International Journal of Innovative Research in Engineering
Volume 4, Issue 2 (March-April 2023), PP: 96-100.
www.theijire.com ***ISSN No: 2582-8746***

Crop Recommendation System Using Machine Learning Algorithm (ANFIS)

Sunil Prabhu D[1], Subas DT[2], Priyanka Ram S[3], Kavitha R[4]
[1,2,3,4]*Computer Science and Engineering, Bannari Amman Institute of Technology, Tamilnadu, India*

How to cite this paper:
Sunil Prabhu D[1], Subas DT[2], Priyanka Ram[3] S, Kavitha R[4]
"Crop Recommendation System Using Machine LearningAlgorithm (ANFIS)",
IJIRE-V4I02-96-100

Abstract: *Data Mining is the maximum plausible method of the prevailing virtual international for reading mass of records units to gain left out relationship. The approach used for the evaluation of statistical records over a time frame is the time collection evaluation. This method is clinical and dependable in forecasting occasions to observe over a period. Probability of manufacturing should continually be expected to theclose to perfection with the aid of using time collection evaluation. In this work, meals manufacturing is the thrust for prediction. The distinguished type strategies on this take a look at are the SUPPORT VECTOR REGRESSION AND ANFIS. In this project, Multi-Layer Perception MLP are the proposed ensemble version used to task the crop manufacturing over a time frame. This ensemble version is as compared to Svr and anfis strategies. The parameters used one at a time for prediction of output are the accuracy andthe type error. The locating yields that Multi-Layer Perception and MLP are agreeable than Svr and anfis for the records set analyzed.*

Key Word: *ANFIS, Data Mining, MLP, Support vector regression*

I.INTRODUCTION

Our Before the arrival of pc age, extraction of sample from statistics units become executed best via way of means of guide methods. But now with the incredible improvement of pc technology, series of statistics set, type and garage as splendidly increased. This has made massive alternate in Pattern recognition. In order to discover specific sample from the massive statistics units, an utility is advanced via way of means of the usage of unique automatic set of rules with inside the area of Data mining. Machine getting to know has been advanced in Data Mining as a version in getting to know idea via way of means of the usage of the pc. Given massive statistics units, prediction of recent units of statistics are advanced theusage of getting to know idea via way of means of this version via way of means of schooling and testing. With the intention of predicting an outcome, growing a version with the item of producing type is popularly known as modeling. The type in statistics mining method is predicting the price of a goal variable via way of means of producing a version primarily based totally on a few attributes' express variable. By this method, type of a given statistics is primarily based totally on elegance labels and schooling. A multi-layer perceptron has one input layer and for each input, there is one neuron (or node), it has one output layer with a single node for each output and it can have any number of hidden layers and each hiddenlayer can have any number of nodes. A schematic diagram of a Multi-Layer Perceptron (MLP) is depicted below. In the multi-layer perceptron diagram above, we can see that there are three inputs and thus three input nodes and the hidden layer has three nodes. The output layer gives two outputs, therefore there are two output nodes. The nodes in the input layer take input and forward it for further process, in the diagram above the nodes in the input layer forwards their output to each of the three nodes in the hidden layer, and in the same way, the hidden layer processes the information and passes it to the output layer. Every node in the multi-layer perception uses a sigmoid activation function.

The time collection statistics is a statistical statistic measured at a selected time c program language period over a period. The evaluation main to end in this statistics for destiny prediction is known as time collection evaluation. A giant vicinity for time collection evaluation is fashion in crop manufacturing. In this paper, crop manufacturing fashion is usually recommended the usage of statistics mining predictive strategies consisting of Support Vector machines and Multi-Layer Perception which also can be referred as classifier strategies with inside the evaluation of time collection statistics units is used. For the cause of decreasing the mistake price and to growth the prediction accuracy, boosting is likewise carried out. Joining irrelevant and vulnerable prediction guidelines for growing system orientated getting to know is known as Boosting.It is used to lessen bias and variance in system getting to know to transform vulnerable novices to sturdy ones via way of means of this ensemble Meta set of rules. Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi-layer perception is a neural network that has multiple layers. To create a neural network, we combine neurons together so that the outputs of some neurons are inputs of other neurons.

The define of this painting is illustrated via way of means of the use of a determine on this section. Fig 1 describesthe assessment of the proposed paintings. In Fig 1, the ancient crop manufacturing and the environmental weather records is taken under consideration for this study. At the start stage, this big records set is accomplished into preprocessing and is referred to as Data Preprocessing. In the subsequent stage, the fashions are generated via way of means of the use of device getting to know algorithm. In the very last stage, validating the version is achieved via way of means of evaluating the end result of current and the proposed technique.

