# Raghvendra Pal

DataScientist / Software Developer

## WORK EXPERIENCE (2+ Year) :

### Data Scientist at Infosys Limited

Location - Pune, Maharashtra 02/2018 - PRESENT

- Responsible data pre-processing of unstructured data (data cleaning, handling missing values and re-sampling to required frequency)..
- Have work on Image Processing and Natural Language Processing projects.
- Developed model using pre-trained deep learning models(VGG16, ResNet50), pre-train text processing vectors like Glove Vectors.
- Used euclidean distance, binary cross entropy to calculate loss.
- Used Bag of Word, TF-IDF, Word2Vec techniques for text processing.
- Have worked on LSTM, CNN, Faster RCNN Model to perform image and text tasks.
- Performed EDA and Feature Engineering to identify most contributing features.
- Used confusion metric, F1 score, precision, recall metrics to measure performance of model.

### 7Eye IT Solutions

Location - Jabalpur, Madhya Pradesh 08/2017 - 01/2018

- Developed a Desktop Based application from scratch, which is used to identify how much tax will be applicable on products.
- Application maintain all the history of user and can generate a report on the basis of dates.
- Worked closely with clients to establish problem specifications and system designs.
- Collaborated with other developers to identify and alleviate number of bugs and errors in the software.
- Built databases and table structures for the application.
- Developed code fixes and enhancements for inclusion in future code releases and patches.

## EDUCATION :

**Government Engineering College, Jabalpur(M.P.)** 07/2013 – 05/2017

Bachelor Of Engineering(Computer Science & Engineering) 6.48 CGPA

**Madhya Pradesh Board, Jabalpur(M.P.)**

Higher Secondary School (XII) 07/2011 – 05/2012 74%

High School (X) 07/2009 –05/2010 86.33%

## PERSONAL INFO:


**Email** : raghvendrapal65@gmail.com

**LinkedIn :** linkedin.com/in/raghvendra-pal-b98228104/

**GitHub :** github.com/RaghvendraPal

**Place :** Pune, Maharashtra (411057)

**Contact no :** +91 8319903602


## TECHNICAL SKILLS :

**Programming Languages :**

Python, Keras, TensorFlow, Java, HTML, CSS.

**Database :** Oracle, MySQL.

**Machine Learning :**

Time Series Analysis, Linear Regression, SVM, Decision Tree, Logistic Regression, KNN, Random Forest, Naive-Bayes Classifier, XGBoost, Stack.

**Deep Learning :**

Convolutional Neural Networks(CNN), Long Short Term Memory(LSTM), Faster RCNN, Neural Network, U-Net.

**Data Visualization :** Box plot, histogram, CDF,Python.

**Statistics :** PCA, t-SNE , Hypothesis Testing, A/B Testing.

## STRENGTH :

Self Confidence

Self-Motivation

Problem Solving

Critical Thinking

Interpersonal Skills

## PROJECTS :

### Human Mask detector to provide security in ATM :

- Performed Pre-Processing
  - RGB to Gray
  - Image Resizing
  - Image Scaling
- Used Faster RCNN with ResNet50 pre-trained model.
- Used RPN layer to create bounding box of mask and ROI layer to classify that the particular portion of image has mask.
- Used Euclidean distance to improve bounding box error.
- Trained the model using ANN to Classify Masks.
- Used Normalization to handle outlier problem.

**TECHNOLOGY USED:**

- Python
- Deep Learning Model(CNN)
- Scikit-learn
- Keras
- OpenCV

### Banking System :

- Have Developed a Deep Learning System which is used to predict whether a customer will remain use his bank account or not.
- Have used LabelEncoder and OneHotEncoding to convert categorical variable into numerical variable.
- Normalized the numerical feature variables.
- Used binary Cross Entropy as loss function because this is classificatior type problem.
- Tried different epoch and batch size to improve accuracy.

- Python
- Artificial Neural Network(ANN)
- Scikit-learn
- Keras

### GST Desktop Application :

- Have Developed Desktop Application which is used to calculate tax applied on products of the shop.
- Used Jasper Report to show Amount paid by Customer on Monthly and Yearly basis.
- Used Swing technology to create the front end of software.
- Used MySQL connectivity api to connect Java environment with MySQL.

- Java
- MySQL
- Jasper Report
- Swing
- OOP(Object Oriented Programming)

## PERSONAL PROJECTS :

### LANL EarthQuake Prediction Challenge (Kaggle Competition):

This competition is much important too get to know when earthquake happen. we will try to predict time left to the next laboratory earthquake based on seismic signal data.

- This is Time Series Data, In Data processing step we have divide time series data into multiple chunks of 1,50,000 size each.
- Used Statistical Feature, Rolling Window Feature and Fourier Feature to get important data like min, max, value from chunks.
- Used MAE(Mean Absolute Error) as performance matrix.
- Used SGDRegressor, RBF SVM, Random Forest, XGBoost algorithms to train the model.
- Have saved rank between 23 to 30.

- Python
- Machine Learning
- Scikit learn
- Random Forest
- XGBoost
- RBF SVM(Kernel SVM)
- Linear Regression
- Keras
- Deep Learning
- LSTM

### BLOG :

https://medium.com/@raghvendrapal/lanl-earth-quake-prediction-challenge-846d47e0a0a0

**Quora Question Pair Similarity (Kaggle Competition) :**

Identify which questions asked on Quora are duplicates of questions that have already been asked.

- Did data preprocessing step for data cleaning.
- Did Feature extraction step to get useful information about data like similarity between two questions, length of each question.
- Used t-SNE for visualization and log-loss, confusion matrix as performance metric.
- Trained different models on the normalized dataset.
- Performed hyperparameter tuning to improve performance of model.

