AI enabled Decarbonization Framework

Ayush Jain*
Manikandan Padmanaban*
ayush.jain@ibm.com
manipadm@in.ibm.com
IBM Research Labs, India
Bengaluru, Karnataka, India

Jagabondhu Hazra
Ranjini Guruprasad
jahazra1@in.ibm.com
rangurup@in.ibm.com
IBM Research Labs, India
Bengaluru, Karnataka, India

Heriansyah Syam
Triputra Agro Persada
Jakarta, Indonesia
heriansyah@tap-agri.com

ABSTRACT

Eight major supply chains contribute to more than 50% of the global greenhouse gas emissions (GHG) [1]. These supply chains range from raw material to end-product manufacturing. Hence, it is critical to accurately estimate the carbon footprint of these supply chains, identify GHG hotspots, explain the factors that create the hotspots and carry out what-if analysis to reduce carbon footprint of supply chains. Towards this, we propose an enterprise decarbonization accelerator framework with a modular structure that automates carbon footprint estimation, identification of hot spots, explainability and what-if analysis to recommend measures to reduce carbon footprint of supply chains. To illustrate the working of the framework, we apply it to the cradle to gate extent of palm oil supply chain of a leading palm oil producer. The framework identified that the farming stage is the hot spot in the considered supply chain. As the next level of analysis, the framework identified the hotspots in the farming stage and provided explainability on factors that created hotspots. We discuss the what-if scenarios and the recommendations generated by the framework to reduce the carbon footprint of the hotspots and the resulting impact on palm oil tree yield.

KEYWORDS

Supply chain, AI, Decarbonization, Carbon accounting, Hotspot identification, what-if-analysis, Palm oil supply chain

1 INTRODUCTION

In the last few decades, supply chains have become more global, multi-echelon, inter-connected, and dynamic leading to benefits in terms of reducing costs, enhanced speed, diversifying operational sourcing, and quality. However, these shifts bring with it the massive contribution of supply chains to GHG emissions. It is estimated that goods and services that are traded internationally contribute to about 22% of the global GHG emissions [2]. Eight major supply chains are contributing to more than 50% of the global GHG emissions. These eight supply chains include food, construction, fashion, fast moving consumer goods (FCMG), electronics, auto, professional services and other freight. Of these, food contributes to more than one third followed by construction which contributes to 10% of the global GHG emissions [1].

Enterprises in general and particularly those who operate/are part of the above-mentioned supply chains are under significant pressure from investors, consumers, and policymakers to disclose their GHG emissions and commit to reduce emissions. Over 20 percent of the world's largest companies have set long term net-zero targets [3]. To achieve net-zero targets, enterprises need technology to measure, track, and decarbonize (reduce their emissions) while building operational resiliency to the effects of climate change.

In this paper, we will discuss about the novel framework/workflow called Enterprise Decarbonization Accelerator (EDA). EDA performs emission computation, hotspot identification with explainability, and what-if analysis to provide recommendations in an automated manner to accelerate decarbonization journey of enterprises. To demonstrate the efficacy of the EDA in accelerating the decarbonization process, we will apply the EDA to palm oil supply chain to measure carbon footprint, identify, explain the factors casuing the hotspots, and use what-if analysis to provide recommendations to mitigate carbon hotspots of the enterprise.

The rest of the paper is organized as follows. Section 2 summarizes the related work. Section 3 describes the EDA framework to estimate GHG emissions, identify hot spots, provide explainability and perform what-if analysis to provide recommendations for reduction/removal of hot spots. Section 4 describes palm cultivation, the plantation data obtained from a leading producer of palm oil and application of EDA to the palm oil supply chain. Section 5 presents the results obtained by applying the EDA to the palm oil supply chains. Finally, we present the concluding remarks in section 6.

2 RELATED WORK

Decarbonization of supply chain is an important area of study, which includes accounting of carbon footprints, identifying an inefficient process, understanding the factors which attributes to low performance and recommending the feasible intervenable actions for overall carbon reductions. This enables businesses to identify sustainability impacts across a range of attributes such as economic, environment, social and governance. It allows decision-makers to identify sustainability opportunities and prioritize reduction actions. In the contemporary literature, most of the decarbonization works revolves around carbon accounting and hotspot identification by using either qualitative [4] or quantitative approaches[5].

Hot Spot Analysis (HSA) [4] is a qualitative approach that uses relative relevance numbers from existing studies to give a rough overview of relevant sustainability aspects. This approach has been used to identify hotspots in supply chains by comparing the relevance numbers. [6] utilised HSA to integrate social and environmental dimensions along the entire value chain and to identify relevant aspects for a product specific sustainability management. [7] used the Social Hotspot Database to study the social hotspots of numerous product categories, while [8] conducted a global hotspot analysis concerning food loss and waste (FLW) with its associated GHG emissions. HSA has also been used to analyze environmental impacts of food supply chains, like [9] performed HSA to identify

*Both authors contributed equally to this research.

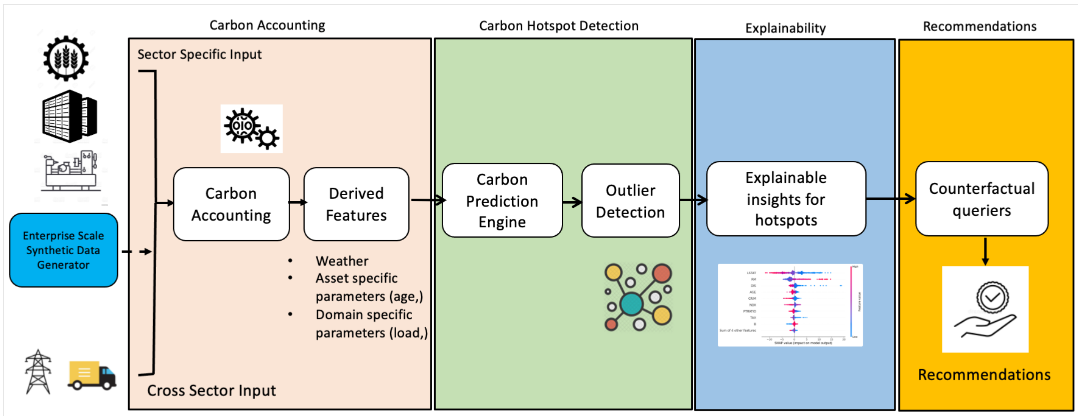

Figure 1: Enterprise Decarbonization Accelerator Framework

resource intensive hotspots in the life cycles of coffee and cream cheese. However, these qualitative approaches have limited applications as they do not provide any actual values for the impact factors but rather use relative relevance numbers to give a rough overview of relevant sustainability aspects.

