Kedar Dabhadkar

2615 First Avenue, Seattle, WA 98121, USA
kedar.dabhadkar@gmail.com • +1 (724) 931-0383 • https://dkedar7.github.io

EXPERIENCE

Revel Consulting, Seattle, WA
Associate Data Scientist Mar 2019 – present

- Microsoft | Information Protection and Governance
  - Provided Microsoft managers key insights into managing high-risk supplier engagements and proposed solutions to reduce the risk of non-compliance using Natural Language Processing (NLP), applied machine learning, and data visualizations in Power BI.
  - Analyzed textual comments using unsupervised NLP to extract themes and understand root causes of engagement risk.
  - Developed machine learning predictive models to proactively identify the type of risk associated with engagements.
- Microsoft | Proactive Recommendations
  - Recommendation engine proof of concept to proactively recommend workshops, seminars to thousands of Microsoft customers.
  - Addressed cold-start problems by matching customer attributes, comments in the open-cases with product description and past recommendations.
- Revel | Internal Projects
  - Actively contributed to building and improving Data Science and Big Data Engineering technical screens using predictive modeling techniques and tools like Docker, PySpark.
  - Conceptualizing positioning of Revel as a leader in data science and AI consulting by identifying key use-cases and building proof of concepts in computer vision, natural language processing and time series analysis.

Carnegie Mellon University and Air Liquide S.A., Pittsburgh, PA
Graduate Researcher Dec 2017 – Feb 2019

- Air Liquide S.A. | Data-driven Modeling of Reactor Temperature Profiles
  - Collaborated with Air Liquide S.A. to develop statistical methods to model performance of reactors.
  - Modeled process data using ARIMA with exogenous parameters (ARIMAX), NARX time-delayed neural networks and LSTM Recurrent Neural Networks (RNN).
  - Empowered Air Liquide engineers with the capability to forecast performance of two key reactors.
  - Assisted transition to data-driven methods from traditional methods by laying down the framework and building data pipeline.
  - Presented this work at ‘Enterprise-Wide Optimization’, Pittsburgh, PA and it was also presented, on my behalf, at the conference ‘Big Data and Process Engineering: Opportunities and Limits’, Paris, France.

EDUCATION

Carnegie Mellon University, Pittsburgh, PA, USA

- Master of Science (M.S.). in Chemical Engineering Aug 2017 – Dec 2018
  - Thesis: Data-driven Modeling of Reactor Temperature Profiles
  - Advisor: Prof. Nikolaos Sahinidis
  - Focus: Sequential Data Modeling, Time Series Analysis, Data Science, Deep Learning.
  - Cumulative GPA: 3.74 / 4.00

Institute of Chemical Technology, Mumbai, India

- Bachelor of Chemical Engineering (B.Chem.Engg) Aug 2013 – May 2017
  - Final Project: Techno-economic Analysis of Condensing Water From Air
  - Advisor: Prof. V. H. Dalvi

Relevant Coursework:
Process Systems Modeling (06-665), Introduction to Machine Learning (10-601), Computer Science in Engineering (06-611), Applied Data Science (16-791), Introduction to Deep Learning (11-785), Mathematical Modeling in Chemical Engineering (06-623), Computational Methods (06-606), Project Economics (CET 1504).

PRESENTATIONS **CONFERENCES**

[1] K. Dabhadkar, K. Ingale, N. Sahinidis, E. Ors, and M. Mighani "Data-Driven Modeling of Reactor Temperature Profiles," in *Enterprise-Wide Optimization*, Pittsburgh, PA, USA, Nov 2018.

SKILLS

Programming Languages:
Proficient: Python, R, SQL; Intermediate: JAVA; Basic: Bash, FORTRAN, C++, HTML.
Software: Apache Hive, Power BI, Tableau, MATLAB, GAMS.
Databases: PostgreSQL, MySQL, MSSQL.
Packages: Pandas, TensorFlow, PyTorch, Keras, scikit-learn, PySpark.
Cloud Platforms: AWS: Entire Stack; Azure: ML Services, ML Studio, VM; GCP: VM.
Certifications: AWS Certified Cloud Practitioner, Jul 2019

PROJECTS

Analysis of Medical Records Using Natural Language Processing , Pittsburgh, PA Jun 2018
Third Prize, Hackathon, North American Association of Central Cancer Registries (NAACCR)
- Analyzed Electronic Medical Records (EMRs) of 10,000 cancer patients to identify tumor site.
- Scored an average F1 score of 0.91 on held-out data with an ensemble of Naïve Bayes, Random Forests and SVM.
- Presented applicability of such a system in practice to a group of physicians by demonstrating a web application.

Pattern Recognition in Electroencephalogram (EEG) of Brain, Pittsburgh, PA Mar 2018
First Prize, Hackathon, Auton Lab, Carnegie Mellon University and Phillips
- Cleaned, pre-processed noisy EEG data to induce stationarity and transformed into a sequential matrix.
- Detected occurrence of Cyclic Alternating Pattern (CAP) with an Area Under the ROC (AUROC) of 58% using logistic regression.
- Published methodology and results in an academic paper (https://arxiv.org/abs/1804.08750).

Extracting Information from Text, Pittsburgh, PA Nov 2017
- Built feature engineered logistic regression models to extract information from 50,000 sentences.
- Deployed the program on an AWS EC2 p2.xlarge instance to parallelize computations using GPU.
- Improved average F1-score from 0.60 to 0.75 by feature modification using time-delay technique.

Speaker Verification, Pittsburgh, PA Oct 2018
- Trained a Convolutional Neural Network (CNN) for speaker identification on 125 GB speech corpus.
- Extracted embedding from penultimate fully connected layer and compared speakers based on cosine similarity.
- Achieved EER (Equal Error Rate) of 13.7 % on held-out data.

Time Series Analysis of Currency Valuation, Pittsburgh, PA Nov 2017
- Implemented time-series analysis, descriptive statistics, various smoothing and stationarity induction methods, and autocorrelations 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.

AWARDS & SCHOLARSHIPS

- Third Prize, Cancer Informatics Hackathon, NAACCR 2018 Annual Conference Jun 2018
  Natural Language Processing Model for Analysis of Medical Records.
- First Prize, HackAuton, Auton Lab, Carnegie Mellon University and Phillips, Mar 2018
  Pattern Recognition in Electroencephalogram (EEG) of Brain.
- Narotam Sekhsaria Foundation Postgraduate Scholarship, May 2017
  For academic and co-curricular excellence

PROFESSIONAL AFFILIATIONS

Society for Industrial and Applied Mathematics
- Member 2017 – Present

CAMPUS ACTIVITIES

Treasurer, Technological Association, Institute of Chemical Technology 2015 – 2016
Event Coordinator, ICT Marathon, Institute of Chemical Technology 2014 – 2015
Coordinator, Awaaz- The Social Drive, Institute of Chemical Technology 2014 – 2015