

KOUSIK S

Data Scientist

skousik789@gmail.com | +44 7501016898 | [LinkedIn Profile](#) | A: London

Professional Summary

Around **4+ years of Machine Learning/Data Scientist** experience, transitioning from Python development to specializing in Big Data Analytics and advanced AI/ML applications. Expertise includes Deep Learning (NLP, LLMs), time series analysis, data engineering (ETL, distributed systems), and deploying models on cloud platforms like AWS, ensuring end-to-end solution delivery

- Experience of all phases of the Software Development Life Cycle (**SDLC**) using methodologies like **Agile and Waterfall**.
- Involved in the entire data science project life cycle and actively involved in all the phases including **data extraction, data cleaning, statistical modelling and data visualization** with large data sets of structured and unstructured data, created ER diagrams and schema.
- Experienced with **Machine learning algorithms. Python (NumPy, Pandas, matplotlib and Sci-kit-learn)**.
- Experience to Perform Information Extraction using **NLP** algorithms coupled with **Deep learning (ANN and CNN, RNN, LSTM, Encoders), Keras and Tensor Flow**.
- Ability to write and optimize diverse SQL queries, working knowledge of **RDBMS like SQL server, NoSQL databases like MongoDB**.
- Experience in building frameworks and automating complex workflows using **Python** for Test Automation
- Good Experience with a focus on **Deep Learning, Machine Learning, Image processing or AI**.
- Experience with time series data, and perform various statistical and **(AI) artificial intelligence, (ML) machine learning algorithm** such as LMS, regression, filtering, correlation, neural network.
- Extensively worked on other machine learning libraries such as **Seaborn, Scikit-learn, SciPy** for machine learning, familiar working with **NLTK, spaCy for Natural Language Processing**.
- Hands-on experience with fine-tuning large language models (**LLMs**), including **GPT**-based architectures, to tailor performance for specific business use cases.
- Good experience of developing web applications implementing **MVC** architecture using **Django and Flask** web application frameworks.
- Having a good working experience of various **Python IDEs using PyCharm, and Jupyter Notebook**.
- Experience in Cloud virtualization and resource management solutions, (e.g., **Docker, Mesos, Kubernetes**).
- Experience dealing with large data sets (**structured, semi-structured and unstructured**), cleaning the data and organizing the data for analyses.
- Experience in designing, implementing, and optimizing end-to-end **ML** workflows, ensuring scalability and performance using tools like **TensorFlow, PyTorch, and Scikit-learn**.
- Very good experience and knowledge in provisioning virtual clusters under **AWS cloud** which includes services like **EC2, S3, and EMR**.
- Experience in **C and C++ languages. Written C and C++ applications** and also migrated many legacy applications to **Python**.
- Designed, developed, and optimized data models and **ETL** pipelines, working with **SQL/NoSQL** databases and distributed systems like **Kafka and MongoDB** to handle high-volume, high-velocity data in real-time applications.
- Good Experience with a focus on **Big data, Deep Learning, Machine Learning, Image processing or AI**.
- Working experience in **SQL Server, MySQL, Spark SQL, PostgreSQL, MongoDB, Cassandra, DynamoDB**.
- Implemented **CI/CD with GitHub** Actions to push code updates to development, staging, and production Docker images, deploying them to the **Amazon EC2 registry**.

- Attend the project review meetings, daily/weekly status meetings and walkthroughs with the project manager in **JIRA**.
- Experienced the full software life cycle in **SDLC, Agile, Devops and Scrum** methodologies including creating requirements, test plans.

Technical Skills

- ❖ **NLP: NLTK, Apache Spark, Scikit-learn**
- ❖ **Python Libraries, NumPy, Pandas,**
- ❖ **Deep Learning: CNN, RNN, LLM, NLTK**
- ❖ **Cloud: AWS, DevOps, CI/CD**
- ❖ **Data Sources: SQL Server, Excel,**
- ❖ **Big Data Ecosystems: Hadoop, MapReduce, HDFS, HBase, Hive, Spark**
- ❖ **Data Visualization: PowerBI, TABLEAU, Matplotlib, Plotly, Seaborn, Matlab**
- ❖ **Programming Languages: Python, SQL, R, No SQL, MongoDB, Postgre SQL, Cassandra, Dynamo DB.**
- ❖ **Scripting Languages: XML, Python,**
- ❖ **Version control: Git, SVN (subversion), Github**
- ❖ **Methodologies: Agile, scrum, waterfall**
- ❖ **Operating System: Linux, Windows, Unix**
- ❖ **Bug tracking Tool: Jira**

Professional Experience

Company: NTT Data

Client: Citi Bank

Role: Machine Learning /Data scientist

Duration: Sep 2022- Till date

Description: I lead initiatives in quality management, where I apply Natural Language Processing (NLP) techniques to streamline processes and deepen business understanding of quality-related events. By analyzing unstructured data from various sources such as customer feedback and quality assurance reports, I uncover actionable insights that drive continuous improvement. This integration of NLP into quality management practices not only enhances operational workflows but also strengthens product quality and customer satisfaction metrics, aligning with organizational goals for excellence and efficiency.

