

BIRAJ PARIKH

Bloomington, Indiana | 812-650-2716 | bparikh@iu.edu | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Summary: Experience Data Scientist skilled at ingesting, analyzing, interpreting large datasets and developing predictive models to unravel business problems. Possessing an extensive analytical skill with a passion to solve complex problems using Deep learning.

EDUCATION

Indiana University, Bloomington

Aug 2019 – May 2021

Master's in Data Science

GPA: 3.8 / 4

Coursework: *Applied Algorithms, Data Mining, Machine Learning, Computer Vision.*

Dwarkadas J. Sanghvi College of Engineering, University of Mumbai, India

Dec 2013 – May 2017

Bachelor of Engineering in Mechanical Engineering,

CGPA: 3.7 / 4

Relevant Coursework: *Linear Algebra, Calculus, Statistics and probability theory.*

PROFESSIONAL EXPERIENCE

Reliance Jio Infocomm Limited, Mumbai

May 2018 – Jun 2019

Data Scientist

- ◆ Worked with cross-functional stakeholders to streamline automated scalable data pipelines resulting in **20% redundancy reduction**.
- ◆ Optimized Real-Time and Robust spark jobs by **15%** to transform the streaming data into the optimal format and making fault-tolerant code on the On-Prem (production) cloud data lake environment.
- ◆ Independently analyzed the performance of complex Hive scripts to find anomalies in the designed architecture.
- ◆ Reporting analyses by building powerful Dashboards on ZoomData to visualize transformed and aggregated streaming data.

Piramal Corporate Service Limited, Mumbai

Oct 2017 – Apr 2018

Data Science Intern

- ◆ Implemented a **predictive and prescriptive model** to predict whether the debtor will repay the loan amount or not based on the past transaction behavior of the dealers which benefitted the organization making a tangible business impact.
- ◆ Innovated web-scraping framework as another measure to validate user information which accounted for **5%** in the overall decision.

PROJECTS

Glassdoor Analytics (Python, Machine learning, Deployment) [[App Link](#)]

Feb 2020

- ◆ Deployed an ML model to predict whether customer will renew the job slot product subscription for the Glassdoor Analytics use-case.

Amazon Fine Food Reviews (NLP, Machine learning classifiers, Pandas, Numpy, Jupyter Notebooks)

Dec 2019

- ◆ Performed Data Cleaning & Data Preprocessing like removed html tags, punctuations, Stop-words and Stemmed the words using Porter Stemmer.
- ◆ Applied K-Nearest Neighbour (NN), Naive Bayes using Bernoulli NB & Multinomial NB, Logistic Regression, SVM with rbf(radial basis function) kernel, Decision Trees, Random Forest and validated the models efficiency.
- ◆ Evaluated and assessing the test data on various performance metrics like accuracy, f1-score, precision, recall, etc. also plotted Confusion matrix using seaborn.

Recommender System (Python, Machine learning, Collaborative filtering, Content-based filtering)

April 2018

- ◆ Designed and developed a system to possibly recommend the insurance policies to the customers based on the historical data.
- ◆ Performed exploratory data analysis on the past transactions of the customers to get meaningful insights.
- ◆ Implemented Item-Item and User-Based collaborative filtering.

Classification using Natural Language Processing (Python, NLP, NLTK, Matplotlib, Machine Learning Classifiers)

July 2017

- ◆ Designed and implemented a multi-label classifier to detect the user intent and classify the user chat into the following categories like travel, recharge, nearby, food, movies, reminders, and others.
- ◆ Pre-processed raw data using tokenizer, stop word removal, stemming and lemmatization using NLTK library.
- ◆ Utilized sklearn (One-vs-Rest Multi-label Classification) and visualized the output through word clouds.

TECHNICAL SKILLS

- ◆ **Machine Learning:** Classification, Regression, Clustering, Feature engineering, Neural Nets, Forecasting, Statistical techniques.
- ◆ **Languages and Scripts:** Python, R, Keras, Scala, PySpark, SQL, BigQuery, shell-scripting, Flask.
- ◆ **Python & R Libraries:** Pandas, Numpy, Scikit-learn, NLTK, Requests, Beautiful Soup, Plotly, Matplotlib, tidyverse, ggplot2, dplyr.
- ◆ **Framework:** Hadoop, Apache Spark, HBase, Hive, Nifi, Apache Kafka, Apache Airflow, Spark Streaming, AWS, GCP.fi
- ◆ **Databases & Tools:** MySQL, NOSQL, MongoDB, Neo4j, Tableau, Git, GitHub, MS Office tools, Linux, Docker.

LEADERSHIP AND EXTRA/CO-CURRICULAR ACTIVITIES

- ◆ Led a team of 30 volunteers for conducting campaigns such as Beach Cleaning drive, Safety health measures, Reduce Water wastage to spread awareness and healthy environment.
- ◆ Worked with Indian Development Foundation (IDF) - a National NGO for organizing various events like Blood Donation Drives and fortunate to be a part of the Swachh Bharat Campaign, world's largest cleanliness initiative.
- ◆ Completed a four-month long Data Science Engineering Certification course at GreyAtom, Mumbai.
- ◆ Wrote blogs on the topic related to machine learning and statistics [@medium.com](#).