
Objective

Seeking a full-time opportunity to work in Data Analytics and Machine Learning.

Education

❖ **Master of Computer Science***Expected Graduation: May 2018***North Carolina State University**, Raleigh, NC

Coursework: Algorithms for Data Guided Business Intelligence; Machine Learning for User Adaptive Systems; Data Intensive Computing; Artificial Intelligence; Automated Learning and Data Analysis; Database Management Concepts and Systems; Foundations of Data Science; Enterprise Storage Architecture; Software Engineering; Design and Analysis of Algorithms

❖ **Bachelors in Telecommunication Engineering****Bangalore Institute of Technology**, Bangalore, India

Skill Set

❖ **Programming Languages:** Python, R, SQL, Java❖ **Machine Learning:** Linear/Logistic Regression, Clustering (Hierarchical, k-means, k-Nearest Neighbors), SVM, Decision Trees and Random Forest, Association Rule Mining, Time-Series Analysis, Statistics❖ **Tools & Frameworks:** Apache Spark, Kafka, Oracle DB, MongoDB, Elasticsearch, Kibana, AWS, Git, Github, Linux, Mac OS X, Windows, Ansible, Tableau, Bash, Selenium, R-Shiny❖ **Libraries:** Pandas, Numpy, Scipy, Scikit-learn, Matplotlib, networkx, Jupyter, regex

Work Experience

❖ **Graduate Research Assistant, North Carolina State University**, Raleigh, NC (January 2018 – Present)

- Creating data platforms, based on CRISP-DM and Agile Data Science software stack, for building and deploying Data Science applications. This platform includes tools such as Apache Spark, Apache Kafka, Mongo DB, Elasticsearch and Deep Learning.

❖ **Software Development Engineer, North Carolina State University**, Raleigh, NC (May 2017 – August 2017)

- Wrote code in Python to build a rule-based machine learning algorithm based on specific grammar rules to discover relations.
- This procedure was meant to find frequent patterns, correlations and associations in graph networks.

❖ **Project Engineer, Indian Institute of Science**, Bangalore, India (July 2015 - June 2016)

- Built electronic systems that mimic percussion instruments in terms of usage, look and feel.
- Piezoelectric sensors were used to capture 15ms audio signals for real-time processing (normalization, time-series and FFT analysis) and identification of fundamental frequency and harmonics for each sound.
- Built a prototype and wrote an algorithm in Python to process and match subsequent vibrations (audio signals) using pattern recognition and play sounds of the percussion instrument from memory.

❖ **Decision Scientist, Mu Sigma**, Bangalore, India (July 2014 - June 2015)

- Performed extensive data mining, data munging and analysis using SQL and Tableau to find patterns and correlations across terabytes of time-series data. Exploratory data analysis provided an understanding of the underlying structure of the data and to analyze the presence of trends, seasonality, and impact of outliers/noise. Machine Learning provided sales/revenue forecasting and patient/drug usage behavior.
- Implementing these methods in projects such as Cost Sensitivity Analysis, Length of Patient Therapy and Patient Behavior Analysis resulted in an immediate improvement in sales and revenue across multiple categories for the pharmaceutical client.

Projects

❖ **Scrum Bot integrated with Amazon Alexa (Java, JavaScript, AWS, Ansible)**

- Implemented a bot as an Amazon Alexa skill to perform tasks in JIRA and modify events using Google Calendar API via simple voice interactions. The whole process was automated with Ansible. The Ansible-playbook creates a new EC2 instance, installs dependencies, pushes and builds the code, and runs the service.
- Application development required the use of Jira, Google Calendar API from G-suite, AWS Lambda function, and AWS EC2 with Java-based spring REST service. Spring-based REST service was used to perform any long-running task since lambda functions have limited resources.

❖ **Data Pipeline for Data Streaming, Processing and Analytics (Python, Spark, MongoDB, Elasticsearch, AWS)**

- Built a data pipeline that handles multiple sources of streaming data for near real-time processing and analytics.
- Distributed processing framework included Apache Kafka for buffering data, Apache Spark for data processing and sentiment analysis, MongoDB for persistent storage of data, and Elasticsearch and Kibana for fast indexing, search and data visualization.
- Amazon Web Services (AWS) was used for creating a VPC (Virtual Private Cloud) along with EC2 instances. Each framework was set up on separate EC2 instances for maximum resource utilization. MongoDB was configured to provide replication while Ansible provided infrastructure automation to easily handle configuration.

❖ **Semi-Supervised and Supervised Learning for Classification (Python, Machine Learning)**

- Implemented and compared three semi-supervised learning techniques (S3VM, Co-Training, and Expectation-Maximization) and three supervised learning techniques (SVM, Logistic Regression, and Artificial Neural Networks).
- The project demonstrated the advantage of combining labeled and unlabeled data to better capture the shape of the underlying data distribution and improve learning accuracy.

❖ **Health Application using Oracle DB (SQL, Java)**

- Designed and implemented Oracle Relational Database for a Personal Health Management (PHM) application.
- Wrote extensive DDL statements to define the database structure and schema. DML statements were written for data manipulation.
- Implementation enforced integrity constraints, concurrency control, access control, and database normalization. Database design used triggers, cursors and functions using SQL.

Leadership Activities

- ❖ Graduate Student Representative, University Dining Committee, North Carolina State University (2016 – 2017)