**Karun Parashar**

[linkedin.com/in/karunparashar](mailto:linkedin.com/in/karunparashar)[|github.](https://d.docs.live.net/Users/karunparashar/Library/Containers/com.apple.mail/Data/Library/Mail%20Downloads/257CB626-DFC0-4959-9590-F86CF1094D02/github.com/karunparashar)com/karunparashar|Parasharkarun.adm@gmail.com|7166039241|Buffalo, NY-14214

***Skills and packages***

Python: Scikit-learn, keras (Tensorflow), TF-IDF, Pandas, Numpy, Selenium, Scikit-image and OpenCV

R: Caret, Rpart, NNET, gbm, glmnet, arules and DNN

Basic regression/classification: Linear, KNN, Regularization, Logistic Regression and Random Forests

Cloud Infrastructure: Google cloud platform AutoML, CloudVision API, BigQuery, Cloud Function Repository management and distributed processing- Git, BitBucket and Hadoop/Spark

Data Visualization: Tableau, PyPlot express, Matplotlib/Seaborn using Python and GGplot

***Professional experience***

Data Scientist**,** Tiger Analytics, Santa Clara, California October 2021 - Present

* Constructed Tableau dashboard for holistic view of medical summary data from 5 different sources in terms of diagnostic, prescriptions and lab tests to make underwriting decisions up to 40% faster
* Conducted data richness analysis from multiple vendors to find gaps in current lab data ordering process to reduce policy result timeline up to 75% for rejection and 45% for acceptance cases and save around $45K for bulk lab data ordering for applicants

Data Scientist**,** MTX Group Inc, Albany, New York March 2021-October 2021

* Sports Analytics application for college basketball teams
  + Built game level line-up analyzer to find best combination of players at any given minute of time to optimize match winning strategies
  + Constructed custom scout report dashboard for microlevel analysis to highlight strengths and weakness to assist coaches in making offense/defense tactics for critical games
* Modelled a sports injury predictor to assist Coaches in making decisions regarding which players to play/rest using cudaML to achieve 91% test accuracy

Artificial Intelligence Intern**,** MTX Group Inc, Albany, New York Jun 2020 - Dec 2020

* Created real-time predictive pipelines using Google Cloud Platform for critical COVID 19 stats - fatalities, recovery rate and hospitals utilization with the help of LSTM and statsmodel.api to enable hospital beds allocation and lockdown planning by state government
* Automated data crawling using selenium and deployed those scripts in Kubernetes and containerized employing Docker to run on cloud clusters to eliminate 70% of project LOEs
* Performed data manipulation/wrangling using SQL, python ETL libraries and constructed visualization tools from python’s matplotlib and ipywidget for representing business insights

Associate Analytics Consultant**,** Capgemini Technology, Bengaluru, India Jul 2018 - Jul 2019

* Implemented relational script to handle monthly data duplication and tape management using SQL based tools and in server console to automate monthly tape ordering/dispatch
* Performed extensive analysis on incident/problem reports to identify redundant events and unresolved archaic open problems to downsize client’s server maintenance budget by 20%

Data Scientist**,** HCL Technologies, Noida, India Mar 2016 - Jul 2018

* Built SVM regression models using maintenance jobs runtime, failure frequency, etc. to forecast and timely communicate system upgrade timings to save 200 project planning hours annually
* Developed robust incident classification models by combining parametric and non-parametric models (Nearest neighbor) and secured more than 95% accuracy

Test Analyst**,** Akon Electronics India, Haryana Mar 2015 - Mar 2016

* Built time series forecasting model using ARIMA on electronic modules from daily manufacturing to estimate power, voltage, impedance and RF/Video tuning ranges
* Analyzed frequently failing electronic units to identify the root cause and worked with cross

functional teams to devise an alternate configuration for power supply across different units leading to 40% reduction in component failure and hence component prices

***Projects and Challenges***

Movie genre prediction-Natural language processing

* Instituted a PySpark based prediction pipeline of Regex, Stop-words removal and Word2Vec vectorizer for multi label movie genre prediction and achieved 93% macro F1 score

Segmenting Building images for Disaster resiliency- Open Cities AI

* Developed a U-Net architecture model for segmentation of satellite building image(partitioning into black and white segment for easier identification) to attain 90% segmentation accuracy