

# Atul Rai

## ML Engineer

**Address:** Nashik, India, 422207

**Phone:** (+91) 9493847067

**E-mail:** iamatul1214@gmail.com

**GitHub:** <https://github.com/iamatul1214>

**LinkedIn:** [www.linkedin.com/in/iamatul1214](http://www.linkedin.com/in/iamatul1214)

3 years of total experience in the IT industry that includes predictive modeling, data processing, Machine Learning and deep learning algorithms(NLP & CV), hands-on experience leveraging machine learning, deep learning models to solve challenging business problems.

## Skills and Expertise

### Key Skills

Predictive Analysis, Machine Learning & Deep Learning Algorithms, statistical modeling, Computer Vision, Natural Language Processing, AIOps, Python scripting and automation, Web scrapping.

### Database and Programming Languages

MySQL, MongoDB, Python, JAVA

### Clouds

Heroku, GCP, Azure (Basic)

### Statistics

Descriptive Statistics, Inferential Statistics

### Libraries

Pandas, Numpy, scikit-learn, Seaborn, Matplotlib, Flask, TensorFlow, PyTorch etc.

### CI/CD tools

Github actions, CircleCi, Docker

### AIOps

MLFlow, Data Version Control and TFX

### IDEs

Eclipse, Google Colaboratory, VSCode, PyCharm

### OS

Windows, Linux

## Work Experience

**2021-03 - Current**

### Software Engineer

*Mindtree Ltd., Bangalore, Karnataka*

#### Project 1: Auto Xpath Reviewer

**Technology used:** Naive bayes, Google cloud platform

**Summary:** This application reduced manual review task of atleast half an hour to few mins. The mindtree inhouse automation framework requires the selenium developers to store all the xpath objects in one resource file. Our system takes the resource file as input which contains all the xpath objects which are required for running automation scripts. It processes the resource file into a data frame with all the xpaths and classifies each xpaths from good and bad xpaths. The report is generated and gets downloaded automatically. The report generated is also shared with the Leads and managers, so that they are aware of the performance of the automation tester.

**Future scope:** We are planning to integrate feedback system to understand the performance of the model from users.

#### Project 2: Digital ML platform

**Technology used:** Flask, Django, Machine Learning Algorithms, MongoDB

**Summary:** Digital ML platform, is an end-to-end machine learning platform which doesn't require code. It is form-based application, which takes the input of the hyper parameters and other related inputs and create a model on click. This system is created ideally for EDA/Pre-model creation before starting any ML project by managers or anyone who wants to precheck before starting actual coding.

#### Project 3: Back-order prediction

**Technology used:** K-Mean Clustering, XG boost and Random Forest, logistic regression, MongoDB, Heroku, docker with circle-ci.

**Summary:** This project was done as part of Mindtree Hackathon fair. Here, we accepted the challenge

of solving backorder prediction for ecommerce organizations and built an end-to-end system which takes the multiple csv files as input and predicts those items are going to be backorder or not. The system is designed to be smart enough to perform retraining automatically.

2019-08 - 2021-03

## Senior Development Associate

*NTT DATA Global Delivery Services, Bangalore, Karnataka*

**Project 1:** Web Scrapping for customer reviews analysis

**Technology used:** Python, Selenium, MySQL

**Summary:** This project was a collaborative effort for creating a system which can scrap all the reviews of any particular product available on Honeywell aerospace website. Store all the reviews in the database (MySQL). Then to retrieve the stored reviews from database and convert it into .csv format so that Data analyst can analyze the sentiments of the users and take actions on the products sale. Created a system with UI which enables the BA to retrieve the reviews from the Honeywell applications.

**Project 2:** Data Cleaning & Storage

**Technology used:** Python, Excel, MySQL

**Summary:** This project was part of HAIL engineering where our task was to clean data reports available in .csv format to a proper format, there were more than 50 reports, so we created a system which will input the csv file and report name, perform the cleaning operations and store each report file into database. So that later on these datasets were retrieved from database by machine learning team and used for modeling to predict the effectiveness of the sensors.

## Personal Projects (<https://github.com/iamatul1214>)

1. **NanDealer**- A library available on PYPI for handling the NaN values in your dataset. Just do Pip install NanDealer and use it.
2. **CatDealer**- A library available on PYPI for handling the categorical values in your dataset. Just do pip install CatDealer and use it.
3. **System volume controller** using hand gesture.
4. **Hand sign recognizer** for special abled people.
5. **Movies recommendation system** using collaborative filtering.
6. **Twitter sentiment analysis** using BERT, Bi-directional RNNs.

## Interests & Hobbies

1. Writing blogs/articles about Data science and statistics. (Head to [inblog.ineuron.com](http://inblog.ineuron.com) to check some of my articles)
2. Reading research papers on Data science.
3. Playing basketball and cricket.

## Education

2015-06 - 2019-05

**Bachelor of Engineering:** Computer Science and Engineering

*Jawaharlal Nehru Technological University - Hyderabad*

**Percentage:** 67%

## Languages

English and Hindi