SUNIL DHAKA

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

□ (+91) 9079292093 · Sunild@iitk.ac.in · A home.iitk.ac.in/ sunild · C sunil-dhaka

EDUCATION

Year	Degree	Institution	Score
2017 – 2023	BS-MS, MnC	Indian Institute of Technology, Kanpur,UP	8.33/10.0
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

COURSE PROJECTS

Relevant Information Retrieval from Text Documents

nuil-dhaka/IR-Project

Jan '22 - Apr '22

Information Retrieval, under Prof. Arnab Bhattacharya

- Implemented various extractive text summarization algorithms like text-rank, Latent semantic analysis, and abstractive self supervised seq2seq architectures like Google PEGASUS, t5-small
- Over 230k news articles from both Hindi and English were included in the Text Corpora, along with their own human-written summaries
- Cleaned and preprocessed text corpora using a varity of NLP techniques like noise removal, tokenization, lowercasing, stemming, lemmetization, stop word removal, and object strandardization
- Used ROUGE-N score to evaluate accuracy of the algorithm summarized text against human text summaries for documents in the dataset
- Tested out all the algorithms on both Hindi and English language datasets and achieved SOTA f1-scores on our datasets for both abstractive algorithms after fine tuning
- Achieved SOTA f1-scores on our datasets for both abstractive algorithms after fine tuning
- Worked in a 5 person team to do all the research, presentation and implementation work
- Created a Flask web application for users to interact with the algorithms and get summaries of articles or blogs in Hindi and English languages

Influential Observations in Linear Regression

O sunil-dhaka/influential-points

Linear and Non-Linear Models, under Prof. Satya Prakash Singh

Aug '21 - Nov '21

- Conducted research on and delivered a class presentation on Professor Dennis Cook's paper "Influential Points in Linear Regression"
- Implemented influential observation detection measures like Cooks distance, studentized residual, and deleted studentized residual from the paper using R
- Validated the discovered outlier points by removing them one at a time and observing the impact on residual plots

Indian Agriculture Data Mining

O sunil-dhaka/Agriculture-Analysis

DATA MINING, UNDER PROF. ARNAB BHATTACHARYA

Aug '21 - Nov '21

- Analysed seasonal Indian crop data for last 20 years and extract key insights into crop production and crop type across states
- Collected around 2.5 lakh data points from data-gov servers using multiple API calls
- Cleaned and processed the data using a variety of data 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 a zone-by-zone and crop-by-crop basis
- Unearthed some unexpected information regarding the Indian agricultural state

1

Racial Disparity in COVID-19 Vaccination in US

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 & pre-processed data by z-score normalization and Principal component analysis (PCA) dimension reduction of feature matrix
- Implemented variety of regression algorithms like Multiple Linear Regression(MLR), Lasso Regression, Ridge Regression , Elastinet Regression, Support Vector Regression(SVR), Random Forest
- Fine tuned the hyper parameters of these algorithms using train and test MSE as metric of measure
- Found Covid-19 immunisation rates (1st dose) in US counties to be significantly affected by socioeconomic status and political ideology

Critical Comparative Appraisal of Monetary Policy Tools and Strategies

nuil-dhaka/monetary-essay

sunil-dhaka/covid-disparity

Monetary Economics, under Prof. Sukumar Vellakkal

Aug '21 - Nov '21

- Structural comparison of monetary authorities in India and the United States
- Monetary policy comparison between these countries based on tools like reserve requirements, discount rate, and open market operations
- Critical comparison of various monetary policy strategies these countries have used over the last few decades to control economic variables like unemployment rate, inflation rate, interest rate, GDP, money supply etc

Repelling-Attracting Metropolis Algorithm

O sunil-dhaka/RAM

MCMC, UNDER PROF. DOOTIKA VATS

Jan '21 - Apr '21

- Implemented tweaked version of MH algorithm called Repelling-Attracting Metropolis Algorithm(RAM) for multi modality
- Used Auxiliary Variable approach to derive the steps of algorithm
- Demonstrated how the RAM model outperforms the MH sampler, by generating MCMC samples for real life numerical examples like sensor network localization, strong lens time delay estimation
- RAM provided better simulations for multimodal distributions, as seen by ACF plots, trace plots, acceptance rate and downhill-uphill average proposal numbers of generated samples

