Prathamesh Pradip Datar

pratt.datar@gmail.com | https://www.linkedin.com/in/datar96 | https://github.com/pratt-datar | (315)728-0287

Experienced Data Scientist, bringing passion, versatility, and strong technical acumen to create business impact

Education

Syracuse University - ***School of Information Studies*** ***GPA: 3.97/4; May '21***
MS – Information Management, Certificate in Advanced Studies – Data Science
Relevant Courses: Big Data Analytics, Text Mining, Cloud Management, Data Analysis & Decision Making

University of Mumbai - ***K.J. Somaiya College of Engineering*** ***GPA: 8.69/10; Jun '18***
BTech - Electronics and Telecommunication Engineering

Skills And Competencies

- **Core Skills**: Data Science, Machine Learning, Neural Networks, Deep Learning, Time Series Analysis
- **Programming Languages**: Python (PySpark, Pandas, Scikit, matplotlib), R (tidyverse, ggplot2, igraph, RShiny)
- **Databases and Tools**: MySQL, Visio, MS Access, Mongo DB
- **Big Data Skills:** Hadoop, BigQuery, Hive, MapReduce, AWS EC2, AWS S3, Docker, Google Cloud Platform, Azure, Spark
- **Data Analysis Tools:** Jupyter Notebook, Tableau, PowerBI, Gephi, MS Excel, MS PowerPoint, Plotly, SAS, Databricks
- **Collaboration tools:** Git, BitBucket, AWS CodeCommit, Jira

Experience

Acoustic Wells (MIT) – *Data Science Intern* ***Boston, US; Jun '20 – Aug '20***
- Identified bottlenecks in oil tank height estimation and implemented a production allocation system to determine financial revenue
- Desensitized well pressure data dependent on temperature fluctuations by 10% using linear regression for temperature estimation
- Improved production rate estimation of oil wells by 5% by writing a custom loss function with L1 norm and regularization
- Coordinated with COO to conduct a parametric study and validate research findings using agile methodology and Jira

Think Analytics – *Associate Data Scientist* ***Mumbai, India; Nov '18 – Jul '19***
- Delivered an 18% improved real-time anomaly detection system using PCA and WOE-IV model to detect issues in the oil well operating conditions, minimize production loss and provide maintenance team better insights for an expedited recovery
- Invented a real-time segmentation system using computer vision techniques and a custom-made algorithm to identify the player's landing point and detected 90% of uncalled no-balls in the game of cricket to assist umpires in making better decisions
- Developed an identity verification system for Banks using RNN, AWS Rekognition, EC2, S3 for smooth customer onboarding

Syracuse University – *Graduate Research Assistant* ***Syracuse, US; Feb '20 – Present***
- Studied a large-scale Google Cluster usage trace dataset 2020 (5 TB) to characterize the temporal correlations in vertical scaling
- Compared latency-sensitive and production priority jobs for scheduling delays and resource requests for jobs and alloc sets
- Investigated a social phenomenon of drinking bleach during a pandemic by extracting 2 TB coronavirus tweets using MongoDB

Projects

Pause & Ponder: Altice USA 2020 Innovation Hackathon *(NLP and Cloud Computing)* ***Project Link; Nov '20 – Nov '20***
- Achieved 2nd Place at the hackathon organized by Altice, Google, Microsoft, and Infosys to help improve user's mental health
- Generated top topics with topic modeling NMF algorithm on the user browsing data and created a web application using Flask
- Hosted Flask app on GCP VM instance and provided a monthly summary of top topics for browsing data on HTML webpage

2020 US Election Analysis *(Big Data and Machine Learning)* ***Project Link; Aug '20 – Dec '20***
- Performed a comparative analysis of tweets before and after the election result to unearth the influence of social media in elections
- Developed a user aggregated hashtag analysis and applied Kmeans clustering and PCA to detect communities and identify outliers
- Applied Logistic Regression and Random Forest in PySpark to recognize feature important words in predicting user sentiment

Sentiment analysis of drug reviews *(NLP and Machine Learning)* ***Project Link; Jul '20 – Aug '20***
- Empowered drug manufacturers to provide better customer service by determining review sentiment of 215K patient reviews
- Predicted sentiment and achieved an F-score of 0.85 by applying LinearSVC model using unigram bigrams and custom vocabulary
- Discovered ambiguous reviews by conducting a comparative study between SVM and Naïve Bayes models for text classification

Southeast airline customer service analysis *(Data Science and Machine Learning)* ***Project Link; Aug '19 – Dec '19***
- Mapped 10.2K flight survey reviews in R to recognize service and operational shortcomings and improve customer satisfaction
- Summarized 5 main attributes of data by performing exploratory data analysis like bar charts, map visualizations, wordcloud
- Led a team of 4 to provide actionable insights by applying association rule mining to find prominent predictors for customer type

Publications

Automatic license plate recognition (ALPR) *(Machine Learning and Deep Learning)* ***Publication Link; Jul '17 – Jun '18***
- Published in Springer Communications in Computer and Information Science (volume 941, Chapter 4: Advances in Data Science)
- Proposed an ALPR system using Python to decongest tollways by 82% in India by creating a custom 43K vehicle dataset
- Enhanced existing ALPR system using Computer Vision and Semantic Segmentation techniques to achieve 82% test accuracy