

EXPERIENCE

AI Engineer Intern

June 2021 - Present

- Designed a scalable system to normalize and assign unique IDs to cloud products across marketplaces, enabling consistent cross-platform mapping.
- Engineered pipeline using APIs to index product documents, reducing internal query time by
- Built full-stack web app (,) with fuzzy search and autosuggest, boosting product matching accuracy by

AI Research Engineer []

January 2021 - June 2021

- Developed task-specific data selection strategies (Uncertainty, Clustering) to identify most informative data points
- Performed Supervised of - LLaMA-2 on samples leveraging (HPC)
- Achieved significant reduction of in training data requirement, cutting down on annotation and training costs
- Evaluated performance using AlpacaEval (LLM-based evaluation tool), outperforming baseline in of test cases
- Conducted to validate task wise data selection substantially improves model performance

Machine Learning Engineer

January 2020 - June 2020

- Fine-tuned on customer support chats to build Gen-AI powered troubleshooting assistant for tech queries
- Streamlined preprocessing pipeline for transcripts, reducing formatting time by through vectorized and regex ops
- Conducted (EDA) on mobile insurance records to uncover root causes behind declining conversion rates using and Python.
- Collaborated with to deliver on afternoon peak-hour losses, where high call volume and customer propensity was combined with low-performing agents
- Designed agent allocation rebalancing strategy, forecasting sales improvement via simulation testing

Machine Learning Intern []

August 2019 - December 2019

- Developed a robust end-to-end Neural Speaker Diarization (EEND) system leveraging SpeechBrain and
- Integrated -based feature encoder, and fine-tuned WavLM encoder, on speech dataset (hours)
- Achieved reduction in the Diarization Error Rate (DER) for above real life dataset

PROJECTS

[]

- Executed advanced queries on Google and trained linear regression model for loan forecasting
- Engineered ETL pipeline and trained Decision Tree classifier with accuracy on loan approvals
- Built pipeline with Airbyte, , and to integrate survey, stock, and forex data for unified data analysis.

[\[Github\]](#)

- Performed using Correlation Coefficient to extract key features (Readability Score, Vocabulary, length)
- Implemented and achieving accuracy of in predicting difficulty levels.

PUBLICATIONS

- "Improving Task Diversity in Label Efficient Supervised Finetuning of LLMs." [Abhinav Arabelly*](#), Jagrut Nemade*, Jifan Zhang, Robert D Nowak

SKILLS

: Python, SQL, R, C++, PyTorch, TensorFlow, Transformers, CNN, XGBoost, LightGBM, Pandas, Numpy, Tableau, Matplotlib, Regression, scikit-learn

: LLMs, RAG, Agentic Frameworks, FAISS, ChromaDB, Hugging Face, SpeechBrain

: AWS, GCP, Spark, Kafka, Hadoop, BigQuery, Databricks, Docker, MLflow, FastAPI, Airbyte, Snowflake, DB

EDUCATION

Master of Science in

August 2021 - May 2022

: Statistical Models, Big Data Systems, Data Structures, Algorithms, Data Management, Data Visualisation

Integrated MSC.

and B.E. Electronics and Communication

August 2019 - May 2021