Arnav Singla

Third-Year Undergraduate - Statistics and Data Science
arnavs21@iitk.ac.in | +91 8427277433 | Arnav Singla | arnav39

Academic Qualifications

Year	Degree/Certificate	Institute	Performance
2021 - Present	BS, Statistics and Data Science	Indian Institute of Technology Kanpur	**9.21**/10
2021	CBSE(XII)	SGGS SR SEC School, Chandigarh	97.4%
2019	CBSE(X)	St John's High School, Chandigarh	96.4%

Academic Achievements

- Secured **All India Rank 1018** in Joint Entrance Examination(JEE) Advanced 2021 among **1.41 Lakh** candidates from across the country
- Secured **All India Rank 1653** in Joint Entrance Examination(JEE) Mains 2021 among **1 million** participants from across the country
- KVPY SX 2020-2021 fellowship awardee with an **All India Rank 1474** conducted by Indian Institute of Science, Government of India
- Received the **Academic Excellence Award** twice for exceptional academic performance in academic sessions 21-22 and 22-23

Key Projects

Sign Language Segmentation for Indian Sign Language | *Course Project* | *Jan'23 - Apr'23*

Prof. Ashutosh Modi, Department of Computer Science and Engineering, IIT Kanpur | CS779 - Statistical NLP

- Built a **deep learning model** for delineating boundaries between signs in sign language to distinguish individual signs in continuous footage
- In order to first extract **spatio-temporal** information from the frames, the model made use of an **inflational 3D convolutional network**. Following that, it made use of a **multi-stage temporal convolutional network** to be able to identify the boundaries of the signs
- Due to lack of annotated data for Indian Sign Language, we used a transfer learning paradigm called **source free domain adaptation**
- We attempted to reproduce the findings of a research paper, Sign Segmentation with Changepoint-Modulated Pseudo-Labelling by Renz et al., that involved training a sign segmentation model for German Sign Language using a model pre-trained on British Sign Language data

Air Quality Index Prediction | *IIT Kanpur Consulting Group* | *Oct'22-Ongoing*

Prof. Sachchida Nand Tripathi, Department of Civil Engineering, IIT Kanpur

- Using **machine learning** models such as **Transformers, LSTMs and ARIMA** to predict the values of **PM2.5** using atmospheric data
- Conducted **error correction** of low-cost sensors by incorporating **spatiotemporal** and **atmospheric data** and reference sensor values
- **Calibrated** low-cost sensors using machine learning techniques such as **Gaussian Process Regression** and **Gradient Boosting Regression**
- Developed a **dashboard** for air quality data of Bihar and UP using low-cost sensors deployed at multiple sites providing live **PM2.5** values

Building Corpus for Indian Sign Language Processing | *Undergraduate Project* | *Jan'23- Ongoing*

Prof. Ashutosh Modi, Department of Computer Science and Engineering, IIT Kanpur

- Built one of the **largest Indian Sign Language (ISL)** datasets to date, containing over **138k** sign language videos and their captions
- Co-authored a research paper (currently under review) submitted to the **NeurIPS Conference 2023 (Datasets and Benchmarks)**
- We successfully implemented a **transformers model** for sign language recognition and translation on our dataset to obtain baseline scores
- Leveraged libraries including **pytube, pydub, youtube-transcript-api**, and **ffmpeg** to scrape and process data from YouTube channels.

Sentiment Classification using Deep Learning | *Course Project* | *Apr'23- Apr'23*

Prof. Ashutosh Modi, Department of Computer Science and Engineering, IIT Kanpur | CS779 - Statistical NLP

- Trained a **Transformer** model without making use of any pre-trained models for sentiment classification using the provided dataset
- In terms of the F1 metric, in the validation phase the model received a score of 0.658, whereas the score for the test phase was 0.646

Analyzing Olympics Data | *Course Project* | *Aug'22 - Nov'22*

Prof. Dootika Vats, Department of Mathematics and Statistics, IIT Kanpur | MTH208A - Data Science Lab I

- Utilised packages such as **Rvest** and **Tidyverse**, we **scraped** data from the web about the Olympics and a variety of other factors.
- **Analysed** it in terms of various factors such as life expectancy, alcohol consumption, literacy rates, unemployment rate and host nation
- Developed a dashboard app to visualize data and compiled a reproducible RMarkdown report outlining key hypotheses and conclusions.

