Shell .ai Hackathon for Sustainable and Affordable Energy

Problem Statement

powered by

ORTEC UDACITY STREAMBA

NIRMAAN

shell.com/hackathon

Shell.ai Hackathon for Sustainable and Affordable Energy

Agricultural Waste Challenge: Problem Statement

Introduction

Welcome to the fourth edition of the Shell.ai Hackathon for Sustainable and Affordable Energy, a platform that brings together brilliant minds passionate about digital solutions and AI, to tackle real energy challenges and help build a lower-carbon world. In the previous three editions, we addressed some of the digital challenges around energy transition, such as windfarm layout optimisation (2020), irradiance forecasting for solar power generation (2021) and optimal placement of electric vehicle (EV) charging stations (2022). This year, we turn our attention to another challenge that, if addressed, has the potential to lower emissions in sectors that are particularly hard to decarbonise, like Aviation and Marine, and applications closer home, like cooking. Sustainable Aviation Fuel, Renewable Diesel, and Renewable Natural Gas bring the advantage of lowering the emissions of these sectors without the need to alter our existing flights, ships, homes, and the supply infrastructure.

As humanity, why aren't we using them at scale yet? Firstly, they are expensive and need ambitious government targets and substantial subsidies to be able to compete with gasoline, diesel, and conventional natural gas.[1] And even with targets and price subsidies, biofuels present another challenge. Today, 100 tons of crude oil can provide approximately 75 tons of gasoline and diesel.[2] But to get the same 75 tons of biofuel, a biorefinery needs at least 375 tons of biomass.[3] And unlike crude oil extracted from a well, these 375 tons will not be available at a single source location. Instead, they will be thinly distributed across hundreds of agricultural lands spread amongst multiple geographies.

To set up a biorefinery in a region, an understanding of the region's current and future biomass produce will be required. This biomass needs to be collected and transported to intermediate depots

1. https://www.shell.com/content/shell/corporate/global/en_gb/energy-and-innovation/the-energy-future/decarbonising-aviation/_jcr_content/root/main/section_copy/promo/links/item0.stream/1667916358181/e4f516f8d0b02333f1459e60dc4ff7fd1650f51c/decarbonising-aviation-industry-report.pdf

   https://www.mckinsey.com/~/media/mckinsey/industries/travel%20transport%20and%20logistics/our%20insights/scaling%20sustainable%20aviation%20fuel%20today%20for%20clean%20skies%20tomorrow/clean-skies-for-tomorrow.pdf

2. https://www.eia.gov/tools/faqs/faq.php?id=327&t=6#:~:text=Petroleum%20refineries%20in%20the%20United,gallon%20barrel%20of%20crude%20oil.

3. https://www.mckinsey.com/~/media/mckinsey/industries/travel transport and logistics/our insights/scaling sustainable aviation fuel today for clean skies tomorrow/clean-skies-for-tomorrow.pdf

for de-moisturisation and densification into pellets. The pellets will then need to be transported to the biorefinery for conversion to biofuel. This incurs high cost of feedstock transportation and associated GHG emissions, which will need to be minimised too.[1] The value it generates lies not only in contribution to the global energy transition, but also benefits farmers as a sustainable source of income.

In this hackathon, we challenge you to form teams, brainstorm ideas, and build digital solutions that can design and optimise this new, complex, and strategic supply chain for biorefineries of the future.

Problem Statement

Let's try to understand the Waste-to-Energy problem using the map shown in Figure 1. It corresponds to the yearly residual biomass available in the state of Gujarat in Western India. The geographic region is divided into equisised grid blocks, and the average biomass production of each grid block is represented on the color scale. The spatial distribution will change year after year, and the biomass available in these grid blocks will serve as feedstock to biofuel production plants i.e., biorefineries.

