**Prathamesh Pradip Datar**

| [pratt.datar@gmail.com](mailto:pratt.datar@gmail.com) | 315-728-0287 | [LinkedIn](https://www.linkedin.com/in/datar96) | [GitHub](https://github.com/pratt-datar/) | [Medium](https://medium.com/@prathamesh.datar1996) | [Google Scholar](https://scholar.google.com/citations?user=cpux2kMAAAAJ&hl=en&oi=ao) |

Experienced Data Scientist, bringing technical expertise, passion, and versatility 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*

**Relevant 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 for tracking

**Think Analytics – *Associate Data Scientist***  ***Mumbai, India; Nov ‘18 – Jul ‘19***

* Developed a Know Your Customer (KYC) verification system for banking and finance clients using RNN, AWS Rekognition, EC2, S3 adhering to AML (Anti-Money Laundering) norms prescribed by regulatory bodies for smooth customer onboarding
* 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

**Syracuse University – *Graduate Research Assistant***  ***Syracuse, US; Feb ‘20 – Present***

* Researched a large Google cluster dataset 2020 (5 TB) to characterize the temporal correlations in vertical scaling of Borg clusters
* Compared latency-sensitive and production priority jobs for scheduling delays and resource requests for jobs and alloc sets in GCP
* 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

**Relevant Projects**

**Pause & Ponder: Altice USA 2020 Innovation Hackathon** *(NLP and Cloud Computing****)*** [***Project Link***](http://35.223.18.187:9090/)***; 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***](https://github.com/pratt-datar/IST_718)**; *Aug ’20 – Dec ‘20***

* Performed a comparative analysis of tweets before and after the election result to unearth the influence of social media on 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-rich words in predicting user sentiment

**CoronaEXT: Tracking COVID cases** *(Cloud and Database Management)*[***Project Link***](https://github.com/pratt-datar/IST615_CoronaEXT)**; *Aug ’20 – Dec ‘20***

* Led a team of 6 to build a web app that tracks COVID-19 cases and helps users assess their symptoms on Google Firebase
* Authenticated users using AWS Cognito for providing read access to data on DynamoDB through AWS API Gateway and Lambda

**Sentiment analysis of drug reviews** *(NLP and Machine Learning)*[***Project Link***](https://github.com/pratt-datar/Sentiment-ambiguity)**; *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

**Listing Management on OrangeHousing.com** *(Database Management)* [***Project Link***](https://github.com/pratt-datar/Database-Management)***; 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

**Skills And Competencies**

* **Core Skills**: Statistical Analysis, Predictive Analytics, Strategy Consulting, Machine Learning, NLP, Database Management
* **Analytics & Visualization Tools:** Tableau, PowerBI, MS Excel, Databricks, Jupyter Notebook, Matplotlib, Plotly, ggplot2
* **Programming languages & ETL Tools:** Python (Pandas, Scikit, PySpark), R (tidyverse), SQL Server, AWS Redshift, Snowflake
* **Big Data & Database Skills:** AWS DynamoDB, Lambda, EC2, S3, MongoDB, Hadoop, MapReduce, Docker, Spark, BigQuery

**Publications**

**Springer Communications: CCIS** *(Volume 941, Chapter 4: Advances in Data Science)*[***Publication Link***](https://doi.org/10.1007/978-981-13-3582-2_4)***; Jul ‘17 – Jun ‘18***

* 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