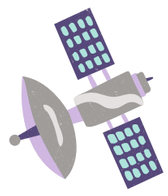
Authors Mattia Cintioli, Natali Peeva, Yuan Su, Sam Tegel and Jesse Wijlhuizen

UNIVERSITY OF AMSTERDAM

Agurotech

Satellite Images in Agriculture



To what extent could future NDVI values be predicted based on weather-connected variables as well as past records of NDVI extracted from satellite images, (using time-series models, supervised learning, and deep learning)?

1 PROBLEM STATEMENT

NDVI prediction in agriculture offers numerous advantages for farmers, society and our stakeholder Agurotech. NDVI is a proven effective indicator of vegetation distribution, growth, health, and productivity changes over space and time. A higher NDVI indicates more vegetation and is also highly correlated with crop yield, making NDVI prediction a valuable tool for more efficient water usage, cost savings and improved profitability for farmers. Moreover, it plays a crucial role in the fight against climate change.

70% of water worldwide is used for agriculture irrigation

2 DATA, PREPROCESSING AND CLEANING

The team utilized the ML4EARTH Hackathon dataset, which was supplied by the project supervisor, covering the period from 2003 to 2021 and comprising climate and satellite image data from 466 American fields. After preprocessing the data, we removed features that were highly correlated or of low importance to our prediction models. To perform supervised machine learning, we also added the previous NDVI value to each sample. Our models were applied to the parcels with no missing data, leaving us with 180 parcels to work with, and the data was then scaled using a min-max scaler.

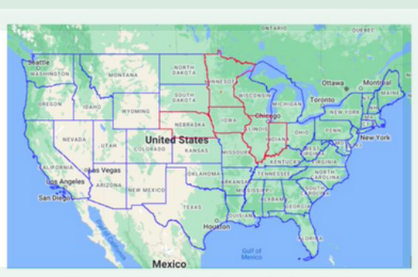


Figure 1: Map of study area (red outline)

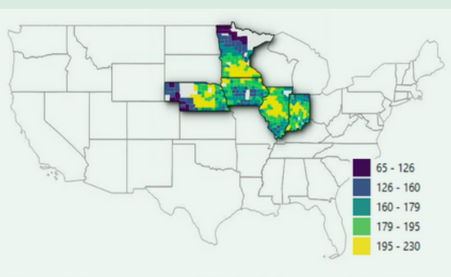


Figure 2: Average corn yield (in bushels per acre) for 2021

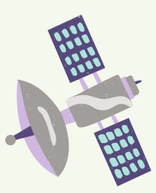
Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
feature	b1 red	b2 nir	b3 blue	b4 green	b5 nir2	b6 swir1	b7 swir2	tmin	tmax	prcp	heat/cold wave index	drought index	NDVI	EVI	NDWI
data type	float							float			float	Categorical (stored as float)	float		
source	MODIS surface reflectance							Daymet variables			Daymet-derived	USA drought monitor	pre-computed features from MODIS		

Table 5: Index of features in .npy files

3 METHODOLOGY

To reach the project objective, the team experimented with various time series, regression, ensemble learning, and neural network models. LightGBM time series emerged as the most accurate and efficient for NDVI value forecasting in the coming months, while Random Forest ensemble learning was best for short-term predictions. Hence, the approach used both models, with Random Forest for short-term predictions and LightGBM for longer term forecasts.

Additionally, based on previous research that showed a strong correlation between NDVI and crop yield, a Random Forest model was employed on the forecasted NDVI values to estimate the expected annual crop yield.

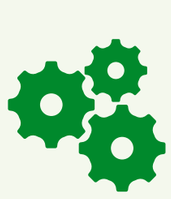


4 RESULTS/FINDINGS

Our results showed that Time Series models were effective for long-term forecasting, while Ensemble learning provided accurate short-term predictions. Neural networks, however, were found to be too time-consuming and computationally expensive for our project and were not necessary. Linear Regression performed well and achieved high accuracy, making it a viable option for future projects if interpretability is a priority.

5 IMPLEMENTATION

The implementation focuses on providing an easy interpretation for the end-user. We have included two visualizations: of the near future and till the end of the year. We included intervals to show when the NDVI of the fields is within the norm of the corresponding season. The implementation also uses color-coding, with red representing danger zones and green representing optimal zones. Based on that the user gets a recommendation whether or not to take action. An end-of-season crop yield estimation is provided, along with automatic profit calculation based on current prices, with the ability to select specific crop fields. Explanations of the graph visualizations of NDVI and crop yield estimation are provided in an interpretable manner, and the interface design is designed to fit well with the current version of the app.



6 USER TESTING

The user interface was tested by students from different background of studies in different levels. First, students freely explored the user interface without knowing what this user interface does. Second, a brief introduction of Agurotech and its appliances were introduced and why it is important to include remote sensing analysis to the current technology. In the last, based on the feed back, some explanations were added to the user interface.

7 CONCLUSION

In conclusion, the team has created a complementary pair of predictive models to forecast the future of NDVI values. LightGBM is used for long-term predictions, while Random Forest is utilized for short-term predictions. The resulting predictions are conveyed to the end-user through a simplified user interface to provide insight into the health of their crops and urge for action if needed. Moreover, using the NDVI forecasts, the team has developed a model that calculates annual crop yield, which is then used to estimate the end-user's potential profits.