

Sanket Patole

+91-9561333389 • snkt.pat2@gmail.com • LinkedIn • Github

SUMMARY

I have completed an **MS in Data Science and AI** and have over **7 years** of experience in **Data Engineering & Analytics**. I have used **modular programming** to develop projects, deployed them on AWS, and built CI/CD pipelines.

SKILLS

- **Languages and Libraries** – Python, SQL, Scikit-Learn, Tensorflow, Keras, NLTK, SPACY, Gensim.
- **Machine Learning** – Linear Regression, Logistic Regression, KNN, SVM, Bagging, Boosting, Random Forests.
- **Unsupervised Machine Learning** – Clustering, Outlier Detection, Dimensionality Reduction, Time Series Forecasting.
- **Deep Learning** – ANN, NLP (RNN, LSTM, CRFs, Attention, Transformer Models, Large Language Models).
- **MLOps and Analytics** – Streamlit, Flask, Docker, GitHub, CI/CD Pipelines, AWS, A/B Testing, EDA, Hypothesis Testing.

WORK EXPERIENCE

Persistent Systems | *Senior Data Engineer*

Mar 2022 - Present

Client: Education Advisory Board (EAB)

- Developed data ingestion pipelines to support college recruitment, utilized by Marketing and Analytics teams.
- Upgraded in-house ETL tools, including AWS Lambda migrations to Python 3.9, enhancing efficiency and performance.

Bitwise Solutions | *Data Engineer*

Oct 2018 - Mar 2022

Client: Discover Bank

- Automated metadata curation & upload tools, & developed data ingestion pipelines, improving data governance for client.
- Facilitated metadata management & analysis by creating pipelines for Snowflake & Tableau, enhancing business insights.
- Led the migration of data storage and ETL processes to AWS, including development of a Kibana dashboard for monitoring, streamlining operations and increasing efficiency.

FIGmd India | *Data Engineer*

Nov 2016 - Oct 2018

Project: Regulatory Compliance Platform

- Developed automated tools for Data Gap Analysis & SQL mapping for 1000+ clinical measures, enhancing data analysis.
- Established a data ingestion pipeline for 100+ US hospitals, optimizing healthcare data collection and processing.
- Provided ongoing support to healthcare providers for data integrity and mapping corrections.

PROJECTS

Next-Generation Educational Platform |

Repo : GitHub

- **Award:** Best use of Azure AI, Persistent System's Annual Hackathon 2024
- **Stack:** Python, Vue.js, AWS, Vector DBs, LangChain, RAG | **Objective:** AI-driven personalized learning for grades 1-10.
- **Achievements:** Developed a child-friendly platform with dynamic content creation using LangChain and RAG, fast query responses with Vector Databases, and ensured a safe, personalized learning experience.

Inshorts | *Generating Summary*

Project Repository: Github

- **Tools used:** Python, Numpy, Pandas, Matplotlib, Tensorflow, Keras, Encoder, Decoder, LSTM's, Hugging Face
- Built an **Encoder-Decoder** LSTM's Architecture and **BERT Transformer** model and achieved an accuracy of **94%** also calculated the rouge score to evaluate. Deployed on AWS and build CI-CD Pipeline and used **modular programming**

Loan Eligibility Prediction |

Project Repository : Github

- Focused on Fraud Analytics and data preprocessing through EDA and feature engineering. It implemented Random Forest with **94%**, deployed on AWS EC2 with a CI-CD pipeline and robust modular code

EDUCATION

MS in Data Science

Expected, 2024

Woolf University (SCALER)

CGPA: NA

B.E. (Computer Science)

Completed, 2016

Maharashtra Institute of Technology (MITWPU)