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, Database Management, Text Mining

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**: Machine Learning, Time-Series Analysis, Neural Networks, Deep Learning, Visualization, Network Science
- **Programming Languages**: Python (PySpark, Pandas, Scikit, matplotlib), R (tidyverse, ggplot2, igraph)
- **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:** Tableau, PowerBI, Gephi, MS Excel, MS PowerPoint, Plotly, Jupyter Notebook, 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 terms
- Managed a team of two to conduct a parametric study and validate research findings using agile methodology and Jira to coordinate

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 teams better insights
- 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***
- Investigated a social phenomenon of drinking bleach during a pandemic by extracting 2 TB coronavirus tweets using MongoDB
- Designed coursework for 'IST 359 – Intro to DBMS' and mentored hybrid class in concepts such as Normalization and ERDs
- Mentored 100 students in 'IST 687 - Introduction to Data Science' to integrate Jupyter notebooks and MIDST tool

Projects

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

Replication study: Designing sustainable online support *(Social Network Analysis)* ***Project Link; Mar '20 – May '20***
- Examined sociotechnical design changes on 49 WebMD forums in 2010 to understand the relationship between design and stability
- Devised a modularity graph analysis in R to validate the division of distinct subgroupings for reply and relationship networks
- Verified and visualized socio-technical and network changes in Gephi by performing a core-periphery analysis from 2009 to 2014

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

Listing Management on OrangeHousing.com *(Database Management)* ***Project Link; Aug '19 – Dec '19***
- Identified database issues on listing management site "OrangeHousing" and developed a SQL application using ERDs and Reports
- Collaborated with a team of 2 analysts to redesign key processes to support historical data access for efficient financial management

Publications

Automatic license plate recognition (ALPR) *(Machine Learning and Deep Learning)* ***Project 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