International Journal of Multidisciplinary Approach and Studies

ISSN NO:: 2348 – 537X

Knowledge Accomplishment and Future Prediction Battle; for the Universe of Big Substances

P. Nithya,

Assistant Professor, Department of Computer Science and Engineering,

Government College of Engineering, Tamil Nadu, India.

ABSTRACT

Each dimension of the Universal Substances is varied based upon every global atomic motion. The prediction over the past statistics for the future enhancement and enrichment is in need to develop the human kind in all aspects. Agriculture, the central system of our country needs severe attention to retain its original position to withstand in the current techno war. Mere prediction and assumption does not help to improve its grade.

Complex information should be framed from the data observed on different angle of correlation. Accurate mining of data should be done with efficient algorithmic methods, should also include perfect presentation scenario too. This is our work on future prediction using the precise knowledge consummated from Distributed Mining Strategy for the universe of big and essential substances.

Keywords: Distributed Mining Strategy, Dimension, Universe, Agriculture.

I. INTRODUCTION

The general idea behind all data base maintenance, data extraction, data mining and all kind of data analysis process is the knowledge accomplishment and future prediction. This kind of analyses always leads to the global development in the concerned field. The universe is at the wealthiest state in the Human-technology relation but not in the Human-environment relation.

Agriculture of the main cores under Human-environment relation which needs concentrated attention to improve its state of instant. Agriculture is the root for living beings. Though all humans are rich with technology we cannot live only with technology have to concentrate on our environmental living dimensions too. The cultivation of crops and all other essential items will be included under this category. The one and only work is to take care of all past statistic and to derive the perfect knowledge from that data to gain more enhancement percentage in the future.

The main basics are to re analyze all mining phenomenon employed for agriculture field. Some of the mining strategies are K-Means, K-Nearest neighbor, artificial neural networks, Support vector machines and multiple linear regressions. The three main divisions are principle component analyses, regression model and bi-clustering technique. Now the research war is only within mining strategy - around the cultivation field. The gist of technology planned to analyze are listed here

1. **K-means method**
2. **Neural Network methods**

2.1 Multilayered perception

2.2 Radial basis function.

3. Regression technique.

3.1 linear regression vector

3.2 Support regression vector

3.3 Decision regression vector

4. Component analyses.

5. Bi-clustering technique.

The listed methods are all existing scenarios in the era of data mining. All we need is generic distributed data mining method for the universe of big substances. The term universe of big substances define agriculture field.

II. OVERVIEW OF THE RESEARCH

Our research circles around the North West zone of the Tamil Nadu. The parameters taken under concern are Timeline, Soil kind, Rain Fall, Location, Irrigation mechanism, Production variety. The data are collected for time period from 2000 to 2010 for in and around North West Zone of Tami Nadu. The complete analyses gave perfect solution for this problem. The major parameters under taken are soil type, weather condition and production factor. Any mining strategy should concentrate on both production and process face. All the algorithm mentioned above are all examined over computational cost and also the integrity level.

This North West Zone of Tamil Nadu is taken with 20 notable places with seven different kinds of soil and 6 different kinds of production and the meteorological locational change per place.

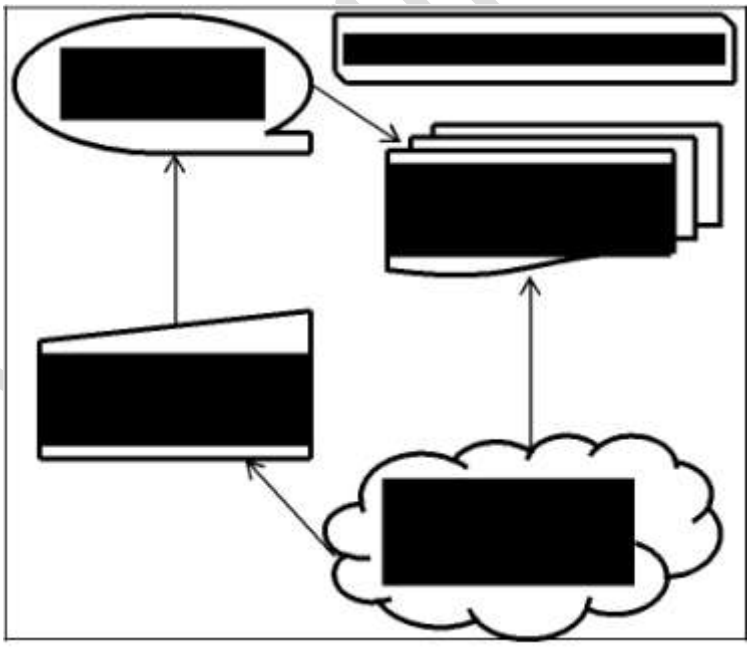


Figure: 1 – Overview of Research

III. DATA SET

As mentioned earlier the dataset taken here is for past ten years 2000 to 2010 and it is composed of seven soil kinds namely, red calcareous, red non calcareous, brown calcareous, brown non calcareous, black soil, alluvial soil and mixed soil . The basic relationship between soil and crop type should be framed as default intro information. This will help to improve the correlation and classification of the data further. For Example it is known that black soil is for cotton, paddy. This method of basic invocation of some messages makes the knowledge extraction process very clear and useful too. Need is to generate default info gallery to abstract some valuable proven knowledge.

