Volume 6, Issue 9, September 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Applying Data Mining Techniques to Evaluate Rural Crops Using kNN

P.Sharmila Devi
MCA., M.Phil., Research Scholar,
Dept. of CS, Nehru Memorial College (autonomous)
Puthanampatti, Tiruchirappalli,
Tamilnadu, India

Dr. R.Periyasamy
M.Sc., (Phy), PGDCA, M.Sc., M.C.A., M.Phil., Ph.D.,
Associate Professor, Dept. of CS, Nehru Memorial College (autonomous), Puthanampatti, Tiruchirappalli,
Tamilnadu, India

Abstract— Extraction of data in rural data is a testing task, from finding designs also, connections also, interpretation. In request to acquire conceivably intriguing designs also, connections from this data, it is therefore vital that a technique be created also, take advantage of the sets of existing strategies also, instruments accessible for data mining also, data very in databases. Data mining is relatively a new approach in the field of agriculture. Precise data in describing crops depends on climatic, geographical, natural also, other factors. These are very imperative inputs to create portrayal also, expectation models in data mining. In this study, an effective data mining technique based on kNN is explored, introduced also, executed to portray rural crops. The strategy draws upgrades to request issues by utilizing Principal Components Analysis (kNN) as a pre preparing strategy also, a changed Genetic Algorithm (GA) as the capacity optimizer. The wellness capacity in GA is changed accordingly utilizing effective separation measures. The approach is to asses, the kNN hybrid data mining method, utilizing diverse rural field data sets, create data mining request models also, and establish significant relationships. The test results show improved request rates also, created portrayal models for rural crops. The area model result may have benefits, to rural analysts also, farmers. These created request models can too be used also, readily incorporated into a decision support system.

Keywords— Classification, data mining, Genetic Algorithm, k-NN, main segment analysis.

I. INTRODUCTION

Data in the rural area are robust, it comes in distinctive formats, complex, multidimensional, also, contains noise. Intriguing designs can be mined from this space in finding knowledge, revealing solutions to particular area issues.

Climatic, geographical, natural also, other elements affects the historical yield of crops, also, these are very imperative inputs to PC created crop yield expectation models. Mathematical also, factual demonstrating are used to discover designs in the data, thru these watched field also, genuine test data, statistically created expectation also, portrayal models are executed also, used by both ranchers also, researchers. These models are too helpful to government organizations in establishing proper policies for decision making process.

Other than statistics, a very intriguing process known as Knowledge Discovery from Databases (KDD) can be used. One of the core errands involved in the KDD process is Data Mining (DM). The framework of the KDD process appeared in Fig. 1, includes several errands or phases.

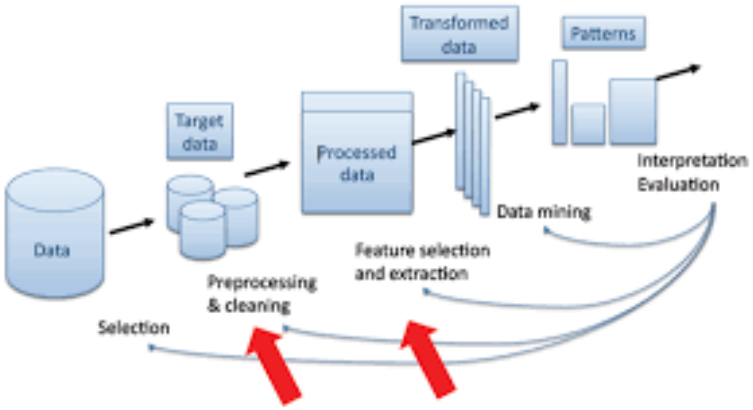

Fig. 1. The KDD process

The explosion of the data revolution also, the proliferation of utilizing computing also, data storage made accessible enormous amounts of data, this led to new strategies such as data mining that can span the data discovered in the data. The extraction of data in the data is now a testing task, from finding intriguing designs also, connections to elucidation of what the data is. In request to acquire conceivably intriguing designs also, connections in the data, it vital that a technique

be developed, taking advantage of the set of existing strategies also, instruments accessible for data mining also, data discovery in databases.

In data mining, request can be seen as design recognition. Each design from the data is reintroduced by a set of measurable highlights or dimensions also, viewed as a point in a given dimensional space. The aim is to choose highlights that permit us to segregate between designs belonging to distinctive classes. Often, the optimal set of highlight is usually unknown, considering every single highlight of an info design in a substantial highlight set makes request computationally complex. Also, the inclusion of irimportant or redundant highlights in the data mining model results in poor predictions also, interpretation, high computational cost also, high memory usage. In general, it is desired to keep a number of highlights as segregating also, little as conceivable to diminish computational time also, complexity, in the data mining process.

