Agricultural Crop Recommendation (ACRE) for Farmers

Rohit Patel (rohitpatel@iisc.ac.in)
Department of Computer Science and Automation, IISc
Advisor: Prof. Y. Narahari

June 17, 2022

Abstract

Agriculture has a significant role to play in the Indian economy. As an important source of income for around 58 percent of the population of the country, it also plays a major role in employment generation. Due to lack of knowledge and technological resources, most Indian farmers do not properly choose crops through a scientific use of information available on soil requirements, weather conditions, and market demands and this results in sub-optimal yield and revenue for the farmers. In this work, our aim is to recommend a near optimal portfolio of crops to the farmers by using available data such as soil characteristics, weather conditions, geography, historical yield data. We develop several machine learning and deep learning models to predict yield using weather conditions, soil characteristics, and geographical location. The distributions of yield, price, and costs are considered, and the crop profit and risk-utility are calculated. The recommendation system generates several portfolios of crops and ranks them using the financial mathematical ratio. We use the Sharpe ratio, commonly used in finance portfolio selection, to recommend a subset of crops to the farmers. We aim to use the data available from public sources to illustrate our approach to crop recommendation.

1 Introduction

India is one of the world's oldest agricultural countries. However, agricultural practices have changed dramatically in recent years due to globalization. Several variables have influenced the state of Indian agriculture. To reclaim its health, many innovative technologies have emerged. The most significant difficulty in the agriculture sector is a lack of understanding of favorable soil characteristics for crops, changing climate conditions, and market demands. Each crop requires its unique set of meteorological conditions and soil characteristics.

Agriculture is one of the most important sources of income for around 58 percent of the inhabitants of our country[7][6]. According to the 2016-17 Economic Survey, a farmer's average monthly income in 17 states is around Rs.1700, resulting in farmer suicides and the conversion of agricultural land for non-agricultural purposes. Furthermore, 48 percent of farmers do not want their children to take care of their farms, preferring to live in cities. The reason for this is that farmers frequently make poor crop selection decisions[17], such as choosing a crop that would not produce a high yield for the soil, planting in the wrong season, and so on.

In this project, we recommend the best suitable portfolio of crops to the farmers based on the input parameters such as soil characteristics, climatic conditions, geographical location, and market demand supply.

The recommendation system calculate utilities according to the farmer and recommend the crops based on these utilities. To calculate these utilities, the system depends upon the yield, price, and cost of cultivation and production of a crop along with their standard deviation, which helps in considering the risk profile of the farmer. We consider the profit and risk as the utility for the farmer. After calculating these utilities, the recommendation system provides the individual crop recommendation for Kharif and Rabi seasons, each based on these utilities. The recommendation system also provides the inter-crop recommendation, generates several portfolios of crops that could be grown in a given season. These portfolios are evaluated by using the financial mathematical Sharpe ratio. The Sharpe ratio considers the expected return and the risk of getting this return. The more the Sharpe ratio of a portfolio, the more the portfolio provides stability in getting the return. The farmer can choose a particular portfolio based on their risk profile.

One of the significant components of this recommendation system is crop yield. Crop yield is a complex variable influenced by several factors, including genotype, environment, and interactions. For accurate yield prediction, a fundamental understanding of the functional relationship between yield and these interaction components is required [12]. We have performed crop yield prediction using an ensemble technique consisting of several machine learning and deep learning regression models. We have considered the parameters for crop yield prediction: soil characteristics, weather conditions, and geographical location. The parameters included in soil characteristics are soil type, pH value, and nutrients such as nitrogen, phosphorous, and potassium. The AESR (Agro-Ecological Subregions) index represents a particular region's type of soil and land characteristics. The weather parameters include rainfall, temperature, humidity, and sunlight. The geographical location is considered using the latitude and longitude of a region.

2 Review of Relevant Work

A significant amount of research has been done on the effects of soil characteristics, climatic conditions, and different geographical locations on agricultural productivity. Machine learning has limitless potential in demonstrating the links between these parameters and agricultural performance. It is a new scientific dimension that uses data-driven ways to increase agricultural productivity. Machine learning techniques assess and help decision-making in intelligent farming.

Artificial intelligence for social good (AI4SG) is a research theme that uses and develops artificial intelligence to address societal concerns and increase global well-being. [18] selects 1176 papers from the AAAI, IJCAI, IAAI, AAMAS, KDD, and COMPASS conferences from the years 2008-2019, and analyzes the trend of AI4SG research. To identify existing efforts in various AI4SG application domains, the authors have proposed a uniform structured AEC (agent – environment – community) architecture. A DPP (descriptive – predictive – prescriptive) framework has been proposed to consider the functionality of the AI intervention.

2.1 Existing Crop Recommendation Systems

[15] has proposed a system to assist the farmers in crop selection by considering all the factors like sowing season, soil, and geographical location. To recommend a suitable crop to the user, the proposed system considers environmental data such as rainfall, temperature, and geographical location in terms of the state and soil characteristics such as soil type, pH value, and nutrient concentration. The key objectives of this paper are, 1. Develop a robust model that can accurately estimate crop sustainability in a given state for a given soil type and meteorological circumstances. 2. Make recommendations for the best crops in the area so that the farmer does not lose money. 3. Analyze the profitability of various crops using data from the previous year. Their Dataset include i) Yield Dataset: contains yield for 16 major crops grown across all the states in kg per hectare. ii) Cost of Cultivation dataset: provides the cost of cultivation for each crop in Rs. per hectare. iii) Modal price of crops: gives the average market prices for those crops over a period of two months. iv) Standard price of crops: gives the current market price of the crops in Rs per hectare. v) Soil nutrient content dataset: contains Nitrogen content, Phosphorous content, Potassium content, and average ph. vi) Rainfall Temperature dataset: contains crops, max, and min rainfall, max and min temperature, and ph values. The proposed system is implemented with Linear regression and Neural network and compared with KNN, KNN with cross validation, Decision Tree, and Naive Bayes.