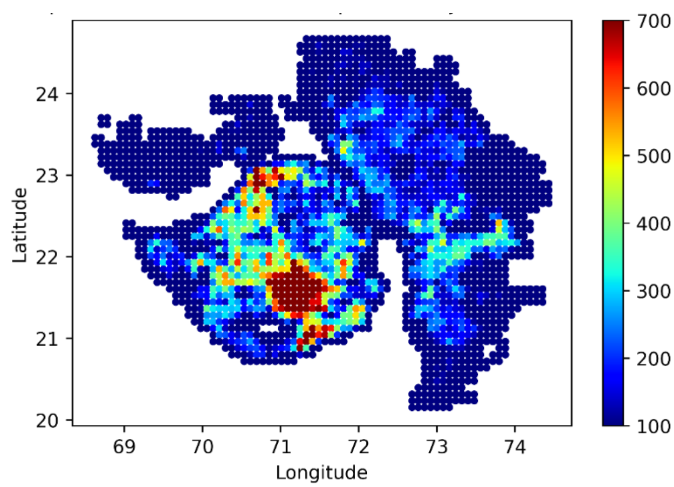

Figure 1: Spatial residual biomass production

The biomass feedstock needs to be optimally collected and transported to the biorefineries' gate, and the grid blocks in Figure 1 i.e., the *Harvesting Sites* serve as the starting point in the supply chain. At each *Harvesting Site*, biomass availability can be forecasted using historical data. External parameters such as weather, rainfall, government regulation etc. may also influence the biomass production. Bales of biomass will be collected from these *Harvesting Sites* and transported to the pre-processing *Depots* for densification and pelletisation. The pellets will then be transported from each of these *Depots* to each *Biorefinery*, in quantities that cater to each *Biorefinery's* demand. The locations of the *Depot* and the *Biorefinery* will remain constant. The schematic of this supply chain is shown in Figure 2.

1. https://www.sciencedirect.com/science/article/pii/S0306261918302022

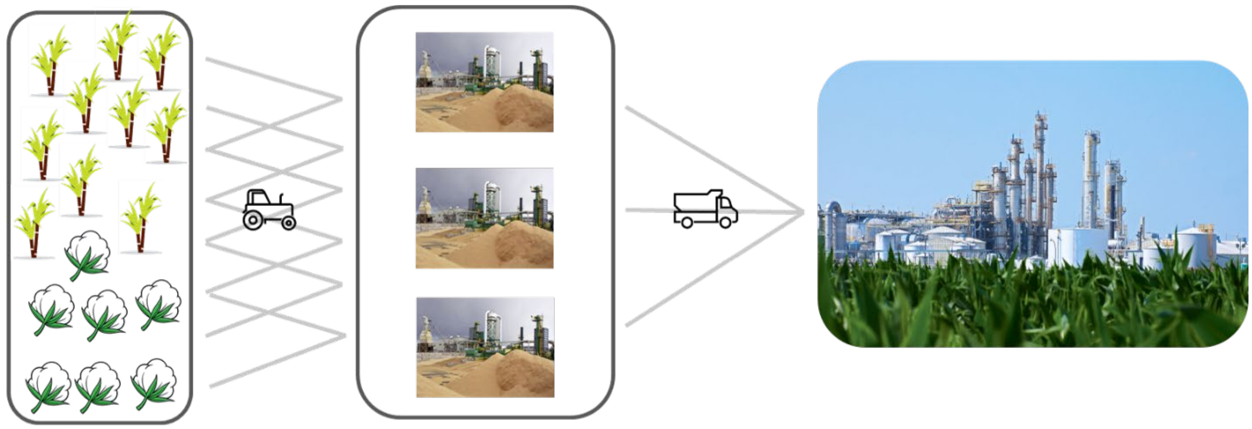

Figure 2: Feedstock-to-biorefinery supply chain schematic

Your solution should accurately forecast the spatial distribution of biomass availability in the given region and select optimal location for all the assets in the supply chain i.e., preprocessing *Depots* and *Biorefineries*, such that the supply chain is robust, while satisfying practical objectives and constraints.

