ISSN: 2393-8528

Contents lists available at www.ijicse.in

International Journal of Innovative Computer Science & Engineering

Volume 4 Issue 3; May-June-2017; Page No. 01-10

A MACHINE TAKING IN POINT OF VIEW FOR FORESEEING AGRICOLA DROUGHTS.

G.Priyadarshini

M.Sc., M.Phil., Department of Computer Application

KG COLLEGE OF ARTS AND SCIENCE

savidpri@yahoo.com

Received 20 March 2017; Accepted 02 May. 2017

ABSTRACT

Dry season influences an extensive number of individuals and make more misfortune society contrasted with other cataclysmic events. Tamilnadu is a dry season catastrophe inclined state in India. Information mining in agribusiness is an exceptionally late research point. This method is utilized for agribusiness in information mining. A related, however not proportionate term is exactness horticulture. The regular event of dry spell gangs an undeniably extreme risk to the Tamilnadu agrarian creation. Dry spell likewise has exceptionally complex wonder that is difficult to precisely measure since it's monstrous spatial and fleeting fluctuation. In the Existing framework, ISDI demonstrate development was actualized for assessing the exactness and the viability. This model application utilizing an assortment of strategies and information, there is still some work to be done in our future research as a result of the complex spatial and fleeting attributes of dry spell. To beat impersonation, it surveys execution of measuring the dry season by utilizing the Spatial and worldly qualities of information's. The dataset is gathered from various areas and furthermore gathers the time differing data's from the dataset. In this, the dry season conditions were anticipated by utilizing the managed learning system. It can be actualized by utilizing the Bayesian administered machine learning calculation. Through this venture exactness and execution can be accomplished, and furthermore execution and adequacy can be made strides.

Keywords: ISDI model, Bayesian supervised machine learning algorithm, WEKA, EM Algorithm.

Introduction

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The data sources can include databases, data warehouses, the web, other information repositories, or data that are streamed into the system dynamically.

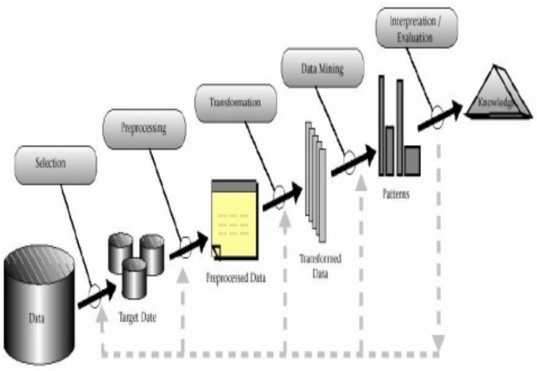

Fig 1: Knowledge extraction process

A. OVERVIEW OF THE PROJECT

Information mining in farming is an extremely late research subject. It comprises in the utilization of information mining procedures to agribusiness. Late innovations are these days ready to give a considerable measure of data on agrarian related exercises, which can then be investigated with a

*Corresponding author: G.Priyadarshini | E-mail: savidpri@yahoo.com

specific end goal to discover vital data. A related, however not proportionate term is exactness horticulture.

Wine is broadly created all around the globe. The maturation procedure of the wine is vital, on the grounds that it can affect the efficiency of wine-related enterprises and furthermore the nature of wine. In the event that it can anticipate how the maturation will be at the early phases of the procedure, so as to ensure a normal and smooth aging, the procedure could be interfaced. Maturations are these days contemplated by utilizing diverse procedures, such as, the k-implies calculation, and a method for grouping in light of the idea of biclustering. Take note of that these works are not quite the same as the ones where a grouping of various types of wine is performed. See the wiki page Classification of wine for more points of interest.

The location of creature's sicknesses in homesteads can affect emphatically the efficiency of the ranch, since debilitated creatures can bring about defilements. Additionally, the early location of the maladies can permit the rancher to cure the creature when the illness shows up. Sounds issued by pigs can be examined for the discovery of ailments. Specifically, their hacks can be contemplated, in light of the fact that they show their affliction. A computational framework is being worked on which can screen pig sounds by mouthpieces introduced in the ranch, and which is additionally ready to segregate among the distinctive sounds that can be distinguished.