The focus of this study is to execute an effective data mining segment based on the blend of Main Segment Investigation (*kNN*) as a preparing strategy also, a changed Genetic Algorithm (GA), as the learning algorithm, in request to diminish computational cost also, time by keeping a number of highlights as segregating also, little as possible. In so doing, creating rural crops request models is effective also, portrayal is improved. The *kNN* data mining segment will be executed for rural crops dataset to identify key characteristic combinations also, qualities that determine crop performance. The result of the data mining demonstrating can be used for decision support in improving rural crops productivity.

II. RELATED LITERATURE

Data mining in farming is new, however there are novel ideas also, studies conducted to explore its applicability from mining also, demonstrating data to developing applications, utilizing the created models based on the calculations used.

The study in reviewed the application of data mining strategies also, found out that there are several calculations also, strategies being connected in the farming domain. Similarly, in, data mining strategies was connected to portray soil data also, found that data mining depends on the amount of data used in the process. Their study connected Innocent Bayes in arranging rural soils. That an increase in dataset size improves accuracy, which may improve the verification of substantial designs compared to standard factual analysis.

On the effectiveness of data mining as a tool, the paper of Raoranne A. A., Kulkarni R. V., discussed how data mining can span data of the data to crop yield estimation. The study assessed new data mining strategies also, was connected to diverse variables to establish if significant connections can be found. It was watched that effective strategies can be created also, examined utilizing fitting data to solve complex rural issues utilizing data mining techniques. Data mining request strategies connected to soil database can be effective in establishing significant connections from the data.

There are distinctive data mining strategies accessible in the writing to improve data mining tasks. Reference used Genetic Algorithm for highlight decision in the setting of a neural framework classifier. GA was configured to use an approximate assessment in request to diminish significantly the computation required. The calculation employed nearest-neighbor (k-NN) classifier to assess highlight sets also, appeared that the highlights chosen by this strategy are effective.

kNN is one of these strategies also, performs well in diminishing complexity in data by diminishing its dimensionality. In they mentioned that, “one of the key steps in data mining is finding ways to diminish dimensionality without sacrificing correctness". They connected *kNN* also, found that it handles sparse data also, created fewer also, and improved association rules. *kNN* is a multivariate technique, that analyses a data table in which perceptions are described by several inter corresponded quantitative subordinate variables. Its goal is to transform the data, represent it as a set of new orthogonal variables called main components. In this case, how many parts should be considered?

In highlight subset decision no new highlights will be created but a, subset of the unique highlights are chosen also, the highlight space is reduced. In cases where there are more highlights than necessary, subset decision helps simplify computational time, enhances also, improves prescient power of classifiers -.

Genetic Algorithm has been appeared in the writing to be an effective instrument to use in data mining also, design recognition. However, GA has issues with premature convergence which inhibit diversity in the populace also, prevent exploration of the whole look space. To address this problem, the work of A. Hassani, also, J. Treijis recommended tweaking the GA to a particular issue also, correctly set all parameters, conversely, L. Na-Na, G. Jun-Hua, also, L. Bo-Ying, used the negative decision strategy also, and appeared promising results.

In the study of A.S. Elden, M. A. Mustafa, H. M. Harb also, A. H. Emara, they designed also, evaluated a fast learning calculation based on GA also, proved to have impressive upgrades on the precision performance, over other classifiers. Also, in, *kNN* was applied, then the k-NN classifier was used as the wellness capacity for the GA also, resulted to diminished request error rates, they further recommended utilizing distinctive classifiers for comparative studies.

III. CONCEPTS AND METHODS

A. The Data Mining Segment

There are two major stages in the data mining segment being presented, the first stage is data preparing utilizing *kNN* also, utilizing GA to find the highlight subset that is the ideal arrangement to the issue being addressed, this process can be considered as an enhancement technique. The second stage is to utilize the ideal results also, rules in creating models of request for the portrayal of crops. This expectation model is then used for decision support.

B. Strategies also, Procedures

1) Data preparing

Data preparing is an imperative errand also, method in the data mining process it transforms data into understandable format. Genuine world data is incomplete, noisy, and inconsistent also, lacking certain trends. Data preparing resolves these issues which includes cleaning, transformation, normalization, highlight extraction also, selection.

kNN is a method that converts a set of perceptions of possibly corresponded variables into a set of values linearly corresponded variables called main components. The changed dataset is characterized in such a way that the first main parts account for much of the variance. Main parts are guaranteed to be insubordinate if the data set is jointly normally distributed.

