Suraj Singh

✉ ss1084si14@gmail.com   ☎ +918130689808   in LinkedIn

PROFESSIONAL SUMMARY

Results-driven Data Scientist with over 5+ years of experience delivering impactful solutions across Generative AI, Agentic AI, intelligent automation, and predictive analytics. Specializing in end-to-end machine learning workflows from data preprocessing and feature engineering to model deployment and real-time integration. Skilled in building Large Language Model (LLM)-based systems, Retrieval Augmented Generation (RAG) pipelines, and semantic search engines for complex enterprise use cases. Adept at translating business problems into data-driven solutions within Agile product teams, driving measurable efficiency and operational intelligence.

SKILLS

Programming: Python, SQL, Flask, FastAPI

- **ML/DL Frameworks:** Scikit-learn, Pandas, XGBoost, Langchain, CrewAI

- **AI/ML Concepts:** Machine Learning, Deep Learning, Natural Language Processing (NLP), Generative AI (GenAI), Large Language Models (LLM), Retrieval Augmented Generation (RAG), Agentic AI, Autonomous Agents, Classification Models, Regression Modeling, Time-Series Forecasting, AWS Bedrock

- **Cloud & Tools:** AWS, Docker

- **Methodologies:** Agile Methodology, Data Visualization, Enterprise Data Analytics, Forecasting

EXPERIENCE

To The New | Data Scientist | 11/2022 – Present

- Led AI/ML model development and Generative AI (GenAI) applications, delivering scalable solutions using LLMs, RAG, forecasting, and classification models across cloud infrastructure, finance, compliance, and operational domains.

Indus Valley Partners | QA Automation Engineer | 05/2020 – 11/2022

- Transitioned from Quality Assurance to Data Science within fintech data platforms, ensuring data integrity and system stability before designing predictive models and analytical dashboards for financial decision-making.

- Gained hands-on experience with regression modeling, data visualization, and real-time financial data ingestion pipelines.

PROJECTS

Newi Intelligent Assistant | 01/2025 – Present

- Developed a GenAI-powered enterprise assistant using LLM-based autonomous agents to streamline operations across HR, IT, and Compliance.

- Designed and integrated multi-agent workflows with a real-time FastAPI backend and CrewAI for cross-functional query resolution.

PWC-GSTR Project | 11/2023 – 12/2024

- Developed an intelligent GST compliance and reconciliation system for automating return filing, transaction validation, and data verification across ERP platforms like SAP, Oracle, and MS Dynamics.

- Integrated a RAG-based Generative AI system for concise document summaries and Q&A, implemented semantic search, and utilized LLMs for keyword extraction to enhance compliance and reconciliation processes.

- Built classification models using historical transaction data to identify invalid, duplicate, and suspicious entries.

AWS Spot Prediction Project | 12/2022 – 10/2023

- Developed a machine learning-based prediction system to forecast AWS Spot Instance terminations, enabling proactive autoscaling and reducing infrastructure disruptions.

- Leveraged historical instance pricing and lifecycle data to train XGBoost models with time-based features like rolling averages and demand spikes.

Data Warehouse Polaris | 06/2022 – 11/2022

- Worked on a financial data warehousing solution designed for hedge funds and asset managers to improve analytics, reporting, and strategic decision-making.

- Developed regression models to forecast security prices using historical market data, enhancing investment planning.

Security and Reference Master | 05/2020 – 05/2022

- Contributed to feature validation, release planning, and quality assurance for 50+ modules spanning data creation, update, and entity search workflows, ensuring end-to-end product reliability.

ACHIEVEMENTS & CERTIFICATIONS

IBM Generative AI Engineering

IBM Data Science

ISTQB Foundation Level

EDUCATION

Bachelor of Technology   07/2014 - 07/2018
PDM College of Engineering

Class XII   04/2013 - 06/2014
CBSE

Class X   04/2011 - 06/2012
CBSE