SHAIK SALMAN BASHA

Salmanbasha.shaik@unb.ca | (506) - 897 - 3381 | **LinkedIn** | **Leetcode**

PROFESSIONAL SUMMARY

Final-year Computer Science Honours student skilled in designing and optimizing data pipelines. Brings hands-on experience developing ETL processes and building applications integrated with enterprise systems like SAP. Proficient in Java, Python, SQL, Docker, and BI tools including Power BI and Tableau. Currently seeking opportunities to contribute to impactful data or software engineering projects in a collaborative team environment.

EDUCATION

University Of New Brunswick, Fredericton **April 2026**

Bachelor of Computer Science, GPA: 3.7 | Dean's list (2022 – 2024)
Relevant Courses: Advanced Algorithms in Java, Software Engineering, Big Data and NLP in python

PROFESSIONAL EXPERIENCE

Agriculture and Agri Food Canada -> Jan 2025 - present

Data Engineer Intern (Bioinformatics)

- Accelerated data ingestion by 23x, improving processing speed from 11 files/sec to 260 files/sec by optimizing parallel processing and I/O operations within the data extraction workflow.
- Designed and implemented a robust ETL pipeline to migrate over 1000 indexed metadata records from Elasticsearch to a structured PostgreSQL database, ensuring data integrity and enabling advanced relational queries.
- Analyzed and benchmarked multiple NLP tools (e.g., Spacy, SciSpacy) to enhance the 'Transform' stage of the ETL process, improving the accuracy of named-entity recognition
- Worked with many raw sequencing files to capture important information and ensure data accuracy, integrity, privacy, security, and compliance through quality control procedures

New Brunswick Power Corporation -> Sep 2023 – Apr 2024

Software Engineer Intern

- Developed and maintained enterprise .NET/C# applications that integrated with core SAP systems, gaining a deep understanding of complex system architecture and how data is generated and consumed in a large-scale business environment.
- Created data-driven solutions using the Microsoft Power Platform (Power Apps, Power BI) and Tableau, building internal tools and interactive dashboards to provide insights from business data
- Developed and maintained various application features, which provided broad technical exposure and strengthened my ability to quickly learn and adapt to complex, interconnected systems.

PROJECTS

Tennis Match Prediction

- Developed a classification model to predict tennis match outcomes given two players as input.
- Model evaluation between Binary naïve bayes, Logistic regression and Neural networks
- Pre-processing steps to make the data clean and usable and report on Accuracy, Precision, F1

Science data Inventory at Agriculture and AgriFood Canada

- Developed an automated pipeline to extract metadata from sequencing files using Tika and shell-based tools.
- Orchestrated scalable ingestion of genomic data into Elasticsearch via Logstash, containerized with Docker for portability and performance.
- Enabled fast, searchable access to high-throughput genomics metadata through a robust ELK-based platform.

SKILLS

Certifications: Hackerrank SQL (Advanced), kaggle Notebooks Expert (Top 6%), Lean Six Sigma
Languages and Libraries: SQL, Python (pandas, matplotlib, seaborn), Java, scikit-learn, Spacy, TensorFlow,
Tools & Platforms: Elasticsearch, Docker, Jupyter, Power BI, Tableau, CKAN (Govt database), API's, Hadoop
Concepts: Data Engineering (ETL), Data Science, Software design & development, Full – stack development