**Unveiling the Future of Carbon Emissions Trading: A Machine Learning and Neural Network Perspective on Regional Markets.**

**GROUP - 07**

Sindhu Nagesha (017419987)

Prayag Nikul Purani (017416737)

Syed Faraaz Ahmed (017428619)

Sai Vivek Chunduri (017435301)

Department of Applied Data Science

San Jose State University

DATA 270: Data Analytics Processes

Submitted to: Dr. Eduardo Chan

April 3, 2024

### **2. Data & Project Management Plan**

### **2.1 Data Management Plan**

A Data Management Plan (DMP) is crucial for any project that involves significant data collection, organization, storage, and utilization. It outlines strategies for effectively collecting, managing, safeguarding, and maintaining data, guaranteeing its availability throughout the entire research process. Essential elements of a Data Management Plan include compliance with data formats, adherence to metadata standards, utilization of reliable storage solutions, and implementation of stringent security protocols. This methodical approach not only ensures the integrity of data but also adheres to ethical standards and meets the requirements of funding agencies.

#### ***2.1.1 Data Collection Approaches***

The data collection approach for this project encompasses three crucial datasets to ensure a thorough analysis of carbon emissions and the factors that influence them. The initial dataset, sourced from the World Bank, covers worldwide carbon emissions data spanning from 1850 to 2022, providing a comprehensive global outlook. The second dataset, sourced from the U.S. Energy Information Administration (EIA), provides comprehensive state-level carbon emission data from 1970 to 2021. Finally, the National Weather Service (NWS) API is used to obtain real-time weather data, which is essential for immediate analysis of meteorological conditions. The incorporation of historical, international, and real-time datasets enables a comprehensive approach to modeling and forecasting carbon emissions. This approach utilizes both past patterns and present environmental conditions to guide the development of efficient carbon management strategies. The extensive data collection is specifically designed to bolster the project's capacity to comprehensively analyze and predict trends in carbon emissions trading, utilizing advanced machine learning and neural network techniques.

***World Data***

The global carbon emissions dataset, obtained from the World Bank, provides detailed parameters such as CO2 emissions from various sources such as coal, cement, gas, and oil, as well as country-specific data including year, ISO code, population, and GDP. These metrics are cataloged over multiple years, offering a comprehensive view of global emission trends. With its renowned reputation for meticulous data collection and extensive coverage, the World Bank's dataset forms a solid foundation for analyzing environmental and economic patterns on a global scale.

During the data collection process, a strong emphasis is placed on precision and scope. Advanced statistical techniques and rigorous data validation protocols are employed to ensure the accuracy of the dataset. For instance, the World Bank utilizes strict data verification methods, such as cross-referencing with other authoritative sources and thorough reviews of data anomalies, to maintain the highest standards of data integrity.

Furthermore, the robust architecture of the dataset supports extensive environmental and economic analyses, making it an invaluable resource for developing informed global strategies. The technical robustness of the dataset, combined with the World Bank's commitment to data transparency and accessibility, ensures that it serves as a reliable resource for policymakers and researchers aiming to understand and mitigate the impacts of carbon emissions worldwide.

***United States Data***

The carbon emissions dataset for the United States, covering the time period from 1970 to 2021, is meticulously arranged in four separate Excel files acquired from the U.S. Energy Information Administration (EIA). The files classify emissions according to different parameters, such as per capita, per unit of energy consumption, and per unit of economic output. This ongoing, supplementary collection of time-series data serves as a strong basis for thorough trend analysis and policy evaluation concerning carbon emissions.

The EIA is an autonomous agency operating under the Department of Energy. Its primary objective is to gather and distribute energy-related data to facilitate well-informed decision-making and enhance public awareness of the effects of energy. The openly accessible data on the EIA website, which does not require any login or special access, is essential for conducting thorough studies on energy consumption and its environmental consequences.

The dataset contains files that are designed for easy analysis, providing comprehensive carbon emissions data for each state over multiple decades. This format not only simplifies direct comparisons between different states and time periods but also improves the usefulness of the data for longitudinal studies that investigate the evolution of emissions over a span of more than five decades. The utilization of a systematic method for gathering data, along with the EIA's dedication to offering impartial and autonomous energy information, highlights the dataset's importance in environmental research and policy-making efforts.

***Weather data***

The real-time data component of the project obtained weather data from the National Weather Service (NWS) API. This API provided important updates such as forecasts, alerts, and observations, which were essential for environmental analysis. This API provides a comprehensive and reliable flow of up-to-date meteorological information, such as temperature, dew point, relative humidity, wind speed, and wind direction. These data are crucial for informing predictive models and trading strategies that specifically target carbon emissions. By integrating this data, the project is able to capture the ever-changing environmental factors that have an impact on market dynamics and carbon emissions. The utilization of the NWS API facilitates the project's objective to exploit machine learning and neural network methodologies for a thorough examination of carbon emissions trading, taking into account tangible factors like weather conditions. Furthermore, the NWS API's daily limit of 5000 requests is suitable for the project's requirements, providing a dependable and economical data solution for ongoing model optimization and validation. The Table below describes the datasets used in this project.

**Table 4**

*Dataset Description*

| Data | Description |
| --- | --- |
| Dataset 1 | Carbon Emissions for the World from the World Bank Website |
| Dataset 2 | Carbon emissions data for each state in the United States from 1970 to 2021. |
| Dataset 3 | Weather data from National Weather Service API. |

Note. Dataset description for carbon emission prediction project.

#### ***2.1.2 Data Management Methods***

Efficient data management is essential for maintaining the accuracy and usability of data in projects that involve extensive data analysis, such as those focused on studying carbon emissions trends. In the beginning, data preprocessing and cleaning are essential steps to improve the quality of the data. This entails the process of detecting and resolving problems such as absent values, repetitions, and discrepancies within the datasets. Methods such as data reduction and normalization are used to guarantee consistency and compatibility among diverse datasets and time periods.

After the data has been cleaned, integration and transformation processes are carried out. Integration is the process of combining datasets that have shared identifiers, such as geographical locations or timeframes, in order to create a complete dataset. Transformation involves the calculation of derived variables and the consolidation of data to enable more efficient analysis. These steps are essential for preparing the data for comprehensive analytical tasks.

Rigorous quality assurance practices are implemented to ensure data quality is maintained throughout its lifecycle. This encompasses the process of conducting validation checks, cross-referencing with external data sources, and meticulously documenting any modifications or assumptions made to the data. Utilizing version control systems is crucial for monitoring modifications in datasets, improving the transparency and dependability of the analysis by guaranteeing that updates and alterations are thoroughly documented and replicable.

Ensuring security measures are of utmost importance, especially when dealing with sensitive information. Strategies encompass implementing audit trails to monitor data access, limiting data access exclusively to authorized personnel, and employing data encryption. Data in the project environment is securely stored on AWS S3 buckets, and can only be accessed by team members who have the required permissions. Access to the data is regulated by IAM (Identity and Access Management) permissions in the AWS ecosystem, guaranteeing that only authorized individuals can access or alter it. Implementing multifactor authentication is necessary to further enhance security.

Moreover, Google Colab is employed for data analysis, where it directly interacts with AWS S3 buckets using the boto3 library. This configuration guarantees both the security and efficiency of data handling by utilizing the advanced capabilities of Google Colab for intricate computations and direct data manipulations from cloud storage.

By following these data management techniques, the project effectively utilizes data to analyze and comprehend patterns in carbon emissions. This enables the creation of predictive models and trading strategies that aid in making informed decisions and promoting sustainable practices in carbon markets. This holistic approach to data management guarantees that the project's data ecosystem is resilient, protected, and favorable for accomplishing its goals.

#### ***2.1.3 Data Storage Methods***

Data storage is an essential element of the data management plan, focusing on the secure, efficient, and easily accessible storage of information throughout the entire project duration. Amazon Web Services (AWS) has been selected as the exclusive platform for data storage in this project. This choice was made after careful evaluation of factors such as data size, sensitivity, and collaboration requirements, and it was preferred over alternatives like Google Drive and Microsoft Data Lake Gen 2. AWS provides a dependable storage infrastructure with a high level of availability, which is crucial for uninterrupted data retrieval.

This organization facilitates the implementation of uniform security measures throughout all stages of data processing. Access to raw, pre-processed, testing, and training data is tightly regulated using IAM (Identity and Access Management) policies within AWS. This ensures that only authorized team members can access or make changes to the data at any point. This comprehensive approach to data security ensures the preservation and secrecy of the project's data at every level.