Life Cycle Assessment (LCA) is a quantitative systems approach aimed at assessing the environmental impact of a product throughout its life-cycle. Works like [10], [11] and [12] utilized LCA to evaluate carbon dioxide equivalent (CO_2e) emissions and identify carbon hotspots in bio-diesel, maize silage and beef supply chains respectively. However, both HSA and LCA methodologies do not provide explainable insights of the identified hotspots. Also, they do not provide the stakeholders with recommendations that can help them in reducing their environmental impact. Attributional and Consequential LCA (ALCA and CLCA) [13] address this gap to some extent. ALCA provides attribution of total emissions from the processes and material flows in a product life cycle, while CLCA provides information about the consequences of changes in the level of output of a product on the total emissions associated with the product. However, these approaches are limited to product life cycle and cannot be used for enterprise level decarbonization.

As we will be evaluating the proposed framework for estimating the carbon footprint and hotspot identification for palm oil supply chain, we next discuss the related work in the space of palm oil supply chains. Several studies have focused on estimating the carbon footprint associated with the production of palm oil. [14] evaluated crude palm oil's GHG balance through an LCA approach, using average data from Brazil region, while [15] assessed water footprint of palm oil supply chain. Few works have focused on specific parts of the cradle to gate extent of the supply chain, such as LCA of oil palm seedling [16] and transportation [17]. Most of the prior work on palm oil supply chain estimate carbon footprints through the LCA approach, using existing databases and process-specific emission data which are then used to identify carbon-intensive phases.

However, these works do not include explainability, [18] or what-if analysis and recommendations [19] that can be helpful for emission reduction. In this work, we seek to address this gap by proposing Enterprise Decarbonization Accelerator, a novel framework that performs emission computation, hotspot identification with explainability insights, and what-if analysis to recommend intervenable measures for reducing carbon footprint and help enterprises accelerate their decarbonization journey.

3 DECARBONIZATION ACCELERATOR FRAMEWORK

We have designed and developed an Enterprise scale Decarbonization Accelerator Framework (as shown in Fig.1) that would be able to perform the processes associated with emission computation, hotspot identification with explainability, and what-if analysis to provide recommendations in an automated manner to accelerate decarbonization journey of enterprises. The proposed Enterprise Decarbonization Accelerator (EDA) Framework consists of an AI workflow with four modules i.e. carbon accounting, carbon hotspot identification, explainability, counterfactual queries and recommendations engine. Next, we describe the four modules.

3.1 Carbon Accounting

This module ingests sector specific (e.g. combustion of fossil fuel, application of fertilizer, fugitive leaks, etc) as well as cross-sector (e.g consumption of electricity, transportation, etc) activity data from any enterprise and leverages GHG protocol compliant Carbon Performance APIs to convert the activity data into emissions leveraging location specific emission factors. This carbon accounting module is built using Scope1 (direct emission from stationary combustion, mobile combustion, and fugitive emissions), Scope2 (usage of electricity) and Scope3 (indirect value chain emissions) APIs. For any supply chain, we convert activities into above mentioned categories and compute carbon footprint using above mentioned APIs. Here is a brief overview of APIs used to compute the carbon footprint of a supply chain.

3.1.1 Scope1:Stationary combustion. Combustion of fuels in stationary (non-transport) combustion sources results in the following greenhouse gas (GHG) emissions: carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). Sources of emissions from stationary combustion include boilers, heaters, furnaces, kilns, ovens, flares, thermal oxidizers, dryers, and any other equipment or machinery that combusts carbon bearing fuels or waste stream materials.

3.1.2 Scope1:Mobile combustion. Mobile combustion means emissions from the transportation of materials, products, waste, and employees resulting from the combustion of fuels in company owned or controlled mobile combustion sources (e.g., cars, trucks, buses, trains, airplanes, ships, etc.). The greenhouse gases CO_2, CH_4, and N_2O are emitted during the combustion of fuels in mobile sources. As per the GHG protocol, we have adopted fuel based approach for accounting carbon footprint from mobile combustion. We consider location, vehicle type, fuel type and amount as input to mobile combustion API.

3.1.3 Scope1:Fugitive emission. Fugitive emissions are leaks and other irregular releases of gases or vapors from a pressurized containment - such as appliances, storage tanks, pipelines, wells, or other pieces of equipment - mostly from industrial activities. We have leveraged sales based approach for computing carbon footprint for fugitive emissions.

While stationary, mobile, and fugitive emissions are common scope1 activities across enterprises, Scope 1 also includes industry specific activities such as applying fertilizer to agricultural fields or production of Ozone depleting gases. In such cases, our framework has flexibility to include industry specific Scope 1 carbon accounting models.

3.1.4 Scope2:Emission from electricity consumption. Scope 2 API accounts for GHG emissions from the generation of purchased electricity consumed by a company. Purchased electricity is defined as electricity that is purchased or otherwise brought into the organizational boundary of the company. Scope 2 emissions physically occur at the facility where electricity is generated and attributed to the users based on their consumption. There are two approaches for Scope 2 i.e. location based approach and market based approach. For simplicity, in this paper, we used location based approach.

