7421 Frankford Rd,
Dallas, TX - 75252

Shubham Kothari

+1 (408) 250-3395
shubhamkothari1211@gmail.com

EDUCATION

MS Computer Science, 3.68/4 **University of Texas at Dallas** **Aug 2017 - May 2019**
Graduate Coursework: Machine Learning, Database Design, Data Structure and Algorithms, Natural Language Processing, Object Oriented Analysis and Design, Internet of Things

BE Computer Engineering, 3.85/4 **University of Mumbai, India** **July 2013 - May 2017**
Undergraduate Coursework: Artificial Intelligence, Computer Networks, Operating Systems, Web Technologies

SKILLS

Languages: Python, Java, C, C++, Swift
Data Science: Pandas, nltk, scikit-learn
Databases: MySQL
Web Technologies: HTML5, CSS3, JavaScript, PHP
Protocols: TCP/IP, UDP, Ethernet, HTTP, SMTP
Tools: Git, Sublime, OpenCV, Anaconda, Tableau

PROJECTS (https://github.com/shubhs1211/Data-Science-portfolio)

FAQ Semantic Matching System, NLP **April 2018**

- Preprocessed Yelp's QA corpus by tokenizing, cleaning and transforming data.
- Extracted semantic features using dependency parsing, WordNet, stemming, pos-tagging and lemmatization.
- Achieved an MRR score of 0.7 for solr model and 0.54 for Bag-of-words model implementation.

Technologies Used: Python, pandas, nltk, stanfordcorenlp, pysolr, scipy, Apache OpenNLP

Automatic Essay Scoring, **Kaggle, NLP, Machine Learning** **April 2018**

- Performed tokenization, stemming, lemmatization and pos-tagging on approx. 13000 essays to extract features like counts of words, spelling mistakes, nouns and vocabulary richness for each essay.
- Predicted essay grades using Neural Networks, Random Forest and Linear Regression Model and achieved a quadratic kappa score of 0.63 after using 5-fold cross validation.

Technologies used: Python, Scikit-learn, Pandas, nltk, scipy, matplotlib

Park Smart, Internet of Things (IOT) | Won 1st Prize – OneM2M Hackathon **Mar 2018**

- Built an application to find and reserve a parking spot, if a user is in vicinity of an available parking spot.
- Implemented by exchanging data through the OneM2M platform and web server between the app and Arduino.
- Data collected was used to gain insights of parking space occupancy with respect to time, weather and location.

Technologies used: OneM2M platform, Node.js, Java, Arduino nodeMCU

Autonomous Greenhouse System, IOT **April 2018**

- Built a prototype to nurture the plants autonomously by forming a wireless sensor network using XBee devices.
- Controlled the lighting and watering from an android app based on data collected from various sensors.

Technologies Used: Arduino, Raspberry Pi, ZigBee, Node.js

Grocery Sales Forecasting, **Kaggle, Machine Learning** **Nov 2018**

- Built classification models including linear regression, neural network, XGBoost, and random forest to forecast sales for over 200,000 products.

Technologies used: Python, Scikit-learn, Pandas

Employment Status Prediction, Machine Learning **Mar 2018**

- Predicted employment status by analyzing time spent by individuals on different activities based on age, education, leisure using Neural Networks and Random Forest with an accuracy of 0.94.

Technologies Used: Python, Scikit-learn, Pandas, Tableau

COMMUNITY SERVICE

Volunteer – National Service Scheme, University of Mumbai **July 2014 – July 2016**

Completed 270 hours of community service which included organizing events like blood donation camps, cleanliness drives, and teaching the underprivileged kids.