Responsibilities:

- Design, build, and deploy **machine learning models and algorithms** to solve complex business problems.
- Perform data exploration, cleaning, and preprocessing to prepare datasets for model training.
- Used **Python (NumPy, SciPy, pandas, Scikit-learn, seaborn)** and R to develop a variety of models and algorithms for analytic purposes.
- Worked on Natural Language Processing with **NLTK** module of python and developed **NLP** models for sentiment analysis
- Developed machine learning models using recurrent neural networks – **LSTM** for time series, predictive analytics.
- Designed the **(AI) artificial intelligence, (ML) Machine learning** data pipeline for regular monitoring and performance evaluation of the deployed **ML** models.
- Developed and implemented generative models using **TensorFlow and PyTorch** to simulate fraudulent transaction patterns, improving fraud detection accuracy.
- Implemented **RAG** (Retrieval-Augmented Generation) to connect **LLMs** with enterprise health data, ensuring relevant, compliant responses.
- Used **AWS Cloud Watch** for performing Monitoring, customized metrics and file logging and successfully completed the task in using **AWS Lambda** to manage the servers and run the code in the **AWS**.

- Implemented **CI/CD** processes using **Jenkins** ensuring efficient and reliable software delivery. Build Back-end support for Application from ground up using **Python, Shell scripts & Perl**.
- Written python code for exploratory data analysis using **Scikit-learn** machine learning python packages- **NumPy, Pandas, Matplotlib, Seaborn, statsmodels, pandas profiling**.
- Collaborate with ML Engineers and Data Scientist to build data and model pipelines and help in running machine learning tests and experiments
- Utilized statistical **Natural Language Processing** for sentiment analysis, mine **unstructured data**, and create insights; analyze and model structured data using advanced statistical methods and implement algorithms.
- Used **LSTM** recurrent neural network (**RNN**) to solve the use case such as natural language processing (**NLP**).
- Extensively involved in all phases of **data acquisition, data collection, data cleaning**, model development, model validation and visualization to deliver data science solutions.
- Used **Jira** for project management teams such as report, & analysis, work customization, issue /task management, project customization and help to team of all types manage work.
- Developed in Software Development Life Cycle (**SDLC**) including Requirements Analysis, Design specification and Testing as per Cycle in both **Waterfall** and **Agile** methodologies

Environment: Python (NumPy, SciPy, pandas, Scikit-learn, seaborn), AI, LLM, Tensor flow, Pytorch, AWS, CI/CD, Shell Scripts, NLP, LSTM, RNN, Jira, Water fall, Agile, NLTK, Jenkins, AWS Cloud watch.

Company: Global Logic

Client: ANZ Bank

Role: Python Developer

Duration: May 2021 to Sep 2022

Description: Provides a variety of banking services to individuals and corporate customers. The main activities of Bank include accepting deposits, establishing investment funds, and providing personal finance, credit card trading in real estate. The Bank deals with all types of transactions.

Responsibilities:

- Developed entire frontend and backend modules using **Python on Django Web Framework**.
- Used **Python (NumPy, SciPy, pandas, Scikit-learn, seaborn)** and R to develop a variety of models and algorithms for analytic purposes.
- Involved in various Object-Oriented Programming (**OOPS**) concepts and implementing the same for Design and Architecture of the application.
- Worked on data preprocessing, developing different statistical machine learning model and data mining solutions to various business, generating data visualizations using **Python, R, Tableau, Microsoft Power BI, version control with GIT**.
- Worked extensively with **Bootstrap, JavaScript, and JQuery** to optimize the user experience. Used **Python Library Beautiful Soup** for web scraping to extract data for building graphs.
- Contributed to the compilation of requirements with various teams, including Retail Product, Cards, Digital, and Corporate, and discussed business issues, data availability limitations, and expected output.
- Led Customer segmentation for cross-sell models such as personal **finance, credit card, prepaid card**, and probability default risk model utilizing machine learning technique
- Generated ad hoc intelligent lists, such as customer spend by merchant insights for supplying the offer and the prime customer with first man power allowance to target for salary account campaigns
- Employed an **NLP** technique to classify the merchants who provided the offer.
- Developed SQL Queries, Stored Procedures, and Triggers Using Oracle, **SQL and PL/SQL**.
- Worked with **Amazon Web Services (AWS)** using **EC2** for hosting and Elastic map reduce (**EMR**) for data processing with **S3** as storage mechanism.

- Research and document advanced solutions architecture for other enterprise customers including building **CI/CD** pipelines and a robust monitoring system for model performance.
- Implemented **Agile** Methodology for building an internal application.

Environment: Python (NumPy, SciPy, pandas, Scikit-learn, seaborn, AWS, EC2, EMR, CI/CD, S3, Bootstrap, Javascript, JQuery, Python, Library, SQL, PL SQL, Tableau, PowerBI, Git, NLP, Agile,