3.1.5 Scope3. Scope 3 emissions are a consequence of the activities of the company, but occur from sources not owned or controlled by the company. Some examples of Scope 3 activities are extraction and production of purchased materials; transportation of purchased fuels; and use of products and service. In this paper, we used Scope 3 API to compute the carbon footprint for logistics using weight distance method.

This module also extracts carbon performance related features such as weather condition (e.g. temperature, humidity), asset specific parameters (e.g. age, size/capacity), and operational parameters (e.g. load, fuel/electricity consumption) associated with carbon performance of the asset/operation. While weather data is pulled automatically based on location information, users need to define the asset specific parameters as well as operational parameters as domain specific knowledge is required here.

3.2 Carbon Hotspot Detection

Carbon hotspot detection module ingests the carbon footprints of assets along with other relevant associated derived features and selects the best prediction and outlier detection algorithm from the library to identify the list of low carbon performing assets/operations. It mainly consists of two sub-modules, namely, (i) emission prediction model and (ii) outlier (note that we use the terms, outlier, anomalous and hotspot interchangeably) detection model as follows:

Algorithm 1 Pseudocode for Multi-dimensional Subset Scan

```
Initialize best_score, i, cur_subgroup;
cur_data_subset = Data|cur_subgroup
repeat
    1. Randomly order the given m features to scan from 1 to M
    for j = 1 to M do
        1. cur_data_subset = Data|cur_subgroup_-j
        (relax the subgroup definition to include all values of feature j)
        2. cur_subgroup = MDSS(cur_dataset)
        (Use MDSS on cur_dataset to identify the exact highest scoring
subset of values of feature j, given cur_subgroup_-j)
        3. cur_data_subset = Data|cur_subgroup
        4. best_score = score_bias(cur_dataset)
    end for
    2. Check end condition, else loop through features in random order again,
i = i + 1
until best_score has not changed between i and i - 1
```

Figure 2: MDSS Pseudocode

3.2.1 Emission prediction model. We learn the functional relationship between carbon footprint of an asset and other independent associated derived factors by selecting best prediction model from the library. We provide a library of well known ML models (such as Linear regression, Decision tree, Random Forest, Gradient boosting, etc.)

3.2.2 Outlier Detection model. In this module, we supports out of the box outlier detection module (e.g. PyOD, Isolation Forest, etc) as well as advanced method such as Multi-Dimensional Subset Scan (MDSS). We used an extension of Multi-Dimensional Subset Scan (MDSS) [20, 21] algorithm for subset scanning of the data. Fig. 2 shows the algorithm provided in [21]. This methodology is able to identify the most anomalous subgroup of feature space in linear time, amongst the exponentially many possible ones, enabling tractable subgroup analysis. The general form of the method is

$$S^* = MDSS(\mathcal{D}, \mathcal{E}, score_{bias}) \quad (1)$$

where S^* is the most anomalous subgroup, $\mathcal{D}$ is a dataset with outcomes Y and features $\mathcal{X}$, $\mathcal{E}$ are a set of expectations for Y, and $score_{bias}$ is an expectation-based scoring statistic that measures the amount of anomalousness between subgroup observations and their expectations. The goal of MDSS is to identify a subset of the data $d(S) \subseteq \mathcal{D}$ corresponding to subgroup S that maximizes $score_{bias}$. For $score_{bias}$, we use log-likelihood ratio defined as $F(S) = \log\left(\Pr(D \mid H_1(S)) / \Pr(D \mid H_0)\right)$. The alternate hypothesis $H_1(S)$ assumes that datapoints $x_i \in d(S)$ are drawn with mean $q\mu_i$ and datapoints $x_i \notin d(S)$ are drawn from mean μ_i, for constant multiplicative factor q>1 known as relative risk. The null hypothesis H_0 assumes that all datapoints, including $x_i \in d(S)$ are drawn with mean μ_i. This definition satisfies the Additive Linear-Time Subset Scanning (ALTSS) property, which is required for MDSS to be tractable [22]. Eqn. 2 gives the general $score_{bias}$ used by MDSS.

$$score_{bias}(S) = \max_{q>1} \sum_{x_i \in d(S)} \left(\log \Pr(x_i \mid q\mu_i) - \log \Pr(x_i \mid \mu_i)\right) \quad (2)$$

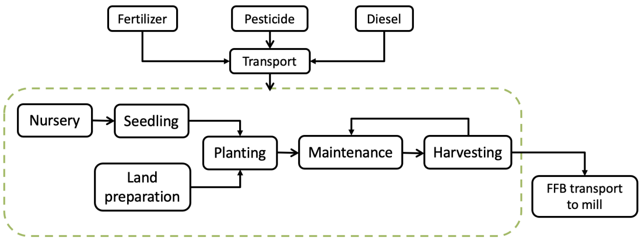

Figure 3: Palm Cultivation: Overview

3.3 Explainability

The interpretation of machine learning model is very essential to understand the feature attribution since most of the models are inherently not apparent. This module uses the well known model agnostic methodology called SHapely Additive exPlanations (SHAP) [18] to derive the insights about local and global interpretability for the carbon hotspots. The SHAP value gives the marginal contribution of each associated feature for the carbon footprint of an asset. We provide a pluggable framework where the other explainable models can be integrated into this workflow.

3.4 Counterfactual queries and recommendation engine

In this module, we provide a framework to leverage the state of the art as well as custom counterfactual queries and recommendation engines. This framework has inbuilt recommendation module which leverages the Diverse Counterfactual Explanations (DiCE) methodology from [19]. DiCE generates counterfactual explanations for any ML model through perturbations within a feasible range that change the output of a machine learning model. It also supports simple constraints on features to ensure feasibility of the generated counterfactual examples. This framework takes an input from user on the set of controllable features with its feasible range and the expected target emissions for intervenable actions, and generates the set of best recommendations.

