Kartik Joshi

India | kartikjoshi0172@gmail.com | +91 6395912621 | linkedin.com/in/kartikj360 | github.com/kartikj360

Education

National Research University Higher School of Economics, Moscow, Russia,   Aug 2023 – Jul 2025
M.S. in Applied Mathematics and Artificial Intelligence, **GPA: 8.1** Link

- **Coursework:** Modern Methods of Decision Making, Ordered Sets in Data Analysis, Linear Algebra for Data Science, Machine Learning and Data Mining.
- **Thesis:** Robust and Accurate Analysis of Data Modalities in Mass Spectrometry . Link

University of Delhi, New Delhi, India, B.S. in Computer Science, **GPA: 8.9** Link   Dec 2020 – Jul 2023

- **Coursework:** Software Engineering, Database Management Systems, Operating Systems, Computer Networks, Design and Analysis of Algorithms , Link

Skills

Data Engineering: Apache Spark, PySpark, Airflow, Snowflake, BigQuery, Dimensional Modeling, Data Pipelines

Programming: Python, SQL, R, C++ , Bash

ML & AI: Machine Learning, Generative AI, PyTorch, TensorFlow, Scikit-learn, XGBoost, Hugging Face

Databases & APIs: PostgreSQL, REST APIs, FastAPI, LangChain, LangGraph

Tools & Platforms: Git, GitHub, Linux, CI&CD, Cloud GCP , Kafka, Databricks, Docker, Kubernetes

Work Experience

Research Assistant (Machine Learning & Data Engineering), Laboratory on AI for   Jan 2024 – Jul 2025
Computational Biology at HSE, Moscow, Russia Link

- Architected and maintained automated ETL pipelines processing 100+ TB of mass spectrometry data using SQL and Python
- Built scalable data pipelines on Cloud Storage to ingest, transform, and load high volume of spectral data daily, enabling downstream ML model training
- Developed scalable ML infrastructure that expanded peptide database diversity by 40%, focusing on data quality and feature engineering for large-scale biological analysis

Centralized Allocation Facilitator, The E.C.A. Commission of the University of   Oct 2022 – Dec 2022
Delhi, Delhi, India Link

- Analyzed student enrollment data using SQL and Python to optimize resource allocation

Projects and Publications

Single amino acid variation identification in high resolution tandem mass spectrometry data in bottom up proteomics @International Journal of Mass Spectrometry(Elsevier) Link

- New method, called SeVa (Sequence Variation), extracts the highest scoring amino acid sequence variation from HR-XPV, called: SeVa-peptides.
- Running the results for custom based genomic sequence , over modulated sequences and PSTM comparison.

Meta-Proteomics ML Pipeline Link

- Engineered batch processing pipeline for 600TB+ raw mass spectrometry data, implementing incremental loading and data deduplication strategies
- Designed schema and data models optimized for analytical queries, reducing storage costs by 30% through intelligent partitioning
- Built data quality validation framework ensuring 99%+ accuracy across 100GB of processed spectral data

Heart Patient Dataset for Classification Link

- Built a classification model achieving 90% accuracy in predicting patient mortality