Sreejith Sreekumar

Availability : May 2019

1 St James Pl., Roxbury, Boston, Massachusetts, 02119

857-399-6443 | sreekumar.s@husky.neu.edu | srjit.github.io | srjit | srjit

Education

Northeastern University, Boston, MA *Jan. 2017 - Present*

Candidate for Master of Science in Data Science, GPA: 3.5/4.0 *Expected Graduation: Apr 2019*

- **Relevant Courses** : Natural Language Processing, Supervised Machine Learning, Applied Probability and Stochastic Processes, Unsupervised Machine Learning, Computer Vision, Special Topics in Artificial Intelligence - Text Mining for Humanities and Social Sciences, Information Visualization

Government Engineering College, Thrissur *Sep. 2007 - June 2011*

Bachelor of Technology

- **Relevant Courses** : Data Structures and Algorithms, Database Management Systems, Programming Paradigms, Numerical Analysis and Optimization Techniques, Design and Analysis of Algorithms

Technical Knowledge

Specialities	Predictive Modeling, Natural Language Processing, Deep Learning, Distributed Computing
Programming Languages	Python, C++, Bash, Java, Groovy, Javascript
ML Tools/Frameworks	Tensorflow, Scikit-Learn, Pandas, Matplotlib
Big Data Ecosystem	Apache Spark and Spark Mllib, Apache Hadoop, Hive, Sqoop, Oozie
Databases	MySQL, MongoDB, HP Vertica
Other Skills	Google Cloud, Tableau, Linux
Certifications	Scalable Machine Learning(edX), Introduction to Big Data with Apache Spark (edX), Machine Learning (Coursera)

Experience

Centre for Complex Networks Research, Northeastern University *Boston, USA*

Graduate Student Researcher *Jan 2019 - Present*

- Studying the effects of the linguistic structure of titles and abstracts of scientific publications in creating impact among researchers, and building quantitative models to estimate it.

Enterprise Risk - Analytics, Fidelity Investments. *Boston, USA*

Data Scientist (Co-Op) *Jan 2018 - July 2018*

- Analyzed network traffic log data, built insightful visualizations, and developed anomaly detection predictive models for abnormal network activity detection.
- Developed a framework for enhanced exploratory data analysis of network connection logs on PySpark.

Data Science Group, Innovation Labs, [24]7.ai Inc. *Bangalore, India*

Senior Data Engineer *June 2016 - Dec 2016*

- Modeled chat transcripts from customer conversations for user intent prediction for customer agent queue routing that achieved a recall of 0.87.
- Designed and developed a Natural Language toolkit on PySpark for chat transcript data analysis and modeling.
- Configured the toolkit on a multi-cluster environment with three apache spark nodes for scalability.

Data Science Group, Innovation Labs, [24]7.ai Inc. *Bangalore, India*

Data Engineer *May 2015 - June 2016*

- Analyzed and modeled user data from website visit behavior for several clients in the e-commerce domain for detection of potential customers who needed help with purchases and predicted their chat propensity with a customer service agent.
- Integrated SVM algorithm into the domain specific custom modeling tool and scaled over a million data points.

Xurmo Technologies Pvt. Ltd. *Bangalore, India*

Software Engineer *July 2011 - May 2015*

- Developed custom analytical functions for Xurmo Big Data Platform for data transformation.
- Programmed analytics applications using Platform as a Service modules - Text exploration engine, Stock market movement prediction, Sentiment analyzer, Customer churn prediction.

Recent Academic Projects

- **Investigating Instances of Gun Violence using Pointer Networks:** Proposed a novel model that employs Attention Mechanism in Sequence-to-Sequence learning and Pointer Neural Net to extract the attributes of gun violence events from news reports.
- **Quantifying Semantic Similarity:** Designed and implemented a Long Short-Term Memory neural network for classifying semantically similar and dissimilar questions from Quora, carrying an accuracy of 83% on validation after tuning.
- **The Fake News Stance Classification:** Achieved an accuracy of 88% on classifying fake news from the genuine ones to four discrete levels - agree, discuss, disagree, and unrelated using handcrafted linguistic features along with distance features from vectorized fields(Word2Vec). Random Forests, Support Vector Machines, and XGBoost algorithms were used for performance comparison.