Before going to market, apples are checked and the ones demonstrating a few deformities are evacuated. Be that as it may, there are additionally undetectable imperfections that can ruin the apple flavor and look. A case of undetectable deformity is the water centers. This is an inside apple issue that can influence the life span of the organic product. Apples with slight or mellow water centers are sweeter; however apples with direct to serious level of water centers can't be put away for any time span. In addition, a couple natural products with extreme water centers could ruin an entire cluster of apples. Therefore, a computational framework is under review which takes X-beam photos of the organic product while they keep running on transport lines, and which is additionally ready to break down (by information mining methods) the taken pictures and gauge the likelihood that the natural product contains water centers.

Late reviews by agribusiness scientists in Pakistan (one of the main four cotton makers of the world) demonstrated that endeavors of cotton product yield augmentation through professional pesticide state approaches have prompted a perilously high pesticide utilize. These reviews have detailed a negative relationship between's pesticide utilize and edit yield in Pakistan. Thus over the top utilize (or manhandle) of pesticides is hurting the ranchers with unfavorable money related, ecological and social effects. By information mining the cotton Pest Scouting information alongside the meteorological recordings it was demonstrated that how pesticide utilize can be streamlined (diminished). Bunching of information uncovered fascinating examples of agriculturist practices alongside pesticide utilize elements and henceforth help distinguish the purposes behind this pesticide manhandle.

To screen cotton development, distinctive government offices and organizations in Pakistan have been recording irritation scouting, farming and metrological information for quite a long time. Coarse assessments of simply the cotton bug scouting information recorded stands at around 1.5 million records, and developing. The essential agro-met information recorded has never been digitized, incorporated or institutionalized to give a total picture, and consequently can't bolster basic leadership, in this manner requiring an Agriculture Data Warehouse. Making a novel Pilot Agriculture Extension Data Warehouse completed by investigation questioning and information mining some intriguing disclosures were made, for example, pesticides showered at the wrong time, wrong pesticides utilized for the correct reasons and transient connection between pesticide use and day of the week.

Dry season influences an expansive number of individuals and make more misfortunes society contrasted with other cataclysmic events. Tamilnadu is a dry season calamity inclined nation. The regular event of dry spell represents an inexorably extreme danger to the Tamilnadu farming creation. Dry spell additionally has extremely complex marvel that is difficult to precisely measure since it's enormous spatial and transient changeability. Along these lines, the reliable meaning of dry season is test. Most researchers considered that dry spell can be characterized into three classifications: meteorological, horticultural, and hydrological dry season. The meanings of horticultural dry spell concentrated on the dampness supply of a specific district underneath the perpetual typical level so that the yield creation or range profitability is antagonistically influenced. It includes an assortment of meteorological dry spell attributes (e.g., the shortage of precipitation, genuine evapotranspiration at just a little part of the potential evapotranspiration rate, and the lack of soil dampness) affect on farming (e.g., yield decrease). Subsequently, the horticultural dry spell condition is influenced by many elements, for example, precipitation, soil dampness, temperature, vegetation sort, soil sort and phenology. In light of the characterized dry season criteria, the power, fleeting and spatial appropriation of farming dry spell can be checked utilizing a dry season file. Rural dry spell lists utilize strategies for scientific displaying to change the elements influencing crop development and improvement into a solitary number and give a far reaching dry season depiction for basic leadership in the farming division.

The initially created files are meteorological dry season records, for example, Standard Precipitation Index (SPI) and Palmer Drought Severity Index (PDSI). The SPI was one of the dry spell lists been broadly utilized around the world, which is intended to be a spatially invariant marker (spatially and transiently equivalent) just in light of in-situ precipitation information. The PDSI is figured utilizing verifiable temperature and precipitation, and data of the accessible water con-tent of the dirt in light of a dirt dampness/water adjust condition. In spite of the fact that the meteorological dry spell files can get more precise and spatially and transiently practically identical dry season conditions, their use is oppressed to the thickness and dissemination of the station organize. This sort of records likewise can't mirror the vegetation condition actuated by the water shortfall.