[16] has proposed a recommendation system through an ensemble model with majority voting technique using Random tree, CHAID, K-Nearest Neighbor and Naive Bayes as learners to recommend a crop for the site specific parameters with high accuracy and efficiency. Their dataset contains the soil specific attributes collected for Madurai district tested at soil testing lab, Madurai, Tamil Nadu, India. The crops considered in this paper include millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. Different machine learning models give the best suitable crop based on the given input and then the crop which is selected by most of the models is recommended to the farmer which is simply a majority voting rule.

In [14], the authors introduce a recommendation system which uses an Artificial Neural Networks (ANN) for recommending the best suitable crop. The crops that they have considered are Maize, Finger millet, Rice, and Sugarcane. In this work, they have considered the data for two locations namely Hadonahalli and Durgenahalli of Doddaballapur (dist.), Karnataka, India. Location specific crop is recommended based on the suitability measures namely: 1. Highly suitable 2. Moderately suitable 3. Marginally suitable 4. Not suitable. They have proposed a four-layered Artificial Neural Network architecture. Crop suitability is recommended in terms of classes. They have the predefined conditions which are favourable for these crops and train their neural network using this data. For the recommendation part, they have ranked these crops based on the values predicted by the model and put these crops in the corresponding classes.

[20] uses multi-objective evolutionary algorithms to select the best crop to plant for sustainable land use based on the soil data. The costs of fertilizing and liming, cultivation, and the expected fluctuation of total return are among the optimization criteria. This work provides a method based on Multi-Objective Evolutionary Algorithms to aid in selecting an optimal cultivation plan by taking into account five crop options and five objectives. This work considers a multi-objective crop selection problem having the following objectives: to minimize the cost of fertilization and lime application, minimize the total cost of production, maximize the average return, maximize the worst-case return and minimize the standard deviation of returns.

2.2 Related Work on Crop Yield Prediction

Many attempts have been made to apply machine learning models to predict the effects of weather patterns on crop productivity to solve concerns such as food insecurity, supply stability, and economic planning. However, because these models are either limited to a small area or a short period (e.g., a few years), they may not be able to generalize spatially or temporally. They also treat each location as an i.i.d sample, disregarding spatial data correlations. In this paper [10], they have introduced a novel graph-based recurrent neural network (GNN) for crop yield prediction, which incorporates both geographical and temporal structure. Their method is trained, validated, and tested on over 2000 counties from 41 states in the US mainland, covering years from 1981 to 2019. They compared 11 representative machine learning models, including GNN and GNN-RNN, on US county-level crop yields for corn and soybean. They have achieved a relative R^2 improvement of 10.44% over the recent CNN-RNN model, 16.16% over the 5-year LSTM, and a relative RMSE improvement of 9.6% over the CNN-RNN model, 13.18% over the 5-year LSTM. These indicate the importance of exploiting geospatial context in making these predictions.

[13] proposes a deep learning framework for agricultural yield prediction based on environmental data and management techniques that uses convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Using historical data, the proposed CNN-RNN model was used in conjunction with other popular methods such as random forest (RF), deep fully connected neural networks (DFNN), and LASSO to forecast corn and soybean yield across the entire Corn Belt (including 13 states) in the United States for the years 2016, 2017, and 2018. The CNN-RNN model was created to capture the temporal dependence of environmental factors and the genetic improvement of seeds over time. Their model is a combination of CNNs, fully connected layers, and RNNs. They have predicted average yield for corn and soybean by using the RNN layer which has a time length of 5 years since they considered a 5-year yield dependencies.

An optimization model created in this research [9] is able to estimate crop yield by increasing remote sensing data availability. The optimization model was created by using a well-known algorithm, Trust-Region Methods for Nonlinear Minimization, that fits available data to an exponential equation. The experimental results proved the accuracy and reliability of the new model in estimating the crop yield in the case of missing remote sensing data (worst case scenario). The potato crop was used to prove the success of the objective of this research. It shows that the lack of remote sensing data is no longer an obstacle for managers and decision makers in the agriculture sector.

A pilot study [8] in India instructed farmers to delay planting by three weeks using predictive models based on climate and weather data and increased the yield by 30 percent. In the Indian state of Andhra Pradesh, they tested a new Sowing Application for farmers combined with a Personalized Village Advisory Dashboard, with the results showing a 30 percent greater average yield per hectare. During the pilot, farmers received ten sowing advisories containing important information such as sowing recommendations, seed treatment, optimum sowing depth, preventive weed management, land preparation, farmyard manure application, harvesting recommendations, shade drying of harvested pods, and storage.

[21] developed a scalable, accurate, and low-cost technique for predicting agricultural yields using pub-

licly available remote sensing data. They suggested a strategy based on contemporary representation learning ideas rather than the hand-crafted features commonly utilized in the remote sensing sector. They also presented a revolutionary dimensionality reduction strategy that enables them to train a Convolutional Neural Network or a Long-Short Term Memory network and learn valuable features automatically even when labeled training material is scarce. They added a Gaussian Process component to model the data's Spatio-temporal structure and increased accuracy explicitly.

2.3 Our Contributions

In most works in current literature, only certain subsets of crops have been taken into account while making recommendations. For example, data has been collected from specific districts of a particular state. The model can not be generalized as those parameters are missing and cannot be easily added. Furthermore, they recommend a single crop to the farmer based on the objective function. They also do not consider inter-cropping recommendation, and do not provide any portfolios of multiple crops after assessing risk to the farmer.

We now look at yield estimation which is a major component of recommendation. Most works only predicted the point or average yield estimates or predictions. They do not consider the variations in the yield, which is much required in agriculture as it helps uncover the risk in the yield. They also use a single regression model for all the crops, which may not capture all the characteristics of the specific crop.

In this project, we have taken the data for several districts of Uttar Pradesh so that our data contains variations in the features. It makes our yield prediction model more generalized. Our model calculates the distribution for yield in terms of expectation and standard deviation of yield. We would be able to consider the variations in yield and recommend the crops based on the risk profile of the farmer. We can use either Bayesian Neural Network [11], Mixed Density Distribution or deep Gaussian Mixture Models [19] for this task. Currently, we consider the triangular distribution for crop yield and other crop-related features by having the mode, maximum, and minimum values. Our system would recommend a single crop or a combination of crops to the farmer. For this purpose, we would require the yield data for the variety of crops and the domain knowledge that tells us which of the multiple crops can grow together. Our recommendation system provides different portfolios for the farmers to choose from and ranks them using the financial mathematical Sharpe ratio by considering their expected return and the stability in return.

