Sriharsh Bhyravajjula

Data Scientist | IIIT Hyderabad
http://darthbhyrava.in | s.bhyravajjula@research.iiit.ac.in | +91-9441161582

EDUCATION

MS in Computational Linguistics
Expected May 2019 | IIIT-Hyderabad

B.Tech in Computer Science
Expected May 2018 | IIIT-Hyderabad
Cumulative GPA : 7.02

SKILLS

Programming & Scripting
Python • C/C++ • Shell

Data Science
TensorFlow • NLTK • Gensim

Frameworks
Flask • web2py • Django • Jekyll

Version Control
Git • Gerrit

POSITIONS

Speaker, AICC 2018
Invited for a technical talk on NLP at MLR College, Hyderabad.

Teaching Assistant, IIIT-H
Natural Language Processing (Monsoon '17)

Mentor, IASNLP 2017
Guided NLP projects at LTRC's summer school. (Jul '17)

COURSEWORK

Natural Language Processing • Information Retrieval and Extraction • Artificial Intelligence • Statistical Methods in AI • Data Structures and Algorithms • Systems Analysis and Design • Distributed Systems

EXTRA CURRICULARS

Chief Editor, College magazine
Aug'16 - Aug'17

Co-ordinator, Literary Club
Aug '16 - Aug '17

POETRY

Published
• 9 poems, Sahitya Akademi's Indian Literature (Issues: 275, 286)
• 3 poems, Muse India (Issue: 62)
• 4 poems, Madras Courier (Sep '17)

WORK EXPERIENCE

Klevu Oy | Data Scientist
Working as a data scientist to improve an e-commerce search engine. Also building an interactive shopping chat-bot which extends the search capabilities for this Finnish startup. (May '18 - current)

Infocrunch Analytics | Data Scientist
Worked on ML and NLP approaches towards identifying user political bias from 100GB of social media datasets. (Jan '18 - Mar '18)

MaoMao Inc. | NLP Engineer
Worked on collection, indexing, topic-modeling and topic classification of over 100 GB of data scraped using a browser extension, for a Singapore-based start-up. (Sep '17 - Jan '18)

Wikimedia Foundation, Inc. | Intern
Built a Python-based automation bot to handle tasks for the "Thanks" extension in MediaWiki as part of my Google Summer of Code project. (Apr '16 - Jul '16)

RESEARCH

Publications (In Progress)
• Retrieval of Character Relations from Text (First author)
• Computational Analysis of the Romantics' Poetry (First Author)

PROJECTS

Online News Article Popularity
Predicted popularity of online news articles using machine learning methods like PCA with linear regression for classification. Accuracies for Naive Bayes, Random Forest, Support Vector Machines (SVMs), Neural Nets were between 53% and 85%.

Factoid QA System over Unstructured Data
Built a factoid question answering system over an unstructured dataset of 150k articles from The Guardian with an accuracy of 74%.

Wikipedia Search Engine
Developed a search engine for a 61 GB Wikipedia dump using merge sort and secondary indexing, and used retrieval techniques like term-frequency (tf-idf) for ranking.

Health Forum Message Classifier
Used Neural Network architectures like LSTMs, CNNs and SVMs to classify messages into categories on a health-forum, with accuracies varying from 55% to 58%.

others
Sentiment analysis on 185k words from a football forum • A rudimentary AI/NLP based football assistant • An IRCbot written in Go • An OpenGL 3D game • A Django portal for EERC lab • A Linux Shell written in C • An Android image-recognition app. • A distributed Windows App hosted on Azure.