

# Shrweta Naik

✉ shrweta10@gmail.com | ☎ 8087696380 | 📍 India

## Education

### Nagpur University

BACHELOR OF ENGINEERING IN ELECTRONICS ENGINEERING  
GPA: 68

Nagpur

July 2011 - Aug 2015

## Work Experience(5+ years of Experience as Data Scientist (Overall IT Exp-8+) \_

### IBM

DATA SCIENTIST

India

Oct 2024 – Present

- Independently designed and implemented a scalable outlier detection pipeline using PySpark to analyze large transactional datasets for financial anomaly detection at AMEX.
- Developed statistical models (Mahalanobis Distance, Adjusted Outlyingness, DBSCAN) to identify potential and clustered outliers
- Automated data pipeline generation for real-time anomaly detection, enhancing accuracy and efficiency. - Built robust Spark transformations to handle high-volume data with minimal computational overhead

### Tech Mahindra

DATA SCIENTIST

India

June 2022 – Oct 2024

- Collaborate with stakeholders to understand business requirements and user needs for chatbot functionality with generative AI.
- Design and develop chatbot solutions using programming languages, frameworks, and platforms suitable for chatbot development
- Deploy trained models using Vertex AI's deployment services, such as AI Platform Prediction
- Collaborated with cross-functional teams to develop innovative solutions at the intersection of data science, machine learning, and generative AI.
- Proficiently used Git, and Jira for version control and project management.

### NTT DATA

DATA SCIENTIST

India

Dec 2020 – May 2022

- Wrote clean, efficient, and maintainable Python code to develop software applications, tools, and systems.
- Create, query, and manage databases to store and retrieve data efficiently.
- Wrote unit tests and conduct debugging to ensure the quality and stability of the software.
- Design machine learning solutions and algorithms to address specific use cases or challenges.

### Livsense Technologies

DATA SCIENTIST

India

Oct 2019 – Dec 2020

- Conduct research to explore new computer vision algorithms, techniques, and methodologies.
- Preprocess and clean image and video data to prepare it for analysis and model training.
- Apply techniques such as data augmentation, normalization, and transformation.
- Train machine learning models for computer vision tasks using labeled datasets.
- Implement and optimize object detection and tracking algorithms for real-time applications.

### BNYMELLON

QA

India

May 2015 – March 2019

- Develop a comprehensive test plan outlining testing strategies, scope, resources, and timelines.
- Create detailed test cases and test scenarios based on project requirements and use cases.
- Execute manual and/or automated tests to validate software functionality and identify defects.

## Skills

Programming languages:

Java, Python, Typescript, Javascript

Machine Learning:

Linear Regression, Ridge and Lasso Regression, Logistic Regression, Random Forest, Decision Tree, SVM,

K Nearest Neighbour, Adaboost, Gradient Boosting, K means Clustering, Naïve Bayes

Natural Language Processing:

Transformers, Bert, Word Embeddings, Sequence model, Sentiment in text

Generative AI:

LLMS(Gemini, LLama2/3, Azure OpenAI), Langchain, RAG, AI-Agents, MCP

Deep Learning:

Deep neural network, Convolutional Neural Network, Recurrent Neural Network, LSTM

Data Science Libraries:

Pandas, Numpy, NLTK, Keras, Tensorflow, Scikit-Learn, Matplotlib, OpenCV, Seaborn, SciPy

Test Techniques:

Functional Testing, Automation Testing

Database:

SQL, NOSQL

Database Tools:

MongoDB, Postgres, MySQL

Big Data:

Hive, PySpark

Deployment Tools and Services:

Docker, Kubernetes

Cloud:

GCP, AWS

Project Management Tools:

JIRA

Web Framework:

Flask, Django, REST, FastAPI

## Projects

---

### 1. GLOBAL RISK SOLUTIONS (GRS)

**Project Description:**

The GRS Reports project focuses on performance and risk analysis to evaluate how accounts are expected to behave in the future or how they have been performing historically.

**Technologies Used:**Functional testing

### 2. QUALITY PREDICTION OF MACHINES IN FACTORIES

**Project Description:**

This project involves implementing predictive modeling techniques to forecast the quality of machines operating in factories.

**Technologies Used:**Python,Scikit-learn,AWS Sagemaker

### 3. VISION BASED COLLISION AVOIDANCE SYSTEM

**Project Description:**

The project involves implementing an object detection model to create a vision-based collision avoidance system.

**Technologies Used:**Python,TensorFlow, OpenCV,YOLO (You Only Look Once) model,AWS Sagemaker

### 4. MCKINSEY ACADEMY CLIENT PORTAL

**Project Description:**

This project involves the implementation of microservices across the McKinsey Academy client portal to enhance functionality and scalability.

**Technologies Used:**Docker, Kubernetes,RESTful APIs,PostgreSQL

### 5. CISCO LEGAL-LTG

**Project Description:**

The project involves developing a multiclass classifier for fraud detection in partner screening processes at Cisco.

**Technologies Used:**Python,Scikit-learn,SQL, Pandas,Flask for deployment,Vertex AI

### 6. CISCO ASK LEGAL BOT

**Project Description:**

The project involves developing an Ask Legal Bot for the Webex chat space to assist users with legal inquiries.

**Technologies Used:**Python,Generative AI,RAG,Docker

### 7. AMEX: LOYALTY OUTLIER DETECTION

**Project Description:**

The project involves developing a robust and scalable outlier detection pipeline for AMEX, designed to efficiently identify anomalous patterns in large datasets

**Technologies Used:**Python,Pyspark

## Certifications

---

Simplilearn Certification in Data Science with python,

Simplilearn Certification in Machine Learning,

Simplilearn Certification in Deep Learning,

Coursera Deep learning Specialization,

Coursera Introduction to Data Science Specialization

Skills Name

Big Data

Kubernetes

React.js

Azure devops

Jenkins

Ansible

Terraform

Artificial Intelligence (AI)  
JavaScript  
Django REST Framework  
Team Leadership  
Microservices  
FastAPI  
Google Cloud Platform (GCP)  
VertexAI  
MySQL  
Application Development  
RAG  
Continuous Integration and Continuous Delivery  
(CI/CD)  
TensorFlow  
Langchain  
Generative AI  
Large Language Models (LLM)  
Django  
Amazon Web Services (AWS)  
Gcp  
Machine Learning  
Deep Learning  
Natural Language Processing (NLP)  
Computer Vision  
OpenCV  
Data Engineering  
MLOps  
Docker  
SQL  
Pandas  
Data Science  
Data Visualization  
Python (Programming Language)  
Flask  
TypeScript

