SHARANYA AKKENAPALLY

Montreal, QC (open to relocate) | +1(514)-569-8534 | sharanyaakkenapally@gmail.com | LinkedIn | GitHub

SUMMARY

- Data Scientist with a proven track record in data analysis, predictive modeling, and machine learning, skilled in leveraging pre-trained models like Large Language Models (LLMs) to deliver actionable insights and accelerate model development.
- Experienced in collaborating with cross-functional teams and developing advanced data solutions for complex business challenges, utilizing Python, R, and cloud platforms.

WORK EXPERIENCE

Cognizant **Hyderabad, India**
Data Science Intern *Mar 2022 - June 2022*

- Collaborated cross-functionally with data engineers and product managers, ensuring scalable data pipelines for seamless integration and utilizing agile development methodologies, promoting iterative progress.
- Developed end-to-end fraud detection systems using LSTM-based RNNs with attention mechanisms, improving detection accuracy by 15%.
- Designed scalable ML pipelines for large-scale dataset, automated workflows using Python and SQL including, reducing deployment time by 30%, conducting ongoing testing to ensure continuous optimization.
- Managed infrastructure with Docker, Kubernetes, and cloud-native tools, deploying and monitoring models at scale using CI/CD pipelines via Git.
- Implemented ETL processes and MLOps to streamline data flow, enhance scalability, and ensure real-time reporting for analytics.
- Created Power BI dashboards to communicate actionable insights to non-technical stakeholders and to monitor KPIs, enhancing data-driven decisions aligned with business needs.

Technocolabs Softwares Inc **Hyderabad, India**
Data Science Intern *May 2021 - Aug 2021*

- Enhanced model accuracy by 15% through data augmentation techniques and temporal transformations.
- Developed feature engineering workflows using AWS and Pyspark to efficiently manage large-scale datasets.
- Achieved 85% classification accuracy in time series forecasting for stock price prediction using Logistic Regression, SVM, and optimized models via Grid/Random Search.
- Proficient in SQL for querying relational databases, data extraction, data wrangling, and performing complex aggregations, and EDA for identifying patterns.
- Utilized Python and R for data preprocessing, data wrangling, data analytics, predictive modeling, feature engineering, and conducted statistical analysis and model evaluation to ensure accuracy and reliability.
- Built interactive dashboards with Tableau to support strategic decision-making and effectively communicate insights, and employed Bayesian inference models to enhance decision-making efficiency by 30%.

The Sparks Foundation **Remote, India**
Data Science and Analytics Intern *Sep 2020 - Dec 2020*

- Processed text data using Natural Language Processing (NLP) techniques such as tokenization and word embeddings to generate high-quality feature vectors, enabling accurate classification.
- Performed descriptive analytics by extracting, exploring, and analyzing large datasets from multiple sources.
- Designed and implemented a spam detection model with Support Vector Machine (SVM), achieving a 92% accuracy rate by fine-tuning hyperparameters through Grid Search for optimal performance.
- Visualized performance metrics and model outcomes using Matplotlib and Seaborn, presenting clear insights into model effectiveness and identifying key areas for further optimization in spam detection.

SmartBridge Pvt Ltd **Hyderabad, India**
AI Developer Intern *May 2020 - Aug 2020*

- Developed a CNN-based algorithm for real-time skin disease detection using supervised models for classification, utilizing OpenCV for image preprocessing and feature extraction in healthcare data analysis.
- Engineered CNN architecture with Max-Pooling, Conv2D, Batch Normalization, and Dropout to enhance feature capture and prevent overfitting.
- Enhanced the model by integrating Transfer Learning, fine-tuning a ResNet-50 architecture, reducing training time by 22% and boosting model accuracy by 4%.
- Deployed the CNN model using Flask, implementing RESTful APIs for real-time image analysis, and asynchronous processing to manage high user traffic, resulting in a scalable and responsive web application.