2) Request

This is an errand performed to generalize known form in data mining to apply to new data. It is too the categorization of data for it's most effective also, effective use. There are numerous data mining request calculations being studied also, executed in distinctive domains. Some of the most popular also, basic are adapted also, introduced herein, based on their capabilities straightforwardness also, robustness.

- K-Nearest Neighbor (k-NN)

The principle behind this strategy is to find characterized numbers of preparing tests closest in the separation to the new point also, predict mark from these. The number of tests can be a user characterized constant or fluctuated based on the local density of points. The separation can be any metric measure. There are separation measures executed in the k-NN, Euclidean, Chebysheb, Manhattan also, Edit Distance, but the Euclidean separation measure is the most basic choice. Despite its straightforwardness it is effective in substantial number of request problems.

- J4.8

J4.8 decision trees calculation is an open source Java usage of the C4.5. It grows a tree also, employments divide-and-conquer algorithm. It is a prescient machine-learning model that decides the target value (subordinate variable) of a new test based on diverse characteristic values of the accessible data.

To characterize a new item, it creates a decision tree based on the characteristic values of the preparing data. When it encounters a set of items in a preparing set, it identifies the characteristic that discriminates. It employments data gain to tell us most about the data instances so that it can characterize them the best.

- Innocent bayes

This classifier is based on the Bayes rule of conditional probability. It employments all of the qualities contained in the data, also, analyses them independently as though they are equally imperative also, insubordinate of each other.

The Innocent Bayes classifier works on a simple, but comparatively intuitive concept. It makes use of the variables contained in the data sample, by observing them individually, insubordinate of each other. It considers each of the qualities separately when arranging a new instance. It assumes that one characteristic works independently of the other qualities contained by the sample.

- Multi -layer perceptron (MLP)

MLP is a feed forward artificial neural framework model that maps sets of info data onto a set of fitting outputs. It consists of multiple layers of nodes, with each layer fully connected to the next one. Each node is a neuron with a nonlinear activation function. It employments a learning method called back propagation for preparing the network.

3) Genetic Algorithm (GA)

Genetic Algorithm is an evolutionary based stochastic enhancement algorithm, proposed by Hollalso, (1973). It is regarded as a capacity optimizer due to its extraordinary execution with optimization. The calculation comprises of three main Hereditary operators: selection, hybrid also, transformation to form a new generation. It converges to the best chromosome, which hopefully represents the ideal or sub-ideal arrangement to a problem.

4) Rural datasets

Important rural data was chosen from the UCI machine learning repository (https://archive.ics.uci.edu/), particular to rural crops, the soybean also, mushroom datasets.

- Soybean dataset

Table I: Soybean Dataset Description

Characteristic Name	Contains
1 Date	April,May,June,July,August,September, October
2 Plant-Stalso,	Normal,Lt-Normal
3 Precip	Lt-Norm,Norm,Gt-Standard
4 Temp	Lt-Norm,Norm,Gt-Standard
5 Hail	Yes,No
6 Crop-Hist	Diff-Lst-Year,Same-Lst-Yr,Same-Lst-Two-Yrs, Same-Lst-Sev-Yrs
7 Area-Damaged	Scattered,Low-Areas,Upper-Areas,Whole-Field
8 Severity	Minor,Pot-Severe,Severe
9 Seed-Tmt	None,Fungicide,Other
10 Germination	90-100,80-89,Lt-80
11 Plant-Growth	Norm,Abnorm

12 Leaves	Norm,Abnorm
13 Leafspots-Halo	Absent,Yellow-Halos,No-Yellow-Halos
14 Leafspots-Marg	W-S-Marg,No-W-S-Marg,Dna
15 Leafspot-Size	Lt-1/8,Gt-1/8,Dna
16 Leaf-Shread	Absent,Present
17 Leaf-Malf	Absent,Present
18 Leaf-Mild	Absent,Upper-Surf,Lower-Surf
19 Stem	Norm,Abnorm
20 Lodging	Yes,No
21 Stem-Cankers	Absent,Below-Soil,Above-Soil,Above-Sec-Nde
22 Canker-Lesion	Dna,Brown,Dk-Brown-Blk,Tan
23 Fruiting-Bodies	Absent,Present
24 External-Decay	Absent,Firm-And-Dry,Watery
25 Mycelium	Absent,Present
26 Int-Discolor	None,Brown,Black
27 Sclerotia	Absent,Present
28 Fruit-Pods	Norm,Diseased,Few-Present,Dna
29 Fruit-Spots	Absent,Colored,Brown-W/Blk-Specks, Distort,Dna
30 Seed	Norm,Abnorm
31 Mold-Growth	Absent,Present
32 Seed-Discolor	Absent,Present
33 Seed-Size	Norm,Lt-Norm
34 Shriveling	Absent,Present
35 Roots	Norm,Rotted,Galls-Cysts
36 Class	Diaporthe-Stem-Canker, Charcoal-Rot, Rhizoctonia-Root-Rot, Phytophthora-Rot, Brown-Stem-Rot, Powdery-Mildew, Downy-Mildew, Brown-Spot, Bacterial-Blight, Bacterial-Pustule, Purple-Seed-Stain, Anthracnose, Phyllosticta-Leaf-Spot, Alternarialeaf-Spot, Frog-Eye-Leaf-Spot, Diaporthe-Pod-&-Stem-Blight, Cyst-Nematode, 2-4-D-Injury, Herbicide-Injury

