



Nitish Yadav

Phone: +91-8010089908

E-Mail: ny11497@gmail.com

Achievement-driven professional targeting assignments in **Machine Learning and Python Programming** with a growing organization preferably in **IT sector**

PROFILE SUMMARY

Having 5.5 years of experience in Data Analyst using technologies Python, Data Visualization Excel, SQL Web Scraping.

- Hands on experience in Web crawling.
- Hands on experience in Data Cleaning and Data Wrangling.
- Hands on experience in Statistical Modeling/Machine Learning
- Hands on experience in using RStudio.
- Hands on experience in Python.
- Adaptive to learn new technologies.
- Quick Learner and hardworking nature.

KEY SKILLS

~ Data Validation & Cleansing ~ Machine Learning ~ Text Mining ~ Web Crawling ~ Data Wrangling
~ Natural Language Processing ~ Time Series Analysis (ARIMA) ~ R ~ Python ~ Sentimental Analytics
~ Hypothesis Testing (ANOVA, f-test, z-test, t-test) ~ ABBYY and Jiffy RPA

ORGANIZATIONAL EXPERIENCE

Aug 2020 – Presents Sofitrapvt ltd as Python Developer

Projects Details:

Project 1: dataCrawling

Write to python scripts for Scraping the data from websites and .

Environment: Python, Pandas, BeautifulSoup, requests, selenium

Role: Python Developer

Dec 2019 – July 2020 Nityo Infotech Services pvt ltd as AI Developer

Projects Details:

Project 1: Summarise the kpi data

Read excel or csv file and manage columns automatically and summarise the kpi data.

Environment: Python, Pandas.

Role: Python Developer

July'18 – Dec 2019 Applied Solar Technologies Pvt. Ltd., Noida as Data Analyst

Projects Details:

Project 1: Web Crawling

Extracting Details of EB Bills from different sites

Extracting Details From PDF files

Environment: Python, Pandas, BeautifulSoup, Selenium, ABBYY, Jiffy RPA

Role: Python Developer

Description:

Download EB Bills from different State Govt. Sites . Extract Information from pdf, using Data Mining tools And Automate by JIFFY RPA and ABBYY for OCR.

Project 2: Battery Age Prediction

Designed, Created and Deployed a battery age and fault predictive model

Environment: Python, Sikikit Learn, Seaborn, Matplotlib

Role- Data Analyst

Algo used:- Linear Regression, k-means

Description:

Using Predictive modelling technique such as Regression Analysis and hypothesis testing the battery age and battery simulation prediction was done.

Based on Battery age and Battery behaviour I was able to predict when the particular site will go down this was done using K-MEANS and KNN. After the model predicted time the task was assigned to site engineer (early fault prediction).

Project 3 – Data Analysis-

- Deployed Predictive Models for energy management, which included techniques like Trend Mining, Decision Trees.

Apr'15 – June'2018

QL2 Software India Pvt. Ltd., Gurugram as Data Analyst

Projects Details:

Project 1:

Jan 2018 to June 2018

Project: Sentimental Analysis

Client: Expedia

Environment: R, RStudio

Algo used:- Random Forest and Naïve Bayes, Tidy text, n-gram, dyplr, tm, nlp, ggplot2,

Role: Data Analyst, R Developer.

Description: Hotel reviews are available in abundance today. This project focuses on providing a text-based search interface for retrieval and recommendation of hotels based on their reviews. The ranking and displaying hotels to the users for the free-text search query entered is based on the aspects present in the query and the hotel reviews. We used a dataset crawled from Expedia to work on the project. The ultimate aim of the project is to integrate our solution with the main system of ReviewMiner. For the purpose of the course project, we have designed our own user interface, query parsing, and indexing system.

Responsibilities: Working as a developer and Data Analyst in a team that analyzes, design and develop the solutions for the project.

- Applying NLP and machine learning algorithm.
- Using dplyr package perform data cleaning and wrangling.
- Applying tidytext and n-gram for preparing training data
- Applying Random forest and naïve bayes algorithm for prediction.

Project 2:

Project :Web Crawling

Title: Hotel Domain

Environment: Webql Studio, Python

Objective: Provides detailed descriptions regarding the products that are available on leading websites for their competitive purpose. With the help of reports details are more clear and useful.

Organization: Expedia, Booking, Agoda, PriceLine, Tripadvisor, Trivago and many more.

Description: All the relevant information that is available on the websites is extracted. Major clients like Expedia is taking data from more than 25 websites of Hotel domain. There are various leading online hotel booking websites like Agoda, Booking, Tripadvisor, Trivagoetc from where data is extracted. Information extracted from websites according to the input provided. Input may be (keyword/Path). There are two types of applications available in Hotel domain (I) that scraps data from the list pages and (II) that scraps data from detail pages. Here data being extracted and further processed according to Client's requirements and then delivered.

Personal projects

- Health care Data Analysis Cloudera challenge Data Visualization.
- Worked on Credit Card Spend Prediction - Cloudera challenge Banking Regression Problem.
- Credit Card Segmentation Factor Analysis

IT SKILLS

- Programming Languages: R,Python, Java, JSP, XML
- MS Office: Excel
- Database: Oracle, SQL
- Others: Text Mining, NLP, Machine Learning

ACADAMIC DETAILS

2018	Master of Computer Application from Amity University(Distance Learning)
2014	Bachelor of Computer Applications fromApar India Collage(Sikkim Manipal University Distance Learning)
2011	12th from Dr.Shyama Prasad MHSS, KajhaKhurd, Mau, U.P.
2008	10th from Dr.Shyama Prasad MHSS, KajhaKhurd, Mau, U.P.

KNOWLEDGE PURVIEW

Statistical Analysis (R)

- Data Normalization, Hypothesis Testing, T-test, F-test, Correlation, ANOVA
- Predictive Modelling & Forecasting Techniques:
 - Generalized Linear Regression
 - Multiple Regression
 - Logistic Regression
 - Decision Tree, Random forest, Naïve Bayes.
- Time Series Analysis, Segmentation (Clustering), Principal Component Analysis
- Natural Language Processing (NLP) using Packages- OpenNLP, TM

PERSONAL DETAILS

Date of Birth: 7th July 1992
 Languages Known: English and Hindi
 Mailing Address: RZ-253, Gali No. 8, Mahavir Enclave Part-2, Delhi-110059