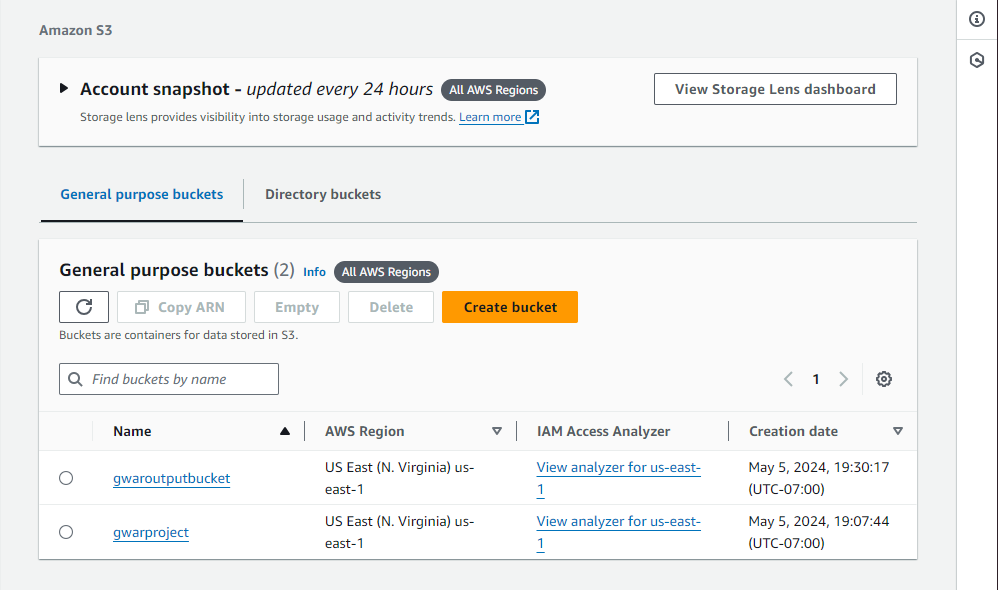
In addition, AWS offers reliable data archival services, such as Glacier, which are designed for storing data over a long period of time. These services ensure that historical data is securely preserved in a cost-effective manner, while still being easily accessible for future analysis. Regular backups and disaster recovery protocols are essential components of the storage strategy, offering robustness against data loss and guaranteeing uninterrupted business operations.

By utilizing AWS for all aspects of data storage, the project gains advantages from a unified and protected platform that facilitates efficient data management, strong security measures, compliance with privacy regulations, and effective disaster recovery strategies. This methodical approach to data storage not only safeguards the accuracy and consistency of data but also improves the project's capacity to analyze patterns in carbon emissions and effectively assist in the development of carbon management strategies. Figure x below

Setting up the AWS infrastructure for the project begins with creating S3 buckets via the AWS Management Console. Two main buckets are created: one for data storage and another for storing Athena query outputs. The data storage bucket acts as a central repository for datasets and model artifacts, offering scalability, durability, and accessibility. Storing data in S3 ensures high availability and reliability for processing and analysis. The second bucket is dedicated to storing Athena query results, facilitating better organization and management of insights. Separating input and output data maintains data integrity and clarity in the workflow.

**Figure 2**

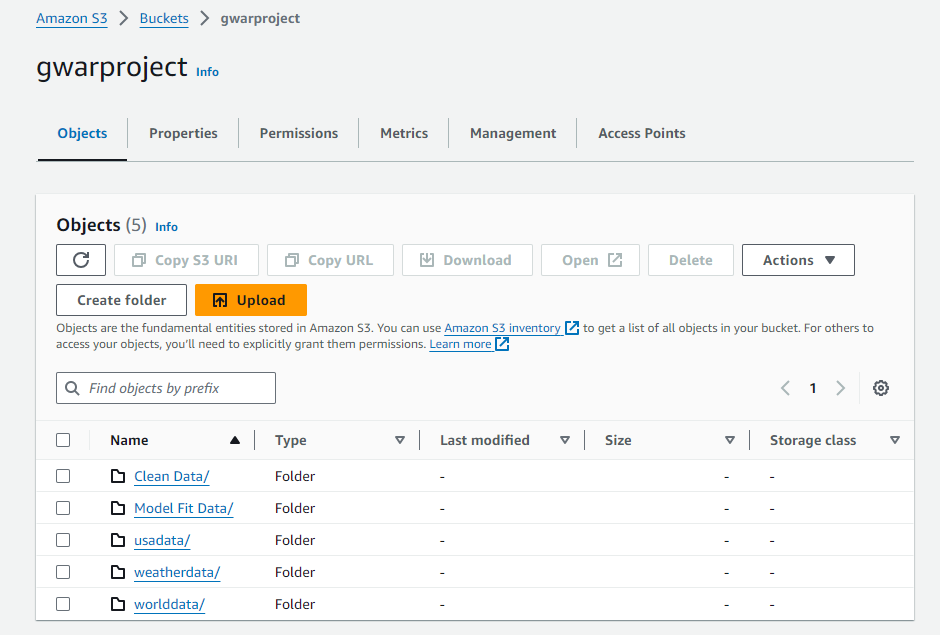
*S3 buckets formation*

**

***Note****.* Use of two buckets for the project.

**Figure 3**

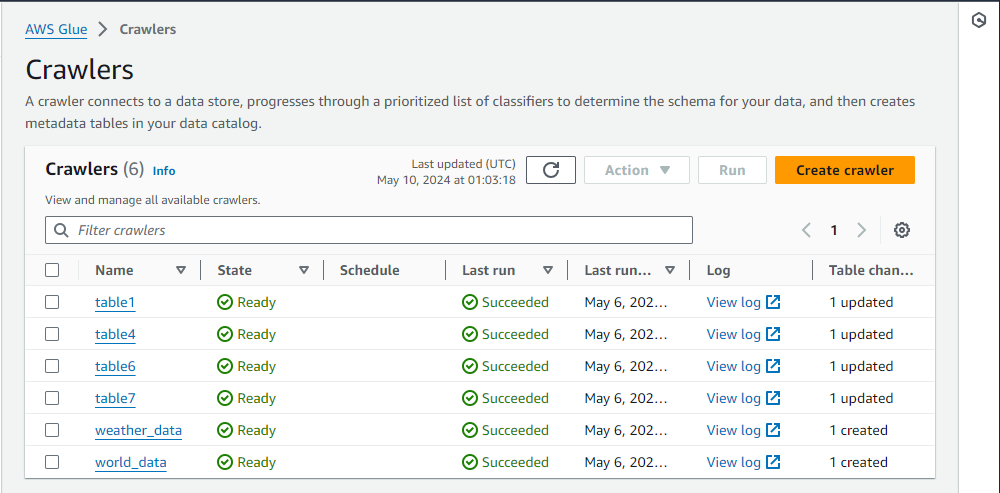
*The folder structure of the “gwarprojcet” bucket*

**

***Note****.* Folder overview.

**Figure 4**

*AWS Glue screen environment*

**

***Note****.* Crawlers for different usage cases.

#### ***2.1.4 Data Usage Mechanisms***

The project's data usage mechanism involves careful scrutiny of details and implementation of robust security protocols. The data sources utilized for this project consist of the World Bank, the U.S. Energy Information Administration, and real-time meteorological data from the National Weather Service. These sources offer both historical emissions data and up-to-date weather conditions. Each dataset is accompanied by thorough documentation that provides detailed information about the sources, methodologies, and collection specifics, as well as comprehensive metadata. The data is stored securely on AWS S3, utilizing robust backup protocols to guarantee data integrity and prevent any potential loss. Access to sensitive information is tightly controlled through AWS IAM policies, guaranteeing that only authorized personnel can gain entry.

Data sharing is conducted in a transparent manner, with the intention of depositing datasets in easily accessible repositories to facilitate future research efforts within the team members. Ensuring the protection of this data is a top priority, with a strong focus on ethical and privacy considerations and compliance with relevant legislation. CSV and JSON are commonly used data formats due to their accessibility and user-friendly nature, while metadata standards aid in the process of discovering and identifying data.

To ensure long-term preservation, strategies have been implemented to maintain data accessibility for a minimum of ten years after the project's completion. Explicit policies establish the parameters for data retention and archival procedures, clearly stating the duration for which data is kept and the circumstances under which it can be archived or deleted. A specifically assigned data manager supervises the compliance with these procedures, guaranteeing that the project's data handling processes are secure, in accordance with regulations, and supportive of achieving the project’s objectives and enhancing its scientific contributions.

**Table 5**

*Hardware Requirements for the project*

| **Data management Phase** | **Performed by** | **Justification** |
| --- | --- | --- |
| Data Collection | Prayag Nikul Purani  Sindhu Nagesha  Sai Vivek Chunduri | Identified carbon emissions data from the World Bank, U.S. EIA, NWS API  and accessed the data. |
| Data Storage | Syed Faraaz Ahmed | Identified cost-effective cloud platforms such as AWS and Google Colab for data storage. |
| Data Security, Privacy and Access | Sindhu Nagesha | Access to the data is controlled through AWS IAM policies, which ensure that only authorized personnel can access the data, maintaining confidentiality and integrity. |
| Data Quality | Sai Vivek Chunduri  Prayag Nikul Purani | Rigorous checks and validations are performed on the data to ensure high quality, |
| Data Format | Prayag Nikul Purani | Data is stored in CSV and JSON formats |
| Data Analysis | Prayag Nikul Purani  Sindhu Nagesha  Sai Vivek Chunduri  Syed Faraaz Ahmed | Perform All cleaning and Exploratory data Analysis |
| Documentation | Syed Faraaz Ahmed  Sindhu Nagesha | Each dataset is accompanied by thorough documentation that describes the source, methodology, and type of data collected, ensuring clarity and usability for future researchers. |
| Ethical and Legal Adherence | Syed Faraaz Ahmed | All data used complies with ethical standards and legal requirements, with particular attention to privacy laws. |
| Data Retention | Sai Vivek Chunduri | Data will be retained for at least ten years post-project completion to allow for extended analysis and review. |

Note. Hardware Requirements for the project.

### **2.2 Project Development Methodology**

Schroer et al. (2021) describe CRISP-DM as a methodology that standardizes the process of data mining across various industries. This methodology provides a structured approach to planning and executing projects that involve extracting actionable knowledge from extensive datasets. It ensures that the insights gained are relevant and beneficial across different business contexts. Due to its adaptability, CRISP-DM can be customized to meet the specific needs of various industries, making it a favored framework among data professionals seeking to incorporate data-driven decision-making into their operations.

