SHIVAM GOEL

Data Scientist

Phone : +91 9911169308 | **Email:** shivam.goel0897@gmail.com | https://www.linkedin.com/in/shivam-goel01/ | https://github.com/ShivamG0897

CAREER OBJECTIVE

Continuously improving to meet goals and be a professional team player, while applying new technologies for real-world solutions in a dynamic environment.

PROFILE

- 2 year of work experience as a Data practitioner.
- Strong foundation in Statistics, Data Mining and Machine Learning with a good understanding of Big Data ecosystem.
- Graduate from Liverpool John Moores University with an MSc in Data Science and a Certified AWS Cloud Practitioner.
- Gathering d a t a from various sources, improving data quality (Bronze, Silver, Gold transformation), generating meaningful insights (visualization using BI tools) and developing predictive models.
- Well versed in developing and understanding the complex SQL queries (multiple joins) and experience working with various python libraries such as PySpark, PyMongo, Pandas and NumPy.
- Proficient in Power BI versatile features such as Power Query, DAX functions for advanced analytics, interactive visualizations, model views, and implementation of row-level security.

SOFTWARE PROFICIENCY

Data Pipeline Tools	HDFS, MapReduce, YARN, Spark, Kafka
Data Science Frameworks	TensorFlow, PyTorch, Scikit-Learn
Data Visualization	Power BI, Tableau, Matplotlib, Seaborn
Database	MySQL, MariaDB, MongoDB
Programming Languages	Python, R, SQL
Additional Tools	Ubuntu-VM, Dockers, NLTK

EDUCATION

Degree	University	Location	Year of Completion
Msc Data Science	Liverpool John Moores University	Liverpool, UK	Sep'23
Foundation in Engineering	Liverpool John Moores University	Liverpool, UK	Aug'22
BBA	OP Jindal Global University	Haryana, India	Aug'20

WORK EXPERIENCE

Current Company	EMIZA Supply Chain Services Pvt Ltd.
Domain	Warehouse and Supply Chain
Role	Data Engineer
Duration	May – Present 2024
Technical Details	Python 3.8, AWS S3, Azure Data Lake, GitLab, JIRA
Responsibilities	Automating Video Management System, gathering data from multiple sources including warehouses into Azure Data Lakes for ELT.

Company	Lovoj Technologies
Domain	E-Commerce
Role	Data Scientist
Duration	Feb – June 2024
Technical Details	TensorFlow, OpenCV, AWS, Flask, Docker, MongoDB
Responsibilities	To develop AI models for Image recognition, similarity search engine and object detection to automate the product description form for B2B product uploads by suppliers.

Company	CEI Americas
Domain	IT Services and AI Solution
Role	Machine learning Intern
Duration	Feb – June 2024
Technical Details	Python 3.9, NumPy, Pandas, Seaborn, Imblearn
Responsibilities	To develop AI models for Image recognition, similarity search engine and object detection to automate the product description form for B2B product uploads by suppliers.

Company	Denave India Pvt Ltd.
Domain	IT Services and Marketing
Role	Data Analyst
Duration	Sep 2020 – Aug 2021
Technical Details	Advance Excel, MSX, explore.ms, LinkedIn Sales Navigator
Responsibilities	Data Profiling & Data Validation of leads for B2B sales as a vendor of Microsoft products to improve the accuracy, and quality of data for the APAC region.

ACADEMIC PROJECTS

Project I: Predictive Maintenance for Overhead Cranes (Manufacturing, Automotive)
Languages/Tools: *Google Cloud, Python 3.9, Time Series, gradient-Descent*
- Designed a condition Monitoring System for DT Engineering Ltd and **Jaguar Land Rover** to make Electric Overhead Travelling (EOT) crane Safer
- Developed and Deployed on Google Cloud , the Monitoring system showcased Fast-Fourier Transform, Lags with autocorrelation, Survival Analysis and Light Gradient Boosting algorithms.
- Ensured timely maintenance interventions by achieving an accuracy of 95% Kappa statistic and quick response time to minimize crane outage

Project II: Sentiment Analysis on Textual data using MapReduce (Social Media, Academic Research, Big Data)
Languages/Tools: *VM-Ubuntu, Hadoop3.3.0, Spark SQL, MongoDB*
Dataset: William Shakespeare's Writing, Covid tweets, FIFA tweets
- Determine the overall sentiment is positive, negative, or neutral to gauge public opinion and make data-driven decisions in a piece of text, such as a tweet, a novel or comment.
- Acquire text from web/ HDFS storage while preprocessing data using tokenization, stop-word removal and stemming.
- Query specific fields from a MongoDB collection and implement MapReduce for sentiment analysis.
- Effectively displayed sentiment scores and insights from the analysis using suitable visualization methods.

Project III: Real-time Data Ingestion Pipeline with AWS and Kafka services (IoT, Sensors, Big Data)
Languages/Tools: Kinesis Data Streams, Lambda, Kinesis Data Firehose, Glue, Athena, Cloud watch
Dataset: Real-Time High velocity IoT and Sensors data
- Build a robust, scalable, and cost-effective data ingestion pipeline, enabling organizations to ingest, process, and analyze real-time data streams for actionable insights and decision-making.
- Integrate diverse data sources such as IoT devices, and sensors with AWS Kinesis Data Streams/Kafka topics to capture real-time data streams.
- Implement data transformation and enrichment processes using AWS Lambda functions/Kafka Connect to cleanse, format, and enrich the incoming data streams.
- Configure AWS Kinesis Firehose/Kafka Connect connectors to route the processed data streams to Amazon S3.
- Integrate Apache Flink for real-time analytics on the ingested data and monitoring mechanism using CloudWatch.

CERTIFICATIONS
- AWS Certified Cloud Practitioner (expiry Dec'26)
- LinkedIn Badge in MongoDB and R.
- Certified Python Programmer – Aptech

EXTRA CARRICULARS/ OTHER EXPERIENCES
- Fostering a collaborative environment among students and faculty as **Student Representative of Msc Data Science 22-23** Astro-physics Research Institute, Liverpool John Moores University.
- Intern at **Kone Elevators India**, effectively utilizing SAP software to streamline inventory management processes resulting in improved efficiency and cost reduction. [2 months]