

PRAVEEN KUMAR BHARTI

Data Scientist

prvnbharti@gmail.com | +91 9113745535 | [in LinkedIn](#) | [Github](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Highly skilled Data Scientist with a Master's in Engineering. Proven ability to develop text pre-processing pipelines, optimize data for chatbot platforms, and ensure data quality. Experienced in Python, SQL, Data Viz., and TensorFlow. Built impactful projects on ML, DL, NLP and Generative AI. Certified by IABAC, NASSCOM, Datamites, and JainX.

WORK EXPERIENCE

AI/ML Engineer

Apr 2024 – Present

Drifko | Noida

- Developed and implemented **data scraping** and **preprocessing** scripts, ensuring high-quality datasets for **training**.
- Fine-tuned** the Llama 3 model with custom scripts, enhancing chatbot performance.
- Created and deployed chatbot code using the **LangChain** framework, crafting **effective prompts** for accurate responses.

Data Science Consultant, Intern

Jul 2023 – Feb 2024

Rubixe | Bengaluru

- Developed robust text pre-processing pipeline: Cleaned, normalized, and segmented customer chat data, handling diverse formats and dialects. Achieved 95% accuracy in data quality assessment.
- Built high-quality pre-processed tokens: Employed **NLP** techniques like **stemming**, **lemmatization**, and **NER** to enrich tokens and improve model training performance. Reduced training time by 30% and increased **F1-score by 15%**.
- Optimized data for Rubixe platform: Tailored processing pipeline for swift integration with Rubixe's **chatbot** builder, facilitating faster customization and **deployment**.

PROJECTS

Automated Machine Learning App: Streamlining Data Science Workflow

- Created an intuitive app leveraging **deep learning** and machine learning techniques for automated data science workflows.
- Utilized modular Python coding** to ensure scalability, maintenance, and efficiency in processing **datasets**, **cleaning**, **preprocessing**, and **visualizing data**, and constructing **predictive models** with **statistical accuracy** metrics.
- Implemented user experience enhancements such as sample datasets to showcase the app's capabilities, resulting in a seamless and engaging environment for users, ultimately advancing the field of data science.

Medical Chatbot

- The medical chatbot uses **the RAG architecture** to dynamically retrieve and incorporate relevant medical information, enhancing the accuracy and specificity of its responses.
- Developed a system to answer user queries efficiently and accurately, utilizing **Python**, **LangChain**, **Flask**, **Meta Llama2**, and **Pinecone**.
- Demonstrated expertise in **AI**, **vector search engines**, and integrating various tools for a cohesive solution.

Movie Recommender System

- Implemented Sentiment Analysis on state-of-the-art techniques to gain valuable insights into personalized customer Movie recommendation
- Improved text summarization by 30% using **RegEx** and **Postags**, resulting in more accurate and concise summaries
- Created a Flask-based web application for user interaction.

ATS Checker and Hand Writing Recognition

- Developed two projects using **Gemini-Pro** and **Gemini-Pro-Vision LLM** models: one for ATS compatibility checking of resumes and another for recognizing handwritten content from images.
- Utilized **Streamlit** to create web applications for seamless user interaction.
- Implemented advanced AI models for resume analysis and **image text recognition**.

EDUCATION

Masters in Environmental Engineering

Sep 2019 - Aug 2021

National Institute of Foundry and Forge Technology, Ranchi

Bachelor of Technology, Civil Engineering

Jul 2013 - Aug 2017

Gandhi Institute for Technology, Bhubaneswar

CERTIFICATIONS

Certified Data Scientist: **IABAC**

Certified Data Scientist: **NASSCOM**

Certified Data Scientist: **Datamites**

Data Scientist: **JainX**

SKILLS

Programming Language: Python, HTML

Python Framework: Pandas, Numpy, Seaborn, Matplotlib, Scikit-Learn, TensorFlow, Flask, Beautiful Soup

Databases: SQL, Faiss, Chroma DB, Pinecone

Data Analysis/Statistics: Data Analysis, Statistics, Predictive Modeling, Predictive Statistical Modeling, Data Mining, Data Analytics, Time Series Analysis

LLM: OpenAI, BERT, Facebook's BART, Hugging Face, Fine Tuning, llama2, Prompt Template

Data Visualization Tools: PowerBI, Tableau

Machine Learning/NLP: ML Algorithms NLP, Deep Learning (ANN, CNNs, Transformers, RNNs), Scikit-Learn, Transfer Learning, Data Preprocessing

GenAI: RAG Architecture, LangChain, Prompt Templating, Agent, Chaining, Embedding,

Deployment: GitHub CI/CD, Streamlit, PythonAnywhere, AWS

Soft Skills/Other: Adaptability, Creativity, Leadership, Project Management, Teamwork, Communication Skill, Time Management, Problem Solving, Collaborative, Innovative