

# SUNIL DHAKA

BS-MS · Mathematics & Scientific Computing  
Indian Institute of Technology, Kanpur

✉ sunild@iitk.ac.in · ☎ (+91) 907-929-2093  
🌐 home.iitk.ac.in/ sunild  
📄 sunil-dhaka

## EDUCATION

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

† : Expected year of graduation

## SCHOLASTIC ACHIEVEMENTS

- Awarded, Higher Education **INSPIRE Scholarship**
- Rank 1954**, JEE Adv. (amongst 0.2 million candidates)
- Rank 2844**, JEE Main (amongst 1.5 million candidates)
- Rank 8**, 2016 RBSE, Rajasthan (XII) Exams

## ACADEMIC PROJECTS

### Relevant Information Retrieval from Text Documents IITK 📄 sunil-dhaka/IR-Project *Apr 2022*

- Implemented and fine tuned text summarization algorithms like **TextRank**, **LSA**, **Google PEGASUS**, **t5-small**
- Preprocessed corpora using variety of NLP techniques like **tokenization**, **stemming**, **lemmetization**
- Used **ROUGE-N** to evaluate accuracy of machine summaries against human summaries for **corpus of size 230k**
- Tested out algorithms and **achieved SOTA f1-scores** on our datasets for 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 IME, IITK 📄 sunil-dhaka/ime692-project *Nov 2021*

- Analysed vaccination rates for **covid-19 by race in US counties**, along with their association with socioeconomic 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**, **Support Vector Regression**, **Random Forest**
- Fine tuned** the hyper parameters of these algorithms using train and test **MSE as metric** of measure

### Indian Agriculture Data Mining CSE, IITK 📄 sunil-dhaka/Agriculture-Analysis *Nov 2021*

- 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 Time Series Data MTH, IITK 📄 sunil-dhaka/time-series-model *Nov 2020*

- 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

## 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

## WORK EXPERIENCE

### Teaching Assistant for Mathematics *Mar '22 - Jun '22* SATHEE, under Prof. Amey Karkare

- Worked as **TA for SATHEE project**, an NTA and IITK endeavour to increase without coaching students in IITs
- Created **600+ relevant problems**, essential formulae and tips & tricks sheets for JEE aspirants to practice

## POSITIONS OF RESPONSIBILITY

### Student Coordinator *May '22 - Present*

RSK, under Rita Singh, Project Executive Officer

- Mentored **70+ ninth-graders** 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

## SELF PROJECTS

### Hello IITK Resources Downloader 📄 *Git Repo*

Tech Used: Python, Requests, BeautifulSoup, PDFkit, API, JSON

- 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 **bs4 parsing and API calls** to collect data link-ages and stored course and file-specific data in JSON
- Developed **automation** script to **download** resources files into their respective folders **once a week**
- Organised and **downloaded forum questions** and instructor announcements as PDFs using **pdfkit tool**

### Flask Web Application 📄 *Git Repo*

Tech Used: Python, Flask, Bootstrap, Jinja3

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

## SKILLS

**Programming:** Python, SQL

**Frameworks:** Flask

**Web:** HTML5, CSS, Bootstrap

**Utilities:** Git,  $\text{\LaTeX}$ , MySQL WB, Linux Shell, Ubuntu

## EXTRA CURRICULAR

- Volunteered in the planning of athletics competitions, organised by Women Cell and Women Association, IITK
- 📄 Automating Boring Stuff Using Computers, 📄 Web Scrapping, Track & Field, Playing Tennis, Reading Books