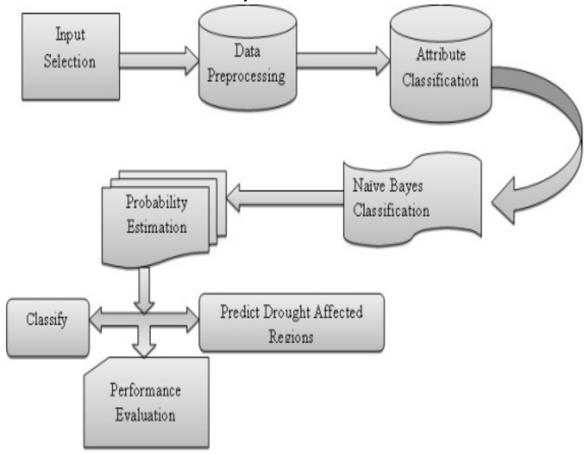

Fig 2: Overview of the project

B. OBJECTIVE OF THE PROJECT

Agricultural drought generally occurred after a period of time since the cessation of rain, when the available stored water will support the evapotranspiration. It involves a variety of meteorological drought characteristics and impact on agriculture. Therefore, the agricultural drought condition is affected by many factors, such as precipitation, soil moisture, temperature, vegetation type, soil type and phenology. Agricultural drought indices use methods of mathematical modeling to transform the factors affecting crop growth and development into a single number and give a comprehensive drought description for decision making in the agricultural sector.

The first developed indices are meteorological drought indices such as standard precipitation index. SPI was one of the drought indices been widely used worldwide, which is designed to be a spatially invariant indicator only based on in-situ precipitation data. The PDSI is calculated using historical temperature and precipitation, and the information of the available water content of the soil based on the soil moisture/water balance equation. Although the meteorological drought indices can get more accurate and spatially and temporally comparable drought conditions, their utilization is enslaved to the density and distribution of the station network.

This type of indices also can not reflect the vegetation condition induced by the water deficit. Satellite based drought indices have obvious advantages compared to station based meteorological drought indices in spatial resolution. The normalized difference vegetation index is one of the early developed and most widely used satellite-based indices. NDVI is the best indicator of the vegetation growth conditions and the vegetation coverage, which has been widely used to estimate vegetation biomass and assess the environmental condition.

Percent of average seasonal greenness is one of the typical drought indexes which is established based on NDVI. PASG provides a measure of vegetation conditions by calculating the percentage between the greenness in the specific period and the average greenness over the same period. Land surface temperature is also closely related with drought. In addition to temperature increases caused by water deficit, the vegetation index decreases along with the vegetation growth status suppressed. Vegetation supply water index have been successfully used to detect drought. This method provides a rapid means to assess drought conditions compared to other remote sensed drought indices for that the strong negative correlation between NDVI and LST is largely due to the changes of vegetation cover and soil moisture. Therefore, no single drought index can be used to adequately monitor the onset of drought and measure drought intensity, duration and impact.

Vegetation drought response index was developed based on data mining technology namely decision making regression tree model. VegDRI can integrate station based meteorological drought indices, satellite-based regional drought indices, as well as biophysical data. The accuracy of VegDRI has been validated by the field observation data, and this index has been used to national wide drought monitoring in America.

Integrated surface drought index based on the concept of VegDRI. ISDI integrates land surface water and thermal environment condition, vegetation growth condition and biophysical information. ISDI includes climate based drought index information, satellite- based drought indices related to vegetation condition and soil moisture content, and several biophysical characteristics used to reflect regional difference of drought.

Drought has many causes. It can be caused by not receiving rain or snow over a period of time. The discussion about the water cycle and weather that changes in the wind patterns that move clouds and moisture through the atmosphere can cause a place to not receive its normal amount of rain or snow over a long period of time. If people live in a place where most of the water comes from a river, a drought in any area can be caused by places upstream is not receiving enough moisture. There would be less water in the river, people who live along the river to use. People can also play a big role in drought. Use of too much water during times of normal rainfall, there may be not enough water when a drought happens.

