Bai Phaneendra

Data Science Trainee

phaneendra.bayi@gmail.com

+91-8106650115

Hyderabad, Telangana

https://www.linkedin.com/in/bai-phaneendra/

https://github.com/Saiphani2105

https://huggingface.co/Phaneendrabayi

Profile

Aspiring Data Scientist with hands-on experience in **Python**, **machine learning**, and **deep learning**. Skilled in **data preprocessing**, **model development**, and **evaluation** using **Scikit-learn**, **TensorFlow**, and **Google Gemini**. Strong foundation in **statistics**, **data visualization**, and **exploratory data analysis (EDA)**. Passionate about building **scalable**, **data-driven solutions** to solve real-world problems and deliver **actionable insights**.

Skills

Python
Pandas, SQL, Java Script

Data Analytics
Data Cleaning, Statistical Analysis, EDA, Visualization using Matplotlib, Seaborn, Plotly, Power BI

Databases
SQLite, MySQL – Schema design, Querying, Joins, User Authentication Systems

Machine Learning & AI
Supervised & Unsupervised Learning ,Predictive Modeling, NLP, Computer Vision, Generative AI, LLM Fine-tuning

Deep Learning
Neural Networks (ANN, CNN, RNN, LSTM), Whisper for Transcription, Text Generation

Tools & Platforms:
Jupyter Notebook, Git, VS Code, Google Colab, Streamlit Cloud

Professional Experience

Machine Downtime Optimization Analyst, *360DigiTMG* (*Intern*)
10/2024 – 11/2024

- Analyzed downtime data using **Python** and **MySQL**, uncovering root causes through correlation analysis and trend forecasting.
- Created **interactive dashboards** to visualize key metrics and trends.
- Delivered a high-impact **presentation** to management, supporting **data-driven decision-making**.

Data Science and Gen AI Intern, *Innomatics Research Labs*
01/2025 – 03/2025

- Built AI apps using Streamlit and **Gemini API**, integrating **LangChain** for **chat memory** and **SQLite** for user login.
- Developed a **Data Science Tutor** with user authentication, chat history, and personalized Q&A.
- Created a **Subtitle Generator** using Whisper for transcription and keyword-based search.
- Built **SQL Coder**, **AI Code Reviewer**, and **AI Travel Assistant** as GenAI tools.
- Developed **Chat with PDF** for querying uploaded documents using LLMs.

Projects

Sign Language Recognition System, (*ML, OpenCV, Mediapipe*)

- **Developed** a **real-time system** to **translate hand gestures into text** for the **hearing-impaired** using **machine learning** and **computer vision**.
- **Extracted 21-point hand landmarks** using **Mediapipe** and **engineered features** for accurate gesture recognition.
- **Trained** a **Random Forest Classifier** on labeled gesture data, achieving over **90% accuracy**.
- **Integrated OpenCV** for **real-time webcam input**, live gesture detection, and text output display.
- **Delivered** a **lightweight**, **responsive application** that supports **accessible communication**.

Tag Prediction for Medium Articles, (*ML, NLP*)

- Developed an **NLP-based multi-label classification model** to suggest relevant tags from Medium article content, enhancing content discoverability.
- Applied **text preprocessing** techniques such as **tokenization**, **stopword removal**, and **lemmatization**, followed by **TF-IDF vectorization**.
- Trained a **Multinomial Naive Bayes classifier** to handle multi-label tag prediction with improved precision and scalability.
- **Deployed the model on Hugging Face Spaces** using **Streamlit**, enabling real-time tag suggestions for user-provided article inputs.
- Delivered a **lightweight, accessible web app** that reduces manual tagging effort and supports authors with **automated, high-quality tag recommendations**.

