Raman Choudhary

Data Scientist

Contact

Address
Bengaluru, KA, 560002

Phone
9962745947

E-mail
raman.iitm2017@gmail.com

LinkedIn
https://www.linkedin.com/in/raman-choudhary-158230

Skills

- Python — Very Good
- Machine learning — Very Good
- Deep learning — Good
- Natural langauge processing — Very Good
- Text mining — Good
- Data cleaning — Good
- Feature engineering

A passionate problem solver seeking an opportunity to apply my skills in solving everyday problems. I have a keen for identifying customer centric problems and providing a solution for the same. I like to keep myself updated with the latest technologies and apply the same in real time situations.

Work History

2021-05 - Current

Sr Data Scientist

koch industries, Bengaluru, Karnataka
Working with pricing analytic of molex(sub division of koch).Building price optimizations model which include segementing the material based on its attribute and coming up with new price .There are few project on forecasting the price based on time series model,building recommendation engine for price in various region.

2020-10 - 2021-04

Sr Data Scientist

huawei technologies india pvt ltd, bengaluru, karnatka
AdSearch:
working with adsearch team for creating traning data so that it can support 16 different langauge.I am working for vitanameses preprocessing like creating token,lemma,pos,ner etc for it.For tokenisation we are trying various technique like standard core nlp,bert tokeniser etc.For pos we are using bert pos tagging and some rule based method like deeppavlov which create rule file and dict file while training and then we can add some other rule based on our resarch to handle number,abbreviation,url etc.Since lemma is same as token in vitanamese ,so no preprocessing on lemma.

2020-03 - 2020-10

Senior Resarch Analyst

inmobi, Bangalore, karnatka
Senior Resarch Analyst
Inmobi , bengaluru, karnatka
Project 1:
Telco project:
We had location data of all user who are using sprint

Good

Computer Vision

Good

AWS SAGEMAKER

Good

Pyspark

Good

Object detection

Good

Recommendation system

Good

Content filtering and collaborative

Very Good

Image classification

Transformer

Good

Bert

Good

Encoder decoder

Good

Software

Python

Good

Pandas

Very Good

Keras

Good

telecom in us.Frequency of data gathering was done in a timestamp of every 15 minutes from pinsightmedia.We need to convert this lat long to feature to make sense out of it.One such feature was number of visit for each store in a day. We need to scrape the public dataset from yahoo finance and then map our visit datapoint based on dates. Then we need to do the prediction using various technique like Random Forest, Xgboost, LSTM etc. We need to compare result of both prediction with or without visit. It was good learning working with LSTM and tuning some of its parameter.

Project 2:

Customer churn prediction: The project aimed at predicting whether the customer of a telecom network will churn or not. It was for a telecom giant sprint in united states. So we had various feature which can give sense of behaviors of customer from historical data. Various technique was tried for model building like logistic regression,Svm,random forest etc. The logistic model was best in predicting .

Project 3:

Object detection model:

We need to scrape the URL of all camera feed for california transportation website and then build a objection model on top of it to count the number of vehicle passing through each traffic signal's have used a OpenCV package named Cvlib which has inbuilt yolv3 trained on coco dateset for 80 different object .Here I had good exposure of scraping dynamic image from website, parsing KML file and extracting location, image link etc.

Project 4:

Telematics solution for motor company:

Given the Telematics data of vehicle ,we need to identify the driver.

Since the Telematics data are time series sensor data so classical machine learning model doesn't work here .So we tried with encoder decoder with attention using lstm layer to capture the sequential information.

2017-05 -

Data Scientist

Pyspark

Good

NLTK

Good

Numpy

Good

Spacy

Good

Languages

English

Very Good

Hindi

Very Good

Assamese

Good

2019-12 *Honeywell, Bengaluru, Karnatka*

Data Scientist I

Honeywell Technology Solution, bengaluru, karnatka

Project 1

Time series forecasting: I was working with aero tool team to build a time series model for labour cost forecasting. We had biweekly data for various team and we need to clean the data and build model which can predict it for next six month. We started with arima and then we tried fb prophet.And with prophet we were able to reduce the Mape to 8 percent.

Project 2

Aircraft tyre life prediction: This project aimed at predicting the life of aircraft tyre so that we can know the inventory. There was various sensor which was attached to an aircraft tires. Whenever the flight lands ,the data get stored in azure sql database. Then we had scrapped the data and created a resource in Azure. We did the cleaning and tried various model like linear regression with lasso and random forest regressor . With RandomForest regressor we were getting better accuracy.

Project 3

Sentiment analysis: This project aimed at creating sentiment analysis model on the comment received from leadership people and partner for one of our product. We wanted to see how it is preforming and what additional improvement we can do on top of it by analyzing negative comments. We did all the text preprocessing and then we tried with various feature engineering technique to convert raw text to feature like bag of word,TFIDF,word2vec etc .Then we build logistic regression and SVM model on top of it .With logistic regression it was working fine.

Project 4:

Topic modelling: There are various program which comes to honeywell and each program has some failure or defects. Those
defects or failure report for each program was documented in the past. Now we need to build a model which will identify the defects from any project/program.

So basically we need to build unsupervised model which can cluster the segment of text data . We used LSA for finding few topic and classify new document to one of those topic. The main idea was to identify defect from new program so that we can train our resources on those defects.

Project 5:

Provider fraud detection in insurance claim:
The goal of this project is to " predict the potentially fraudulent providers " based on the claims filed by them.along with this, we will also discover important variables helpful in detecting the behaviour of potentially fraud providers. further, we will study fraudulent patterns in the provider's claims to understand the future behaviour of providers.

Education

2015-05 - 2017-06

M.TECH

IIT MADRAS - Chennai

2009-06 - 2014-05

B.Tech: Mechanical Engineering

SRM University - CHENNAI

Certifications

2020-05	AWS Sagemaker
2018-02	Neural Network and Deep Learning
2017-06	Introduction to Datascience Python