Fisseha Berhane, PhD
Data Scientist at Aurotech
Phone: 443-970-2353 Website: *http://datascience-enthusiast.com/*
Email: *fisseha@jhu.edu* Regular writer for DataScience+ and R-bloggers

Current Employment

Data Scientist at Aurotech Sep 2015-

Solving various problems using data analytics and machine learning with Spark, R, Python, Hadoop ecosystem, and Tableau.

Projects:

- Hadoop Data Lake for analytics and machine learning with big data
  - Created a Hadoop cluster on AWS EC2
  - Ingested desperate data from various sources and in different formats to the lake
  - Cleaned and transformed the data for downstream analytics pipeline
  - Developed machine learning applications using Spark's MLlib library
  - Connected Tableau with the data lake and created visualizations using Spark SQL with ODBC connector.

- PDF data mining with R
  - Created an R-shiny application that mines useful insights from disparate and massive PDF documents
- Bayesian drug-adverse reaction signal detection
  - Using all the adverse reactions reported to the FDA, created an R-Shiny application that helps to detect signals using Bayesian techniques
- Predicting drug recall potential using various machine learning techniques and various data sources
  Architecture diagram available here
- Frequentist pharmacovigilance signal detection with Spark and shiny
  - Created a Shiny application with Spark that helps to detect drug adverse event signals using various frequentist techniques including Proportional Reporting Ratio (PRR) and Reporting Odds Ratio (ROR)
- Interactive drug adverse event knowledge discovery with R and Shiny using unsupervised machine learning techniques
  Cleaned and merged lots of adverse event datasets and stored them in a database.
  Developed an R-shiny application that clusters (using optics and hierarchical clustering) drug adverse events to discover new insights interactively.
  Architecture diagram available here

- Real-time tracking of disease outbreaks using social media with R and Tableau
  Created a complete pipeline that automates social media data collection, cleaning and processing, sentiment analysis, trend analysis and creates a Tableau dashboard
  Architecture diagram available here
- R-Shiny dashboard API that helps to download the FDA adverse events data
  Created an API that helps users to download data based on search query from the FDA adverse events database
- Social media mining to track natural hazards at real-time
  Created a Tableau dashboard that helps to track flooding
- Google Trends Analytics with R-Shiny
  Created an R shiny application that closely listens to google search trends and identifies anomalies in disease related google searches.

Education

Johns Hopkins University, Baltimore, MD --- Ph.D. in Atmospheric Physics 2016
Johns Hopkins University, Baltimore, MD ----M.A. in Atmospheric Physics May 2013
University of Connecticut, Storrs, CT ----------M.S. in Hydro-climatology May 2011
Mekelle University, Ethiopia ----------B.Sc. in Civil Engineering June 2006

Research Positions

Graduate Research Assistant, Department of Earth and Planetary Science, Johns Hopkins University, Baltimore, Maryland. August 2011 – 2015

- Built semi-automated rainfall prediction models for the globe, with various machine learning techniques such as Tree-based ensemble models (**Bagging**, **Random Forest** and **Boosting**), **Support vector Machines** and **Artificial Neural Network**, with **R (Shiny),** HTML, JavaScript, and CSS.
- Employed various Machine Learning techniques, statistical analysis and data mining methods using **Python** and **R** to understand interactions of atmospheric waves and their impacts on rainfall using large volume climate data.
- Analyzed large volume climate data, using **Python** and **R,** to investigate future climate conditions
- Completed many side-projects on big data using **Spark** (e.g., movie recommendation, web server log analysis, text mining and entity resolution and click-through prediction; available on my website)
- Worked on many other side-projects using **R** (available on my website)
- In addition to the data science courses I have done in grad school, I have taken more than 20 edx, coursera and Udacity data science courses (including data science specialization from Johns Hopkins University and big data XSeries from Berkeley) with **R, Spark, Python, Matlab, and Hadoop and MapReduce** (certificates on my website)

Graduate Research Assistant, Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 2009 – May 2011

- Built and evaluated a model that predicts Nile River flow. Further, examined possible impacts of climate change on river flow using different climate scenarios.
- The main tools I used in this study: **R**, **Python** and GIS.

Publications and Presentations

Three peer-reviewed publications in the Journal of Climate (JCL), which is among the most prestigious Journals in Atmospheric Science, one in preparation and a master's thesis.
More than 12 presentations, including in prestigious international conferences such as the American Geophysical Union (AGU) and the American Meteorological Society (AMS).

Teaching Experience

Teaching assistant (TA), Department of Earth and Planetary Science, The Johns Hopkins University, Baltimore, Maryland. Spring 2013
Assistant Lecturer, Department of Civil Engineering, Mekelle University, Ethiopia 2006-2009

Skills

Python, R, Matlab, Spark, MySQL, T-SQL, Teradata, Tableau, Ferret, NCL, HTML, CSS, JavaScript, Hadoop ecosystem.