Apart from this, our system would also consider the demand and supply of the market for the recommendation. The crop price prediction model takes these features into account, and our system would use this model such that the recommendation would have the demand and supply into consideration.

3 Problem Formulation

We aim to build a crop recommendation system that recommends the best suitable portfolio of crops to the farmers based on environmental conditions such as temperature, rainfall, sunlight, and humidity and soil characteristics such as soil type, pH-value, nitrogen, phosphorous, and potassium contents. We predict the yield for different crops with the help of these parameters and rank the portfolio of crops by calculating the farmer utilities such as profit, yield, and risk for different crops. We recommend the portfolio of crops in a specific ratio based on the risk profile of the farmers.

In this work, an important aspect is to predict the yield for different crops. Most papers in the literature compute the point estimate or average yield for other crops. They do not consider the variance associated with it. We take the triangular distribution by considering mode, maximum, and minimum values for the crop yield. This helps recommend the crop portfolio to the farmers based on their risk profile. We try to recommend a specific ratio (Crop Portfolio) in which a farmer should grow this combination of crops to optimize its utility.

4 Data Collection and Curation

Various datasets are used to predict the yield and recommend the best suitable portfolio of crops based on soil characteristics, climatic conditions, and geographical location. The datasets include:

- Yield Data : We are using the yield data provided by the VDSA, Icrisat. This data is uniform and available for various crops from 1966 to 2011. We are working on Uttar Pradesh yield data, and the granularity for this is district-wise.

  The agmarknet website provides the arrival data, which has the production quantity of different crops in the mandi on a particular data.

The granularity of this data is market-wise in every district. The information is abundant, but they have not provided the area (in hectares) from where this product has arrived. Either we could divide the area uniformly across these mandis or in the production ratio.

We have considered another data for yield available at [5], and its granularity is district-wise. It includes state, district, area in hectares, and production in tonnes. It contains the average yield of a crop for a particular year. The data is not uniform, and values are missing, so we could use interpolation or extrapolation to the yield.

- Weather Data : We are using the weather data from [3]. This data includes environmental parameters such as temperature, rainfall, humidity, and sunlight. The weather data at [3] is available for every 0.5-degree change in latitude and longitude. We have taken the weather data month-wise for Kharif and Rabi seasons separately.
- Soil Characteristic Data : This data is available at [2]. This parameter is the texture (or classification) of soil used by the land surface scheme of the ECMWF Integrated Forecasting System (IFS) to predict the water holding capacity of soil in soil moisture and runoff calculations. The data [2] is available for every 0.5-degree change in latitude and longitude. The Agro-ecological Subregions (AESR) index represents which type of land and soil is present in a particular region according to latitude and longitude.

These data are from the Indian Government websites. The yield data is from [1] [4], the price data is from [1] and the weather data is from [3].

5 Modelling and Methodology

The basic idea of this project is to collect the inputs from the farmer (direct or inferred inputs), which could be based on their traditional or cultural practices, and then predict the yield of the crop. The farmer can choose the utility (profit or risk), and we calculate the utility distribution and recommend the best suitable portfolio of crops based on the financial mathematical ratio considering expected return and the risk associated with it.

