Bhuvnesh Sahu

Mail • +91-9407241555 • LinkedIn • GitHub

Profile

Background in Statistics and Mathematics. Interested in applying Machine Learning and Statistics to solve real world problems.Willing to leverage analytical and problem-solving skills to drive innovation and enhance decision making.

Education

Chennai Mathematical Institute, Chennai, India **2023 - 2025**
M.Sc. Data Science Grade: 8.31/10.00

Hindu College, University Of Delhi **2020 - 2023**
B.Sc. (Hons) Statistics Grade: 8.7/10.00

Work Experience

Data Science Intern, Algolabs **May - Jul 2024**
Real-Time Detection and Segmentation in Computer Vision **Link**

- Developed an end-to-end pipeline with **Streamlit App**, REST API and deployed with **Docker**.
- **Dataset:** Created custom dataset using **LabelMe**, enhancing data quality for model training.
- **Fine-tuned** advanced computer vision models (**YOLOv8, SAM, Detectron2**) on custom data to improve performance.
- Implemented a **comparative analysis** to evaluate model performance on KPIs.

Gen AI Intern, Coriolis **Jan - April 2025**

- Built a system-agnostic AI assistant that enables users to interact with any API-based system using natural language.
- Converted OpenAPI documentation into a structured API knowledge base, enabling intelligent understanding of system.
- Designed an LLM-chained pipeline to decompose user intents, select relevant APIs, and generate multi-step executable workflows.
- Reduced support team workload by automating repetitive task execution and minimizing manual API handling.
- Ensured robust task orchestration with optimized execution sequencing and dependency-aware API chaining.

Projects

End-to-End NLP-based application using LLMs and Gradio to analyze TV series content. **Link**

- **Dataset:** Web-Scrapped three datasets—Subtitles, Jutsu descriptions, and Transcripts.
- Utilized **BART-L-MNLI** for zero-shot classification to extract Themes/Genres of the series.
- Developed an interactive **character network** using **SpaCy, NetworkX**, and **PyVis** to depict character relationships.
- Fine-tuned **DistilBERT** to classify Jutsu types from Jutsu descriptions.
- Fine-tuned **LLaMA 3.1 (8B)** using **PEFT/QLoRA** to build a **Chatbot** with **BufferMemory** for interactive conversations with the main character.

Rag Based Multi Documents ChatBot with Voice **Link**

- **Developed** an interactive Q&A application to chat and converse with Documents; deployed using Hugging Face Spaces.
- Integrated **Open-Source Embeddings** with **ChromaDB** vector storage.
- Enabled voice interface using **Whisper** for audio input and Gemini TTS for audio output, flow built with LangChain.
- Employed LLaMA Scout LLM using **ChatGroq** for dynamic conversations with **ConversationBufferMemory**.

Music Genre Classification **Link**

- **Dataset :**10-genres with 100 samples per genre, each 30 seconds audio file.
- Extracted features from the audio, followed by **dimesionality reduction** using **PCA**.
- Explored various ML models like **LR, SVC, Decision Trees, Naive Bayes,KNN** with Ensebles methods **Random Forest**, **XGBoost** and **Deep ANN**, Employed **CNN** with mel spectogram images.
- Hybrid approach of **Functional API**, integrated **ANN** for audio features, **CNN** for mel spectrograms, merging with **FNN**.
- Did **Bayesian Hyperparameter tuning** with a **comparative analysis table** to evaluate performance of models.

COURSES & TECHNICAL SKILLS

- **Key Courses** – Machine Learning, Deep Learning, NLP, Descriptive & Inferential Statistics, Probability Theory, Linear Algebra, RDBMS & SQL, Time Series Analysis, Financial Statistics, Stochastic Processes, Econometrics, Reinforcement Learning.
- **Programming Languages** –C, Python, SQL, R.
- **Frameworks/Libraries** – NumPy, XGBOOST, Pandas, Matplotlib, Seaborn, Plotly, Scikit-Learn, NLTK, OpenCV, Tensorflow+Keras, PyTorch, LangGraph, LangSmith, LangChain, HuggingFace, Transformers.
- **Database Systems** – MySQL, PostGreSQL, MongoDB, Neo4j.
- **Miscellaneous** – MS Office, Git, Jupyter, Linux, Docker, kubernetes, REST API, Databricks, FastAPI, AWS, CrewAI.

ACHIEVEMENTS/AWARDS

- Department of Science & Technology (DST) INSPIRE Scholarship 2020.

ONLINE COURSES AND CERTIFICATIONS

- Finance and Quantitative Modeling for Analysts Specialization- Coursera
- Finance Risk Management Specialization - Coursera
- Machine Learning Specialization, Andrew NG - Coursera
- Advanced Learning algorithms, Andrew NG - Coursera
- AWS Cloud Solutions Architect Professional Certificate - Coursera