The cassava dataset was taken from reference also, Agrinet. The crude data was chosen also, cleaned based on important fields with the assistance of an expert.

Utilizing a text editor, all the datasets were encoded, formatted also, converted into an characteristic relation record group (.arff) for info compatibility to the DM programming also, to permit the machine learning calculations be connected to create important outcomes. The descriptions of the datasets, the qualities also, the data type that it may contain are appeared in the following tables.

- Mushroom dataset

Table II: Mushroom Dataset Description

Qualities	Contains
1 cap-shape	bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s
2 cap-surface	fibrous=f, grooves=g, scaly=y, smooth=s
3 cap-color:	brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y
4 bruises	bruises=t, no=f
5 odor	almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s
6 gill-attachment	attached=a, descending=d, free=f, notched=n
7 gill-spacing	close=c, crowded=w, distant=d
8 gill-size	broad=b, narrow=n
9 gill-color	black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
10 stalk-shape	enlarging=e, tapering=t
11 stalk-root	bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?

12 stalk-surface-abov	fibrous=f, scaly=y, silky=k, smooth=s
13 stalk-surface-belov	fibrous=f, scaly=y, silky=k, smooth=s
14 stalk-color-above-ring	brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
15 stalk-color-below-ring	brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
16 veil-type:	partial=p, universal=u
17 veil-color:	brown=n, orange=o, white=w, yellow=y
18 ring-number:	none=n, one=o, two=t
19 ring-type:	cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z
20 spore-print-color	black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y
21 populace	abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y
22 habitat	grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d
23 class	edible=e, poisonous=p

- Cassava dataset

Table III: Cassava Dataset Description

	Qualities	Contains
1.	f_root_yield	low, moderate, high
2.	Root_dry_matter_content	REAL
3.	Root_starch_content	REAL
4.	Root_skin_color	dark-brown, cream, light-brown
5.	Root_HCN_content	low, moderate, high
6.	Root_flesh_color	white, yellow, cream
7.	unexpanded_apical_leaves	purple, green-purple, dark-green, light-green, green
8.	first_fully_expanded_leaves	green-purple, light-green, dark-green, purple, green
9.	Petiole_color	purple, green-purple, light-green
10.	Stem_color	dark-brown, light-brown, silver-green, reddish-brown
11.	Plant_type	slightly-or-non-branching, erect-and-non-branching, moderately-branching, erect-or-slightly-branching, erect-and-late-branching, erect, slightly-and-late-branching,
12.	Red_spider_mites	HR, MR, R
13.	White_peach_scale_insects	HR, MR, R
14.	Cassava_bacterial_blight	HR, MR, R
15.	Maturity	low, high
16.	Recommended_Uses	1, 2, 3
17.	Recommended_for_release	IPB-UPLB, PhilRootcrops-VSU
18.	Year_released	1982,1986,1987,1988,1990,1993,1994,1996,1997,1999,2000,2001,2003, 2004,2006,2007,2009
19.	Class	UPLCa-1(Datu1),UPLCa-2(Lakan1), UPLCa-3(Sultan1)(G50-3),UPLCa- 4(Vassourinha),UPLCa-5(Sultan2)(G29r-3),PSBCv-9(VC-4)(CM4014), PSBCv-11(Lakan2)(CMP3419-2A), PSBCv-12(Lakan3)(SM972-20),PS BCv-13(CMP62-15),PSBCv-15(Lak an4)(CM3422-1),PSBCv-17(Sultan3)(CG87-03-01),PSBCv-18(Sultan4)(CG87-02-13),PSBCv-19(SM808-1),