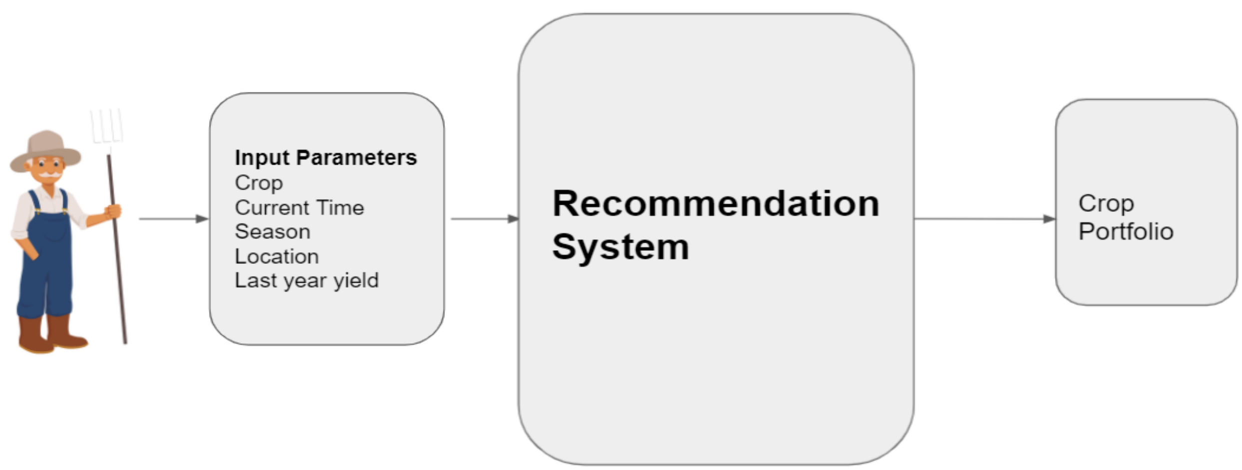

Figure 1: Farmer's point of view on the recommendation system

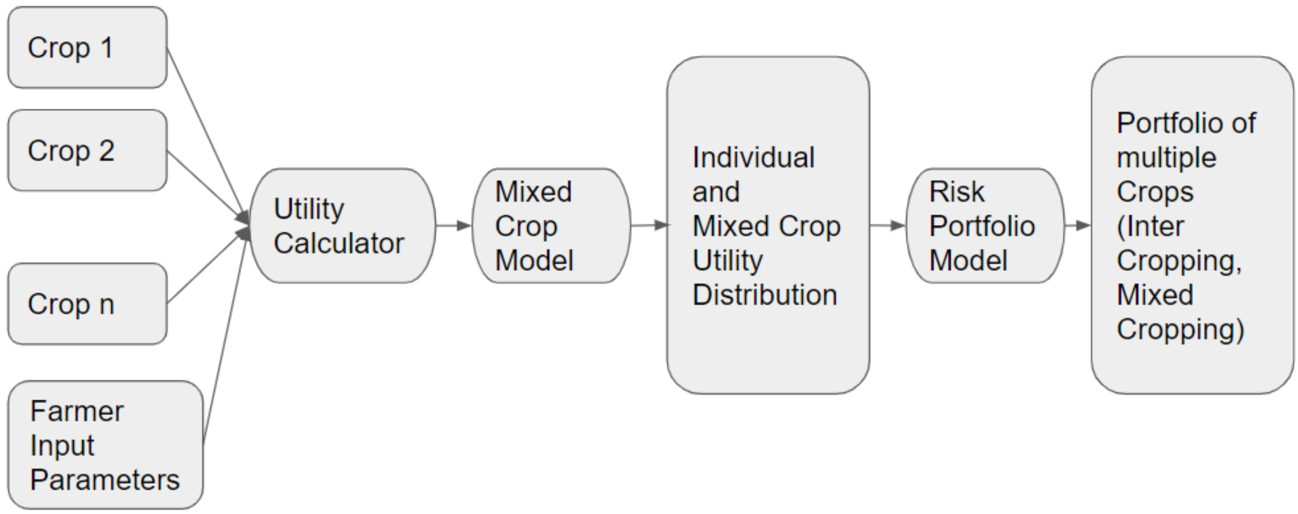

Figure 2: Internal Components of a Recommendation System

5.1 Overview of the System

Figure 1 shows how the farmer looks at the recommendation system. The farmer sees the recommendation system as the black box, which takes the input from the farmer and outputs a crop portfolio.

5.1.1 Input Parameters

- Direct Parameter : The farmer directly selects these parameters. These include State and District, Subset of Crops, Season (Kharif, Rabi, Whole Year), Utility Functions, Soil Type, Soil nutrients, irrigation, Facilities, Crop rotation cycle, Area(in hectares), human labor, etc.

- Inferred Parameter : These parameters can be derived from the direct parameters given by the farmer. These include Temperature, Rainfall, Sunlight, Humidity, Soil Type and Nutrients, and Crop Price. Our recommendation system infers these parameters.

5.1.2 Recommendation System

This system takes all the inputs and builds models such as yield prediction, price, cycle period, cost of production, and cultivation. These models are used to calculate the utility distribution (triangular distribution) based on the utility function chosen by the farmer.

5.1.3 Crop Portfolio

Based on the utility distribution and risk profile of the farmer, we would recommend a combination of crops in a specified ratio to the farmer. Cultural and Traditional Practices of the farmer can be taken into account in this portfolio of crops by asking the farmer which crops he is interested in growing. Also, we could take last year's yields for his crops and check whether our recommendation is aligned with his practices or not.

5.2 Recommendation System

The recommendation system would take either a subset of crops from the farmer or the crops from the block selected by the government. Figure 2 shows the internal components of the recommendation system. The utility calculator calculates the utility distribution in terms of yield and price for different crops given as input. To recommend the best suitable combination of crops, we require combined utility distribution for the crops. The Mixed Crop Model calculates the combined utility distribution based on the constraints of mixing the crops. There are some crops that cannot be grown together, so the mixed crop constraint take all these constraints into account.

The risk and portfolio crop model would take all these crops utility distribution and risk profile of the farmer and recommend the best suitable crop or combination of crops. The crop portfolio is a ratio vector that denotes how much ratio a particular crop should grow in one acre.

5.3 Utility Calculator

It calculates the utility distribution for the farmer based on the parameters given and the utility function chosen by the farmer. It would give us the expectation and standard deviation of the utilities used in recommending the best suitable portfolio of crops to the farmer. Figure 3 shows the basic architecture of the utility calculator.

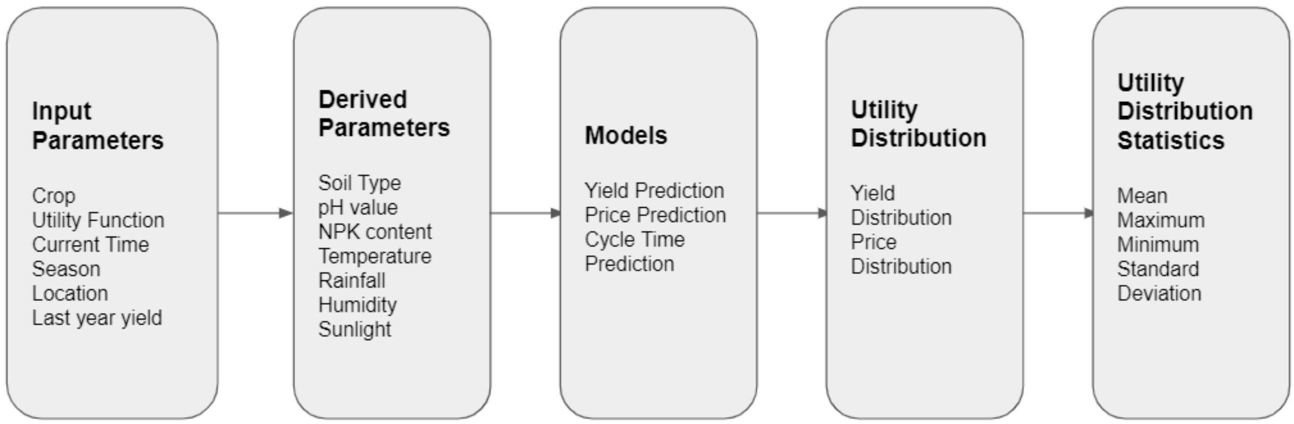

Figure 3: Basic Architecture of the Utility Calculator

5.3.1 Utility Functions

- **Profit (considering cycle time)**

  $$Utility = \frac{Y * (P - C_p) - C_c}{T}$$

  Here, Y denotes yield in Tonnes per Hectare, P denotes crop price in Rupees per Tonne, C_p denotes the cost of production in Rupees per Tonne, C_c denotes the cost of cultivation per Hectare, and T denotes cycle time for a crop in months.

  This utility function calculates the profit for the farmer on a particular crop. It helps maximize the farmer's profit by considering the crop logistic cost and cycle time (monthly). This utility function assumes that farmers could grow any crops with any combination. We would also consider the variance in the yield and price and generate the profit utility distribution for various crops.

