AMRITESH PRAVEEN

amritesh.praveen@gmail.com +91 98865-29292 linkedin.com/in/amritesh-praveen github.com/Amritesh-878

ABOUT ME

AI and ML enthusiast with hands on experience in **neural networks**, **transformers**, and **retrieval augmented generation**. Passionate about developing scalable **AI systems** and applying **data centric learning** to impactful, real world domains.

KNOWLEDGE AND SKILLS

- **Languages:** Python, Java, SQL
- **Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Scikit learn, TensorFlow, PyTorch, LangChain.
- **Core AI Competencies:** RAG, NLP, Deep Learning, Computer Vision, Transformers
- **Visualization Tools:** Power BI, Tableau
- **Other Tools:** Git, GitHub

EXPERIENCE

AI & ML Intern – Impact Solutions Lab **Oct 2025 – May 2026 (Ongoing)**

- Supporting the development of the **AI for Social Impact (AI for SI)** DataHub, enabling data accessibility for research projects.
- Assisting in **AI tool development**, designing and optimizing models for real world social applications.
- Contributing to **data preprocessing** and **pipeline optimization** to ensure high quality datasets for AI experimentation.
- Collaborating with cross functional teams to enhance **AI model integration**, evaluation, and deployment workflows.

Machine Learning Intern – SixPhrase **Feb 2025 – May 2025**

- Preprocessed datasets by handling **missing values**, **scaling features**, and applying **encoding techniques** to ensure model readiness.
- Applied **dimensionality reduction** methods (**PCA**, **LDA**) to enhance performance and reduce training time by **25%**.
- Implemented **supervised learning models** (**Logistic Regression**, **Decision Trees**, **SVM**, **KNN**, etc.) achieving up to **87% accuracy**.
- Evaluated models with **cross validation** and metrics (**precision**, **recall**, **F1-score**, **RMSE**) to ensure reliability.
- Used **regularization** techniques (**L1**, **L2**, **Dropout**) to mitigate overfitting and improve generalization.

Data Analysis Intern – Miles Education **Apr 2024 – May 2024**

- Wrote **SQL queries** to clean and manage data, improving consistency and reducing redundancy by **30%**.
- Analyzed structured data using **Pandas**, **Matplotlib**, and **Seaborn**, delivering insights into trends and outliers.
- Conducted weekly **data analysis** and presented findings, enabling data driven decision making.
- Completed a **Hotel Bookings Project**, uncovering customer behavior patterns and seasonal trends using **Python** and **SQL**.

PUBLICATIONS

Railway Track Monitoring and Repairing Device **07/03/2025**

- Application No: 202541017469 A
- Title: Railway Track Monitoring and Repairing Device
- Abstract: A railway track monitoring and repair device comprises a trolley positioned on a railway track, equipped with **L-shaped telescopically operated rods**, acoustic sensors for defect detection, a **robotic link with a motorized cutter**, a GPS module, a **magnetic impedance (EMI) sensor**, and an **artificial intelligence based imaging unit** to monitor gravel distribution.

PROJECTS

Wildfire Smoke Detection (Ongoing)

- Implemented baseline **SmokeyNet architecture** combining **CNNs**, **LSTMs**, and **Vision Transformers (ViT)** for early wildfire smoke detection.
- Utilized the **FigLib dataset** of labeled images/videos to train and test the base model.
- Focus: exploring model optimization strategies to improve **accuracy**, **F1-score**, and robustness.

Mental Health RAG Chatbot **GitHub**

- Designed and deployed a **retrieval augmented chatbot** using **Groq LLM**, **LangChain**, and **Chroma**.
- Processed domain specific PDFs, generated embeddings with **Sentence Transformers**, and enabled **similarity search**.
- Built both **CLI** and **Chainlit UI** with environment based config.

Cognizant Hackathon – RAG Application **GitHub**

- Developed a competitive **retrieval augmented generation (RAG)** system during the Cognizant hackathon for document intelligence.
- Integrated **LangChain**, **vector databases**, and **embedding models** for semantic document retrieval and question answering.
- Enhanced RAG pipeline efficiency through better **prompt management**, **context chunking**, and **model response evaluation**.
- Focused on scalability and inference speed for real world application deployment.

YouTube Trending Video Analysis Dashboard **GitHub**

- Analyzed the **Kaggle Trending YouTube Dataset** to identify patterns across the US and India.
- Explored engagement metrics (**views**, **likes**, **comments**) using Python data workflows.
- Visualized insights with **Matplotlib**, **Seaborn**, **Plotly**, and **WordCloud**, with an interactive **Streamlit dashboard**.

CERTIFICATIONS

- Machine Learning Specialization – **Stanford** University, Coursera
- Natural Language Processing with Attention Models – **DeepLearning.AI**, Coursera
- Introduction to TensorFlow for AI, ML, and Deep Learning – **DeepLearning.AI**, Coursera
- Fundamentals of Reinforcement Learning – **University of Alberta**, Coursera
- Artificial Intelligence and Machine Learning Fundamentals – **IBM** SkillsBuild
- Machine Learning with Python – Udemy
- Generative AI Tools using Python – MSME **Government of India**
- Python Programming Essentials – **Rice University**, Coursera
- SQL & PostgreSQL – Udemy

EDUCATION

Jain University

B.Tech – Computer Science Engineering – Artificial Intelligence and Machine Learning **CGPA: 7.3**