Saumya Gaurang Shah

Senior Year Undergraduate

: *(+91) 7990848092*
: *sshah@iitk.ac.in, saumyagshah@gmail.com*
: *saumyagshah*
: *saumyagshah*
: *saumyagshah*
: *saumyagshah.github.io*

Education

Sep'19- Feb'20 **Semester Exchange in Computer Science**, *École Polytechnique Fédérale de Lausanne, Switzerland*
Relevant Coursework: Natural Language Processing, Mathematics of Data: From Theory to Computation, Algorithms

- Selected for a **semester exchange at EPFL, Switzerland**: **One among 3 students** to be selected from IITK

2016– Present **Bachelor of Technology**, *Indian Institute of Technology, Kanpur, India*
Major: Electrical Engineering
CPI: 9.6/10
Relevant Coursework: Probabilistic Modeling and Inference, Visual Recognition, Machine Learning, Database Systems (ongoing), Probability and Statistics, Linear Algebra and ODE, Data Structures and Algorithms, Computer Organization, Fundamentals of Computing, Control Systems Analysis, Digital Signal Processing (ongoing), Detection and Estimation Theory (ongoing), Nonlinear Systems (ongoing), Partial Differential Equations, Real Analysis

- Received the Academic Excellence Award for 3 consecutive years (2016-2018), awarded to **Top 10%** students
- Awarded A* in 4 courses (for **outstanding performance**) including Complex Analysis

Internships

May'19- Jul'19 **Computer-Human Interaction in Learning and Instruction (CHILI) Lab, EPFL, Lausanne, Switzerland**
Summer@EPFL 2019 Research Intern, Prof. Pierre Dillenbourg, EPFL
report *Diagnosing Dysgraphia Using Handwriting Data Analysis*

- Contributed to a project for diagnosing dysgraphia, a learning disability that affects written expression, within seconds using writing data from a consumer tablet
- Implemented and explored oversampling techniques to enable better learning for our imbalanced dataset
- Implemented an additional 60 new features for our time series data based on position, time, pressure, tilt, age, gender and laterality of the strokes
- Improved prediction accuracies by about 4-5% on the minority dysgraphic class alone in our imbalanced dataset
- Increased interpretability of the diagnosis by providing new discriminative features capturing the irregularity and shakiness in handwriting

Feb'18- May'18 **Auquan, Bengaluru, India**
Data Science Intern, Machine Learning Platform for Financial Services
github *Predicting Stock Prices to Develop Trading Strategies for the NSE stock market index*

- Developed predictive models for stock prices in Python using the fundamentals of quantitative finance research
- Designed, back-tested and optimized a data-driven quantitative trading strategy on real-world data
- Developed an intra-day mean reversion strategy for a partner firm to give greater than 30% return on capital (RoC) using Hurst values and Autoregressive Integrated Moving Average (ARIMA) models

Nov'17- Dec'17 **Kritsnam Technologies, Kanpur, India**
Winter Intern, IoT Company with a Focus on Water Resource Management Solutions
Analysis of Analog Signals from Atlas Scientific Sensors, Prof. Ketan Rajawat

- Explored the working and physical principles of Atlas Scientific pH, Electrical Conductivity (EC), Dissolved Oxygen (DO) and Temperature sensors
- Integrated sensors with Particle Electron Embedded Platform via mobile data (2G) to retrieve data for these sensors using Atlas Scientific EZO Circuits
- Utilized this data to propose a regression technique to cut costs by establishing a relationship between the analog signals received from the sensors and the corresponding physical quantities

Projects

Sep'19- Present **Analysing Relationship Between Writing and Drawing Skills**
Semester Research Project, Prof. Pierre Dillenbourg, Computer-Human Interaction in Learning and Instruction Lab, EPFL

- Calculated correlations between features obtained from drawing and writing among children from different grades
- Implemented and explored data splitting techniques to obtain better performance for predicting dysgraphia using time series data obtained from drawing samples alone
- Compared predictions and correlations across grades to analyse the transferability of handwriting skills from writing to drawing to allow predicting dysgraphia for younger children who have not yet learned to write