- **Profit (without considering cycle time)**

  $$Utility = Y * (P - C_p) - C_c$$

  This utility function is not considering the cycle time because sometimes a farmer could grow only one crop in the entire year due to some scarcity of resources or some governing body's policies.

- **Yield** This utility function provides the yield per hectare. There are some scenarios where only the yield of a particular crop is essential. The farmer is not interested in profit for those crops because they grow them either to add nutrients to the soil or for their personal usage.

- **Risk** In agriculture, several risk factors affect the yield and price of the crop. Environmental conditions such as rainfall, temperature, sunlight, and humidity affect the yield of a crop, resulting in a loss for the farmer. Our Machine Learning models try to learn these factors and predict the yield. We have calculated the risk in profit by dividing the standard deviation by the expected profit value.

$$Risk = \frac{Standard\ Deviation\ in\ Profit}{Expected\ Profit\ Value}$$

5.3.2 Models

Most works in literature either predict the point estimate of crop yield or recommend the crops to maximize the objective function. In this project, we would calculate the utility distribution and find the variance in the yield and price. For this task, one way would be to either use Bayesian Neural Network [11] or Mixed Density Distribution, or deep Gaussian Mixture Models [19]. Currently, we are using triangular distribution by considering modal, maximum, and minimum values to find the variance in the yield and price of a crop.

6 Experiments and Results

In this project, we have worked on profit maximization utility for farmers. To obtain the profit utility distribution of a particular crop, we required yield, price, cost of cultivation, cost of production, and cycle time of a crop. Currently, we have considered the yield and price distribution of a crop. We have considered the price of those months in which the farmers harvest and sell a particular crop. We have obtained the price distribution using the price variations in these months. For getting the yield distribution of a crop, we have used the ensemble technique in which we have trained multiple Machine Learning models and took 10% deviation from the predicted yield to get maximum and minimum values.

6.1 Crop Yield Prediction

In this recommendation system, we have considered three crops for the Kharif season and three crops for the Rabi season. Rice, Maize, and Groundnut crops are for the Kharif season, and Wheat, Barley, and Masur Dal crops are for the Rabi season. We have taken several machine learning and deep learning regression models to predict yield, which are Polynomial Regression, Random Forest Regression, Deep Neural Network Regression (DNN), Convolutional Neural Network on Deep Neural Network Regression (CNN-DNN), Long Short Term Memory Regression (LSTM), and Convolutional Neural Network on Long Short Term Memory Regression (CNN-LSTM) where Convolutional Neural Network is used for extracting the embeddings for the features. The features used to predict the yield are latitude, longitude, weather parameters such as rainfall, temperature, sunlight, and humidity, and AESR index for soil and land characteristics.

The yield prediction regression model results are measured and compared on different metrics with each other. These metrics are Root Mean Square Error (RMSE), 95% Confidence Interval, 90% Confidence Interval, R^2 Score, Correlation Coefficient, and Coefficient of Variation. The Confidence Interval represents how much percentage of the predicted values are within this range of actual test values. The Correlation Coefficient represents how much the predicted and actual test values are related. The Coefficient of Variation is the ratio of the standard deviation of predicted values and the expectation of actual test values.

For training all the regression models, we have taken the yield data for Uttar Pradesh from the year 1982 to 2008 and for testing from the year 2009 to 2011. We have developed all the regression as mentioned above models for all the crops separately. The crop wheat is showing its stability in crop yield over the years, which helped the regression models to predict its yield very efficiently. All the regression models have better prediction results for crop wheat than other crops.

Table 1 compare all the metrics for all the regression models for the crop wheat. The Random Forest, DNN, and CNN-LSTM regression models perform better than other regression models. Still, the Random Forest and DNN models perform consistently better for all the other crops.

Similarly, we have developed and compared all the above regression models separately for the previously mentioned crops. Random Forest Regression Model performs the best compared to all other regression models for almost all the crops.

Table 2 is showing the results for Random Forest Regression models for all the other crops. The crops groundnut and maize are not performing as good as compared to the other crops because their yield varies a lot as compared to the other crops.

6.1.1 Ensemble Technique Results

Ensemble techniques are machine learning techniques that integrate numerous base models to create a single best-fit predictive model. Generally, Ensemble methods have higher predictive accuracy compared to the individual models.

We have used the ensemble technique in which we have taken all the combinations of Random Forest, DNN, CNN-DNN, and LSTM regression models for ensembling. For almost all the crops, the ensemble technique performs better than the individual models.

The Ensemble technique which consists of Random Forest and Deep Neural Network Regression Models is performing better than other ensemble technique for almost all the crops.

Regression Models	RMSE	95% Confidence Interval	90% Confidence Interval	R2 Score	Correlation Coefficient	Coefficient of Variation
Polynomial Regression	0.33	35.53	63.16	0.76	0.89	0.14
Random Forest	0.22	45.39	76.64	0.89	0.94	0.09
DNN	0.25	46.38	74.34	0.86	0.93	0.10
CNN-DNN	0.27	44.41	71.38	0.84	0.92	0.11
LSTM	0.21	44.69	72.67	0.89	0.95	0.09
CNN-LSTM	0.22	47.91	74.60	0.88	0.94	0.09

Table 1: The table shows the yield prediction results for the Wheat crop for all the regression models.

Crops	RMSE	95% Confidence Interval	90% Confidence Interval	R2 Score	Correlation Coefficient	Coefficient of Variation
Barley	0.28	31.35	54.13	0.81	0.9	0.16
Masur Dal	0.14	33.93	57.14	0.49	0.71	0.17
Rice	0.23	40.79	66.78	0.81	0.9	0.12
Maize	0.29	18.92	39.53	0.55	0.74	0.24
Groundnut	0.29	20.65	33.77	0.48	0.7	0.44

Table 2: The table shows the Random Forest Regression Model results for all the crops.

