APARNA BHUTANI

Contact:
(201)208-8412
ab8473@nyu.edu
LinkedIn
GitHub

Skills:

BI & Warehousing Tools:
Tableau, Power BI, QlikView, SSIS, MS Excel, SAP BI

Programming:
Python(numpy, pandas, matplotlib, scikit-learn, pyspark), Scala, SQL, R, Java, GitHub

Databases:
MySQL, Redshift, PostgreSQL, MS SQL Server, MongoDB

Specialization:
Data Analysis, Data Modeling, Feature Engineering, Hypothesis testing, A/B Testing, Predictive Modeling & Analysis, Descriptive statistics

Big Data Skills:
Spark, MapReduce, Hadoop, Hive

Machine Learning
Linear Regression, Logistic Regression, Clustering, KNN, Decision Trees, Random Forest

Relevant Coursework:
Relational Databases, Data Mining, Data Warehousing, Machine Learning, Applied Data Science, Statistics, Big Data

Education:
Master of Science
Major: Informatics
New York University
Aug 2020
GPA: 3.556

Master of Technology
Major: Information Systems
Delhi Technological University
June 2019
GPA: 3.61

Bachelor of Engineering
Major: Information Technology
University of Pune
June 2013
GPA: 3.42

Experience

GRADUATE TEACHING ASSISTANT | NYU- COURANT | NEW YORK CITY *Aug 2019–June 2020*
- Conducted recitation for course CORE-UA 111: From Data to Discovery incorporating **programming and data analysis using R language,** for 60 students
- Taught students quantitative and algorithmic thinking, **statistical modeling**

DATA SCIENCE INTERN | SIEMENS LTD | INDIA *June 2018–Aug 2018*
- Developed first in house Recommendation Engine prototype for Siemens Generator Services department to automate task/issue assignment process to employees
- Performed data exploration, text wrangling & processing. Implemented **K nearest neighbors** to find similar issues, **Cosine similarity** to find top 10 employees to solve issues

DATA ANALYST | GB ENTERPRISES| INDIA *Jan 2016– May 2017*
- Developed Tableau dashboards providing insights on various milestone KPIs to help business make strategic decisions
- Worked extensively with SQL to improve meeting customer commitments and increase operational efficiencies by utilizing a data driven root cause analysis approach
- Analyzed customer behavior patterns and recommended beneficial changes to the process. **Increased customer retention rate by 15%**
- Eliminated spreadsheet reporting, **reduced processing time by 50%** and **increased reporting accuracy**

BUSINESS INTELLIGENCE ANALYST | ACCENTURE SERVICES PVT LTD | INDIA *Nov 2013-Jan 2015*
- Performed **data analysis, data visualization** for **large volume of historical data**
- Built and tested **Extraction, Transformation, Loading** process to migrate client data from multiple source systems into the data warehouse using SAP BI tool
- Retrieved and aggregated data from multiple sources and compiled it into actionable format
- Collaborated with cross functional teams for defect prevention activities related to business issues and critical operations
- Presented, reported key findings and issues in data reconsolidation, ETL to the client in a simple intuitive format

Projects

PARKING VIOLATIONS IN NEW YORK CITY *Feb 2020–May 2020*
- Analyzed the number of parking violations in New York City from 2015 to 2019. Performed **geocoding, geospatial joins** of 50 million parking violation records using **MapReduce**, **Spark**
- Used **Ordinary Least Squares** method to determine rate of change in total number of violation over the four years. Presented the output and loaded to **HDFS**

PREDICTING VACCINE UPTAKE FOR H1N1 *Feb 2020–May 2020*
- Performed data analysis, data cleaning, feature engineering of H1N1 Vaccine dataset. Used KNN and MICE Imputation model to deal with missing values
- Developed Decision Tree model to predict vaccine uptake using H1N1 vaccine, demographic data. **AUC score = 0.822**. Used Random Forest model to check robustness of result, **AUC score = 0.836**
- Found key factors responsible for influencing decision to take vaccine. Conducted **sentiment analysis** of Twitter data as an additional study of people's view towards vaccine uptake in pandemic

ANALYZING FACTORS FOR SUCCESS OF RESTAURANTS *Sep 2019–Dec 2019*
- Analyzed success of restaurants in Phoenix, Arizona and developed recommender system to recommend top 10 restaurants to users
- Performed data analysis and processing of Yelp data, urban data. Used **DBSCAN, Gaussian Mixture, K- means** to cluster areas based on income, population. Used **folium** to create interactive map representing clusters
- Conducted **sentiment analysis** of Yelp reviews using NLP for every restaurant in Phoenix
- Followed classification approach; Used **SVM, Xgboost, Random Forest** to determine key factors for restaurant success. **Increased F1 score from 0.53 to 0.76**

Publications

- Susan, Seba, and Aparna Bhutani. "Data Mining with Association Rules for Scheduling Open Elective Courses Using Optimization Algorithms." In *International Conference on Intelligent Systems Design and Applications*, pp. 770-778. Springer, Cham, 2018
- Susan, Seba, and Aparna Bhutani. "A Novel Memetic Algorithm Incorporating Greedy Stochastic Local Search Mutation for Course Scheduling." In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, pp. 254-259. IEEE, 2019