

SUNIL DHAKA

BS-MS · MATHEMATICS & SCIENTIFIC COMPUTING
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

☎ (+91) 9079292093 · ✉ sunild@iitk.ac.in · 🌐 home.iitk.ac.in/~sunild · 📄 [sunil-dhaka](#)

EDUCATION

Year	Degree	Institution	Score
2017 – 2023	BS-MS, MnC	Indian Institute of Technology, Kanpur, UP	8.33(PG)/7.6(UG)
2016	RBSE – XII	Navjeevan Public School, Sikar, Rajasthan	96.00%
2014	RBSE – X	Thar Public School, Salasar, Rajasthan	94.67%

SCHOLASTIC ACHIEVEMENTS

- **Scholarship** for Higher Education (SHE) from the Indian Government's **INSPIRE Program**
- Secured All India Rank **1954 in JEE Advanced** 2017 in among around 2 lakh students
- Secured All India Rank **2844 in JEE Mains** 2017 in among around 15 lakh students
- Secured State Rank **8 in 2016 RBSE**, Rajasthan (XII) in among around 2 lakh students

ACADEMIC PROJECTS

Relevant Information Retrieval from Text Documents

📄 [sunil-dhaka/IR-Project](#)

INFORMATION RETRIEVAL, UNDER PROF. ARNAB BHATTACHARYA

Jan '22 - Apr '22

- Implemented text summarization algorithms like **TextRank**, **LSA**, seq2seq architectures like **Google PEGASUS**, **t5-small**
- Preprocessed text corpora using a variety of **NLP techniques** like **tokenization**, **stemming**, **lemmetization**
- Used **ROUGE-N score** to evaluate accuracy of the algorithm summaries against human summaries for **corpus of size 230k**
- Tested out algorithms and **achieved SOTA f1-scores** on our datasets for both abstractive algorithms **after fine tuning**
- Created a **Flask web application** for users to interact with the algorithms and get summaries of articles or blogs

Racial Disparity in COVID-19 Vaccination in US

📄 [sunil-dhaka/covid-disparity](#)

ADVANCED STATISTICAL METHODS, UNDER PROF. SHANKAR PRAWESH

Aug '21 - Nov '21

- Analysed vaccination rates for **covid-19 by race in US counties**, along with their association with socioeconomic status and other factors, using a number of statistical techniques
- Transformed and pre-processed data by **z-score** normalization and **PCA** dimension reduction of feature matrix
- Implemented variety of regression algorithms like **MLR**, **Lasso-Ridge-Elastinet Regression**, **SVR**, **Random Forest**
- **Fine tuned** the hyper parameters of these algorithms using train and test **MSE as metric** of measure

Indian Agriculture Data Mining

📄 [sunil-dhaka/Agriculture-Analysis](#)

DATA MINING, UNDER PROF. ARNAB BHATTACHARYA

Aug '21 - Nov '21

- Analysed **seasonal crop data** for last 20 years and unearthed surprising information about the Indian agricultural
- Pre-processed 2.5 lakh data points using **cleaning techniques** to get robust dataset for subsequent steps
- Developed new categorical variables of zones and crop to make analysis more understandable and comparative
- Used **geopandas** library to discover the geographical situation of Indian agriculture on zone and crop basis

Modelling and Forecasting of Monthly Temperature Time Series Data

📄 [sunil-dhaka/time-series-model](#)

TIME SERIES ANALYSIS, UNDER PROF. AMIT MITRA

Sep '20 - Nov '20

- Modeled a **seasonal ARIMA model** to forecast next years temperatures based on historical data
- Checked data stationarity through **Dickey-Fuller test**, and performed seasonal differencing to ensure stationary TS
- Optimized the model parameters using **Box Jenkins Method** for forecasting purposes
- Performed residual analysis and **information criterion tests** to check model adequacy on the dataset
- Yielded absolute **MAE of 2.2%** on test data for optimized S-ARIMA model

Expectation-Maximization & Metropolis-Hastings Algorithms

📄 [sunil-dhaka](#)

STATISTICAL SIMULATION, UNDER PROF. DOOTIKA VATS

Aug '19 - Nov '19

- Developed **EM algorithms** to fit multivariate **Gaussian mixture models** for latent variable
- Used **MH algorithm** for implementing the **MCMC** method in **Bayesian logistic regression** model
- **Cross Validated** the model on test data to get optimum values of hyper parameters

SELF PROJECTS

Hello IITK Resources Downloader

 [sunil-dhaka/autodownloader-helloiitk](#)

TECH USED: PYTHON, REQUESTS, BEAUTIFULSOUP, PDFKIT, API, JSON

2021

- Created a **python command line tool to automate** the downloading of resources files from the hello iitk portal
- Used requests **sessions object** to keep the user signed in via **cookies**
- Deployed **beautifulsoup parsing and API calls** to collect data linkages and stored course and file-specific data in JSON
- Developed **automation** script to **download** resources files into their respective folders **once a week**
- Downloaded and **organised forum questions** and instructor announcements as PDFs using the **pdfkit tool**

Flask Web Application

 [sunil-dhaka/flask-apps](#)

TECH USED: PYTHON, FLASK, BOOTSTRAP, JINJA3

2022

- Developed a **dynamic website** to display NYT best sellers list as well as book reviews organised by author, title, and ISBN
- Used **Bootstrap** to create responsive front end and employed **Flask** to develop backend component of the application
- Utilized **jinja templates** to render the dynamic data returned by **API calls**

WORK EXPERIENCE

Teaching Assistant for Mathematics

SATHEE, UNDER PROF. AMEY KARKARE

Mar '22 - Jun '22

- Worked as **TA** for **SATHEE project**, an NTA and IIT Kanpur endeavour to increase without coaching students in IITs
- Created **600+ relevant problems** for JEE aspirants to practice, for over 40 Mathematics lectures from grades XI and XII
- Created essential **formulae and tips & tricks** sheets for these lecture videos delivered by IIT professors

POSITIONS OF RESPONSIBILITY

Student Coordinator



RSK, UNDER RITA SINGH, PROJECT EXECUTIVE OFFICER

 [rsk.iitk.ac.in](#)

May '22 - Present

- **Mentored 70+ ninth-graders** of Sri Ram Janki Inter College Bithoor in Mathematics and Physics
- Conducted various **workshops and discussion sessions** with volunteers to enhance subject instructions
- Collaborated with **15+ volunteers** to smoothly conduct regular online and offline classes for students

EXTRA CURRICULAR

- Participated as a **volunteer** in the planning of **athletics competitions** organised by Women Cell and WA, IITK
-  Automating Boring Stuff,  Web Scrapping, Track & Field, **Playing Tennis**, Reading Books

SKILLS

Programming	Python, JavaScript, SQL, R, C
Web	HTML5, CSS, Bootstrap, Node.js, MongoDB
Frameworks	Flask, React, Express
Utilities	Git, \LaTeX , MySQL WorkBench, Linux Shell, Bash Scripting, Ubuntu

RELEVANT COURSES

Fundamentals of Computing	Data Structures and Algorithms	Information Retrieval	Data Mining
Probabilistic Modeling and Inference	Introduction to Machine Learning	Time Series Analysis	Bayesian Analysis
Probability and Statistics	Applied Stochastic Processes	Mathematical Logic	Sampling Theory