4 CASE STUDY: PALM CULTIVATION

Sustainable sourcing of palm oil has gained significant interest over past years. It is thus important for palm producers to identify factors responsible for higher emissions, while also develop explainable intervention plans to lower their carbon footprint. We use our Enterprise Decarbonization Accelerator (EDA) Framework to estimate GHG emission of palm plantation, perform explainable hotspot identification and develop recommendation plans for reducing carbon emissions.

4.1 Overview

Fig.3 gives an overview of palm cultivation. The first step is to produce seedlings from pre-germinated seeds in nurseries, along with land preparation. Then the seedlings are planted manually and fertilizers containing elements such as nitrogen (N), potassium (K) and phosphorus (P) are applied. Fertilizers used are manufactured in factories from where they are taken to the port nearest to the plantations via shipping tankers. Then they are delivered by heavy good vehicles (HGV) such as trucks to the plantations. Along with fertilizers, diesel is also transported to the plantations, where it is used in irrigation, transportation of fertilizers and workers to the farms, land preparation and maintenance. It takes about 3-4 years for oil palms to produce fruits suitable for harvest. Palm trees continue to produce fruit for around 30 years and they are harvested periodically. Palm is harvested manually wherein fronds are cut off to dislodge fresh fruit bunches (FFB), which fall to the ground and are then collected. The fruit bunches are transported to oil mills by light good vehicles, where they are subjected to industrial processing to obtain crude palm oil (CPO) and crude palm kernel oil (CPKO), as well as by-products such as empty fruit bunches (EFB), fibers, shells and palm oil mill effluent (POME). These by-products are returned to the field as manure. Electricity is used in oil extraction phase to power machinery such as threshers and motors of conveyors.

GHG emissions from the palm supply chain can be divided into the following stages - manufacturing, agriculture, and transportation and electricity usage. Manufacturing includes all the emissions resulting from the production of fertilizers and electricity and the extraction of crude palm oil. Agriculture includes all the emissions resulting from activities related to planting and harvesting of palm produce. Transportation includes emissions resulting from transporting fertilizers and diesel to palm plantations, and FFB produce from plantations to mills. Lastly, emissions due to electricity usage includes all the indirect emissions resulting from the consumption of electricity in the supply chain.

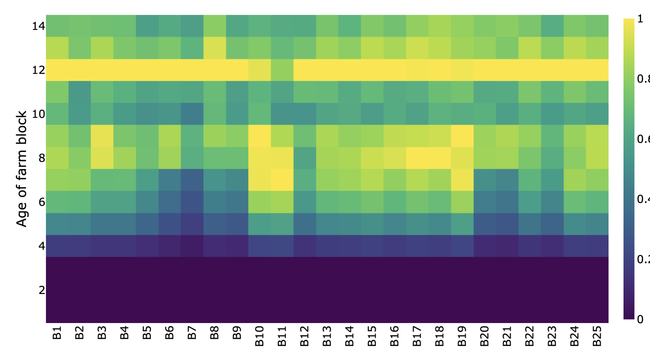

Figure 4: Heatmap showing normalised annual yield from the farm blocks at different age

4.2 Palm Plantation data

Palm plantations are divided into smaller units called blocks. We use our framework to analyze the performance of palm plantation at the block level, using data from 25 palm blocks for 14 years, from the year of plantation, upto the 14th year of cultivation of each block. We consider important factors which impact palm cultivation and captures the carbon emission performance of blocks, such as nitrogen content in the fertilizer application, carbon and nitrogen content in manure application (EFBh and pruned fronds), age of the farm block, initial soil organic carbon content at the depth of 0-15cm and 15-45cm, annual yield and the weather parameters (annual precipitation and temperature statistics). Data is obtained from a leading producer of palm oil. For the sake of anonymity, we

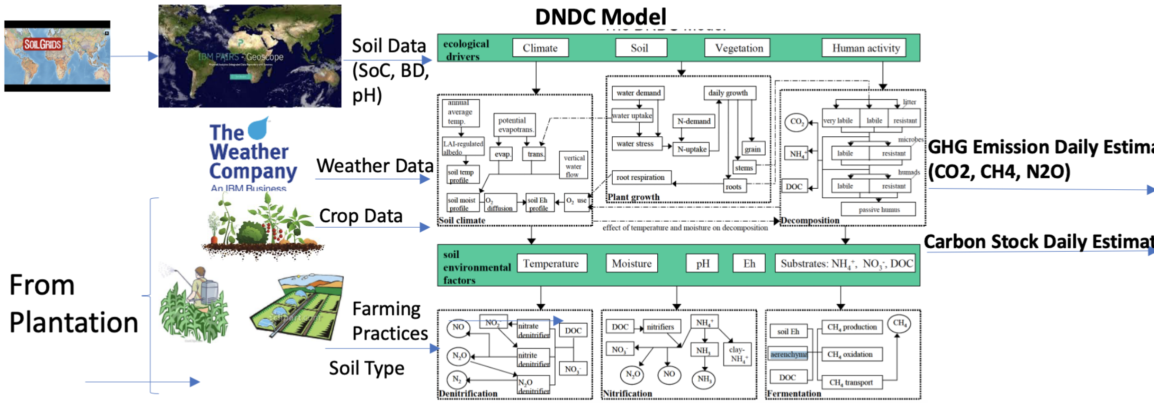

Figure 5: Denitrification-Decomposition (DNDC) model workflow [23]

will not disclose the name of the palm oil producer. Along with the farm plantation data, we also use relevant data from cradle to gate extent of palm oil supply chain for other stages - manufacturing, transportation and electricity

Fig. 4 shows the heatmap of annual yield of the 25 blocks at different ages, where yield of each block has been normalised using min-max scaling. Since palm plantations produce their first harvest in the 3rd-4th year after plantation, the yield is near zero for the first 3 years across all the blocks. We see that the farm blocks have more yield as their age increases, with maximum yield observed around the 12th year of plantation across most of the blocks.

