7421 Frankford Rd, Apt 1128,
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: Proficient: Python, Java, HTML5, CSS3, Bootstrap, PHP, SQL **Familiar:** C, C++, Swift, JavaScript
Data Science: Pandas, nltk, scikit-learn, pytesseract, PyTorch, Apache Spark
Tools: Git, Sublime, OpenCV, Xcode, Anaconda, Tableau

EXPERIENCE

Software Developer, Intern **Thakur College of Engineering & Technology** **Sept 2016 – Dec 2016**

- Developed a web-based attendance management system to keep a record of more than 2000 students.
- Reduced time to generate monthly attendance by approx. 75% by automating the task.
  Technologies used: Bootstrap, JavaScript, JQuery, PHP, MySQL

Data Science Intern **KlassDiamond, India** **May 2016 – Aug 2016**

- Performed data preprocessing and exploratory data analysis to identify key factors affecting sales of the company including number of people, exports, imports, cost price, selling price, diamond price, current affairs.
- Analyzed and presented data using graphs to gain insights to improve return on investment by 6%.
  Technologies used: Python, Scikit-learn, Pandas, Plotly, Seaborn, matplotlib

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

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

- 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, ML** **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

Grocery Sales Forecasting, **Kaggle, Machine Learning (ML)** **Nov 2018**

- Built machine learning models based on linear regression, neural network, bagging, and random forest and compared them to find the least error on the regression problem.
  Technologies used: Python, Scikit-learn, Pandas

Courier Service Robot, Computer Vision, IIT Bombay | Secured 6th position all over India **Oct 2016**

- Decreased parcel delivery time by 8%, using an algorithm based on A* search algorithm.
- Implemented Image Processing techniques and computer vision for object detection and path planning.
  Technologies used: Python, OpenCV, Raspberry Pi, Computer Vision

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

- Developed a smart parking app using OneM2M, Node.js, IoT, and sensor networks to get parking availability.

Decision Tree Implementation, Machine Learning **Sept 2017**

- Implemented the ID3 machine learning algorithm from scratch, without using any packages, and trained the model on a binary classification dataset.