



# Vipul Rai

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

## Summary

Analytical and process-oriented engineer with a bachelor in computer science degree, in addition to 7+ years of professional experience handling multiple projects and teams. Python expert, passionate about using analytics to drive strategic business decisions. Excellent interpersonal skills, able to collaborate effectively with technical and non-technical people.

## Blog

<https://vipulrai.me>

## Skills

- DL Frameworks - Tensorflow, Pytorch, fast.ai
- Big Data - Spark, Kafka, Hive
- DL Models - Transformers, BERT, GAN, DETR
- Cloud - AWS Lambda, S3, EC2, API Gateway, SageMaker
- Dev Tools - Git, Github, Jupyter, VSCode, JIRA
- Linear Regression, KNN, K-means Clustering
- Python
- DL Layers - CNN, RNN, LSTM
- ML libraries - Pandas, Numpy, scikit-learn, OpenCv
- Decision Trees - Random Forest
- Logistic Regression, SVM, XGBOOST, KNN

## Experience

### Senior Data Scientist

SmartNomad • Bengaluru, Karnataka  
09/2017 - 05/2020

Designed and built algorithm which calculates the scores for each point of interest

Developed a microservices architecture and built the entire pipeline beginning from data aggregation to the final itinerary generation

Used Python and Spark to scrape, clean, and analyze point of interest data

Developed video creator app using opencv-python and moviepy

Used multiple Google APIs(distance matrix, maps, places) and integrated partner APIs

Managed a team of 7-9 people and undertook project management

Collaborated with UI/UX designers, app developers and business to create an optimal customer experience

Onboarded flight, hotel and booking partners to create inventory and enable booking experiences

### Consultant

Affine Analytics  
10/2015 - 05/2017

Joined as Senior Business Analyst, got promoted within 9 months to Consultant

Daily routine included - Code Review, managing scrum updates, fixing pending issues, planning for the day/s, updating business requirements, scheduling new tasks, reporting the progress to higher management/stakeholders

### Selected Projects:

- **Gaming Analytics** – fortune 100 client, 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 company ban players using cheats and violating game rules using RandomForestClassifier
  - Wrote an automated Pyspark program which calculated user stats on daily, weekly and monthly level
- **Retail Store Revenue Optimization**– fortune 100 american retail giant
  - The pricing team was tasked with the challenge to identify optimal promo/clearance sale prices which would result in maximum sales.
  - Identified multiple factors which impact unit sales during promotions such as Price Elasticity, Seasonality, % sales in each Day-of-the-Week, effect of circulars/promotions, breakage due to unavailability of different sized SKUs, impact of public holidays, etc.
  - Developed the algorithm using Spark – Scala
- **Travel Portal Optimization** – To reduce the number of unqualified searches on portal's meta channel
  - Gave an optimized solution for a leading online travel brand 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 (Pyspark) and Random Forest from MLib

## Big Data Developer

AIG (American International Group)  
01/2015 - 11/2015

Worked as Big Data Developer to internal AIG teams across the globe.  
Designed and built a POC around Spark and Big Data Ecosystems

Achieved Databricks certified Spark Developer Certification

## Software Engineer

Blue Star Infotech  
07/2013 - 02/2015

Worked as Software Engineer, joined here as a fresher.  
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 dataset digits from scratch
- Binary classifier – transfer learning, hosting as web service using VGG16 and Flask on 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, participant in discussions, bug fixes and documentation, test and generic topics

### Kaggle

Participant in multiple competitions, some being

- 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
- oogle 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 model to classify mixed patterns in protien images uisng PyTorch and resnet34 – secured top 23% rank (493/2169)

---

## Education

### Computer Science and Engineering

Jawaharlal Nehru Technological University

---

## Languages

- English
- Hindi