ABHIJIT GOKHALE

agokhale@syr.edu | (315) 278-4351 | LinkedIn | Handshake | GitHub

EDUCATION

SYRACUSE UNIVERSITY, SCHOOL OF INFORMATION STUDIES, SYRACUSE, NY, USA Jan. 2021 - Dec 2022
Master in Applied Data Science **(GPA: 4)**
- Relevant Courses**:** Quantitative Reasoning for Data Science(stats), Introduction to Data Science, Applied Machine Learning, Big Data Analytics, Financial Analytics.

MUMBAI UNIVERSITY, SARDAR PATEL INSTITUTE OF TECHNOLOGY, MUMBAI, INDIA July. 2012 - May 2016
Bachelor in Electronics Engineering **(GPA: 3.3)**

KEY SKILLS & QUALIFICATIONS
- **Programming & database:** Python (NumPy, Pandas, scikit-learn, seaborn, TensorFlow, PyTorch), PySpark (PyTorch), R, Microsoft SQL server, Azure Data Studio, TOAD, Ambari for HiveQL
- **Data Science & Analytics Concepts:** Stochastics and Statistics, Data Mining, Advance Machine Learning, Clustering, SQL server and relational database management, Business Intelligence, Pipelines, HDFS, Map reduce, Time series.
- **Data Visualization & Analytics tools:** Power BI, AWS IAM, MS PowerApps, Informatica ETL, MS Excel, Draw.io, Jira.
- **Soft-Skills**: Leadership, Strategic Planning, and Task/Project management.

EXPERIENCE

SYRACUSE UNIVERSITY, SYRACUSE, NY, USA Sept 2021 – Dec 2021
Faculty Assistant

- Assisting professor for **IST 687 coursework: Introduction to Data Science** in creating lab and homework solutions.
- Grading and helping students to resolve assignment queries on statistical concepts and implementation in R language.

ACCENTURE, MUMBAI, INDIA Dec 2016 – August 2020
Application Development Analyst

- Worked with large and complex dataset for **'US and UK based clients – Pharmaceutical and Energy Production'** following **Agile and Collaborative approach** to perform data analysis for enhancing and deploying ETL jobs in production linked with 80+ Jira items, using relational DBMS, to improve the project flow performance by 30%.
- Analyzed data for inaccuracies and errors to resolve 60% of major daily issues while **running ETL jobs to store data in Data Warehouse database** in regulating overall profits, sales for various regions, and severity of disease for pharma client; payroll and employee details in various departments using approved financial budget for energy production client.
- Developed **Python and R language** script to automate the initial data analysis process for handling the data formats for 15 raw files.

ACADEMIC PROJECTS (SYRACUSE UNIVERSITY, SCHOOL OF INFORMATION STUDIES)

(I) European Hotel Industry Analysis – Revenue, Market Segment and Reservation analysis – R squared, RMSE (R)
- Executed Hypothesis testing and found the associated rules aligning with the results of SVM, Decision Tree and regression techniques giving 89% accuracy of definite correlation among 120,000 observations spread across 24 features
- Advised the solution to improve customers reservations by 30% annually and to expand at city locations for 50% more revenue by focusing on mode of online reservations done by family and corporate groups.

(II) Vaccination Rates in Californian School Districts – Time series trends, Feature Importance for vaccine costs (R)
- Analyzed 15 years incremental trend and changes in features such as literacy, outreach, population growth.
- Executed Hypothesis testing, Baysian Factor Analysis, MCMC test to highlight the maximum likelihood of significant factors affecting the vaccination rates to develop business strategy and product roadmaps based on analytical insights.
- Recommended actionable insights from features obtained through regression, classification and exploratory analysis focusing on Cook's distance, Q-Q plot, Multicollinearity, R squared, RMSE to predict the vaccination rate with 80% accuracy.

(III) Credit Card Fraud Detection –Fraud or legitimate transaction - Area under PR curve. (PySpark & Python – NumPy, Pandas)
- Identified customer behavior through daily, weekly, and monthly transaction history of 1.8 million records.
- Designed 3-fold cross validated model using Random Forest and Gradient Boosting techniques available in scikit-learn and ML packages by tunning hyperparameters which can detect 99% of highly imbalanced transactions as fraud or legitimate by maximizing the F1 score with 90% accuracy and interpreted insights to reduce frauds.

(IV) Covid-19 Patient's Pre-Condition Analysis – ICU, Mortality and COVID classification - AOC, Accuracy, F-1 (Python)
- Analyzed 1 years' trend on patient's Covid-19 chronology data with 560K observations and interpreted 3-fold cross validated Random Forest and Gradient Boosting models to understand features such as ICU requirement, Covid-19 result, and death based on symptoms, hospitalization, and patient died datetime data.
- Showcased actionable insights to treat all age category patients with the importance of ICU requirement based on certain pre-conditions such as Covid-19, Pneumonia, Immunosuppression to reduce the mortality rate by 70% and save hospitalization costs for less concerned pre-conditions.