Amoli Rajgor

amolir.github.io
amoli.rajgor@gmail.com
Ahmedabad, India

• Linkedin - amolirajgor
• GitHub - AmoliR
• Gitlab - amoli13
• Researchgate - Amoli_Rajgor

ABOUT

Data analyst with significant experience in designing, developing and deploying data science projects. Possess understanding about complete Machine learning project life cycle. Keen to get in-depth knowledge about the problem and provide an appropriate solution.

WORK EXPERIENCE

Data Analyst **| Project - Shoppr.ai** | *Apr 2018 - Dec 2020*

Dataone Innovation Labs, Ahmedabad, India

• Implement statistical models as APIs for customer profiling, segmentation, retention and other customer related analytics.
• Design, create and implement the data schema for storing the insights generated by algorithms running on the analytics platform.
• Design and develop machine learning models as *REST* services using any *Python - Flask* framework.
• Develop *MongoDB* query models for Shoppr dashboard metrics.
• Create and evaluate indexing strategies to enhance MongoDB query performance.
• Create documentation of the database architecture.

Research Analyst **| Project - Shoppr.ai** | *Nov 2017 - Mar 2018*

Dataone Innovation Labs, Ahmedabad, India

• Research database and API structure of various e-commerce platform vendors and identify key metrics and data mining techniques that can be used to provide analytical solutions to business problems.
• Understanding and mining data from the sources and then reorganising it into a data model format that can be easily processed by the application.
• Propose and develop e-commerce related statistical metrics and machine learning models for the analytics platform Shoppr.
• Develop database, API and dashboard prototypes (in *R Shiny*) and proofs of concept for the proposed metrics.
• Develop APIs for statistical models using *R* & utilise Apache *Spark* for improving performance of R APIs for large scale data processing.

PERFORMANCE BASED RECOMMENDATION OF IPL PLAYING XI TEAM | POC | Dec 2017 **|Paper - dataone.io/research | Statistics / ML:** *K-Means Clustering*

• Determine and model the performance metrics that influence selection of a player

EDUCATION

M.TECH - ***Computer Science & Engineering***
Ahmedabad University
2014 - 2016 . Ahmedabad
Specialisation in *Data Science & Analytics*
CGPA - *3.52 / 4.33*

B.E - ***Information Technology***
Gujarat Technological University
2010 - 2014 . Gandhinagar
CGPA - *8.76 / 10.0*

SKILLS

Data Science

• ***Python*** **-** pandas, NumPy, scikit-learn, NLTK, PyMongo, SQLAlchemy, Matplotlib, seaborn, Altair

• ***R*** **-** Shiny, ggplot2, dplyr, tidyr, XGBoost, caret, RMarkdown, Plotly, reshape2, scatterplot3d

• ***Deployment*** **-** Docker, Google Cloud Platform, AWS (EC2)

• ***Other Tools*** **-** MongoDB Atlas, BigQuery, Anaconda, Jupyter, Apache Spark

Programming Languages

Python, R, MATLAB, Java, LaTeX

Databases

MongoDB, MySQL

and suggest a set of players for aiding the team selection process.

• Used unsupervised machine learning technique k-Means clustering to categories players and then rank them in their assigned category to evaluate their chance of selection.

THESIS

1. DESIGN AND ANALYSIS OF ALGORITHMS FOR STREAMING DATA | Jun 2015 – Jun 2016

Paper - arxiv.org/abs/1707.08369 | **Codebase - rank1-svd-update**

The purpose of this research was to efficiently modify the conventional algorithms to make them adaptive to streaming data. We reviewed methods such as *Linear Regression, MSE, PCA, PCR, EVD and SVD*. We proposed an efficient algorithm for updating *SVD* of a *rank-1 perturbed matrix.*

PROJECTS

1. NLP BASED BOOK RECOMMENDER | POC | Apr 2022

Codebase - nlp-for-book-recommendation | **NLP:** *keyBERT, TF-IDF*

Content-based book recommendation system that uses NLP. Keywords that most describe the book are extracted from the book description using BERT-embeddings, that is further reduced using the frequentist feature extraction method TF-IDF that ranks the words based on their frequency in the book and the corpus.

2. PAYMENT DEFAULT PREDICTION | POC | Dec 2021 - Mar 2022

Codebase - payment-default-prediction |**Demo - AWS Web App** |

Statistics / ML: *Logistic Regression, Binary Classification, SMOTE*

Predict the probability of an account holder defaulting their next payment based on the information of their account.

3. CLICK-STREAM DATA ANOMALY DETECTION | POC | Nov 2017

Codebase - anomaly-detection | **Statistics / ML:** *Markov Model*

Identify anomalous session activities based on the click-stream collected for sessions. Statistical analysis was carried out using R.

4. PREDICTIVE MODELLING FOR RENTAL PROPERTIES | POC | Sept 2016

Codebase - predictive-analytics | **Statistics / ML:** *Multiple Linear Regression, Support Vector Machine, Regression Trees, Random Forest*

Identify key attributes that influence pricing of the rental properties. Implement various predictive models to predict features like rent price, cost of hosting extra people etc. using real estate data.

5. RISK FACTOR PREDICTION OF CARDIAC DISEASE | Sep 2014 – Dec 2014

Statistics / ML: *Fuzzy C-Means Clustering* | Implemented Clustering algorithm on lipid profile data of patients to predict their likelihood of getting cardiac disease. Statistical analysis was carried out using R.

Web

REST APIs (Python-Flask) & (R - Plumber), Nginx, HTML, CSS3, JavaScript

Source Code Management

Git, Gitlab

LANGUAGES

English - Proficient (IELTS *General score* 8.0) | 2020

Hindi

Gujarati