Harsh Shah

Senior undergraduate student interested in Deep Learning, Natural Language Processing and Information Retrieval

Education

BITS Pilani Aug. 2017 – Present
Bachelor of Engineering in Electrical and Electronics, Master of Science in Economics *Pilani, Rajasthan*

Technical Skills

Languages: Python, C, C++, MySQL, Java
Developer Tools: Git, Docker, VS Code, Jupyter, PyCharm, Eclipse, Anaconda
Libraries and Frameworks: PyTorch, Pandas, NumPy, Keras, TensorFlow, NLTK, Transformers, Spacy, scikit-learn

Relevant Courses

Neural Networks and Fuzzy Logic, Object Oriented Programming, Operating Systems, Computer Programming, Linear Algebra, Probability and Statistics, Applied Econometrics, Derivatives and Risk Management

Publications

A graph-aware temporal framework with influence, identity, and interpersonal cues for email reply prediction *(Under Review)*
Harsh Shah, Kokil Jaidka, Lyle Ungar, Jesse Fagan, Travis Grosser

Experience

Visiting Researcher Aug 2021 – Present
National University of Singapore *Singapore*

- Working under the guidance of Dr. Kokil Jaidka and Prof. Lyle Ungar (University of Pennsylvania)
- Exploring the use of OpenAI **Codex** API while developing code and models for different projects
- Researching Multimodal and Graph-based approaches for context aware email reply prediction
- Detecting implicit microaggressions towards minorities in corporate communication

Research Intern May 2021 – July 2021
Philips *Bangalore*

- Built a document information extractor using open-source libraries in Python and Java
- Finetuned Multimodal transformer model **LayoutLM** for Document Understanding on FUNSD dataset
- Achieved Recall and Precision over **0.95** in tasks like Key-Value pair and Table extraction from medical documents

Natural Language Processing Intern May 2020 – July 2020
FlexiEle *Gurgaon*

- Information extraction from unstructured data using **NLTK** and **OpenCV** and stored data in JSON format
- Built and deployed a **RASA** based chatbot for tasks like leave management and stress assesment
- Secured the chatbot with a custom HTTP input channel having JWT authentication to allow authorized access

Data Science Intern May 2019 – July 2019
Belief Systems *Chennai*

- Wrote scripts to scrape data using **BeautifulSoup** and Selenium to create database for offline use
- Built a people counter app using OpenCV to keep track of people inside the shop using video from CCTV

Teaching

Head Teaching Assistant, Neural Networks and Fuzzy Logic January 2021 – May 2021
BITS Pilani *Pilani*

- Conducted workshops to familiarise students with NumPy, Pandas, PyTorch and TensorFlow
- Designed and evaluated Neural Networks based Python assignments to test their practical knowledge
- Supervised students as they implemented and reproduced results from research papers

Projects

Intro to Codex | *Python, OpenAI API* | GitHub Link — Sep. 2021 – Oct. 2021

- Explored the Codex based language models of code generation and completion available through OpenAI API
- Demonstrated the use of this API for tasks like converting code from one framework to another and error solving

Transfer Learning in NLP | *Python, Transformers, PyTorch Lightning* | GitHub Link — Jan. 2021 – May 2021

- Researched current trends in NLP like Self-Attention and Pretraining large Language Models
- FineTuned **t5-small** on different tasks like Sentiment Analysis and Emotion Classification

ML for Option Pricing | *Python, Keras, Pandas, NumPy* | GitHub Link — Jan. 2021 – Present

- Explored the effectiveness of different Neural Network architectures for the task of Option Pricing
- Compared the results of ML based techniques with traditional models like the Black-Scholes and the Heston model

Attention based LSTM | *Python, PyTorch, NumPy* | GitHub Link — Nov. 2020 – Dec. 2020

- Implemented the paper *Attention-based LSTM for Aspect-level sentiment classification* on SemEval 2014 dataset
- Reproduced the results from the paper and additionally visualized attention weights for different aspects

YouTubeNLP | *Python, Docker, Flask, Gensim, Transformers* | GitHub Link — May 2020 – Oct. 2020

- Worked on a web application which provided insights about comments and transcript of YouTube videos
- Used different transformer based models like **BERT** and **GPT** for tasks like NER and Emotion Analysis

Fashion Intelligence System | *Python, PyTorch, React* | GitHub Link — July 2020 – Aug. 2020

- Developed a Fashion Intelligence System which scrapes and ranks products from e-commerce websites
- Products were ranked based on similarity between image embeddings of scraped and trending products

Competitions

Flipkart GRiD 2.0 — July 2020

Finished in the Top 1.2 percentile out of 22,000+ participants in a Software Development competition

Extra Curricular Activities

Group Leader — March 2018 – March 2021

Department of Sponsorship and Marketing, BITS Pilani

- Headed data analysis and helped the team to close various deals and make better decisions
- Set up data pipelines and macros to facilitate effective and targeted communication with potential sponsors

Volunteer — Aug. 2018 – Aug. 2019

National Service Scheme (NSS)

- Taught English and Personality Development to underprivileged children of Pilani village
- Conducted speaking and listening assignments for students in government schools to improve their English skills

Referees

Dr. Kokil Jaidka
Assistant Professor
Department of Communications and New Media
National University of Singapore
jaidka@nus.edu.sg

Prof. Surekha Bhanot
Professor
Department of Electrical & Electronics Engineering
BITS Pilani
surekha@bits-pilani.ac.in

Certifications

- Applied Social Network Analysis in Python | Ongoing
- Social-Behavioral-Educational Foundations | Certificate Link
- Natural Language Processing in TensorFlow | Certificate Link
- Neural Networks and Deep Learning | Certificate Link
- Convolutional Neural Networks | Certificate Link