Sentiment Analysis & Text Generation, *TensorFlow, Keras, LSTM*

- **Built** an end-to-end NLP solution to **classify customer sentiments** and **generate sample content** using deep learning.
- **Developed an LSTM-based sentiment classifier** with **Keras** and **TensorFlow** to label reviews as **positive**, **negative**, or **neutral**.
- **Trained a text generation model** using **LSTM networks** to produce contextually relevant and grammatically coherent sentences.

Education

Kakatiya University,
Masters in Computer Application (*MCA*)
2022 – 2024 | Hanumakonda, Warangal

DonBosco Degree College,
Bachelor of Science (*MSCS*)
2019 – 2022 | Erraggada, Hyderabad

Courses

Python

Exploratory Data Analysis

FLASK

Power BI

MySQL

Machine Learning

Deep Learning

Generative AI

Certificates

- Oracle Cloud Infrastructure 2024 Generative AI
- Advanced Data Analysis Internship Program
- Power BI
- Advanced Data Science with Python
- ISRO Participation Certificate of AI and ML for Geodata Analysis
- Data Visualisation : Empowering Business with Effective Insights
- Accenture North America - Data Analytics and Visualization Job Simulation
- Workshop on AI and ML with Data Science
- Hugging Face Agents Course

- **Deployed both models** in a **Streamlit app**, enabling real-time input, prediction display, and text generation for users.
- **Result:** Delivered a tool for fast, accurate sentiment insights and content creation, improving **customer feedback analysis** and **copywriting support**.

OCR Engine for Handwritten Text, (*Keras, CNN + RNN*)
- Built a robust **deep learning-based OCR pipeline** to accurately recognize **noisy handwritten text**, overcoming limitations of legacy models.
- Designed a hybrid architecture combining **Convolutional Neural Networks (CNN)** with **Recurrent Neural Networks (RNN)** and **CTC loss** for sequence prediction.
- Trained the model using the **EMNIST** and **IAM handwriting datasets**, incorporating **synthetic handwritten samples** for generalization.
- Implemented character-level accuracy evaluation, overlay visualizations, and **stress-tested** with rotated, noisy, and blurred text inputs.
- **Deployed the application via Streamlit** for real-time handwritten text recognition, supporting digitization in diverse workflows.

LLM-Powered RAG Chatbot, (*Google Gemini & LangChain*)
- **Built** an enterprise-grade **PDF-based Q&A chatbot** using a **Retrieval-Augmented Generation (RAG)** pipeline.
- **Embedded PDF documents** using **GoogleGenerativeAIEmbeddings** and stored them in a **FAISS vector database** for semantic retrieval.
- **Integrated LangChain** to handle context-aware interactions and manage **chat memory** for multi-turn conversations.
- **Developed** the **frontend in Streamlit**, allowing users to upload PDFs and query them interactively in real time.
- **Tested the system** through unit testing and simulated end-to-end conversations to ensure accuracy and robustness.
- **Result:** Reduced HR onboarding and training time by **80%**, replacing manual PDF lookups with **instant, reliable AI-powered responses**.

Publications

ML Zero To Hero, *Self-Published via Hugging Face*
2025
Designed and developed a comprehensive web-based learning platform to simplify **Machine Learning (ML)**, **Natural Language Processing (NLP)**, and **Computer Vision (OpenCV)** concepts through real-world, hands-on projects.
- **Covered foundational to advanced topics**: Supervised & Unsupervised Learning, model selection, NLP pipelines (tokenization, stemming, vectorization), and image classification workflows.
- Implemented **Image Augmentation techniques** (rotation, flipping, brightness/contrast adjustment, Gaussian noise) to improve model robustness in vision tasks.
- Delivered **step-by-step interactive tutorials** and **code walk-throughs** using **Streamlit**, enabling self-paced learning with real-time feedback and visualizations.
- Used **Python**, **Scikit-learn**, **NLTK**, **OpenCV**, **Matplotlib**, and **Pandas** to create engaging examples and practical modules for each concept.
- Aimed to support **students, early-career data scientists**, and **AI enthusiasts** with no prior experience, helping them transition from theory to application.