Dinesh Vennapoosa

+1 (317) 720 7226 – Portfolio – dinesh.vennapoosa@gmail.com – Linkedin – Github

EDUCATION

Indiana University Purdue University — **Indianapolis, Indiana, USA**
Master's in Applied Data Science — *Aug 2022 - Dec 2023*
National Institute of Technology, Mizoram — **Aizwal, Mizoram, India**
Bachelor's in Mechanical Engineering — *Aug 2016 - July 2020*

TECHNICAL SKILLS

Programming Languages: Python, Object Oriented Programming, R, HTML, CSS, JavaScript, SQL
Machine Learning: Linear Regression, Computer Vision, SVM, KNN, XGBoost, PCA, CNN, RNN, TensorFlow, PyTorch
Natural Language Processing: NLTK, SpaCy, BERT, Transformers, LLMs, LangChain
Statistical Methods: Hypothesis Testing, Time Series, Correlation
Cloud Computing & Big Data: MongoDB, Hadoop, Spark, AWS
Data Analysis: Statistics, Numpy, Pandas, SciPy, Scikit-learn, Matplotlib, Seaborn, PowerBI
Dev Tools: Linux, Git, CI/CD, Django, Docker

WORK EXPERIENCE

Data Scientist
Indiana University, Indiana, USA — *Aug 2022 - Dec 2023*

- Led a pioneering research project using **GPT-3.5** and other **LLMs** to rank cultural holidays across 58 countries, creating a novel dataset of 459 holidays for **NLP** cultural analytics.
- Enhanced data richness by integrating **ChatGPT** (using the "**text-davinci-003**" model) and Wikipedia descriptions, facilitating nuanced global cultural comparisons.
- Implemented a sophisticated few-shot Question and Answering framework to assess and compare the performance of three leading **LLMs , (GPT-3.5, Bloom 7b, Llama 13b)** showcasing **GPT-3.5**'s exceptional ability in **cultural nuance recognition**.

Machine Learning Engineer
Ugaan Labs Pvt Ltd, Karnataka, India — *Aug 2020 - June 2022*

- Implemented a state-of-the-art **object detection** model using the **YOLOv4 tiny** model to detect the type of KYC document with an accuracy of **92**%.
- Used pytesseract (**OCR**) to extract text features from each **KYC** document type and implemented **age progression** technique using **conditional Adversarial Autoencoder** (CAAE), to achieve a **15**% increase in match accuracy for live video-based customer verification
- Deployed project using **Django REST API** on **AWS**, incorporating **EC2, DynamoDB, Elastic Load Balancing, Auto Scaling, IAM, and CloudWatch** for efficient hosting, data storage, scalability, security, and performance monitoring.

Data Analyst Intern
Indian Institute of Technology Tirupati, Andhra Pradesh, India — *July 2019 - Nov 2019*

- Leveraged **machine learning**, specifically **Random Forest** and **SVM** algorithms, to enhance fluid dynamics and thermal system energy efficiency by **23**% and improve understanding of system operations through **CFD** data analysis.

Data Analyst Intern
Indian Institute of Technology Guwahati, Assam, India — *July 2018 - Dec 2018*

- Led a predictive maintenance project using machine learning (**Decision Trees**) and **vibration** analysis for **95**% accurate **fault detection**, in collaboration with mechanical engineers for enhanced system performance.

PROJECTS

- **Image Steganography**, Developed a deep learning-based Image Steganography model, enhancing security with superior encoder-decoder techniques and high SSIM scores, demonstrating advanced ML skills. GitHub
- **Sentiment Analysis of Stock Market Tweets using PySpark**, Led analysis on social media sentiment's effect on stock trends with PySpark and ML, achieving high accuracy with a Feed Forward Neural Network. GitHub

PUBLICATIONS

- By **artificial intelligence** algorithms and machine learning models to **diagnose cancer**. Materials Today: Proceedings Link