Virali Thakkar

www.linkedin.com/in/virali1 (213) 292-7905 | vcthakka@usc.edu

EDUCATION

University of Southern California, Los Angeles *August 2017 – May 2019*
MASTER OF SCIENCE IN COMPUTER SCIENCE | GPA: 3.71/4.0
Coursework: Algorithms, AI, ML, Natural Language Processing, Information Retrieval and Web Search Engines, Data Mining

Gujarat Technological University, India *June 2013 – May 2017*
BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING | GPA: 3.9/4.0

TECHNICAL SKILLS

Programming Languages: Python, JAVA, C, C++, R, Scala, HTML, CSS, PHP, JavaScript, SQL, MATLAB
Frameworks and Libraries: Pandas, NLTK, Hadoop, Spring MVC, Hive, Pig, Hibernate, Apache Spark, HBase, AngularJS, TensorFlow
Tools and Technologies: MongoDB, MySQL, Tableau, Docker, Elasticsearch, Kibana, Git, AWS (EC2, S3, Redshift, Lambda, EMR)

PROFESSIONAL EXPERIENCE

Pulse Secure - Software Engineering Intern *May 2018– August 2018*
Data Visualization | Docker | Machine Learning | Elasticsearch | Kibana | Python

- Designed a **Flask based Web Application** to analyze JSON-formatted Network and Software logs pertaining to *Pulse Connect Secure* (mobile VPN) employing **Elasticsearch** coupled with **Kibana** for effective visualization.
- Reduced the time required for identifying the **root cause** of various issues (Cluster Split, Node Failure), from two hours to few minutes.
- Developed a classification framework to perform Anomaly Detection of HTTP request data by detecting DDOS attacks through classifiers which include Naïve Bayes and Neural Network.

PROJECTS

Recommendation System and Market Basket Analysis *September 2018 – November 2018*
Data Mining| Scala | Apache Spark

- Performed **Item-based Collaborative Filtering** on the Yelp dataset containing 452353 reviews corresponding to 30,000 users and 30,000 businesses (Data: 90% Train and 10% Test).
- Predicted User-Business Ratings (RMSE: 1.08) and implemented **Matrix Factorization** CF employing Spark MLlib package (RMSE of 1.06).
- Implemented **SON (Apriori)** algorithm using the Apache Spark Framework for identification of frequent words in Yelp Reviews having execution time 120 seconds for 100000 Support Threshold.

Web Search Engine *September 2018 – November 2018*
Google Cloud | Apache Lucene | JAVA | Python

- Developed a Java-based **Multithreaded Web Crawler** to crawl *LA Times* and extracted total 2.7 million HTML pages.
- Created an **Inverted Index** (word docID: count) of 86554 unique pages within *LA Times* using Hadoop cluster on Google Cloud Platform.
- Leveraged **Apache Solr** to index web pages from *LA Times* and enhanced the search engine with *Auto-Suggest* functionality.
- Calculated PageRank Score of web pages using NetworkX library and compared the search capabilities of Apache Lucene and **PageRank**.

VMware CodeHouse - Hackathon *July 2018*
OpenFaaS | Natural Language Processing | Python

- Orchestrated a web application (Flask, D3.js) comprising of different modules which present scientific facts related to entities identified in an image and provide interactive video tutorials incorporated with assistive chatbot, to encourage children to pursue STEM education
- Deployed modules on **OpenFaaS** framework (Functions as a Service) for consumption of functions asynchronously or collectively.

Autonomous Tagging of Stack Overflow questions *January 2018 – May 2018*
Natural Language Processing | Python

- Created a multi-label classification system to automatically assigns tags for questions posted on Stack Overflow forum.
- Compared Algorithm adaptation techniques with base algorithms being k-Nearest Neighbor, Random Forest classifier and SVM.
- Achieved an optimal F1 score of 62.35% for SVM classifier, on a subsampled portion of original data restricted to 100 tags and a minimum of 500 instances of each tag across data.

Deep Search - Crime Mystery Exposed! *March 2018 – May 2018*
Computer Vision | Python | Flask

- Extracted keyframes by creating an average brightness histogram and calculating entropy.
- Obtained a highly contextual description of each scene employing Microsoft's Cognitive services and acquired a most likely description of a search by cosine similarity and jump to resulted frame.

MENTORSHIP & ACHIEVEMENTS

- Mentor, **Viterbi Graduate Mentorship Program,** University of Southern California *Fall 2018*
- Winner, **TrojanHacks,** USC ACM Hackathon *Spring 2018*
- Mentor, **AthenaHacks**, University of Southern California *Spring 2018*