SHAHRUKH SOHAIL

srk.soh7@gmail.com | ▮▮▮▮▮▮▮ | [LinkedIn](#) | [GitHub](#) | [Website](#) | Chicago, IL

SKILLS

Programming Languages: Python, Java, SQL, R, GIT, Scala.
Deployment: AWS, GCP, Docker, Kubernetes, Fast API, Flask, Django, Apache, Snowflake.
Statistics: ANOVA, Hypothesis Testing, A/B Testing, Statistical Analysis, Time Series Analysis.
Data Analysis: Tableau, Power BI, R-Studio, Excel, Pandas, NumPy, Scikit-learn, Matplotlib, Neo4j.
Data Engineering: ETL, Hadoop, Hive, Spark, Airflow, Kafka, Cassandra, MongoDB, BigQuery, S3, EC2.
Machine Learning: Regression, Classification, Clustering, XgBoost, Decision Trees, Arima, NLP, Pytorch, Tensorflow.
Certifications: AWS Certified Cloud Practitioner, Google Data Analytics, Foundational Data and AI in Google Cloud

WORK EXPERIENCE

Ground Network Solutions — Chicago, IL
Data Analyst — Aug 2023 – Present
- Developed and deployed highly efficient and scalable data extraction pipelines on AWS, utilizing Redshift for processing and analyzing a substantial 5TB dataset, resulting in a 20% improvement in data processing workflows.
- Processed 1TB of unstructured data with ETL pipelines using Apache Airflow to automate data validation and monitoring, leading to a notable 20% enhancement in forecasting performance metrics.
- Designed and maintained a collaborative CI/CD DevOps data pipeline, ensuring seamless infrastructure and code deployment for a new product, and successfully scaled it from 0 to 100,000 daily active users.

Pure Platform — Chicago, IL
Machine Learning Intern — Jan 2023 – May 2023
- Implemented high-performance text extraction solutions using OpenCV and OCR on 10TB of product images, reducing manual effort by 15% and achieving 75% accuracy.
- Developed a Puppeteer-based E-Commerce Product Details Extractor using Python, reducing manual effort by 50% and achieving an impressive extraction accuracy rate of 95%.
- Led a 6-person team in developing a Product Details Extractor, achieving 15% revenue growth through impactful communication with key stakeholders and effective conflict resolution.

OMS — Lahore, PK
Data Scientist — July 2018 – April 2021
- Developed and deployed an interactive Tableau dashboard in production environment, saving 20 hours per week of manual reporting work and enabling data-driven strategy making for 2020-2021.
- Managed a 500 GB data warehouse for the Information Systems department which included financial, project tracking and equipment stock data reducing annual operational cost by $12k.
- Predicted future resource requirement using multivariate regression, to reduce contractor spendings by $10k.
- Worked closely with cross functional engineering teams to forecast demand and power production of a solar farm using ML regression and improved predictions by 40%.
- Led the design of a data intelligence tool for aggregating unstructured data from 7+ sources, enabling deep analysis which directly resulted in revenue growth by 7%, showcasing collaboration and creative problem-solving skills.

PROJECTS

Real-time fraud detection system using Apache Kafka & Apache Spark — May 2023 - Aug 2023
- Developed a real-time fraud detection system using Kafka and Spark, processing over 10,000 data events per second from diverse sources, resulting in a 25% reduction in fraudulent transactions.
- Designed and implemented custom fraud detection logic, reducing false positives by 15% while maintaining a 98% accuracy rate, enhancing the system's efficiency and effectiveness in safeguarding financial transactions.

EDUCATION

ILLINOIS INSTITUTE OF TECHNOLOGY — Chicago, IL
Master of Science in Data Science — Aug 2021 – May 2023
Cumulative GPA: **3.92/4.0**

GIK INSTITUTE — Topi, PK
Bachelor of Science in Electronics Engineering — Aug 2014 – June 2018

ADDITIONAL

Honors: Upsilon Pi Epsilon Recognition of Outstanding Talent in Computing at Illinois Tech
Activities: Art Exploration, Photography, Photoshop, Swimming, Tennis, Travel