Sajeev Singh

GitHub | LinkedIn | Portfolio | + 1 732-558-9741 | sajeevysingh@gmail.com | New Brunswick, NJ, 08901

EXPERIENCE

Robert Wood Johnson University Hospital | Research Specialist - Data Science **Apr 2025 - Present**

- Extracted and transformed large-scale ECG data from MuseNX (raw XML) into structured, model-ready datasets using Python and SQL. Built an automated ETL pipeline integrating clinical metadata. Developed predictive models (CNN, LSTM, SVM) to detect early cardiac arrest risks, achieving 85% sensitivity, with insights visualized through dashboards.
- Developed a deep learning pipeline to automate frame-wise classification of echocardiogram (.dcm) images as contrasted or non-contrasted, including custom script to extract frames and a Convolutional Neural Network (CNN) model for contrast enhancement detection, streamlining image triage for downstream cardiac analysis

Med-Kick | Data Science Intern **Oct 2024 - Apr 2025**

- Designed and deployed machine learning solutions using AWS SageMaker and Lambda to process large-scale audio data, enabling real-time transcription, AI-driven summaries, interaction analysis scorecards. Reduced manual processing delays and ensured HIPAA compliance.
- Developed a Python-based pipeline to create AI-voice avatar reminders triggered by CRM notifications, increasing patient engagement and reducing no-show rates by 30%.
- Automated ETL pipelines for real-time data integration with GoTo (VOIP) and HubSpot (CRM) via webhooks enabling data-driven decisions using Power BI dashboards and improving enrollment metrics by 15%.

DataFlow Group | Data Science Engineer **Dec 2022 - Jul 2023**

- Automated quality checks for client support websites using Selenium and Robot Framework, and developed Power BI dashboards to track operational metrics, enhancing monitoring efficiency.
- Designed and deployed a texture based tampering detection model using XGBoost and Random Forest to verify critical documents (e.g., transcripts, certifications). Improved fraud detection accuracy and reduced manual reviews by 30% through precision tuning and workflow integration with GitLab CI/CD.
- Developed a computer vision solution for passport detection and data retrieval, leveraging MRZ and ElasticSearch for scalable data indexing, improving applicant data processing efficiency by 50%.

Larsen & Toubro | Graduate Engineer **Jul 2022 - Nov 2022**

- Built a time series forecasting model (Prophet, XGBoost) with Bayesian tuning to predict rail project material demand, reducing stockouts by 35% and improving procurement planning.

SKILLS

Core Expertise	**:** Deep Learning, Predictive Modelling, Statistical/Analysis, Natural Language Processing, Computer Vision, Gen AI Applications, Causal Inference, Survival Analysis, A/B Testing
Programming & MLOps	**:** Python, R, SQL, Docker, Kubernetes, Jenkins, Git/Github, Data Pipelines
Data Analysis & Visualization	**:** Power BI, Tableau, Advanced Excel, Qlik Sense, IBM Cognos
Databases	**:** PostgreSQL, MySQL, MongoDB, Firebase, Neo4j, SAS, Cassandra, Snowflake, Hadoop(Hive, HDFS), Spark, Map Reduce
Cloud & Distributed Computing	**:** AWS (SageMaker, RedShift, Neptune, Lambda, Route 53, EC2), Google BigQuery, Microsoft Azure, Databricks
Version Control & OS	**:** Git, Linux, macOS, Windows

PROJECTS

Agentic RAG System for AI Content Creation | *Python, LangChain, OpenAI, Airflow, Github, AWS EC2*

- Architected a modular Retrieval-Augmented Generation (RAG) framework using LangChain agents and OpenAI APIs, enabling dynamic document-grounded content generation at scale.
- Deployed the system on AWS EC2 with orchestration via Airflow, reducing manual content engineering time by **30%** and enhancing production pipeline efficiency.

Knowledge Graph Based Movie Recommendation System | *SPARQL, RDF/OWL, RDFlib, Flask, Python, Protégé, DBpedia*

- Designed and implemented a semantic recommendation engine using RDF ontologies and SPARQL queries over linked data from DBpedia and TMDb.
- Built an interactive web app with Flask, enabling users to query for complex relationships (e.g., "Sci-fi movies by Christopher Nolan") with context-aware results.

Optimized Resource Allocation via Time Series Forecasting | *Python, R, ARIMAX, Neural Networks, Gurobi, Pyspark*

- Forecasted demand using ARIMAX and neural network models on large-scale operational data in Databricks, improving prediction accuracy under seasonality and exogenous drivers.
- Integrated Gurobi optimization to allocate resources efficiently, leading to a 30% reduction in operational overhead while preserving SLA adherence.

EDUCATION

Rutgers, The State University of New Jersey **New Brunswick, NJ, USA**

Master of Science in Data Science, CGPA- 3.7/4.0 Aug 2023 - May 2025

G.B.Pant University of Technology **Uttrakhand, India**

Bachelor of Technology in Electronics & Communication Engineering, CGPA- 8.00/10.0 Aug 2018 - Jul 2022