Drought Risk Management Photo: UNDP People who live in dry areas are vulnerable to disasters of various kinds. They are subject to recurrent droughts, and when the rains come, they are often affected by serious floods. In the past, crisis preparedness and management often focused on man-made disasters and acute natural disasters. Recurrent exposure to natural hazards, especially drought, has been largely ignored. This is changing, and UNDP-DDC is helping the change to come about. Drought risk management (DRM) is the concept and practice to avoid, lessen or transfer the adverse effects of drought hazards and the potential impacts of disaster through activities and measures for prevention, mitigation and preparedness.

It is a systematic process of using administrative directives, organizations and operational skills and capacities to implement strategies, policies and improving coping capacities. In recognition of the multiplicity of drought challenges in the context of uncertainties surrounding climate change, UNDP-DDC is focusing on building long-term resilience to climate shocks and change as well as mitigating immediate disaster risks and impacts.

II. SYSTEM ANALYSIS

A. EXISTING SYSTEM

Using data-mining technology, this paper established a new method, named the Integrated Surface Drought Index (ISDI). ISDI integrates traditional meteorological data, remotely sensed indices, and biophysical data, and attempt to describe drought from a more comprehensive perspective. The evaluation results indicated that the construction models for three phases of growth season have very high regression accuracy. The drought condition can be predicted using the independent variables. The practical application of ISDI in Tamilnadu also demonstrated that ISDI has good application accuracy in mid-eastern China. ISDI results were corresponding to the disaster observation records of agro-meteorological sites.

B. PROPOSED SYSTEM

In this project, integrate multi-source information is used to achieve the purpose of accurately monitoring agricultural drought. Agricultural drought generally occurred after a period of time since the cessation of rain, when the available stored water will support the actual evapotranspiration. It involves a variety of meteorological drought characteristics (e.g., the scarcity of precipitation, actual evapotranspiration at only a small fraction of the potential evapotranspiration rate, and the shortage of soil moisture) impact on agriculture (e.g., yield reduction). Therefore, the agricultural drought condition is affected by many factors, such as precipitation, soil moisture, temperature, vegetation type, soil type and phenology. Based on the defined drought criteria, the intensity, temporal and spatial distribution of agricultural drought can be monitored. By using this type of attributes, we are going to predict the drought conditions by using the supervised machine learning methods.

Fig 3: The building flowchart of ISDI based on the Supervised Classification and Regression-Tree (CART) model.

- ISDI Inputs Calculation

Rigorous pre-treatment has been applied to meteorological data before they are used to build SPI and PDSI. The sites with more than one year missing records were abandoned. The short term missing records were replaced by linearly interpolated values at the corresponding period of the adjacent two years. SPI provides a measure of precipitation deficit compared to historical precipitation record. PDSI accounts for the effect of both precipitation and temperature and their combine effect on soil water available to drought conditions.

The independent variables in the original VegDRI model include PASG, start of season anomaly (SOSA), Land Cover, Global Irrigation Area Map (GIAM), and Eco-region data.

Ten years (2000-2009) long 16-day interval PASG and yearly SOSA series were obtained based on the MODIS NDVI products.

- ISDI Construction and Cross-Validation Method

For the same vegetation type, the vulnerability under drought condition is significant in different phases of growth season. Therefore, we divide the growing season into three seasonal phases, namely spring, summer, and fall, and ISDI was constructed for the three phases separately. According to India meteorological four seasonal division methods the spring is generally considered from March to May, June to August for the summer, and from September to November for fall.

A commercial Supervised Classification and Regression-Tree (CART) algorithm called cubist 2.07 was used in this investigation to analyze historical drought indices and biophysical variables and build the three seasonal, rule-based, linear regression models. This approach can handle a variety of data types (e.g., discrete and continuous), so it suitable for constructing ISDI which integrates site-scale meteorological drought indices and regional scale remote sensed data. The inputs of the cubist model are the three seasonal historical variables records.