II.LITERATURE REVIEW

PATTERN BASED SEQUENCE CLASSIFICATION

Cheng Zhou, Boris Cule et.al has proposed In this paper, Sequence classification is an important task in data mining. We address the problem of sequence classification using rules composed of interesting patterns found in a dataset of labelled sequences and accompanying class labels. We measure the interestingness of a pattern in a given class of sequences by combining the cohesion and the support of the pattern. We use the discovered patterns to generate confident classification rules, and present two different ways of building a classifier. The first classifier is based on an improved version of the existing method of classification based on association rules, while the second ranks the rules by first measuring their value specific to the new data object. Experimental results show that our rulebased classifiers outperform existing comparable classifiers in terms of accuracy and stability. Additionally, we test a number of pattern feature based models that use different kinds of patterns as features to represent each sequence as a feature vector. We then apply a varietyof machine learning algorithms for sequence classification, experimentally demonstrating that the patterns we discover represent the sequences well, and prove effective for the classification task.

We introduce a sequence classification method based on interesting patterns named SCIP. We present four concrete classifiers, using various pattern types and classification strategies. Through experimental evaluation, we show that the SCIP rule based classifiers in most cases provide higher classification accuracy compared to existing methods. The experimental results show that SCIP is not overly sensitive to the setting of a minimum support threshold or a minimum confidence threshold. In addition, the SCIP method proved to be scalable, with runtimes dependent on the minimum supportthreshold and the number of data objects. What is more, by using the discovered patterns as input for a number of learning-based classification algorithms, we demonstrate that our pattern mining method is effective in finding informative patterns to represent the sequences, leading to classification accuracy that is in most cases higher than the baselines. Therefore, we can conclude that SCIP is not only an effective and stable method for classifying sequence data, but also that its first, patternmining, step provides a valuable tool for discovering representative patterns.

A BAYESIAN CLASSIFICATION APPROACH USING CLASS-SPECIFIC FEATURES FOR TEXT CATEGORIZATION

Bo Tang et.al has proposed In this paper., we present a Bayesian classification approach for automatic text categorization using class-specific features. Unlike the conventional approaches for text categorization, our proposed method selects a specific feature subset for each class. To apply these class-dependent features for classification, we follow Baggenstoss's PDF Projection Theorem to reconstruct PDFs in raw data space from the class-specific PDFs in low-dimensional feature space, and build a Bayes classification rule. One noticeable significance of our approach is that most feature selection criteria, such as Information Gain (IG) and Maximum Discrimination (MD), can be easily incorporated into our approach. We evaluate our method's classification performance on several real-world benchmark data sets, compared with the state-of-the-art feature selection approaches. The superior results demonstrate the effectiveness of the proposed approach and further indicate its wide potential applications in text categorization.

we have presented a Bayesian classification approach for automatic text categorization using class-specific features. In contrast to the conventional feature selection methods, it allows to choose the most important features for each class. To apply the class specific features for classification, we have derived a new Multi-Layer Perception rule following Baggenstoss's PDF Projection Theorem. One important advantage of our method is that many existing feature selection criteria can be easily incorporated. The experiments we have conducted on several data sets have shown promising performance improvement compared with the state of-the-art feature selection methods.

VISUALLY COMPARING WEATHER FEATURES IN FORECASTS

P. Samuel Quinan et.al has proposed in this paper Meteorologists process and analyze weather forecasts using visualization in order to examine the behaviors of and relationships among weather features. In this design study conductedwith meteorologists in decision support roles, we identified and attempted to address two significant common challenges inweather visualization: the employment of inconsistent and often ineffective visual encoding practices across a wide range of visualizations, and a lack of support for directly visualizing how different weather features relate across an ensemble ofpossible forecast outcomes. In this work, we present a characterization of the problems and data associated with meteorological forecasting, we propose a set of informed default encoding choices that integrate existing meteorological conventions with effective visualization practice, and we extend a set of techniques as an initial step toward directly visualizing the interactions of multiple features over an ensemble forecast. We discuss the integration of these contributionsinto a functional prototype tool, and also reflect on the many practical challenges that arise when working with weather data.

In this work We present a characterization of both the problems and data associated with meteorological forecasting. We outline a system for informed defaults that allow meteorologists without visualization expertise to generatea wide variety of effective visualizations based on current meteorological conventions and visualization principles. We alsoextend state-of-the-art visualization techniques in order to allow users to effectively relate multiple isocontour features. Asa proof-ofconcept for these ideas we present Waver, an open-source tool for interactively visualizing weather forecasts. Asfuture work we are interested in conducting a formal evaluation of both interactive contour boxplots and interactive spaghetti plots as mechanisms for looking at multiple isocontour based features simultaneously. We are also interested in investigating how to design a user study that can account for highly individualized analysis processes like those found in the meteorological community. Additionally, there remains no current visualization methods that enable the direct visualization or comparison of non-isocontour based features

across an ensemble, providing another interesting avenue for future.