For the development of the carbon emissions prediction project, the CRISP-DM methodology was adopted. This approach structures the project into several key phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Following this framework ensures a well-defined pathway for the project, facilitating better planning and more efficient results. The CRISP-DM methodology enhances project readability, aids in anticipating potential challenges, and improves overall project management.

This methodology's iterative nature allows for flexibility and adaptability as new information emerges and project requirements evolve. In a dynamic field like carbon emissions, where data and conditions can change rapidly, the ability to refine and improve models continuously is crucial. CRISP-DM emphasizes understanding the project objectives and stakeholder requirements early in the process, ensuring that subsequent data analysis is aligned with overarching business goals.

By guiding the team through a structured process of problem understanding, data preparation, modeling, evaluation, and deployment, CRISP-DM proves indispensable. It includes a rigorous evaluation phase to assess model performance, which is critical for projects where decision-making relies on accuracy. Tailored specifically for data mining and not just software development, CRISP-DM adeptly handles the complexities of data-centric tasks, making it the optimal choice for this strategic, data-driven project with significant implications for business and environmental policy.

This is how the project development aligns with the phases of the CRISP-DM methodology.

***Business Understanding***

Martinez-Plumed et al. (2020) emphasized the significance of business understanding in data science projects, highlighting that the primary goal of these initiatives is to enhance business operations. Despite this, it was observed that many data science learning resources often neglected the importance of business acumen, focusing predominantly on technical proficiencies like machine learning and programming. However, Martinez argued that a robust grasp of business concepts substantially bolstered employability within the data science sector.

In the initial phase of CRISP-DM, the project team identified key objectives for their carbon emission prediction project, such as reducing environmental impact or informing policy decisions. Concurrently, project planning, resource allocation, and scope definition were undertaken. Crucially, all project stakeholders were identified, including governments, businesses, academics, and data providers, essential for aligning activities towards common objectives. Additionally, metrics and success criteria were established to guide project management, evaluation, and goal achievement. Success criteria provided explicit, quantitative requirements to gauge whether project objectives were met, guiding project activities throughout its lifecycle.

For instance, achieving a high level of accuracy in carbon emission forecasts was a primary success criterion. This was measured using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared values, which offered concrete measures of prediction accuracy. Metrics were numerical measures used to track and assess performance, ensuring that the predictions were reliable and useful for stakeholders.

These criteria and metrics enabled stakeholders to make data-driven decisions, identify trends, and monitor progress toward environmental sustainability goals, particularly in the context of carbon emissions trading and market dynamics. By setting clear, quantifiable targets, the project aimed to provide actionable insights that could inform policy decisions and strategies for carbon emissions reduction.

***Data Understanding***

In the second phase of CRISP-DM, the focus shifted to data identification and understanding, a critical step in the carbon emission prediction project. The team encountered challenges in acquiring relevant data due to discrepancies in carbon footprint tracking across countries. However, the team has successfully obtained three comprehensive datasets, first the World Bank dataset providing global carbon emissions data from 1850 to 2022, the U.S. Energy Information Administration (EIA) dataset encompassing carbon dioxide emissions data from various energy sources across U.S. states from 1970 to 2021, and data integrating real-time weather data through the National Weather Service's API.

Exploring subcategories within these datasets, such as emissions by sector, fuel type, and geographic region, enabled the team to discern nuanced emission patterns and their underlying causes. The objective was to identify subtle trends that could inform targeted policy decisions and interventions. In addition to standard meteorological parameters, intricate elements like atmospheric conditions, precipitation, and air pressure were incorporated, allowing for an understanding of dynamic weather variations and their impact on emission levels.

Prioritizing accuracy, timeliness, and reliability, the team meticulously vetted APIs and data sources to ensure the integrity of their real-time weather information. By leveraging reputable sources and robust API endpoints, they guaranteed the currency and accuracy of their weather data, enabling adaptation of strategies based on short-term trends and anomalies. Vigilant monitoring of data quality metrics, including relevance, consistency, and completeness, allowed identification and rectification of inconsistencies or anomalies, thereby ensuring the reliability of the study's findings. Adhering to stringent data quality standards enhanced the validity and credibility of the results, bolstering the project's overall integrity.

***Data Preparation***

In the Data Preparation phase of the CRISP-DM methodology, comprehensive steps were undertaken to ensure the datasets were clean, consistent, and ready for analysis in the carbon emission prediction project. This phase involved meticulous data cleaning, transformation, and preparation to facilitate accurate modeling and reliable predictions.

For the World Data obtained from the World Bank, the initial raw dataset underwent rigorous cleaning to handle missing values, address data noise, and remove duplicates or inconsistencies present. Following cleaning, the data was transformed through aggregation, regularization, and standardization techniques to ensure uniformity and enhance its usability for modeling purposes. Finally, the dataset was split into training, testing, and validation sets using the K-fold cross-validation method to optimize model training and evaluation.

The USA Data, sourced from the U.S. Energy Information Administration (EIA), underwent similar preprocessing steps. The raw data was cleaned to remove unnecessary information, handle duplicates and missing values, and ensure correct column formatting. Subsequent transformation steps included data smoothing and normalization to prepare the dataset for analysis. The cleaned and transformed data was then divided into training, testing, and validation sets based on chronological segmentation, facilitating a robust evaluation of the models.

The Weather Data, obtained from the National Weather Service (NWS) API, was cleaned to handle missing values and select the appropriate timeframe for analysis. Following cleaning, the data underwent discretization to categorize continuous variables into discrete bins, enhancing the interpretability of the model predictions. This data was prepared for analysis purposes, by integrating it with the other datasets, although it did not require the same level of segmentation due to its nature.

Overall, the Data Preparation phase ensured that all datasets were thoroughly cleaned, transformed, and appropriately divided for effective modeling and analysis. This meticulous approach to data preparation was crucial for the project's success, enabling accurate predictions of carbon emissions and providing a solid foundation for subsequent modeling and evaluation phases.

***Modeling***

In the fourth stage of the CRISP-DM methodology, the processed data is utilized to predict carbon emissions employing machine learning models. For this project, four appropriate machine learning models were selected to tackle this challenge.

Initially, a sophisticated model that combines Convolutional Neural Networks with Iterative Bayesian Filter Adaptation (CNN-IBFA) was used. This novel fusion utilizes CNNs' ability to capture intricate spatial patterns and Bayesian filters' iterative refinement capability. By integrating deep learning with Bayesian inference, this approach aimed to uncover subtle relationships in the data, enabling more accurate predictions of carbon emission footprints.

Next, a hybrid model that combines Convolutional Neural Networks (CNNs) with Artificial Neural Networks (ANNs) was employed. This hybrid approach leverages CNNs' ability to extract spatial features from data and ANNs' capacity to learn complex relationships between variables. By combining these two architectures, the aim was to enhance the model's performance in predicting carbon emissions.

The project also utilized Support Vector Machines (SVMs), a potent supervised learning algorithm renowned for its effectiveness in classification and regression tasks. SVMs were employed to identify the optimal hyperplane that segregates data points into distinct classes or forecasts continuous outcomes. Within this context, SVMs were leveraged to model the correlation between input variables and carbon emissions, with the goal of furnishing precise predictions based on historical data and pertinent features.

Along with the above models, the Long Short-Term Memory (LSTM) network was incorporated as a cutting-edge technique for time series forecasting. LSTM networks, a type of recurrent neural network (RNN), excel at capturing temporal dependencies and patterns in sequential data. By leveraging LSTM architecture, the project aimed to effectively model the dynamic nature of carbon emissions over time, thereby enabling accurate predictions of future emission trends. Through the utilization of LSTM networks, the project endeavored to enhance the predictive capabilities by efficiently capturing long-term dependencies and temporal intricacies present in the emission data.

***Evaluation***

The fifth stage of the CRISP-DM methodology involved identifying the model that best addressed the problem statement of the carbon emission prediction project. This was accomplished through a comparative analysis of four models, evaluated based on the metrics of Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

The evaluation phase commenced with a rigorous assessment of each model's performance on the predefined metrics. MAE provided a direct measure of average prediction errors, indicating the closeness of predictions to actual values. Lower MAE values signified better performance and higher accuracy. MSE emphasized larger errors due to its squaring component, making it sensitive to significant deviations between predicted and actual values. A lower MSE was desirable as it suggested fewer large errors and more accurate capture of underlying data trends. R-squared indicated the proportion of variance in the dependent variable that was predictable from the independent variables, with values closer to 1 suggesting a good model fit.

In conclusion, this comprehensive evaluation phase ensured that the selected model not only met the performance criteria but was also robust and ready for real-world deployment. This ultimately supported the project's goals of accurate carbon emission forecasting and effective carbon management strategies.

**Deployment**

During the deployment phase of the CRISP-DM methodology, four models—LSTM, CNN-IBFA, Hybrid CNN-ANN, and SVM—were executed and evaluated on Google Colab, leveraging its robust computational capabilities and seamless integration with cloud storage solutions like AWS. After a comprehensive evaluation using metrics such as MAE, MSE, and R-squared, the model with the best fit was selected for deployment and integrated into a user-friendly, web-based dashboard designed to present carbon emissions and offset data.

The interactive dashboard provided stakeholders with an accessible interface, displaying key metrics, trends, and predictions related to carbon emissions for the top four emitting countries: China, India, Russia, and the United States. This dashboard illustrates the projected costs to offset these countries' carbon footprints from 2023 to 2026, and predicted carbon emissions for the same period.

AWS services played a pivotal role in the deployment process. AWS Lambda facilitated running the model predictions in a serverless environment, ensuring scalability and cost-efficiency. The data was securely stored and managed in AWS S3, with access controlled via IAM policies to maintain security and compliance.

