# KEDAR DABHADKAR

dkedar@cmu.edu | (734) 819-0242 | linkedin.com/in/dkedar7 | dkedar7.github.io

## EDUCATION

**Carnegie Mellon University | GPA: 3.65/4.0** — **Pittsburgh, PA**
Master of Science in Chemical Engineering — **Dec. 2018 (expected)**
(Specialization: **Data Analytics, Machine Learning**)

**Relevant Coursework:**
Process Systems Modeling (06-665), Introduction to Machine Learning (10-601), Computer Science in Chemical Engineering (06-611), Applied Data Science (16-791)*, Introduction to Deep Learning (11-785)*. **Currently enrolled*

**Institute of Chemical Technology** — **Mumbai, India**
Bachelor of Chemical Engineering — **May 2017**

## SKILLS

**Programming Languages:** *Proficient*: Python, R, SQL; *Intermediate*: JAVA; *Basic*: Bash, FORTRAN, C++, HTML.
**Software:** Apache Hive, Apache Spark, Tableau, MATLAB, GAMS, ALAMO. **Databases:** PostgreSQL, MySQL, MSSQL.
**Packages:** Pandas, TensorFlow, PyTorch, Keras, scikit-learn. **Cloud Platform**: AWS EC2.

## WORK EXPERIENCE

**Carnegie Mellon University and Air Liquide S.A.** — **Pittsburgh, PA**
*Graduate Researcher, Master's thesis, advised by Prof. Nikolaos Sahinidis* — **Jan 2018-present**

- Collaborated with Air Liquide S.A. to develop statistical methods to model performance of chemical reactors.
- Modeled process data using ARIMA with exogenous parameters (ARIMAX) and NARX time-delayed neural networks.
- Assisted transition to data-driven methods from traditional methods by laying down the framework and data pipeline.
- Was presented (on my behalf) at **'Big Data and Process Engineering: Opportunities and Limits', Paris**, France.

## PROJECTS

**Analysis of Medical Records of Cancer Patients Using Natural Language Processing |** Python — **Pittsburgh, PA**
*Third Prize, Hackathon, North American Association of Central Cancer Registries (NAACCR)* — **June 2018**

- Analyzed Electronic Medical Records (EMRs) of 10,000 cancer patients to classify according to cancer site.
- Scored an average F1 score of 0.91 on held-out data with an ensemble of Naïve Bayes, Random Forests and SVM.
- Presented the applicability of such a system in practice to a group of physicians by demonstrating a web application.

**Pattern Recognition in Electroencephalogram (EEG) of the Brain |** Python | MATLAB — **Pittsburgh, PA**
*First Prize, Hackathon, Auton Lab, Carnegie Mellon University and Phillips* — **March 2018**

- Cleaned, pre-processed noisy EEG data to induce stationarity and transformed into a sequential window matrix.
- Detected the occurrence of Cyclic Alternating Pattern (CAP) with an accuracy of 58% using logistic regression.
- Published the methodology and results in an academic publication (https://arxiv.org/abs/1804.08750).

**Time Series Analysis of Currency Valuation |** Python — **Fall 2017**

- Implemented time-series analysis, descriptive statistics, various smoothing and stationarity induction methods, and auto-correlations to analyze valuation of the Indian National Rupee against the US Dollar.
- Employed web-scraping to perform live one-day-ahead predictions.
- Achieved a mean squared loss of 0.05 with ARIMA and improved it to 0.03 using an LSTM Recurrent Neural Network.

**Ideal Locations of Fire and Rescue Stations in Montgomery County, MD |** Python — **Summer 2018**

- Inspected the current locations of all fire and rescue stations in Montgomery County, MD.
- Gathered and analyzed the information of about 300,000 9-1-1 calls arising within the bounds of the county.
- Recommended revisions to the locations of stations to improve response time by about 5 seconds using clustering.

**Extracting Information from Text |** Python | AWS — **Spring 2018**

- Built feature engineered logistic regression models to extract information from about 50,000 sentences.
- Deployed an AWS EC2 p2.xlarge instance to handle heavy computations and got an average F1 score of 0.75.
- Improved the average F1-score from 0.60 to 0.75 by feature modification using the time-delay technique.

**Data-driven Scheduling of a Manufacturing Facility |** GAMS — **Fall 2017**

- Improved process scheduling by hypothesizing and adopting a global optimization strategy.
- Formulated the MINLP problem on GAMS and trained constrained models using operation data.
- Cut down costs by 30% and summarized the findings in a detailed report.

## AWARDS AND LEADERSHIP

**Narotam Sekhsaria Foundation Postgraduate Scholarship** — **June 2017**
**Treasurer, General Student Body, Institute of Chemical Technology** — **July 2015-July 2016**
**Event coordinator, ICT Marathon** — **Dec 2015**