Table 1: Equations and references of indices as inputs for vegdri

Drought indices	Formula	Source and Reference
Palmer Drought Severity Index(PDSI)	PDSI=dK D = P - $\hat{P}$	Palmer,1965
Standard Precipitation Index(SPI)	F $(x<x_0)=\int_0^{x_0} f(x)dx$ F(x=0)=m/n	McKee etal.,1993
Percent of Average Seasonal Greenness (PASG)	$PASG_{PnYn}$ = $(SG_{PnYn}/xSG_{Pn)}$ * 100 SG = $\int_{P_1=SOS}^{P_n=EOS} NDVI$	Brown etal., 2008
Start of Season Anomaly (SOSA)	SOSA=SOST-SOST	Brown etal., 2008
Vegetation Supply Water Index(VSWI)	$VSWI_{ij}$= $NDVI_{ij}$/ LST_{ij}	McVicar and Bierwirth, 2001;

The three seasonal ISDI models are then used to calculate the spatial drought intensity results in Tamilnadu through 2000 to 2009. Before this, the site-scale SPI variable was interpolated to 1 km raster images using the Inverse Distance Weighting (IDW) method. If the conditions of a certain pixel are met one or more rules, then use the associated linear regression models to calculate the averaged ISDI value. From the perspective of construction principle, the new established drought monitoring method ISDI is a regressed and predicted PDSI using the optimum variables. Therefore, ISDI has the same value range (6 to 6, typically 4 to 4) and drought degree classification as PDSI.

The f-fold cross-validation approach integrated in the software was used to evaluate the statistical accuracy of the three seasonal models. The cross-validation is an iterative process. In each cycle computing, the 10 years historical records were divided into 10 blocks of roughly the same size, one block for test, and the remaining blocks as constructing cases. The accuracy of constructed model is estimated by averaging results on the hold-out cases.

Table 2: Drought degree classification of ci and isdi

Grades	Drought	CI	ISDI
1	Normal	-0.6 < CI	-1 < ISDI ≤ 1
2	Milds	-1.2 < CI ≤ -0.6	-2 < ISDI ≤ -1
3	Moderate	-1.8 < CI ≤ -1.2	-3 < ISDI ≤ -2
4	Sever	-2.4 < CI ≤ -1.8	-4 < ISDI ≤ -3
5	Extreme	CI ≤ -2.4	ISDI ≤ -4

ISDI Inputs Calculation:

The relative error is the ratio of average error magnitude to the error magnitude that would result from always predicting the mean value. The smaller the values indicate that the higher accuracy of the models and the values less than 1 proved the models are useful.

Practical Application of ISDI:

The generated three seasonal models are applied to the gridded image input data for each 16-day period to calculate the series of 1 km ISDI maps across the 2000–2009 growing season in Tamilnadu. The site scale SPI series was interpolated with 1 km spatial resolution using the IDW method in ArcGIS software. In the process of calculation, the values of all the independent variables for each pixel were considered to determine which rule with corresponding linear regression equation should be applied to calculate the ISDI value.

III. MODULE DESCRIPTION

A. Dataset collection & pre-processing

For the agricultural drought monitoring, the main factors to reflect the drought degree include the precipitation, soil moisture, vegetation condition, as well as drought onset season and so on. The information used to generate the new drought index (ISDI) includes climate-based drought index information, satellite-based drought indices related to vegetation condition and soil moisture content, and several biophysical characteristics used to reflect regional difference of drought.

Meteorological Data

The basic meteorological data set used in the ISDI contains50 years (1960–2009) long daily series of air temperature and atmospheric precipitation at 130 weather stations in Tamilnadu. All the data set are derived from the China Meteorological Data Sharing Service System.

MODIS Data

The MODIS Terra data set was used to calculate the regional remote sensed drought indices. High quality, consistent and well-calibrated MODIS data set provides a means for quantifying land surface characteristics such as NDVI, LST, and land cover type. The 16-day interval MODIS NDVI products (MOD13A2, 1 km 1 km) were used to monitor vegetation dynamics. The LST data is derived from8-day interval MODIS products (MOD11A2, 1 km 1 km).The NDVI and LST products are both from 2000 to 2009.MODIS land cover data (MOD12Q2, 1 km 1 km) for 2008was used to distinguish the drought characteristic differences in various land cover types.

