

# PROJECT PROPOSAL



<b>Completion Date:31/10/2021</b>
<b>Version:2.0</b>
<b>State: In Complete</b>
<b>Author: Osuntuyi Michael</b>

## Table of Contents

<b>Domain understanding</b> .....	4
Project description .....	4
AI Functionality and Features .....	4
<b>Impact assessment</b> .....	5
What problem does this technology want to solve: .....	5
Can you motivate why you are sure that this technology is solving the right problem: .....	5
In what way is this technology contributing to a world you want to live in: .....	5
<b>Stakeholders: Have you considered all stakeholders?</b> .....	6
What are the main users/target groups/stakeholders for this technology? .....	6
Stakeholders Involved .....	6
How is this stakeholder affected? .....	6
Did you consult the stakeholder? .....	6
Are you going to take this stakeholder into account? .....	6
Did you really consider all stakeholders, even the ones that might not be a user or target group, but still might be of interest? .....	6
<b>Privacy: Are you considering the privacy &amp; personal data of the users of your technology?</b> .....	7
Does this technology register personal data? If yes, what personal data? .....	7
Do you think this technology invades someone's privacy? If yes, in what way? .....	7
Do you think this technology is compliant with prevailing privacy and data protection law and can you motivate why? .....	7
In which way can you imagine a future impact of the collection of personal data? .....	7
<b>Data: Is data in your technology properly used?</b> .....	8
Are you aware of the limitations and subjectivity of data and is this reflected in this technology? ..	8
<b>Transparency: Are you transparent about how your technology works?</b> .....	8
(How) is explained to the users how a technology works and how the business model works? .....	8
Is it possible to file a complaint or ask questions/get answers about this technology? .....	8
Is the technology (company) clear about possible negative consequences or shortcomings of this technology? .....	8
<b>Inclusivity: Is your technology fair for everyone?</b> .....	9
Will everyone have access to this technology? .....	9
Does this technology have a built-in bias? .....	9
Does this technology make automatic decisions and how do you account for them? .....	9
Is everyone benefitting from this technology or only a small group? Do you see this as a problem? Why/why not? .....	9
Does the team that creates the technology represents the diversity of our society? .....	9
<b>Hateful and criminal actors: What can bad actors do with your technology?</b> .....	9

In which way can this technology be used to break the law or avoid the consequences of breaking the law?.....	9
<b>Human values How does the technology affect your human values? .....</b>	<b>10</b>
How does your technology affect the identity of the user(s)? .....	10
How does the technology influence the user(s) autonomy?.....	10
What is the effect of the technology on the health and/or wellbeing of the user(s)?.....	10
<b>Future Did you consider future impact?.....</b>	<b>11</b>
What could possibly happen with this technology in the future?.....	11
Utopian scenario: .....	11
Dystopian scenario.....	11
Would you like to live in one of these scenarios? Why? Why not?.....	11
What happens if your technology (which you have thought of as ethically well-considered) is bought or taken over by another party? .....	12
<b>Modelling.....</b>	<b>18</b>
Is the choice of model(s) appropriate for the target variable c.q. the problem at hand? .....	18
Should several models be compared? Why yes/no? .....	18
<b>Evaluation and deployment.....</b>	<b>18</b>
Is there a plan for domain knowledge verification? Who can say whether the model works well or not? .....	18
Questions about Deployment.....	18
What are acceptable evaluation results? (Think about accuracy, precision, recall, false positives/negatives etc.) .....	18
Will the model be deployed? If so, how? .....	18

## Domain understanding

### Project description

The project I have decided to work on is structured towards trying to help farmers especially small and upcoming ones in India have better opportunities in taking up agriculture.

According to (Devenyns, 2021) consumers around the world are swaying towards organic produce due to the current pandemic (Covid19) and the hopes of living healthier lives, but further research has shown that demand of organic produce surpasses the supply (Faber, 2020).

A factor which aids in the low supply of organic produce is, due to the reduction of individuals who want to take farming as an occupation, people in rural areas which have more opportunity in this field end up forgoing this for a life in the city areas.

### Project Goal

The goal of this project is to make an AI/Machine learning model application which can predict the total amount of crops that will be produced, thus helping the farmer know when the best period for farmers in India to plant crops will be.

This will enable the small farmers have a better edge in selling their products and reducing loss due to crops wilting away because of improper planting season, thus helping to increase the supply of organic produced crops.