May'19- Aug'19 **Julia Seasons of Contributions**
Student Developer, Model Zoo for Turing.jl, Open Source Contribution
blog github
- Implemented the following graph and time series models in Julia using Bayesian inference with the probabilistic programming language (PPL) Turing:
  - Mixed Membership Stochastic Blockmodel
  - Autoregressive (AR(p)) and Moving Average (MA(q)) models
  - Autoregressive Integrated Moving Average (ARIMA(p, d, q)) model
- Wrote blog posts, most of which are published in the Medium publication Towards Data Science, describing the work done in detail

Feb'19- Apr'19 **Scaling Up Gaussian Processes and Learning Kernels from Data for Gaussian Processes**
Course Project for Topics in Probabilistic Modeling and Inference (CS698X), Prof. Piyush Rai, IIT Kanpur
report github presentation
- Explored the approach of scaling Gaussian Processes via inducing point methods
- Implemented various approximations via inducing point methods and compared these methods
- Implemented the learning of Gaussian Process kernels from data using the Spectral Mixture kernel

Feb'19- Apr'19 **Model Zoo for Unsupervised Transfer Learning**
Course Project for Visual Recognition (CS783), Prof. Vinay P. Namboodiri, IIT Kanpur
report presentation
- Developed a model zoo of unsupervised learning algorithms on a vehicle dataset from surveillance cameras at IITK
- Implemented Object Detection, Object Classification, Image Segmentation, Object Tracking, Pose Detection, Super Resolution and Future Frame prediction in an unsupervised manner

Sep'18- Nov'18 **Explainable Machine Learning**
Course Project for Introduction to Machine Learning (CS771), Prof. Piyush Rai, IIT Kanpur
report github presentation
- Developed a web application to explain the prediction of any classifier on the user's dataset using LIME
- Implemented feature visualisation using matrix factorisation by generating adversarial examples using BFGS method
- Studied state of the art techniques for visualising CNNs using Lucid and neuron group methods

Aug'18- Nov'18 **SLAM: Extraction of Backend Data and Comparison of Backend Solvers**
Undergraduate Project, Prof. Ketan Rajawat, IIT Kanpur
report github
- Modified and implemented Google Cartographer to retrieve backend data for our own dataset in the form of raw nodes and edges before being published as visual marker arrays
- Implemented the solvers Toro and g2o on this retrieved backend data
- Compared error metrics for these solvers on the Manhattan3500 and city10000 datasets

May'18- Jul'18 **Multiple Sensors Dataset Repository**
Summer Project, Prof. Ketan Rajawat, IIT Kanpur
github
- Collaborated to record and benchmark a new dataset comprising of multiple sensors for research in short and long term Simultaneous Localisation And Mapping (SLAM)
- Developed a ROS Node in Python to calibrate LIDAR and odometry parameters for this dataset by implementing CSM and calibration packages developed by Andrea Censi
  - Developed a Bash + Python script to automate the entire calibration process
- Implemented and compared the results of Google Cartographer, Hector-SLAM and open-karto

Awards and Achievements

2019 **Team Captain:** Won Deloitte TechnoUtsav 2.0 (national level tech competition with **9500+ participants**) along with 2 team members - Received Cash Award of INR 500,000
- cleanAI: Developed an economically viable machine learning model to predict wind and solar power output from the weather forecast
- Showcased our prototype at Deloitte US India Analytics 2019 at Hyderabad, among other novel AI ideas

2019 **Summer@EPFL 2019 Fellowship**: Awarded to about **1%** students among 1200+ applicants for pursuing a three-month fellowship at EPFL, Lausanne, Switzerland

2016 Ranked in **National Top 1%** (amongst 1,200,000 candidates) in JEE Main

2016 Ranked in **National Top 4%** (amongst 150,000 candidates) in JEE Advanced

Technical Skills

Languages *Proficient:* Python, C, C++, Julia *Comfortable:* Java, SQL, R, Shell(bash), MATLAB/Octave, Verilog, MIPS

Frameworks Pytorch, Keras, TensorFlow, ROS

Data Science Libraries NumPy, Pandas, Scipy, Scikit-Learn

Other Git, Travis CI, LaTeX, Arduino, Particle Electron, Autodesk Fusion 360, Adobe Premier Pro

Miscellaneous

2017 Table Tennis- Member of IIT-Kanpur Table Tennis team; awarded Player of Summer Camp for exceptional performance in Table Tennis