

Jay Thakur

jay.thakur.job.work@gmail.com | +91-8605601777 | [jaydataScience-projects](#) | [jay-thakur-data-science](#) | <https://jaydatascience-projects.github.io/>

Education

| | |
|--|--|
| NBN Sinhgad School of Engineering Bachelor of Engineering in Computer Engineering • Courses: OOP in C++, Compilers, Algorithms, Operating Systems, Data Structures, Machine Learning, Database Management System | Aug 2014 - May 2020 CGPA: 6.64 / 10 |
|--|--|

| | |
|---|-------------------------------------|
| International Institute of Information Technology Bangalore Post Graduate Diploma in Data Science • Courses: Python for Data Science, SQL, MongoDB, Visualization Tools, Machine Learning, AWS, Data Storytelling | May 2020 - Present GPA: 3.09 / 4 |
|---|-------------------------------------|

Skills

| | |
|---------------|---------------------------------|
| Languages: | Python |
| Technologies: | SQL, AWS, Git, HiveQL, Big Data |
| BI Tools: | Tableau, Power BI, Excel |

Certifications

| | |
|---|------------|
| • Getting Started with SAS Programming Coursera | 08/30/2020 |
| • The Complete Python Course for Machine Learning Udemy | 02/06/2020 |
| • Complete Python Bootcamp: Go from Zero to Hero in Python 3 Udemy | 02/06/2020 |

Projects

| | |
|---|--|
| Credit Approval & Disapproval Analysis Brief: Analyzing the credit card data & predicting whether the income exceeds USD 50K/year via end-to-end modelling process • Performed Data Cleaning and Manipulation for making data understandable and ready for analysis. • Performed Exploratory Data Analysis for treating redundant and missing values. • Finally, analyzed the data using statistical analysis , correlation matrix and other plots for results. | 3rd Aug, 2020 - 10th Aug, 2020 |
| Bike Selling Analysis Brief: Analyze the bike sharing data and to predict using model which factors could help the company to understand the demand of people and increase their sale of renting or selling the bike. • Performed Exploratory Data Analysis and established the hypothesis of available data. • Tested for Multicollinearity , managed outliers and processed missing data. • Performed Feature selection and VIF , p-value to check every features significance in model building • Model building using Multiple Linear Regression . • Performed evaluation using Residual Analysis & finalized the final model with 10 Features in final model . | 7th Sep, 2020 - 14th Sep, 2020 |
| Online Education Lead Score Analysis Brief: Considered the dataset from a X Education (Online Education Platform) to improve the conversion rate from 30% to 80% of converting/making at least 80% people who visited their site to take online course. • Performed Data Cleaning, EDA for making data ready for analysis. | 20th Oct, 2020 - 27th Oct, 2020 |

- Implemented Feature selection and other metrics for analysis important features.
- Model building using **Logistic Regression**.
- Validation and evaluation of model using **ROC Curve** and **Optimal Cutoff Point**.

HELP International NGO Fund Disburse Analysis

12th Oct, 2020 - 19th Oct, 2020

Brief: Analyze the bike sharing data and to predict using model which factors could help the company to understand the demand of people and increase their sale of renting or selling the bike.

- Performed Data Cleaning, EDA for making data ready for analysis.
- Implemented preliminaries like **Hopkins's measure**, **Silhouette curve**, **Elbow curve**, **Single linkage & Complete linkage** methods for **generating optimized 'k' value**.
- Model building using **K-Means** and **Hierarchical Clustering**.
- Model evaluation using evaluation metrics.

IMDb Movie Analysis

23rd Nov, 2020 - 30th Nov, 2020

Brief: Analyzing the movie data to analyze which movie had made most profit, best top Trio cast and most loved Genre among different age groups and genders.

- Analyzed movies using **Window function**, **Case Statement** and **Query Optimization** from **MySQL**.

Click Stream Data Analysis

11th Jan, 2021 - 18th Jan, 2020

Brief: Analyze a click stream data collected by an Online cosmetic selling company to understand most selling product and customer behavior.

- Implemented **HiveQL** in **AWS Cluster** and loading data to **HDFS**.
- Implemented **Joins**, **Table creation (External Table)**, **Bucketing**, **Partitioning**, etc. for result generation.

IPL Stats Visualization

11th Feb, 2021 - 18th Feb, 2021

Brief: Building a Tableau dashboard of IPL statistics over the years since its inception in order to create an infographic for a newsletter that their team is working on.

Achievement

-
- **Pratyaksh - Data Visualization Competition | Parigyan - The Data Science Society of Goa Institute of Management | 1st Prize (Team - Data Wizards)**

25th Feb 2021 - 8th March 2021