



## Shubham Kumar Rai

[rai.shubhammech52@gmail.com](mailto:rai.shubhammech52@gmail.com)

+91 9425575287

Bangalore, KA

[http://www.linkedin.com/in/shubham-kumar-rai-720b6315b\\_](http://www.linkedin.com/in/shubham-kumar-rai-720b6315b_)

<https://github.com/raishubham23>



### Technical Skills

- ❖ Power Bi \*\*\*\*
- ❖ My SQL\*\*\*
- ❖ Advanced Excel \*\*\*\*
- ❖ Tableau\*\*\*
- ❖ Python\*\*\*
- ❖ Data Preprocessing\*\*\*
- ❖ NLP\*\*\*
- ❖ KNN\*\*\*
- ❖ SVM\*\*\*
- ❖ Data Modeling\*\*\*

### Objective:

To secure a challenging position as a **Data Scientist & Data Analyst** which utilizes my knowledge and skill set, while allowing me the opportunity to grow professionally.

### Experience Summary:

Internship Data Analyst **AI Variant** (from **Oct 2022 to July 2023**)

#### Internship Project

**Project (P136):** - Employee Retention

- HR ANALYTICS can provide insights into employee engagement, performance, retention and more.
- By analysing data, we can identify trends, make predictions and even proactively address issues.
- By leveraging the data, HR professionals can make informed decisions that drive organizational success.
- This presentation will explore the benefits of HR ANALYTICS and provide insights on how to create a data-driven culture.

([https://github.com/raishubham23/DA\\_Project\\_1](https://github.com/raishubham23/DA_Project_1))

**Project(P168):** - FMCG Analytics

- Our FMCG Sales Dashboard provides a comprehensive view of sales performance, enabling data-driven decision-making.
- By analysing store-wise sales, we can identify top-performing stores and areas for improvement. Month-wise sales growth highlights peak and slow periods, allowing us to optimize resources and plan targeted strategies.
- Area Manager-wise sales analysis helps identify top performers and support underperformers for overall sales improvement.
- Product-wise sales analysis enables us to leverage successful categories, address challenges, and optimize our product offerings.

([https://github.com/raishubham23/DA\\_Project\\_2](https://github.com/raishubham23/DA_Project_2))

**Project(P200):** - Health Care Analytics

- Preventive Care Importance: One of the key findings in healthcare is the significance of preventive care.
- Rise of Chronic Diseases: Chronic diseases such as diabetes, heart disease, and certain cancers have been on the rise.
- Digital Health Innovation: The integration of technology into healthcare (telemedicine, wearable devices, health apps) has transformed patient care and remote monitoring, making healthcare more accessible and efficient.

([https://github.com/raishubham23/DA\\_Project\\_3](https://github.com/raishubham23/DA_Project_3))

## Certification

Certification Course is completed in Data Analyst & Data Science From ExcelR Solutions (Collaboration with IBM) BTM 1st Stage Bangalore (Karnataka).

## Education

B.E (First Division) C.S.V.T.U **2013 – 2017**  
**Mechanical Engineering Percentage 70.20**

## NON-IT

- ❖ Worked as Site Engineer in “B.P. Gupta & Comp.” From May 2022 to Sep 2022.
- ❖ Worked as Site Engineer in “Sahicon Engineering” From April 2021 to April 2022.
- ❖ Worked as Site Engineer in “M.C.K Kutty Engineers Pvt. Ltd” From July 2020 to March 2021.
- ❖ Worked as Site Engineer in “Hira Power & Steel” From June 2019 to June 2020.
- ❖ Worked as Site Engineer in “Shankar Prasad Sharma” From April 2018 to May 2019.

## Hobbies

- ❖ Playing Chess
- ❖ Solving Puzzle

## Languages

- ❖ English \*\*\*
- ❖ Hindi \*\*\*\*

Internship Data Science **AI Variant** (from **July 2023** to till date)

## Internship Project

**Project(P272):** - Real and Fake News Analysis

Challenges Faced While Building a Model

### Dataset

Both the datasets were having some discrepancy due to which loading datasets and analyzing the issue took time. From this we can say that the quality of the data impact the model's performance.

### Feature Extraction

Extracting meaningful features from text data requires careful preprocessing and consideration of techniques like tokenization, lemmatization and vectorization.

### Model and Hyperparameter Tuning

Choosing the right algorithm considering accuracy as well as precision. Hyperparameter tuning, to achieve optimal performance was quite time and memory consuming.

(<https://github.com/raishubham23/Project1>)

**Project(P284):** - Internet Traffic Data

**Non-Stationary Data:** The initial challenge was dealing with non-stationary like ARIMA and SARIMA, which assume stationarity. time series data. Non-stationary data can make it difficult to build accurate time series models

**Forecasting Horizon:** The original model was designed to forecast for 7 days, but there was a need to extend the forecasting horizon to 30 days. This required a modification in the model's input data and possibly its configuration.

### How We Overcame These Challenges:

**Stationarity Check:** To address the issue of non-stationary data, stationarity tests like the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) and ADF (Augmented Dickey-Fuller) tests were performed.

**Differencing:** To make the data stationary, differencing was used. Differencing involves subtracting each value from its previous value. It helps remove trends and seasonality, making the data more amenable to modelling.

**Forecasting Horizon Extension:** When the model encountered an error while trying to forecast for 30 days, the input data was changed from 'daily\_data\_last\_7' to 'daily\_data\_last\_30'. This change allowed the model to forecast for a longer time horizon.

([https://github.com/raishubham23/DS\\_Project\\_2](https://github.com/raishubham23/DS_Project_2))

**Project(P298):** - Bankruptcy Prevention

**The variables were Categorical Variables:** Applied label encoding to convert categorical data into numerical format.

**Evaluation of Metrics, choosing appropriate evaluation metrics:** Utilized methods like precision, recall, F1-score and ROC curve for comprehensive model evaluation.

**Platform Selection for Deployment:** Deployed the model on Streamlit Sharing for easy and free hosting.

([https://github.com/raishubham23/DS\\_Project\\_3](https://github.com/raishubham23/DS_Project_3))