Jay Kumar

Phone: +1 902-989-1704
Email: Jay_tharwani1992@yahoo.com
Jay.kumar@dal.ca
Links: GScholar | LinkedIn | Github

I am a data scientist with extensive experience in designing machine learning, deep learning and probabilistic models. Worked on real-time sequential **text** and **Geo-spatial data**. With solid programming and problem-solving skills, my significant contributions include the development of open-source software for variety of industries such as cargo shipping transport, finance, health-care facilities and academic institutes. While contributing on 25+ research papers publications in top peer-reviewed journals and conferences in computer science, I have completed 3 research projects.

CERTIFICATIONS

- Data Visualization with **Tableau** (Tableau)- 2025
- **Cybersecurity** Training Course in Privacy/ Info Security Awareness (Dalhousie University) **-** 2024
- **Java** Techonology (Netlync Research Lab) **–** 2014
- **Office Automation** (AIMS Business & IT Solutions)**–** 2011

WORK EXPERIENCE

- **AI Developer** (Dalhousie University, CIOOS Atlantic) | Halifax, Canada August 2024 –Present
  - Communicating with data integrators, front-end developers, and data management team.
  - AI model integration in data ingestion pipeline
  - Investigation of user access logs
  - Integration of recommendation system for data catalogue repositories.
- **Postdoctoral Researcher** (Dalhousie University) | Halifax, Canada Jan 2022 –May 2024
  - Designing data pipelines of 6 years of Cargo and Tanker Ship Geospatial Automatic Information System (AIS) data and conducting data quality assessment with statistical analysis.
  - Optimized LLM for accurate long term vessel route forecasting to integrate with decision-making system to upgrade the overall system efficiency and performance.
  - Developed and maintained open-source Python and RUST libraries for scalable AIS data preprocessing and manipulation. Led the development, troubleshooting, and designing test-case suits of Python library.
  - Collaborated with cross-functional teams on developing flowcharts, component integration, data reconciliation, and code documentation.
  - Data reconciliation w.r.t Integration of weather data, ocean data and AIS data.

SKILLS

- **Communication:** Interpersonal Skills, teamwork, active listening
- **Analytical:** Problem solving
- **Soft / Technical:**
  - **Programming Language:** Python | Java (J2SE, J2EE) | R | PL/SQL
  - **Data Science:** Statistical model | Machine Learning | NLP (Natural language Processing) | Data Engineering
  - **Frameworks:** Pytorch | Tensorflow | Numpy | Pandas | Jasper Reports | Hibernate | Spring | Git
  - **Database:** Relational database design | MySQL | Oracle | Postgres | SQLite | Microsoft Access | Data Profiling | Entity Relation Diagram (ERD)
  - **Tools:** Excel | QGIS (similar like ArcGIS) | RapidMiner | Linux | Docker | Tableau
  - **Hardware:** Arduino micro-controller programming | Raspberry pi micro-controller | RFID

EDUCATION

- **Ph.D.** (Computer Science and Technology) Sep 2018 – Dec 2021
  University of Electronic Science and Technology of China
  Published 5 publications in top venues including ACL, IEEE Transactions on Cybernetics, and Expert Systems with Applications.
  - Proposed Non-parametric Dirichlet models for stream of short text for clustering to capture semantic similar words with evolving distribution
  - Proposed semi-supervised classification of multi-label short text stream suitable for real world setting where number of classes, feature-class distribution, and relationship among class labels may change over time.
  - Developed supervised and unsupervised models based on ML models including KNN, K-Mean, Random Forest, HDBScan, Linear Regression, Parametric stochastic models, Markov Chain model, non-parametric stochastic model, and Dirichlet mixture models.
  - Received runner up award from IEEE Sensors on a research article.

- **M.Phil.** (Computer Science) Jul 2015 – May 2018
  Quaid-i-Azam University
  - GPA 3.5/4.0
  - Performed feature engineering on textual and user behavioral features.

  Applied classification algorithm including Decision trees, Support vector machines, and Random forest algorithms for detecting fake reviews or spam opinions. Thesis

- **B.S.** (Computer Science) Jan 2011 – Dec 2014
  University of Sindh
  - GPA 3.30/4.0
  - School management System Software product development
  - Problem-solving algorithmic approach for admission system in University of Sindh.
  - Algorithm designing for admission selection of student, fee payment, student record management, data migration and reports.

  Major courses including networking, data structures, object-oriented programming, statistics, financial accounting, advanced algorithms, database systems, software engineering, computer graphics, web programming, and scientific modeling and simulation.

