Pratyush Kumar

pratyushkumar.me
me@pratyushkumar.me | +91 7895395395

EDUCATION

IIT ROORKEE
B.Tech in Industrial Eng.
2013 - 2017 | Roorkee, India

UDACITY
Data Analyst Nanodegree
June 2016 - Sep 2016

SKILLS

DATA ANALYSIS
Pandas • NumPy
DATA WRANGLING
Python
DATA VISUALIZATION
R • D3 • Matplotlib
MACHINE LEARNING
Scikit-learn • TensorFlow
BIG DATA
Spark • Map Reduce
PROGRAMMING
Java • C++ • Scala
DATABASE
SQL
VERSION CONTROL
Git
SOFTWARE PACKAGES
MATLAB • MySQL • PyCharm • Eclipse • MS Office
OPERATING SYSTEM
Linux • MacOS • Windows

COURSEWORK

Machine Learning (Coursera)
Advanced Statistics
Linear Algebra
Algorithm and Data Structure
Graph Theory
Database Management System
Engineering Optimization
Operation Research

REFERENCES

• P.M. Pathak
Associate Prof., IIT Roorkee
pushpfme@iitr.ac.in
• Kaushik Pal
Associate Prof., IIT Roorkee
palkfme@iitr.ac.in

EXPERIENCE

DUKE IT | Software Developer Intern
May 2016 – July 2016 | Bengaluru, India

- Used AWS cloud computing platform to analyze raw data containing data of users. Uploaded the data on Amazon S3 bucket, used Amazon Elastic Map Reduce built in features to load the data onto the cluster.
- Queried the database as per company requirements.
- Explored the dataset using EDA and visualized it using R, and fulfilled the company demands for required insights about the data.

PROJECTS

HAPTICS ROBOT FOR MOTOR REHABILITATION
July 2016 – Present | IIT Roorkee

- The project involves design and development of Haptics planner robot hand.
- Virtual environment for providing force feedback mechanism to the haptic hand is being developed.
- Machine learning algorithms like the Neural Network has been implemented to train the Haptic hand.
- Finally, the hand is used for motor rehabilitation purposes.

IDENTIFYING FRAUD FROM ENRON EMAILS AND FINANCIAL DATA | July 2016 – August 2016 | Udacity | GitHub Link

- Aim was to explore the Enron dataset, use data wrangling and visualization techniques to clean and visualize the data, find the correlation between various features, identify extreme outliers and finally use appropriate machine learning algorithm to predict Person of Interest 'POI' i.e., employees who committed fraud.
- The accuracy of algorithms was calculated by F1 score; the logistic regression with PCA gave the best results.
- Achieved the highest accuracy among all the code submissions.

OPENSTREETMAP DATA WRANGLING WITH SQL
May 2016 – June 2016 | Udacity | GitHub Link

- Implemented data munging techniques like assessing the quality of the data for validity, accuracy, completeness, consistency and uniformity.
- Used ElementTree to parse the OSM file which had lots of inconsistencies in the dataset viz., abbreviations, lowercase, misspelling etc.
- Created regex to clean and standardize the dataset. Queried the dataset to extract useful information viz., number of unique users, common amenities, popular places etc.
- Created an input data model so that new users follow it to reduce the number of inconsistencies.

DESIGN AND ANALYZE AN A/B TESTING
August 2016 – September 2016 | Udacity | GitHub Link

- Udacity tested a "free trial screener" for the new enrollment.