Shardul Suryakant Rane

shardul.rane0913@gmail.com |+1(716)602-3269 | https://github.com/shardul0913 | https://www.linkedin.com/in/shardul-rane-06142b115

EDUCATION

- **Masters of Science in Data Science** **GPA: 3.56/4.0**
  University at Buffalo, State University of New York Aug 2019 - Feb 2021
  Relevant Coursework: *Machine Learning, Statistical Data Mining, Deep Learning, Data Intensive Computing, Probability, Modelling and Query Language*
- **Bachelor of Engineering, Computer Engineering** **GPA: 7.52/10.0**
  University of Mumbai Jun 2013 - May 2017
  Relevant Coursework: *Artificial Intelligence, Analysis of Algorithms, Software Engineering, Operating Systems, Data Structures, Database Management*

TECHNICAL SKILLS

- **Skills:** Predictive Modelling, Exploratory Data Analysis, Hypothesis Testing, Time-Series Modelling, Anomaly Detection, Topic Modelling
- **Programming:** Python, R, MATLAB, Java, C, JavaScript, SQL, NoSQL, Spark, Hadoop, Docker, RESTful API, BigQuery, Sagemaker
- **Libraries and Technologies:** Scikit-Learn, PyTorch, Keras, TensorFlow, Scipy, Pandas, Numpy, Matplotlib, dplyr, DAAG, Vowpal wabbit, R-Shiny, Dash Python, IBM SPSS, Qlik, Tableau, GitHub, Google Cloud Console, Adobe Analytics, Google Analytics, Excel

PROFESSIONAL EXPERIENCE

Artificial Intelligence Intern, *Hewlett Packard (HP) -Boise, USA* **Jun 2020 – Present**

- **Developed Printer Toner Usage Forecasting Tool for Print Products at 93% Accuracy**
  - Built Forecasting model with R-shiny dashboard for 5 years monthly usage prediction in 10 markets to help product managers with NPI features. **Tools** : Logarithmic Regression, R-Shiny, AWS Redshift, MS SQL
  - Tweaked original accuracy of 82% by 11% with Stepwise Selection and Regularization on error metrics RMSE, Adjusted R-Squared
- **Implemented Solution to Analyze Patterns in Counterfeit Product Reviews with 85% accuracy**
  - Designed NLP Bi-gram Topic Model and Sentiment Analysis Engine to intimate PMs with Clone Product Usage Insights.
  - **Methods**: Sklearn, NLTK, Truncated SVD, Non-Negative Factorization, Term Freq-IDF, Watson sentiment
- **Analyzed Product Usage and Share differences in 10 Markets Worldwide**
  - Performed Statistical Significance Tests on Multivariate Time-Series Dataset with Granger Causality, Dickey-Fuller test

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

- **Developed User Behavior and Intent Prediction Model Based on Web Traffic data**
  - Implemented Ecommerce solution to track user purchase intent based on search query and historical data, categorizing in 8 segments with 84% accuracy. **Methods**: K-Means Clustering, PCA Feature Extraction, Random Forest Classifier, GCP, Data studio
- **Designed Website Performance Analysis Frameworks For Ecommerce, Travel and Finance Firms**
  - Tracked Clickstream data to analyze sales, product performance, A/B test results using Python, jQuery, Google/Adobe Analytics
- **Established Custom Ecommerce Revenue Attribution Solution and Reduced Task Time by 200%**
  - Updated Traditional First touch attribution model based on user session duration and conversion rules with an automated reporting dashboard. **Tools :** Python, Google Analytics, BigQuery, JavaScript, Google App Engine

ACADEMIC PROJECTS

Fraud Detection Web-tool for Employees Travel Expenses Auditing [**Link**]
- Tracked false travel expenses with One Class SVM with 87% and predicted expenses with Ridge L2 Regression at 96% accuracy.
- Gained improvement 13% in accuracy by using SMOTE to tackle outlier data imbalance.

Anomaly Detection in Machine Sensors Using Supervised Techniques [**Link**]
- Traced machine failures with 89% accuracy for varying threshold with ensemble of Auto-encoder LSTM Neural Network and ARIMA model on 'Pressure sensor' data. Improved accuracy by 15% by Stabilized Time-series with Time Differencing.

Transfer Learning with Convolutional Neural Network to Identify Stars with Exoplanets [**Link**]
- Classified the stars with 97% accuracy using Transfer Learning between CNNs by Identifying light flux Spatial Pattern
- Implemented Transfer Learning between VGG16 and traditional CNN in Keras with Flask API for model support

Content filtering using Multi-label Classifier in Apache Spark to track Movie Genres [**Link**]
- Overcame the limitation of Multi-label classifier in Spark with collection of multiple Binary Classifiers of Logistic Regression
- Improved accuracy of 71% by 13% with feature engineering by validating TFIDF, Word2Vec in Spark Mllib

Classification Model Testing to Predict Soccer Game Outcome [**Link**]
- Implemented Ensemble Learning with 78% Recall accuracy using XGBoost, Random Forest Classifier and SVM on multiclass dataset with 18 features to predict game outcome. Implemented Grid Search for Parameter Tuning for the models.

Examined Effect of 'US China Trade War' on Domestic and International Companies [**Link**]
- Investigated effect of trade war tariffs in sectors of Electronics, Textile, Distribution and Energy by EDA, Stock data analysis

CERTIFICATIONS

Neural Networks and Deep Learning, AWS Machine Learning, Apache PySpark, Tableau Certification, IBM SPSS, Google Analytics