I've decided that there's a lot of coding and programming I want to learn. There's a lot I want to work on but I want I've always struggled with getting work done and continuing the motivation and flow to complete it. A struggle I'm sure most of us have faced. I've got a long list of projects I've started or drafted ideas on yet many are either not started, on halt etc.

For the current work, I've decided to work on a simple 1 day project, well the goal is to work in it for one day. Todays project, as selected and supported by my good ol' friend GPT.

AI Supported Farm Resource Management System

So lets begin.
The "Farm Resource Management System" is a data-driven tool designed to optimise the utilisation of resources on farms. This project aims to enhance farm efficiency and sustainability by smartly allocating resources like water, fertiliser, and labour based on various agricultural data inputs.

The first step I would assume is figured out how we utilise the data and our task.
The goal for now, will be deciding how to allocate appropriate resources of water, fertiliser and labour to various areas of the crops. For instance, proximity to a lake would decrease water need. I am not entirely sure how I would even begin to approach this idea, its not clear, but that's what makes this interesting.
I felt that given I don't know much about agriculture and farming it would be an interesting challenge as I've always found it to be quite interesting.
I'm thinking the context and use case we might focus on is a bee farm and a lavender farm. Or possible a wheat farm given its great presence. Tea would also be an interesting crop, but lets keep it simple.
My current goal is just to fix some possible parameters and cases in order to direct my research and narrow the scope before we can bring it back to a generalisation.

Given my little understanding of the material, I felt it would be natural to start looking at what current work and research has already been done. GPT can only take you so far.

After performing some research on two papers, I've got a better idea of what we're trying to predict.
Fertiliser will be crucial and critical, for that, we aim to determine for a given area given the context of soil and crop the values of N,P,K which is the 3 primary chemicals used in fertiliser. They will be numerical values.
With regards to water, we'll consider volume of water required per area, so how much water is required. Future considerations could consider irrigation schedule, such as how much we'd need given a set day or week. But for now based on some set time parameter, we'll only look at and assess the volume of water needed for a given area.
Labour we'll consider as simply the number of manhours of work required, this will be a general parameter we'll consider. It may prove useful to consider hours required as a time dependent variable, as planting and harvesting would lead to higher or lower workload. Labour would be abstracted to a ratio of number of people needed. Though there is the possibility of abstracting how much of it and what tasks can be automated or taken over by machines though for simplicities sake, we're not gonna bother with that.

To conclude the params we'll output are:
Fertiliser:
N, P, K values for a given soil, therefore given us a ratio of what we need
Volume of water usage per area (Most likely per square metre, though I need to confirm this is granularity is appropriate)
Hours of work required, we'll assume simply a base amount of work required for now, regardless of crop planting and harvesting, that can be determined later on.

To simplify the data, we'll consider the model to assess the total water, hours, N,P,K for one cycle of the plant, so one plant to harvest.
If this proves successful, we'll then consider a more granular time scale. I'm unsure if time scale should be 1 day/24 hours or 1 week, though that can be determined later.

Now that we have the outputs figured out, its time to assess what data we need and relevant data sources.
With the annoying work done, I've got a much clearer picture of what it is I'm trying to achieve. Now for lunch before we get back to work!

The data sources

Now that we've decided what we are trying to do, its time to think about how we are gonna get the data for our model.
I've broken up the data into a few general needs:
Crop Growth and Needs:
- Growth cycle data for the chosen crop (lavender or wheat), including various stages from planting to harvesting. I would like to expand to future crops, but lets keep it simple to two edge cases.
- Specific nutrient requirements (N, P, K) for the crop at different growth stages.
- Water requirements throughout the growth cycle.

Soil Conditions:
- Soil type and texture data, as these affect water retention and nutrient availability.
- Soil pH and existing nutrient levels (N, P, K).
- Historical data on soil conditions, if available, for trend analysis.

Landscape and Geospatial Data:
- Proximity to water sources, which can influence irrigation strategies.
- Field size and layout for accurate resource allocation.

Climate and Weather Data:
- Historical and current weather data, including rainfall, temperature, humidity, and evaporation rates.
- Seasonal weather patterns to anticipate water needs and plan irrigation.

Labour Requirements:
- Historical data on labour hours required for various tasks (planting, maintaining, harvesting) for the specific crop.
- Task-specific labour needs considering crop and farm size.

Water Availability and Sources:
- Information on local water sources and their reliability (e.g., wells, rivers, rainfall).
- Data on existing irrigation infrastructure and its efficiency.

Agricultural Practices and Management Data:
- Historical crop yield data to correlate with past resource use and practices.
- Information on current and past farming practices, including crop rotation, fertilisation, and irrigation methods.

Regulatory and Environmental Data:
- Local agricultural regulations regarding resource usage, especially for water and fertilisers.
- Environmental impact data related to farming practices in the region.

