Rahul Yadav

+1(716)238-0835 | rahulyad@buffalo.edu | linkedin.com/in/rahul-yadav-704 | Github | Website

EDUCATION

University at Buffalo, Buffalo — Aug. 2022 – Dec. 2023
Master of Science in Computer Science and Engineering — New York, US
Coursework: NLP, Computational Linguistics, Pattern Recognition, Machine Learning, Algorithms Design — GPA: 3.79 / 4

Indian Institute of Technology (Banaras Hindu University), Varanasi — Jul. 2014 – May 2018
Bachelor of Technology in Electronics Engineering — Uttar Pradesh, India

EXPERIENCE

Apex Analytix — May. 2023 – Aug. 2023
Machine Learning Engineer Intern | Archimedes Team — Greensboro, US

Knowledge Specific Conversational Agent

- Developed task-specific generative chatbot, leveraging a knowledge base derived from cross-team documents
- Built efficient retrieval system by utilizing LangChain for custom document preprocessing and ALBERT for vectorization.
- Deployed service across 12 teams within the organization, facilitating seamless and intelligent access to the 110 documents

Oracle — Sept. 2020 – Aug. 2022
Senior Application Engineer | OFSS Machine Learning Team — Bengaluru, India

Preemptive Anomaly Prediction in Corporate Billing

- Engineered in-memory multivariate anomaly prediction service with model explainability, for insights into billing anomalies
- Optimized database performance with indexing and parallelism, processing 1.2M bills and 5M bill segments within 20 minutes.
- Integrated anomaly prediction service into Oracle Revenue Management and Billing (ORMB) product

Wipro Limited — June 2018 – Sept 2020
Project Engineer | iX Team CTO Office — Bengaluru, India

Chatbot Services for Employee Helpline Portal's Ticketing System

- Implemented chat service for the employee helpline portal that accurately classified user query intent and used BERT embedding to provide relevant responses based on similarity with historical resolutions
- Released service significantly reduced workload on agents, resulting in decreased average wait time from 18 to 4 minutes

PROJECTS

Topic-Based Empathic Chatbot — Spring Term 2023

- Implemented open-domain chatbot with retrieval-based and casual conversation capabilities.
- Developed logical and rule-based dialog manager for effective query redirection based on user interaction
- Integrated retrieval and generative models, improving conversational accuracy by 35% while maintaining resource efficiency.

Network-based Intrusion Detection System (NIDS) — Research Assistant, Dr. Hongxin Hu | Spring Term 2023

- Implemented an Intrusion Detection System using deep neural detectors to efficiently identify and respond to potential security threats in the network, mitigating the risk of data loss and downtime.
- Conducted comprehensive analysis on 12 different network attack datasets and evaluated the performance of 4 deep neural detectors. Reported insights into the limitations and effectiveness of methods.

PATENTS and PUBLICATIONS

Paper Published: Virtual Conversation with Real-Time Prediction of Body Moments/Gestures — **ICMLIP 2019**
*Gopichand Agnihotram, Rajesh Kumar, Pandurang Naik, **Rahul Yadav*** — springerLink

Patent Granted: Method And System For Multimodal Analysis Based Emotion Recognition — **16/795840 [PDF]**
*Inventors: **Rahul Yadav**, Gopichand Agnihotram* — Applicant: **Wipro Limited**

SKILLS

Languages : Python, Java, C++, C, PLSQL, SQL, MongoDB

Technologies : Retrieval Augmented Generation(RAG), Large Language Models(LLMs), Natural Language Processing(NLP), Machine Learning, Deep Learning, Web Services, Data Structures, Algorithms, Prompt Engineering, Indexing, Quantization

Cloud : Oracle, Azure, Google Cloud Platform (GCP), Amazon Web Services (AWS), Atlas

Frameworks and Libraries : Pytorch, Tensorflow, MLFlow, Kafka, SentenceTransformers(embeddings, re-rankers), PySpark, Langchain, LlamaIndex, VectorDBs(ChromaDB, Faiss, elasticsearch), OpenAI, MLFlow, Git, Jenkins, Flask, Docker