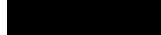


# SHAHRUKH SOHAIL

Chicago, IL (open to relocate at own expense) | [srk.soh7@gmail.com](mailto:srk.soh7@gmail.com) |  | [LinkedIn](#)

## SUMMARY

Data Science professional with over 3 years of experience, specializing in Data Engineering, Machine Learning and Advanced Analytics. Demonstrated expertise in driving decision-making and operational efficiency through innovative data solutions across various global industries. Possesses strong communication skills, adept at translating complex data insights into clear, actionable strategies for cross-functional teams and stakeholders.

## WORK EXPERIENCE

### Opportunity AI Data Engineer

Chicago, IL

April 2024 – Present

- Deployed a pipeline leveraging **Mistral 7B LLM** on AWS, fine-tuned for domain-specific applications to establish a robust process for creating product attributes, resulting in a **30%** increase in product attribute generation accuracy.
- Developed **ETL** workflows utilizing **AWS Glue** and **Lambda** to automate complex data extraction, transformation, and loading processes from multiple data sources, improving data consistency and reducing manual intervention.
- Collaborated with data scientists to optimize **machine learning** models by ensuring data quality and consistency, leading to a **15%** increase in model accuracy.

### Ground Network Solutions

Columbia, MD

#### Data Analyst II

Aug 2023 – May 2024

- Streamlined the monthly revenue reporting system by integrating automated data processing and deploying interactive **Tableau** dashboards, leading to a **93%** reduction in calculation time for management.
- Led the integration of multiple database systems into **Amazon Redshift**, successfully eradicating over **50%** of data inconsistency issues, which significantly enhanced customer engagement.
- Engineered solutions using extensive knowledge of **Machine Learning** algorithms and predictive models to investigate problems, detect patterns, and implement strategies for increasing customer footprint.

### Pure Platform

Chicago, IL

#### Machine Learning Engineer

Jan 2023 – May 2023

- Developed and deployed a real-time fraud detection system with Apache **Kafka** and **Spark**, leveraging data stored in **Snowflake**, processing millions of transactions and reducing fraudulent transactions.
- Designed and executed comprehensive **A/B tests** and formulated data-driven hypotheses to evaluate and enhance the performance of consumer-facing products, resulting in a **25%** increase in user engagement.
- Led a 6-person team in developing a **Product Details Extractor** to extract product information from e-commerce pages, facilitating a **15%** revenue growth by improving data accuracy and enhancing product listings.

### OMS

Calabasas, CA

#### Data Scientist

July 2018 – April 2021

- Created dynamic, real-time **Power BI** dashboards integrated with Trino and **Spark** to effectively communicate strategic insights to senior leadership for informed decision-making.
- Utilized advanced **SQL** functions, including JOIN, Aggregation, and Windowing, on **BigQuery** to solve complex business problems and deliver actionable insights, enhancing data-driven decision-making and operational efficiency.
- Developed **Python scripts** to aggregate unstructured data from multiple sources, enabling advanced statistical analysis and improving data processing efficiency by **30%**, facilitating faster and more accurate insights.
- Collaborated with cross-functional engineering teams to **forecast** demand and power production for a solar farm using advanced predictive analytics, resulting in a **40%** improvement in prediction accuracy.

## EDUCATION

### ILLINOIS INSTITUTE OF TECHNOLOGY

Chicago, IL

Master of Science in Data Science | **CGPA: 3.92/4.0**

Aug 2021 – May 2023

### GIK INSTITUTE

Bachelor of Science in Electronic Engineering

Topi, PK

Aug 2014 – June 2018

## SKILLS

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

**Statistics:** ANOVA, Hypothesis Testing, A/B Testing, Microsoft Excel, Time Series Analysis, matplotlib.

**Data Engineering:** Hadoop, Hive, Spark, Kafka, Apache Airflow, Databricks, Athena, Glue, S3, EC2, Sagemaker.

**Machine Learning:** Classification, NLP, TensorFlow, PyTorch, XGBoost, Deep Learning, Reinforcement Learning.

**Certifications:** AWS Certified Cloud Practitioner, Google Data Analytics, IBM Data Science Professional Certificate.