		PSBCv-20(Sultan5)(CG91-13-01),N SICCv-25(Sultan6)(CG91-08-05),N SICCv-26(Sultan7)(CG91-14-01),N SICCv-27(Datu2)(CM9158-4),NSIC Cv-28(LSUCv-14)(OMR36-05-09), NSICCv-30(LSU Cv-15)(Rayong5),NSICCv-35(LSU Cv-18)(CMR37-24-1), NSICCv-36(LSUCv-19)(OMR36-62 -03),NSICCv-37(Sultan9)(CG97-17- 02),NSICCv-39(Rajah3)(CG97-05R r-01),NSICCv-40(Sultan10)(CG97-0 3-04),NSICCv-41(Sultan11)(CG97- 09-23),NSICCv-42(Rajah4)(CG97-0 1r-04),NSICCv-43(LSUCv-21)(OM R40-40-03),NSICCv-44(LSUCv-22)(CMR39-50-18),NSICCv-45(LSUC v-23)(OMR39-48-02),NSICCv-46(VSUCv-24)(CMR40-09-34)

IV. USAGE OF THE *KNN* MECHANISM

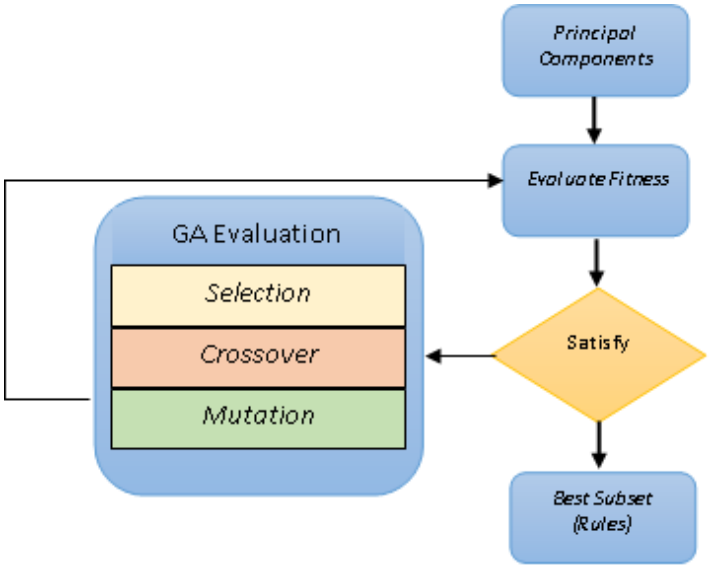

Fig. 2. Exploded view of the *kNN* method.

The idea is to execute the application of Main Segment Investigation to diminish the dimensionality of a dataset to a highlight set called main components. The main parts are then used as info populace in the look space of the GA in searching for the ideal solution. This segment proficiently simplifies the data mining process utilizing the delegate data of the unique dataset, to which diminishes computational time also, improves request execution of classifiers

However, the *kNN* method has a tendency to lose data interpretability but has high discriminative power. To overcome the shortcomings of this process, a highlight subset decision method based on a changed GA is used. In this context, utilizing other classifiers is investigated also, received as the wellness function. The wellness capacity in GA is changed accordingly utilizing effective variation of separation measures between features, this provides better separation of the design classes, which, in turn, diminishes complexity also, improves the execution of classifiers also, diminish computational costs.

A. The kNN Calculation Main parts as populace

- Compute also, assess the wellness of each main segment in the populace
- If the end condition is satisfied, stop also, return the best arrangement in current population, otherwise,
- Create new populace by repeating the following steps until the new populace is complete
- Select two parent chromosomes from a populace according to their wellness (the better fitness, the bigger chance to be selected)
- With a hybrid probability, cross over the guardians to form a new posterity (children). If no hybrid was performed, posterity is an exact copy of parents.
- With a transformation likelihood mutate new posterity
- Place new posterity in a new populace
- Use new created populace
- Go to step b

V. TEST RESULTS

Utilizing the classifiers introduced is received in the test as the wellness capacity for the GA. The k-NN request calculation was too tested also, approved utilizing fluctuated separation measures and, results are compared accordingly.

The test used the WEKA version 3.6.10 data mining programming in the usage also, utilization. A PC with 2 Gigabyte of memory, equipped with a 2.80 Ghz Processor, also, a proprietary 32 bit Operating Framework was utilized. The default settings in the data mining programming also, in the calculation configurations, was used in the experiment.

Three (3) rural crops field data, soybean, mushroom also, cassava datasets was used in the experiment. The cassava also, mushroom datasets were converted to a compatible record group for info to the DM software.