This deployment strategy ensured that the carbon emission predictions were accessible, actionable, and continuously updated, supporting informed decision-making and strategic planning for carbon management and sustainability goals. Additionally, the deployment phase included planning the deployment strategy, risk management, and future scope. Comprehensive documentation of the entire process, findings, and implications, along with identifying future plans, was crucial. By extensively training and validating models with diverse datasets, accurate predictions of future emission patterns were ensured, aiding decision-making for academics, energy analysts, and policymakers. A standardized project pipeline ensured consistency, transparency, and efficacy in dataset usage, streamlining the analytical process and enabling rapid decision-making.

Cloud service providers played a vital role in risk and access control, ensuring data confidentiality, privacy, and compliance. Leveraging modern analytics tools provided a comprehensive understanding of how energy usage impacts the environment, empowering stakeholders to drive sustainable change and make informed decisions for a greener future. This comprehensive deployment approach not only facilitated the immediate application of the carbon emission predictions but also laid a foundation for continuous improvement and adaptation in response to evolving data and environmental conditions.

### **2.3 Project Organization Plan**

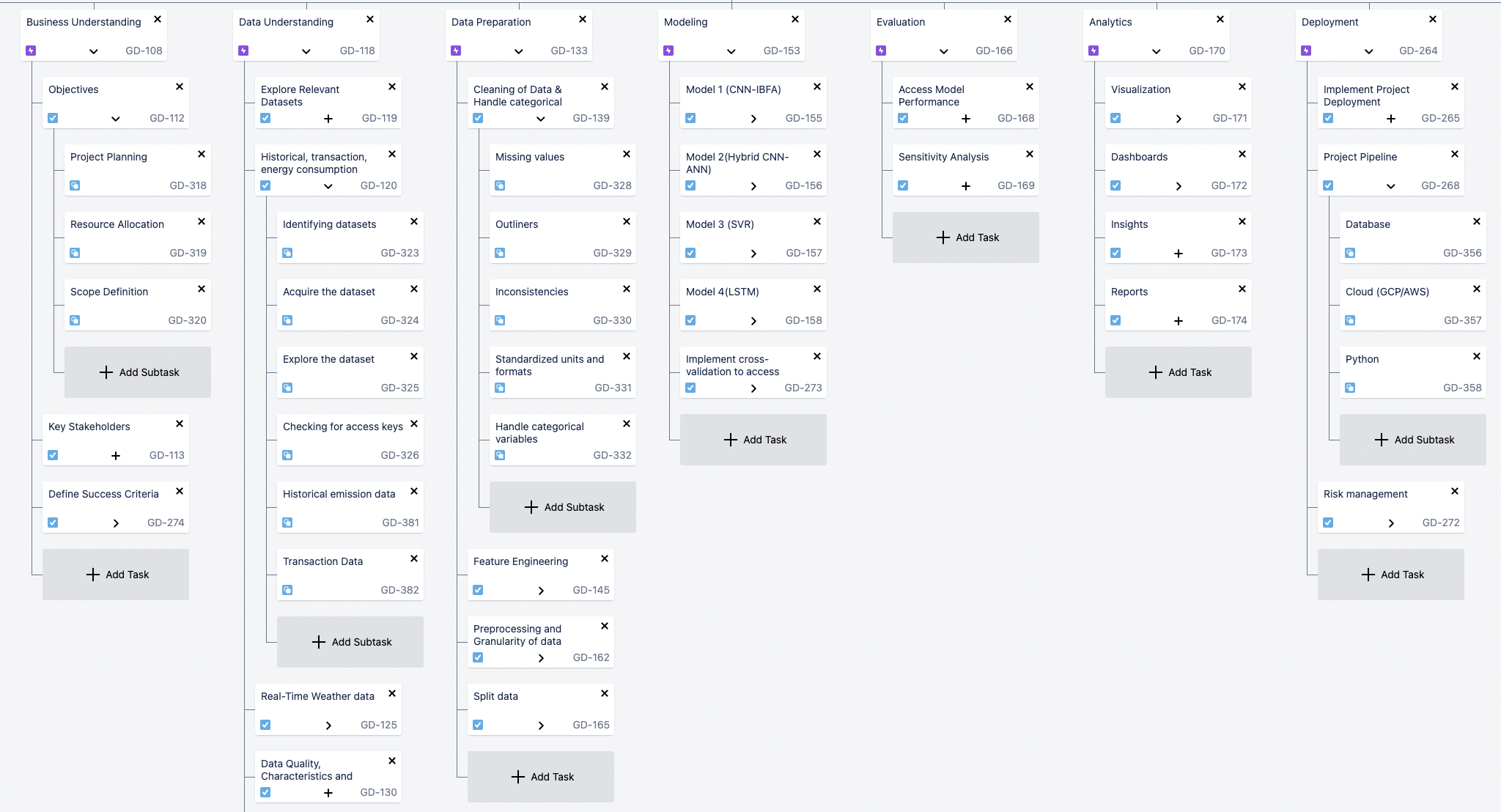
#### ***Work Breakdown Structure(WBS)***

Tausworthe (1979) emphasized the importance of the Work Breakdown Structure (WBS) as a vital tool for dividing a complex engineering project into smaller, more manageable subprojects, tasks, subtasks, and work packages. This planning tool establishes a coherent structure that connects project goals with the necessary resources and actions, functioning as a crucial monitoring mechanism throughout project execution. It enables the measurement of subtask completion in relation to the overall project plan. Although the Work Breakdown Structure (WBS) has been extensively used in different engineering fields, its formal implementation in software projects has been relatively restricted. Nevertheless, recent instances of triumphant integration of WBS in software projects have showcased its suitability and advantages for efficient project management.

The CRISP-DM methodology was adhered to in this project, and the WBS was employed to partition the project into smaller, more feasible tasks. The Work Breakdown Structure (WBS) depicted all project deliverables in a hierarchical and incremental format, systematically dividing larger tasks into smaller, more detailed subtasks. Afterward, these smaller tasks were allocated to individual team members or groups to be completed, which helped achieve the overall project goals. This representation method proved advantageous not only for the team members but also for stakeholders, clients, and project managers, offering a lucid and structured perspective of the project's advancement and individual obligations.

**Figure 5**

*Work breakdown structure*



Note: A high-level view of the project's WBS created in JIRA.

As shown in Figure 1, the entire project was broken down into multiple steps aligned with the CRISP-DM methodology, where each step represents a unique phase. In addition to the six primary phases of CRISP-DM, the Work Breakdown Structure (WBS) provided a detailed framework for executing the project efficiently.

The first step in the WBS is Business Understanding, which involves defining the project goal of predicting carbon emissions. Sub-tasks included project planning, resource allocation, and defining the scope. Identifying key stakeholders and success criteria, measured by carbon emission prediction accuracy, was crucial and informed by literature surveys.

The subsequent stage is Data Understanding, which focuses on identifying and sourcing appropriate data from reputable organizations such as the World Bank, the U.S. Energy Information Administration, and the National Weather Service. The dataset, comprising weather information, was securely stored in AWS and was accessible only to team members. Given the reliance on real-time API-based data, rigorous checks were conducted to ensure its quality. Descriptions of the data's structure, size, and quality were provided, and the data was further explored through visualizations. This phase also involved identifying potential data challenges and assessing data quality.

The third stage is Data Preparation, involving the processing of data to prepare it for input into machine learning models. This included handling missing values, outliers, inconsistencies, and categorical variables as needed. The granularity of the input data was determined based on the expected output value. Feature engineering was performed to extract important features for target value prediction. Subsequently, the data was divided into training and testing sets. Both the raw data and the processed data obtained in later stages were stored separately.

The fourth stage in the WBS chart is Modeling, which involves training models using pre-processed data. The selected models for this project included CNN-IBFA, LSTM, SVM, and CNN-ANN models. Following model training, performance evaluation was conducted based on metrics such as MSE, MAE, and R-squared error. Subsequently, model evaluation was performed, followed by hyperparameter tuning and optimization techniques like gradient descent to minimize the loss function.

The subsequent step is Evaluation, involving the comparison of predicted values from all four models with actual emission data to calculate error metrics. The model that best aligned with the success criteria was chosen through a comparative analysis of MSE and MAE. Additionally, sensitivity analysis was performed to assess the models' adaptability to variations in input data.