### AI Functionality and Features

The AI's functionality will work based on data inputted by users in India(farmers), stating the period (i.e season) they would like to plant their seeds, which will then enable the model created to predict if this would have a good return or would be the best time to plant this seed.

Some of the features which I believe would be required for making this AI are:

- Data on seasons when seeds are planted.
- Data on how much yield happened during the season.
- Data on important factors that happened that affected the amount produced e.g(rainfall, temperature)

### Research Question

- What affects the growth of crops in India?
- Why is coconut the most produced crop but not one of the most popular crops
- What are the most and least popular crops produced in India?
- What periods of time yield the best for crops produced in India?
- Why is kharif a season with high production
- Why there was a huge spike in the year 2011 and 2014

### Physical Factors that affect Crop Growth in India

- Temperature

- Rainfall

## Impact assessment

### What problem does this technology want to solve:

Research has shown that around the world there is a decline in the number of Individuals taking up farming as a business, thereby leading to the low supply of organic produced food.

According to the Welp Magazine (Welp Magazine, 2020), common problems that small farmers face are “Insufficient funds for starting up or continuing their farms, No land availability and No or inadequate knowledge about marketing their products”.

This technology will be aimed at trying to solve some of these problems in India, by assisting the small farmers know when is best to plant their seeds for a good yield, thereby improving their farming knowledge and providing adequate income after selling their products. This will enable the reduction of naturally grown crops shortage and reduce poverty in India due to unemployment.

### Can you motivate why you are sure that this technology is solving the right problem:

According to (IFAD, n.d.), supporting small farmers can be a part of the solution to global food security (i.e all people, at all times have a physical and economic access to sufficient, safe and nutritious food to meet their dietary and food preferences for an active and healthy life) which in this case is organic produce.

Although the production of artificially grown crops has helped with preventing worldwide food shortage, research has found that long term ingestion of these foods can cause health problems in the long run for humans and animals (Defence Food Research Laboratory, 2012), therefore it would be healthier to consume more organic foods.

Moreover, the unemployment rate in India is at a high, individuals in rural areas become jobless after the realization that there is little to no support for farming which is the main source of income in rural areas.

The technology I plan on developing will be focused on trying to predict past information of when was the best time crops were. If this model prediction is then possible and to a good degree of accuracy, there is a chance of it assisting with predictions in the future, therefore helping to solve the problems stated above.

### In what way is this technology contributing to a world you want to live in:

A world I would like to live in is one whereby there would be less artificially grown crops for food, thus reducing health problems in both animals and humans, also a world where there would be less rampant poverty.

The technology which I am making contributes to achieving this world by its functionality of predicting when is best for farmers to plant, therefore small farmer or prospective farmers knowing of its existence will be more encouraged to go into farming thus increasing more naturally grown crops/organic crops and more people having jobs.

## Stakeholders: Have you considered all stakeholders?

What are the main users/target groups/stakeholders for this technology?

This technology is meant for farmers but mainly targeted to assisting small and upcoming farmers.

### Stakeholders Involved

Osuntuyi Michael (Developer & Internal Stakeholder)

Mr Adeyeye (Secondary School Agricultural Science Teacher & External Stakeholder)

Mr Raj (Software Developer from India & External Stakeholder)

Welman Nick (Semester coach & Internal Stakeholder)

Konings Hans (Semester coach & Internal Stakeholder)

### How is this stakeholder affected?

All internal Stakeholders listed to produce this technology require its functionality and process to be at a proficient level. A lot of time and effort has been given out by each individual stake holder to ensure the product is realized.

All External Stakeholders listed to produce this technology are affected in a little amount whereby it is not as imperative for the product to be fully realized.

### Did you consult the stakeholder?

Yes, all stakeholder regarding the production of this technology have been contacted and consulted.

### Are you going to take this stakeholder into account?

Yes, every stakeholder opinion about the product and during development will be considered.

### Did you really consider all stakeholders, even the ones that might not be a user or target group, but still might be of interest?

Yes, I have made sure to consider individuals which might not be users of this product example is other developers from my school.

## Privacy: Are you considering the privacy & personal data of the users of your technology?

Does this technology register personal data? If yes, what personal data?

No personal data of users is utilized in the development of this technology.

Do you think this technology invades someone's privacy? If yes, in what way?

No, users and none users' privacy is not being invaded with the use and development of this technology.

Do you think this technology is compliant with prevailing privacy and data protection law and can you motivate why?

Due to no personal data of the public and user is being utilized, the privacy laws do not apply to this technology.

