Raka Dalal

Aspiring Data Scientist/ Machine Learning Engineer

253 Amherst Road Apt A2, Sunderland, MA 01375

(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) *June 2018 - Aug 2018*

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

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.

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 clustured 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.

Publications

Generating Digital Twin models using Knowledge Graphs for Industrial Production Lines *Troy, NY, USA*

Industrial Knowledge Graphs, co-located with the 9th International ACM Web Science Conference 2017.

- We introduced a simple way of formalizing knowledge as digital twin models coming from sensors in industrial production lines by presenting a way to extract and infer knowledge from large scale production line data. We also enhanced manufacturing process management with reasoning capabilities by introducing a semantic query mechanism. We created an end-to-end pipeline to automate the process of extracting semantic relation from sensor input.

Major Projects

Auto-Completion System [Github] *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] *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 2016 - July 2017*

- 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.

Technical Skills

Languages	Python(Most Experienced), C, C++, C#, Java, MATLAB, Octave
Tools/Frameworks	PyTorch, Keras, TensorFlow,NumPy, SciPy, Pandas, NLTK, Gretl, ***ETL Tools***: SAS, SSIS, ***Big Data***: Hadoop, Hive
Data Visualization	Tableau, Microsoft Excel, ***Web Visualization***: Flask
Database Concepts	SQL, PL/SQL, PostgreSQL