5 EXPERIMENTATION AND RESULTS

In this section, we discuss the results that are obtained by following the methodology outlined in Sec. 3 for palm plantation blocks and draw useful insights.

5.1 Carbon Footprint of Palm Blocks

We have used the physics-based Denitrification-Decomposition (DNDC) model [23] to estimate the carbon footprint of palm plantation. The model bridges the chemical reactions in soil with the ecological drivers (climate, soil, vegetation and farming practices) and environmental factors (temperature, moisture, pH, redox potential (Eh), and simulates carbon and nitrogen dynamics. The DNDC model predicts crop growth, soil temperature and moisture, soil carbon sequestration, emission of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) along with other trace gases.

The fig. 5 shows the inputs, outputs and components of the model. The inputs to the model are the soil data (type, clay fraction, amount of initial soil organic carbon), weather data (temperature, precipitation, wind speed, solar irradiation), crop data, farming practices (fertilizer, manure and irrigation schedules, tillage information), planting and harvesting schedules. The two-component DNDC model has six submodules, namely, soil climate, crop growth, denitrification and decomposition, nitrification and fermentation.

Our framework computes carbon footprint at each stage from cradle to gate, i.e. raw material extraction to CPO production leveraging the APIs mentioned in Sec. 3.1 and the DNDC model. Emissions related to activities within the organizational boundary were

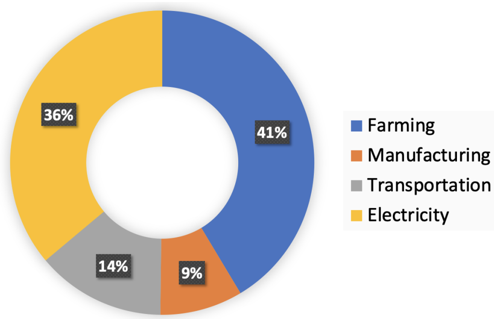

Figure 6: Stage-wise distribution of carbon emission

computed using Scope1, Scope2 APIs and DNDC model. Upstream emissions (i.e. raw material extraction and manufacturing of fertilizer) and emissions related to transportation of upstream raw material and fertilizer were computed using the Scope3 logistic API. At the farming stage, DNDC tool has been used to precisely compute the carbon footprint due to farming activities. The framework uses "Carbon Dioxide Equivalent" or "CO_2e" as the standard unit for measuring carbon footprint. For any quantity and type of greenhouse gas, CO_2e signifies the amount of CO_2 which would have the equivalent global warming impact. A quantity of GHG can be expressed as CO_2e by multiplying the amount of the GHG by its Global Warming Potential (GWP), where GWP is the heat absorbed by any greenhouse gas in the atmosphere, as a multiple of the heat that would be absorbed by the same mass of carbon dioxide (CO_2). For example, the GWP value for N_20 is provided as 265 in IPCC Fifth Assessment Report, 2014 (AR5) [24, 25]. Therefore, if 1kg of N_2O is emitted, this can be expressed as 265kg of CO_2e (1kg N_2O * 265 = 265kg CO_2e).

The emissions were divided into four stages - farming, manufacturing, transportation and electricity as mentioned in Sec. 4.1. Fig. 6 shows the stage-wise distribution of carbon emissions in the cradle to gate palm oil supply chain, averaged across the 14 years of plantation data. The plot shows that the farming stage is the major source of emission in the supply chain. Therefore, in this paper, we will focus on the farming stage, and use our framework for block level analysis of palm plantation.

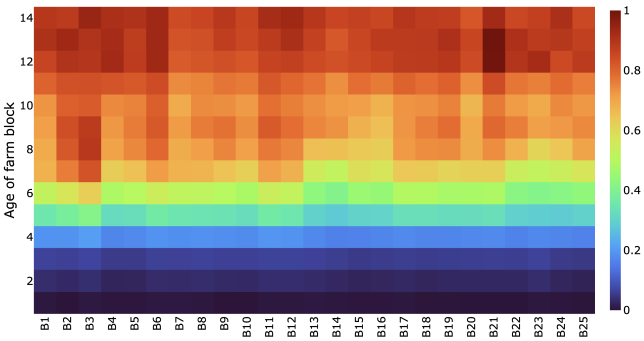

Figure 7: Normalized annual equivalent carbon emission (CO_2e) from farm blocks at different age

The DNDC model gives the emission of GHG gases such as CO_2, CH_4 and N_2O from the farming stage at daily intervals. For our analysis, we have aggregated the GHG emission to annual temporal resolution. Since all the palm plantation blocks are located in the upland mineral soil, the methane emission from the farming stage is very negligible. We have considered only the soil CO_2 emission from decomposition process and N_2O emission from nitrification and denitrification process to compute the total carbon footprints (CO_2e).

Fig. 7 shows the heatmap of annual CO_2e emission of the 25 palm plantation blocks at different ages, where the emission numbers are normalized across all the blocks using min-max scaling. All 25 blocks receives an equal amount of fertilizers for the first four years and for subsequent years, the amount of fertilizers and manures are determined based on the age and soil testing. We can see that the annual CO_2e emission are low and almost in similar range for the first four years and starts to show higher value as the age increases across all the blocks.

5.2 Block level Hotspot Identification

The annual carbon footprint of the farm blocks along with the plantation data detailed in Sec. 4.2 is used for hotspot identification. We first use the emission prediction module to model the block level carbon footprint of palm blocks using Random Forest Regressor.

