

## WEEKLY REPORT

### **Work done in last week ( Attach supporting Documents):**

#### *Data Cleaning*

##### **1. Developed an in-depth understanding of Data Cleaning techniques and best practices.**

- Explored handling of missing values, duplicate records, inconsistent formats, and outliers.
- Practiced ensuring data integrity, standardization, and reliability across datasets.

##### **2. Practical experience with real-world datasets to implement data cleaning concepts.**

- Dataset 1: Nye York city Airbnb Dataset
- Dataset 2: YouTube Dataset

##### **Key Concepts Applied:**

- Missing Data Handling: Imputation and logical exclusion
- Duplicate Removal: Ensured data uniqueness
- Standardization: Formatted variables uniformly
- Outlier Detection: Identified & treated skewed entries

#### *Sentiment Analysis*

##### **Description:**

Designed a sentiment analysis system to classify text data sentiment, aiding in understanding customer feedback and social media trends.

##### **Datasets:**

- Dataset 1 Link
- Dataset 2 Link
- 

##### **Key Concepts and Challenges:**

- Sentiment Analysis: Text classification into positive/neutral/negative sentiment.
- NLP Techniques: Tokenization, stop-word removal, vectorization
- ML Algorithms: Used models such as Naive Bayes and SVM
- Feature Engineering: Extracted relevant features using TF-IDF

- Data Visualization: Created plots to depict sentiment distribution

**Reason for incomplete work:** N/A – All planned tasks completed for Week 2.

**Plans for next week:**

**1. Project: Predicting House Prices with Linear Regression**

- Goal: Build a model using linear regression to estimate numerical outcomes based on selected features.
- Focus: Data cleaning, feature selection, model training using Scikit-Learn, and evaluation using MSE/R<sup>2</sup>.

**2. Project: Wine Quality Prediction**

- Goal: Predict wine quality based on chemical attributes.
- Approach: Use Random Forest, SGD, and SVC classifiers, evaluate using accuracy and F1-score.

**3. Project: Fraud Detection**

- Goal: Detect fraudulent transactions using machine learning models.
- Techniques: Anomaly detection, feature engineering, and real-time fraud monitoring using Logistic Regression, Decision Trees, or Neural Networks.

**References:**

- Oasis Infobyte: <https://oasisinfobyte.com/>
- Sentiment Analysis Guide: <https://www.geeksforgeeks.org/sentiment-analysis/>
- Data Cleaning Techniques: <https://www.kaggle.com/learn/data-cleaning>

**Signature of External Guide**



---

Program Coordinator

**Signature of Internal Guide**

**Student Id: 23AIML014**

**Student Name: DHANDHUKIYA DEVANG**