Optimizing Investments on Agricultural Land by Data Analysis

Group-U

Pachava, Venkata Goutham - 11667698

Gummalla, Bharath - 11665138

Introduction

It is important that agriculture is food secure and economically stable. To engage in farmland investment, decision is not easy to make. Farmers and agri-businesses need to verify number of factors before leasing or buying land for farming. This will include fertility of soil, land cost and what crop is previously produced. This is because the wrong choice of land will result in low production and leads to financial losses. Therefore, it is vital to have data information on the investment to be made. The main aim of the project is to assess the historical crop yield data, soil characteristics, farmland prices and crop markets to establish the most viable areas for cultivation of different crops. The findings of this study can be useful to farmers, investors and businesses dealing with agriculture to make better decisions on the farmland to choose in order to enhance production and earnings.

Due to changing soil conditions crop productivity is not that easily investable as in other industries. In most industries where data is rather stable, agriculture is a process of constant adjustment as the project is significant in the real world due to the understanding of these variables. This is because farmland investment is not as straightforward as that of other forms of business investment. Crop rotation which is a conventional means of checking soil fertility results in rather erratic yields. This is a useful project for decision making especially for financial risks and historical trends and soil conditions where data is collected over a period of time.

This project is scalable. This analysis may be extended over several crops and locales whereas many agricultural studies concentrate on particular crops or geographical areas. The information may help investors find lucrative land farmers choose the best crops and legislators create agricultural regulations. Farmers & investors may use data-driven methods to make well-informed agricultural as well as financial decisions. This research combines agricultural & economic analysis to make sure land investments can be both economical and productive.

The aim of this research is to connect efficiency with choices effectively by integrating past patterns in agriculture with soil information and market rates to give useful advice on enhancing returns from investing in farmland.

The Data

The analysis of farmland investment potential in this project is based on historical agricultural data. The data was obtained from the National Agricultural Statistics Service (NASS) and government database of agricultural statistics in the United States part of the United States Department of Agriculture.The USDA Quick Stats Database (https://quickstats.nass.usda.gov/) that contains comprehensive agriculture statistics spanning several years & regions which is the source of the dataset. To maintain relevance as well as consistency alongside current farming practices. we have chosen data for this project from 2010 to the present but the USDA Quick Stats Database has agricultural records going back to 1920. We may examine current agricultural investment prospects while preserving enough historical context for analysis of trends by concentrating on recent data. The dataset was collected by government records, satellite observations along with farmer surveys. Crop production, the value of land, soil conditions and market prices are all recorded by USDA structured agricultural assessments

which are used to generate this information. The approach used for data collecting ensures outstanding precision as well as reliability which makes it appropriate for a thorough examination of farmland investment potential at the regional and national levels. The data set consists of approximately 50,000 observations and 15 important variables that are relevant to the study of agricultural investment. A wide range of variables such as productivity, land value, environmental circumstances as well as economic considerations influence agricultural decision-making.

Below is the structure of dataset:

Attribute	Details
Total Observations	~50,000
Total Variables	15 .
Time Period	2010 - Present
Geographic Scope	County and State-Level Data (USA)
Data Collection Methods	Farmer Surveys Government Reports Satellite Observations
Primary Data Source	https://quickstats.nass.usda.gov/

Crop Yield Dataset (County Level) provides historical agricultural yields per acre for important crops such as wheat, barley, soybeans and maize.

Land Value Dataset (County Level) gives information on typical farmland prices every acre which may be used to determine the possibilities for land investment.

Soil Quality Data (County Level) includes important soil parameters that affect agricultural yield, such as pH, moisture content and fertility indicators.

Crop Market Prices (State Level) keeps track of important crops past selling prices to support profitability analysis.

Research questions Exploratory Data Analysis (EDA) Questions:

1.What effects do soil characteristics have on the yield of crops in various geographical areas?

- Finding the ideal soil conditions for various crops can be made easier with an understanding of soil properties and how they relate to crop productivity.

2.Over the last ten years, which states and counties have produced the best yields of various crop types?

- By examining past yield patterns, consistently high yielding areas may be identified which makes them perfect future agricultural investment.

Advanced Analysis Questions

3. Which factors affect rising or falling yields and how have patterns in agricultural yields altered over time?

- Long-term yield trends & outside factors including shifting land use, degraded soil and prevailing economic conditions are examined in this question.

4. In order to achieve maximum profitability in various farming sites. how can we best choose crops and make investments?

- This topic deals with financial decision-making, analyzing land value , crop productivity & market trends to choose farming areas that yield the greatest return on capital.

Challenges in Data Collection and Integration:

The intricacy of gathering and combining agricultural data is one of the study main obstacles. Careful preprocessing is necessary to combine disparate variables into a single dataset since they are gathered at different geographic levels (county-level for soil and crop production data, state-level for crop market price). Crucial phases in the study include addressing missing values & guaranteeing data consistency.

Crop rotation techniques may add unpredictability to yield statistics. To preserve soil fertility farmers frequently switch crops every season, which makes it challenging to identify reliable long-term trends. When examining past yield trends this element needs to be taken into account.

There will likely be 15 variables and more than 50,000 rows in the finished dataset. which will put a heavy burden on processing capacity and performance. To manage such a massive dataset, effective data processing methods are needed such as streamlined data pipelines for data cleansing, transformation and merging. In order to assure seamless processing powerful statistical models & geographical studies may increase computational burden and need for scalable computer resources. memory-efficient strategies will be used to avoid performance snags when doing intricate analysis. Python will be utilized for data processing and modeling in order to address these issues. It will make use of libraries like Pandas and NumPy for manipulating large amounts of data, Scikit-learn for statistical modeling as well predictive analytics, Geo pandas as well as Folium for mapping and geospatial analysis & Matplotlib for visualizing trends.

Exploratory data analysis (EDA) is utilized to discover patterns in productivity geospatial analysis is used for mapping high-yield areas, trend analysis is used for looking at past yield changes & profitability modeling is used to evaluate agricultural investment. Interactive stakeholder insights dashboards will be made possible via Tableau.

Missing data, uneven data levels and changes in market prices are major obstacles. Techniques for imputation will be used to handle missing records. We will use weighted average to modify land value data which is only available at the state level. Market price volatility will be taken into consideration using sensitivity analysis.

Using the agricultural databases and AI driven research this initiative aims to change farmland investment choices. We will offer useful suggestions for boosting agricultural earnings and crop productivity through thorough data integration & predictive modeling. in addition to being pertinent to farmers and agribusinesses this initiative offers substantial academic and industrial significance and lays the groundwork for future precision agricultural uses of Ai.