Biophysical Data

The biophysical data inputs for ISDI include ecological zoning data (Eco-region), Available Water-holding Capacity(AWC) data, and the irrigation water management distribution data and Digital Elevation Data (DEM). The Eco-region data is obtained from a Chinese eco-geographical zoning map. This map was transformed from vector type to grid type with1 km spatial resolution before input into the ISDI. AWC is provided by the International Geo-sphere-Biosphere Programme (IGBP). The original AWC map was interpolated from 10 km spatial resolution to 1 km resolution using the K ringing Interpolation Method provided by ArcGIS software.

B. Attribute classification

WEKA's pre-processing capability is encapsulated in an extensive set of routines, called filters that enable data to be processed at the instance and attribute value levels. These filters have a standard command-line interface with a set of common command-line options. Many of the filter algorithms provide facilities for general manipulation of attributes for example, to insert and delete attributes from the dataset. When experimenting with learning schemes in the development of a data mining application, one of the most common activities involves building models with different subsets of the complete attribute set. WEKA provides three feature selection systems to aid in choosing attributes for inclusion in an experiment: a locally produced correlation based technique (Hall and Smith, 1998); the wrapper method (John and Kohavi, 1997); and Relief (Kira and Rendell,1992).In some cases it can be beneficial to apply a transformation function to an entire column in the dataset—for example, to normalize each value in an attribute.

- **Classification**

Classification is a form of data analysis that extracts models describing important data classes. Such models, called classifiers, predict categorical class labels. Classification has numerous applications, including fraud detection, target marketing, performance prediction, manufacturing and medical diagnosis. Classification is a two step process, consisting of a learning step and a classification step. In the first step, a classifier is built describing a predetermined set of data classes or concepts. This is the learning step where a classification algorithm builds the classifier by analyzing or learning form.

Machine Learning Techniques

Classifiers

The output from this type of learning scheme is, literally, a classifier usually in the form of a decision tree or set of rules that can be used to predict the classification of a new data instance. One attribute in the input table is designated as the category or class for prediction; the rest of the attributes may appear in the "if" portions of the rules (or the non-leaf nodes of the decision tree).The most primitive learning scheme in WEKA is Zero R. This scheme models the dataset with a single rule. Given a new data item for classification, Zero R predicts the most frequent category value in the training data for problems with a nominal class value, or predicts the average class value for numeric prediction problems. Zero R is useful for generating a baseline performance that other learning schemes are compared to. In some datasets, it is possible for other learning schemes to induce models that perform worse on new data than Zero R—an indicator of substantial over fitting.

The next scheme, One R, produces very simple rules based on a single attribute. One R is also useful in generating a baseline for classification performance indeed, this algorithm was found to perform as well as more sophisticated algorithms over many of the standard machine learning test datasets (Holte, 1993)! It appears that at least part of the reason for this result is that many of the standard test databases embody very simple underlying relationships in the data. Real world databases may also contain very simply structured information about a domain as well, and these simple relationships can be parsimoniously detected and represented by One R

Naive Bayes

Implements a Naïve Bayesian classifier, which produces probabilistic rules—that is, when presented with a new data item, the Naive Bayes model indicates the probability that this item belongs to each of the possible class categories (Langley et al,1992). The Bayesian classifier is 'naïve' in the sense that attributes are treated as though they are completely independent, and as if each attribute contributes equally to the model.

If extraneous attributes are included in the dataset, then those attributes will skew the model. Despite its simplicity, Naive Bayes, like One R, can give surprisingly good results on many real world datasets.

Decision Table

Decision Table summarizes the dataset with a 'decision table'. In its simplest state, a decision table contains the same number of attributes as the original dataset, and a new data item is assigned a category by finding the line in the decision table that matches the non-class values of the data item. This implementation employs the wrapper method (JohnandKohavi, 1997) to find a good subset of attributes for inclusion in the table. By eliminating attributes that contribute little or nothing to a model of the dataset, the algorithm reduces the likelihood of over-fitting and creates a smaller, more condensed decision table. Instance-based learning schemes create a model by simply storing the dataset. A new data item is classified by comparing it with these 'memorized' data items, using a distance metric. The new item is assigned the category of the closest original data item (its 'nearest neighbor'). Alternatively, the majority class of the k nearest data items may be selected, or for numeric attributes the distance-weighted average of the k closest items may be assigned.