A. Highlight Sets Chosen by kNN

The soybean dataset has originally thirty six (36) qualities counting the class label, the mushroom also, cassava datasets has twenty three (23) also, eighteen (18) qualities respectively. After preparing utilizing *kNN*, the changed dataset contained forty one (41) main parts for the soybean, fifty nine (59) for the mushroom also, twenty two (22) for the cassava datasets, counting the class labels. GA was connected to the resulting pre-processed data, further diminishing the datasets. In the enhancement process, GA chosen highlight sets that are considered the optimum, thereby further diminishing the data into a littler delegate dataset.

Table IV: Number Of Highlight Sets By Executing *kNN*

Datasets	*kNN*	k-NN-GA	J4.8-GA	Innocent Bayes-GA	MLP-GA
Cassava	22	2	2	3	---
Mushroom	58	2	21	17	---
Soybean	41	2	18	26	---

As can be seen in Table IV, data preparing utilizing *kNN* also, highlight decision utilizing GA resulted into a littler number of highlight sets, which are considered the best delegate highlight sets of the data.

B. Request Execution Results

The following tables are the results of the request process.

Table V: Request Rates With K-NN Both As Wellness Capacity In GA Also, Classifier Utilizing Distinctive Separation Measures

Dataset	k-NN	*kNN*-Changed GA (Euclidean/Chebysheb/Manhattan)
Soybean	99.85%	99.85%
MAE	0.0029	0.0026
RMSE	0.0152	0.0105
RAE	2.99	2.74
RRSE	6.93	4.77
Tine	0.0sec	0.0sec
Mushroom	100%	99.98%
MAE	0.0001	0.0002
RMSE	0.0001	0.0078
RAE	0.0246	0.0458
RRSE	0.0246	1.572
Time	0.01sec	0.0sec
Cassava	100%	100%
MAE	0.0317	0.0317
RMSE	0.0898	0.0898
RAE	50.8475	50.8475
RRSE	50.8406	50.8406
Time	0.0sec	0.0sec

Table VI: Request Rates Uisng Distinctive Classifiers With J4.8 As Wellness Capacity In GA

Classifier	Soybean Unique	*kNN*- GA	Cassava Unique	*kNN*- GA	Mushroom Unique	*kNN*- GA
k-NN	91.22% 0.0sec	99.85% 0.0sec	100% 0.0sec	100% 0.0sec	100% 0.01sec	99.85% 0.01sec

J4.8	91.51% 0.02sec	98.68% 0.1sec	53.33% 0.0sec	46.67% 0.0sec	100% 0.05sec	100% 0.52sec
Innocent Bayes	92.97% 0.01sec	92.53% 0.0sec	100% 0.0sec	60.00% 0.0sec	95.83% 0.02sec	97.22% 0.12sec
MLP	93.41% 112sec	98.83% 18.2sec	100% 5.32sec	6.67% 0.0sec	100% 1876sec	99.96% 72.4sec

Table V appears the execution of the changed GA, utilizing the k-NN as the classifier also, at the same time the wellness capacity in fluctuated separation measures. It can be seen there is no huge distinction in the request accuracy, compared to the request rate on the unique dataset. This can be ascribed to the nature of similarities in the separation measurement functions. However, further analysis, the watched errors for the soybean dataset was diminished after executing *kNN* on the unique dataset. For the mushroom dataset, precision was contrarily affected but preparing time improved.

In the case of the cassava dataset, there is no huge distinction observed, this perhaps ascribed to the experts data in selecting important qualities in the encoding also, creation of the dataset.

Table VI appears otherwise the resulting effect of executing the J4.8 classifier as a wellness capacity in GA also, utilizing distinctive classifiers in the request process. The results, infers that request execution can be improved by utilizing GA as an enhancement method in the request process. With the exemption of Innocent Bayes, further investigation appears that its execution is subordinate on the nature of the dataset, also, wellness capacity used.

Table VII: Request Rates Synopsis Both As Classifier Also, As Wellness Capacity In GA

Classifier	Soybean Unique	*kNN*- GA	Cassava Unique	*kNN*- GA	Mushroom Unique	*kNN*- GA
k-NN	91.22% 0.0sec	99.85% 0.0sec	100% 0.0sec	100% 0.0sec	100% 0.01sec	99.98% 0.0sec
J4.8	91.51% 0.02sec	98.68% 0.1sec	53.33% 0.0sec	46.67% 0.0sec	100% 0.05sec	100% 0.52sec
Innocent Bayes	92.97% 0.01sec	94.44% 0.01sec	100% 0.0sec	100% 0.0sec	95.83% 0.02sec	97.89% 0.07sec
MLP	93.41% 112sec	---	100% 5.54sec	20% 0.72sec	100% 1876sec	---

The MLP performed excellent on the datasets, particular to the speed of preparing on the mushroom also, soybean datasets, which appears a very huge distinction between the unique also, the *kNN* diminished dataset. Intriguing too to note, the request rates on the mushroom dataset, though the classifier performs extraordinary with the unique dataset, the indicated request process took longer to perform as compared to the other classifiers but is excellent in speed on the *kNN* diminished dataset.