The predicted carbon emissions serve as the expected emissions $\mathcal{E}$, to be used by MDSS based outlier detection module. We use Gaussian scoring function [22] as $score_{bias}$, which uses Gaussian distribution to model the log likelihood ratio statistic defined in equation 2. This serves as the statistical measure of divergence between subgroup observations and their expectations. Gaussian scoring function for a subgroup S is given by [22]:

$$score_{bias}(S) = \max_{q>1} \sum_{x_i \in d(S)} y_i \hat{\mu} \frac{(q-1)}{\hat{\sigma}^2} + \sum_{x_i \in d(S)} \hat{\mu}^2 \left(\frac{1-q^2}{2\hat{\sigma}^2} \right) \quad (3)$$

where, y_i are the observed values in the subset $d(S)$ belonging to subgroup S, $\hat{\mu}$, and $\hat{\sigma}^2$ are the expected mean and variance of the subgroup. We perform MDSS based outlier detection, as outlined in 3.2.2 to obtain hotspot farm blocks, with $score_{bias}$ as defined in equation 3. B2 and B3 are identified as hotspot blocks across multiple years. In Fig. 8 the red region depicts feature space of the anomalous subgroup S^* identified by MDSS. From this, we observe that blocks with very low or high fertilizer application, high manure application and moderate yield are identified as hotspots. However, this does not provide explainable insights or help with identifying opportunities for reducing emissions. To address this, we will next analyze the hotspot blocks using explainability and what-if scenario.

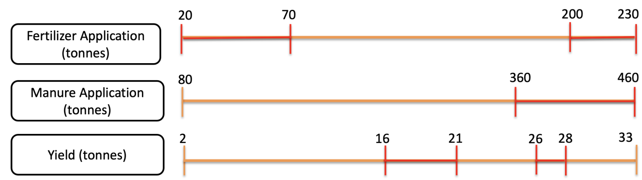

Figure 8: Anomalous subgroup identified by MDSS

5.3 Hotspot Explainability Analysis

To understand the factors that dictates the variation in carbon footprints from palm plantations, we have used SHAP value to identify the marginal contribution of all the relevant features. We begin with learning the prediction model with farming practices, soil properties and weather related parameters as input features for predicting the soil CO_2 and N_2O emission independently. The individual emission regressor model will give us more insights about the dominating features that influences the respective emissions. Figs. 9 and 10 shows the holistic summary of feature importance for CO_2 and N_2O emission regressor model respectively using SHAP value.

From fig. 9, we observe that for soil CO_2 emissions, the most dominating features are plantation age, the amount of nitrogen content in the manure and fertilizers, annual yield, annual precipitation and the initial soil organic carbon (SOC) at the depth of 15cm-45cm and 0-15cm. Since the CO_2 emission is the product of soil decomposition and it is clear that the above factors are influencing the rate of decomposition process. The pruned palm fronds and the EFBs are the main source of manure applications and it has high carbon to nitrogen ratio. Because of high carbon content, the manure takes precedence over fertilizer in the feature importance plot. The amount of manure and fertilizers tends to increase with age and thus the increase in annual yield. For SOC at the depth of 0-15cm and 15cm-45cm, we see the modest impact with a higher value leading to increase in CO_2 while the lower value having neutral impact. This can be attributed to the capacity of soil to hold more carbon without getting saturated. The higher SOC content in the soil tends to have low holding capacity and thus more free carbon to participate in the decomposition process.

Similarly from fig. 10, we see that for N_2O emission, the dominating features are same as of CO_2 emission but the order of precedence and the impact are quite different. The nitrogen content in fertilizer takes the precedence followed by annual precipitation and then the manure application. The N_2O emission is the product of nitrification and denitrification process as shown in fig. 5. It is evident from the DNDC model workflow, the presence of nitrate and ammonium ion with high precipitation tends to be a conducive environment for N_2O emission.

Since the dominating features of soil CO_2 and N_2O emissions are similar, it will be appropriate to investigate the feature importance of total carbon footprint (CO_2e). This will also capture the interdependency among soil CO_2 and N_2O emissions and helps the enterprise to focus on one metric for the overall carbon footprint reduction. Fig. 11 shows the global interpretation of feature importance for the CO_2e emission regressor model using Shapley value. The order of feature importance is same as that of soil CO_2 and it shows that the proportion of soil CO_2 is larger than N_2O. Based on the features SHAP value, we infer that in general, following factors lead to an increase in CO_2e - (i) high fertilizer and manure application results in higher emission. (ii) older blocks tend to have a higher carbon footprint associated with them. (iii) Soil organic carbon at depth 0-15cm (SOC_45) and 15-45cm (SOC_15) has a modest impact, with high soil organic carbon leading to increase in carbon impact, while low SOC having neutral impact.

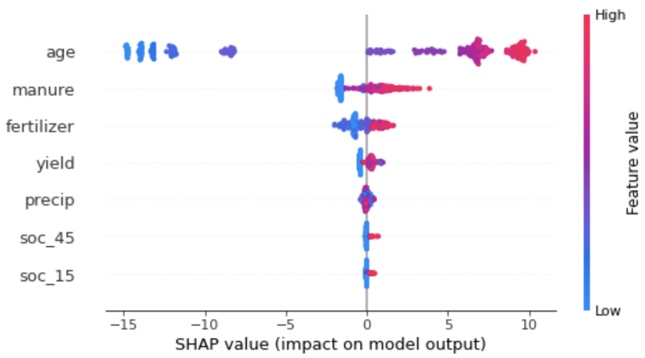

Figure 9: Global Explainability for CO_2 emission

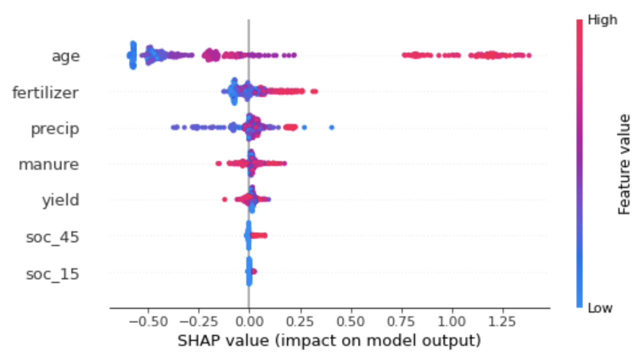

Figure 10: Global Explainability for N_2O emission

We also perform explainability analysis of the identified hotspot blocks to identify factors behind low performance of the blocks. We use local explainability SHAP plots as mentioned in Sec. 3.3. The plot shows which factors contribute to the increase or decrease in CO_2e of the hotspot blocks, when compared to the baseline of average carbon footprint. Fig. 12 and Fig. 13 provide the plots for two hotspot blocks.

