotlib
- Collaboration/Project Management: JIRA, Cross-functional team collaboration, Communication of technical concepts to non-technical stakeholders

## PROFESSIONAL EXPERIENCE

---

**Artificial Intelligence Developer** (CIOOS Atlantic | Halifax, Canada) Aug/2024–Jul/2025

- ETL pipeline maintenance for ocean data analysis enhancing advisor tools for data-driven decision-making.
- Integrating recommendation systems for data lake, improving operational efficiency (Python, numpy, scikit-learn, SQL)
- Debugging and testing python REST APIs, code and services.
- Troubleshooting data and model design workflow
- Performed statistical analysis of user access logs, enabling actionable insights for catalog improvements (visualization: matplotlib, plotly).
- Maintained, created test cases (A/B Testing), and improved current code function libraries using Git, CI/CD , and data validation techniques.

**ML Research Scientist** (Dalhousie University | Halifax, Canada) Jan/2022 – May/2024.

- Integrated multi-source data (AIS, weather, ocean biophysical data) to perform advanced analytics, providing actionable insights to stakeholders (Docker, postgres, SQL, selenium) .
- Enhancing data quality, implementing outlier detection, duplicate checks, and validation methods (Python, Geopandas, Github CI/CD).
- Debugging and testing python REST APIs, code and services.
- Real-time monitoring and anomaly detection, significantly improving data usability and decision-making efficiency.
- Building runbooks and documentation for API usage (Streamlit).

▪ **Research Scientist** (University of Electronic Science and Technology of China | Chengdu, China) Sept/2018 – Dec/2021

- Machine learning model development for classification and grouping unstructured short-text stream clustering using non-parametric Dirichlet processes to detect evolving text data patterns.

- Feature engineering of unstructured text data to extract semantics and contextual dependency to improve **text similarity** indexing and developed interpretable ML model experiments for dynamic data environments.
- Generated **0.1M\$** project funding for the supervisor.
- Applied deep learning models to enhance text embeddings for evolving vocabularies, supporting NLP-based solutions.

<b>Software Developer</b> (Sapphire Consulting Services   Jamshoro )	Jan/2015 –May/2015
<ul style="list-style-type: none"> <li>- <b>Java</b> web and desktop application development with MVC frameworks including <b>Spring</b>, Hibernate, japer report, crystal report and Struts.</li> <li>- <b>Database:</b> Optimizing <b>SQL queries</b>, database communication, designing query <b>framework</b></li> </ul>	
<b>Software Developer Intern</b> (University o f Sindh  Jamshoro )	Jun/2014 –Dec/2014
<ul style="list-style-type: none"> <li>- <b>Java</b> desktop application development for student admission system.</li> <li>- Designing algorithms for admission selection based on complex catalog criteria.</li> <li>- Front-end and back-end programming for student admisison form. (includes <b>data migration</b>)</li> <li>- <b>Database:</b> Optimizing <b>SQL queries</b>, JDBC database communication, fee management</li> </ul>	

## CERTIFICATIONS

---

- **AWS Cloud Practitioner** Essential Training- 2026
- **Docker** - 2025
- **Learning Hadoop** - 2025
- **Google Cloud Foundation** - 2025
- **MLOps with databricks** - 2025
- **SQL Essential Training** - 2025
- **Designing REST APIs** - 2025
- Data Visualization with **Tableau** (Tableau) – 2025

## EDUCATION

---

▪ <b>Ph.D.</b> in Computer Science and Technology (UESTC, China)	Sep 2018 – Dec 2021
▪ <b>Masters</b> in Computer Science (Quaid-I-Azam University, Pakistan)	Jul 2015 – May 2018
▪ <b>Bachelor</b> in Computer Science (University of Sindh, Pakistan)	Jan 2011 – Dec 2014

## INVITED TALKS

---

- CIOOS Building Bridges [Workshop](#) - 2024
- MiTE International Conference on Evolving Technologies – 2024
- Merdian AIS Analysis and Visualization [Worshop](#)  2023