It can too be examined from the table that utilizing a particular classifier, as a wellness capacity infers that the same wellness capacity should be used in the request process in request to have impressive upgrades in the results of request process. This can too be ascribed to the qualities of the GA.

It can be seen from Table VII, that a blend of *kNN* also, a changed GA improves request accuracy, particular to Innocent Bayes for all the datasets, utilizing it both as wellness capacity also, classifier. The k-NN also, J4.8 too performed well for the mushroom also, soybean dataset. The J4.8 also, MLP classifiers performed contrarily after the *kNN* was connected to the cassava dataset.

The MLP classifier as it has been watched in the experiment, poorly performed in preparing time both as classifier also, as wellness capacity with the GA in the enhancement process, although very precise in arranging the unique datasets, this perhaps ascribed to the MLP characteristics.

The speed of preparing as can be seen does not have huge distinction for all the datasets, with the exemption of the MLP. The execution rates watched may be ascribed to the dependency also, qualities of the classifiers on the nature of the datasets: large, small, clean or noisy.

C. Request Models Visualization Utilizing Innocent Bayes as Wellness Capacity in GA also, as Classifier after applying kNN

Now, that we have appeared the execution results of the data mining process, for its excellent execution we select the Innocent Bayes classifier in the presentation of the models created in executing the *kNN* data mining segment for the portrayal of the rural crops. Appeared in the following sections, are models of nineteen (19) soybean disease classification, cassava varietal profitability also, mushroom edibility classification, utilizing the Innocent Bayes classifier which performed excellent in the data mining process with close great precision for all of the datasets.

The reader is introduced a view of the results of the experiment, which classifier is best suited in describing crops based on the *kNN* mechanism. It illustrates that these are conceivable strategies also, the decision is to have the most effective also, precise model in describing crops.

The visuals presented, proves the capability of the classifiers based on the *kNN* diminished dataset to segregate also, categorize the data on distinctive classes introduced in the unique datasets. This infers that, data mining request based on the *kNN* segment is effective also, advantageous as compared to raw, substantial also, complex datasets. Further investigation of the models, validates the efficiency also, straightforwardness of executing the segment in mining on delegate set resulting to improved classifier execution also, establishing huge connections among the variables that impact higher precision rates, thus simplifying the errand of describing crops.

D. Removed Request Rules Utilizing JRIP also, PART

To further demonstrate, appeared in Table VIII, are the number of removed rules utilizing JRIP also, PART from the *kNN* segment utilizing the Innocent Bayes as the wellness capacity in GA with the corresponding precision rates. It can be seen that the removed request rules from the mushroom dataset is highly accurate, hence establishing substantial connections among the variables from the diminished delegate dataset based on *kNN*.

Table VIII: Discovered Request Rules From The *kNN* Diminished Dataset Based On The Innocent Bayes As Wellness Function

Classifier	Soybean % Rules	Accuracy	Cassava % Rules	Accuracy	Mushroom % Rules	Accuracy
JRIP	25	91.25	1	96.67	9	99.99
PART	32	98.68	11	63.33	9	99.99

To illustrate the established connections that impact higher precision rates in the portrayal process, let us consider the mushroom dataset. Careful investigation of the rules created by JRIP also, PART, the Synopsis in Table IX appears the qualities that portray the mushroom crop, with close great accuracy. The results imply that the rules removed utilizing the qualities introduced can be used proficiently in the usage of a canny framework for rural crops characterization.

The simplest rules were of the JRIP involving sixteen (16) out of twenty two (22) attributes. The rules for palatable mushrooms are obtained as negation of the removed rules. In the case of PART, the same qualities were found to have high impact in the edibility of mushroom, though slight variations on the rules exist for non-palatable also, palatable mushrooms.

VI. CONCLUSION

Introduced in this paper, is a proposed hybrid data mining strategy based on *kNN*. The segment was appeared to have impressive impact in improving request execution rates of classifiers. Utilizing both the classifiers as wellness capacity in GA also, in the data mining request process, improves the execution of the data mining process.

