**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**



**A REPORT ON**

CELEBAL TECHNOLOGIES INTERNSHIP

Submitted in Partial Fulfillment of the requirements for the Degree in B.Tech in Computer Science & Engineering

BY

Adarsh Jain

Roll: 2005355

**ACKNOWLEDGEMENT**

I am immensely grateful to the KIIT Training and Placement for granting me the internship opportunity. Their support, guidance, and inclusive environment have fostered my professional growth. I appreciate the trust placed in me, and the experiences gained will undoubtedly shape my future career. Thank you for this invaluable opportunity.

**Student Details**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Roll No** | **Branch** | **Semester** | **Signature** |
| Adarsh Jain | 2005355 | CSE | 7th | adarsh |

**Internship Details**

|  |  |  |  |
| --- | --- | --- | --- |
| **Organization Name** | **Position** | **Duration** | **Mentor** |
| Celebal Technologies | Data Science Intern | 20 May 2023- Present | Amey Patil |

1. **Programme Scope and Internship Details**

Currently undergoing my summer internship at celebal technologies as a Data Science Intern which started at 20th may.

There we are assigned groups and respective mentors who guide us in completing the projects and other HR activities which are assigned to us on a weekly basis.

1. **Project Portfolio**

Project 1: Air Quality Predicton for Urban Areas using Machine Learning

The project involves various stages, including web scraping, exploratory data analysis (EDA), training a machine learning model, and saving the model for future use. Dataset was Scraped from aqi.in/india which included features like PM2.5, PM10, CO, NO2 and meteorological conditions like temperature, humidity, wind speed, and timestamp. Exploratory Data Analysis was performed which included libraries like Matplotlib, Seaborn etc.

We normalized the data and implemented Sequential Tensorflow Model and evaluation of the trained model's performance was done using suitable evaluation metrics (e.g., mean squared error, R-squared). Finally we deployed out model using Streamlit providing a clean user interface where user can input the parameters and model predicts the AQI and its corrosponding AQI bucket and then we saved the model for future use.

Project 2: Toxic Comment Classification using NLP (ongoing)

The main goal of this assignment is to develop a predictive model that can accurately classify toxic comments into one or more toxicity classes in the dataset, using some innovative methods.

1. **Software Tools and Libraries Used :** Python, Pycharm, Jupyter Notebook, VSCode Machine Learning libraries like Tensorflow, Seaborn, Matplotlib, Numpy, Pandas, CountVectoriser, NLTK, Web scraping tools like Selenium and Beautiful Soup.