VARUN CHAUDHARY

• (480)930-9286 • varunchaudharycs@gmail.com • in/varun-chaudhary-cs • /varunchaudharycs

EDUCATION

Master of Science in Computer Science **May 2021**
Arizona State University GPA: 3.9

Bachelor of Technology in Computer Science & Engineering **July 2018**
Manipal University Jaipur GPA: 3.7

SKILLS

Languages	Java, Python, Scala, MySQL, AspectJ, LaTeX, *nix shell scripting
Libraries	pandas, numpy, scikit-learn, opencv, spark-streaming, nltk, torch, scrapy transformers, tensorflow
Tools	Springboot,Git,Solr,MS SQL,Jupyter-notebook,IntelliJ,Pycharm,Colab
Machine Learning	Classification(Logistic Regression,Decision Trees,Naive-Bayes, SVM), Clustering(DBSCAN, K-Means), NLP(BERT), PCA, SVD
Big Data	Hadoop, Spark, MapReduce, Kafka, Kudu, Hive, Impala, HDFS, YARN

WORK EXPERIENCE

Software Engineering Intern, *Kuebix(a Trimble Company)* May '20 - *present*

- Involved in aspect-oriented development using spring-boot.

Software Engineer (Big Data), *Infoobjects Inc.* Jan '18 - Apr '19

- Designed two real-time ETL frameworks and APIs to process and store GBs of data within a second.
- Increased robustness by handling point of failures and reduced downtime by 15% via alarm system.
- Actively presented and participated in the "Big-data Club" to share technologies among project teams.

Software Engineering Intern, *Tata Consultancy Services* Jun - Aug '17

- Created a real-time sentiment analysis model using tweets on a social issue.
- Analysed change in public opinion over an year via sentiment scoring and phrase popularity for insights.

SELECTED PROJECTS

Social media analysis of RedForEd for educational policies Jan '20 - *present*

- Voluntary DCEdEx project studying the impact of RedForEd in Arizona by monitoring social media.

`aNswER` - Question-Answering for MultiRC dataset(novel) May '20

- Created `aNswER`, an NER approach for MultiRC, a multi-hop multi-choice question answering dataset.
- Achieved F1 score of 60, improvement over a baseline model score of 58(using BERT-base).

Healthcare Mining May '20

- Developed a symptom, disease, discussion forum and drug-based search-engine after mining and indexing data from healthcare websites with 86% precision.

Bio-metric identification on 11K hands data set Sep '19

- Implemented feature extraction models and LSH for CBIR and user-based relevance-feedback.
- Accurately predicted labels using personalized page rank (92%), SVM (90%) and decision tree (89%).

Event sequence prediction of medical events Nov '17

- Trained an LSTM network over 2-year medical record history of 1500 patients.
- Successfully predicted future medical conditions of patients with an accuracy of 87%.