The Naive Bayes also, k-NN classifiers both performed excellent as wellness capacities also, classifiers in the *kNN* data mining mechanism. Likewise, the findings accessible in the writing is further approved which appeared huge results with other separation measures in the k-NN.

Based on the results of the experiment, the usage of the calculation based on *kNN* is effective in optimizing the data mining process, creating request models also, rules for rural crops characterization. This may be ascribed to the enhancement qualities of the GA in the data mining process. Further observation, the classifiers may have dependencies on the nature of the datasets, particular to the characteristic data sorts also, in the pre-preparing method used in utilizing genuine field data of rural crops.

VII. RECOMMENDATION ALSO FUTURE WORK

It is recommended that comparative studies may too be undertaken in utilizing other preparing strategies also, further study on the *kNN*. Further validation also, assessment of the proposed strategy is too recommended utilizing other rural datasets with varying characteristic data types, also, utilizing the results herein as benchmark data. The request models also, rules created also, introduced can be used as baseline framework in the development of canny systems for precision agriculture.

Future work includes further study to use other effective separation measures in the k-NN data mining request calculation not introduced herein also, utilizing only effective separation measures as the wellness capacity is being considered.

REFERENCES

[1] Hongqi Li; Haifeng Guo; Haimin Guo; Zhaoxu Meng, "Data Mining Techniques for Complex Formation Evaluation in Petroleum Exploration and Production: A Comparison of Feature Selection and Classification Methods", Computational Intelligence and Industrial Application, 2008. PACIIA '08. Pacific-Asia Workshop on Year: 2008, Volume: 1 Pages: 37 – 43.

[2] Sarojini Balakrishnan; Ramaraj Narayanaswamy; Nickolas Savarimuthu; Rita Samikannu, "SVM ranking with backward search for feature selection in type II diabetes databases", Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on Year: 2008 Pages: 2628 – 2633.

[3] Liangpei Zhang; Yanfei Zhong; Bo Huang; Jianya Gong; Pingxiang Li, "Dimensionality Reduction Based on Clonal Selection for Hyperspectral Imagery", IEEE Transactions on Geoscience and Remote Sensing Year: 2007, Volume: 45, Issue: 12, Pages: 4172 – 4186.

[4] Abdelaziz Kallel; Catherine Ottle; Sylvie Le Hegarat-Mascle; Fabienne Maignan; Dominique Courault, "Surface Temperature Downscaling From Multiresolution Instruments Based on Markov Models", IEEE Transactions on Geoscience and Remote Sensing Year: 2013, Volume: 51, Issue: 3 Pages: 1588 – 161.

[5] Shuang Hong Yang; Bao-Gang Hu, "Discriminative Feature Selection by Nonparametric Bayes Error Minimization", IEEE Transactions on Knowledge and Data Engineering Year: 2012, Volume: 24, Issue: 8 Pages: 1422 – 1434.

[6] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir,Kallal Das, Faridur Rahman, Rashedur M Rahman "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh " IEEE paper Issue 1, june 2015

[7] D. Diepeveen and L. Armstrong, "Identifying key crop performance traits using data mining" World Conference on Agriculture, Information and IT, 2008.

[8] Mohammad Motiur Rahman, Naheena Haq and Rashedur M Rahman "Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield", 17th. IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, 2014.

[9] Soils,Plant growth and crop production- Vol.I-Climate and its Effects on Crop Productivity and Management-S Mark Howden,Devid H White.

[10] A Survey on Data Mining Techniques in Agriculture.M.C.S.Geetha Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.

[11] Sudarshan Reddy S, Vedantha S, Venkateshwar Rao B, Sundar Ram Reddy and Venkat Reddy. Gathering Agrarian Crisis Farmers Suicides in Warangal district. Citizens Report, 1998.

[12] Anderson, W. K. (2010). Closing the gap between actual and potential yield of rainfed wheat. The impacts of environment, management and cultivar.. Field Crops Research, 116(1), 14-22.

[13] Asseng, S., & Pannell, D. J. (Adapting dryland agriculture to climate change: Farming implications and research and development needs in Western Australia. Climatic Change, 1-15, 2012.

[14] Phillip McKerrow, Member, IEEE, and Neil Harper, "Plant Acoustic Density Profile Model of CTFM Ultrasonic Sensing", IEEE SENSORS JOURNAL, VOL. 1, NO. 4, DECEMBER 2001.

[15] Raorane A.A, Kulkarni R.V, "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 1, Issue 2, July – August 2012

[16] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