Modelling and Forecasting of Monthly Temperature Time Series Data

O 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
- Obtained monthly average temperatures for India during the past 117 years in Celsius from data-gov website
- Checked data stationarity through the augmented Dickey-Fuller test, and performed seasonal differencing to ensure stationary time series
- 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 mean percentage error(MAE) of 2.2% on test data for optimized S-ARIMA model

Expectation-Maximization & Metropolis-Hastings Algorithms

O sunil-dhaka

STATISTICAL SIMULATION, UNDER PROF. DOOTIKA VATS

Aug '19 - Nov '19

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

Mechanized Paper Cutter with Counter

🖸 sunil-dhaka

Manufacturing Processes II, under Prof. Shantanu Bhattacharya

Jan '19 - Apr '19

- Used Fusion 360 Software to design a mechanical paper cutter with a counter mechanism
- Developed the suggested prototype with a team of 7 people

Simple Rube Goldberg Machine

O sunil-dhaka

MANUFACTURING PROCESSES I, UNDER PROF. SUDHANSHU S. SHEKHAR

Aug '18 - Nov '18

• Constructed a 40x40x40 cm simple Rube Goldberg contraption employing fundamental mechanical techniques such as welding, moulding, etc. in the Mechanical lab

SELF PROJECTS

Hello IITK Resources Downloader

nil-dhaka/autodownloader-helloiitk

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

2021

- Create a python command line tool to automate the downloading of files from the hello iitk portal, such as videos and PDFs
- 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 and beautifulsoup libraries
- Been using this utility for over a year for personal use

Flask Web Application

O sunil-dhaka/flask-apps

TECH USED: PYTHON, FLASK, BOOTSTRAP, JINJA3

2022

- Develop a dynamic website using Flask framework to display the New York Times best sellers list as well as book reviews organised by author, title, and ISBN
- Utilized jinja templates to render the dynamic data returned by API calls
- Used Bootstrap to create responsive front end of application
- Employed flask to develop compact backend componant of the application
- Project Demo

WORK EXPERIENCE ___

Teaching Assistant for Mathematics

SATHEE, UNDER PROF. AMEY KARKARE

Mar '22 - Jun '22

- Worked as a teaching assistant (TA) for the SATHEE project, an NTA and IIT Kanpur endeavour to increase the percentage of candidates who are chosen without coaching
- 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

SKILLS ____

Programming Python, JavaScript, SQL, R, C

Web HTML5, CSS, Bootstrap, Node.js, MongoDB, Docker

Frameworks Flask, Django, React, Express

Utilities Git, LTFX, MySQL WorkBench, Linux Shell, Bash Scripting, Ubuntu

RELEVANT COURSEWORK

- **Computer**: Data Structures and Algorithms, Fundamentals of Computing, Topics in Probabilistic Modeling and Inference, Introduction to Machine Learning, Data Mining, Information Retrieval, Bayesian Models & Data Analysis, Computational Linguistics for Indian Languages**
- **Economics:** Introduction to Economics, Macroeconomics-I, Monetary Economics, Financial Market*, Trading Basics*, Trading Algorithms*, Advanced Trading Algorithms*, Creating a Portfolio*, Introduction to Financial Markets*, Investment Strategy*, Behavioral Investing*
- Mathematics: Linear Algebra, Abstract Algebra, Mathematical Logic, Real Analysis, Complex Analysis, ODE, PDE, Several Variable Calculus, Numerical Analysis
- Statistics: Probability & Statistics, Time Series Analysis, Statistical Simulation and Data Analysis, Statistical and AI Techniques in Data Mining, Applied Stochastic Processes, Inference-I, Introduction to Sampling Theory, Bayesian Analysis, Markov Chain Monte Carlo, Linear and Non Linear Models, Non-linear Regression, Statistical Inference**

 *: Coursera, **: Ongoing

POSITIONS OF RESPONSIBILITY

Student Coordinator

rsk.iitk.ac.in

RSK, UNDER RITA SINGH, PROJECT EXECUTIVE OFFICER

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, such as races and walks, organised by Women Cell and Women Association
- Automating Boring Stuff Using Computers, Web Scrapping & Data Hoarding, Track & Field, Playing Tennis, Reading Books