Data

We are providing the following datasets:

1. Biomass_History.csv: A time-series of biomass availability in the state of Gujarat from year 2010 to 2017. We have considered arable land as a map of 2418 equisized grid blocks (*harvesting sites*). For ease of use, we have flattened the map and provided location index, latitude, longitude, and year wise biomass availability for each *Harvesting Site*.

Index	Latitude	Longitude	2010	:	2017
0	24.66818	71.33144	8.475744	:	5.180296
1	24.66818	71.41106	24.02978	:	42.12695
:	:	:	:	:	:
2417	20.15456	73.16282	0.621228	:	0.226953

2. Distance_Matrix.csv: The travel distance from source grid block to destination grid block, provided as a 2418 x 2418 matrix. Note that this is *not* a symmetric matrix due to U-turns, one-ways etc. that may result into different distances for 'to' and 'from' journey between source and destination.

Distance Matrix		Destination Index					
		0	1	2	3	:	2417
Source Index	0	0	11.3769	20.4557	38.1227	:	681.4235
	1	11.3769	0	9.0788	28.9141	:	679.1758
	2	20.4557	9.0788	0	22.3791	:	679.7786
	3	38.1227	28.9141	22.3791	0	:	678.969
	:	:	:	:	:	:	:
	2417	679.2328	676.9851	677.5878	677.9406	:	0

Using the above data, you are required to forecast biomass availability as well as design the optimal supply chain for the years 2018 and 2019, following the objective and constraints given in the subsequent sections.

Note: All quantities/values provided in the above datasets are dimensionless. However, the data set has all the characteristics of a practical use case.

Notations and Constants

i = Index of the harvesting site. i within 0 to 2417.
j = Index of the depot. j within 0 to 2417.
k = Index of the biorefinery. k within 0 to 2417.
$Dist_{i,j}$ = Distance from the i^{th} harvesting site to the j^{th} depot
$Dist_{j,k}$ = Distance from the j^{th} depot to the k^{th} biorefinery
$Biomass_{i,j}$ = Biomass demand-supply matrix i.e. the amount of biomass procured and transported from the i^{th} harvesting site to the j^{th} depot
$Pellet_{j,k}$ = Pellet demand-supply matrix i.e. the amount of pellets transported from the j^{th} depot to the k^{th} biorefinery
$Biomass_{forecast,i}$ = The forecasted amount of biomass available at the i^{th} harvesting site
$Biomass_{true,i}$ = The true amount of biomass available at the i^{th} harvesting site
Cap_{depot} = Maximum yearly processing capacity of a depot (20,000)
$Cap_{refinery}$ = Maximum yearly processing capacity of a biorefinery (100,000)

Objective

The objective of the problem is to minimize the overall cost which comprises of 3 components:

1. Cost of transportation ($Cost_{transport}$): This cost depends on the quantity of biomass/pellets being transported and the respective distance.

$$Cost_{transport} = \left(\sum_{i,j} Dist_{i,j} \times Biomass_{i,j}\right) + \left(\sum_{j,k} Dist_{j,k} \times Pellet_{j,k}\right)$$

2. Cost of biomass forecast mismatch ($Cost_{forecast}$): This cost depends on how far the forecasted biomass is to the actual biomass availability for every harvesting site.

$$Cost_{forecast} = \sum_{i} abs(Biomass_{forecast,i} - Biomass_{true,\,i})$$

3. Cost of underutilization ($Cost_{underutilization}$): This cost depends on whether each asset i.e., the depots and the refineries are utilized to their maximum operating capacity

$$Cost_{underutilization} = \sum_{j} (Cap_{depot} - \sum_{i} Biomass_{i,j}) + \sum_{k} (Cap_{refinery} - \sum_{j} Pellet_{j,k})$$

So overall cost ($Cost$) $= \text{a} \times Cost_{transport} + b \times Cost_{forecast} + \text{c} \times Cost_{underutilization}$, where a = 0.001, b = 1, c = 1 are constants.

