

# SHAHRUKH SOHAIL

srk.soh7@gmail.com |  | [LinkedIn](#) | [GitHub](#) | [Website](#)

## EDUCATION

### ILLINOIS INSTITUTE OF TECHNOLOGY

Master of Science in Data Science

Cumulative GPA: **3.92/4.0**

Relevant Coursework: Data Structures, Algorithms, Linear Algebra, Database Organization, Applied Statistics, Big Data Technologies, Data Preparation & Analysis, Online Social Network Analysis, Project Management

Chicago, IL

Aug 2021 – May 2023

## WORK EXPERIENCE

### Ground Network Solutions

Chicago, IL

#### Data Analyst

Aug 2023 – Present

- Orchestrated business operations by streamlining the monthly revenue reporting system and creating multiple Tableau dashboards, resulting in a 93% reduction in calculation time for management.
- Pioneered integration of multiple database systems in Amazon Redshift to eradicate more than 50% of inconsistency issues, resulting in enhanced customer engagement.
- Engineer solutions using strong knowledge in ML algorithms and predictive models to investigate problems, detect patterns and recommend solutions for increasing customer footprint.

### Pure Platform

Chicago, IL

#### Machine Learning Engineer

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.
- Architect data pipelines to fetch data from multiple data sources like SQL Server and create custom tables and views to meet business needs.
- Processed unstructured data with ETL pipelines using Apache Airflow to automate data validation and monitoring, leading to a notable 20% enhancement in forecasting performance metrics.
- 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.
- 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.

## SKILLS

**Programming Languages:** Python, Java, SQL, R, GIT, Scala.

**Deployment:** AWS, GCP, Azure, Docker, Kubernetes, Fast API, Flask, Django, 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, Apache, 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

## 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.

## ADDITIONAL

**Honors:** Upsilon Pi Epsilon Recognition of Outstanding Talent in Computing at Illinois Tech

**Activities:** Art Exploration, Photography, Photoshop, Swimming, Tennis, Travel