Sree Aurovindh Viswanathan

| sviswa10@asu.edu | 480-399-3390 | linkedin.com/in/sreeaurovindh | github.com/sreeaurovindh |

Summary Computer Science Ph.D. Student with a strong background in machine learning, big data analytics, software development and statistics. Four years of professional experience in building machine learning models with Scientific Python Stack, R and have worked on extensive independent data science projects.

Education

Doctor of Philosophy, Computer Science **Dec 2018**
Arizona State University, Tempe. (GPA 3.63/ 4.00)

Master of Science, Computer Science **Aug 2014**
Arizona State University, Tempe. (GPA 3.52/4.00)

Bachelor of Engineering, Computer Science **May 2010**
Anna University, Tamilnadu, India. (GPA 78/100)

Skills

Machine Learning and Statistical Modelling: Experimental design, classification, clustering, regression and model selection, neural nets, support vector machines, and recommender systems
Big data Technology: Map Reduce (Hadoop), Apache Spark, ETL (Pig and Hive) and AWS
Scientific Python Stack: Numpy, Scipy, Matplotlib, Scikit-learn and Jupyter Notebooks
Programming Languages and Databases: Python, R, Java**,** PHP, SQL, MySQL and MongoDB

Patents

Systems and methods for estimating an impact of changing a source file in a software (US 20140123108)

Professional Experience

Graduate Research Assistant **Dec 2012- Present**
Arizona State University, Tempe, United States
- Hypothesized and established different qualitative levels of collaborative human behavior after careful analysis of over 130+ hours of observational study involving student's work on digital tablets
- Developed supervised and unsupervised models that classified quality of user interactions between student pairs using data from a combination of user interaction logs and acoustic and prosodic fingerprint
- Performed feature engineering in R using log features and extracted time-series based features using raw audio data using scientific Python Stack that improved balanced accuracy (BAC) 15% above baseline

System Engineer (Research) **June 2010- Jul 2012**
Software Engineering and Technology Labs, Infosys Labs, Bangalore, India
- Designed and developed bug prediction system that provides various metrics that enabled product managers to gauge the release readiness of the software system
- Clustered millions of commit data and developer profiles using Java and SQL based on their source code version history to provide better resource allocation to various bug fixes

Projects

Experimentation to Evaluate and Enhance Student Experience **Aug 2017 – present**
- Evaluated various invariant and evaluation metrics to measure the effect of the screening test
- Estimated number of samples and statistical power required to conduct a randomized trial
- Evaluated the impact of trial screener on student experience using effect size, significance and sign tests
- Recommended a follow up to reduce early cancellations by measuring total number of hours spent

Search for Sensitive Information in Enron Email Corpus **July 2017- Aug 2017**
- Clustered emails based on dense vector representations of words learned by Word2Vec model and evaluated quality of clusters obtained by various clustering algorithms using Silhouette Coefficient
- Uncovered five high level categories of emails that contains sensitive information such as credit card numbers and bank account information and flagged them for careful analysis

Deducing Structure of Web using Common Crawl **July 2015- Dec 2015**
- Sampled and analyzed **154 TB** dataset and validated an aggregation algorithm using Java and Python
- Grouped web pages with similar structure across different geo locations with information extracted from raw unstructured html files using Apache Pig

Yelp Food Recommender **Jan 2017- April 2017**

- Mapped Yelp user reviews to specific food items by combining data from Yelp and Wikipedia by performing text search on Yelp reviews using Apache Solr
- Created a dashboard that compares quality of food items of different restaurants in a specific neighborhood and enabled business owners to understand food quality changes over time

Data Exploration and Recommendation System **Jan 2015- May 2015**

- Created a recommendation system that predicts the expert user who would likely answer the question along with the reasonable time estimate from a corpus of over 2 million posts
- Developed code in R and python in order to perform preprocessing of data by using standard Natural language processing techniques and calculated TF-IDF Score for each term in the document
- Extracted cosine similarity across the entire document corpus to find documents with similar characteristics using apache mahout and amazon web services

Mining Prediction Challenge **Feb 2015- May 2015**

- Built machine learning models using ensemble method with a combination of random forests and support vector machines and validated it using ten-fold cross validation
- Ranked fourth out of fifty teams with 94% balanced accuracy score on classification of the dataset.

Publications

Sree Aurovindh Viswanathan, and Kurt VanLehn. "High Accuracy Detection of Collaboration from Log Data and Superficial Speech Features." International Society on Learning Sciences 2017.

Sree Aurovindh Viswanathan, and Kurt VanLehn. "Using the tablet gestures and speech of pairs of students to classify their collaboration." IEEE Transactions on Learning Technologies (2017)

Girish Makseri Rama, Deepthi Karnam, **Sree Aurovindh Viswanathan**, Srinivas Padmanabamuni, “Bug Prediction Metrics based decision support for Preventive software maintenance”, Software Engineering Conference(APSEC) 2012, 2012 19th Asia- Pacific

Girish Maskeri Rama, Deepthi Karnam, **Sree Aurovindh Viswanathan**, Srinivas Padmanabamuni,”Version history based source code plagiarism detection in proprietary systems”, 2012 28th IEEE International conference on Software Maintenance(ICSM), pp.609-612

G Shankar, **Sree Aurovindh Viswanathan**, Gowthamie B, “Swarm based intelligent transition of control from Manned to Unmanned vehicular system using Sun SPOT”, International Journal of Computer Applications, February 2010, Volume I, 82-89

Relevant Coursework Data Mining, Big Data Computing, Statistical Pattern Recognition, Spoken language processing, Intelligent systems, Design and analysis of algorithms

Online Coursework Statistical Learning Online Certificate from Stanford, R Programming, Getting and Cleaning Data, Reproducible Research, Machine Learning

Masters Thesis Using the tablet gestures and speech of pairs of students to classify their collaboration.
Advisor: Dr. Kurt VanLehn

Undergrad Thesis Dynamic agent discovery and Migration in Small Programmable Object Technology (SPOT)
Advisor: Dr. Conor Muldoon (University of Oxford, United Kingdom)