

Neeraj Kumar

DATA ANALYST

Data analyst with 2 years of experience adept at collecting, analyzing and interpreting large datasets. Visualizing the data, developing new forecasting models and solve challenging business problems. possessing an extensive-analytical skills and significant to work in a team environment. Self-motivation and passion are two key factors that catalyze my process of learning by inclining myself towards tough and challenging paths.

nk6022427@gmail.com



+91 9997467070



Noida, Uttar Pradesh, India



[linkedin.com/in/neeraj-kumar-](https://www.linkedin.com/in/neeraj-kumar-)



github.com/Neeraj-DS



WORK EXPERIENCE

Data analyst

I3INFOSOFT PVT LTD | NOIDA

09/2020 - Present

- Using automated tools to extract data from primary and secondary sources.
- Performed high-quality data collection along with data preprocessing operation like **data wrangling, data cleaning, data profiling** on large volume data.
- Successfully interpreted data in order to draw conclusion using **Exploratory Data Analysis (EDA)**.
- Exploring and visualization of data to drive insights.
- Perform **Data Cleaning, features Scaling, features Engineering** using **python**.
- To build and trained **Machine Learning model** using **supervised machine learning model**.
- Hyper-parameter tuning to improve the model Validation metric for Regression and classification algorithms.
- Preparing final **analysis reports** for the stakeholders to understand the data-analysis steps, enabling them to take important decisions based on various facts and trends.

PROJECTS

Airline Customer Service Prediction

(11/2021 - 07/2022)

Problem Statement : An airline company collects information on all passengers who have flown in their airlines and are interested to do an analysis of customer satisfaction. They've recorded the satisfaction of the customer over several of their in-flight services.

As a solution specialist the client has approached to analyze the data and arrive at solution using machine learning to predict the customer satisfaction rate.

Tasks:-

- Perform data cleaning and data preparation using **python** libraries on more than one Million Records.
- Handling outliers and Null values.
- Collecting the data and implementing **Exploratory Data Analysis** with different types of graphs.
- Select appropriate features which are important to create best models and apply suitable algorithms to get better accuracy.
- Technologies used: Pandas, NumPy, MySQL, Matplotlib, Seaborn, Machine Learning models and techniques.

SKILLS

Python

NumPy

Pandas

MySQL

Ms-Excel

Statistical Analysis

Data Visualization

Tableau

Matplotlib

Seaborn

Power BI

Machine Learning (supervised models)

Scikit-Learn

EDUCATION

Bachelor of Technology

(Mechanical Engineering)

G L Bajaj Group of Institution
Mathura

(08/2016 - 07/2020)

LANGUAGES

English *Full Professional Proficiency*

Hindi *Native or Bilingual Proficiency*

HOBBIES & INTERESTS

Artificial Intelligence

Reading books and news paper

travelling

Healthcare industry No-show Prediction (03/2021 - 09/2021)

Problem statement : In the healthcare industry, there is a constant challenge of overbookings, causing delay in services offered to the patients. 'No Shows' is one of the factors leading to the overbookings.

A patient may book for a doctor's appointment at a health center, but does not show up on the appointment date/time. This is called as 'No Show'.

Client has approached to analyze the data and arrive at solution using machine learning to prediction of patients whether They attend their appointments or not which were scheduled by these patients.

Tasks:-

- Performed data cleaning and data pre-processing by merging three different datasets of more than one million records.
- Collecting the data and implementing **Exploratory Data Analysis** with different types of graphs.
- To build and trained **Machine Learning model** using Logistics Regression, Random Forest Classifier, Decision Tree Classifier, Cross Validation.
- To handle imbalanced classification data modeling with oversampling.
- To find the accuracy of the model prediction using Confusion Matrix, AUC Score.
- Technologies used: Python, ML techniques and models, MySQL, NumPy, Pandas, Scikit Learn, Matplotlib and Seaborn.

Multiple Analysis on SCM (01/2021 - 02/2021)

The client was facing challenge in dynamically change in the market and

Want to analyze sales insights and want to know proper SCM.

Tasks:-

- Extract the data using MySQL with various condition.
- Perform **ETL** process, handling null and filtering of data using Tableau.
- Create various charts for visualization of data using **Tableau**.
- Used Tableau to build interactive BI dashboard that can show revenue declining.
- Prepared data story that clearly displays the insight.