

# SREEJITH SREEKUMAR

+1 (857)-399-6443

<http://github.com/srjit> <http://srjit.github.io>

[sreekumar.s@husky.neu.edu](mailto:sreekumar.s@husky.neu.edu)

[www.linkedin.com/in/srjit](http://www.linkedin.com/in/srjit)

## Education

**Northeastern University**, Boston, MA

Jan 2017– May 2019

*Master of Science, Data Science* – GPA: 3.5/4.0

*Coursework:* Natural Language Processing, Supervised & Unsupervised Machine Learning, Applied Probability & Stochastic Processes, Computer Vision, Information Visualization, Text Mining, Science of Science (Research)

**Government Engineering College**, Thrissur, Kerala

Sep 2007 – June 2011

*Bachelor of Technology, Computer Science*

*Coursework:* Data Structures and Algorithms, Database Systems, Numerical Analysis and Optimization Methods

## Technical Skills

<b>Specialties:</b>	Predictive Modeling, NLP, Deep Learning, Distributed Computing, Data Visualization
<b>Machine Learning:</b>	Tensorflow, PyTorch, Scikit-Learn, Pandas, NumPy, Matplotlib, Plotly
<b>Programming:</b>	Python, C++, R, Bash, Java, Javascript
<b>Big Data:</b>	Apache Spark, Mllib, Hadoop, Hive, Sqoop
<b>Databases:</b>	MySQL, Vertica
<b>Other Skills:</b>	Google Cloud, Tableau, D3.js, Linux

## Professional Experience

**Centre For Complex Networks Research**, *Research*, Boston, MA

Jan 2019 – Present

- Built quantitative models to estimate the influence of title lengths of scientific articles on their popularity.
- Estimated the temporal interdisciplinary novelty in scientific publications as a function of occurrence of new words in their titles.
- Characterized the fluctuations in word usage in subsequent years with different distributions to identify the words whose usage is increasing or decreasing more than by noise.

**Fidelity Investments**, *Data Scientist*, Boston, MA.

Jan 2018 – Jul 2018

- Developed anomaly detection predictive models, visualizations to analyze abnormal network traffic activities.
- Built exploratory data analysis framework for network log analysis on PySpark.
- Designed probabilistic models for classifying files containing potential threats and achieved a recall of 0.88.
- Created intuitive visualizations using Matplotlib and presented visual stories to executive management.

**[24]7.ai**, *Senior Data Scientist / Engineer*, India

May 2015 – Dec 2016

- Modeled chat transcript data to predict the intent of customer care calls and re-route them to the concerned agent. Achieved a recall of 0.86 for the model.
- Designed and developed a Natural Language Toolkit for chat transcript data exploration and modeling.
- Configured the toolkit on a multi-cluster environment with three Spark nodes for scalability.
- Developed models to predict chat propensity of customers with agents based on their website behaviour data.
- Deployed propensity models in production using Javascript & integrated it on customer engagement platform.
- Integrated SVM and Random Forest algorithms as Vertica R UDFs and scaled them on multiple nodes.

**Xurmo Technologies**, *Software Engineer - Analytics*, India

July 2011 – May 2015

- Built custom analytical functions for data transformation as Apache Hive function extensions.
- Programmed analytics applications using Platform as a Service REST APIs – Text exploration engine, Stock market movement prediction, Sentiment analysis, and Customer churn prediction.

## Projects

- **Investigating Instances of Gun Violence using Pointer Networks:** Extracted attributes of gun violence events using Attention Mechanism and Pointer Neural Nets (Tensorflow) from news reports.
- **Fake News Stance Classification:** Tuned and achieved 88% accuracy in classifying fake news from genuine ones. Random Forest, XGBoost and SVM algorithms were used for performance comparison.
- **Sales Time Series Forecasting:** Forecasted the sales fluctuations of 10 stores using a supervised regression approach with ARIMA and XGBoost and contrasted it with LSTM (PyTorch) Multi-step Time Series Forecasting.