SARANSH GUPTA

+91 9530277421 ◇ Rajasthan, India

Email ◇ Linkedin ◇ GitHub ◇ Google Scholar ◇ Website

EDUCATION

Indian Institute of Technology Kharagpur 2017 - 2022
Integrated Master of Technology in Engineering Product Design **Grade:** 8.09 / 10

Master's Dissertation:
- A Spatial Optimization Approach for Identifying the Optimal Hospital Locations Jan 2022 - Apr 2022
- Addressing Bottlenecks in Healthcare Accessibility Jul 2021 - Nov 2021

Activities and societies:
- Safety Analytics Virtual Reality (SAVR) Lab Jan 2019 - Jul 2019
- Autonomous Underwater Vehicle Research Group IIT Kharagpur May 2018 - Dec 2018
- TeamKART: Official FSAE of IIT Kharagpur Sep 2017 - May 2018

PUBLICATIONS

- S. Gupta et al. "Integrative Network Modeling Highlights the Crucial Roles of Rho-GDI Signaling Pathway in the Progression of Non-Small Cell Lung Cancer," in **IEEE - JBHI, 2022**, doi: 10.1109/JBHI.2022.3190038
- Entity-aware Question-Answer Extraction for Shopping Guidance, **Amazon Machine Learning Conference**

SKILLS

Technical Skills Python, PyTorch, Transformers, BERT, Transfer Learning, scikit-learn, TensorFlow
Soft Skills Critical thinking, Problem-Solving, Team Player

WORK EXPERIENCE

American Express Aug 2022 - Present
Engineer-III *Gurugram, Haryana, India*
Project 1: Failure cause identification of applications on generated Incidents for their automated resolution

- Implemented a **Question-Answer** based strategy on top of raw dataset to identify failure cause of applications
- Achieved **F1 Score of 0.84** by finetuning a pre-trained **BERT based Question-Answering model**

Project 2: Automation of various repetitive tasks to save the manual efforts

- Improved the **data security** by developing an automatic PII data identification and encryption tool
- Reduced **12 business hours per month** by automating the application availability report generation process
- Automated resolutions for certain repetitive Incidents saving upto **2 business hours** every day

INTERNSHIPS

Amazon Development Centre India Jan 2022 - June 2022
Applied Scientist - Intern *Bengaluru, Karnataka, India*
Project: Generate Pre-curated Question Bank (PCQB) by Question and Answer extraction from articles

- Developed a **Transformers-based** two-step model for Question Generation followed by the answer extraction
- Scrapped Texts, People Also Ask (PAA) questions and answers using queries related to the E-Commerce domain
- Achieved a **Perplexity score of 82.3** on Question Generation by fine-tuning a **T5 model** on the PAA dataset
- Attained **F-1 score of 0.79** on the answer extraction task by fine-tuning the T5-large model on the PAA dataset
- Deployed the two-step model pipeline on the **streamlit-based** demo web application that accepts user input

ZS Associates Jan 2021 - June 2021
Data Science Associate - Intern *Bengaluru, Karnataka, India*
Project 1: Extract biomedical text dataset, identify entities, and classify if there exists a relation between entities

- Created a pipeline to extract texts from PubMed database, identifying entities using **Selenium** and **PubTator**
- Implemented Binary Classification rules, devised **four** labeling functions using bio-verbs, co-occurrence of entities
- Generated a training dataset utilizing the four labeling functions in **Snorkel** by applying the **Weak Supervision**
- Achieved **F1 score of 0.88** on the test dataset in relation-classification by fine-tuning **RoBERTa base** model

Project 2:: Identify the type of relationship between two entities if it exists from the results of the Project-1

- Created a new set of **three** labeling functions for relation-type identification by using the results of the project-1
- Attained **F1 score of 0.83** on the test dataset using XGBoost Model followed by feature engineering

RESEARCH EXPERIENCE

Emory University Jul 2022 - Present
Volunteer Researcher (remote) *Atlanta, GA, USA*
Project: Predict the type of Venous thromboembolism (VTE), from the medical diagnosis and clinical Impressions

- Reduced manual adjudication of dataset by **20 times** using pegasus paraphrasing model on sample dataset
- Achieved **F1 score of 0.97** in predicting the type of VTE on test dataset by fine-tuning a **Bio-BERT** model
- Improved F1 score on test dataset by **20 percent** by deploying paraphrasing and Bio-BERT finetuning pipeline

Osaka University Jan 2020 - Dec 2020
Research Assistant (remote) *Ibaraki, Osaka, Japan*
Project: Predict Non-Small Cell Lung Cancer (NSCLC) using Machine Learning, identify potential drug targets

- Extracted **412** essential genes out of **10,077** by applying Boruta Feature selection on gene expression dataset
- Obtained **F-1 score of 1.0** on validation, **0.98** on test dataset by using the **XGBoost** model to predict NSCLC
- Predicted drug targets for the NSCLC by simulating a **Bayesian Network Model** on Rho-GDI signaling path

ACHIEVEMENTS

- Conferred **Blue-Award** at the **American Express** for impactful contribution to the organization in **Jan 2023**
- Featured as one of the **Top 30** Undergraduate Achievers of IIT Kharagpur in the UG Achievers Directory 2020
- Awarded scholarship of **2200€** by The A*Midex Foundation of **Aix-Marseille University, France**, Feb 2020
- Selected among **Top 5 percent** out of all for the summer fellowship at Institute of Science Technology Austria
- Featured in the **ISE Newsletter Autumn-2020** under Department Spotlight of ISE fights COVID-19, 2020

COMPETITIONS AND CONFERENCES

- Annual Amazon Machine Learning Conference (AMLC) – Bengaluru, Karnataka Aug 2022
- 23rd World Business Dialogue, Creation Lab at Evonik - Cologne, Germany Jun 2022
- Amazon ML Summer School 2021: Offered PPI Jul 2021
- International Conference on Human Interaction Emerging Technologies: Future Applications Aug 2020
- Young Data Scientists annual meetup at Kaggle - days, Dubai World Trade Centre Mar 2020
- Winner at Databuzz 2020 conducted by DoMS, IIT Madras Jan 2020

LEADERSHIP

- Mentored **5 Undergraduate students**, actively involved in facilitating their one-one sessions and group doubt-clearing sessions to ensure their steep learning curve and right career trajectory