Rishikesh Ajay Ksheersagar

Ann Arbor, MI | +1(734)489-2596 | rishiksh@umich.edu | https://www.linkedin.com/in/rishikeshksheersagar/ | https://rishiksh20.github.io/

PROFILE

- *Skills:* Natural Language Processing (NLP), Large Language Models (LLMs), Machine Learning (ML), Deep Learning, Statistical Analysis, Reinforcement Learning (RL), Information Retrieval, Search Engines, Quantitative Methods, Big Data
- *Languages:* Python (Pandas, Dask, NumPy, ScikitLearn, Tensorflow, PyTorch, Keras, NLTK, Spacy, Vader, StreamLit), SQL, R, PySpark, C++
- *Tools / Platforms:* Snowflake, Hadoop, GCP, AWS, Jenkins, Tableau, PowerBI

EDUCATION

University of Michigan - Ann Arbor   **August 2023 – April 2025**
Masters in Data Science, GPA 4.0/4.0   **Ann Arbor, MI, USA**
Subjects: CSE 595 (NLP), ECE 598 (LLMs), CSE 545 (ML), SI 650 (Information Retrieval), STATS 510 (Probability Distributions)

Savitribai Phule Pune University   **June 2015 – June 2019**
Bachelor of Engineering in Computer Engineering, GPA 3.7/4.0   **Pune, India**

PROFESSIONAL EXPERIENCE

UNIVERSITY OF MICHIGAN   **May 2024 – Present**
Research Assistant @ LIT Lab   **Ann Arbor, MI, USA**
- Conducted NLP research on "Climate Change, Demographic Shifts, and Socio-Political Stability in Sub-Saharan Africa" under the Minerva Initiative. Automated metadata retrieval, PDF scraping, and text extraction for 50k+ research papers. Analyzed 20k+ causal sentences linking climate change and social unrest using POS tagging and LLMs.
- Explored Fake News Detection and User Perception by leveraging LLMs to model semantic and demographic similarities across 5.1 million Reddit posts and 1,350 Qualtrics survey responses using word embeddings to investigate Misinformation spread.

Graduate Student Instructor
- Conducted weekly lab sessions for 60+ students in the QMSS 301 course during Fall 2024 and Winter 2025 semester, topics include – Geospatial Analysis in R, Predictive Modeling and Sentiment Analysis in Python, Web Scraping, and Research Methodologies.

MU SIGMA INC.   **July 2019 – June 2023**
Data Science Manager   **Bangalore, India**
- Managed 2 teams consisting of 16 data scientists working with Fortune-100 clients in Telecom domain, spearheaded the growth and management of engagements generating $1.5M annually.
- Engineered an advanced unbalanced multi-class classifier using RxMER data, stacking XGBoost and sequential Neural Network models, resulting in 45% decrease in probable outages by precisely identifying causation of Modem Network Impairments in near real-time.
- Developed a PoC tool to simulate Patient Journeys in Clinical Trials by integrating Therapeutic Area, Site, PI, Patient, and Trial attributes, utilizing Bayesian Networks and Agent-Based Models to achieve 79% similarity with previous real trials and enabling proactive planning and mid-trial adjustments for Phase 3 Clinical Trials.
- Drove RFP connects with CXOs of 2 Fortune-100 Telecom clients for adding new engagements to Mu Sigma's portfolio successfully.

Data Scientist
- Led a team of 7 data scientists in identifying key features for degraded network service for the Data Science and Data Engineering team of a Fortune-100 Telecom clientele.
- Conducted EDA on 7 datasets including Cable Modem Registration, Speed Tests, Modem Utilization, and PNM (Proactive Network Maintenance, RxMER, FEC), achieving 98.7% accuracy in detecting degraded network service events and ~37% fewer customer complaints.
- Enabled Digital Transformation for a legacy Store Planning tool for the FP&A team of the world's largest Home Improvement Retailer, reducing tool execution time by 60%.
- Utilized Washout 2 and 3 constraint Optimization, along with Time Series models (ARIMA) to design Financial Plans of 11 retail metrics resulting in the client organization exceeding planned Gross Margin by ~2% in FY 2020-21.
- Created 7 Tableau Dashboards to provide detailed insights and flag anomalies in Financial Plans, empowering Store Managers and Region Finance Leads to meet Sales targets in 63% more instances.

BMC SOFTWARE   **August 2018 – April 2019**
Project Intern   **Pune, India**
- Worked on a PoC which involved implementation of private Blockchain with voting-based consensus mechanism by leveraging Hyperledger Composer, in addition to a traditional Structured Database, in the backend of a globally used legacy ITSM Software.

ACADEMIC PROJECTS

- **PapeRet** (Sept - Dec 2024) - Designed a research paper retrieval system, processing 98,000+ academic papers using recursive metadata extraction, web scraping, PDF download and text extraction. Leveraged LLaMA for Retrieval-Augmented Generation (RAG) to create summaries. Achieved significant performance improvements, with MAP@10 of 0.539 and NDCG@10 of 0.81. [GitHub][Report]
- **Register Augmented LLM Fine-Tuning** (Oct - Dec 2024) - Developed a register-augmented fine-tuning approach for LLMs, enhancing global context management and interpretability. Implemented RegBERT for QA tasks, improving F1 and Exact Match scores on the TyDiQA GoldP dataset, with attention analysis using Layer-wise Relevance Propagation (LRP) and Integrated Gradients. [GitHub][Report]
- **Few-Shot Preference-Based RLHF** (Jan - May 2024) - Implemented and refined few-shot preference-based reinforcement learning algorithms, including MAML, iterated MAML, and REPTILE, to optimize human feedback efficiency on Metaworld datasets. Developed a generalized reward function adaptable to new tasks with minimal human queries and ~90% reduction in training time. [GitHub][Report]
- **Is it easy to be Multilingual** (Nov - Dec 2023) - Explored mBERT's transfer mechanics, emphasizing syntactic, morphological, and phonological similarities as key predictors. Displayed language model performance's critical role in cross-lingual transfer. Proposed a framework achieving 62.5% accuracy in selecting optimal source language for multilingual cross-transfer. [GitHub][Report]

HONORS AND AWARDS

- **Mu Sigma Inc.:** Received SPOT Awards in 3 consecutive years (Aug 2022, Aug 2021, Oct 2020) for exceeding project goals, delivering exceptional results and designing optimal solutions.