

Prathamesh Pradip Datar

Passionate about applying Data and Strategy for Social Good

pdatar@syr.edu | <https://www.linkedin.com/in/datar96>

+1-315-728-0287 | <https://github.com/pratt-datar>

EDUCATION

Syracuse University - School of Information Studies

GPA: 3.96/4; May '21

MS – Information Management, Certificate in Advanced Studies – Data Science

Achievements: Graduate Program Merit Tuition Scholarship

Relevant Courses: Data Analysis & Decision Making, Social Networks, Data Admin Concepts & Database Management, Text Mining

University of Mumbai - K.J. Somaiya College of Engineering

GPA: 8.69/10; Jun '18

BTech - Electronics and Telecommunication Engineering

Relevant Courses - Big Data Analytics, Neural Networks, Operations Research, Cloud Computing

SKILLS AND COMPETENCIES

- **Core Skills:** Machine Learning, Time-Series Analysis, Neural Networks, Deep Learning, Visualization, Network Science
- **Programming and Scripting Languages:** Python (Pandas, Scikit, matplotlib), R (tidyverse, ggplot2, igraph), SQL
- **Databases and tools:** MS SQL Server, MySQL, Visio, MS Access, Mongo DB, Jira
- **Big Data Tools:** HDFS, MapReduce, AWS EC2, AWS S3
- **Data Analysis Tools:** Tableau, PowerBI, Gephi, MS Excel, MS PowerPoint, Plotly, Jupyter Notebook, SAS
- **Statistical Skills:** Descriptive and Inferential Statistics, Data Modelling, Hypothesis Testing, ANOVA, Regression
- **Version Control:** Git, BitBucket

EXPERIENCE

Acoustic Wells (MIT) – Data Science Intern (Full-time)

Boston, US; Jun '20 – Aug '20

- Implemented a production allocation system to determine financial revenue and monitor oil production using casadi optimization
- Conducted a parametric study to replicate the synthetic production allocation system using random walks and white noise in Python
- Remodeled loss function using L1 norm and validated research findings using regularization coefficients in matplotlib
- Applied Gaussian mixture models for on-off detection of oil wells and tracked agile team progress and coordination using Jira

Syracuse University – Graduate Research Assistant (Part-time)

Syracuse, US; Feb '20 – May '20

- Extracted Coronavirus tweets (80M or 1 TB) using Twitter API to identify social phenomenon in pandemic using Python and R
- Processed and tracked improvements using MongoDB, PyMongo, and Git and interpreted underlying social structures using Gephi
- Mentored over 100 students in IST 687 to help understand Jupyter Notebooks and Data Science Tool (MIDST) within their projects

Think Analytics – Associate Data Scientist (Full-time)

Mumbai, India; Nov '18 – Jul '19

- Performed time-series analysis to predict oil pressure anomalies using Principal Component Analysis and WOE-IV model
- Improved F-score of existing prediction system by 18% and provided business insights using Plotly Visualization tool in Python
- Invented a real-time semantic segmentation system to eliminate uncalled no-balls in cricket using FloydHub, and YOLOv3
- Developed an identity verification system using RNN, Amazon Rekognition, EC2, S3 for speech and face recognition in Python

PROJECTS

Sentiment analysis of drug reviews (NLP and Machine Learning)

[Project Link](#); Jul '20 – Aug '20

- Analyzed UCI Drug Review Dataset to identify ambiguous reviews using conjunction counts in Jupyter Notebook
- Conducted a comparative study between SVM and Naïve Bayes models for text classification to predict patient review sentiment
- Achieved a test accuracy of 85.48% and F-score of 0.85 on LinearSVC model by applying unigram bigrams and custom vocabulary

Replication study: Designing sustainable online support (Social Network Analysis)

[Project Link](#); Mar '20 – May '20

- Examined the sociotechnical design changes in WebMD forums to understand the relationship between design and stability
- Devised a modularity graph analysis to validate the division of distinct subgroupings for reply and relationship networks in R
- Performed a core-periphery analysis to examine replies to and from each class and visualized network change in Gephi

Southeast airline customer service analysis (Data Science and Machine Learning)

[Project Link](#); Aug '19 – Dec '19

- Extracted and cleaned airline customer's survey dataset for providing actionable insights to increase customer satisfaction using R
- Conducted exploratory analysis using bar charts, and ggplot2 map visualizations and implemented text mining technique wordcloud
- Worked as a Team leader to predict customer satisfaction with 81.38% test accuracy using SVM and apriori association rule mining

Listing Management on OrangeHousing.com (Database Management)

[Project Link](#); Aug '19 – Dec '19

- Identified database issues for listing management on OrangeHousing and developed an application to support historical data access for financial management using Entity Relationship Diagram, Forms, and Reports on SQL Server, MS Visio, and MS Access.

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

[Project Link](#); Jul '17 – Jun '18

- Published in Springer (CCIS, volume 941, Chapter 4: Advances in Data Science)
- Created a 43K image car dataset to deploy an Automatic License Plate Recognition system to decongest tollways by 95% in Python
- Built ALPR using Convolutional Neural Networks, Deep learning algorithm YOLOv2, and Semantic Segmentation to segment and recognize number plate characters with 82% accuracy on test data with significant breakthroughs than image processing algorithms