The area North West Zone is covered over twenty places Salem, Veerapandi, Panamarathuppatti, Ayothiyapattinam, Valappadi, Yercaud, Attur, P.N. palayam, Thalaivasal, Sankagiri, Magudanchavadai, Edapadi, Konganapuram, Kolathur, Mecheri, Nagavalli, Omalur, Gangavalli, Tharamangalam and Kadayamapatti.

Major crop varieties are paddy, pulses, millets, oil seeds, cotton and sugarcane. The rainfall level, production estimation level and coverage production area are all observed and tabulated in the following.

The complete dataset we taken is shown here as snapshot of our observance is pasted here which shows the variation with all predictors and predictant. The basic calculation is always for Yield prediction (YP). The Total Production Factor (TPF) derived from all resources and target will decide the future yield.

IV. ANALYSES PROCESS

For processing the agriculture dataset different kind of algorithms are in use for example k-means method is employed for data which having no prior information. The neural network terminology is for identifying the quality of the product by x-ray method. All research work under this category continues to debate for its on method now is the time to generalize all the results to form the centralized.

The following tabulation shows our observed value with all dimensional change. Each table illustrated here pictures the importance of Dependent and independent parameters. Merging all tables the important and the unified goal is to predict necessary and distributed mechanism situations for high yield prediction (YP) through the total production factor (TPF). Total Production factor for agriculture dataset mainly depends on the quality and nature of the algorithm employed to derive.

International Journal of Multidisciplinary Approach and Studies

ISSN NO:: 2348 – 537X

SALEM DISTRICT

SOIL TYPE /FERTILITY INDEX

Sl. No.	BLOCK	pH	EC	Texture	Nutrient status N	P	K
1	Salem	7-8.7	0.50	sc(,c)	90-103	3.9-5.1	80-100
2	Veerapandi	7-8.7	0.50	s(,sc)	80-100	2.8-4.9	69-98
3	Panamarathuppatti	6.6-9.0	0.10	Cl,SI,Rl	95	4.6-8	49-115
4	Ayothiappattinam	6.6-8	0.02	Scl,Sl,Cacl	95	5.1	85
5	Valappadi	7-8.7	0.50	Sl, Scl, cl	94	4.4	98
6	Yercaud	5.6-6.5	0-1	Cl,	195	4.2	79
7	Attur	7-8.5	0.50	s(,c)	96-140	4.1-8.5	79-235
8	Pethanaickenpalaya m	7-8.1	0.50	Cl, Sl, Scl	78	3.3	67
9	Gangavalli	7.0-8.5	0.50	s(,sl	103	4.9	76
10	Thalaivasal	7-8.5	0.50	Cl. Sl. Scl	106	4.8	80.2
11	Sankagiri	8.1-8.7	0.50	sl	116	18	164
12	Magudanchavadi	7.8	0.50	sl	100	4	78
13	Edappadi	7-8.7	0.50	SL, SCL, CL	104	6	85
14	Konganapuram	7-7.8	0.50	scl	82-112	4.4-6	71-120
15	Kolathur	7.1	0.10	s	130	4.5	48
16	Mecheri	7.1	0.50	s	82-109	3.3-4.5	70-110
17	Nangavalli	7-7.8	0.50	sl	23-115	2.4-6.5	67-147
18	Omalur	7.9	0.20	sl	74-140	2.4-6.6	48-116
19	Tharamangalam	8.5-2.8	0.37	scl	83-152	3.5-5.2	54-105
20	Kadayampatti	7-7.8	0.50	sl	101	4.5	76

sl-Sandy loam; cl - clay loam; c - clay; scl - sandy clay loam

Normal Range	N	P	K
Low	<113	<4.6	<48
Medium	113-181	4.6-9	48-113
High	>181	>9	>113

