Ajayeswar Reddy Peddyreddy

+1 (341)-732-8139 ♦ ajayeswarb2a@gmail.com ♦ linkedin.com/in/ajayeswarreddy/ ♦ GitHub ♦ Blog

Full-stack data scientist with 3 years of experience and a keen interest in building end-to-end scalable ML products, from development to deployment, committed to designing and developing scalable end-to-end data pipelines to enable data-driven decision making.

Education

Master of Science in Data Science – The University of San Francisco, USA ***Jul 2022 – Jun 2023***

Courses: Machine Learning, Deep Learning, Regression Analysis, Time Series, A/B Testing, Data Structures & Algorithms, Statistics, Exploratory Data Analysis, Distributed Data Systems (Apache Spark), Relational Databases (SQL), Data Acquisition, MLOps

B.E Electrical and Electronics – BITS PILANI, India ***Jul 2016 – May 2020***

Courses: Neural Networks and Fuzzy logic, Probability and Statistics, Linear Algebra, Differential Calculus, Operating Systems, Object Oriented Programming, Operations Research

Professional Experience

Amazon Machine Learning Solution Lab & USF data institute **San Francisco, USA**

Machine Learning Engineer Intern ***Oct 2022 – Present***

- Developing new algorithms and training pipelines to accelerate large-scale neural network training, optimizing model performance and accuracy, mentored by Amazon Machine Learning Solutions Lab and USF data institute.
- Implemented a sampling strategy for language models, resulting in faster convergence than traditional SGD and improved model efficiency.

GEP Worldwide **Hyderabad, India**

Data Scientist ***Jul 2020 – Jun 2022***

Work involved handling end-to-end ML and engineering tasks for the product **OCR for invoices and credit memos.**

- Designed and implemented rule-based NLP pipeline for invoice processing, achieving up to 100% accuracy for select attributes and an overall accuracy range of 70-100%. Expanded language support to **15** additional languages, resulting in improved global usability and expanded market reach
- Revamped the architecture of the system by migrating the API from **flask** to **FastAPI** and transitioned from synchronous to **asynchronous** processing, which resulted in a significant improvement in the reliability and cut down the error rate for processing complex invoices by more than **90%.**
- Successfully decoupled a monolithic API into a **microservices** architecture, resulting in seven independent services with improved scalability, fault tolerance, and agility.
- Engineered the product using Docker and Azure Kubernetes serving ~10000 requests per day.

Data Scientist Intern ***Jul 2019 – Dec 2019***

- ***Reverse Image Search:*** Used pre-trained Res-net 50 architecture to encode the images and used modified K-nearest neighbors to classify the images and recommend the appropriate item.
- Reduced false-positive rate from 25% to 5% using classification models and utilized YOLO for accurate object detection, to identify multiple objects from one image and recommend accordingly.

Projects (selected)

- ***Job recommendation*** – Developed an end-to-end distributed deep learning pipeline to match jobs with resumes. GitHub
  - Scraped and stored job postings from multiple sources to GCP (Google cloud), preprocessed data using PySpark on Databricks, and stored collections in MongoDB (**ELT pipeline**). Utilized Airflow for scheduled scraping.
  - Generated embeddings for job postings and resumes using Word2Vec and **Sentence Transformer models** (large and small).
  - Performed vector similarity search between job and resume embeddings using Pinecone achieving perfect accuracy.
- ***ML from scratch*** – Implemented Decision Trees, Random Forest, Gradient Boosting, Adagrad, k-means clustering algorithms.
- ***Recommendation systems*** – Developed Matrix Factorization recommendation system using PyTorch to predict movie ratings.
- ***Search Engine*** – Implemented using Object Oriented hash table to retrieve matched documents from a corpus in constant time.
- ***Mixed code sarcasm detection*** – Developed and implemented traditional ML algorithms (Naive Bayes, SVM, logistic regression, random forest) and a bi-directional LSTM attention-based model to detect sarcasm in Hindi-English code-mixed tweets, achieving 92% accuracy. GitHub

Technical Skills

Python (NumPy, Pandas, Matplotlib, Scikit-learn, NLTK, Tesseract, regex), TensorFlow, PyTorch, Databases, SQL(Postgres), NoSQL (MongoDB), Apache Spark, Airflow, data pipelines (ELT, ETL), Version control (Git, GitHub), Cloud (Azure, GCP), Databricks, Fast API, Flask, Docker, MLFlow, Weights and biases, DVC.