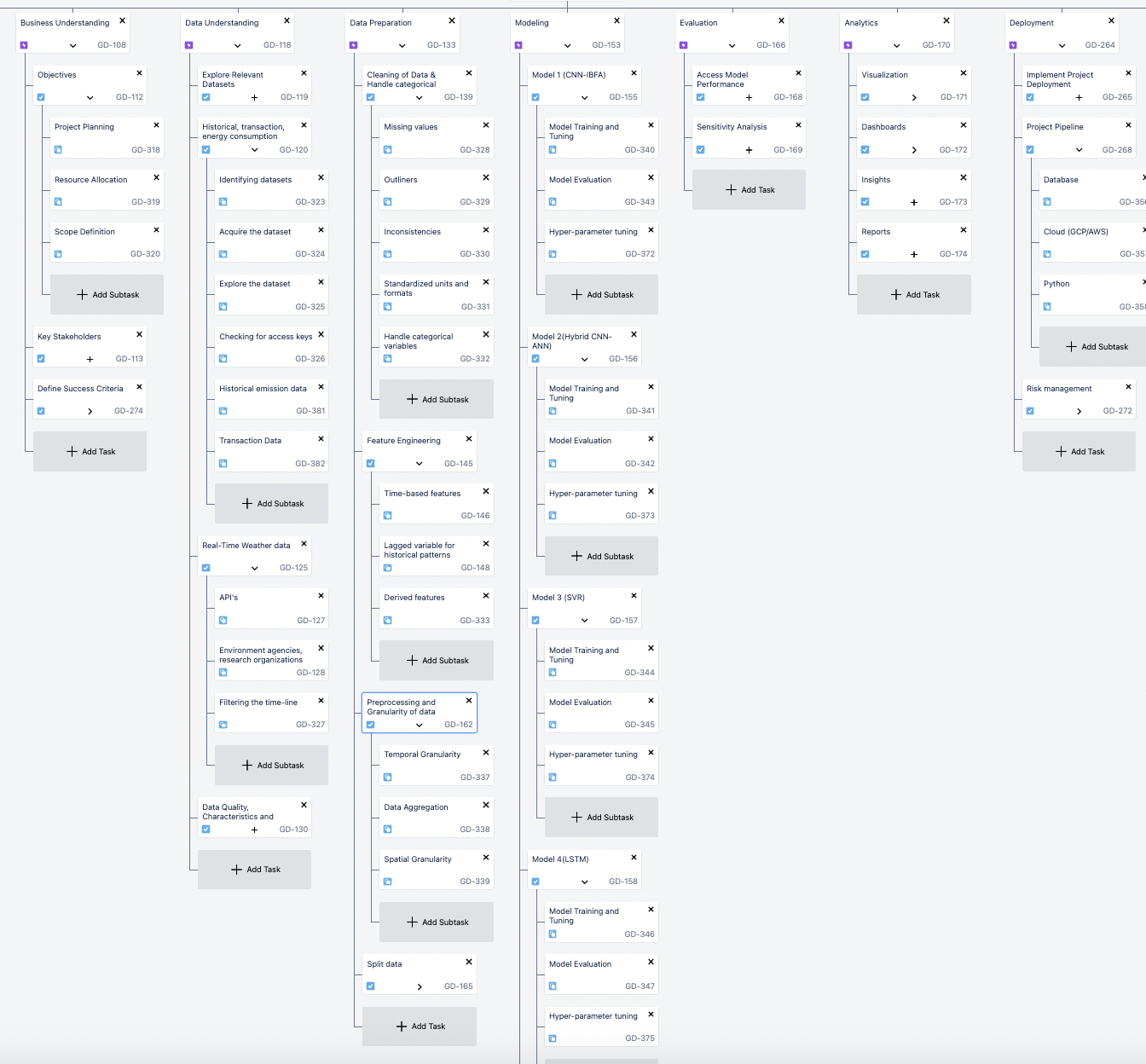
During the deployment phase of the CRISP-DM methodology, the selected models were executed on Google Colab, leveraging its high computational power and integration with AWS cloud storage solutions. AWS Lambda and S3 services were critical for running the models and securely managing data through IAM policies, ensuring scalability and cost-efficiency. The deployment process integrated the models into existing workflows, making them accessible to stakeholders like policymakers and energy analysts, thus facilitating informed decision-making on emission reduction and sustainable development. A standardized project pipeline from data collection to maintenance was established, with cloud service providers managing access and security. These measures ensured that carbon emission predictions were actionable and aligned with strategic carbon management goals, minimizing deployment risks.

In the analytics phase, Python was utilized to generate detailed visualizations and interactive dashboards, showcasing predictions for major carbon-emitting countries and their associated offset costs from 2023 to 2026. These tools illustrated carbon emissions and the financial implications of offsetting them, aiding stakeholders in monitoring sustainability objectives and evaluating energy efficiency measures. The visual aids enhanced decision-making, assisting in the prioritization of emission reduction strategies and assessment of policy impacts. Comprehensive documentation was maintained alongside these visualizations to ensure meticulous record-keeping, supporting the project’s aim of providing actionable insights for sustainable development and carbon neutrality.

The figure below displays the initial segment of a detailed Work Breakdown Structure, which includes all components and work packages for each phase of the CRISP-DM methodology within the WBS. The figure below displays the remaining segment of a detailed Work Breakdown Structure, which includes all components and work packages for each phase of the CRISP-DM methodology within the WBS.

**Figure 6**

*WBS Hirachical*

**

*Note.* WBS Sub-levels.

### **2.4 Project Resource Requirements and Plan**

#### ***2.4.1 Hardware Requirements***

Any machine learning (ML) project must consider the hardware requirements because ML models usually need a lot of processing power. These models sometimes have intricate algorithms that require handling several data sets simultaneously, which can be resource-intensive and time-intensive. Project delays and higher expenses may result if processing capacity constraints force model training time frames to be extended.

Thus, having the right hardware specs, such as a strong CPU, plenty of Memory, and an acceptable GPU, will have a significant influence on the speed and efficiency of this ML-based project. It also expedites model training, enhances model accuracy, and improves project results. Table 2 states all the hardware requirements that will be required in terms of hardware.

**Table 6**

*Hardware Requirements for the project*

| **Hardware** | **Configurations** | **Justification** | **Purpose** | **Cost** |
| --- | --- | --- | --- | --- |
| Laptops | Intel Core i7 with 16GB RAM, 512GB SSD, and GPU of Nvidia GTX 1660Ti with 8GB VRAM | Required for parallel processing and handling intensive machine learning and deep learning model computations. | Training, testing, and validation of models. | 4 units at $700 each |
|  |  | Total |  | $2800 |

Note. Hardware Requirements for the project.

#### ***2.4.2 Software Requirements***

In our technology-focused research project, software tools are crucial for data analysis and modeling. It's essential to ensure that the software versions, including Python libraries and AWS services, are compatible with our datasets. This compatibility optimizes performance and prevents issues during data processing. The project uses a range of software from data manipulation with Pandas to advanced machine learning with TensorFlow and Keras. AWS services such as Glue, S3, IAM, and Athena are utilized for efficient ETL processes, data storage, and access management, ensuring seamless data handling and security. This comprehensive software setup supports our analytical procedures and enhances the reliability of our project outcomes. Below is a table outlining the various software packages and libraries utilized in the project.

**Table 7**

*Software / Libraries Requirements*

| **Specification** | **License** | **Justification** | **Cost (per unit)** |
| --- | --- | --- | --- |
| PyTorch | Open Source | Flexible framework for building and training deep learning models. | Free |
| Pandas | Open Source | Data manipulation and preprocessing. | Free |
| Numpy | Open Source | Numerical operations | Free |
| Matplotlib and Seaborn | Open Source | Used for data visualization to clearly communicate the results and insights derived from the data analysis. | Free |
| TensorFlow | Open Source | Supports deep learning models, facilitating the development of sophisticated neural network architectures needed for this project. | Free |
| Keras | Open Source | Making neural network testing and prototyping quick. | Free |
| Scikit-learn | Open Source | To train and tune the models. | Free |
| spaCY | Open Source | Provides a range of tools and pre-trained algorithms for handling and interpreting text data. | Free |
| stats models | Open Source | Provides comprehensive statistical modeling capabilities, facilitating in-depth data analysis and hypothesis testing | Free |
| CodeCarbon | Open Source | Tracks and estimates the amount of CO2 emissions generated by the compute resources used during model training or any computation task | Free |
| Flask | Open Source | Visualization purpose in Localhost | Free |
| Boto | Open Source | To connect with AWS | Free |
| AWS (glue, S3, IAM, Athena) | Paid Version | AWS services will be used for ETL processes to clean and transform data, with storage in S3 for easy access along with Athena. | $100/ month |
|  |  | Total | $400 |

Note. The Software / Libraries Requirements of this project.

#### ***2.4.3 Tools and Licenses***

The project leverages a variety of tools to support its various phases, each selected for its specific benefits and functionality. Jupyter Notebook is employed for interactive data exploration, analysis, and model prototyping, providing a robust environment for real-time code execution and visualization. Google Colab is utilized to share code among teammates seamlessly, ensuring that dependencies and environments are managed without hassle. For project management, JIRA is used to create detailed roadmaps, helping to organize and track project progress efficiently. GitHub serves as the version control system, making the project publicly available and fostering a collaborative improvement environment. Zoom facilitates communication and coordination among team members, essential for maintaining project momentum. Grammarly’s paid version is employed to ensure all documentation is clear and professional, enhancing the readability and quality of project reports. Google Docs supports real-time collaborative document editing, enabling multiple team members to work simultaneously and efficiently. Collectively, these tools streamline the project's execution and enhance team productivity and project quality.

**Table 8**

*Tools Used*

| **Tools** | **License** | **Purpose** | **Cost** |
| --- | --- | --- | --- |
| Jupyter Notebook | Open Source | Interactive environment for conducting data exploration, analysis, and model prototyping. | Free |
| Google Colab | Open Source | Share the code work with teammates and no need to worry about any dependency and environment. | Free |
| JIRA | Open Source | A project management tool that helps in making a roadmap for the project. | Free |
| GitHub | Open Source | Version control tools enable the project to be made publicly available, allowing others to learn from and make improvements to the existing model. | Free |
| Zoom | Open Source | Connect with the team members and plan for the task and the coordination management. | Free |
| Grammarly | Paid Version | To make the reports free from all grammatical errors | $12/ month |
| Google Doc | Open Source | To make documents simultaneously in a group. | Free |
|  |  | Total Cost | $48 |

*Note.* Tools and Licenses required for the project.

