Sajeev Singh

+1 (732) 558-9741 | sajeevysingh@gmail.com | Linkedin | Github

SUMMARY

Data Scientist with expertise in **machine learning**, **statistical modeling**, and **data-driven decision-making**. Proficient in developing scalable decision systems, optimizing predictive models, and creating actionable insights to improve business outcomes. Skilled in Python, R, SQL, and modern software development practices, with hands-on experience in agile methodologies and DevOps principles.

EDUCATION

Rutgers University — ***New Brunswick, New Jersey***
Master of Science in Data Science, **CGPA: 3.8/4.00** — *Aug 2023 - May 2025*

G.B.Pant University of Technology — ***Uttrakhand, India***
Bachelor of Technology in Electronics & Communication Engineering, **CGPA: 3.5/4.00** — *Aug 2018- Jul 2022*

TECHNICAL SKILLS

Core Domain Expertise: Machine Learning, Statistical Analysis, Predictive Modelling, NLP, Deep Learning, Reinforcement Learning, Data Analytics
Programming Languages: Python, R, Julia, SQL, C++
Database Technologies: MongoDB, Cassandra, ScyllaDB, Redis, Neo4j, MySQL, OracleSQL, Hadoop, Apache Spark, Apache Kafka, Databricks, Snowflake
Other Computing and programming: Scala, SPSS, ETL, STATA, RestAPI, Selenium, Robot Framework, SAS, Pytorch, Pyspark, Meta Learning, MS-Excel, Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, Streamlit
Data Visualization Tools: PowerBI, Tableau, Qlik, Looker**,** Google Data Studio

PROFESSIONAL EXPERIENCE

Med-Kick LLC — *Oct 2024 – Present*
Data Science Intern — *Florida, USA*

- Deployed a machine learning model using **AWS SageMaker**, triggered by an **AWS Lambda** function to process audio recordings uploaded to Amazon S3, ensuring real-time transcription generation, AI summary and scorecard for QA of patient-nurse interaction and eliminating manual delays following **HIPAA**.
- Created a pipeline that generates **AI-voice cloned voicemails** and emails triggered by **CRM notifications**. Designed personalized reminders for office visits, improving patient engagement and reducing no-show rates by **30%**. Streamlined the workflow to enhance efficiency and accuracy in patient outreach.
- Developed comprehensive operational dashboards using **Google Data Studio** to track key performance metrics, including patient enrollment progress, communication effectiveness, and team productivity. Integrated data from **CRM and VOIP** systems to provide actionable insights, streamlining reporting processes and improving enrollment outcomes by **15%.**

DataFlow Group — *Dec 2022 – Jul 2023*
Data Science Engineer — *Noida, India*

- Created monitoring scripts using **Prometheus** to track pipeline health and reduce downtime. Optimized ETL processes for real-time data movement, leveraging **Airflow** and **Python** to reduce pipeline downtime by **20%.**
- Developed and deployed **XGBoost**-based tampering detection model, achieving a **30% reduction in manual labor** for document verification, meeting operational efficiency and risk reduction goals.
- Independently engineered an end-to-end computer vision solution for passport detection, leveraging MRZ and data retrieval techniques. Verified over **1000 applicants' details** across roles, reducing manual workload by **50%** and enhancing process efficiency.

Larsen & Toubro — *July 2022 – Nov 2022*
Graduate Engineer, Transport — *Noida, India*

- Project management including planning and execution tracking for **overhead electrification** and **signaling system** in new rail line introduced between Dadri-Khurja.
- Automated stock management workflows using Python and SQL, reducing manual data entry errors by 50% and enhancing vendor tracking accuracy.

RESEARCH PUBLICATION

Using inception modules to improve the accuracy of image classification (6th international conference ICCVBIC 2022) — *2022*

- Implemented an image classification model using inception modules, achieving an accuracy improvement of **16%** over baseline models, demonstrating efficacy of leveraging advanced convolutional architecture. Datasets included were CIFAR, MNIST and ImageNet

Object detection chest band and cane stick for assistance of visually impaired people (IEEE) — *2023*

- Engineered a computer vision-based system for detecting obstacles and providing voice output mentioning distance to obstacle connected via RaspberryPi Bluetooth module. Integrated sensor equipped cane sticks for pothole detection through tactile feedback.

PROJECTS

- **Bikeshare Optimizer-** Developed a system integrating geospatial analysis to optimize ride-sharing routes and reduce travel time in urban areas. Leveraged open-source geospatial datasets like **OpenStreetMap**, and applied **PostgreSQL** with **PostGIS** for data management. Implemented machine learning models, including **K-Means clustering** and **Dijkstra's algorithm**, to analyze traffic patterns and dynamically adjust routes.
- **PHI Masking and Patient FAQ Chatbot Application-** Developed an end-to-end secure application for patient engagement by integrating **PHI masking** and **chatbot functionalities**. Fine-tuned Microsoft's **Presidio model** with **regex** to accurately detect and mask PHI in unstructured data, ensuring compliance with **HIPAA** and **GDPR** standards. Leveraged **OpenAI's GPT API** to create a conversational chatbot for answering **FAQs and general patient queries**. Deployed the application on an **AWS EC2** free instance, incorporating **SSL** encryption and secure **API** gateways for robust data security.
- **Real-time Supply Chain Monitoring System**- Developed a real-time data pipeline using **OEC International Trade Database** for predicting demand for an e-commerce supply chain using **Apache Kafka** for streaming data ingestion and **Airflow** for orchestrating batch processes. Simulated inventory and shipment data using synthetic data sources, storing results in **Google BigQuery** and creating **Looker** dashboards to visualize insights on inventory status, shipment delays and demand forecasts.