

AbderRahman Sobh

Data Scientist / Software Engineer

187 Acalanes Dr
Sunnyvale, Ca 94086
(217) 979-4127
abbysobh@gmail.com

EXPERIENCE

Software Engineer - University of Illinois at Urbana-Champaign

Jan-2012 - PRESENT

Data Science Platform:

Designed data science pipelines using Jupyter Notebooks, Python ML libraries, and Spark to provide single-click analysis tools for educating future data scientists.

Built distributed machine learning tools using Python, PySpark, and Pandas which simplifies the use of NLP and Text feature extraction.

Optimized AWS Elastic MapReduce by leveraging Spark and increasing the distributed performance.

Collaborated with Full Stack developers to integrate data science pipelines using Docker containers to encapsulate various data science software stacks providing end-users with quick usage of data-science tools and visualizations.

Simulation Workflow Development:

Collaborate and integrate software applications within openVZ containers providing rapid deployment and accessibility of custom built GUI within Nanobio portal webapp.

Develop graphical user interfaces using TCL and RAPPTURE (XML framework) for the containerized applications to ease the usage of applications for portal end-users.

Optimize simulation memory-handling and runtime on millions of data points using Python, TCL- specifically using associative arrays/hash mapping.

Connected developed workflows to remote cluster systems and optimized parallel threading on serial coding procedures to accommodate scaling.

Integrated visualized workflow results using the VTK toolkit for robust manipulation on custom displays.

Online Database for Yeast Strain Mutations and Replications:

Developed web application in PHP and MySQL that provides end users with an interface for creating, uploading, curating, and visualizing data for gene expression analysis.

SKILLS

AWS(EC2,Lambda, S3, EMR),Python, Spark, TCL, MySQL, SQL, NoSQL, R, SAS, Matlab, Mathematica, Tableau, Microsoft VBA/Excel, Fortran90, C, C++, HTML,PHP, JavaScript, CSS, XML, OSX, Linux, Windows

Groups

Data Scientist
Engineers Project -
Burlingame, Ca

New York Open
Statistical
Programming -
Manhattan, NY

LANGUAGES

English, Arabic,
Japanese

EDUCATION

University of Illinois at Urbana-Champaign *Bachelors of Science in Statistics*

August 2007 - May 2012

Dean's Honor List (2008), (2012)

Course Work

Advanced Data Analysis

Statistical Data Management

Applied Regression and Design

Biostatistics

Computer Engineering

Applied Linear Algebra

PROJECTS

Kaggle Data Science — Competitor

During the Kaggle competition I cleaned data sets, normalized data, and predicted missing values. I leveraged the use of NLP and clustering to investigate text features. In addition, I performed feature selection, created models to predict new data values using Python/Pandas, GraphLab, scikit-learn, XGBoost, and Amazon Web Services for parallel data processing workflows. My efforts received public recognition and a key part of the script is open source (aka a "Kaggle Kernel").

<https://github.com/itg-abby/KaggleScripts/>

Automated Level Creator — In the Groove 2/Stepmania

Development of level files for rhythm-based video games is something that is typically done by hand. The aim of this open source script is to reduce the amount of human effort spent on analyzing songs for relevant sound features. Specifically, relevant features are mapped according to a given time signature and presented in an output format readable by programs such as Stepmania. Time series analysis is powered by the AUBIO library for Python.

<https://github.com/itg-abby/StepGen>

Portfolio

<https://itg-abby.github.io/portfolio/>