From Fig. 12, we observe that high SOC_15 and SOC_45 content, high manure application is resulting in increase in emissions, while low fertilizer application is lowering the carbon footprint. The block also has a moderate yield which could be due to low fertilizer application. From Fig. 13, we can infer that very high fertilizer and manure application are responsible for high emissions. The block has a good yield performance, however, it has a low carbon performance when compared to other blocks. The observations are in accordance with the insights drawn from Fig. 11 earlier.

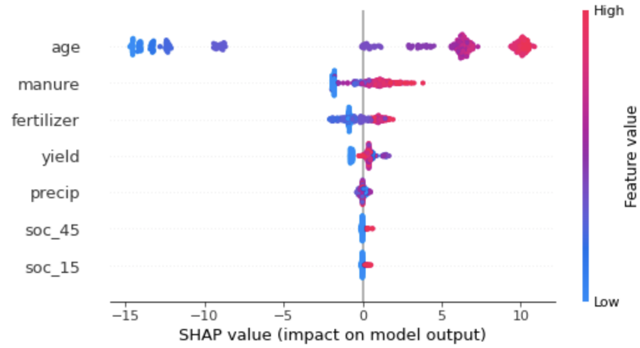

Figure 11: Global Explainability for CO_2e emission

The two explainability plots share some common factors, we can observe that high fertilizer and high manure application negatively impact the carbon performance of the farm blocks.

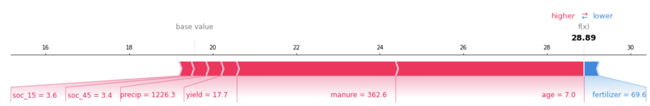

Figure 12: Palm Plantation block B2 - Local Explainability

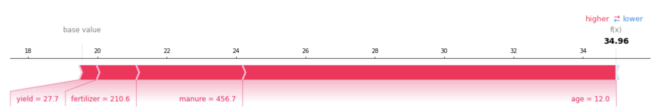

Figure 13: Palm Plantation block B3 - Local Explainability

5.4 Counterfactual Recommendation

Recommendation Engine introduced in Sec. 3.4 is used to generate counterfactual recommendation plans for the hotspot blocks to reduce their carbon footprint. We focused on finding the set of intervenable features which can be perturbed to reduce the CO_2e emission without compromising on the yield produced by palm plantations. Figs. 14 and 15 shows the intervention plans generated by the engine for the two hotspot blocks. We observe that for both the blocks, the fertilizer and manure application are identified as the intervenable areas. For B2, we generated the counterfactual query to produce a set of three best diverse recommendation to restrict the CO2e emission within the range of 10 to 25 tonnes of CO_2e. From fig. 14, we observe that reducing fertilizer application by around 10% and manure application by around 30% can help in reducing the carbon emission by roughly 20%, with negligible impact on the yield. Similarly, for B3, we generated the counterfactual query to produce the diverse set of four best recommendation without impacting the yield. From fig. 15, we see that reducing manure application by 25%-30% can reduce the carbon emission by around

10% with negligible impact on the yield for B3 palm plantation block.

fertilizer	manure	precip	age	soc_15	soc_45	yield	CO2e
69.56	362.58	1226.3	7.0	3.55	3.35	17.721434	28.891819

Diverse counterfactual set - CO_2e : [10,25]

fertilizer	manure	precip	age	soc_15	soc_45	yield	CO2e
63.3	248.2	-	-	-	-	17.5	23.27286577079987
58.0	266.5	-	-	-	-	18.4	24.36235165660021
63.5	241.7	-	-	-	-	17.1	22.794548305000127

Figure 14: B2 Counterfactual recommendations

fertilizer	manure	precip	age	soc_15	soc_45	yield	CO2e
210.58	456.69	1074.6	12.0	1.01	1.03	27.698867	34.865427

Diverse counterfactual set - CO_2e : [10,32]

fertilizer	manure	precip	age	soc_15	soc_45	yield	CO2e
213.3	326.9	-	-	-	-	27.8	31.083256241399727
194.1	368.6	-	-	-	-	28.6	31.1429671037998
192.9	333.6	-	-	-	-	28.5	30.29656883919958
216.5	321.8	-	-	-	-	26.5	31.29766918619998

Figure 15: B3 Counterfactual recommendations

6 CONCLUSION AND FUTURE WORK

In this paper, we presented an unified novel framework called Enterprise Decarbonization Accelerator (EDA) to accurately estimate the carbon footprint at enterprise or process level across all types of asset classes, identify GHG hotspots and explain the factors that create the hotspots and carry out what-if analysis to reduce the carbon footprint. The efficacy of the framework is demonstrated with a palm oil enterprise data. Results presented in this paper indicate that agriculture stage is the most carbon intensive and blocks in which high fertilizer amounts were applied and low yields were obtained are the hot spots. This enabled customization of farming practices such as right amount of fertilizer or manure for low performing blocks which would not only improve their yield but also reduce their carbon footprint and ensures profitability in a sustainable manner. The proposed framework is generic, industry agnostic, and can be used seamlessly across enterprises (IT, oilgas, Energy Utility, etc). While this framework provides short term operational recommendations, in future, we plan to extend this framework for long term strategic recommendations to accelerate net-zero decarbonization journey of enterprises.

