Adirtha Borgohain

91-8723912165 adirthaborgohain@gmail.com linkedin.com/in/adirtha-borgohain github.com/AdirthaBorgohain

Experience

InveniAI **Bengaluru, Karnataka/Remote**

Senior Data Scientist *March 2022 – Present*

- Building a **Named Entity Recognition**, **Entity Linking**, **Relation Extraction** pipeline for InveniAI's proprietary AI platform, **AlphaMeld** to extract biomedical entities from texts, disambiguate them to their normalized forms and assign relations between them.
- Built **automated ETL pipelines** for sourcing, extracting, and making machine learning inferences on data and writing to **Postgresql** and **neo4j** databases using **RabbitMQ** and **Redis** which reduces the time and cost from the previous flow by more than 2x. Accumulated over 1 billion data points which are being leveraged to build robust and cutting-edge solutions that contribute to drug discovery processes.
- Led the development of **ChatAlphaMeld**, a biomedical chat engine, leveraging **sentence-transformers**, **Langchain**, and **LlamaIndex** with **LLM** models. It dynamically decomposes user queries and structures plans to accurately address user queries, while intelligently utilizing web resources to generate well-informed and relevant responses. This streamlined the research process for drug researchers, optimizing time and enhancing productivity.

Netenrich **Bengaluru, Karnataka**

Data Science Analyst (prev. Intern) *Feb 2020 – March 2022*

- Deployed text classification models using word and character embedding methods: **chars2vec, AlBERT** to filter out piracy-related text content from curated articles for a cyber-intelligence news platform, **KNOW** (know.netenrich.com). Designed the **clustering algorithm** for clustering similar news articles together in the platform.
- Built NLP based **multi-task regression and classification** solutions to improve efficiency and reduce manual overload in SLA response times for incoming issues and tickets
- Built **Named Entity Recognition Models (NER)** using **BERT** and **Attention** based deep learning architecture for detection of cyber threat entities in raw text sources, improving the accuracy and efficiency of the task from the previous implementation of rule-based models.
- Designed a probabilistic graph based **Bayesian network architecture** to determine the root causes behind a failure or an issue to speed up the process of root cause analysis.

Threatlandscape **Bengaluru, Karnataka**

Machine Learning Intern *June 2019 – July 2019*

- Built a source code based phishing detection model using **monte carlo dropout estimation** to capture uncertainty in outputs of the model. Successfully **dockerized** the project and deployed in a server using **nginx** for load balancing.
- Developed **Graph Convolutional Network** model and **Poincare Embeddings** of a threat intelligence-based graph for capturing its overall structure so that information could later be used for text generation and entity prediction.

Technical Skills

Languages: Python, C, C++, SQL, JavaScript, Typescript, HTML/CSS
Tools/Frameworks: Git, Tensorflow, Keras, spacy, Docker, Kafka, RabbitMQ, ElasticSearch, transformers, Redis, FastAPI, Flask, Postgres, langchain, llama-index

Projects

art-critiq | *transformers, FastAPI, pyTorch*

- A multi-modal pipeline to generate three tones of reviews for an artwork using pre-trained **blip2** and **Flan-T5** models.

intelliweb-GPT | *langchain, llama-index, trafilatura, transformers*

- Intelligent search engine/QA module that uses **GPT** models to provide accurate, relevant & recent answers from Google News/Web, & can also directly answer user queries using GPT's training knowledge.

Education

Tezpur University **2016 – 2020**

Bachelor of Technology in Computer Science, ***CGPA:*** *8.51/10* *Tezpur, Assam*

Research and Publications

Detection of Malicious Network Traffic using Machine Learning

- **Borgohain, A.**, Sarmah, S., Bhattacharrya, D.K., *"Detection of Malicious Network Traffic using Machine Learning"*, Proceedings of International Conference on Recent Trends in Science and Technology (ICRTST)-2020, ISBN: 978-93-5396-830-4. Submitted to BJIT – "International Journal of Information Technology", ISSN: 2511-2112 for review.