RIKESH PATEL

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

SKILLS

Programming	Python • R • SQL • SAS • C++
DB/Big Data	Oracle • MySQL • Hadoop (HDFS, Hive, Pig)
Data Analytics	Python (scikit-learn, pandas, IPython, seaborn) • R (tidyverse) • Mahout • SPSS
Other Technologies	AWS • Linux • Git • Tableau • Excel

EDUCATION

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

Related Coursework: Mining Big Data, Fundamentals of Data Science, Database Systems, Data Analysis and Regression (SAS), Data Visualization, Data Structures

WORK

Data Science Intern, Urban Outfitters Summer 2020
Offer accepted. Cancelled due to COVID-19 *Philadelphia, PA*

Would have:
• 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.

PROJECTS

Sales Data Cluster Analysis Spring 2020
• Transformed raw sales data using Hive, Pig, and Hadoop streaming in order to prepare data for ML.
• Performed unsupervised learning by applying K-means clustering using Mahout, in order to partition the data into 7 clusters.
• Familiarized with setting up multi node clusters in AWS EC2, reading and writing data to/from S3, and working in a Linux environment.

Improving Bank Telemarketing using ML Winter 2019
• Responsible for building 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)
• Also conducted data cleaning and transformation in order to prepare data for ML.

Forecasting Medical Expenses Spring 2019
• Responsible for building a regression model in SAS to forecast annual medical expenses of Americans based on data provided by an insurance company. Stepwise regression test was used to determine the best fit model. (Can be viewed here)
• 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.
• Also was responsible for exploratory data analysis, in which I created visualizations to understand the attributes and used indicators such as VIF to address multicollinearity in our model.

Visualizing Gun Ownership Winter 2018

• Created an interactive streamgraph showing the increase in firearm background checks initiated over the past two decades using the D3 Javascript library. (Can be viewed here)