REFERENCES

[1] WEF, "Net-Zero Challenge: The supply chain opportunity ," 2021.
[2] GEI, "Supply Chain Carbon and Energy Footprint ," 2016.
[3] R. Black, K. Cullen, B. Fay, T. Hale, J. Lang, S. Mahmood, and S. Smith, "Taking stock: A global assessment of net zero targets, energy climate intelligence unit and oxford net zero," 2021.
[4] M. Kuhndt *et al.*, "Hot spot analysis in practice–a case study focusing on mnc, confidential report," 2002.
[5] A. Rudnicka, T. B. Kalinowski, and G. Wieteska, "Co2 hotspots identification in supply chains of different products," *Studia Ekonomiczne*, vol. 321, 01 2017.
[6] K. Bienge, J. v. Geibler, M. Lettenmeier, B. Biermann, O. Adria, and M. Kuhndt, "Sustainability hot spot analysis: A streamlined life cycle assessment towards sustainable food chains," in *Conference proceedings of the 9th European International Farming System Association Symposium*, pp. 4–7, 2010.
[7] C. B. Norris, G. A. Norris, and D. Aulisio, "Efficient assessment of social hotspots in the supply chains of 100 product categories using the social hotspots database," *Sustainability*, 2014.
[8] X. Guo, J. Broeze, J. J. Groot, H. Axmann, and M. Vollebregt, "A worldwide hotspot analysis on food loss and waste, associated greenhouse gas emissions, and protein losses," *Sustainability*, vol. 12, no. 18, p. 7488, 2020.
[9] C. Liedtke, C. Baedeker, S. Kolberg, *et al.*, "Resource intensity in global food chains: the hot spot analysis," *British Food Journal*, 2010.
[10] A. A. Acquaye, T. Wiedmann, K. Feng, R. H. Crawford, J. Barrett, J. Kuylenstierna, A. P. Duffy, S. L. Koh, and S. McQueen-Mason, "Identification of 'carbon hot-spots' and quantification of ghg intensities in the biodiesel supply chain using hybrid lca and structural path analysis," *Environmental science & technology*, vol. 45, no. 6, pp. 2471–2478, 2011.
[11] G. Piringer, A. Bauer, A. Gronauer, M. K. SAYLOR, A. Stampfel, and K. Iris, "Environmental hot spot analysis in agricultural lifecycle assessments–three case studies," *Journal of Central European Agriculture*, 2016.
[12] A. Singh, N. Mishra, S. I. Ali, N. Shukla, and R. Shankar, "Cloud computing technology: Reducing carbon footprint in beef supply chain," *International Journal of Production Economics*, vol. 164, pp. 462–471, 2015.
[13] M. A. Thomassen, R. Dalgaard, R. Heijungs, and I. De Boer, "Attributional and consequential lca of milk production," *The International Journal of Life Cycle Assessment*, vol. 13, no. 4, pp. 339–349, 2008.
[14] T. O. Rodrigues, A. Caldeira-Pires, S. Luz, and C. A. Frate, "Ghg balance of crude palm oil for biodiesel production in the northern region of brazil," *Renewable Energy*, vol. 62, pp. 516–521, 2014.
[15] V. Subramaniam, Z. Hashim, S. K. Loh, and A. A. Astimar, "Assessing water footprint for the oil palm supply chain-a cradle to gate study," *Agricultural Water Management*, vol. 237, p. 106184, 2020.
[16] H. Muhamad, I. B. Sahid, S. Surif, T. Y. Ai, and C. Y. May, "A gate-to-gate case study of the life cycle assessment of an oil palm seedling," *Tropical life sciences research*, vol. 23, no. 1, p. 15, 2012.
[17] F. Arshad, Y. A. Tan, and S. Yusoff, "A cradle-to-gate study of ghg emissions from the transportation of palm oil, palm olein and palm stearin using the life cycle assessment approach," *Journal of Oil Palm Research*, vol. 29, no. 1, pp. 120–129, 2017.
[18] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," NIPS'17, (Red Hook, NY, USA), Curran Associates Inc., 2017.
[19] R. K. Mothilal, A. Sharma, and C. Tan, "Explaining machine learning classifiers through diverse counterfactual explanations," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 607–617, 2020.
[20] D. B. Neill, E. McFowland III, and H. Zheng, "Fast subset scan for multivariate event detection," *Statistics in medicine*, vol. 32, no. 13, pp. 2185–2208, 2013.
[21] Z. Zhang and D. B. Neill, "Identifying significant predictive bias in classifiers," *arXiv preprint arXiv:1611.08292*, 2016.
[22] S. Speakman, S. Somanchi, E. McFowland III, and D. B. Neill, "Penalized fast subset scanning," *Journal of Computational and Graphical Statistics*, vol. 25, no. 2, pp. 382–404, 2016.
[23] S. L. Gilhespy, S. Anthony, L. Cardenas, D. Chadwick, A. del Prado, C. Li, T. Misselbrook, R. M. Rees, W. Salas, A. Sanz-Cobena, P. Smith, E. L. Tilston, C. F. Topp, S. Vetter, and J. B. Yeluripati, "First 20 years of dndc (denitrification decomposition): Model evolution," *Ecological Modelling*, vol. 292, pp. 51–62, 2014.
[24] G. Myhre, D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, *et al.*, "Climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change," *K., Tignor, M., Allen, SK, Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, PM, Cambridge University Press Cambridge, United Kingdom and New York, NY, USA*, 2013.
[25] G. Myhre, D. Shindell, and J. Pongratz, "Anthropogenic and natural radiative forcing," 2014.