Likith Venkatesh Gowda Prathima

likithvp21@gmail.com | 857-746-8045 | Brookline, MA | LinkedIn

EXPERIENCE

Khoury Graduate Teaching Assistant, Khoury college of computer science *Jan 2024 – Present*

- Achieved a **20%** increase in student interest through innovative course restructuring, focusing on Large Language Models for real-world NLP applications.
- Implemented engaging activities leading to a **30%** rise in student engagement and participation levels in the NLP.

Data Scientist Intern, Lifesight *Jan 2022 - Jul 2022*

- Developed Data-Driven attribution models and performed predictive and behavioral analysis on mobility, audience and location data using **PySpark**, and cloud data platforms (**GCP**), resulting in a **24%** increase in revenue.
- Develop and manage pipelines (**ETL/ELT jobs**) using cloud data platforms and tools such as **Apache Airflow**, lowered Proof of Concept (POC) delivery time by over 60% by bridging the gap between the Sales/CSM teams and the Data Engineering team, providing ad-hoc data, **data enrichments**, and **analytical solutions**.
- Worked closely with the Data Engineering team to explore and utilize technologies/tools such as **Python**, **SQL**, **Jenkins**, **Hadoop Spark**, & cloud data platforms to design complex data modeling scenarios, automate processes, and perform transformations for optimal data engineering solutions thereby reducing the overall lead time by 40%.

ML Engineer Intern, ML Labs *Sept 2021 - Nov 2021*

- Contributed to a deep tech product using Python, TensorFlow, and OpenCV for real-time performance assessment of city services, enhancing service quality through AI-driven automation. Implemented a machine learning model for a US-based inventory management company, using Image Classification and Semantic Segmentation to minimize accidents and enhance workplace safety on the factory floor.

EDUCATION

Northeastern University *Sep 2022 – May 2024*
Master of Science in Computer Science GPA: 3.91
Related Course: Algorithm, Natural Language Processing, Machine Learning, Program Design Paradigm, Foundations of AI, AI in Human-Computer Interaction, Database Management Systems and Mobile Application Development.

Visvesvaraya Technological University (VTU) *Aug 2018 – Jul 2022*
Bachelors of Engineering in Information Science and Engineering GPA: 3.66
Related Course: Design and Analysis of Algorithms, Data Mining and Warehousing, Object Oriented Programming, Machine Learning & Artificial Intelligence, Operating Systems, Complex Analysis using Probability & Statistics, UNIX.

SKILLS

Programming Languages: Java, Python, R (Programming Language), C++, PostgreSQL, MySQL, UNIX, NoSQL.
Tools: GitHub, Tableau, Docker, Kubernetes, Anaconda, Hadoop, AWS S3, Apache Airflow, Apache Spark, Postman.
Python Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, Seaborn, SciPy, Plotly, OpenCV, PyTorch.
Certifications: Foundations of Data Science (PadhAI), Essentials of Data Science using R software (NPTEL), Business Analytics for Decision Making (NPTEL), Data Science for Engineers (NPTEL).
Leadership: Education Support Mentor (Make A Difference), Student Body Placement Head (CMRIT), Mentor (NEU).

PROJECTS

Automated Medical Recommendation System using ML & NLP for Cancer Treatment (Publication) |
Python, DenseNet121, Random Forest, Long Short-Term Memory, BERT, Shapley Additive Explanations (SHAP)

- Engineered a medical support system that analyses medical reports and radiological images to predict the occurrences of abnormalities in the body with an accuracy of around 78%, complemented by an interactive chatbot.
- Employed the DL framework, DENSENET, RF regressor and NLP methodologies for developing the various modules.

Toxicity Tune: Dual Model approach for generating and scoring toxic comment |NLTK, Tensorflow, PyTorch (**Git**)

- Built a dual-model approach utilizing BERT and miniGPT to identify toxic comments, achieving an impressive 85.44% accuracy through rigorous comprehensive hyperparameter tuning and optimization algorithms, batch sizes and vocabulary sizes to reduce model loss by 55%, effectively utilizing A100 GPUs for resource-efficient training.
- Implemented a robust k-fold cross-validation (k=14) methodology to rigorously evaluate 14 distinct models, achieving consistent loss values between 0.27 and 0.35, ensuring statistical reliability in assessing model performance.

