

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

CANDIDATE FOR MASTER OF SCIENCE IN DATA SCIENCE, GPA: 3.5/4.0

Jan. 2017 - Present

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

BACHELOR OF TECHNOLOGY

Sep. 2007 - June 2011

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

Technical Knowledge

| | |
|------------------------------|---|
| Specialities | Natural Language Processing, Classification, Clustering, Regression, 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 |
| Certifications | Scalable Machine Learning(edX), Introduction to Big Data with Apache Spark (edX), Machine Learning (Coursera) |

Experience

Centre for Complex Networks Research, Northeastern University

GRADUATE STUDENT RESEARCHER

Boston, USA

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.

DATA SCIENTIST (Co-Op)

Boston, USA

Jan 2017 - July 2017

- 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.

SENIOR DATA ENGINEER

Bangalore, India

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.

DATA ENGINEER

Bangalore, India

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.

SOFTWARE ENGINEER

Bangalore, India

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.