**Sanjeev Gaikwad Data Scientist | Machine Learning Engineer | NLP, Deep Learning, Cloud (AWS/GCP), MLOps 9137109402 |gaikwad.sanjeev@outlook.com|** [**LinkedIn**](https://www.linkedin.com/in/gsv18/) **|** [**Portfolio**](https://gaikwadsanjeevv.github.io/) **|** [**Github**](https://github.com/gaikwadsanjeevv)

**Summary**

Versatile and results-driven Data Scientist with 6+ years of software engineering experience and a strong academic foundation in Computer Science. Skilled in building scalable ML models, end-to-end data pipelines, and real-time analytics platforms using Python, scikit-learn, TensorFlow, Spark, AWS, and GCP. Strong command of data structures, algorithms, and system design principles, with hands-on experience deploying cloud-native, production-grade solutions. Passionate about solving complex problems through data and currently seeking impactful roles in Data Science or ML Engineering within high-performance product teams.

**Education**

University of Central Missouri, MS Computer Science (2022-2024) – Lee Summit, Kansas.   
SGGSIE&T- B.Tech, Production Engineering | GPA 7.77 – Nanded, Maharashtra, India.

**Skills and Coursework**

**Coursework:** Big Data Analytics, Machine Learning, Advanced Algorithms, Stats for Data Science & AI

**Skills:** Python (NumPy, Pandas, scikit-learn, TensorFlow, Keras, Statsmodels), SQL (PostgreSQL, MySQL, BigQuery), PySpark, XGBoost, LightGBM, CatBoost, Airflow, Apache Kafka, Apache Spark (Batch and Streaming), Hadoop Ecosystem (MapReduce Basics), AWS (S3, Lambda, EC2, ECS, RDS), GCP (BigQuery, Cloud Run, Cloud Functions, Pub/Sub), Docker, Kubernetes (GKE basics), FastAPI, Flask.

Data Structures & Algorithms (Recursion, Trees, Graphs, Dynamic Programming, Sliding Window), System Design (Microservices, Load Balancing, Stateless APIs, Data Partitioning, Event-Driven Architecture, High Availability).

Feature Engineering, Data Preprocessing (Encoding, Scaling, PCA), Model Evaluation (Accuracy, AUC, RMSE, MAE, R²), Cross-Validation, Hyperparameter Tuning (GridSearchCV), MLOps Practices (Model Deployment & Monitoring), ETL Workflow Design, Machine Learning Model Deployment, Deep Learning (CNN, RNN, LSTM), Natural Language Processing (SpaCy, TF-IDF, Text Classification), Cloud Deployment, Distributed Systems Concepts, Git, GitHub, Agile (Scrum), Jupyter Notebooks.

**Work Experience**

**EnCiv- Full Stack Developer | Irvine, CA | July 2024-Present**

• Designed ML-based user intent classification models using SpaCy and scikit-learn to categorize civic participation posts in real-time.  
• Integrated NumPy + Pandas pipelines for feature engineering on large-scale civic discourse data (100K+ monthly entries).  
• Built real-time prediction services using FastAPI and deployed them via Google Cloud Run, reducing latency by 35%.  
• Leveraged Apache Kafka and GCP Pub/Sub to collect and process user activity logs for behavior analytics.  
• Applied text vectorization, TF-IDF, and logistic regression for user intent modeling, reaching 87% F1-score.  
• Improved model accuracy through cross-validation and hyperparameter tuning using GridSearchCV and XGBoost.  
• Led system design to scale ML predictions, reducing prediction response time from 450ms to 150ms for 50K+ concurrent users.  
• Applied scalable system design principles (microservices, load balancing, stateless APIs) to deploy ML inference systems with high availability.

**Sysark Datasol Pvt. Ltd (Study-Hub)- Software Developer| Hyderabad, India | April 2018-May 2022**

• Built student performance prediction models using multivariate regression and decision trees in scikit-learn.  
• Processed time-series learning behavior data using Pandas and NumPy to optimize ETL workflows.  
• Designed academic dashboards with Seaborn and Plotly to help educators drive targeted interventions.  
• Tuned models using Lasso and Ridge to reduce overfitting and improve predictive accuracy.  
• Developed modular ML services in Flask, deployed on AWS ECS and secured using API Gateway and JWT.  
• Streamlined data transformation using SQL and Pandas merges for faster processing.  
• Integrated AWS Lambda functions to automate periodic model retraining with new student data.  
• Architected microservice-based ML components to align with the platform’s evolving infrastructure.

**Swashtech Software Paradigm (Cure.fit)- Software Developer| Pune, MH| July 2016-March 2018**

• Developed a wellness recommendation engine using classification models (KNN, Decision Tree, SVM) for user fitness goals.  
• Preprocessed biometric and user-activity datasets using Pandas and NumPy; applied K-Means clustering for segmentation.  
• Designed a basic deep learning model (Keras LSTM) to predict workout attendance trends.  
• Deployed ML models using AWS EC2 and S3 to integrate with cloud-hosted fitness tracking data.  
• Built batch ETL workflows using PySpark to process wearable device data for training pipelines.  
• Used TensorFlow and Keras to create prototype nutrition models informed by health history and biometric inputs.  
• Documented and evaluated models with R², MAE, and RMSE for performance validation across user cohorts.