In which way can you imagine a future impact of the collection of personal data?

Due to this technology not dealing with personal data of users and the public, this question does not apply to it.



## Data: Is data in your technology properly used?

Are you aware of the limitations and subjectivity of data and is this reflected in this technology?

Yes, I am aware of the limitations and subjectiveness of data.

Yes, the subjectiveness and limitations of data can be seen in this technology being that for creating the machine learning model I would have to use data about: which seasons produced high yield for farmers and these data entries are given by the farmers themselves.

## Transparency: Are you transparent about how your technology works?

(How) is explained to the users how a technology works and how the business model works?

Explanation on how the technology works in form of text can be found in this document, each section giving more insight about the technology, its purpose and how it will function.

Is it possible to file a complaint or ask questions/get answers about this technology?

The developer of this technology will be ready to answer any questions regarding his technology, to ensure transparency

Is the technology (company) clear about possible negative consequences or shortcomings of this technology?

All possible negative consequences and shortcomings of this technology will be listed in this document if any.



## Inclusivity: Is your technology fair for everyone?

Will everyone have access to this technology?

Regardless of this technology mainly targeting small farmers and upcoming farmers, all individuals interested in its use will be given the opportunity to access the technology and make use of it.

Does this technology have a built-in bias?

No, this technology has no built in bias and is fair to every user

Does this technology make automatic decisions and how do you account for them?

Yes, all automatic decisions made by this technology are because of the data it will be trained with.

Is everyone benefitting from this technology or only a small group? Do you see this as a problem? Why/why not?

I believe everyone will benefit from this technology, I do not see this as a problem because the benefits from this technology are directed to giving benefits to the user and the public.

Does the team that creates the technology represents the diversity of our society?

Yes, the developer does, to represent diversity means to get various views from different ethnic background, different genders and so- on, as the developer of this technology I have ensured I got views from different backgrounds as seen from the data which will be used in training my machine learning model is gotten from farmers from around the world.

## Hateful and criminal actors: What can bad actors do with your technology?

In which way can this technology be used to break the law or avoid the consequences of breaking the law?

Due to this technology's main goal is to try to predict information of the past and not of the future, it will be of no relevance to bad actors at its current stage/functionality.

## Human values How does the technology affect your human values?

How does your technology affect the identity of the user(s)?

I do not think this technology affect its user's identity/

How does the technology influence the user(s) autonomy?

This technology will influence the user's decision making heavily, because it is centered around helping its user make statically good decisions based on data it has been trained with.

But to make it clear this technology is designed to make suggestions, therefore it is up to the user to accept such suggestion.

What is the effect of the technology on the health and/or wellbeing of the user(s)?

Since this technology will be assisting its users in making decisions, the user will be less on the edge and most likely be more relaxed after its application.

## Future Did you consider future impact?

What could possibly happen with this technology in the future?

Utopian scenario:

After the long use of this technology, more naturally grown food/crops is in stock in comparison to lab grown food/crops, countries which suffer right now from food/crop shortage would be thinking of this problem as the past, more job opportunities open up in the farming sector, small farmers of 20/50 years before now own large businesses due to them working with the technology assistant and upcoming farmers would be confident about going into the agricultural sector.

Lastly, governments in Africa and Underdeveloped countries which have good agricultural land would have seen the profit and benefits of this sector and are currently making full use of it.

Dystopian scenario.

After the long use of this technology, greedy business owners which have noticed the benefits/profits to be made from this technology, would have gotten full ownership of it and now not making it available for those with lower financial status, making the cost to acquire the technology inflated.

Due to only wealthy and already established businesses having access to this technology, more people especially small farmers begin to close from farming, because they are at a disadvantage in the market when pit against these companies.

Lastly, there is an increase in artificially manufactured food/crops, and this has caused humanity to have gotten more unorthodox diseases which can withstand the natural protective measure of the human antibodies.

## Would you like to live in one of these scenarios? Why? Why not?

I would love to live in the Utopian scenario, mainly because, there would be less crime due to more people having jobs via the agricultural sector, human beings would be eating more healthy grown food/crops therefore, I would not have to worry about shortened health for myself, my family and friends due to the food we are eating, small farmers around the world are not at a disadvantage in the market and government officials especially in my home country Nigeria would be properly utilizing its most valuable sector and chances of Nigeria being a developed country would have become way higher or maybe Nigeria would already be a developed country at this time due to my technology assistance.

