Jay **Thakur**

jay.thakur.job.work@gmail.com | +91-8605601777 | ⬤ jaydataScience-projects | 🔗 jay-thakur-data-science | 🖼 Portfolio

Education

NBN Sinhgad School of Engineering	Aug 2014 - May 2020
Bachelor of Engineering in Computer Engineering	CGPA: 6.64 / 10

• **Courses:** OOP in C++, Compilers, Algorithms, Operating Systems, Data Structures, Machine Learning, Database Management System

International Institute of Information Technology Bangalore	May 2020 – Present
Post Graduate Diploma in Data Science	GPA: 3 / 4

• Courses: Python for Data Science, SQL, MongoDB, Visualization Tools, Machine Learning, AWS

Skills

Languages: C++, Python, HTML
Technologies: MySQL, SQL, MongoDB, AWS, Git, HiveQL, CSS, Big Data, Cloud

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   3rd Aug, 2020 – 10th Aug, 2020
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.

Bike Selling Analysis   7th Sep, 2020 – 14th Sep, 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 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**.

Online Education Lead Score Analysis   20th Oct, 2020 – 27th Oct, 2020
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.
- 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 prelimaries like **Hopkins 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 – 22nd 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.

Titanic Data Analysis   25th Feb, 2021 – 2th March, 2021

Brief: To understand the behavior, importance and correlation of all the variables against **target variable Survived**. EDA (Exploratory Data Analysis) steps will be performed to attempt a solution for given problem statement.
- 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.**