Work Authorization: H1B
Bangalore, India
+91 (819) 741 1464
tarunrs@gmail.com
https://github.com/tarunrs

Tarun Ronur Sasikumar

Objective

To obtain a Full-time position in data science and software engineering

Education

2011 – 2013 **Ohio State University**, *Columbus, OH*.
MS in Computer Science and Engineering. GPA: 3.7/4.0

2004 – 2008 **Vishweswaraiah Technological University**, *Belgaum, India*.
BE in Computer Science and Engineering.

Professional Work Experience

2016 – Current **Data Scientist**, *Crediwatch*, Bangalore, India.
Implemented a distributed crawling infrastructure capable of crawling and indexing 100+ pages/minute per bot using scrapy, frontera, elasticsearch, kafka and HBase. Designed and developed a real-time news search system with sentiment analysis, named entity extraction and clustering of news stories. Built a system to detect and generate Breaking News. Improved the precision of a date extractor library by 20%. Built multiple classification models to tag news articles with metadata. Invented a new pre-clustering algorithm based on Canopy clustering algorithm that also optimizes on the density of the clusters. Built a Lead Generation system using machine learning techniques that improved our client's convertion rate by 2-7x.

2013 – 2015 **Software Developer**, *Epic Systems Corp.*, Verona, Wisconsin.
Resposible for design and development of modules used to document Oncology related information in the EMR. Instrumental in adding support to interface with third-party Radiation Oncology systems.

2008 – 2011 **Software Developer**, *Atlantis Computing*, Bangalore, India.
Responsible for design, development and maintenance of user profile and application virtualization components in the Atlantis ILIO product stack. Built expertise in the core filesystem level and driver development of both Windows and Linux platforms.

Research Projects

tCTR - Network attention predictor for tweets, Prof. Srinivasan Parthasarathy.
Studied the effect of language syntax and sentiments in sparse data like tweets to predict network attention. Showed that using these lower dimensional language features contributed to network attention more than word based features or novelty of subject.

Realtime Analysis of Tweets, Prof. Arnab Nandi.

Implemented a system for real-time analysis of sentiments and surfacing of contextual tweets related to political events like the Presidential debates. The supervised model achieves a precision of 85% with a response time of less than 200ms on the 1% Twitter firehose.

Link Prediction - KDD Cup 2012, Prof. Srinivasan Parthasarathy.

Part of a team of 3 that took part in the KDD Cup link prediction challenge in social networks. Wrote multiple scripts to extract semantic and network features from the social graph which helped achieve a weighted average precision of 81%.

pSnipSuggest, Prof. Arnab Nandi.

Implemented a system that provides on-the-go, context-aware assistance in the SQL composition process, based on SnipSuggest by Khoussainova et al. Increased the average precision by 21% and the response time by 30% compared to the original implementation. Written in Python

Open Source and Personal Projects

OSU Events.

Part of a team of 6 that developed applications on multiple mobile platforms for events happening at OSU. Designed and developed the crawler and screen scrapping algorithm for extracting events, as well as the webserver for the APIs used by the mobile clients. Written in Python and Ruby.

jamMm.in.

Co-founded jamMm.in, an online music collaboration application. Served as the User Experience developer. Wrote the framework for mixing multiple tracks as well as the library to generate waveform images for the tracks. Written in Python and Ruby.

Movie Showtimes.

Developed an Android application for Movie showtime/theatre details based on the user's location. Downloaded over 36,000 times in the Android Market. Written in Java

These projects are listed with source code at: https://github.com/tarunrs

Technical Skills

Languages	Experience: Python, C Basic: HTML/CSS, Javascript, Java, C++ Familiar with distributed/parallel processing systems like Spark, MapReduce/Hadoop ecosystem, dask
Datastores	MongoDb, HBase, Elasticsearch, MySQL, postgres
Message Queues	RabbitMQ, Kafka, Redis
Machine Learning Libraries	scikit-learn, nltk, spacy, gensim, tensorflow, pandas, CoreNLP, numpy, matplotlib, numba
Tools	Git, SVN, Balsamiq, Weka.

References available on request