Sai Manoj Akondi

manoj-akondi | LinkedIn | GitHub |

Email: manojakondi25@gmail.com
Mobile: 6303364524

Roles and Responsibilities

Data Scientist

Decimal Point Analytics

- Processing financial data and developing ML/DL models on the processed data
- Translate business requirements into ML Problem statements
- Primarily focused on areas such as NLP, CV and RL
- Improve AI/Machine learning models by conducting experiments and analyzing results to address business requirements.
- Coordinate with various teams to solve problems efficiently and effectively.

Senior Analyst - Machine Learning

Tiger Analytics

- Deploying and Monitoring Machine learning models
- Developing and enhancing algorithms and models to solve business problem
- Engage with clients to understand their business context
- Collaborate with a team of data scientists and engineers to embed AI and analytics into the business decision processes.

Experience

Senior Analyst - MLE

Tiger Analytics *Hyderbad, Telangana*

* **Bug Predictability:** Deployed a model which predicts chance of getting an out of sprint bug from a story. Used Jenkins software to schedule a cron job and run predictions on new stories created everyday.

Data Scientist

Decimal Point Analytics *Mumbai, Maharashtra*

* **CSV Agent:** Developed a CSV agent utilizing large language models to facilitate natural language interaction with Excel files. Designed the agent to understand user queries and generate goals based on their input. The agent writes code to execute actions corresponding to the defined goals, allowing seamless interaction with CSV data.
* **Dolat Summarization** Contributed to the fine-tuning of a large language model on consumer GPUs to generate summarizations of earnings calls. Leveraged advanced techniques, such as LoRA, to optimize the training process for large models on consumer-grade hardware. Collaborated with a team to work with state-of-the-art models like Pythia, MPT, GPT-Neo-X, and T5 Flan, achieving significant progress in the project.
* **Semantic Data type Detection:** Designed and developed a high-accuracy transformer model for automated classification of column data types, leveraging a dataset of 800K data points from both online and internal sources. Achieved an impressive 94% accuracy on the test set, demonstrating the model's robustness and effectiveness. In addition, trained a LSTM model with an embedded layer to learn character embeddings, which achieved a solid accuracy of 85%. Conducted rigorous experimentation and analysis to optimize model performance and improve accuracy.
* **ESG Classifier:** Worked on a project with the objective of developing a text classifier capable of accurately categorizing text into three distinct genres: environmental, social, and governance. Incorporated the GradCam technique to visualize the specific phrases that the BERT model focused on during the classification process, enhancing interpretability and transparency of the final output.
* **Tiger automation:** Worked on a project to automate the appraisal process by developing a robust data processing pipeline using PySpark. Leveraged Snowflake database as the primary data source and performed all calculations and transformations using PySpark's distributed processing capabilities.
* **PDF2Excel:** Worked on development of a cutting-edge ML pipeline for converting PDF tables to Excel, utilizing OCR and advanced layout detection models. Leveraged DiT (document image transformer) to accurately detect the layout of PDFs and extract tabular data.

* **Auto ML** Working on an end-to-end pipeline that automates diverse ML processes, including preprocessing, feature selection, model selection, and hyperparameter tuning. Developed with a vision to democratize model training, the product aims to empower individuals with non-technical backgrounds, making it effortless and accessible for them to engage in machine learning.

Software Development Intern at Decimal Point Analytics May 2021- July 2021
Decimal Point Analytics *Mumbai, Maharashtra*

* **Blockchain based Chat and Bid Application:** Designed and developed a cutting-edge blockchain-based chat and bidding application, leveraging the power of Hedera Hashgraph. Utilized Hedera Hashgraph's JavaScript API to build a secure and robust application that enabled seamless messaging and bidding transactions.

Technical Skills

Languages	:	C, CPP, Python, SQL
Cloud Technologies	:	AWS S3, AWS EC2, AWS Sagemaker
Frameworks	:	TensorFlow, Keras, Pytorch, JAX, Pyspark, Dask
Libraries	:	Scikit-Learn, numpy, cv2, xgboost, Numpy, Pandas
Databases	:	Snowflake
Technologies	:	CV, NLP, RL, ML, DL
Dev Tools	:	Git, Docker, Jenkins, Visual Studio Code
Other Skills	:	Linux, Data Structures

Education

National Institute of Technology, Calicut Calicut, Kerala
B.Tech Mechanical Engineering, CGPA: 8.66/10 *Aug 2018 – May 2022*

Publications

* Bijoy et al "Deep Cleaner - A Few Shot Image Dataset Cleaner using Supervised Contrastive Learning" in IEEE Open Access Journal, 2023 [paper]
* Bijoy et al "Cervix type detection using a self-supervision boosted object detection technique" in IMA, 2022 [paper]
* P. Limna Das et al "Early Detection of COVID-19 from CT Scans Using Deep Learning Techniques", Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 736. Springer, Singapore[paper]

Projects

* **Automation of Cleaning cervical data using deep learning techniques:** Developed a supervised contrastive model to filter outliers in a cervical image dataset resulted in superior performance when compared to human cleaning. These impressive findings were published in the prestigious IEEE Access journal.
* **EfficientCenterDet: A novel Self supervision boosted RoI proposal network for cervix type detection[code]:** A fully automated self-supervised pipeline has been developed for the detection of cervical cancer. This impressive feat was achieved by leveraging a novel object detector, which drew inspiration from both the efficientdet architecture and centrenet loss. These impressive findings were published in the prestigious International Journal of Imaging Systems and Technology.
* **Covid-19 detection from CT scans[code]:** I successfully designed and implemented an advanced EfficientNet architecture that accurately predicts Covid-19 infection through CT scans. To ensure optimal performance, I employed a BCD U-net for efficient segmentation of the region of interest. These findings were communicate to a conference.

* **Cassava Leaf disease classification[code]:** I undertook a challenging Kaggle competition by implementing a variety of advanced models, including Vision Transformer, EfficientNets, and ResNets, all trained using Bi-Tempered Loss. To achieve even greater accuracy, I utilized an ensemble of these models in conjunction with Test Time Augmentation (TTA).
* **Stock Market prediction with tabnet[code]:**Successfully trained tabnet architecture, (original developed by google AI cloud) for regressing over a complex tabular data. Along with tabnet, I also trained gradient boosted tree algorithms like xgboost, catboost. Also, trained RNN for puts call ratio from historical data. I leveraged self supervised methods to handle missing values and ensure the highest level of model accuracy.
* **Tweet Sentiment Extraction[code]:** Worked on a project to extract key phrases given the sentiment from tweets, utilizing multiple advanced transformers, including XLNet, RoBERTa, and alBERT. To achieve even greater performance, I implemented an ensemble of these models, further enhancing my model's predictive power
* **Human Activity Recognition using 2D pose[code]:** For this project, I tackled the challenging task of detecting human activities from video data. To achieve this, I utilized the powerful pose recognition model, Posenet, as a starting point, and built a custom Convlstm head on top of it. This model was then fine-tuned using a data input of 20 frames at a time, allowing for greater accuracy in activity detection.
* **Multi task learning for self driving cars:** Developed a single neural network that can perform object detection, segmentation and depth perception using IDD dataset.

Certifications

* **Coursera:** Convolutional Neural Networks in TensorFlow, Convolutional Neural Networks, Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization, Introduction to TensorFlow for Artificial Intelligence, Machine Learning and Deep Learning, Natural Language Processing in TensorFlow, Neural Networks and Deep Learning, Structuring Machine Learning Projects, Introduction to Structured Query Language (SQL).
* **Electives:** Control Systems