RIKESH PATEL

(847) 400-7500 ◇ rikeshp98@gmail.com
rikeshp.github.io ◇ Chicago, IL 60614

SKILLS

Programming	Python • R • Javascript • SQL • SAS • Java
DB/Big Data	Oracle • MySQL • MongoDB • Hadoop (MapReduce, Hive, Pig, Spark, Mahout)
Data Analytics	Scikit-Learn • SciPy • NumPy • Pandas • Gensim • SpaCy • NLTK • Tableau • Excel • SPSS
Web Development	Node.js • Express • WebSockets • HTML • D3
Environments/Tools	AWS (EC2, S3) • Jupyter Notebooks • Linux • Git

EXPERIENCES

Research Assistant - DePaul Center for Data Science, Chicago, IL July 2020 - Present

• Developed unsupervised learning model to segment a large dataset of medical teaching files by their literary uncertainty levels into 4 clusters. This helped my research team develop a classification model based on the uncertainty levels my model identified.

• Responsible for building NLP topic modeling pipeline using NLTK and Gensim, and building a machine learning model using Scikit-Learn.

Data Science Intern - Urban Outfitters, Philadelphia, PA June - August 2020

• Built a classification model to determine which marketing channels were most effective in leading to a sale.

• Transformed raw customer data into MySQL and conducted dimension reduction and attribute encoding in order to prepare data for machine learning.

• Visualized my results using Python and R and presented my finding to the marketing team.

PROJECTS

Bitcoin Ocean Summer 2020

• Bitcoin Ocean is a Node.js/Express app that visualizes bitcoin transactions in real time as ocean creatures. Uses Blockchain's WebSocket API to stream data and runs on Heroku cloud. (Can be viewed **here**)

Sales Data Cluster Analysis Spring 2020

• Processed large volume sales transaction data using hadoop cluster on AWS EC2, used Hive and Pig for data transformation, and performed unsupervised learning by applying K-means clustering using Mahout.

• Resulted in 7 partitions, segmenting the data into groups of customers based on similar purchasing behavior.

• Gained experience setting up multi node hadoop clusters in AWS EC2, reading and writing data to/from S3, and working in a Linux environment.

Predictive Bank Telemarketing Winter 2019

• Built classification models in Python using decision trees and k-nearest neighbors to predict whether or not a bank client will subscribe to a term deposit. Resulted in 84% of positive cases accurately determined and showed that a client's occupation type was the most important feature, among others. (Can be viewed **here**)

• Responsible for cleansing and transforming data using one hot encoding to prepare data for machine learning.

Forecasting Medical Expenses Spring 2019

• Built a regression model in SAS to forecast annual medical expenses of Americans using insurance data. Used stepwise regression test to determine the best fit model. Ultimately our model predicted 75% of variability in our data and showed that age and BMI were the most influential features on the regression model. (Can be viewed **here**)

• Responsible for understanding project requirements, building data pipeline in order to gather, enrich, and cleanse the data, and creating exploratory data visualizations such as correlation matrices and box plots.

EDUCATION

DePaul University, Chicago, IL 2020

B.S, Computer Science, Minor in Data Science, GPA: 3.4

Relevant Courses: Mining Big Data, Fundamentals of Data Science, Advanced Data Analysis, Regression Analysis, Database Systems, Data Visualization, Data Structures, Statistics