Sweta Kumari

Junior Undergraduate
Materials Science and Engineering
Indian Institute of Technology Kanpur

Email: swetak20@iitk.ac.in (**Homepage**)
Phone: +91-7667570374

Academic Qualifications

Year	Degree	Institute	CPI/%
2020 - 2024	Bachelor of Technology (B. Tech)	Indian Institute of Technology Kanpur	8/10
2020	Class XII (CBSE)	The Pentecostal Assembly School	96.60%
2018	Class X (CBSE)	The Pentecostal Assembly School	97.40%

Honors and Awards

- **She++ Uber**
  - Selected as one of the top 75 female candidates from over 2500+ registrations from 19 of the best engineering colleges in the country.
- **ABinBev Ale:Go-Rhythm 2022**
  - Selected at AbInBev GCC India for a data analyst internship for summer'23 based on the coding tests and recommender system built for the Ale:Go-Rhythm hackathon.

Work Experience

- **Data Analysis Intern,** *ABInBev* (*Nov'22 - Present*)
  - Conceptualize succession plans for people turnover using inferential statistical results on employee's accumulated data.
  - Implement and improve NLP models to predict plans without bias adapting to deficient data.
  - Perform first principal analysis to define health and risk of a pipeline for employees in leadership positions.
- **Software Developer Intern,** *Uber* (*May'22 - Jul'22*)
  - Worked on a service to automate calculation of vendor consolidated spend by integrating with mutliple data sources.
  - Build schema-less table to hold dissimilar data from Hive, GSheets,etc integrated with Orchestrator for dataflow management.
  - Reduced latency for fetching data from google sheets by integrating the Gsheet API to the backend service.
- **UG Fellow Research Intern,** *C3i Hub* (*Aug'21 - Mar'22*)
  - Worked on Cyber Security with the aim to address the issue of cyber security of Cyber Physical Systems in its entirety.
  - Implemented **Peerclear** - A botnet detection tool for IoT devices on attacks including **DDoS, Trojan**, etc using **PyShark** and **Wireshark**.Fabricated a dataset by setting up IoT system and process the pcap files using **Zeek - Bro**.

Key Projects

- **Trading Strategies in Emerging Markets Specialization,** *Coursera , ISB* (Link) (*Aug'22 - Dec'22*)
  Applied ready-made trading strategies, based on rigorous academic research - strategies based on momentum, momentum crashes, price reversal, persistence of earnings, quality of earnings, underlying business growth, behavioral biases and textual analysis of business reports about the company.Pertained basic asset pricing theories to calculate the expected returns of a stock or a portfolio. In the culminating project, developed new trading strategies, based on Pairs Trading evaluated and integrated them with the existing portfolio.
- **P2P Botnet Detection System,** *Mentor: Sr. Research Engineer, Anand Handa, C3i Hub* (Link) (*Nov'21 - Jan'22*)
  - Analyzed botnet network traffic patterns to derive inferences and strategies to prevent intrusion attacks. Proposed and implemented an IDS model which is **content-agnostic** and can also handle **encrypted data packets**.
  - Using **Benign and Malicious P2P** network data packets, built a model that is independent of payload - uses only the header information of **TCP control packets**.
  - For sophisticated P2P traffic categorization, used a more straight forward approach for categorizing P2P traffic like **failed connections threshold, destination diversity threshold,** etc.
  - For feature selection, used a **classification and regression tree method**.
- **Accelarated Computing with CUDA Python,** NVIDIA Deep Learning Institute (Link) (*Sep'21*)
  - The project aimed at implementing fundamental techniques needed to GPU-accelerate Python applications using **Numba**.
  - **Custom CUDA Kernels**, **multidimensional Grids** and **Shared Memory** for CUDA Python were used for initial optimizations.
  - Utilized **grid stride** loops for working in parallel over large data sets and leveraging **memory coalescing**.
  - Used **atomic operations** to avoid race conditions when working in parallel.
  - Used shared memory to coordinate threads within a block and **facilitate coalesced memory access patterns** and **resolved shared memory bank conflicts.**
- **Fundamentals of Deep Learning,** NVIDIA Deep Learning Institute(Link) (*Sep'21*)

- Initially prepared **Image Classification** of an American Sign Language Dataset, then taught our model to be more robust when looking at new data, using **Data Augmentation**.
- Used **Keras** to load a very well-pretrained model. Preprocessed the images to work with the model and performed accurate inference on the images.Performed **transfer learning** with our own small dataset on a pretrained model and further **fine tuned** the model for even better performance.
- Compiled the model with **categorical crossentropy** on a dataset from Kaggle and procured **95.74% accuracy.**

- **Intrusion Detection System,** HCL-Hackathon (Link) (*Dec'21 - Feb'22*)
  - Trained an ML model which takes inputs from the sensor measurements at a time stamp 't' and classify it either in "ATTACK" or "NORMAL" categories using dataset provided.
  - Performed **feature extraction using Principal Component Analysis** to implement dimensionality reduction of the feature vectors. Then developed an **unsupervised time-series ML model** for better accuracy is suggested as first part of the phase.
  - In the second phase, developed a **Linux based malware detection and classification tool** that extracts features from **ELF Files** using **Command-Line Tool - objdump** and parse the output using python script to obtain the dataset.
  - Then build the Detection System by implementing **PCA**, and **Random Forest**.
- **Deep Into CNNs,** Summer long Project, Programming Club, IITK(Link) (*May'21 - Jul'21*)
  - Built Multi layer Perceptron Convolutional Neural Networks **(MLP-CNNs)** on datasets like **CIFAR10, MNIST** and analyse the effectiveness of proposed Networks , and compare their performances.
  - Executed **AlexNet, Inception, Xception, VGG11**, as State-of-the-art models. And optimised the models using **Adam gradient descent, weight initialization, Adagrad**, etc.
  - Applied **Simple, Convolutional and denoising Auto-encoders** for **Unsupervised learning**.
  - Participated in **Tabular Playground Series - Jun 2021** at Kaggle (Link).

Technical Skills

Languages	C/C++, Python3, Golang, Proto, grpc, HTML/CSS/JS, Bootstrap, LaTeX
Softwares/Tools	Github, Postman, Phabricator, MySQL, MATLAB
Machine Learning	PyTorch, Scikit-Learn, Tensorflow, Seaborn

Relevant Coursework

NSM-Computer Architecture Winter School	Linear Algebra and ODE	Neural Networks and Deep Learning**
Amazon ML Summer School	Probability and Statistics	Web Developer Bootcamp 2020 *
Malware Analysis and Intrusion Detection	Partial Differential Equations	Fundamentals of Computing

**: Coursera | *:Udemy

Positions of Responsibility

- **Science and Technology, Secretary** Hall 6, IIT Kanpur (*Apr'22 - Ongoing*)
- **Web Secretary,** Games and Sports Council, IIT Kanpur (*May'21 - Mar'22*)
- **Company Coordinator,** Students' Placement Office, IIT Kanpur (*Jun'21 - Feb'22*)

Extra Curricular

- 2022 — Secured **24th position** in HCL-Hack IITK 2021 among **12496 participants** from 2100 colleges and 12 countries.
- 2021 — Came in **top 15** teams at Techweek conducted by SNT council among more than 100 teams.