SANSIDDH JAIN

A budding researcher with a keen interest in creating social impact through AI

EDUCATION

M.Tech, Computer Science & Engineering *2017 - 2019 (Expected)*
Indian Institute of Technology, Delhi Overall GPA: 8.45/10
Courses Studied - Machine Learning, Artificial Intelligence, Deep Learning, Theoretical Foundations of Data Science, Database Management, Advanced Data Structures, Functional Programming
∗ - Ongoing

B.Tech, Industrial & Production Engineering *2014 - 2019 (Expected)*
Indian Institute of Technology, Delhi Overall GPA: 7.99/10
Courses Studied - Operations Research, Stochastic Modelling, Statistics, Linear Algebra, Real Analysis, Numerical Computation

WORK EXPERIENCE

Tracking Social Development From Satellite Data Aug 2018 - Present
Advisor - Prof. Aaditeshwar Seth, IIT Delhi *Master's Thesis*

- Working on developing models for estimating social development indicators (employment, literacy, etc) from satellite images and nightlight data with Census 2011 data serving as ground truth
- Trained different deep learning models via transfer learning (developing on Oshri et al (KDD 2018)) to predict labels (extracted from Census 2011) given an input of a satellite image
- Clustered spatial variation of village labels from district hotspots to assign different labels to districts
- Implemented 2-Step Floating Catchment Analysis (2SFCA) to obtain accessibility metric for every village in dataset of 2 Lakh villages

Neonate Weight Estimation From Images May 2018 - Jul 2018
Wadhwani AI, Mumbai

- Developed a computer vision pipeline for estimating volume of an object given a set of images
- Designed for determining whether newborn baby is low birth-weight or not using a mobile phone
- Used Mask R-CNN trained on MS COCO dataset to segment out object from background in image
- Applied Structure-from-Motion (SfM) on segmented images to generate sparse 3D point cloud of object
- Applied Multi-view Stereo (MVS) depth map fusion on above resultant to generate dense point cloud
- Implemented "Touch-Expand" graph-cut algorithm to generate closed surface for volume calculation
- Also advised student team working on estimating soyabean yield from satellite images in Maharashtra

Data-driven Legal Reforms May 2017 - May 2018
Advisors - Prof. Mausam & Prof. Nomesh Bolia, IIT Delhi *Undergraduate Thesis*

- Scraped large dataset (> 15TB) of court cases summaries from Indian district courts. 420+/594 districts (∼ 71%) scraped. Analysed districts across several metrics to identify good and bad anomalies
- Used KL divergence to come up with aggregate measure encompassing performance across all metrics
- Implemented lexical cosine distance clustering to obtain a standardised list of casetypes across courts
- Trained several random forest, gradient boosted trees, and MLP models to predict case duration time
- Trained mixture density networks to output probability distribution over case duration time instead
- Modeled case (sequence of hearings) as a Markov Reward Process with reward being case duration
- Website designed to showcase project work. Presented to senior SC judge Justice Madan Lokur, and NITI Aayog

Wi-Fi Based Location Tracking Jan 2018 - May 2018
Advisor - Prof. Mausam, IIT Delhi *In collaboration with i2e1*

- Developed algorithm to estimate indoor location of multiple devices using solely Wi-Fi RSSI values
- Necessary for algorithm to be scalable, energy efficient, for deployment in several indoor environments
- Initial approaches included triangulation heuristics, and data agnostic signal propagation models
- Experimented with fingerprinting models; supervised nature inhibited scalability
- Trained unsupervised multinomial Hidden Markov Model (HMM) and unsupervised Gaussian HMM
- Developed 2 Kalman Filter models - one before RSSI to distance conversion, one after

Deep Learning Aug 2016 - Nov 2016
Instructor - Dr. Raghavendra Singh, IBM Research Delhi

- Designed a Neural Network which analyses simulated hadron collision data, to predict whether particular collision results in interim exotic particle formation (Baldi, et. al. (2014))
- Performed ablation studies on activation unit, learning algorithm, and model depth to get best-fit model
- Designed and trained a CNN for classifying apparel images. Also trained a Siamese CNN for returning the visually similar images for an input image from the apparel data (Bell and Bala (2015))
- Designed and trained LSTMs which correct grammatically incorrect sentences

Machine Learning Feb 2018 - May 2018
Instructor - Prof. Parag Singla, IIT Delhi

- Trained multiple Naive Bayes classifiers on the Large Movie Review Dataset to predict movie ratings from movie reviews. Alterations included incorporating TF-IDF frequencies, and character bigrams
- Trained decision trees and random forest classifiers on US Census data, to predict whether individual earns more than $50,000 a year
- Trained k-means, PCA-SVM, NN, and CNN classifiers to predict object given it's hand-drawn sketch
- Above dataset posted as internal Kaggle competition. Positioned 23/160 in leaderboard

Artificial Intelligence Jul 2018 - Nov 2018
Instructor - Prof. Mausam, IIT Delhi

- Used Minimax with advanced heuristics to develop AI bot for two-player abstract strategy game Yinsh
- Experimented with Q-Learning and Monte Carlo Tree Search (MCTS) based approaches
- Modeled BlackJack as a Markov Decision Process (MDP) to calculate strategy for every possible hand
- Solved problem of finding subgraphs in graph given constraints by converting it to Boolean SAT problem

Miscellaneous Computer Science Projects

- Developed agri-commodity price prediction system for 20+ commodities and 3200+ markets in India
- Designed and developed an information automation and management system for the IIT Delhi hospital
- Used Lex/Yacc to design an HTML-to-Latex Converter via synthesis of Abstract Syntax Tree
- Developed enhanced version of Paratrooper game in Unity with modern graphics and modified controls
- Modified Firefox source code to incorporate features such as customized tab-management

TECHNICAL SKILLS

Python	Keras, TensorFlow, Pandas, Numpy, Matplotlib, Scikit-Learn
Other Programming Languages	C/C++, Java, OpenMP, R, Prolog, SML
Web Development	HTML, CSS, Javascript (Leaflet, Plotly)
DBMS	MySQL, PostgreSQL
Other Tools	MATLAB, Shell scripting, HPC Systems