SIDDHANTH UNNITHAN

4th year Systems Design Engineering Student

sidunnithan@gmail.com
416 - 660 - 8773
github.com/SiddhanthUnnithan

EXPERIENCE

Data Insights Engineer, Flatiron Health — Summer 2017

- Worked closely with community oncology practices to design and develop data infrastructure for Value-Based-Care analytics products. Optimized data consolidation workflows and improved overall data quality
- Built application framework, via Flask, Pandas, and scikit-learn, to streamline the development of biomarker classification models
- Led weekly calls to manage product needs of multiple clients, and conducted onsite visits for user research and training sessions

Data Scientist, Capital One Data Science Lab — Fall 2016

- Pioneered merchant scoring model allowing credit card holders to determine the likelihood of fraud using Redshift, Hive, Pandas, and Boruta
- Constructed an application deployment framework to enable public release of features to 700+ mobile users, using Terraform and Docker

Data Engineer, WatrHub Inc. — Winter 2016

- Designed and implemented data model for petabyte storage of permit documents and utility data using S3, MongoDB, and PostgreSQL
- Built document search engine to perform content searching and dataset filtering using Elasticsearch and Kibana
- Wrote 23 web scrapers to consolidate wastewater utility data from 13 states across the USA
- Designed, developed, and benchmarked Python ETL pipelines

Analytics Developer, TD Asset Management — Fall 2014 & Summer 2015

- Built financial security modelling tool with real-time and historical stock/bond data-retrieval capabilities
- Enhanced portfolio optimization tool through the implementation of new cash-flow and key-rate duration constraints
- Developed tool to automate creation and aggregation of LDI curves

PROJECTS

Bot-Engine

- Developed a Python framework enabling non-developers to rapidly prototype and deploy Facebook Messenger bots through a simple JSON specification

Bosch Production Line Performance - Kaggle Competition

- Built model using Random Forest classifier to predict at-risk parts of Bosch production line with precision of 89.2%
- Analysed feature set and formulated data cleaning methods to improve pre-processing step for continuous features

Allstate Claims Severity - Kaggle Competition

- Wrote a regression model to predict insurance claims' severity using Support Vector and Random Forest estimators with ensemble-based feature selection

REDEFINE - Waterloo DECA Conference

- Led a team of 20 towards organization of national business and technology conference attended by 150 students and 48 industry professionals
- Raised $27,550 in sponsorship funding (28% increase from 2015)

SKILLS

Programming Languages
Python, C++

Analytics and ML Tools
Elasticsearch, Kibana, Pandas
Numpy, scikit-learn, BorutaPy

Databases
PostgreSQL, MongoDB, MSSQL
Hive, Redshift, Teradata

Infrastructure
Docker, Terraform
AWS – EC2, S3, ELB, EMR

Web
HTML/CSS, JS, Dojo
Django, Flask, Apache Nutch

EDUCATION

Systems Design Engineering
University of Waterloo
Class of 2018
Dean's Honours List - Winter '17

RELEVANT COURSES

Machine Intelligence
Applied Linear Models
Data Structures and Algorithms
Operating Systems

EXTRACURRICULARS

Co-President
University of Waterloo DECA
May 2015 - April 2016

HACKATHONS

Deep Health Hackathon
Gift The Code

INTERESTS

Podcasts
Kaggle Competitions
Science Fiction Novels