

# Sandeep Bandi

Hyderabad, Telangana | +91 94944 20777 | bandisandeep1423@gmail.com | LinkedIn | GitHub | Website

## SUMMARY

Versatile **AI/ML Engineer** and **MLOps Practitioner** with over 3 years of hands-on experience in architecting, developing, and deploying scalable, production-grade **Machine Learning** and **Generative AI** solutions—specializing in **deep learning**, **natural language processing (NLP)**, and **intelligent automation** across cloud-native environments to drive innovation, operational efficiency, and measurable business value.

## TECHNICAL SKILLS

**Programming Languages** : Python, R, SQL, Flask

**Data Science & Analytics**: Data Preprocessing, Feature Engineering, Statistical Modeling, Probability Theory

**Machine Learning & Deep Learning**: scikit-learn, TensorFlow, Keras, PyTorch, ANN, CNN, RNN

**Natural Language Processing (NLP)**: Transformers, Hugging Face, BERT, GPT

**MLOps & CI/CD**: MLflow, DVC, GitHub Actions, Jenkins, Apache Airflow

**Cloud & Computing**: AWS (EC2, S3, Lambda), GCP (BigQuery), Distributed Systems Concepts

**Databases & Big Data**: MySQL, PostgreSQL, SQL Server, MongoDB, AWS Glue

**Data Visualization & BI Tools**: Power BI, Matplotlib, Seaborn, Excel

**Development Tools**: Git, GitHub, Docker, Kubernetes, Jupyter Notebook, VS Code

**LLM Tools & Frameworks**: LangChain, OpenAI APIs, Hugging Face Models

**Methodologies & Concepts**: Agile, Data Warehousing, Business Analytics

## EXPERIENCE

### MLOps Engineer

May 2025 – Present

*PairSoft · Full-time, Remote*

- **Papersave (GL Coding Automation)**: Built and deployed an end-to-end deep learning pipeline for multi-tenant GL account prediction using TensorFlow, Docker, and Azure.
- Implemented MLflow for experiment tracking, versioning, and monitoring across tenants.
- Designed scalable architecture to support multiple clients with minimal configuration overhead.
- **APRO (Invoice–PO Matching, GenAI)**: Leveraged LangChain and LLM APIs to automate invoice–PO data matching using GenAI, improving data reconciliation efficiency.
- Integrated GenAI services into enterprise AP workflows, enhancing automation and accuracy.
- **Tech Stack**: Python, TensorFlow, Scikit-learn, Hugging Face, LangChain, MLflow, Docker, Azure

### Freelancer (NLP/GenAI Engineer)

Oct 2024 – Present

*Freelancer · Remote, Contract*

- Worked on research-driven NLP and GenAI projects with clients and academics exploring LLMs in real-world and experimental use cases.
- Built LangChain pipelines for document retrieval, question answering, and semantic search across diverse datasets.
- Implemented prompt engineering, zero-shot classification, and LLM fine-tuning using Hugging Face, OpenAI, and PaLM.
- Deployed FastAPI interfaces for rapid prototyping, client testing, and feedback loops.
- Authored whitepapers and PoCs bridging cutting-edge NLP research with enterprise solutions.
- **Tech Stack**: Python, LangChain, Hugging Face, OpenAI API, FastAPI, Flask, Docker, Git, Weaviate, Pinecone

### AI/ML Engineer

Jun 2024 – Sep 2024

*BeWorld App · Part-time · Hyderabad, India*

- Designed and deployed a personalized feed ranking system for social media engagement using hybrid recommendation (collaborative + content-based).
- Built ML models (Logistic Regression, XGBoost) to predict post interest scores using MongoDB-sourced engagement and metadata features.
- Served ranked content via FastAPI endpoints and deployed scalable services on AWS EC2.
- Implemented CI/CD pipelines with Docker and Jenkins for reliable model deployment.
- Collaborated with product and backend teams for smooth integration into the live app.

