Saransh Gupta

Contact Email LinkedIn Github Website

ACADEMIC PROFILE

Year	Institution	Degree/Examination	Percentage/CGPA
2022	Indian Institute of Technology Kharagpur	M.Tech. (Engineering Product Design)	9.26 / 10.00
2022	Indian Institute of Technology Kharagpur	B. Tech. (Engineering Product Design)	8.07 / 10

PUBLICATIONS

- (Gupta et al.) An integrative machine learning and Bayesian modeling approach highlights the crucial roles of the Rho-GDI signaling pathway in the progression of non-small cell lung cancer (NSCLC); implications for drug target discovery, (under review), **IEEE Journal of Biomedical and Health Informatics (JBHI-01644-2021), 2021**
- (Gupta et al.) Development of a virtual reality-based fire training simulator and machine learning-based path guidance system (working paper), **IHIET-AI, 2020**, Centre Hospitalier Universitaire Vaudois, **Lausanne, Switzerland**

INTERNSHIPS AND PROJECTS

Amazon Inc. | Applied Scientist - Intern **Jan'22 – June '22**

Project - 1: *Build a demo tool to help in the navigation and exploration of the Pre-curated Question Bank (PCQB)*

- Created a dashboard using **streamlit** enabling a user to input their query and get relevant questions accordingly
- Integrated the frontend with the backend and a **BERT** based model to fetch relevant questions based on queries input
- Demonstrated the coverage of PCQB with respect to user queries using the query-question relevance feature
- Developed a navigation feature to allow searching and navigating the PCQB based on pre-determined tags

Project - 2: *Pre-curated Question Bank (PCQB) Question and Answer extraction from articles (**ongoing**)*

ZS Associates Inc. | Data Science Associate - Intern **Jan'21 – June '21**

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 the 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 gold-standard dataset in relation-classification by training **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** labelling functions for **relation-type identification** by using the results of the project-1
- Attained **F1 score** of **0.83** on the gold-standard dataset using **XGBoost** Model followed by feature engineering

Tools and Software: Python, TensorFlow, Transfer Learning, Medline-Plus API, PubTator, Selenium, Snorkel

Osaka University, Japan | Remote Research Assistant **Jan '20 – Dec '20**

Guide: Dr. Kenji Mizuguchi, Mizuguchi Lab, Osaka University, Osaka, Japan

Project: *Predict the Non-Small Cell Lung Cancer (NSCLC) using Machine Learning, identify its potential drug targets*

- Extracted **412** essential genes out of **10,077** by applying **Boruta** Feature selection on their gene expression dataset
- Obtained **F-1 score** of **1.0** on validation and **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 the Rho-GDI signaling pathway
- Discovered methodology leads to an accurate treatment of the disease impacting **85%** of the lung cancer patients

Tools and Software: Python, TargetMine, scikit-learn, smote, NetworkX, NumPy, pandas, Plotly, joblib

ACHIEVEMENTS

- Featured as one of the **Top 30 Undergraduate Achievers** of IIT Kharagpur in the UG Achievers Directory 2020
- Conferred merit-based scholarship of **2200 €** by The A*Midex Foundation of **Aix-Marseille University, France**
- Selected among **Top 5%** out of all for the summer fellowship at **The Institute of Science & Technology Austria**
- Got featured in the ISE Newsletter Autumn-2020 under the Department Spotlight of **ISE fights COVID-19, 2020**
- Awarded as **Intern of The Month** for my contribution as a Data Analyst at Sapio Analytics by the CEO in July 2020

COMPETITIONS / CONFERENCES

Young Data Scientists annual meetup at Kaggle - days, Dubai World Trade Centre **[Mar 2020]**

International Conference on Human Interaction & Emerging Technologies: Future Applications **[Aug 2020]**

- Virtually Presented the research paper on "Development of a virtual reality based fire training simulator and machine learning based path guidance system: A case of hospital fire breakouts" at Lausanne, Switzerland

Runner Up | Databuzz (Data Analytics Competition) DoMS IIT Madras **[Jan 2020]**

Objective : Prediction of the defaulters on lending credit cards to minimize loss incurred to the banks (credit risk)

- Visualised key features governing target variable, removed outliers, imputed missing values
- Implemented undersampling for handling imbalanced classes, applied decision trees as baseline model

Improved the validation AUC-ROC from 0.56 to 0.79 by applying the Bayesian optimization model

Gold Winner| Inter-Hall Data-Analytics competition| IIT Kharagpur **[Feb 2019]**

- Worked in the Visualization of data-set, making interpretation of trends followed by different variables
- Applied **XGBoost** followed by feature engineering, obtained **82.55%** on training, **80.66%** on test data
- Awarded best in the category of best F-1 Score, Overall stood second in terms of Business Case Study