ADITYA CHINDHADE

LinkedIn: **adityachindhade** | GitHub: **adityagc** | Email: **adityagc@cmu.edu** | Phone : **+1 (412) 596-3250**

EDUCATION

Carnegie Mellon University | Pittsburgh, PA   *December 2018 (expected)*
*MS (Research) Chemical Engineering (Focus on **Machine Learning** and **Mathematical Optimization**)* | GPA: 3.62/4
Thesis: Structured sparsity-based regularization techniques for group-wise feature selection and estimation.
Graduate Coursework: Cloud Computing* | Data Structures | Java Programming | Machine Learning | Computational
Methods (Optimization) | Mathematical Modeling | Computers in ChemE (Fortran and Parallel Programming) (*=ongoing)

Birla Institute of Technology and Science (BITS) | Pilani, India   *May 2017*
BE (Honors) Chemical Engineering | GPA: 8.59/10
Thesis: Design of experiments and data-driven modeling of fouling in a heat-exchanger.
Undergraduate Coursework: C and Shell Programming, Probability and Statistics, Numerical Methods for engineers.

SKILLS

Languages: Python, Java, R, MATLAB, Octave, Fortran, Scala, Q (Kx), Shell, C, HTML, XML.
Data Science/Machine Learning: Scikit-Learn, Keras, Pandas, TensorFlow, Weka, Ggplot, Matplotlib.
Big Data/Distributed Systems: Spark-TS-Flint (Two-Sigma), Spark-TS (Cloudera), Kafka, Kafka Streams,
Airflow, AWS (S3, EC2, Kinesis), Hadoop MapReduce, Docker, OpenAcc, OpenMP, MPI.
Databases: MySQL, SQLite, MongoDB, KDB+, TimeScaleDB.

PROFESSIONAL EXPERIENCE

Insight Data Science   New York, NY
Data Engineering Fellow (ETL pipeline)   *Fall 2018 - Present*
- Built a scalable data pipeline with batch processing, which involved extraction of financial data from an API, storage using S3 as data lake, followed by time-series processing using Two-Sigma's Flint and Cloudera's Spark-TS libraries; consequently storing data in Timescale and KDB+ databases for querying statistical metrics for S&P 500 companies.
- Presently automating the data extraction, transformation and loading using Airflow and comparing with stream processing approaches using Kafka, Kinesis and Kafka Streams in terms of scalability, elasticity and fault tolerance.
- Consulting a major financial company in New York City by developing a unified analytics platform for data processing.

Sahinidis Optimization Group, Carnegie Mellon University   Pittsburgh, PA
Research Assistant (Machine Learning)   *Spring 2018 - Present*
- Benchmarked feature selection techniques like the group-lasso and elastic-net using non-linear optimization solvers.
- Achieved 10% improvement in accuracy with ALAMO, an integer programming tool, while reducing model complexity.
- Achieved significant speedups in predictions by making a 4-layer Deep Neural Network sparser by 65%, using the Group-Lasso regularization which is a deterministic equivalent of the Dropout.

Harvard-MIT Health Sciences and Technology, MIT   Cambridge, MA
Summer Intern (Statistical Analysis)   *Summer 2016*
- Ran experiments to test bio-activity of anti-cancer drugs; followed by cleaning and visualization of spectroscopy data.
- Modeled biological drug release in R and identified key features using ANOVA and Forward Stepwise Regression.

PROJECTS

Benchmarking Optimization Methods for Deep Convolutional Neural Networks   Fall 2017
- Trained a 7-layer deep Convolutional Neural Network (CNN) on 60,000 grayscale images using Keras with TensorFlow.
- Benchmarked first-order gradient methods like SGD w/o momentum, Adam and AdaDelta with various batch sizes; concluded that Adam converged the fastest with an accuracy of 99.3%, followed by SGD, AdaDelta and SGD+momentum.

Time-Series Analysis for Identification of Irregular Sleep Patterns | CMU Auton Lab | Philips   Spring 2018
- Implemented a rolling-window approach for feature selection on 22 GB of EEG data on a supercomputing cluster; trained a classifier for predicting occurrence of irregularities in sleep patterns and visualized ROC.
- Publication / Preprint: **Chindhade, A.**, Alshi, A., Bhatia, A., Dabhadkar, K., Menon, P. S. (2018). *A machine learning model for identifying cyclic alternating patterns in the sleeping brain.* arXiv preprint arXiv:1804.08750 (2018).

Semantic Classification of Electronic Health Records | North American Central Cancer Registry   Summer 2018
- Developed a system to automatically detect cancer types from raw pathology reports which was hosted as a web-app.
- Built a custom bag-of-words from medical ontologies; trained an ensemble of models such as Naive-Bayes, Logistic Regression and Random Forests to predict cancer type with an accuracy of 94.09 % and F-score of 0.936.

Mapping Neuron Firing to Physical Motion using Neural Networks | CMU BrainHub | Google   Summer 2018
- Performed exploratory data analysis to illustrate the difference in neuron activity between the Cortex and Striatum.
- Trained a multi-layer perceptron to predict movement from neuron activation data with an accuracy of 92.22%.
- Identified clusters of neurons that fire simultaneously, thus helping in detection of neurosignaling pathways.