

Education

Maharshi Dayanand University

Bachelor Of Technology in Computer Science; GPA: 7/10.0

Gurugram, India

Aug. 2013 – June. 2017

Experience

QuickCompany (Startup)

Software Developer (Data Science)

New Delhi, India

Feb, 2018 - Mar, 2018

- Involved in product development by building data scraping API and deploying to cloud and storing into database (AWS technologies), also involved in building companies trademark (image) search engine using Machine Learning and other data science analytical skills.

Xavient Information Systems

Intern (Big Data and Hadoop)

Noida, India

Jan, 2017 - June, 2017

- Involved in building and deploying data ingestion pipeline using big data technologies.

Projects

- **Yellow Taxi Demand Prediction in NYC** : Predict the pick up density of yellow cabs at a given particular time and a location in new york city using **Linear Regression, Random Forest, XGBoost, Time Series and Fourier Transformation**.
- **Personalized Cancer Diagnosis**: Classify the given genetic variations/mutations based on evidence from text-based clinical literature using **Logistic Regression, Random Forest, TF-IDF and Feature Engineering hacks**.
- **Microsoft Malware Detection**: In the past few years, the malware industry has grown very rapidly that, the syndicates invest heavily in technologies to evade traditional protection, forcing the anti-malware groups/communities to build more robust softwares to detect and terminate these attacks. The major part of protecting a computer system from a malware attack is to **identify whether a given piece of file/software is a malware**. Techniques used : **Logistic Regression, Random Forest, XGBoost**.
- **Netflix Movie Recommendation System** : Given list of ratings provided by user per a movie on particular date. We need to predict the rating will be given by an user on a particular day. Based on those predictions to build movie recommendation systems. Techniques used : **XGBoost, SVD++**.
- **Quora Question Pair Similarity** : An automated way of detecting if pairs of question text actually correspond to semantically equivalent queries using **Logistic Regression, Linear SVM and XGBoost**.
- **Stackoverflow Tag Predictor** : Since there are a huge number of tags, it is often a cumbersome process to search the correct tags. It may be useful to have an auto-tagging system that suggests tags to users depending on the content of the question. Techniques used : **Logistic Regression (One vs Rest Multilabel Classifier)**.
- **Web Scraper** : Build a dynamic web scraper using Scrapy, PhantomJS, OpenCV & Python while deploying scraper on AWS DynamoDB and EC2.
- **Hadoop Administration** : Description Configure and Manage Hadoop Cluster on CentOS 7. Setup Single Node, Pseudo-distributed, & Multi-node Cluster. Setup Hadoop Cluster using Hortonworks Data Platform. Configure and managed data Node, setup security and failover. Enabled high availability on cluster.

Skills

Proficient with : Python, Java, SQL, Scrapy, Scikit-learn, Pandas, Scipy, Seaborn, Matplotlib, Naive Bayes, Logistic Regression, Random Forest, XGBoost, SVM, SGD, K-Means, Hierarchical Clustering, TensorFlow, Git, Linux.

Familiar with : R Programming, Hadoop, Apache Kafka, AWS, Hive, MySQL