Farjana Sultana Samia

Ph.D. Student, Iowa State University

Website ⋄ LinkedIn ⋄ Github ⋄ fssamia@iastate.edu

RESEARCH EXPERIENCE

Ph.D. Research /Data and Knowledge Mining Lab

- NLP, LLM, Information Retrieval, Knowledge Graph, Entity Linking, RAG
- Currently working in advanced research work on Natural Language Processing (NLP), Large Language Models (LLMs), Information Retrieval, Knowledge Graphs, Retrieval-Augmented Generation (RAG) and Entity Linking.

Bangla Text Document Categorization

- *"A Resource-efficient NLP-based Online Bengali News Categorization Model with Outlier Removal",* (On review Elsevier IPM publication)

Undergraduate Research Work

- *Prediction of COVID-19 from Chest X-ray Images Using CNN And Pre-Trained Models*

PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant /Iowa State University January 2024 - Present

- COM S 327: Advanced Programming Techniques.
- My responsibilities are grading and helping students to solve problems.

Software Engineer (NLP) /Technometrics Limited December, 2022 - December, 2023

- Developed Bengali Multi-label Topic classification model using pre-trained BERT. which can predict 17 different topics with 85 percent accuracy (Multi-class, BERT, Flair, Pytorch, Transfer Learning, Fasttext)
- Developed a Bengali aspect-based sentiment analysis model to predict different polarity scores based on different aspects from the same sentence (PyABSA, Predicting aspect and sentiment of that aspect, Writing Annotation Guideline)
- Developed a social media and news media data collecting bot to extract all types of data from Facebook and several news portals and analysis application, including Word Cloud, Trending Topic, Most Active User, Viral Post Analysis, etc. This bot can automatically complete scraping tasks as per the scheduled timetable, and store the data in MongoDB. The collected data can be made available through an API as required (Selenium, MongoDB, Scraping, Multiprocessing)
- Developed a scraping pipeline that can scrape 25 Bengali blogs and 15 news portals (Scrapy, MongoDB)

TECHNICAL SKILLS

- **Programming Languages:** Python, C/C++, C#
- **Libraries:** LlamaIndex, LangChain, Neo4j, Pinecone, Huggingface Transformers, TensorFlow, Keras, Numpy, OpenCV, Matplotlib, Pandas.
- **Web Technologies:** HTML, CSS, Bootstrap, Asp.net, Flask, FastAPI
- **Web Scraping:** Scrapy, Selenium, Request, Beautifulsoup
- **Database:** MySQL, MongoDB, Neo4j
- **Tools:** Github, VS Code, SPSS, Jupyter Notebook, Pycharm, Matlab, Photoshop, SQL Server, OS (Ubuntu, windows)

EDUCATION

Ph.D. in Computer Science (Jan 2024 – Present)
Iowa State University, Ames, IA, USA
Cumulative GPA: 4.00 out of 4.00

B.Sc. in Computer Science and Engineering (Feb 2017 – Dec 2021)
Manarat International University (MIU), Dhaka, Bangladesh
CGPA: 3.90 out of 4.00
Latter Grade: Graduated with distinction; 1st position among 48 students

PROJECTS

Project 1 Retrieval-Augmented Generation (RAG) GitHub

- Used GPT & Llama openAI LLM for generation and Chunking, Re-ranking was applied to get accurate response. Funix was utilized for turning Python functions/classes into accessible web applications.

Project 2 CIFAR-10 - Object Recognition in Images GitHub

- Object Recognition in CIFER-10 Images(CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes). (Using Python Keras library)

Project 3 Red blood cell counting from images of blood cells using transfer learning GitHub

- Faster RCNN, SSD, YOLO