I would not like to live in the Dystopian scenario, because I would feel unsafe both health wise and physically

More artificially grown foods would be in the market, crime rate is high because most individuals who are not really educated can't find jobs because the easiest sector for them to go to has been taken over by lab science, fields and land for planting crops would have reduced making the ozone depletion increase.

All around this scenario would not be favorable.

What happens if your technology (which you have thought of as ethically well-considered) is bought or taken over by another party?

If my technology has been bought or taken over by another party,

if I have a chance to still be connected to the technology, I will find every way possible to ensure the company or party does not make the technology stray away from the main objective it was designed for.

If I am not given the opportunity to be connected at all after it might have been bought or taken over, I would have to just hope and pray that the party which has the technology stays true to the tech main/initial functionality and goal.

# Data Sourcing

## Data collection

The dataset which I am currently using was gotten from the

[Open Government Data \(OGD\) Platform India](#)

This website contains datasets on information related to India.

## Data Enriching and Storage

My approach to more enriching and storing my dataset will be with the application of google cloud, I found this to be an easy and reliable way of both adding information to my dataset and long-term safe keeping.

I will enriching my dataset by adding data on India rainfall and weather, I feel these factors might play a factor in the agriculture in India.

The screenshot shows the Google Cloud Platform console with a BigQuery query editor and results. The query is a SELECT statement joining crop production data with rainfall data for India. The results table displays data for Madhya Pradesh, SHAHDOL district, for the years 2001, 2002, and 2003, showing crop production and monthly rainfall.

Row	State_Name	District_Name	Crop_Year	Crop	Season	Area	Production	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	
1	Madhya Pradesh	SHAHDOL	2001	Sugarcane	Whole Year	23.0	590.0	4.40	20.40	0.00	104.60	187.30	283.90	198.90	144.30	213.50	105.20	1
2	Madhya Pradesh	SHAHDOL	2001	Sugarcane	Whole Year	23.0	590.0	0.30	1.90	32.80	49.40	175.80	384.20	269.30	248.70	205.90	163.60	8
3	Madhya Pradesh	SHAHDOL	2002	Sugarcane	Whole Year	27.0	690.0	74.20	37.20	126.30	248.20	197.10	396.80	604.00	290.70	257.90	94.00	4
4	Madhya Pradesh	SHAHDOL	2002	Sugarcane	Whole Year	27.0	690.0	41.50	4.90	12.20	3.30	32.10	196.30	138.40	352.40	122.30	57.80	1
5	Madhya Pradesh	SHAHDOL	2002	Sugarcane	Whole Year	27.0	690.0	3.80	7.90	0.30	14.70	40.60	108.20	54.70	118.20	52.20	126.40	3
6	Madhya Pradesh	SHAHDOL	2003	Sugarcane	Whole Year	20.0	510.0	10.60	56.40	19.40	3.00	3.90	40.70	136.70	94.70	52.20	0.00	2

## EDA

### 1. Findings(s)

Crop seasons in India

The dataset contains a column called seasons which consists of

- Spring.
- Autumn.
- Winter.
- Summer.
- Kharif.
- Rabi.
- Whole year.

After noticing in the graph below **Kharif** had the highest total amount produced over the years and **Rabi** at one point in time was the highest, I began to wonder what those seasons are actually and why do they have such high production amount in the dataset.

### Research result:

I found that Kharif and Rabi are classification of crops/cropping seasons, in simple terms Rabi crops grow during winter and Kharif crops during Monsoon which is in the period of Autumn

For further reading and explanation: (General knowledge Chronicle, n.d.).

### Research result confirmation:

After finding information about the cropping seasons from the website I decided to confirm from one of the external stakeholders listed in this document

Mr Raj, confirmed that the cropping seasons are as follows

- Kharif is Monsoon Crops (July to October).
- Rabi is Winter Crops (October to March).
- Zaid is summer crops (March to June).

Thus, this confirms my research result.

## Proof of Discussion with External stakeholder:



## 2. Findings(s)

After more EDA on this dataset, I found that Coconut was a crop which had the highest production amount in total, but when I visualized data on which crops are the most popular mentioned in the dataset, I found coconut was not one of the top 25.

Another strange occurrence which happened after performing more EDA, I found that in the year 2011 and 2014 there was a huge spike in the total amount produced.

