Varun Prakash
Data Scientist
Hyderabad ● varunpatnabh@gmail.com ● +91-9872491129 ● www.linkedin.com/in/varun-p19

Professional Experience

Innominds **Hyderabad, IN**
Jr. Data Scientist (Full time) Aug 2022-Present
- Hands on Experience data extraction from PDFs using Tabula and database (Postgres) using MySQL for optimizing transformations.
- Collaborated with the data science team to design a streamline analysis of extracted clean and pre-processed data using Pandas, NumPy and SciPy libraries.
- Developed and implemented processes for automated data extraction and image generation using Matplotlib.
- Achieved 95% accuracy in ECG signal arrhythmia classification by custom model making and used deep learning tools like PyTorch, Keras and TensorFlow framework.
- Used AWS EC2 for the model building and also setup a system to check model performance on completely new data.

Domain: Healthcare, Language and Other: Python, MySQL, GPU, CNN, Hyperparameter Tuning, Linux, OpenCV, Scikit learn

Innominds **Hyderabad, IN**
Jr. Data Scientist June 2022-Aug 2022
- Utilized advanced NLP models like BERT and algorithms to generate high-quality synthetic data for NLP applications.
- Developed a POC for privacy-enhancing Masking Feature to safeguard data using Spacy, NLTK and Regex

Domain: Synthetic data, Other Technology and Language used: Python, Pandas, NumPy, NLTK, Hugging Face

Innominds **Chennai, IN**
Data Science Intern March 2022- June 2022
- Hand on experience in synthetic Audio data generation using libraries like Librosa, Pydub to address data insufficiency.
- Implemented a voice-changing feature for voice modulation.

Domain: Data Generation and Data Augmentation, Technology: Python, Pandas, NumPy, PyTorch, Hugging Face

IIITDM Kancheepuram **Chennai, IN**
Teaching Assistant Aug 2020- April 2022
- Created teaching materials and digital presentations to enhance learning experiences using software like Excel.
- Conducted classes and provided problem-solving guidance to undergraduate students using MS office and PowerPoint.

Technologies
- Programming Languages : Python, Java (Basic), MySQL
- Data Science: PySpark, SQL, Machine Learning Algorithms, Deep Learning, NLP, Feature Engineering, Generative AI
- Tools & Frameworks: TensorFlow, PyTorch, Flask, Tableau Visualization and Analytics, LLM, GCP, Docker, Flask, Streamlit
- Cloud Services: S3, AWS Sagemaker, AWS EC2, GCP containerization, Azure Cloud (Basic), Git, GitHub

Education

IIIT Chennai **Chennai, IN**
PG: M.Tech Specialisation in Smart Manufacturing(related to Data Science and Analytics) GPA-9.4 2020-2022
Relevant Coursework: Machine learning, Artificial Intelligence (AI), Big Data, IoT, Robotic Sensor

CGC College Engineering **Punjab, IN**
UG: B.Tech in Mechanical Engineering GPA -8 2014-2018

Other Personal Projects

Teaching Assistant Module
- Currently Working on making study course book easy by utilizing advance LLM models GPT, Gemini Pro using RAG Feature.
- Using Gemini Pro Vision also making to extract information using images.

Bike Rental Count
- Performed EDA on bike rental service data with storytelling and tested various model like Linear regression, SVM, Random Forest, XGBoost etc models achieving an 89% accuracy rate in predicting bike demand surges on test data.
- I've crafted a Machine Learning solution to predict bike rentals throughout the year. This project takes into account various factors such as weekdays, weekends, seasons, and current weather conditions to provide precise predictions.
- Converted this into full fledged end-to-end ML project with modular and optimized code. Later also deployed it on GCP where I used ChatGPT OpenAI to build frontend HTML page and running at https://bike-count-prediction-b2ftsooxcq-uc.a.run.app/

Food Delivery Time Prediction
- Analyzed an Indian food delivery dataset from Kaggle, conducted exploratory data analysis, and interpreted findings using various visualization techniques using Power BI in my master's.
- Engineered new features and developed a regression model achieving 90% accuracy in delivery time estimation.

Surge Prediction
- Conducted EDA on sigma cabs service data and built a classification model to predict fare price surges Category.

E commerce Traffic Analysis
- Performed advanced analysis using MySQL on an e-commerce website dataset, providing insightful interpretations.

Smart Air Monitoring System
- Leveraged IoT and cloud computing to develop a versatile air quality monitoring system.
- Created a Tableau-based dashboard for visualizing air quality parameters, enhancing data insights.