

Raka Dalal

ASPIRING DATA SCIENTIST/ MACHINE LEARNING ENGINEER

☎ (443)515-7912 | ✉ rak2@umbc.edu | 🏠 rakadalal1409.wordpress.com | 📷 RakaDalal | 🌐 rakadalal

Education

University of Maryland Baltimore County- Baltimore, MD

Aug 2016 - Aug. 2018

MASTERS IN COMPUTER SCIENCE

GPA: 4.0/4.0

- **Relevant Courses:** Algorithms, Machine Learning, Natural Language Processing, Database Systems

Jadavpur University - Kolkata, India

Aug 2012 - May 2016

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

GPA: 8.16/10.0

Online courses (Udemy)

Data Science A-Z, Deep Learning A-Z, Hive for Processing Big Data, Complete & Practical SAS, AWS Concepts, AWS Essentials

Technical Skills

Languages Python(Most Experienced), C, C++, C#, Java, MATLAB, Octave

Tools/Frameworks PyTorch, Keras, TensorFlow, Scikit-learn, NumPy, Pandas, NLTK, Gtrel, **ETL Tools:** SAS, SSIS, **Big Data:** Hadoop, Hive

Data Visualization Tableau, Microsoft Excel, **Web Visualization:** Flask

Database Concepts SQL, PL/SQL, PostgreSQL

AWS Services IAM, VPC, S3, EC2, RDS, SNS, CloudWatch, ELB, Auto Scaling, Route 53, Lambda

Experience

UMBC

Baltimore, MD

GRADUATE RESEARCH ASSISTANT

Jan 2017 - May 2018

- Created an end-to-end pipeline that uses a semi-supervised bootstrap learning model to automatically extract different relations from cybersecurity text with limited training samples and populate a knowledge graph. We evaluated our model on the CVE dataset and achieved high accuracy. [[Github](#)]
- Created an end-to-end pipeline to automate the process of extracting semantic relation from sensor input by formalizing knowledge as digital twin models coming from sensors in industrial production lines and by introducing a semantic query mechanism. We published this work at IKG 2017. [[Publication](#)]

General Electric

GRC, Niskayuna, NY

RESEARCH FELLOW INTERN

May 2017 - Aug 2017

- I built a semi-supervised bootstrap learning based approach to extract relations from unstructured text with an iterative client feedback loop. Evaluations over diverse datasets, including aircraft engine maintenance records and a Google relation extraction corpus, showed promising results. I also designed a user Interface using Python Flask to showcase my work and facilitate users to give feedback.

Samsung Research & Development

Bangalore, India

SUMMER RESEARCH INTERN

May 2015 - July 2015

- Detected the level of stress of user with high accuracy using raw accelerometer and gyroscope data. Filtering and Principal Component Analysis were applied on the data to get the resultant signal. Fast Fourier Transform of this final data gave the heart rate which was used to infer the stress level.

IIT Kharagpur

Kharagpur, India

SUMMER RESEARCH INTERN

May 2014 - July 2014

- The project involved analysis of GPS data of trucks to detect hotspots based on stopping time. We clustered the data using Density-Based Spatial Clustering (DBSCAN) and subsequently typecast the data with respect to stopping time distribution, busyness distribution and finally visualized them on Indian roads.

Major Projects

Auto-Completion System [[Github](#)][[Blog Post](#)]

Aug 2018 - Sept 2018

- The project involves building an auto-completion system for customer service representatives by suggesting sentence completions. For this I used the trigram Katz Backoff model along with some other novel ideas. Finally, this system was wrapped in a local HTTP server.

House Price Prediction [[Github](#)] [[Kernel](#)]

Aug 2018 - Sept 2018

- I worked on the Ames Housing dataset which consisted of 79 explanatory variables describing every aspect of residential homes in Ames, Iowa. The training data has 1461 rows of 81 variables. After extensive feature engineering I predicted the final price of each home by using XGB Regressor.

Prediction of academic references for Wikipedia articles [[Github](#)]

Jan 2015 - Aug 2016

- Crawled wikipedia for academic CS articles starting from Wikipedia Books to gather interesting statistics about the references in wikipedia articles till 2012. The data was cleaned, analyzed and finally, a statistical model was built using N-grams to predict references that will be added in the future.

Intelligent Cybersecurity Recommender (UMBC)

Feb 2017 - Apr 2017

- Designed an effective recommender system that analyzes the probable vulnerabilities and recommends the least vulnerable products. It works on question-answer(Q/A) based model where the system admin is expected to ask product vulnerability questions to which the system suggests a right option.

Detecting Stance in Tweets (UMBC)

Feb 2017 - Apr 2017

- Reproduced the work by Saif M. Mohammad et. al. in the paper SemEval-2016 Task 6: Detecting Stance in Tweets. Designed a system that for any given tweet and target entity, determines whether the tweeter is in favor of the given target, against the given target or neutral to it.

Web Server Development of Jadavpur Server for CMATER (Jadavpur University)

Feb 2014 - Apr 2014

- Built a web server application where user inputs two or more uniprot id. The fasta sequence is fetched from the uniprot server, parsed and subsequently provided as input into a SVM. The output is parsed and the interaction network is plotted. The image is then sent back to client for visualization.