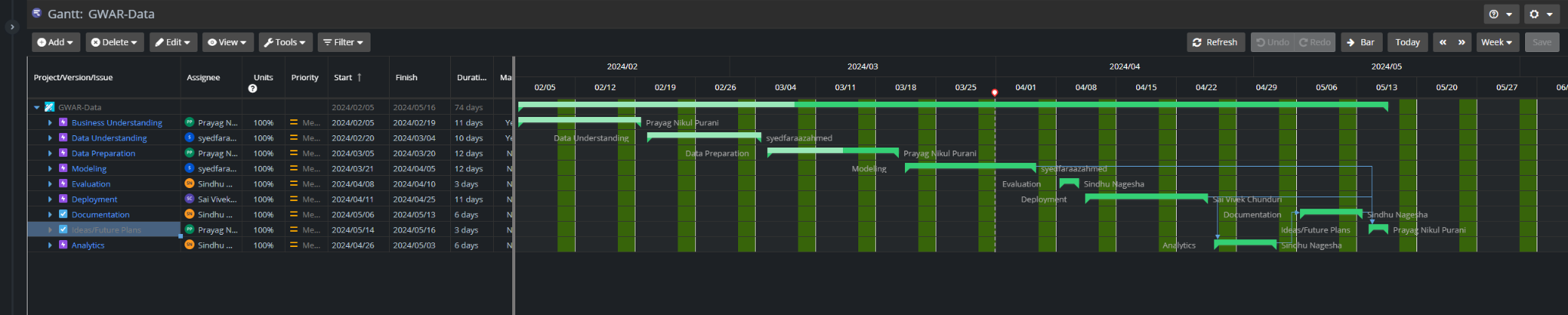
### **2.5 Project Schedule**

#### ***2.5.1 Gantt Chart***

A WBS (Work Breakdown Structure) Gantt chart is a graphical project management tool that integrates the comprehensive hierarchical breakdown of tasks provided by a Work Breakdown Structure with the timeline-based scheduling format of a Gantt chart. It serves as a means to meticulously plan, organize, and schedule every aspect and activity involved in a project. The figure below outlines the overall Gantt chart for this project.

**Figure 8**

*Gantt chart for prediction of carbon emission using ML*

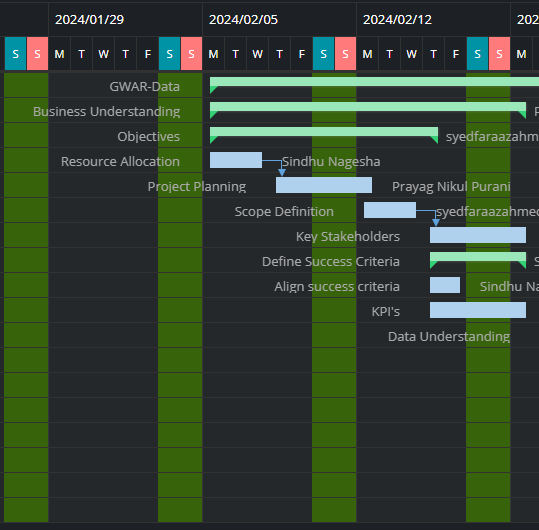


*Note.* Gantt chart for overall carbon emission prediction.

Figure X below outlines the stages of the project from Business Understanding. It includes goals such as assessing emission patterns, understanding and preparing data, modeling with techniques like regression, evaluating models, and deploying them into decision-making processes. Documentation and future planning are also highlighted. Analytics plays a crucial role throughout the project, generating insights to address carbon emissions and environmental concerns, and providing actionable information for climate change mitigation.

**Figure 9**

*Gantt chart for prediction of carbon emission using ML*



*Note.* Gantt Chart of the Business Understanding Phase.

Figure 3 outlines the initial phase of the project, defining objectives such as estimating future emissions patterns or analyzing historical trends. Resource Allocation involves identifying and assigning resources, both tangible and intangible, to ensure project success. Project Planning entails creating a detailed plan with tasks, schedules, and communication channels. Important Stakeholders, including policymakers and community groups, are involved early to consider their interests. Success Criteria are quantifiable standards indicating project success, such as lower emissions or increased efficiency. Aligning Success Criteria with stakeholder expectations ensures project relevance and support. Key Performance Indicators (KPIs) are measurable metrics like model accuracy or emission reduction, guiding project performance assessment and decision-making. This structured approach ensures efficient resource utilization, stakeholder engagement, and progress tracking throughout the project lifecycle. This Phase is scheduled to be completed by February 12, 2024, as shown in the above figure.

**Figure 10**

*Gantt chart for prediction of carbon emission using ML*

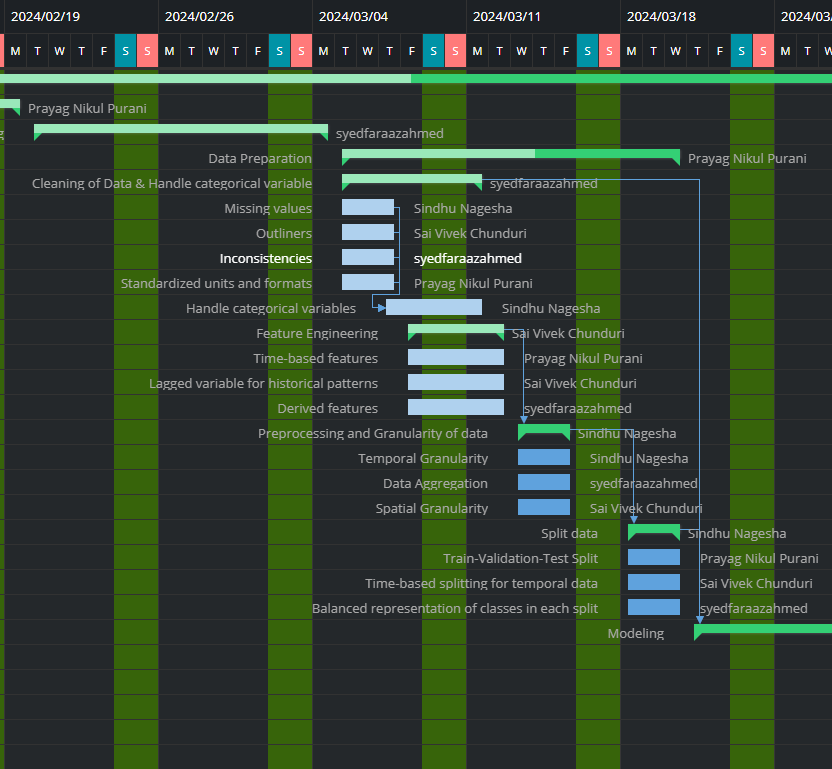


*Note.* Gantt Chart of the Data Understanding Phase

Figure 4 outlined the Data Understanding phase of the project, highlighting the need for historical carbon dioxide emissions data related to state energy programs to identify long-term trends. These datasets, ideally spanning several decades, were crucial for analyzing market dynamics and consumption patterns linked to carbon trading or energy use. Data was sourced from credible government organizations, ensuring access to regularly updated and well-curated datasets. Additionally, the phase incorporated the use of APIs to gather real-time meteorological data, essential for assessing the impact of weather conditions on emissions and energy consumption. This task will be completed by March 4, 2024, as shown in the above figure.

**Figure 11**

*Gantt chart for prediction of carbon emission using ML*



*Note.* Gantt Chart of the Data Preparation Phase.

Figure 5 details the data preparation phase, crucial for our study on state energy-related carbon dioxide emissions and carbon intensity indicators. Managing missing values is addressed through imputation or deletion to prevent bias. Outliers are handled using smoothing and winsorizing techniques to maintain data integrity. Standardizing units and formats ensures coherence, while feature engineering involves creating or modifying features to capture relevant information and relationships. Time-based variables, lagged variables, and derived properties are utilized for deeper insights. Spatial and temporal granularity considerations ensure pertinent patterns and changes are captured accurately in the analysis. This phase is scheduled to be completed by March 18, 2024.