## Research:

After noticing these strange occurrences, I went ahead to try to find out why

- Coconut is not a popular crop planted but has the highest production
- The years 2011 and 2014 have huge spikes in the total amount of crops produced

## Research result:

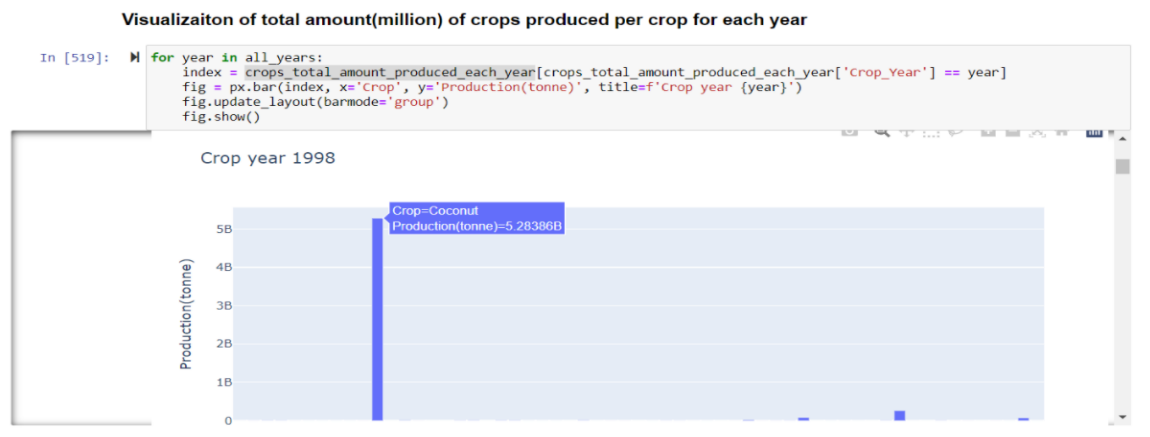
I found that these strange occurrences are as a result of the season whole year in the dataset.

The season whole year contains the total amount produced for certain crops and years, which included the crop Coconut and the years 2011 and 2014

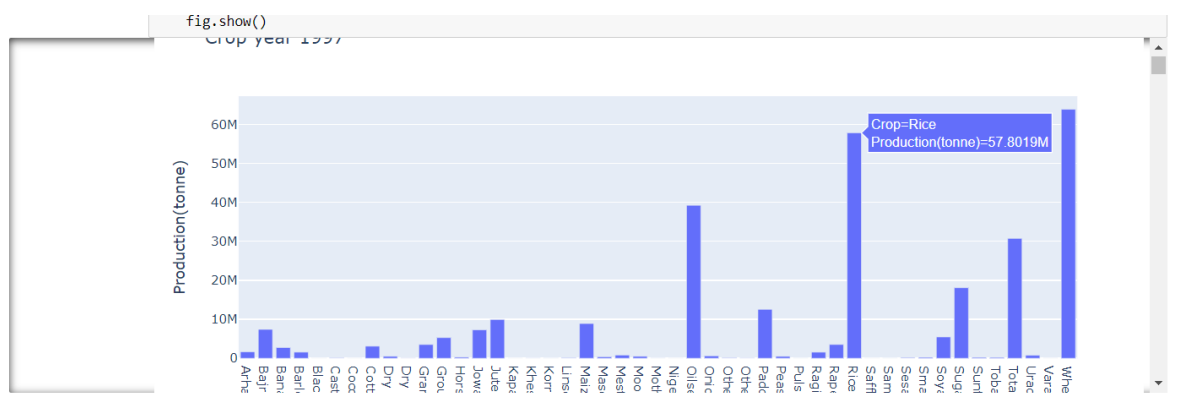
In conclusion, after removing data entries which were associated with whole year the strange occurrences were solved



Visualization of total amount of crops produced per year before removal of whole year season  
(Crop year 1998)