PERSONAL PROJECTS

- **RFID-based Attendance System:** ETL UI developed integrated with RFID technology to automate attendance system.
- **Fertilizer E-commerce System:** Java desktop App with DB for managing the sales and distribution of products.
- **Pathology Laboratory Repo System:** Reports management system for pathology laboratories, facilitating the management of test results, patient records, and compliance with healthcare standards.
- **Admission Management System:** Collaborated to develop student admission and record management system.

AWARDS

- Best Paper Award IEEE Sensor- **2021**
- Academic Achievement Award 2nd Prize – **2020**
- Excellent Performance Award 3rd Prize – **2019**

INVITED TALKS

- CIOOS Building Bridges Workshop - **2024**
- MiTE International Conference on Evolving Technologies - **2024**

OTHERS

- Class 5 Driving License
- **Volunteering:** volunteer work at Maritime Sikh Society Halifax

JOURNAL PUBLICATIONS

Title	Journal	Year
[1] Privacy-preserving blockchain-based federated learning for brain tumor segmentation	Computers in Biology and Medicine	2024
[2] Multi-path long-term vessel trajectories forecasting with probabilistic feature fusion for problem shifting	Ocean Engineering	2024
[3] Maritime tracking data analysis and integration with AISdb	SoftwareX	2024
[4] Enhancing Sindhi Word Segmentation Using Subword Representation Learning and Position-Aware Self-Attention	IEEE Access	2024
[5] Online Semi-supervised Classification on Multi-label Evolving High-Dimensional Text Streams	IEEE Transactions on Systems, Man, and Cybernetics: Systems	2023
[6] A Context-enhanced Dirichlet Model for Online Clustering in Short Text Streams	Expert Systems with Applications	2023
[7] Dynamic context management in context-aware recommender systems	Computers and Electrical Engineering	2023
[8] FedStream: Prototype-Based Federated Learning on Distributed Concept-drifting Data Streams	IEEE Transactions on Systems, Man, and Cybernetics: Systems	2023
[9] An Online Semantic-enhanced Graphical Model for Evolving Short Text Stream Clustering	IEEE Transactions on Cybernetics	2022
[10] Learning High Dimensional Evolving Data Streams with Limited Labels	IEEE Transactions on Cybernetics	2022
[11] Blockchain and Homomorphic Encryption based Privacy-Preserving Model Aggregation for Medical Images	Computerized Medical Imaging and Graphics	2022
[12] Inferring context with reliable collaborators: A novel similarity estimation method for recommender systems	Applied Intelligence	2022
[13] Multi scale and direction target detecting in remote sensing images via modified YOLO-v4	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	2022
[14] Trends in Vehicle Re-Identification Past, Present, and Future: A Comprehensive Review	Mathematics MDPI	2021
[15] Context-Aware Bidirectional Neural Model for Sindhi Named Entity Recognition	Applied Science	2021
[16] Blockchain-federated-learning and deep learning models for covid-19 detection using CT imaging	IEEE Sensors Journal	2021
[17] An Integration of Blockchain and AI for Secure Data Sharing and Detection of CT images for the Hospitals	Computerized Medical Imaging and Graphics	2021
[18] Data stream classification with novel class detection: a review, comparison and challenges	Knowledge and Information Systems	2021
[19] Online Reliable Semi-Supervised Learning on Evolving Data Streams	Information Sciences	2020
[20] A multimodal malware detection technique for Android IoT devices using various features	IEEE Access	2019

CONFERENCE PUBLICATIONS

Title	Conference	Class	Year
[1] An Online Semantic-enhanced Dirichlet Model for Short Text Stream Clustering	Association for Computational Linguistics (ACL)	A	2020
[2] Neural Joint Model for Part-of-Speech Tagging and Entity	International Conference on Machine	EI	2021

Extraction	Learning and Computing		
[3] A Non-Parametric Multi-Lingual Clustering Model for Temporal Short Text	International Computer Conference on Wavelet Active Media Technology and Information Processing	EI	2020
[4] H3DNN: 3D Deep Learning Based Detection of COVID-19 Virus using Lungs Computed Tomography	International Computer Conference on Wavelet Active Media Technology and Information Processing	EI	2020
[5] Malicious code detection based on image processing using deep learning	International Conference on Computing and Artificial Intelligence	EI	2018
[6] SiPOS: A Benchmark Dataset for Sindhi Part-of-Speech Tagging	Proceedings of the Student Research Workshop associated with RANLP	EI	2021
[7] Effective and explainable detection of Android malware based on machine learning algorithms	International Conference on Computing and Artificial Intelligence	EI	2018