RIKESH PATEL

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

SKILLS

Programming	Python • R • SQL • SAS • Java • Scala
DB/Big Data	Oracle • MySQL • Hadoop (MapReduce, Hive, Pig, Spark, Mahout)
Data Analytics	Scikit-Learn • NumPy • Pandas • MATLAB • Tidyverse • Excel • SPSS
Environments/Tools	AWS (EC2, S3) • Jupyter Notebooks • Linux • Git

EDUCATION

DePaul University, Chicago, IL 2020
B.S, Computer Science, Minor in Data Science, GPA: 3.3

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

WORK

Data Science Intern - Urban Outfitters, Philadephia, PA Summer 2020
• Built classification model using Naive-Bayes algorithm to optimize marketing channel preferences.
• Transformed raw customer data into MySQL and conducted dimension reduction and attribute encoding in order to prepare data for machine learning.
• Presented my findings to marketing team, which included exploratory and explanatory visualizations created in Python and R.

Pro Bono Consultant - HealedPeopleHealPeople.com Spring 2020
• Proposed and designed solution to automate an online retailer's large inventory datasets by migrating Excel spreadsheets to MySQL tables and building data pipeline to refresh MySQL tables.

PROJECTS

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.

Improving Bank Telemarketing using ML 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 and addressed multicollinearity with indicators such as VIF. (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.
• Ultimately our model predicted 75% of variability in our data and showed that age, body mass index, and number of children were the most important features in predicting an individual's medical expenses.