EDUCATION

Master of Applied Computer Science - *Concordia University, Canada* *Sep 2022 - May 2024*
B. Tech in Information Technology - *Sreenidhi Institute of Science and Technology, India* *Aug 2018 - Jul 2022*

TOP SKILLS

- **Programming languages & Databases:** C, Data Structures, C++, Python, R, Java, MySQL, NoSQL
- **Web Technologies**: HTML, CSS, JavaScript

- **Machine Learning:** PyTorch, Keras, TensorFlow, Hugging Face, Deep learning, Generative AI, NLP, Transformers, LLM Fine Tuning, MLOps, Sci-kit learn, EDA, SpeechBrain, OpenCV, CUDA
- **Data Visualization & Analysis:** Tableau, Power BI, MS Excel, R-Studio, Matplotlib, Seaborn, Scipy, Plotly, Pandas, Numpy
- **Cloud & DevOps:** Azure ML, AWS (Sagemaker), PySpark, HQL, HBase, Oracle cloud (OCI), Docker, Kubernetes, CI/CD pipelines, Git, Flask

PROJECTS

Transformer based TTS System for Speech Synthesis

- Developed a Transformer-based Text-to-Speech (TTS) model with the LJSpeech dataset, improving speech quality and performance using the SpeechBrain framework.
- Accelerated training using CUDA, cuDNN, and GPU optimizations, achieving 2.5x faster speeds than Tacotron2. Experience in distributed training and optimization of language models to solve resource allocation to enhance model efficiency.
- Applied scaled positional encodings and teacher forcing, reducing Mel Error to 8.27e-02 and lowering Stop Error by 10%.
- Enhanced efficiency with dynamic batching and Noam Scheduler for more natural and fluid speech synthesis.

Dialogue Summarization: Fine Tuning LLM using Prompt Engineering and PEFT

- Explored and Fine-tuned FLAN-T5 Language Model from Hugging Face for dialogue summarization using prompt engineering and zero/few-shot learning to improve adaptability for LLM-based solution.
- Applied PEFT (LoRA), updating just 2-12% of parameters, reducing model size for fine-tuning and optimizing the LLM.
- Improved ROUGE-1 by 17.47% and ROUGE-2 by 8.73% over human baseline summaries, ensuring efficient compression and accuracy.

Automated Retail Product Classification using CNN

- Designed CNN architectures (ResNet-18, GoogleNet, AlexNet) to classify grocery products, addressing class imbalance with weighting and oversampling.
- Performed data augmentation, model scaling and utilized grid search for hyperparameter optimization, fine-tuning learning rates, batch sizes, and dropout rates with focus on complexity analysis and enhancing performance metrics.
- Leveraged transfer learning by fine-tuning ResNet-18 with pre-trained ImageNet weights, resulting in an 8% increase in classification accuracy.

A Diagnostic Assistant for Prediction of Post-Covid Skin Allergies

- Led research on the predictive modeling of post-Covid skin allergies using image processing, focusing on forecasting disease severity.
- Employed a combination of ensemble methods with Neural Networks, including Gradient Boosting, AdaBoost, and XGBoost, to improve model accuracy and robustness.
- Achieved 89% classification accuracy, with XGBoost outperforming other models, offering a more reliable assessment of disease severity, thus significantly improving the model's and predictive capabilities.

CERTIFICATIONS AND ACHIEVEMENTS

- Oracle cloud infrastructure 2024 generative AI certified professional
- Machine Learning certification - Stanford University
- The Data Scientist Toolbox certification, Johns Hopkins University
- AWS Academy- Cloud Foundations course - Amazon
- Silver Award at Ennovate - The International Innovation Show, Poland, 2021

ACTIVITIES AND SOCIETIES

- Volunteer Team Leader, ConUHacks VIII, Concordia University — Jan 2024
- The Techvision club- President and Executive — Aug 2020 - Sep 2022
- AIESEC in Hyderabad - Junior Marketing Manager, Junior Manager(Incoming Social Sector) — Aug 2021- Aug 2022