Constraints

You are required to solve the problem under the following practical constraints:

1. All values (forecasted biomass, biomass demand-supply, pellet demand-supply) must be greater than or equal to zero.
2. The amount of biomass procured for processing from each harvesting site $'i'$ must be less than or equal to that site's forecasted biomass.
3. Total biomass reaching each preprocessing depot $'j'$ must be less than or equal to its yearly processing capacity (20,000).
4. Total pellets reaching each refinery $'k'$ must be less than or equal to its yearly processing capacity (100,000).
5. Number of depots should be less than or equal to 25.
6. Number of refineries should be less than or equal to 5.

7. At least 80% of the total forecasted biomass must be processed by refineries each year.
8. Total amount of biomass entering each preprocessing depot is equal to the total amount of pellets exiting that depot (*within tolerance limit of 1e-03*).

Evaluation

You are required to submit one file (solution.csv) on the HackerEarth portal for evaluation. This file will comprise of your solution with the following columns in .csv format,

Column Names	Valid Entries
year	2018, 2019, 20182019
data_type	depot_location, refinery_location, biomass_forecast, biomass_demand_supply, pellet_demand_supply
source_index	Within 0 to 2417
destination_index	Within 0 to 2417
value	Any value following problem constraints

Notes

- We have provided a sample solution.csv file for your reference. Your submission must follow the same format.
- If you don't provide values for all valid indices, a default value i.e. zero will be considered as the value for those indices. This may result into constraint violation.
- For *data_type* depot_location and refinery_location, entries under *destination_index* and *value* are not required and will be disregarded. You may choose to keep these entries blank.
- For *data_type* biomass_forecast, entries under *destination_index* are not required and will be disregarded. You may choose to keep these entries blank.
- Optimized supply chain infrastructure proposed in your solution must be the same for both year 2018 and 2019. Consequently, the entries for *data_type* depot_location and refinery_location will be agnostic to year. Use *year* = 20182019 for such common entries.
- You can only place one depot per grid block/location. Similarly, You can only place one biorefinery per grid block/location.
- Your solution will be eligible for ranking only if it satisfies all the constraints for 2018 and 2019.
- You do not need to submit your source code files. When you submit your solution, you can ignore the "Upload source code" field.

- We will keep the first year (2018) of your solution for the public leaderboard. You can test your solution any time and see how it ranks.
- We will keep the second year (2019) of your solution for the private leaderboard and it will be used to determine the finalists.

Scoring

If your solution satisfies all constraints, we will first calculate your solution's cost using the overall cost function. Your cost (the lower the better) will then be converted to a score (the higher the better) between 20 to 100 using the following transformation function for the leaderboard ranking:

$$Leaderboard\ Score = \max\left[20, \left(100 - \frac{80 \times cost}{500000}\right)\right]$$

Scores between 0 to 18 are reserved for the error codes detailed below.

Error Codes:

If your solution doesn't satisfy any of the constraints or submission format, you will get the following error codes on the leaderboard:

0 → Format error: solution.csv not following the correct format. Check sample solution file for reference
1 → Constraint 1 violated
2 → Constraint 2 violated
3 → Constraint 3 violated
4 → Constraint 4 violated
5 → Constraint 5 violated
6 → Constraint 6 violated
7 → Constraint 7 violated
8 → Constraint 8 violated
9 → Index error: Harvesting site location index i should be an integer value between 0 and 2417
10 → Index error: Depot location index j must be an integer value between 0 and 2417
11 → Index error: Biorefinery location index k must be an integer value between 0 and 2417
12 → Index error: Harvesting site location index i out of bound in biomass demand-supply matrix
13 → Index error: Depot location index j out of bound in biomass demand-supply matrix
14 → Index error: Depot location index j out of bound in pellet demand-supply matrix
15 → Index error: Biorefinery location index k out of bound in pellet demand-supply matrix

16 → Index error: You can only specify one value of biomass forecast per location. Multiple found.

17 → Index error: You can only place one depot per location. Multiple found.

18 → Index error: You can only place one biorefinery per location. Multiple found.