Predicting Stock Prices using CAPM and Fama French Models | *Finance and Analytics Club, IIT Kanpur* | *May'22 - Jul'22*

- Implemented the **Capital Asset Pricing model** and the **Fama French model** (with three and five factors) in Python on various stocks
- Sourced the data from Yahoo Finance API and used Pandas, Matplotlib and Numpy to build statistical models for **time series analysis**

Defining the Market Research Problem | *Course Project* | *May'23 - Jul'23*

Prof. Amit Shukla, Department of Industrial and Management Engineering, IIT Kanpur | MBA631A - Marketing Management

- Presented a **comprehensive analysis** of the market research conducted by **GEF India**, assessing the effectiveness of their **launch strategy**

Text to Image Generation using GANs | *Project Mentor* | *Science & Technology Council IITK* | *Apr'23 - Jul'23*

- Taught a variety of concepts in **deep learning** including **ANNs**, **regression**, **CNNs**, basics of **NLP**, and **GANs** to a cohort of 24 freshmen
- A text-to-image **generative adversarial network** (GAN) model was meticulously developed employing these effective methodologies.

Technical Skills

Programming	Machine Learning	Data Science	Utilities
Python, R, C++, C	PyTorch, Numpy, Pandas, Matplotlib, Scikit-Learn, OpenCV, pytube, pyaudio	Rshiny, Rcpp, Rvest, Tidyverse, ggplot2, plotly, dplyr, imager	git, LaTeX, Markdown, bash, docker

Relevant Courses

**: ongoing courses*

Data Science Lab I, II	Statistical Computing	Theory of Statistics
Statistical NLP	Theory of Statistics	Elementary Stochastic Processes($A*$)
Regression Analysis*	Time Series Analysis*	Fundamentals of Computing
Data Structures and Algorithms*	Introduction to Probability Theory	Marketing Management

Positions of Responsibility

Secretary | *IIT Kanpur Consulting Group* *Aug'22 - May'23*

- Applying the power of data science to initiatives and collaborations with a wide variety of corporations and organizations to promote community welfare and social good by implementing solutions that are both efficient and inventive through the use of **deep learning** and **machine learning**
- Conducted a comprehensive literature review of numerous research papers in the field of AI, health care (smart ICUs), Air Quality Monitors
- Participated in case study contests, including the Global Management Challenge (qualified for round 2) and the IIMB White Paper competition

Secretary | *Dance Club, IITK* *Jun'22 - May'23*

- Participated in a week-long workshop organized by **Professional Artists**, contributing to its conduction and learning diverse forms and styles.
- As a cohort of about 30 individuals, participated in college-level performances such as Thomso'22, IIT Kanpur Orientation, and Antaragni'22

Student Guide | *Counselling service, IIT Kanpur* *Sep'22 - Present*

- Helped as a part of a **team** to conduct a week long Orientation Session for incoming batch consisting of over **1200** students
- Assisting a group of 5 freshmen academically and emotionally, to get acclimatized with the new college environment

Extra-Curricular Activities

- Actively engaged in **boxing training** as a personal pursuit during my leisure hours at college over the course of the past few months.
- I was part of the **winning dance** team in Antargni'22 at IIT Kanpur, as well as the **2nd Runner Up** dance team in Thomso'22 at IIT Rorkee
- Implemented and fine-tuned multiple Neural Networks in **PyTorch**, encompassing diverse models such as RNNs, CNNs, LSTMs, Transformers, and more. Acquired comprehensive understanding through studying influential books like **Dive into Deep Learning by arXiv**