**Figure 12**

*Gantt chart for prediction of carbon emission using ML*

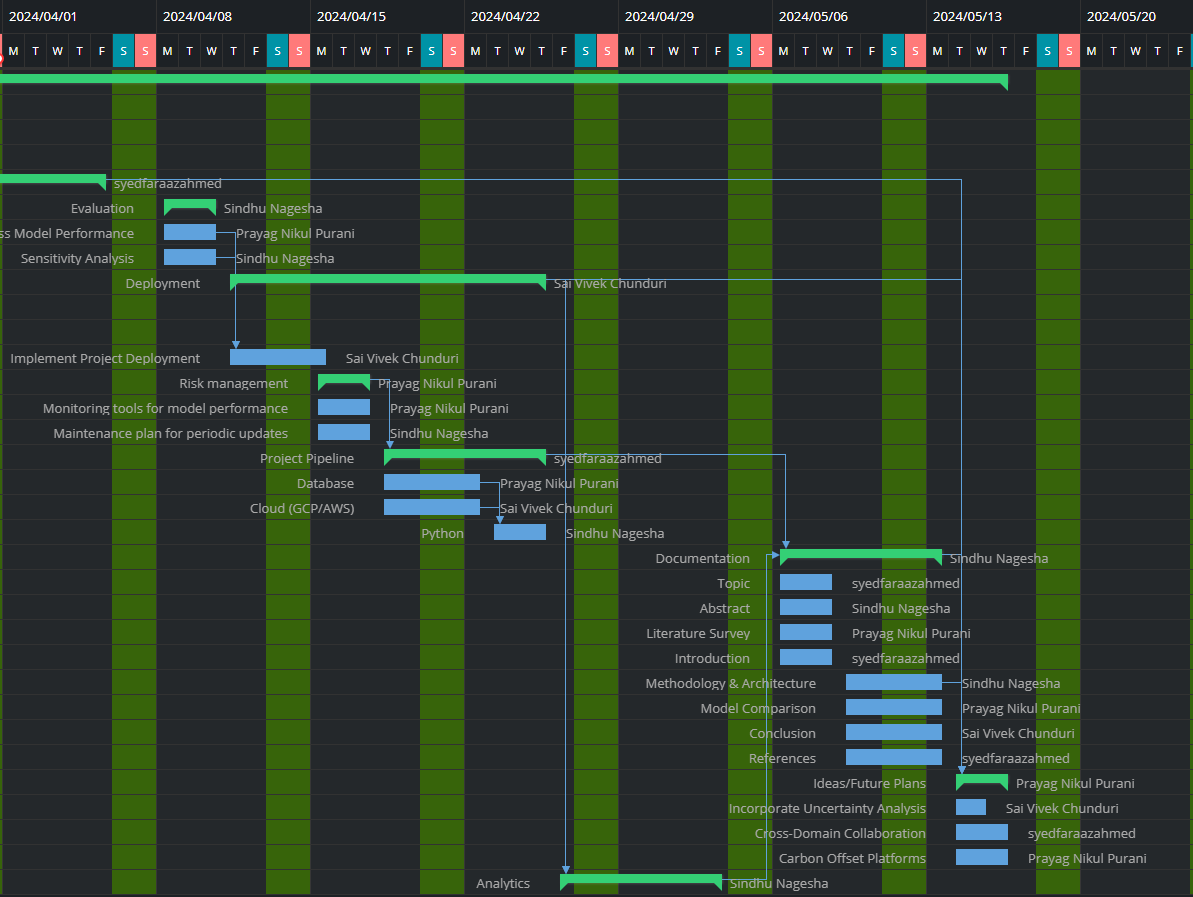


*Note.* Gantt Chart of the Modeling Phase.

Here, the data previously prepared and split into test, train, and validation datasets is ready for modeling. This phase involves developing models using four distinct algorithms: Convolutional Neural Network with Iterative Bayesian Filter Adaptation (CNN-IBFA), Long Short-Term Memory (LSTM), Support Vector Machines (SVM), and Convolutional Neural Network-Artificial Neural Network Hybrid (CNN-ANN). Each model is implemented by individual team members who then tune their models using hyperparameters to enhance performance. This task is scheduled to be completed by April 8, 2023, as evident from the above Figure.

**Figure 13**

*Gantt chart for prediction of carbon emission using ML*

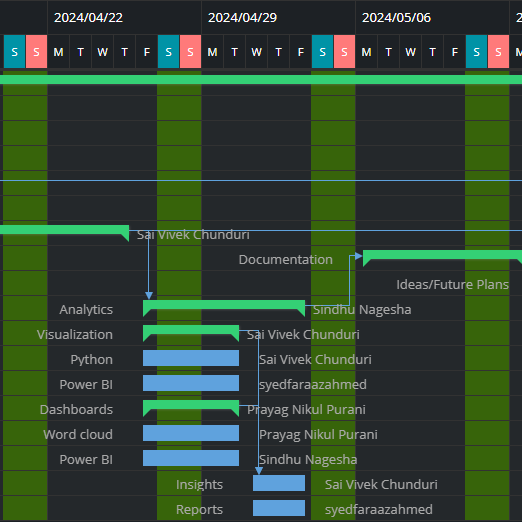


*Note.* Gantt Chart of the Deployment Phase.

Figure 7 illustrates the deployment phase, where parameters influencing emissions and intensity measurements are systematically altered to understand their impact. Stakeholders utilize this information for policy development and emissions reduction strategies. Risk management techniques mitigate potential setbacks, while monitoring tools ensure model performance and maintenance schedules guarantee relevance over time. Defined tasks, dependencies, and workflows to expedite project execution and team participation. Storage and management of project data in a structured database facilitate retrieval and analysis. Cloud infrastructure like AWS enhances project capabilities, providing cost-effectiveness and scalability. Python, with libraries like NumPy and Pandas, enables efficient data analysis, modeling, and visualization for examining carbon emissions and intensity measures. This phase is scheduled to be completed by April 22, 2024.

**Figure 14**

*Gantt chart for prediction of carbon emission using ML*



*Note.* Gantt Chart of the Analytics Phase.

Figure 8 highlights the analytics phase, where Python tools such as Matplotlib, Seaborn, and Plotly are used to create visualizations for examining correlations, trends, and regional patterns in carbon emissions. Python along with Flask facilitates dynamic dashboards with various visualizations and filters that support data-driven decision-making. Word clouds are employed to analyze textual data regarding industry trends and policies. Reports provide comprehensive insights, narratively explaining trends and patterns in the data to aid decision-making and policy development aimed at emissions reduction and sustainable energy practices. This phase is scheduled to be completed by May 6, 2024, including all documentation and presentations of the project.

**2.5.2 Pert Chart**

PERT (Program Evaluation Review Technique) chart is a visual representation that depicts the timeline of a project by showing all the individual tasks that must be accomplished to complete the project. Creating a PERT Chart for the project provides a clearer understanding of project timelines, identifies potential risks, and makes informed decisions about resource allocation and project management. A task-based pert chart for the whole project has been developed.

The visual representation of this project's PERT chart, depicted in Figure X below, serves as a concise overview. It aids in strategizing the approach and resource allocation, ensuring efficient task management. By leveraging this visual tool, members have optimized resource utilization and streamlined the pipeline to yield maximum efficiency in output.

A task-oriented PERT Chart has been developed, encompassing a total of 59 tasks. This chart aligns with the CRISP-DM methodology for project management. Each task within the chart is meticulously structured to ensure clarity in project planning and execution. By adhering to the CRISP-DM methodology, the project team aims to streamline the data mining process, from understanding business objectives to deploying the final model. The comprehensive nature of the PERT Chart allows stakeholders to visualize the project timeline, dependencies, and critical path, facilitating effective project monitoring and control.

The critical path is the sequence of tasks that determines the shortest possible duration for completing the project. The red arrowed path in Figure 9 below depicts the critical path for this project, highlighting the tasks that must be completed on time to prevent delays in the overall project timeline.

1. Forward Pass (Calculate Earliest Start (ES) and Earliest Finish (EF))
   * EF = ES + duration (of the activity)
2. Backward Pass (Calculate Latest Start (LS) and Latest Finish (LF))
   * LS = LF - duration (of the activity)

Additional Formulas (to calculate slack):

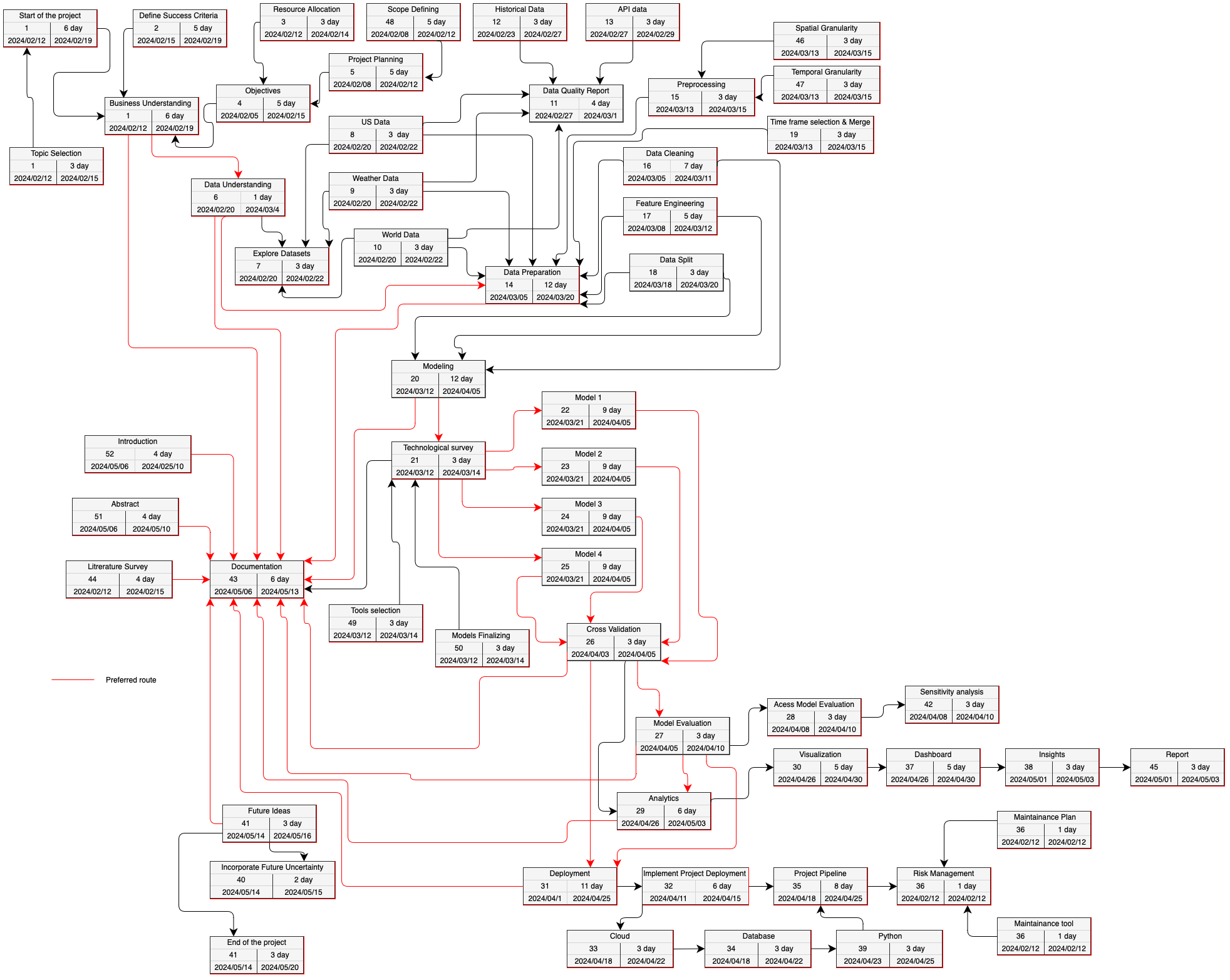
Total Float (TF): LF - ES (represents the buffer time an activity can be delayed without impacting the project deadline).