I've started with labour estimates as I thought once some data had been found, I hope it's a pretty simple parameter, as we can consider it as a simple ratio based on the crop and area, a simple rule based approach.
I started my approach by considering how many hours of work, lets refer to this as WH for work hours is required per square metre, though I quickly realised that no body measures crops based on square metre, and instead acre might be more appropriate.
In searching for an appropriate measure, I've tried finding data on man-hours for various activity on a farm, really any survey or metrics I can use, I simply need a starting point for this, a reasonable reference. Its not the best practice, but as we are starting at a very general, high level approach before developing a more robust and granular model I felt this approach would be sufficient.
Whilst my parameter looks at research papers, an interesting approach could be researching present and historical employment at farms and correlating that with crops, farm size, yields and other time and geographic data. Job listings would also be interesting to assess. One could potentially also assess cars, lodging and other assets to house people as a crude way to predict potential number of people required to operate in a farm. With all this data, I would like to hypothesis that one could generate various bounds and estimates of numbers, combine all these factors to determine a more accurate measure, though such a level of data needs is not presently required for this work, though I would imagine labour needs and WH would be fascinating study for further causal inferences in the local and wider economy.

Two papers on this matter are linked blow which have helped me in my research
1. Agrarian origins of Individualism and collectivism
2. Google search source, from the 80's

Paper 1 proved to be incredibly interesting and insightful. In short the paper discusses the relation between work in agriculture production and relation to individualism, this doesn't quite do the paper justice and is an over simplification however its quite insightful. Whilst I would love to go much deeper into the topic, I should force myself to remain focused on the present task at hand, the following figure will serve as a very useful general benchmark.

Hand-Method		Machine-Method	
Operations	Man-Hours	Operations	Man-Hours
Breaking ground	6h 40m	Breaking ground	1h
Sowing seed	1h 15m	Sowing seed	15m
Pulverizing	2h 30m	Pulverizing	12m
Reaping, binding, and shocking	20h	Reaping, thrashing, and sacking	1h

Whilst other papers are discussed and specific formula's are outlined detailing how to estimate and model WH, we'll take this example for wheat operations. We'll assume machine operations, though the difference in hours is fascinating and staggering. The specific hours yields a total time per acre as below

Title	Hand-method	Machine-method
Hours, minutes	30,25	2, 27
Hours	30.42 hours	2.45 hours
Hours square metre (H/m^2)	0.00752	0.000605

Note 1 acre is 4046.86 square metre[1]

Some days later and a nice cup of tea its time to get back to my project. Whilst it was planned to be a simple one day project, that, did not go as planned, there was more work than expected though it was quite fun and enjoyable with what I've learnt so far. Given this was intended to be a more simple project, perhaps starting with more mock data would have been more suitable and more robust defined problem and context to develop a solution, though diving deep without much knowledge has been a valuable experience.

Now that we've got some basic data on labour requirements, its time to find soil and environmental data, growth/crop requirements, water data and landscape data.
To develop the crop data, I made some pseudo data with the help of GPT,

Field ID	Soil N (ppm)	Soil P (ppm)	Soil K (ppm)	Soil pH	Avg Temp (°C)	Annual Precipitation (mm)
000	20	30	150	6.5	16	750
001	10	25	100	7.0	14	650
002	15	20	120	6.8	17	700
003	5	15	80	7.2	19	500
004	25	35	160	6.3	15	800
005	30	40	170	7.4	18	550

Crop Type | Growth Stage | Water Requirement (L/m²) | Elevation (m) | Slope (%) | Proximity to Water (m) |

[1] Hours calculated as hours/4046.86, for data source, see 1.1 in research

Wheat	Seeding	0.4	120	2	100
Lavender	Flowering	0.2	200	3	250
Wheat	Vegetative	0.6	100	5	150
Lavender	Harvest	0.1	90	10	300
Wheat	Harvest	0.7	110	4	80
Lavender	Germination	0.3	85	6	400

Through this data, I would then extrapolate the data based on the distribution and created another 40 data points. Creating what is essentially an infinite set of new pseudo data is quite a fun and something I would like to dive deeper into, see how much one can push artificial and pseudo data.
Now we've got crop data, we need to develop GIS data on soil, water proximity etc.
Its easy to get lost from the goal of this project, finding data for the sake of data so its been helpful to regularly reflect on what I wrote previously and reference what's been written in the research papers I've found
"The "Farm Resource Management System" is a data-driven tool designed to optimise the utilisation of resources on farms. This project aims to enhance farm efficiency and sustainability by smartly allocating resources like water, fertiliser, and labour based on various agricultural data inputs."
Whilst I have been attempting to find data of N, P and K levels in soil, I have been struggling to find any datasets that provide such GIS data, perhaps it would be difficult to find such data, I'm sure there would be some datasets but as I wanted this to be a GIS focused project, with mapping, I see I'm somewhat inexperienced with this and struggling to find such data, I've decided to forgo that for now, and we'll assume there's simply a base amount of fertiliser needed, irrespective of present soil conditions.
With focus as soil moisture, I ended up finding this dataset which was very useful
https://shiny.esoil.io/SMIPS/
https://portal.tern.org.au - Another useful data source I'd like to explore in greater detail.

