

Venkata Phani Siva Sankar Emany

AI/ML ENGINEER · SENIOR DATA SCIENTIST · RESEARCH ENGINEER

📞 (+91) 9121016133 | ✉️ sankaremany@gmail.com | 🏠 <https://blog.eternalblue.co.in/> | 📺 Siva227 | 🌐 siva-sankar-evp

Summary

Recent Master's graduate in Computer Science with a focus on Probabilistic Modeling, Deep Reinforcement Learning, and Generative AI. Proven track record with 4.5 years of diverse data science experience as a Research Engineer at Byjus and a Senior Data Scientist at Quadratyx. Specialized in Natural Language Processing, Computer Vision, MLOps, and building Machine Learning tools for Healthcare, Banking and Finance, and Education sectors. Adept at programming in Python, C/C++, and Java, with expertise in Scikit-Learn, PyTorch, Tensorflow, etc. Passionate about Generative AI and Explainable AI, aiming to increase human trust in AI/ML models. Seeking opportunities to apply skills and experience to build innovative and impactful AI/ML solutions for real-world problems.

Education

CWRU(Case Western Reserve University)

M.S IN COMPUTER SCIENCE

Cleveland, OH, U.S.A

Jan. 2023 - Dec. 2023

INSOFE(International School of Engineering)

POSTGRADUATE SPECIALIZATION IN BIG DATA ANALYTICS AND OPTIMIZATION

Hyderabad, Telangana, India

Feb. 2018 - Aug. 2018

DA-IICT(Dhirubhai Ambani Institute of Information and Communication Technology)

B.TECH IN INFORMATION AND COMMUNICATION TECHNOLOGY

Gandhi Nagar, Gujarat, India

Aug. 2010 - Jul. 2015

Work Experience

Byjus

RESEARCH ENGINEER

Bangalore, India

Jan. 2022 - Nov. 2022

- Trained character and word-level OCR models for English and Spanish to enable hands-on writing exercises for pre-school kids on a phygital platform called Osmo. The models were designed to run on Android and iOS devices
- Trained OCR models to transcribe hand-written answer sheets from middle and high school students attending Byjus Tuition Centers.
- Explored and implemented word-OCR models based on CNN, LSTM and Transformer architectures using CTC loss

Quadratyx

SENIOR DATA SCIENTIST

Hyderabad, India

Mar. 2020 - Dec. 2021

- Built a service for automating prescription-handling in pharmacies owned by a Canadian healthcare services provider. Involved in designing and implementing APIs for prescription recognition, OCR, information retrieval and automated data entry using Django, Apache Airflow, Tensorflow and Scikit-learn.
- Explored and created multiple demonstrations about using DataBricks, Azure Data Factory, Azure Purview, Apache Nifi, ELK stack, Apache Superset etc.
- Helped devise AI driven use cases to create a unified platform of services for a Canadian healthcare services provider. Defined data requirements and analysed data availability for implementing said use cases.
- Built pipelines for a Colombian asset management company to automate the ingestion of data, creation of reports and projection of recovery curve patterns using Multi-Output Regression Models and Time Series Analysis in order to aid the decision makers in purchase of Non-Performing loans and reduce the data science effort spent in data preparation
- Designed the material and conducted a week long intensive training session with the data science and data engineering teams of a Colombian energy distribution company to introduce them to Big Data and teach them applications of PySpark in a variety of use cases

Quadratyx

DATA SCIENTIST

Hyderabad, India

May. 2018 - Feb. 2020

- Built REST APIs using Django for a large Indian private bank to extract named financial entities and classify tables into multiple relevant categories based on text extracted from scanned financial documents
- Designed and implemented a custom data genome to integrate electronic health records and dispensation records for a Colombian pharmaceutical organization using Hadoop
- Extracted key medical information from raw Spanish text and performed context analysis on extracted elements using PySpark
- Extracted rules indicating multi-drug usage, medical condition and demographic information associated with adverse medical outcomes using the above genome using PySpark

Academic Projects

Explaining Link Prediction using Ensemble Learning

COURSEWORK FOR MACHINE LEARNING ON GRAPHS

- Built a scikit-learn based toolkit for Link Prediction on graph/network data
- Exposed scikit-learn like APIs for networkX and graph-tool based link scoring techniques
- Implemented DeepWalk and Node2Vec node embedding algorithms
- Implemented Cannistraci-Hebb link scoring algorithms
- Analysed the ensemble models using SHAP explainability scores

Detecting Artificial Text

COURSEWORK FOR DEEP GENERATIVE MODELS

- Identified data sources to discriminate between human text and text generated by Large Language Models
- Created stylometric features using Spacy, NLTK, empath and neuralcoref
- Trained classifiers using XGBoost, LightGBM etc and used SHAP to analyse the predictions
- Trained classifiers using RoBERTa backbones
- Explored the transferability of the above models from text from one domain to the other

Miscellaneous Explorations

COURSEWORK FOR PROBABILISTIC GRAPHICAL MODELS AND REINFORCEMENT LEARNING

- Explored various Offline Reinforcement Learning algorithms and implemented Implicit-Q learning
- Explored the applications of Variational Autoencoders in semi-supervised learning
- Explored and implemented various mixtures of Linear and Logistic Regression, and Density Networks using Expectation-Maximization algorithm

Skills

Programming	Python, JAVA, C, C++, LaTeX
ML tools	Scikit-learn, Pandas, Numpy, PySpark, XGBoost, LightGBM, Catboost
Deep Learning	Tensorflow, PyTorch, Keras, Pyro, Tensorflow-Probability
NLP	Spacy, NLTK, Huggingface Transformers
Computer Vision	OpenCV
Time Series	Statsmodels, sktime, Prophet, ts-learn, pmdarima
MLOps	AWS, Docker, Azure, Django, Flask, FastAPI
Dashboarding	Matplotlib, Seaborn, Streamlit, Tableau, Superset, Kibana, Plotly
Data Stores	HDFS, Hive, MongoDB, Oracle, PostgreSQL, MS-SQL Server
Languages	English, Telugu, Hindi, German