Models	RMSE	95 % Confidence Interval	90 % Confidence Interval	R2 Score	Correlation Coefficient	Coefficient of Variation
"['RF', 'DNN']"	0.13	66.15	89.74	0.96	0.98	0.06
"['RF', 'CNN-DNN']"	0.14	61.03	84.87	0.95	0.98	0.06
"['RF', 'LSTM']"	0.12	64.10	89.74	0.96	0.98	0.06
"['DNN', 'CNN-DNN']"	0.18	53.33	76.41	0.92	0.96	0.08
"['DNN', 'LSTM']"	0.15	53.85	82.56	0.94	0.97	0.07
"['CNN-DNN', 'LSTM']"	0.16	51.79	78.97	0.93	0.97	0.08
"['RF', 'DNN', 'CNN-DNN']"	0.14	59.74	84.87	0.95	0.97	0.07
"['RF', 'DNN', 'LSTM']"	0.13	62.56	88.21	0.96	0.98	0.06
"['RF', 'CNN-DNN', 'LSTM']"	0.13	61.28	85.38	0.95	0.98	0.06
"['DNN', 'CNN-DNN', 'LSTM']"	0.16	53.85	80.00	0.94	0.97	0.07
"['RF', 'DNN', 'CNN-DNN', 'LSTM']"	0.14	59.74	86.41	0.95	0.98	0.06

Table 3: The table shows the ensemble technique results for the crop Wheat.

Crops	RMSE	95% Confidence Interval	90% Confidence Interval	R2 Score	Correlation Coefficient	Coefficient of Variation
Barley	0.14	43.55	72.93	0.9	0.94	0.09
Masur Dal	0.14	32.73	53.57	0.43	0.66	0.17
Rice	0.13	52.82	84.35	0.92	0.96	0.07
Maize	0.14	33.34	62.69	0.78	0.9	0.13
Groundnut	0.28	12.45	30.81	0.49	0.7	0.43

Table 4: The table shows results for the ensemble technique consists of RF and DNN for all the crops.

Crop	Season	Predicted Profit	Maximum Profit	Minimum Profit	Actual Profit	Variance (Risk)
Rice	Kharif	2803.04	3397.28	-1190.22	3130.29	0.32
Maize	Kharif	2135.74	2410.5	1879.57	2446.84	0.04
Groundnut	Kharif	1734.28	16187.87	-1071.52	2004.84	1.88
Barley	Rabi	1929.79	6327.63	1346.01	1894.66	0.58
Wheat	Rabi	1873.97	2484.14	1128.7	1979.72	0.13
Masur Dal	Rabi	12695.37	13377.07	11812.59	14638.82	0.02

Table 5: The table shows the profit utility distribution for different crops for the year 2011.

Crop	Season	Predicted Profit	Maximum Profit	Minimum Profit	Actual Profit	Variance (Risk)
Rice	Kharif	3198.42	3775.29	2713.28	3927.83	0.06
Maize	Kharif	3075.17	3390.24	2777.18	4099.84	0.04
Groundnut	Kharif	1223.04	15836.24	-1724.26	1491.48	3.13
Barley	Rabi	1261.25	1657.19	594.94	1217.37	0.17
Wheat	Rabi	2329.54	3254.93	1741.26	2812.26	0.13
Masur Dal	Rabi	5841.83	6355.12	5228.05	14383.27	0.03

Table 6: The table shows the profit utility distribution for different crops for the year 2010.

Table 3 shows the different ensembles of several regression models and their performance for the crop wheat. The ensembles (RF, DNN), (RF, CNN-DNN), and (RF, LSTM) perform better.

Table 4 shows the results of the ensemble method consisting of (RF and DNN) for all the other crops.

6.2 Profit Utilities Results

We have experimented with three crops of Rabi and Kharif seasons each. We have calculated the profit utility distributions for the farmer which consists of average, maximum, minimum and actual profit of the crops. We have also calculated the coefficient of risk in profit by dividing the standard deviation in profit by expected profit value for the crops.

Table 5 shows the profit utility distributions for the Kharif and Rabi seasons for the year 2011. Here, the negative value represents that there are chances of incurring loss to the farmer if something terrible happens during the growth period of the crop, like the weather conditions, soil nutrients deficiency, etc. The maximum and minimum profit are calculated using the 10% deviation in predicted yield. The profit is calculated in rupees per hectare per month where different crops have different growth period duration. In the Kharif season, the crop groundnut has the maximum profit but also shows the highest deviation in profit, representing the highest risk. In contrast, the crop maize has the lowest maximum profit but the lowest risk. Similarly, in the Rabi season, the crop Masur Dal shows the highest maximum profit and also the lowest risk and turns out to be the best crop for Rabi season.

Table 6 shows the profit utility distributions for the Kharif and Rabi seasons for the year 2010. Here, again the crop groundnut has the highest maximum profit and risk, and the crop maize has the lowest maximum profit but is also the most stable crop in the Kharif season. In the Rabi season, the crop Masur Dal is again showing the highest maximum profit with the lowest risk as compared to all the other crops in the Rabi season.

6.3 Individual Crop Recommendation

Based on the profit utility distribution and the variance or risk associated with the profit, we have recommended the best suitable crop to the farmer for the particular season. There are some farmer who wants the crop which maximizes their profit irrespective of the risk associated with that crop whereas some farmers are risk averse who wants a stable return from the crop without any risk.

Year	Season	Maximizing Profit	Minimizing Risk
2011	Kharif	Groundnut	Maize
2011	Rabi	Masur Dal	Masur Dal
2010	Kharif	Groundnut	Maize
2010	Rabi	Masur Dal	Masur Dal

Table 7: The table shows the individual crop recommendation based on maximum profit and minimum risk.

Table 7 shows the one crop one season recommendation for the years 2010 and 2011 for Kharif and Rabi seasons.

6.4 Inter Crop Recommendation

In this experiment, we wanted to analyze the combination of crops in different ratios. We have used the Sharpe ratio to find which combination or portfolio of crops is best suitable for the farmers for a particular season.