III.PROPOSED SYSTEM AND METHODOLOGY

In this paper, forecasting of crop production is done by using the time series data set precisely than the existing models. By using Boost technique, ensemble models such as Multi Layer Perception and MLP are developed. To bring weak learners who are slow in learning, Prediction technique helps their understanding. Multi Layer Perception when joinedwith Prediction (Multi Layer Perception) will make superior classification by giving weak learners with appropriate training. A like method is used for Multi Layer Perception classifier in which Prediction based Multi Layer Perception (MLP) is usedto generate superior classified data. Fig 2 depicts the system implementation where the mass of historical crop production data and climate data is gathered and is made to data preprocessing work. In the data preprocessing, the data's are combineand selected for the study. The models are generated by classifying the mass of input data by using Svr and anfis modelingtechniques. It may be seen that it is further classified in order to improve the performance with the help of Multi Layer Perception and Multi Layer Perception (MLP). The models are evaluated by comparison.

Advantages

- Validating the model is done by comparing the result of existing and the proposed technique.
- Forecasting of crop production is done by using the time series data set precisely than the existing models.
- Accuracy in time series analysis is the value forecasted which is very near to the actual value
- System implementation where the mass of historical crop production data and climate data is gathered and is madeto data preprocessing work.
- Crop yield prediction is subjected to analysis by the existing techniques Svr and anfis with the proposed techniques.

Methodology Module Description

- Data preprocessing
- Feature selection
- Feature extraction
- Classification algorithm
- Crop yield prediction analysis

Data Preprocessing

Hear the raw data in the crop data is cleaned and the metadata is appending to it by removing the things which areconverted to the integer. So, the data is easy to train. Hear all the data. In this pre-processing, we first load the metadata intothis and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will bemoved further and remove the unwanted data in the list and it will divide the data into the train and the test data For this splitting of the data into train and test we need to import train_test_split which in the scikit-learn this will help the pre- processed data to split the data into train and test according to the given weight given in the code. The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively.

Feature Selection

Feature selection refers to the process of applying statistical tests to inputs, given a specified output. The goal is todetermine which columns are more predictive of the output. Apply the machine learning techniques which are helpful for finding crop yield for any of new data occurred in the data. After this data acquisition suitable machine learning algorithm must be applied to compute efficiency and capability of the model, here we have applied various machine learning algorithms.

Feature Extraction

Feature extraction involves reducing the number of resources required to describe a large set of data. So we analyzed that proposed model has got more efficiency than the existing model for finding crop yield. The implementation of above system would help in better cultivation of the agricultural practices of our country. Further it can be used to reducethe loss faced by the farmers and improve the crop yield to get better capital in agriculture. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.

Classification AlgorithmMulti-Layer Perceptron

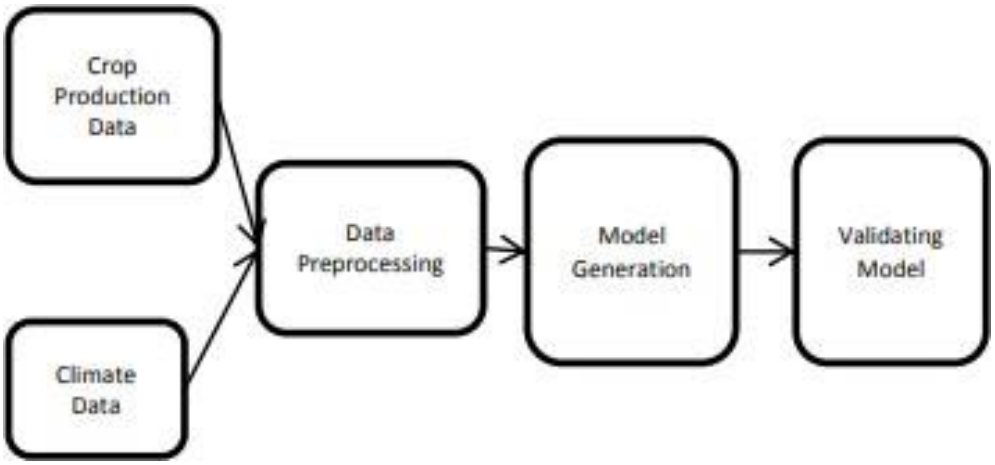
Multi-layer perception is also known as MLP. It is fully connected dense layers, which transform any input dimension to the desired dimension. A multi-layer perception is a neural network that has multiple layers. To create a neural network wecombine neurons together so that the outputs of some neurons are inputs of other neurons. A gentle introduction to neural networks and TensorFlow can be found here: Neural Networks, Introduction to TensorFlow, A multi-layer perceptron has one input layer and for each input, there is one neuron (or node), it has one output layer with a single node for each output and it can have any number of hidden layers and each hidden layer can have any number of nodes. A schematic diagram of a Multi-Layer Perceptron (MLP) is depicted below. In the multi-layer perceptron diagram above, we can see that there are three inputs and thus three input nodes and the hidden layer has three nodes. The output layer gives two outputs, therefore there are two output nodes. The nodes in the input layer take input and forward it for further process, in the diagram above the nodes in the input layer forwards their output to each of the three nodes in the hidden layer, and in the same way, the hidden layer processes the information and passes it to the output layer. Every node in the multi-layer perception uses a sigmoid activation function. The

sigmoid activation function takes real values as input and converts them to numbers between 0 and 1 using the sigmoid formula.

