

mgalarny@eng.ucsd.edu
(858) 999-7481
La Jolla, CA

Michael Galarnyk

github.com/mGalarnyk
linkedin.com/in/michaelgalarnyk
mgalarnyk.github.io

EDUCATION

University of California, San Diego

M.S. Computer Science (Data Science)

2017

B.S. NanoEngineering, Minor Biology

2015

SKILLS

Programming:	Python (pandas, sklearn, PySpark), SQL, MATLAB, Bash Shell Scripting, JavaScript, HTML, CSS, C++
Database Management:	Amazon AWS (S3, EC2, EMR), MSSQL, PostgreSQL, MySQL
Operating Systems	Linux (Ubuntu), Windows, Mac
Other Technologies:	Hadoop, HBase, Spark (PySpark), Tableau, IPython, Jupyter, Git, SolidWorks, AutoCAD

EMPLOYMENT

AWS/Python Data Science and Architecture Intern at Chicago Mercantile Exchange

2016

- Utilized AWS (S3, EC2, EMR), Git, Python (Pandas, SciPy, PyQt), and Spark to model large-scale time series data
- Regression, Clustering, and Distribution Fitting for the Discovery of Fraud.
- Utilized Tableau for Data Visualization

DUV Systems Engineering Intern (Data Analysis) at Cymer

2015

- Designed MATLAB GUI tools to automate SQL queries (MSSQL) and to automatically generate data reports.
- Reduced average SQL query time by ~27% for Cymer GUI tools.
- Presented to CEO and earned my coworkers recognition at the Cymer All-Hands Meeting

Analyst at UCSD NanoBioElectronics Lab

2013-2015

- 9 coauthored publications in peer reviewed journals, 145+ citations (h-index: 7), 1 publication featured on cover
- Gathering and analyzing data using MATLAB and Python (NumPy, Pandas)
- Graphic Design/SolidWorks Modeling for schematic illustrations in peer-reviewed journals and new organizations (BBC, Nanowerk)

RELEVANT COURSEWORK SAMPLE

Python for Data Analysis	Machine Learning
Probability and Statistics using Python	Data Analysis using Hadoop and Spark
Data Management Systems (DBMS)	Python for Informatics
Tableau for Data Visualization	Data Integration and ETL

TECHNICAL WORK SAMPLE

Text Mining Twitter with Spark (PySpark)

- Cleaned data using Regex and performed sentiment analysis on 20GB of raw Twitter Data on US elections
- Carried out analysis using TF-IDF on Twitter key words to estimate preferred candidate for each group

ETL & Uplift Modeling on IRI Dataset (Consumer Goods Dataset)

- Extracted, filtered, and merged 3649 files totaling 140GB into a single 1.8GB HDF5 file
- Performed clustering analysis on customer transactions on yogurt sales for market segmentation

ETL & Analysis of Ebola Data

- Scraped and cleaned Ebola data from the World Health Organization
- Python (NumPy, Pandas, matplotlib) based analysis on the data

Relational Database for Sales Data

- Designed and optimized database schema using E/R diagrams (Cube Schema)
- Created database on Amazon AWS instance (EC2)
- Reduced query and maintenance cost by using optimized relational algebra, indices, and materialized views.

Pokemon Battle Game

- HTML, CSS, JavaScript, and JQuery
- Game: mgalarnyk.github.io/web_development/pokemon/index.html