Portfolios	Crops	Ratio 2009	Ratio 2010	Ratio 2011
”(0.0, 0.0, 1.0)”	”['Rice', 'Maize', 'Groundnut']”	0.00	0.00	0.00
”(0.0, 0.2, 0.8)”	”['Rice', 'Maize', 'Groundnut']”	0.40	0.74	0.73
”(0.0, 0.4, 0.6)”	”['Rice', 'Maize', 'Groundnut']”	1.07	1.97	1.96
”(0.0, 0.6, 0.4)”	”['Rice', 'Maize', 'Groundnut']”	2.40	4.42	4.40
”(0.0, 0.8, 0.2)”	”['Rice', 'Maize', 'Groundnut']”	6.35	11.70	11.65
”(0.0, 1.0, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	55.72	87.53	100.00
”(0.2, 0.0, 0.8)”	”['Rice', 'Maize', 'Groundnut']”	1.28	0.72	0.96
”(0.2, 0.2, 0.6)”	”['Rice', 'Maize', 'Groundnut']”	2.24	1.95	2.24
”(0.2, 0.4, 0.4)”	”['Rice', 'Maize', 'Groundnut']”	4.16	4.39	4.79
”(0.2, 0.6, 0.2)”	”['Rice', 'Maize', 'Groundnut']”	9.86	11.67	12.09
”(0.2, 0.8, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	89.91	100.00	50.88
”(0.4, 0.0, 0.6)”	”['Rice', 'Maize', 'Groundnut']”	3.41	1.92	2.49
”(0.4, 0.2, 0.4)”	”['Rice', 'Maize', 'Groundnut']”	5.91	4.36	5.05
”(0.4, 0.4, 0.2)”	”['Rice', 'Maize', 'Groundnut']”	13.30	11.57	11.65
”(0.4, 0.6, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	100.00	94.43	27.88
”(0.6, 0.0, 0.4)”	”['Rice', 'Maize', 'Groundnut']”	7.64	4.31	5.18
”(0.6, 0.2, 0.2)”	”['Rice', 'Maize', 'Groundnut']”	16.63	11.43	10.74
”(0.6, 0.4, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	94.52	77.08	19.13
”(0.8, 0.0, 0.2)”	”['Rice', 'Maize', 'Groundnut']”	19.81	11.23	9.72
”(0.8, 0.2, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	87.07	60.82	14.64
”(1.0, 0.0, 0.0)”	”['Rice', 'Maize', 'Groundnut']”	80.95	48.76	11.92
”(0.33, 0.33, 0.33)”	”['Rice', 'Maize', 'Groundnut']”	6.70	5.83	6.47
”(0, 0.5, 0.5)”	”['Rice', 'Maize', 'Groundnut']”	1.60	2.95	2.93
”(0.5, 0, 0.5)”	”['Rice', 'Maize', 'Groundnut']”	5.11	2.88	3.64
”(0.5, 0.5, 0)”	”['Rice', 'Maize', 'Groundnut']”	98.03	86.19	22.67

Table 8: The table shows the Sharpe ratio for different portfolio of crops for Kharif season for the year 2009, 2010, and 2011.

Portfolios	Crops	Ratio 2009	Ratio 2010	Ratio 2011
"(0.0, 0.0, 1.0)"	"['Barley', 'Wheat', 'Masur Dal']"	0.10	0.74	0.06
"(0.0, 0.2, 0.8)"	"['Barley', 'Wheat', 'Masur Dal']"	0.00	0.09	0.00
"(0.0, 0.4, 0.6)"	"['Barley', 'Wheat', 'Masur Dal']"	0.12	2.03	0.29
"(0.0, 0.6, 0.4)"	"['Barley', 'Wheat', 'Masur Dal']"	0.79	9.33	1.62
"(0.0, 0.8, 0.2)"	"['Barley', 'Wheat', 'Masur Dal']"	2.57	26.04	5.85
"(0.0, 1.0, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	6.84	62.23	22.31
"(0.2, 0.0, 0.8)"	"['Barley', 'Wheat', 'Masur Dal']"	1.76	0.43	1.32
"(0.2, 0.2, 0.6)"	"['Barley', 'Wheat', 'Masur Dal']"	2.30	0.00	1.96
"(0.2, 0.4, 0.4)"	"['Barley', 'Wheat', 'Masur Dal']"	3.43	4.30	3.65
"(0.2, 0.6, 0.2)"	"['Barley', 'Wheat', 'Masur Dal']"	5.81	18.67	8.25
"(0.2, 0.8, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	11.50	54.57	25.70
"(0.4, 0.0, 0.6)"	"['Barley', 'Wheat', 'Masur Dal']"	6.55	1.58	5.96
"(0.4, 0.2, 0.4)"	"['Barley', 'Wheat', 'Masur Dal']"	8.92	2.81	9.07
"(0.4, 0.4, 0.2)"	"['Barley', 'Wheat', 'Masur Dal']"	13.37	13.71	16.16
"(0.4, 0.6, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	23.88	48.70	40.91
"(0.6, 0.0, 0.4)"	"['Barley', 'Wheat', 'Masur Dal']"	15.46	7.19	15.25
"(0.6, 0.2, 0.2)"	"['Barley', 'Wheat', 'Masur Dal']"	22.93	15.06	25.44
"(0.6, 0.4, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	41.53	49.07	59.80
"(0.8, 0.0, 0.2)"	"['Barley', 'Wheat', 'Masur Dal']"	34.46	25.70	35.11
"(0.8, 0.2, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	65.72	63.33	79.75
"(1.0, 0.0, 0.0)"	"['Barley', 'Wheat', 'Masur Dal']"	100.00	100.00	100.00
"(0.33, 0.33, 0.33)"	"['Barley', 'Wheat', 'Masur Dal']"	7.89	5.26	8.57
"(0, 0.5, 0.5)"	"['Barley', 'Wheat', 'Masur Dal']"	0.36	4.81	0.76
"(0.5, 0, 0.5)"	"['Barley', 'Wheat', 'Masur Dal']"	10.30	3.51	9.84
"(0.5, 0.5, 0)"	"['Barley', 'Wheat', 'Masur Dal']"	32.02	47.69	50.13

Table 9: The table shows the Sharpe ratio for different portfolio of crops for Rabi season for the year 2009, 2010, and 2011.

6.4.1 Sharpe Ratio

In finance, the Sharpe ratio measures the performance of an investment such as security or portfolio compared to a risk-free asset after adjusting for its risk. It is calculated by dividing the difference between the investment's returns and the risk-free return by the standard deviation of the investment returns.

Here, we have used the Sharpe ratio on profit maximization and it is calculated by dividing the expected profit value by the standard deviation in the profit.