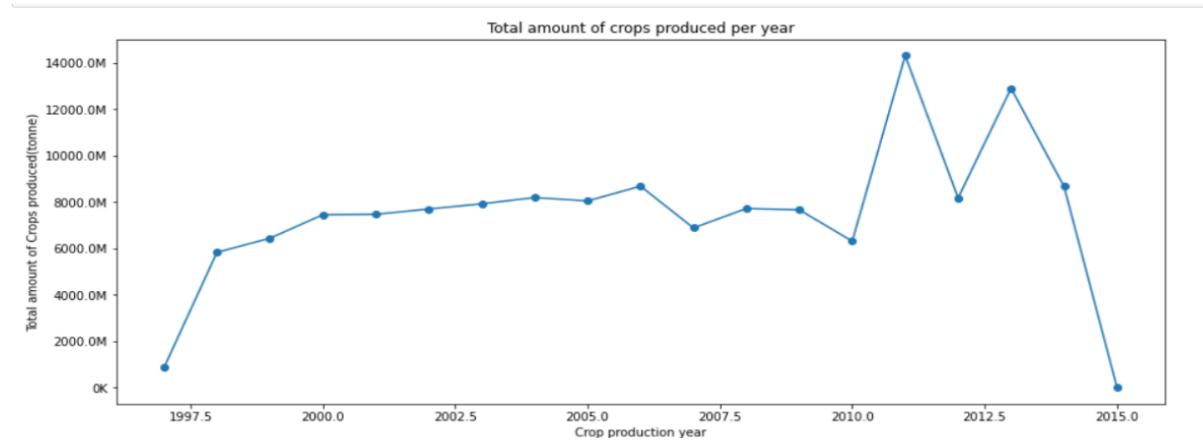
Visualization of total amount of crops produced per year after removal of whole year season  
(Crop year 1998)



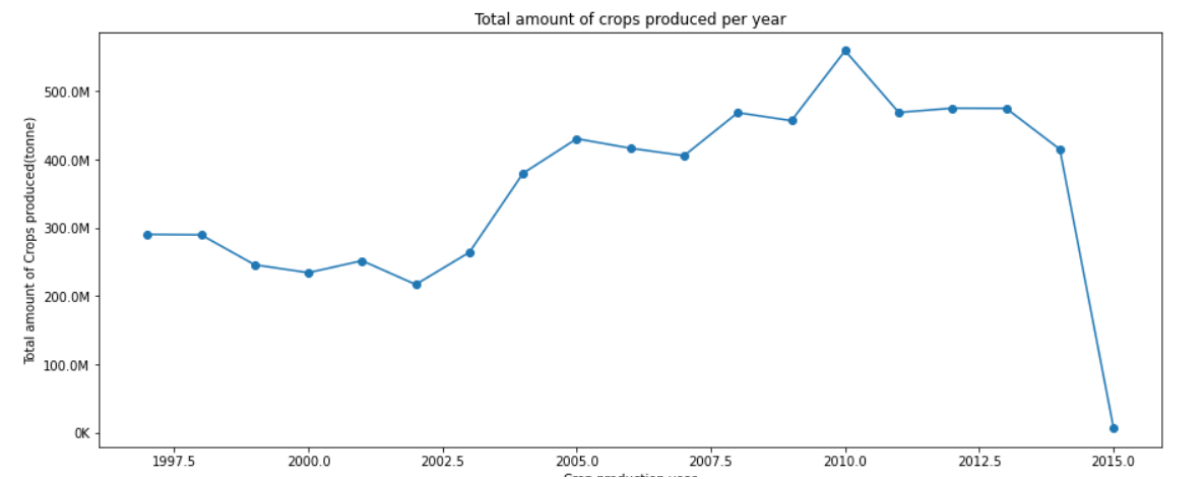
Visualization of total amount of crops produced for each year before removal of whole year season

Visualization of total amount of crops produced per year

Image of Crops produced per year before whole year data removal



Visualization of total amount of crops produced for each year after removal of whole year season



## Modelling

Is the choice of model(s) appropriate for the target variable e.q. the problem at hand?

My current choices for models to use for predicting my outcome of total amount that will be produced.

- Random Forest
- Decision tree
- K-Nearest Neighbor

These models are based on regression, I have chosen these models because my outcome is under regression, and these are the best models for working with regression predictive problem.

Should several models be compared? Why yes/no?

Yes, I believe several models should be compared, this is because by comparing several models I will be enabled to find the best model with the highest accuracy/will produce the best prediction

## Evaluation and deployment

Is there a plan for domain knowledge verification? Who can say whether the model works well or not?

Regarding domain knowledge verification, I plan on showing this document to my teachers which are stakeholders in this project and consulting both my agricultural teacher from secondary school and the developer from India who are noted as external stakeholders.

Questions about Deployment

What are acceptable evaluation results? (Think about accuracy, precision, recall, false positives/negatives etc.)

An acceptable evaluation would be if my Model can predict my outcome based on the past data with an at least 60-80% accuracy.

Will the model be deployed? If so, how?

I believe I would be deploying my model with a web application solution; this would be done with the use of the flask framework.

- Is there a feasibility deduction?

At the current stage of the technology development, I will not be able to answer these questions yet and would give an accurate answer when the technology development has reached this stage.

