Sreejith Sreekumar

Portfolio: http://srjit.github.io
Linkedin: https://www.linkedin.com/in/srjit/
Email: sreekumar.s@husky.neu.edu
Mobile: +1-857-399-6443
Github: http://github.com/srjit

Education

- **Northeastern University** — Boston, MA
  Master of Science in Data Science — *Jan. 2017 – Present*
- **Government Engineering College** — Thrissur, India
  Bachelor of Technology in Computer Science and Engineering — *Sep. 2007 – Apr. 2011*

Skills

- **Specialities**: Classification and Clustering, Regression, Deep Learning, Natural Language Processing & Distributed Computing
- **Programming**: Python, R, Scala, Shell Scripting, Java, Groovy, Javascript
- **ML Tools/Frameworks**: Tensorflow, Keras, Scikit-Learn, Pandas
- **Big Data Ecosystem**: Apache Spark and Spark Mllib, Apache Hadoop, Hive, Flume, Sqoop, Oozie
- **Databases**: MySQL, MongoDB, HP Vertica
- **Certifications**: Scalable Machine Learning(edX), Introduction to Big Data with Apache Spark (edX), Machine Learning (Coursera)

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 Nets to extract the attributes of gun violence events from news reports.
- **Quantifying Semantic Similarity of Sentences using Long Short-Term Memory Networks**: Designed and implemented a sequence-to-sequence model (LSTM network) for classifying semantically similar and dissimilar questions from Quora, carrying an accuracy of 83% on validation after tuning.
- **Domain Specific Classification using AlexNet**: Tuned the layers of a pre-trained AlexNet model for binary classification task on images that obtained an accuracy 94% for the new task.
- **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.
- **Home Value Prediction**: Modeled Zillow's house rent prediction problem using Microsoft's LightGBM algorithm with a mean absolute error of 0.064.

Experience

- **Enterprise Risk - Analytics, Fidelity Investments** — Boston, US
  Data Scientist (Co-Op) — *Jan 2017 - July 2017*
  - Modeled the detection of anomalies in SOCKS proxy logs for suspicious network activity using Isolation Forest and Local Outlier Factor.
  - Developed a framework for enhanced exploratory data analysis of SOCKS connection logs on PySpark
- **Data Science Group, Innovation Labs, [24]7.ai Inc** — Bangalore, IN
  Senior Data Engineer — *Jun 2016 - Dec2016*
  - Modeled chat transcripts from customer conversations for user intent prediction for customer agent queue routing that achieved an accuracy of 90%.
  - 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, IN
  Data Engineer — *May 2015 - June 2016*
  - Analyzed and modeled user data from web for several clients in the e-commerce domain for increasing chat propensity of potential customers with customer agents and uplifting purchases.
  - Integrated SVM algorithm into the domain specific custom modeling tool and scaled over a million data points.
- **Xurmo Technologies Pvt. Ltd.** — Bangalore, IN
  Software Engineer — *July 2011 - May 2015*
  - Developed and maintained machine learning modules of the flagship product of the company - Xurmo big data analytics platform.
  - Developed and integrated machine learning algorithms on Apache Spark (Java).
  - Developed custom analytical functions as a platform functionality for data transformation.
  - Programmed analytics applications using the Platform as a Service - Text exploration engine, Stock market movement prediction, Sentiment analyzer, Customer churn prediction.