Rajasvi Vinayak Sharma

+1(858)319-9642 | rvsharma@ucsd.edu | linkedin.com/in/rajasvi | rajasvi | rajasvi.github.io

EDUCATION

University of California - San Diego | California, US Sep 2021 – Mar 2023
M.S. in Electrical & Computer Engg. ***(Major: Machine Learning & Data Science)*** GPA: **3.88 / 4.0**

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

Indian Institute of Technology (Banaras Hindu University), Varanasi | IN Jul 2014 – May 2018
B.Tech. in Electronics Engineering GPA: **8.81 / 10**

SKILLS

Languages : Python, SQL, R, Java, C++
Big Data : Apache Flink, PySpark, SparkSQL, Hadoop, Map Reduce, Kafka, Redis, Hbase, HDFS, Yarn
ML Frameworks/Libraries : PyTorch, TensorFlow, Pandas, Numpy, scikit-learn, CausalML, DoWhy
MLOps/Dev Tools : Kubeflow, MLflow, Databricks, Kubernetes

EXPERIENCE

Nvidia
Jun 2022 – Sep 2022
Data Scientist Intern | *GeForce Now - Cloud Gaming Analytics Team* Santa Clara, CA

- Developed Time-series Anomaly Detection tool to alert about malicious activities in 1000+ categories across 10M+ gaming sessions using Z-Score thresholding; Reduced response time from few months to 1 week.
- Improved in-place A/B test analysis by creating tool to identify most affected sub-population using causal inference ML models like S-Learner, T-Learner, Double ML methods etc. (using CausalML)
- Built user engagement metrics using regression analysis & game completion modelling to identify & target disengaged down-grading users. Able to track user's local & absolute engagement while progressing in a game.

Goldman Sachs
Jun 2018 – Aug 2021
Data Scientist, Analyst | *Core Machine Learning - Search Engineering Team* Bengaluru, IN

Entity recognition: Email's Salutation, Disclaimer, Signature (SDS) block extraction | Spacy, SparkNLP

- Developed scalable Conditional Random Field (CRF) model to extract Signature, Salutation and Disclaimer blocks along with contact related entities from >8M emails/day. Model achieved 85.7% accuracy.
- Enriched Goldman's knowledge graph using extracted entities, improving graph surveillances for external Bloomberg contacts.

ML features pipeline : Cross-language ML infrastructure for real-time Big Data Analytics | PySpark, Flink

- Developed a PySpark ML feature extraction pipeline & integrated with existing (Apache Flink) Big Data ETL pipeline. Primarily responsible for processing >10M emails per day and providing real-time predictions a rate of >24k data points per min.
- Deployed ML models using this pipeline for spam classification and entity recognition use-cases.

Samsung R&D Institute
May 2017 – Jul 2017
Machine Learning Intern | *Bixby AI Team* Noida, IN

- Developed offline image-classification android app, integrating custom-build optimized CNN models using Tensorflow.
- Final model achieved accuracy of 82% and occupied mere 7kb on phone with offline prediction capability.

PROJECTS

Neural Collaborative Filtering for Recommendation Systems | PyTorch [code] [report] Jan 2022 – Apr 2022

- Built 3 Neural Collaborative Filtering models: Generalized Matrix Factorization (GMF), Multi-Layered Perceptron (MLP) & Neural Matrix Factorization (NeuMF) following architecture & key-metrics evaluation from original paper.
- Compared traditional recommendation algorithms like Matrix Factorization with state-of-art Neural collaborative filtering for Movie Recommendation task. Calculated nDCG and HR@10 scores to study the effect of latent hidden factors on performance.

Sequence Tagging with Hidden Markov Model | Python, Pandas, Plotly Mar 2022 – May 2022

- Developed Tri-gram HMM class with Viterbi algorithm decoding and finding emission and transition probabilities for sequence tagging. Extended HMM model with various smoothing techniques like Laplace, Katz Back-off, & Linear Interpolation.
- Implemented context aware N-gram language model (LM) class and performed analysis on Out-of-Domain & In-Domain text.

Named Entity Recognition (NER) with BiLSTM CRF | PyTorch, Pandas, Plotly Mar 2022 – May 2022

- Built custom model with BiLSTM for feature representations & combined with CRF for CoNLL-2003 NER Shared Task. Implemented forward algorithm utilizing emission & transition potentials to compute partition function & Viterbi algorithm for decoding.