SMO

The "sequential minimal optimization" algorithm for support vector machines (SVMs), which are an important new paradigm in machine learning (Burgess,1998). SVMs have seen significant application in learning models for text categorization. While SMO is one of the fastest techniques for learning SVMs, it is often slow to converge to a solution—particularly with noisy data. WEKA contains three methods for numeric prediction. The simplest is Linear Regression. LWR is an implementation of a more sophisticated learning scheme for numeric prediction, using locally weighted regression (Atkeson et al, 1997). M5Prime is a rational reconstruction of Quinlan's M5 model tree inducer (Wang and Witten, 1997). While decision trees were designed for assigning nominal categories, this representation can be extended to numeric prediction by modifying the leaf nodes of the tree to contain a numeric value which is the average of all the dataset's values that the leaf applies to. Finally, Decision Stump builds simple binary decision "stumps" (1-level decision trees) for both numeric and nominal classification problems. It copes with missing values by extending a third branch from the stump—in other words, by treating "missing" as a separate attribute value. Decision Stump is mainly used in conjunction with the Log it Boost boosting method.

Clustering

EM Algorithm

The EM algorithm is used to find the maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. Typically these models involve latent variables in addition to unknown parameters and known data observations. That is, either there are missing values among the data, or the model can be formulated more simply by assuming the existence of additional unobserved data points. For example, a mixture model can be described more simply by assuming that each observed data point has a corresponding unobserved data point, or latent variable, specifying the mixture component that each data point belongs to finding a maximum likelihood solution typically requires taking the derivatives of the likelihood function with respect to all the unknown values.
In statistical models with latent variables, this usually is not possible. Instead, the result is typically a set of interlocking equations in which the solution to the parameters requires the values of the latent variables and vice versa, but substituting one set of equations into the other produces an unsolvable equation. The EM algorithm proceeds from the observation that the following is a way to solve these two sets of equations numerically. One can simply pick arbitrary values for one of the two sets of unknowns, use them to estimate the second set, then use these new values to find a better estimate of the first set, and then keep alternating between the two until the resulting values both converge to fixed points.

- **Association Rules**

WEKA contains an implementation of the apriori algorithm (Agrawal, et al, 1993) for generating association rules, a type of learning scheme commonly used in "market basket analysis"(MBA). MBA algorithms have recently seen widespread use in analyzing consumer purchasing patterns specifically, in detecting products that are frequently purchased together. These algorithms were developed in response to the vast flood of transaction data produced by barcode-based purchasing/ordering systems. This data was quickly recognized by the business world as having immense potential value in marketing, but traditional data analysis techniques could not cope with the size of the hypothesis spacethat these datasets engender.

For this type of analysis, data is logically organized into "baskets" (usually records in which the items purchased by a given consumer at a given time are grouped together).MBA algorithms such as Apriori discover "association rules" that identify patterns of purchases, such that the presence of one item in a basket will imply the presence of one or more additional items. A hypothetical example of such a rule might be that shoppers who purchase toothpaste are also likely to buy bananas on the same trip to the grocery store. This result can then be used to suggest combinations of products for special promotions or sales, devise a more effective store layout, and give insight into brand loyalty and co-branding.

C. Probability estimation

The drought research by using the supervised classification algorithm called Naive Bayes Classification algorithm was implemented. A **Bayes classifier** is a simple probabilistic classifier based on applying Bayes theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood. An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

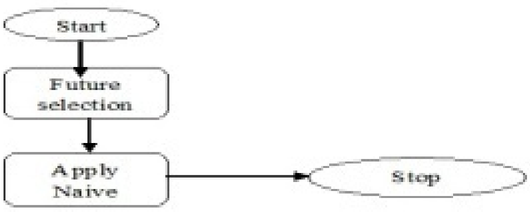

Fig 4: Future selection of agricultural drought

D. Evaluation

Predicting the drought affected region by using the naïve bayes algorithm and evaluate the performance by using the parameters of the process.
These are estimated by using the probability values in the dataset. The graph can be evaluated by using this type of parameters form probability estimations.

