Shardul Suryakant Rane

shardul.rane0913@gmail.com | (716)602-3269 | github.com/shardul0913 | https://shardul0913.github.io/ | https://linkedin.com/in/shardul-rane-06142b115/

EDUCATION

- **Master of Science in Data Science** **GPA: 3.62/4.0**
  University at Buffalo, State University of New York Aug 2019 - Feb 2021
  Relevant Coursework: *Machine Learning, Statistics, Data Mining, Probability, Deep Learning, Data Intensive Computing, Data Modelling*
- **Bachelor of Engineering in Computer Engineering** **GPA: 7.52/10.0**
  University of Mumbai Jun 2013 - May 2017
  Relevant Coursework: *Artificial Intelligence, Software Engineering, Data Structures, Database Management, OS, Analysis of Algorithms*

TECHNICAL SKILLS

- **Skills:** Predictive Models, Classification & Regression, Supervised & Unsupervised learning, Hypothesis Testing, EDA, Time-Series Modelling, Anomaly Detection, Marketing Analytics, Fintech Machine Learning
- **Programming:** Python, R, SQL, NoSQL, Java, JavaScript, Apache Spark, Hadoop, Docker, Flask, RESTful API, Google Cloud Console, AWS Lambda, Sagemaker, IBM SPSS, Qlik, Tableau, GitHub, Excel
- **Libraries:** Scikit-Learn, Scipy, Pandas, Pyfolio, Statsmodels, Keras, TensorFlow, Numpy, Matplotlib, dplyr, DAAG, R-Shiny, Dash

ACADEMIC RESEARCH PROJECTS

Fraud Detection Web-tool for Employees Travel Expenses Auditing [**Link**]
- Reduced Manual Fraud Analysis time by tracking incorrect expenses with SVM at 87% and predicted expenses with L2 Regression at 96% accuracy. To tackle data imbalance in univariate time series, implemented SMOTE for regression.

Quantitative Stock Analysis to Examine Industry-wise Effect of 'US China Trade' [**Link**]
- Performed thorough research using Dynamic Time Warping, Hierarchical Clustering and Exploratory Data Analysis on stock returns for Electronics, Textile, Distribution services and Energy companies (Jan 2019 - Sep 2019).

Strategical Arbitrage technique to predict Investment Returns using Unsupervised Bayesian Modelling [**Link**]
- Time series prediction using Gaussian Mixture Modelling and Expectation Maximization on investment returns.
- Tackled non-stationarity and data distribution differences by citing 'Machine Learning in Asset Management' paper

Classification Modelling with 'Dublin Business School' students to Predict Game Outcome [**Link**]
- Implemented Ensemble Learning with 78% Recall accuracy using XGBoost, Random Forest and Support Vector on multiclass dataset with to predict soccer game result. Implemented Grid Search for Parameter Tuning for the models.

Apache Spark Multi-label Natural Language Classifier to track Movie Genres [**Link**]
- Overcame the limitation of Multi-label classifier in Spark by implementing Logistic Regressors for each class
- Improved accuracy of 71% by 13% with feature engineering by validating TFIDF, Word2Vec in Spark Mllib

PROFESSIONAL EXPERIENCE

Data Scientist Intern, *Hewlett Packard (HP) – Boise, USA* **Jun 2020 – Dec 2020**

Developed Forecasting Tool for Printer Toner Usage in 10 markets and gained 93% Accuracy
- Built Logarithmic Regression model with L2 Regularization to predict in-market printer usage to help PMs with feature selection in NPIs. Published R-Shiny model dashboard to service HP's NA, EU and Asia regions. **Tools**: AWS Redshift, SQL, Qlik

Investigated the Impact of Pandemic on Product Sale and Usage in 10 Markets Worldwide
- Presented Quantitative research report by performing Hypothesis Testing on Time-Series dataset with Granger Causality, Dickey-Fuller test to explain the relationship between impact of COVID-19 on personal and enterprise printing volume

Traced Patterns in Counterfeit Product Reviews with 85% accuracy with NLP
- Designed Bi-gram Topic Model and Sentiment Engine for Clone Product Usage. Intimated PMs by highlighting the topic wise keywords for positive/negative reviews. **Methods**: Sklearn, NLTK, Non-Negative Factorization, SVD, TF-IDF, Watson sentiment

Data Analyst, *Performics, Publicis Media – Mumbai, India* **Dec 2017 – Jun 2019**

Engineered User Intent Machine Learning Model to Improve Display Targeting Revenue by 12% for Client
- Developed predictive model with 84% accuracy, using Machine Learning and Google Cloud Platform. Tracked user purchase intent based on search term & demographics. **Methods**: Python, Random Forest Classification, KMeans Clustering, PCA

Formalized Ecommerce Revenue Attribution Model to Reduce Task Time by 79%
- Updated Traditional First touch attribution model based on user session duration and conversion rules with an automated reporting dashboard in Google Analytics. **Tools :** Python, BigQuery, Rest API, Google Tag Manager, Google Analytics,

Gained 14% user traffic for News Website by Introducing Content Data Acquisition Framework and Web Dashboard
- Tracked and analyzed micro features like article impressions, clicks with Python, JavaScript to consult client's content strategy

Boosted conversions by ~16% for 10+ Ecommerce and Finance Firms by Designing BI Framework
- Instituted analytics dashboard and deployed APIs for product performance, A/B tests results, and conversion funnel reports.

ACHIEVEMENTS & LEADERSHIP

- **Certified**: Coursera Neural Networks, Apache PySpark, IBM SPSS, Tableau, Investment Management with Machine Learning
- Achieved AI Explorer badge at HP for print Usage Prediction model, Trainee for employees on Advanced Analytics at Performics