I got some lunch, went for a walk and assessed the progress I've been making and I strongly feel that it's really not worth it to integrate my GIS data into this project. At the end of the day I just need data to create the model, the GIS and real world mapping data would more so serve a technical prototype, something with more usability, something very cool to have, but simply not worth the effort and cost, though I have still learnt some presently, perhaps I'll tackle GIS in another project, but today is not the day.

With this much simpler approach, we created various mock data ranging from farm size, terrain slope, waterfall and n,p,k levels in the soil as well. Though the data did not meet the requirements as I originally planned to set out. After some thought, given some data would suit different scales and would not blend well, so data like labour and the size of the farm would remain as independent values to the rest of the data, simply acting as a ratio for expected hours and labour required. All other data could be modelled through assessing the present data, averages and trends in the crops, a generally rule based approach. Decision Trees would most likely be the optimal strategy. Given much of the theory had been developed and all the data was, well not particularly applicable, I felt that this has achieved some level of completion.

Whilst the original outcome was not what I intended, I went into this project with the assumption that data would be much easier to find and my project would be somewhat simple, the difficulty of sourcing the data proved to be greatly troubling, I overestimated the work required for this “one-day project” but I’ve learnt a bit about finding and considering the use of data in the real world, how difficult it can be, and an understanding of the critical importance of defining the problem, task as well as thinking about the relevant inputs and outputs. Less technical achievements have been developed than planned, though the general skills for approaching open ended problems, finding research papers and sources on relevant sources and understanding how data can operate together has proved as a useful exercise.

Useful research findings

Micro simulation model of Australian farms

https://www.agriculture.gov.au/sites/default/files/sitecollectiondocuments/abares/farmpredict-micro-simulation-model-of-australian-farms.pdf

site-specific climate data including various measures of rainfall, temperature and soil moisture. A statistical model is estimated linking the production of outputs (e.g., wheat, beef cattle, wool etc.), the use of inputs (e.g., fuel, fertiliser, labour etc.) and changes in farm stocks (e.g., livestock and grain) with farm fixed inputs, input and output prices, climate variables and other control variables. The model is estimated using a non-parametric machine learning method, which combines a gradient boosted regression tree algorithm (xgboost) with multi-target stacking (two-stage regression). The resulting model can be used to forecast or simulate production, financial outcomes and stock changes for individual farms given scenarios for climate conditions and commodity prices
Variables are considered

Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices

https://www.mdpi.com/2077-0472/13/11/2141

Use both a NN and a rule based approach. Useful reference for recommendation fertilizer use.
A fertilizer recommendation system is a specialized tool aimed at aiding farmers in making informed decisions about the suitable amounts of fertilizers to use for their crops. This system aims to increase agricultural productivity while minimizing adverse environmental impacts.
This study proposes a rule-based fertilizer recommendation system to guide farmers on the most beneficial types and amounts of fertilizers for specific crops. The system’s foundation is rooted in the well-established scientific principles of soil chemistry and plant biology. Recognizing that different crops have varied nutrient requirements and that these needs are influenced by soil pH, the system seeks to bridge the knowledge gap by providing specific

fertilizer recommendations based on these factors. The fertilizer recommendation system considers various parameters such as soil quality determined based on pH level, crop type, and specific nutrient requirements of each crop. By analyzing these factors, the system provides recommendations for the optimal amounts of **N, P, and K, the primary nutrients needed by crops**. It is to be noted that soil pH is a crucial parameter because it affects the solubility of nutrients, which has a high impact on plant growth. A pH of 7 is considered neutral, while anything below 7 is acidic and anything above is alkaline. The soil's pH can influence the crop's health, yield, and disease resistance. Different crops prefer different pH ranges.

Labour papers

1. https://www.researchgate.net/publication/357909077_Agrarian_Origins_of_Individualism_and_Collectivism
    1.1. https://www.researchgate.net/figure/Required-man-hours-per-acre-for-producing-wheat-by-operations_tbl3_357909077
2. Labor Used to Produce Field Crops, Estimates by States
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiVg43cgZaEAxVXcmwGHeB2AzQQFnoECBQQAQ&url=https%3A%2F%2Fageconsearch.umn.edu%2Frecord%2F153771%2Ffiles%2Fsb346.pdf&usg=AOvVaw3Y7O5-dM8_7AOKsOXCFqYM&opi=89978449