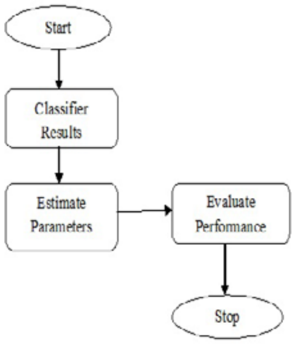

Fig 5: Evaluation of agricultural drought

IV. CONCLUSION AND FUTURE ENHANCEMENT

A. Conclusion

The accuracy and the effectiveness of ISDI for model construction was evaluated, and model application using a variety of methods and data, there is still some work to be done in our future research because of the complex spatial and temporal characteristics of drought. In this project, for the limitation of field observation data, a case study of qualitative and quantitative evaluation of ISDI was presented and more in-depth validation and assessment in our future research with the field observation data increasing and improvement. Multi-year, multi-seasonal meteorological and observed vegetation condition and soil moisture data at different sites and region will be used to assess the ISDI performance from both spatial and temporal aspects.

TABLE 3: REGRESSION PRECISION EVAUATION OF VEGDRI AND ISDI

SEASON	MODEL	AVERAGE ERROR	RELATIVE ERROR
Spring	VegDRI ISDI	0.3688 **0.3569**	0.24 **0.23**
Summer	VegDRI ISDI	0.7152 **0.7064**	0.42 **0.42**
Autumn	VegDRI ISDI	0.3984 **0.4105**	0.20 **0.22**

Future research with the field observation data increasing and improvement. Multi-year, multi-seasonal meteorological and observed vegetation condition and soil moisture data at different sites and region will be used to assess the ISDI performance by using the different set of environmental characteristics. For the different year annual rate and soil moisture are calculated for various seasons. According to the season the annual rate and soil moisture helps to find which crop will be suitable for soil to aggregate.

REFERENCES

1. Boltenetal.J.D, *"Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring,"* IEEE J. Sel. Topics Appl. Earth Observe Remote Sens., vol. 3, pp. 57–66, 2010.
2. Brownetal.J, *"The Vegetation Drought Response Index (VegDRI): A new integrated approach for monitoring drought stress in vegetation,"* GIS Remote Sens., vol. 45, pp. 16–46, 2008.
3. Dracup.J.A,Lee.K.S.,andE.G.P.Jr,*"On the definition of droughts,"* Water Resources Res., vol. 16, pp. 297–302, 1980.
4. Guiqin.J, Fuliang.Y, and Yong.Z, *"An analysis of vulnerability to agricultural drought in China using the expand grey relation analysis method,"* Procedia Eng., vol. 28, pp. 670–676, 2012.
5. Lian.F,Chuanmin.H, and Xiaoling.C, *"Satellites capture the drought severity around China's largest freshwater lake,"* IEEE J. Sel. Topics Appl. Earth Observe Remote Sens., vol. 5, pp. 1266–1271, 2012.
6. Li.K.R, GuoQ.Y, and J. C. Zhang, "*China Drought Research and Disaster Mitigation Measures (in Chinese)*". Zhengzhou, China: Henan Science and Technology Press, 1999.
7. McKee.T.B, Doesken.N.J, and Kleist.J, *"The relationship of drought frequency and duration to time scales,"* in Proc. 8th Conf. Applied Climatology, Anaheim, CA, USA, 1993, pp. 179–184.
8. Palmer.W.C, *"Meteorological drought,"* US. Weather Bureau Res. Paper, p. 58, 1965.
9. Quiring.S.M and Ganesh.S, *"Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas,"*Agricult. ForMeteorol., vol. 150, pp. 330–339, 2010.
10. Quiring.S.M and Papakryiakou.T.N, *"An evaluation of agricultural drought indices for the Canadian prairies,"*Agricult. ForMeteorol., vol. 118, pp. 49–62, 2003.
11. Sivakumar.M.V.K, *"Agricultural drought—WMO perspectives,"* in Proc. WMO/UNISDR Expert Group Meeting on Agricultural Drought Indices, Murcia, Spain, 2010, pp. 22–34.
12. Sonetal.N.T, *"Monitoring agricultural drought in the Lower Mekong Basin using MODIS NDVI and land surface temperature data,"* Int. J. Appl. Earth Observ. Geoinf., vol. 18, pp. 417–427, 2012.