# ARUN JYOTHI PRAKASH BOOMINATHAN

**Phone:** (919) 995-4516 **Email:** arun.j.prakash@gmail.com


## TECHNICAL SKILLS

Languages: Python, C, C++, R, MATLAB
Tools: ipython, Amazon Web Services (Elastic MapReduce, EC2, S3, Dynamo), Hadoop stack (MapReduce, YARN, Hive, Pig)
Toolkits: numpy, pandas, scikit-learn, scipy, matplotlib, CUDA, nltk, orange, Theano, Torch, peach, mlpack
Proficient in Machine Learning, Data mining, Predictive and Regression Analysis, Pattern Recognition, Classifiers, Deep Learning

## EDUCATION

**Colorado State University,** Fort Collins, CO — Spring 2015
Master of Science in Electrical Engineering — GPA – 3.185
**Courses:** Machine Learning, Big Data, Neural Networks and Adaptive Systems, Applications of Random Processes, Robot Motion Planning, Topics in Robotics, Engineering Risk Analysis, Foundations of Systems Engineering, STEM Communications

**Visveswaraya Technological University,** India — Spring 2012
Bachelor of Engineering in Electronics and Communication Engineering — GPA – 7.06/10

**Other Courses:** Algorithms, Natural Language Processing, Image Processing, Artificial Intelligence, Game Theory, Distributed Computing

## RECENT PROJECTS

**Project 1: Prediction and Estimation of Factors affecting Redshift of Quasars (Team: 2)**

- Used the Sloan Digital Sky Survey dataset which consisted of detailed parameters of 45000+ Quasars
- Designed and built a Random Forest and implemented Multivariate Adaptive Regression Splines (MARSplines) in Python to perform predictive analysis of Redshift values of Quasars
- Designed and coded a Naïve Bayes' classifier in Python to perform classification analysis of Quasars based on various parameters of the electromagnetic spectrum

**Project 2: Weather Prediction using Regression Analysis (Team: 2)**

- Deployed Amazon's Elastic MapReduce (EMR) to reduce a 20GB daily global weather measurements dataset (years 1929-2009) from National Climate Data Center (NCDC) to manageable proportions (<1Mb)
- Developed a Multiple Linear Regression (MLR) model that predicts future weather of a location in Python

**Project 3: Diagnostic Analyzer for Frequent Errors in HTTP Logs using MapReduce**

- Developed software that stores data in the Hadoop Distributed File System (HDFS) and uses Apache Hadoop's MapReduce implementation
- Designed and implemented a log analysis system to perform access, peak, error and correlation analysis of HTTP logs

**Project 4: Cloud based Source-aware Key-value storage**

- Designed and implemented a simple key-value storage that runs on Amazon's Elastic Compute Cloud (EC2) instance cluster
- The system uses a distributed hash table, Chord protocol, as the underlying algorithm for dispersing datasets
- The key-value store stores/ retrieves data in a decentralized style and is aware of multiple data sources

**Project 5: Classification of Brain Tumor and Alzheimer's Disease using Artificial Neural Networks**

- Conducted feature extraction from magnetic resonance images using Principal Component Analysis
- Performed classification of Brain Tumor and Alzheimer's, and further classification into benign or malignant (in case of Brain Tumor) or Type 0, Type 0.5, Type 1, Type 2 (in case of Alzheimer's)
- Coded a Feed Forward Neural Network to classify Alzheimer's disease types and K-map Clustering to classify Brain Tumor images into its types

**Project 6: Image Data Compression using Self-Organized Feature Maps (SOMs)**

- Designed and tested self-organized maps to perform vector quantization on digital images
- Employed a diverse number of neurons to find the optimal number of bits for efficient transmission of images

**Project 7: Supervised Learning for Parameter Estimation and Stock Market Prediction**

- Time series modelling of procured stock data using linear and non-linear autoregression models
- Usage of parameters such as end-of-day prices to train a developed neural network to predict future stock prices

**Project 8: Stock Market Trend Predictor**

- Used real time stock data to predict the rising and falling of (trends) future stock prices
- Singular Value Decomposition (SVD) using Principal Components is used as the prediction algorithm

**Project 9: Implementation of Mixture of Gaussians algorithm**

- Used various datasets from UCI Repository as testing grounds for Mixture of Gaussians; to find optimal number of Gaussians
- Displayed projections onto two dimensions using Principal Component Analysis and modified Sammon Projection mapping