Date	Topic	Description	Time Spent	Challenges faced	Solution
11/12/2024	Group meeting	Discussed about the topic, Problem statement	2 hour	NIL	NIL
17/12/2024	Group meeting	Discussed about the Dataset for the problem statement and deciding on the model.	2 hour	NIL	NIL
21/12/2024	Finding Dataset (Weather - India)	Was able to get the India Weather Dataset from visualcrossing.com. Took 3 days to download the data due to data limit for free account.	1 hour	As it is a group project, each member should have a proper and similar dataset	After searching multiple websites.
22/12/2024	Finding Dataset (Crop Dataset - India)	Downloaded datasets for crop yield,production and harvested area for the Year 2020- 2023. (**Our World in Data- Agricultural Production**)	2 hour	Search for a dataset with required columns and rows	After searching multiple websites.
23/12/2024	Weather Dataset	Downloaded datasets for Indian weather and crop data for the year 2018-2019	1 hour	Because to increase the data points	Downloaded the file
24/12/2024	Loading the Dataset - Jupyter Notebook	Loaded the weather and crop datasets into Python using Jupyter Notebook. Additionally, Merged the multiple weather datasets into a single dataset.	2 hour	Unable to merge due to difference in the column datatypes, additional columns.	Apply code and completed the merge.
25/12/2024	Loading Database - MongoDB	The merged weather dataset, along with the crop dataset, has been loaded into the database separately using MongoDB, creating two separate collections within the same database connection.	2 hour	NIL	NIL
26/12/2024	Retrieve Data from Database	The stored data has been retrieved using a MongoDB query and converted into a DataFrame for preprocessing.	1 hour	NIL	NIL
27/12/2024	Data Characteristics and Preprocessing	Data characteristics and preprocessing are conducted on the datasets separately (Weather and Crop) to check for missing values, duplicates, and outliers.	4 hour	NIL	NIL
28/12/2024	Data Manipulation	Filtered the 2024 weather data since Crop dataset only includes data up to 2023 for consistency between the datasets, which is crucial for analysis and modelling.	1 hour	The crop dataset contains data from 2018-2023. While the weather weather dataset contains data from 2018-2024	Removed 2024 data from weather dataset.
29/12/2024	Data Manipulation and Visualization	Removed outliers and filled null values from both datasets. I also utilized various visualizations (**Histogram, Density plot, Box plot,line plot, Pie chart**) to understand the distribution of features and the target variable. Additionally, I applied log transformation to the target variable, which was right-skewed.	2 hour	NIL	NIL
30/12/2024	Feature Engineering	Merged the datasets and conducted correlation analysis after aggregating the weather features to identify the best-correlated features for the target variable. After this, used MongoDB to store the merged data. Once stored, I retrieved only the necessary features based on the correlation and statistical analysis to prepare for machine learning.	3 hour	Identifying the highly correlated features with target variable was a little challenging as the dataset contains limited features that correlates well with the target variable	To overcome this challenge, had to aggregate the features before and after the correlation and check for best possible features.
31/12/2024	Machine Learning Algorithm	Using the selected features and target variable, employed the **Random Forest Regressor, Decision Tree and, Gradient Boosting Regressor** to assess the model's performance on the chosen dataset.	3 hour	Tried to implement Linear, logistic and SVM, but the performance of the model was very weak compared to the other tree based models.	Removed it from the code.

Date	Topic	Description	Time Spent	Challenges faced	Solution
1/12/2025	Evaluation Process	Calculated performance metrics for **Random Forest, Decision Tree and, Gradient Boosting Regressor** models, comparing their results to identify the best-performing model. Additionally, I visualized the comparison of metrics to provide a clear understanding of each model's effectiveness and accuracy.	4 hour	I attempted to visualize the yield, production, and harvested area on an India map. However, this was not feasible as the dataset lacks state-wise details, limiting the granularity required for geographical representation.	Used a one-way ANOVA test to analyze the statistical significance of differences among "Area harvested," "Yield," and "Production" data.
2/1/2025	Evaluation Process	Worked on creating visual representations to analyze yearly trends in **area harvested**, **production**, and **yield** for key crops (Barley, Coconut oil, Coffee, Cotton seed, Groundnut oil, Maize, Millet, Rice, Soyabean, Tea leaves, and Wheat) from 2018 to 2023. The graphs provided a comparative overview of these metrics, helping to identify patterns and variations across crops and years. This analysis aims to support decision-making for resource allocation and agricultural strategy optimization.	3 hour	NIL	NIL
3/1/2025	Report - Related works	Started work on report writing for the project - Added content in related works about the Decision Tree algorithm. Also, added the conclusion and future work section.	3 hour	Future work was a concern.	Dashboards will be tailored to specific regional farming needs, integrating localized data such as crop patterns, weather predictions, and soil conditions
4/1/2025	Report - Conclusion and Future Work	Added Evaluation metrics for my dataset in the report.	30 minutes	NIL	NIL
	System Requirements file	Created a system requirement file using markdown in Jupyter Notebook.	1 hour	NIL	NIL
	Output storage - MongoDB	The data for selected grains from 2018 to 2023 has been stored in the MongoDB database, including temperature, humidity, and key metrics like yield, production, and area harvested, for easy comparison and analysis.	30 minutes	NIL	NIL
5/1/2025	Creating a video	Created a video about the project. Started with the introduction of the project.	8 min35 sec	NIL	NIL
	GitHub - Uploading files	Uploaded the code and Data files in Wanpin's Github collaborated link.	5 minutes	NIL	NIL
	Report - Formatting	Final formatting of the Report	1 hour	NIL	NIL