We have scaled the values of the Sharpe ratio for different portfolios from 0 to 100, where 0 is the lowest and 100 is the highest. There are four possible extreme categories for a portfolio. A portfolio could either have the highest return and least risk or lowest return and least risk or highest return and highest risk or lowest return and highest risk. The former two categories are favorable for small-scale marginal farmers. The former two categories would take the highest value for the Sharpe ratio, and the latter two would take the lowest value.

Table 8 shows the Sharpe ratios for different portfolios of crops for the Kharif season for the years 2009, 2010 and 2011. The portfolio of Rice, Maize, and Groundnut with the combination of (0.2, 0.8, 0.0) for the Kharif season is giving the highest Sharpe ratio for the year 2010 and with (0.0, 1.0, 0.0) for the year 2011.

Table 9 shows the Sharpe ratios for different portfolios of crops for the Rabi season for the years 2009, 2010 and 2011. The portfolio of Barley, Wheat, and Masur Dal with the combination of (1.0, 0.0, 0.0) for the Rabi season is giving the highest Sharpe ratio for the year 2010 and with (1.0, 0.0, 0.0) for the year 2011 also.

7 Conclusion

This project recommends the best suitable portfolio of crops by calculating the profit utility distribution based on the crop characteristics such as yield, price, cost of cultivation and production, and crop cycle time. We have predicted the crop yield using the ensemble technique, consisting of different Machine learning and Deep learning regression models such as Random Forest and DNN. The Utility Calculator provides the utility distribution by calculating and merging the yield and price distribution of the crop.

The individual crop recommendation is performed using profit maximization and risk minimization because there are farmers who want the crop with a maximum profit irrespective of the risk associated with it. Also, there is a risk-averse farmer who wants a stable return from the harvest without any risk. We have used the Sharpe ratio to perform the inter-crop recommendation to consider these factors together. In this project, we have considered the small-scale marginal farmers because they constitute many. The marginal farmers are risk-averse, so we used the Sharpe ratio, which gives the portfolio the maximum profit possible with the least risk possible. In this way, we have recommended the best suitable crop portfolio based on the farmer's risk profile.

8 Future Work

In this project, we have worked with three crops for each season and analyzed the inter-cropping profit values of these crops. We would try to work with more crops and make our system more generalized in the future. We would also try to add more parameters to make our system more robust, for example, variations in temperature and rainfall, fertilizers data for yield prediction, etc. We would also need to improve the yield prediction models. We have considered the yield data till the year 2011, and for the state Uttar Pradesh, we would like to increase this data till the year 2022 and for different states of India. Currently, this recommendation system works for the Uttar Pradesh state. In the later part of this project, this recommendation system would work for different states in India.

In the future work for this project, we would recommend a mixed crop strategy to the farmer by applying the mixed crop constraints and models by using expert advice on how the soil and yield are affected by the growth of particular crops nearby. Currently, we are not using the price prediction model, but we would like to include it later in this recommendation system. We would also try to improve the cost of cultivation and production data by directly interacting with the farmer and developing synthesized data representing more realistic characteristics. We would like to include other financial mathematical ratios considering other factors like maximum acceptable return, etc., for recommending a portfolio. Our main aim would be to recommend a non-trivial portfolio of crops to the farmer. Currently, the recommendation system is optimizing the utilities for a particular season. Still, in the future, we would like to optimize the utilities for the whole year by considering the summer season and perennial crops.

References

[1] "https://agmarknet.gov.in/."

[2] "https://cds.climate.copernicus.eu/cdsapp/dataset/reanalysis-era5-single-levels?tab=form."

[3] "https://climate.copernicus.eu/climate-data-store/."

[4] "https://data.gov.in/."

[5] "https://data.gov.in/sector/agriculture."

[6] "http://www.ibef.org/industry/agriculture-india.aspx."

[7] "http://www.india.gov.in/topics/agriculture."

[8] ICRISAT, "Microsoft and icrisat's intelligent cloud pilot for agriculture in andhra pradesh increase crop yield for farmers," 2017.

[9] M. M. Awad, "Toward precision in crop yield estimation using remote sensing and optimization techniques," *Agriculture*, vol. 9, no. 3, p. 54, 2019.

[10] J. Fan, J. Bai, Z. Li, A. Ortiz-Bobea, and C. P. Gomes, "A gnn-rnn approach for harnessing geospatial and temporal information: Application to crop yield prediction," *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI*, 2022.

[11] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, "Hands-on bayesian neural networks—a tutorial for deep learning users," *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 29–48, 2022.

[12] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in plant science*, vol. 10, p. 621, 2019.

[13] S. Khaki, L. Wang, and S. V. Archontoulis, "A cnn-rnn framework for crop yield prediction," *Frontiers in Plant Science*, vol. 10, p. 1750, 2020.

[14] J. Madhuri and M. Indiramma, "Artificial neural networks based integrated crop recommendation system using soil and climatic parameters," *Indian Journal of Science and Technology*, vol. 14, no. 19, pp. 1587–1597, 2021.

[15] A. Priyadharshini, S. Chakraborty, A. Kumar, and O. R. Pooniwala, "Intelligent crop recommendation system using machine learning," *Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021) IEEE Xplore Part Number: CFP21K25-ART*, 2021.

[16] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," *2016 Eighth International Conference on Advanced Computing (ICoAC), IEEE*, pp. 32–36, 2017.

[17] M. T. Shakoor, K. Rahman, S. N. Rayta, and A. Chakrabarty, "Agricultural production output prediction using supervised machine learning techniques," *2017 1st international conference on next generation computing applications (NextComp), IEEE*, pp. 182–187, 2017.

[18] Z. R. Shi, C. Wang, and F. Fang, "Artificial intelligence for social good: A survey," *arXiv preprint arXiv:2001.01818*, 2020.

[19] C. Viroli and G. J. McLachlan, "Deep gaussian mixture models," *Statistics and Computing*, vol. 29, no. 1, pp. 43–51, 2019.

[20] C. Von Lücken and R. Brunelli, "Crops selection for optimal soil planning using multiobjective evolutionary algorithms," *Proceedings of the Twentieth Innovative Applications of Artificial Intelligence Conference*, pp. 1751–1756, 2008.

[21] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," *Proceedings of the Twentieth Innovative Applications of Artificial Intelligence Conference*, pp. 1751–1756, 2008.