

Rajasvi Vinayak Sharma

📞 +1(858)319-9642 | ✉️ rajasvi777@gmail.com | 🌐 linkedin.com/in/rajasvi | 🔄 rajasvi | 🌐 rajasvi.github.io

EXPERIENCE

Moveworks (acquired by ServiceNow)

Jan 2025 – Present

Senior Machine Learning Engineer | Enterprise Search Relevance Team

Mountain View, CA

- Implemented an **LLM-based relevance annotator** for a **search ranking task** with **batched listwise** labeling, yielding a **10× speedup** and enabling rapid creation of large training datasets with minimal human effort.
- Reduced human labeling effort by **~60%** by building a **hybrid annotation workflow** that uses LLM pre-annotation and escalates only ambiguous cases to human annotators for high-confidence eval sets.
- Built a summary-quality evaluation pipeline for **search-result summaries** using LLM-as-judge and reference-less metrics (summary relevance, groundedness, citation validity), improving summary relevance from **60% → 90%**.
- Shipped a user-facing **keyword-highlighting** feature for search results by extracting document entities and query tokens to surface relevant terms in the UI.
- Created a **synthetic query generation** pipeline to augment feature-specific queries (e.g., temporal, author) and add training signal for search ranking models.

Signos

Apr 2023 – Dec 2024

Machine Learning Engineer

Remote, CA

- Designed deep learning model for post-meal glucose spike prediction, dual-purposed for exercise recommendations to effectively minimize post-meal glucose spikes by suggesting exercises to **>10,000 users** & help them lose weight.
- Built Retrieval Augmented Generation (RAG) pipelines using LangChain for feature extraction from **> 1M free text meal logs** to improve post-meal glucose spike prediction model accuracy.
- Achieved a **12%** boost in user meal logging engagement by creating a real-time anomaly detection algorithm for identifying point of interest events & rapid glucose level spikes.
- Developed a **meal recommendation engine** using a hybrid collaborative & content filtering model based on features from multiple signals like continuous glucose data, meal/exercise/sleep logs, to help **>10K users** make healthy food choices.

Nvidia

Jun 2022 – Sep 2022

Data Scientist Intern

Santa Clara, CA

- Developed and deployed an end-to-end Time-series Anomaly Detection tool using Z-Score thresholding and Kubeflow pipelines to alert about malicious activities in 1000+ categories across 10M+ gaming sessions, reducing the alert response time from a few months to just 1 week.

Goldman Sachs

Jun 2018 – Aug 2021

Data Scientist, Analyst

Bengaluru, IN

- Fine-tuned BERT, RoBERTa, and DistilBERT Transformer models using Hugging Face to extract named contact-related entities from over 8 million emails per day, achieving an impressive 85.7% accuracy rate.
- Built XGBoost & LGBM based models for classifying jurisdiction violations among Bloomberg trader conversations (>6M daily) by leveraging semantic & temporal information extracted from the data as features, achieving an impressive 78% precision.

EDUCATION

University of California - San Diego | California, US

Sep 2021 – Mar 2023

M.S. in Electrical & Computer Engineering (Major: Machine Learning & Data Science)

GPA: 3.91 / 4.0

- Coursework:** Recommender Systems & Web Mining, Deep Learning for Natural Language Understanding, Statistical Natural Language Processing (NLP), Big Data Mining & Spark Analytics, Statistical Machine Learning (ML), Linear Algebra

Indian Institute of Technology (Banaras Hindu University) Varanasi | India

Jul 2014 – May 2018

B.Tech. in Electronics Engineering

GPA: 8.81 / 10

PROJECTS

Clickbait Spoiler Generation (SemEval'23 Shared task) [code] [report] | PyTorch, Hugging Face

Sep 2022 – Dec 2022

- Created a 2-stage process to generate Clickbait spoilers using text classification, question answering, and ranking models. Utilized DeBERTa, RoBERTa, and DistilBERT Transformer models, fine-tuning them to achieve optimal results. Evaluated using BLEU-4, exact match, and F-1.
- Contributed by surpassing benchmark results via techniques from organizers' paper, achieving 36.4 BLEU-4.

SKILLS

Languages: Python, SQL

ML/LLM: LLMs (Large Language Models), NLP, Information Retrieval (search/ranking), recommendation system, LLM Evaluation

Frameworks: PyTorch, Hugging Face, scikit-learn

Data/MLOps/Cloud: Spark (PySpark/SparkSQL), Kafka, Kubeflow, MLflow, Databricks, W&B, AWS, GCP