

Kumaramangalam Vangavolu

kumarvrsec@gmail.com

[linkedin.com/in/kumarvangavolu](https://www.linkedin.com/in/kumarvangavolu)

<https://kumarama3.github.io/portfolio/>

Summary

Senior Data Scientist with 7.5 years of experience in building efficient Machine Learning, CV and NLP models from scratch to deployment, across multiple domains such as Insurance, Healthcare, Finance, Banking, Retail and Security etc.

Experience



Senior Software Engineer

GE HealthCare

Dec 2021 - Jan 2023 (1 year 2 months)

- Led the Research and Development of Clinical Language Processing Engine using Nashbio and NCCS data to retrieve significant clinical information from clinical text reports, using Spark NLP library in Python
- Developed and fine-tuned BiLSTM, Clinical BERT based Named Entity Recognition models on 9000 Oncological clinical reports, achieving a 90% F1 score across 42 entities using clinical embeddings.
- Performed iterative Text mining on annotated clinical datasets, and subsequently built models for Relation and Assertion Extraction, achieving an accuracy rate of 85%.
- Built a data de-identification model with 90% accuracy, to de-identify Personal Identifiable Information (PII) and Personal Health Information (PHI) for enhanced privacy and HIPAA compliance.
- Conducted code reviews, deployed the engine with Docker, and optimized model response times while addressing security vulnerabilities.
- Implemented MLOps practices for the NLP engine development, incorporating the MLFlow library and custom-built components for seamless experiment tracking, drift measurement and comprehensive model management
- Assumed responsibility for Text Summarization and Generation projects, using Large Language Models (GPT-2, GPT-3) and customized techniques
- Collaborating with domain and testing teams to improve the business logic and quality of the Engine pipeline



Information Technology Analyst

Tata Consultancy Services

Jan 2020 - Dec 2021 (2 years)

Natural Language Processing, Computer Vision, Machine Learning, Trustworthy AI - Explainable AI, Robust AI, Adversarial Machine Learning



System Engineer

Tata Consultancy Services

May 2017 - Jan 2020 (2 years 9 months)

NLP, Machine Learning, Time Series Forecasting



Assistant System Engineer

Tata Consultancy Services

May 2015 - May 2017 (2 years 1 month)

Education



University at Buffalo

Master's degree, Data Science



Velagapudi Ramakrishna Siddhartha Engineering College

B.Tech, Electronics and Communications Engineering

Licenses & Certifications



Applied Data Science with Python - Coursera



Hands On Essentials - Data Warehouse - Snowflake



Spark NLP for Healthcare Data Scientist - John Snow Labs



Spark NLP Data Scientist - John Snow Labs

Skills

PyTorch • Hadoop • Generative AI • Natural Language Processing (NLP) • Cloud Computing • Large Language Models (LLM) • Deep Learning • Big Data • Data Warehousing • MLOps