Vipul Rai

Linkedin - https://www.linkedin.com/in/vipulrai/
Blog - https://vipulrai.me

vipulrai8891@gmail.com
+91-8892598819
Bengaluru, Karnataka

7+ years of experience in handling petabytes of data, implementing engineering pipelines, predictive modeling, machine learning, and deep learning. Python expert, passionate about using analytics to drive strategic business decisions.

Skills/Algorithms/Platforms

- TensorFlow, Pytorch, Keras, Fast.ai
- Decision Trees, Random Forest, Logistic Reg
- Supervised, Unsupervised & Transfer Learning
- AWS (S3, EC2, Lambda, API Gateway)

- CNN, RNN, LSTM, GAN, Transformers, BERT, DETR
- KNN, K-means, SVM, XGBoost
- Python, Django, MySQL, Hive, Spark

Professional Experience

Senior Data Scientist
Jun 2017 – May 2020
Bengaluru

SmartNomad

- Designed and implemented end to end pipeline to create personalized itinerary generation engine (Flights, Hotels, Restaurants, POIs) for travelers
- Object Detection to select similar images and discard images containing unqualified objects using YOLOv3 Topic analysis for tagging categories of each point of interest (POI) using Latent Dirichlet Allocation in sci-kit learn
- Multi-class Sentiment Analysis for POI reviews using BERT in PyTorch and HuggingFace Transformers
- Developed Video Creator app using OpenCV and Moviepy
- Integrated multiple Google APIs (distance matrix, maps, places) and other partner APIs Implemented a serverless microservices architecture on AWS to support recommendation modules
- Led and mentored a team of 7 people (Data scientists and Data Engineers)

Consultant
Oct 2015 – May 2017
Bengaluru, Seattle & Los Angeles

Affine Analytics

- **Gaming Analytics** - Fortune 100 client, the biggest game publisher in the world
  - Developed the end to end pipeline for data orchestration to consumption using PySpark and Kafka
  - Developed and deployed a fraud detection model to help client ban players using cheats and violating game rules using Random Forest Classifier
  - Wrote an automated PySpark program which calculated user stats on daily, weekly, and monthly level
- **Retail Store Revenue Optimization** - Fortune 100 Retail giant
  - The pricing team was tasked to identify optimal promo/clearance sale prices which would result in maximum sales.
  - Identified multiple feature which impact unit sales during promotions such as Price Elasticity, seasonality, % sales in each day of the week, the effect of circulars/promotions, breakage due to unavailability of different sized SKUs, the impact of public holidays, etc. using Spark-Scala
- **Travel Portal Optimization** - Fortune 100 Online Travel brand
  - Gave an optimized solution to reduce the number of unqualified searches it is getting from meta channels, to improve the business KPIs like ROI/profitability/efficiency
  - Wrote the Optimization Program using LPsolve from Pulp package using PySpark and logistic regression

Big Data Developer

AIG (American International Group)

Jan 2015 – Oct 2015
Bengaluru

- Designed and built a POC and projects around Spark and Big Data Ecosystem

Software Engineer

Blue Star Infotech

July 2013 – Jan 2015
Bengaluru

- Learned and worked around Big Data tools such as Hadoop, Hive, etc.

Personal/Open Source Projects

Deep Learning - https://github.com/vipulrai91/my-projects/wiki/Projects-Home

- MNIST digit classifier from scratch - built a classifier using Keras, TensorFlow to classify MNSIT data from scratch
- Binary classifier on browser - transfer learning, hosting as web service using VGG16 and Flask on a live browser which allows user to upload an image and it returns predictions whether it was a cat or a dog
- Sign Language digit classifier using MobileNet Keras - built a classifier to classify the sign language hand digits and deployed to as web service using TensorflowJS

pandas-dev/pandas

- Active contributor to pandas-dev, a participant in discussions, bug fixes and documentation, test, and generic topics

Kaggle

- Google Research - quickdraw-doodle-recognition - used Pytorch and resnet34 model to build a classifier - secured bronze medal and rank 119/ 1316 - top 10%
- Tweet Sentiment Extraction - Extracting support phrases from sentiment labels - built model using TensorFlow, Keras and Transformers API roBERTa - ongoing
- Google Analytics Customer Revenue Prediction - build the model to predict revenue per customer using XGBOOST - secured rank 717/3611 - top 20%
- Human Protein Atlas Image Classification - build a model to classify mixed patterns in protein images using PyTorch and resnet34 - secured top 23% rank (493/2169)

Certifications/Courses

- Databricks Developer Certification for Apache Spark
- Deeplearning.ai - Coursera Deep Learning Specialization

Education

Computer Science and Engineering Jawaharlal Nehru Technological University   2008 – 2012