## Assistant System Engineer

Jan 2022 – Feb 2024

*Tata Consultancy Services · Full-time · Hyderabad, India*

- Built and deployed ML models (Scikit-learn, XGBoost) for risk scoring and fraud detection in insurance workflows.
- Designed data pipelines for claims and policyholder data processing, improving reporting accuracy and efficiency.
- Conducted EDA, feature engineering, and validation to ensure model quality in production environments.
- Deployed solutions on AWS EC2 and integrated into enterprise batch-processing pipelines.
- Collaborated across data, business, and cloud teams to deliver actionable ML tools.
- **Tech Stack:** Python, Pandas, Scikit-learn, XGBoost, AWS EC2, SQL, Jupyter, Git

## Project Intern

November 2021 – February 2022

*iNeuron.ai*

*Remote*

- Developed a machine learning system for income price prediction through an end-to-end ML pipeline.
- Executed core stages including **data ingestion**, **EDA**, **transformation**, **model training**, **prediction**, and **deployment**.
- Built and containerized the solution using **Python** and **Docker**; deployed it on AWS for real-time inference.
- Gained hands-on experience with VS Code, GitHub version control, and cloud-based model delivery.

## EDUCATION

---

### Lovely Professional University

Punjab, India

*Bachelors of Technology in Computer Science Engineering*

*2016 - 2020*

## PROJECTS

---

### Hotel Reservation Cancellation Prediction | *Python, MLflow, Jenkins, Docker, Flask*

[GitHub](#) | [Live App](#)

- Developed an end-to-end machine learning solution to predict hotel reservation cancellations, enabling proactive customer engagement and revenue optimization.
- Designed and implemented the entire ML pipeline, from data ingestion (GCP) and preprocessing to model training and deployment.
- Conducted feature engineering, data versioning (DVC), and model versioning for reproducible experiments.
- Implemented experiment tracking (MLflow, CometML) and automated model deployment using CI/CD (Jenkins, Docker, GCP Run).
- Built a Flask-based web application with an interactive UI (HTML, CSS) for real-time predictions.

### Hybrid Anime Recommendation System | *Python, Scikit-learn, Pandas, Flask, Streamlit*

[GitHub](#) | [Live App](#)

- Developed a hybrid recommendation system combining collaborative filtering and content-based filtering techniques to provide personalized anime suggestions.
- Utilized user-item interaction matrix and anime metadata to generate tailored recommendations.
- Designed a modular and scalable ML pipeline ensuring clean code practices and easy experimentation.
- Implemented a Flask-based backend and integrated with a Streamlit front-end for deployment-ready application.

### Adult Census Income | *Python, Flask, scikit-learn, Pandas, MongoDB, AWS, S3, SQL*

[GitHub](#) | [Live App](#)

- Designed and implemented a binary classification system to predict whether individuals earn more than \$50K annually, leveraging Logistic Regression for the best performance with an accuracy of 85%. Developed a modular workflow comprising Data Ingestion, Data Transformation, and Model Training pipelines.
- Built a user-friendly web application for real-time predictions, incorporating a preprocessing pipeline for handling numeric and categorical variables and integrating Logistic Regression models for seamless deployment.

### Gemstone Price Prediction | *Python, LightGBM, XGBoost, scikit-learn*

[GitHub](#) | [Live App](#)

- Built and deployed a predictive model leveraging supervised learning techniques to estimate gemstone prices, optimizing performance using cross-validation and advanced regression algorithms.
- Analyzed and visualized key factors influencing gemstone prices using tools like Python, Pandas, and Matplotlib, enabling clear interpretation of market trends and model outputs.

## CERTIFICATIONS

---

**Full Stack Data Science Certificate** | iNeuron

[Certificate Link](#)

**Complete ML, DL, NLP Bootcamp Certification** | Udemy

[Certificate Link](#)

**Python** | Kaggle

[Certificate Link](#)