Free Float (FF): Minimum of LS of successor activities - ES (represents the buffer time without impacting dependent activities).

The total duration of the critical path is about 77 days, and any slack for these tasks will delay the project deliverable hence these are considered critical tasks. Identifying the critical path is essential to prioritize activities and allocate resources effectively to ensure timely project completion. The critical path follows these task numbers “1—>6–>14—>20—>26—>27—>29—>31—>40—>41—>44—>51—>52”

**Figure 15**

*Pert Chart for the project*

****

*Note.* PERT Chart for prediction of carbon emission project.

### **References**

Acheampong, A. O., & Boateng, E. B. (2019). Modeling carbon emission intensity: Application of artificial neural network. Journal of Cleaner Production, 225(225), 833–856. https://doi.org/10.1016/j.jclepro.2019.03.352

Alshatri, N., & Hussain, F. K. (2023, October 19). Data-driven approach of external factors influencing carbon credit prices: A systematic literature review. Social Science Research Network. https://doi.org/10.2139/ssrn.4602609

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. ArXiv:2007.03051 [Cs, Eess, Stat]. https://arxiv.org/abs/2007.03051

Bhatt, H., Davawala, M., Joshi, T., Shah, M., & Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. Cleaner Chemical Engineering, 5, 100095.

Boateng, E. B., Twumasi, E. A., Darko, A., Tetteh, M. O., & Chan, A. P. C. (2020). Predicting building-related carbon emissions: A test of machine learning models. Enabling AI Applications in Data Science, 247–266. https://doi.org/10.1007/978-3-030-52067-0\_11

Carbon Credits. (2022, January 26). The ultimate guide to understanding carbon credits. Carbon Credits. https://carboncredits.com/the-ultimate-guide-to-understanding-carbon-credits/

Chi, Y. J., Pan, H.-Y., Cheng, F.-J., Chang, Y.-I., & Chuang, P.-C. (2022). Experience of carbon monoxide poisoning and the outcome predicting score: A multicenter retrospective study. The American Journal of Emergency Medicine, 58, 73–78. https://doi.org/10.1016/j.ajem.2022.05.012

Ding, F., Wu, H., Zhu, G., & Shi, Y. Q. (2020). METEOR: Measurable energy map toward the estimation of resampling rate via a convolutional neural network. IEEE Transactions on Circuits and Systems for Video Technology, 30(12), 4715-4727.

Ethayarajh, K., & Sadigh, D. (2020, October 12). BLEU Neighbors: A Reference-less Approach to Automatic Evaluation. ArXiv.org. https://doi.org/10.48550/arXiv.2004.12726

Guðbrandsdóttir, H. (2011). Predicting the price of EU ETS carbon credits: A correlation, principal component, and latent root approach. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0411203a5f91aa5cc8af9ccda46694032e52f5ec

Hu, B., & Cheng, Y. (2023). Predicting regional carbon price in China based on multi-factor HKELM by combining secondary decomposition and ensemble learning. PLOS ONE, 18(12), e0285311–e0285311. https://doi.org/10.1371/journal.pone.0285311

Huang, Y., & He, Z. (2020). Carbon price forecasting with optimization prediction method based on unstructured combination. Science of the Total Environment, 725(138350), 138350. https://doi.org/10.1016/j.scitotenv.2020.138350

Kenton, W. (2023, November 17). Carbon credit. Investopedia. https://www.investopedia.com/terms/c/carbon\_credit.asp

Kunda, D., & Phiri, H. (2017). An approach for predicting CO2 emissions using data mining techniques. International Journal of Computer Applications, 172(13), 7-10.

Li, H., Huang, X., Zhou, D., Cao, A., Su, M., Wang, Y., & Guo, L. (2022). Forecasting carbon price in China: a multimodel comparison. International Journal of Environmental Research and Public Health, 19(10), 6217.

Liu, L., Chen, C., Zhao, Y., & Zhao, E. (2015). China׳s carbon-emissions trading: Overview, challenges and future. Renewable and Sustainable Energy Reviews, 49, 254–266. https://doi.org/10.1016/j.rser.2015.04.076

Lu, H., Ma, X., Huang, K., & Azimi, M. (2020). Carbon trading volume and price forecasting in China using multiple machine learning models. Journal of Cleaner Production, 249, 119386. https://doi.org/10.1016/j.jclepro.2019.119386

Mao, Y., & Yu, X. (2024). A hybrid forecasting approach for China’s national carbon emission allowance prices with balanced accuracy and interpretability. Journal of Environmental Management, 351(119873), 119873. https://doi.org/10.1016/j.jenvman.2023.119873

Mardani, A., Liao, H., Nilashi, M., Alrasheedi, M., & Cavallaro, F. (2020). A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. Journal of Cleaner Production, 275, 122942. https://doi.org/10.1016/j.jclepro.2020.122942

Peterdy, K. (n.d.). Carbon credits. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/esg/carbon-credit/

Usman, M., Mendes, E., Weidt, F., & Britto, R. (2014). Effort estimation in agile software development. Proceedings of the 10th International Conference on Predictive Models in Software Engineering - PROMISE ’14. https://doi.org/10.1145/2639490.2639503

Wang, Z.-X., & Ye, D.-J. (2017). Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. Journal of Cleaner Production, 142, 600–612. https://doi.org/10.1016/j.jclepro.2016.08.067

Wen, L., & Cao, Y. (2020). Influencing factors analysis and forecasting of residential energy-related CO2 emissions utilizing optimized support vector machine. Journal of Cleaner Production, 250, 119492. https://doi.org/10.1016/j.jclepro.2019.119492

What are carbon markets? (n.d.). Www.youtube.com. https://www.youtube.com/watch?v=dHj-H1giztA

Wijaya, C. Y. (2021, December 31). Data Science Project to Improve Your Business Understanding. Medium. https://towardsdatascience.com/data-science-project-to-improve-your-business-understanding-776386abbf63

Wu, L., Liu, S., Liu, D., Fang, Z., & Xu, H. (2015). Modelling and forecasting CO2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. Energy, 79, 489-495.

Youcef Djenouri, Asma Belhadi, Srivastava, G., & Lin, J. (2022). Hybrid graph convolution neural network and branch-and-bound optimization for traffic flow forecasting. Hybrid Graph Convolution Neural Network and Branch-And-Bound Optimization for Traffic Flow Forecasting, 139, 100–108. https://doi.org/10.1016/j.future.2022.09.018

Zhao, T., & Liu, Z. (2022). Drivers of CO2 Emissions: A Debt Perspective. International Journal of Environmental Research and Public Health, 19(3), 1847. https://doi.org/10.3390/ijerph19031847

Zhao, Y., Liu, L., Wang, A., & Liu, M. (2023). A novel deep learning based forecasting model for carbon emissions trading: A comparative analysis of regional markets. Solar Energy, 262, 111863. https://doi.org/10.1016/j.solener.2023.111863

Zhou, W., Zeng, B., Wang, J., Luo, X., & Liu, X. (2021). Forecasting Chinese carbon emissions using a novel grey rolling prediction model. Chaos Solitons & Fractals, 147, 110968–110968. https://doi.org/10.1016/j.chaos.2021.110968

Zhu, F., Liu, P., & Hu, P. (2023). Prediction of Carbon Emission Right Price Based on XGBoost Algorithm. Frontiers in Business, Economics and Management, 7(1), 61-67.

### **Appendix**

| Context | Links (Requested to open with SJSU email only) |
| --- | --- |
| Document | <https://docs.google.com/document/d/1IvdLsrZbVTQyOaKAgUWdF7Nvjr4MHAbIvgxvsXrvfuw/edit> |
| WBS | <https://sjsu-prayagnikulpurani.atlassian.net/jira/software/projects/GD/apps/787143b2-23f5-41c2-ba67-91994c652ab7/47e35f85-18b2-4af3-a1ba-88ce6e019565> |
| Gantt-chart | <https://sjsu-prayagnikulpurani.atlassian.net/projects/GD?selectedItem=com.atlassian.plugins.atlassian-connect-plugin:jp.ricksoft.plugins.wbsgantt-for-jira__edit-link-gantt-software> |
| Effort Estimation | <https://docs.google.com/spreadsheets/d/1TP3OounvqWlNBDF06v7TrnS78NOyy5nm3C39h9qwrcw/edit?usp=sharing> |
| Pert - chart (.sdr file) | <https://drive.google.com/file/d/1wFOzpdHeHHoOgW5CebTKCH0Z2rK0UrnG/view?usp=drive_link> |
| Dataset | <https://www.eia.gov/environment/emissions/state/> |
| API(National weather service) | <https://www.weather.gov/documentation/services-web-api#/default/alerts_query> |