- Python
- Machine Learning
- Scikit learn
- Logistic Regression
- XGBoost
- Linear SVM(Kernel SVM)
- TF-IDF vectors
- Log-Loss

**Personalized Cancer Diagnosis (Kaggle Competition) :**

- Did data preprocessing step for data cleaning.
- Did Univariate Analysis on data to get the importance of each feature.
- Used features which performed well on univariate analysis.
- Used Multi class log-loss and confusion matrix as performance metric.
- Trained different models on the normalized dataset.
- Performed hyperparameter tuning and different Feature Engineering Techniques to improve the performance of model.

- Python
- Machine Learning
- Scikit learn
- Naive Bayes
- KNN
- Logistic Regression
- OneVsRest
- Random Forest
- Linear SVM(Kernel SVM)
- Stacked Model
- Multi Class Log-Loss

**Facebook Friend Recommendation Problem (Kaggle Competition) :**

- Used different types of techniques to get features from data like page rank, cosine distance, number of followers etc.
- Used F1 Score and confusion matrix as performance metric.
- Trained different models on the normalized dataset.
- Performed hyperparameter tuning and different Feature Engineering Techniques to improve the performance of model.

- Python
- Machine Learning
- Scikit learn
- Random Forest
- XGBoost
- F1 Score

**New York Taxi Demand Prediction (Kaggle Competition) :**

- Performed Clustering techniques to identify the different areas in new york.
- Tried different features from data like latitude, longitude, trip distance etc.
- Used Fast Fourier Transform method to extract features from data.
- Used F1 Score, MAPE and confusion matrix as performance metric.
- Trained different models on the normalized dataset.
- Performed hyperparameter tuning and different Feature Engineering Techniques to improve the performance of model.
- Have reduced the MAPE 12% from previous results.

- Python
- Machine Learning
- Scikit learn
- Random Forest
- XGBoost
- Linear Regression
- F1 Score
- MAPE

**Netflix Movie Recommendation (Kaggle Competition) :**

- Did Data Preprocessing step to clean the data.
- Used User-User similarity and Item-Item Similarity matrix to extract features.
- Handled Cold Start Problem.
- Used RMSE as performance metric.
- Trained different models on the normalized dataset.
- Performed hyperparameter tuning and different Feature Engineering Techniques to improve the performance of model.
- Have reduced the MAPE 12% from previous results.

- Python
- Machine Learning
- Scikit learn
- Random Forest
- XGBoost
- Surprise Baseline
- F1 Score
- MAPE

**E-commerce Item Recommendation System :**

- Did Data Preprocessing step to clean the data.
- Used Content Based approach to get similarity between products.
- Tried TF-IDF, IDF,Word2Vec, Bag of Word techniques to get features from text data.
- Used Euclidean distance to get most similar products.
- Used Images to get features.
- Used Machine Learning as well as Deep Learning techniques to get similar products on the basis of image.

- Python
- Machine Learning
- IDF, TF-IDF, Word2Vec
- Deep Learning(CNN)
- Machine Learning + Deep Learning
- Euclidean distance

**Handwritten MNIST Digit Identification System :**

- Performed Pre-Processing
  - Image Resizing
  - Image Scaling
- Used Data Augmentation step to create more images from existing data at run time.
- Used BatchNormalization, Dropout to avoid underfitting and overfitting.
- Used Adam optimizer and Categorical Cross Entropy to improve performance of model.
- Used Different architecture of model.

- Python
- Deep Learning(CNN)
- Dropout
- BatchNormalization
- Euclidean distance

**Human Activity Recognition :**

This project is to build a model that predicts the human activities such as Walking, Walking_Upstairs, Walking_Downstairs, Sitting, Standing or Laying.

- Did analysis on time series data.
- Used different techniques like FFT , sliding window protocol to extract features from data.
- Data Visualization :
  - Histogram
  - Box plot
  - t-SNE
- Applied Machine Learning and Deep Learning model on data..
- Used BatchNormalization, Dropout to avoid underfitting and overfitting.
- Used Adam, Rmsprop optimizer and Categorical Cross Entropy to improve performance of model.
- Used Different architecture of model.

- Python
- Machine Learning
- Logistic Regression
- Linear SVC
- Kernel SVC
- DecisionTree
- Random Forest
- GradientBoosting DT
- Deep Learning(CNN)
- TimeDistributed+CNN+RNN
- Dropout
- BatchNormalization

**Self Driving Car Steering Angle Prediction :**

- Used continuous images of road.
- Applied different architecture of CNN with ‘relu’ and ‘tanh’ activation function to predict the angle of steering.
- Performed Pre-Processing
  - Image Resizing
  - Image Scaling
- Used mean square error as loss function.
- Used Adam and Rmsprop as optimizer.

- Python
- Deep Learning(CNN)
- TensorFlow
- Dropout
- BatchNormalization

**Object Detection :**

- Performed Pre-Processing
  - Image Resizing
  - Image Scaling
  - Three channel image
- Used bounding boxes to tell the model about image at training time.
- Used square error(euclidean distance) as loss function.
- Used faster RCNN model to identify the object.
- Used RPN(Region Proposal Network) to create bounding box and identify whether there is an object or not.
- Used ROI layer to classify object.

- Python
- Deep Learning(CNN)
- Faster RCNN
- Dropout
- BatchNormalization

## DECLARATION :

I hereby declare that the information and facts furnished above are correct and true to the best of my knowledge and Belief.

RAGHVENDRA PAL