Crop Yield Prediction Analysis

Agricultural data is being produced constantly and enourmosly. As a result, agricultural data has come in the era of big data. Smart technologies contribute in data collection using electronic devices. In our project we are going to analyse and mine this agricultural data to get useful results using technologies like data analytics and machine learning and this result will be given to farmers for better crop yield in terms of efficiency and productivity. The work will help farmers to increase the yield of their crops. Also implemented Multi Layer Perception algorithm for finding out the exact crop. Thus, we implemented a system which will predict the crop name and approximate yield in a particular farm. Multi Layer Perception uses probability of crop being grown in those circumstances. Hence in the following result probabilities are calculated and most probable crops are chosen for further final accumulation.

Proposed System Flow Diagram



IV.EXPERIMENTAL RESULT

Multi-Layer Perception, Multi-Layer Perception, are the classification methods used for time series forecast in this paper. Two groups are separated from the data set for training and for testing the algorithms of classification. In order to implement the classification algorithms, the tool used is RapidMiner data analysis. “Read CSV” operator of RapidMiner tool is first loaded for Secondary data retained in CSV file. For classification process only a subset of data is selected from the loaded data. To select a subset from original data, “Select Attributes” are utilized by the operator. The chosen subset is then subjected to “X-Validation” operator. It develops the classification model which is validated by the test data. Svr and anfis are implemented for classification by using “X-Validation” operator. The performance of the classification algorithm is evaluated by using the performance operator. Performance evaluation achieved for both the classification algorithms (existing and proposed) are given in Table 1 and Table 2.

Dataset Discribtion

The historical climate data of the CSV type with a record for every line of text alike to the data for a given month of a year.The crop production data of the CSV type with a record for each line of text belongs to the data for a specific month of a year. In the preprocessing stage, though there are many measured parameters available in the raw climate dataset, the less relevant features responsible for the study are ignored and the important features are only taken into account. Thus the two different types of historical records are preprocessed and combined together so that it can be used for this study. The time series historical data over 100 years is taken into study for this experiment. For the purpose of training and testing, separation of the dataset has been made where 60% of the dataset is used for training and 40% of the dataset is used for testing. By using the classification algorithm, the training dataset is used and models generated. The created model is used on the testing dataset in order to test the accuracy. A representation for the environmental demand of evapotranspiration can otherwise be called “Reference Crop Evapotranspiration". This represents both height of the short green crop and sufficient water level in the soil. In other words, it is a reflection of the energy available to evaporate water and the wind available to transport the water vapor from the ground to the lower atmosphere.

V.CONCLUSION

The time series analysis of crop yield prediction is subjected to analysis by the existing techniques svr and anfis with the proposed technique. It may be concluded from the results that there is good amount of perfection in accuracy of prediction and also good amount of fall in the percentage of accuracy in both the proposed techniques. Future research can enlighten the study whether by changing the technique produces better results or by increasing the input data set for the same technique results change in the findings. Importance of crop prediction is highly needed for agriculture and economy. Continuous research for improving new methods of prediction would be fruitful. This project is a beginning for further research in forecasting.

VI.FUTURE ENHANCEMENT

The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of fertilizer so that he can use only the required amount of fertilizers for that field. This way we can help the farmers to grow the crop which gives them better yield.

References

[1] Cheng Zhou, Boris Cule, Bart Goethals "Pattern Based Sequence Classification", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, No.5, 2016, pp.1285-1298.

[2] B. Tang, H. He, P. M. Baggenstoss and S. Kay "A Bayesian Classification Approach Using Class-Specific Features for Text Categorization", IEEE Transactions on Knowledge and Data Engineering, Vol.28, No.6, 2016, pp.1602-1606.

[3] P.Samuel Quinan, Miriah Meyer "Visually Comparing Weather Features in Forecasts", IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No.1, 2016, pp. 389-398.

[4] Y. Chen and Y. Li "Entropy-Based Combining Prediction of Grey Time Series and Its Application" , IEEE International Conference on Intelligent Computation Technology and Automation (ICICTA), 2009, pp. 37-40.

[5] G. Chen, X. Xu, G. Wang and H. Chen "The corn output in a time series prediction model", IEEE International Conference on World Automation Congress (WAC), 2010, pp. 283-286.