Healthify Living | Java, Figma, Google Analytics, Google Firebase (**Git**)

- With the motive to address sharing of delicious recipes for people with dietary constraints, ideated and created a mobile application for people to share nutritional recipes on a community forum.
- With over 45% of the American population following a dietary regime, I devised a platform to share their experimentations and experiences to help individuals alike connect and help stay aligned with their diets and goals.

Agri-Commerce Platform | Kotlin, Dart, MS-Excel, Figma, Power BI, Google Firebase (**Git**)

- Our novel business model in the agricultural sector has not only predicted a 1.5 times higher crop yields and a substantial 37% reduction in expenses but has also revolutionized supply chain management, earning us a spot among the top 5 finalists in the Nanban Foundation Organic Food Supply Chain Hackathon in India.
- Through our application's farmer empowerment program, we aimed at connecting unskilled farmers with experts for modern techniques (which stands as 65% of India's total farming populous).

Stock Portfolio Management System | Java, SOLID

- Crafted a robust Java-based stock trading platform incorporating advanced mechanisms such as sharding and MVCC for managing user portfolios and executing stock transactions.
- Implemented transaction mechanisms and master-slave replication for data consistency and high availability. Employed query optimization techniques to enhance performance and scalability. Utilized storage engines optimized for efficient data storage and retrieval. Reduced development time by over 20% by employing various design patterns.

Paddy Disease Classification using ML | PyTorch, Gradient Accumulation, Convnext, Swinv2, Vit (**Git**)

- Leveraged the power of transfer learning by fine-tuning pretrained models on the Paddy disease dataset, leading to accurate disease classification. Implemented sophisticated ensemble models, with a primary focus on vit models, to significantly improve accuracy and ensure robust predictions.
- Implemented advanced test time augmentation strategies, augmenting model accuracy by evaluating flipped and rotated versions of test images, boosting the model's adaptability to diverse real-world scenarios.

Project Upload | Vue.js, Arduino Uno R3 (Git)

- Developed a digital Signage platform using Vue.js and Arduino Uno R3, enabling universities and offices to efficiently broadcast notifications, important information, and company/university videos, fostering transparency in operations.
- Achieved a 54% increase in employee/student support by providing a dynamic communication channel, enhancing engagement, and keeping stakeholders consistently informed about university/company updates.

Hybrid Approach to deriving context from long documents | PyTorch, Gradient Clipping, Weights&Biases (Git)

- Implemented a novel hybrid NLP model combining Shortformer's local attention efficiency with a Longformer-inspired modification, achieving a competitive cross-validated accuracy of 0.633.
- Outperformed the baseline Longformer transformer model with an enhanced modification, by a little over 1%, emphasizing the effectiveness of local attention integration.

Sleep Cycle estimation in wearable devices | LightGBM, CatBoost, XGBoost, Random Forest, KFold CV (Git)

- Implemented a robust feature engineering pipeline for accelerometer data, incorporating key metrics such as ENMO, anglez, and their variations over specified periods, optimizing for efficient sleep onset and wakeup event detection.
- Conducted a comprehensive analysis and comparison of LightGBM, CatBoost, XGBoost, and Random Forest models for sleep cycle-based time series logs through rigorous fine-tuning and feature engineering, evaluating their individual strengths and performance characteristics in the context of wrist-worn accelerometer data.

StoryGen | PyTorch, DPR, DPO, QLoRA, RAG-Token

- Directed a team in achieving a 30% increase in human evaluators' ratings for narrative coherence and engagement compared to baseline models. Spearheaded the implementation of novel techniques resulting in a 25% improvement in user engagement metrics, including reader retention and social media shares.
- Instrumented an innovative approach leading to a 40% reduction in coherence errors and a 20% increase in thematic consistency, as measured by quantitative metrics.