Samyak Anand

Data Scientist | ML Engineer | AI Enthusiast

✉ samyak.g.anand@gmail.com 📞 +91 9623719948 ○ GitHub in LinkedIn

PROFESSIONAL SUMMARY

Data Scientist with 6+ months of internship experience building predictive analytics models, clinical ML systems, and MLOps pipelines. Skilled in Python, scikit-learn, TensorFlow, and cloud deployments on AWS. Passionate about transforming complex data into actionable insights and expertise in developing and deploying machine learning models to solve real-world problems.

SKILLS

- Python • SQL • HTML/CSS • Git • Docker • VS Code • Jupyter Notebook • pandas • NumPy • scikit-learn • XGBoost • Feature Engineering
- PCA • Predictive Modeling • NLTK • spaCy • Regression • Classification • Clustering • Data Preprocessing • TensorFlow • Keras • PyTorch • CNNs
- RNNs • NLP • Hugging Face Transformers • FastAPI • Flask • REST APIs • MLflow • DVC • DAGsHub • CI/CD • GitHub Actions • Jenkins
- AWS • Azure • Google Cloud • Drift Detection • PostgreSQL • MySQL • MS SQL Server • MongoDB • VectorDB • Power BI • Tableau • Streamlit
- Matplotlib • Seaborn • Plotly • Excel • Machine Learning • Deep Learning • Data Visualization • Statistical Analysis • API Development
- Cloud Deployment • Data Ingestion • Data Cleaning • Model Deployment • Experiment Tracking • Web Scraping • ETL • Natural Language Processing
- Data Analysis • Data Mining

EXPERIENCE

Evoastra Ventures Inc
Jun 2025 - Present

Data Scientist Intern

- Led a team of 5 interns to build end-to-end Machine Learning (ML) workflows using Python, SQL, scikit-learn, and Power BI, covering data ingestion, cleaning, modeling, and interactive dashboards.
- Developed a CARS24 web-scraping & analytics pipeline to track real-time pricing trends and uncover market patterns, enhancing market analysis capabilities.
- Created an AI-powered resume parser (Natural Language Processing (NLP) + Flask) that extracts candidate details and matches resumes to job descriptions, accelerating shortlisting processes.

Gilbert Research Center
Mar - Jul 2024

Data Science & ML Intern

- Engineered clinical Machine Learning (ML) models and modular pipelines with drift detection and API deployment using Flask, MLflow, and DAGsHub.
- Applied advanced statistical methods to improve accuracy and interpretability of clinical data insights.
- Collaborated cross-functionally to turn data insights into clinical recommendations, improving patient outcomes.

PROJECTS

AI Resume Analyzer

- Developed a Flask application that parses PDF/DOCX resumes and pasted job descriptions to automatically extract contact information, skills, education, and experience.
- Computed a hybrid relevance score (50% semantic via Sentence Transformer, 30% TF-IDF cosine similarity, 20% skill overlap) in seconds, improving candidate matching efficiency.

Disease Predictor: Multi-Disease ML System

- Built modular Python pipelines to predict heart disease, breast cancer, kidney disease, liver disease and women's diabetes, covering data ingestion, schema validation, transformation, model training and drift detection.
- Integrated MLflow for experiment tracking and DagsHub for version control; exposed predictions via Flask/Waitress APIs on Render with a responsive HTML/CSS/Bootstrap UI and MongoDB backend.

Car24 Hyderabad Web Scraper & Analysis

- Built a Python pipeline using requests and BeautifulSoup to crawl Cars24's Hyderabad listings and extract vehicle details (year, mileage, fuel, transmission, engine, ownership, insurance).
- Structured a modular codebase (scrape.py & analyze.py) for clear ETL stages, enabling seamless future upgrades (Parquet/SQL storage, headless-browser support).

Network Security ML Pipeline

- Built real-time intrusion detection ML pipeline covering data ingestion, drift detection, transformation, training & evaluation (Python, FastAPI, Docker, MongoDB).
- Integrated MLflow for tracking, DAGsHub for version control, and deployed on AWS; achieved 99.17% F1-score (train) and 97.61% F1-score (test).

EDUCATION

B.Tech in Computer Science & Engineering
Sep 2021 - Aug 2024

Ballarpur Institute of Technology | Maharashtra, India

Diploma in Computer Science
Jun 2018 – Aug 2021

Bajaj College of Polytechnic | Maharashtra, India

ACHIEVEMENTS & CERTIFICATIONS

Full Stack Data Science and AI – Naresh IT

Data Science & ML (Basics to Advanced) – Udemy