SALEM DISTRICT

SOIL TYPE /FERTILITY INDEX

Sl. No.	BLOCK	Soil series
1	Salem	slm-49.54%, smy-9.37%, mal-6.09%, tlk-5.63%, plm-1.84%
2	Veerapandi	slm-49.54%, smy-9.37%, mal-6.09%, tlk-5.63%, plm-1.84%
3	Panamarathuppatti	slm-49.54%, smy-9.37%, mal-6.09%, tlk-5.63%, plm-1.84%
4	Ayothiappattinam	slm-49.54%, smy-9.37%, mal-6.09%, tlk-5.63%, plm-1.84%
5	Valappadi	slm-49.54%, smy-9.37%, mal-6.09%, tlk-5.63%, plm-1.84%
6	Yercaud	oty-53.49%, kpg-12.25%, ycd-5.96%, ngr-0.14%, soil associations-21.16%
7	Attur	lgr37.18%,Pyk12.92%, Tlk-8.97%, Plm4.82% & forest-36.11%
8	Pethanaickenpalayam	lgr37.18%,Pyk12.92%, Tlk-8.97%, Plm4.82% & forest-36.11%
9	Gangavalli	lgr37.18%,Pyk12.92%, Tlk-8.97%, Plm4.82% & forest-36.11%
10	Thalaivasal	lgr37.18%,Pyk12.92%, Tlk-8.97%, Plm4.82% & forest-36.11%
11	Sankagiri	tlk-54.66, irg-18.36%, chi-13.54%, upi-2.72%, mam-2.71%, cvp-1.2%, amy-1.1%, slm-0.2%, mlr-0.97% & forest 4.54%
12	Magudanchavadi	tlk-54.66, irg-18.36%, chi-13.54%, upi-2.72%, mam-2.71%, cvp-1.2%, amy-1.1%, slm-0.2%, mlr-0.97% & forest 4.54%
13	Edappadi	tlk-54.66, irg-18.36%, chi-13.54%, upi-2.72%, mam-2.71%, cvp-1.2%, amy-1.1%, slm-0.2%, mlr-0.97% & forest 4.54%
14	Konganapuram	tlk-54.66, irg-18.36%, chi-13.54%, upi-2.72%, mam-2.71%, cvp-1.2%, amy-1.1%, slm-0.2%, mlr-0.97% & forest 4.54%
15	Kolathur	lgr-59.3%,tlk-4.21%, smy-3.68%, cvp-1.66, mal-0.35, plm-0.18, upi-0.17 & forest 30.53%
16	Mecheri	lgr-59.3%,tlk-4.21%, smy-3.68%, cvp-1.66, mal-0.35, plm-0.18, upi-0.17 & forest 30.53%
17	Nangavalli	lgr-59.3%,tlk-4.21%, smy-3.68%, cvp-1.66, mal-0.35, plm-0.18, upi-0.17 & forest 30.53%
18	Omalur	lgr-56.46%, tlk-18.21%, mal-2.66%, plm-2.25%, amy-1.23% & forest - 19.19%
19	Tharamangalam	lgr-56.46%, tlk-18.21%, mal-2.66%, plm-2.25%, amy-1.23% & forest - 19.19%
20	Kadayampatti	lgr-56.46%, tlk-18.21%, mal-2.66%, plm-2.25%, amy-1.23% & forest - 19.19%

lgr	Irugur	Plm	Paalamedu
Pyk	Periyanaickenpalayam	Forest & Hills	
Tlk	Tulukkanur	upi	Upparappatti
smy	Somayanur	chi	chittalandur
cvp	Chevadiparai	mam	mallasamuthiram
mal	Mallur	mlr	misc. land typ
oty	ooty	kpg	kombughooti
ngr	nagalur		

Class	Depth (in cm)	Soil series
Moderately deep	26-50	tlk, lgr, chi,

Table 1: Salem city with soil type and fertility index

AREA COVERAGE BLOCKWISE 2009-10 (in ha.)

SL. NO	BLOCK	PADDY	MILLETS CHOLAM	CUMBU	RAGI	MAIZE	THINAI	VARAGU	SAMAI	[illegible] R CHOLAM	OTHER MILLETS	TOTAL MILLETS	PULSES RED GRAM	GREEN GRAM	BLACK GRAM	HORSE GRAM	COWPEA
1	SALEM	1045	2857	0	2	27	0	0	0	0	0	2886	12	0	72	0	10
2	VEERAPANDI	1320	3464	13	14	888	15	15	0	0	0	4409	44	721	1005	38	2302
3	PANAMARATHUPATTI	1367	2763	0	4	425	0	0	0	0	0	3192	0	1	309	206	10
4	AYOTHIAPPATTINAM	2439	4679	353	386	1822	214	0	100	0	266	7820	234	213	1019	412	218
5	VALAPPADI	1070	2662	123	147	2769	5	0	93	0	2	5801	34	1	640	305	84
6	YERCAUD	293	756	0	913	366	0	125	98	0	0	2258	34	0	0	0	7
7	ATTUR	3645	4264	0	0	6236	0	0	0	0	1	10501	4	0	362	0	0
8	P.N.PALAYAM	2421	4434	34	159	5683	181	173	164	0	0	10828	250	0	416	386	85
9	GANGAVALLI	3820	1402	0	0	10061	0	0	0	0	0	11463	1	0	196	0	0
10	THALAIVASAL	4516	1242	12	0	10101	0	0	0	0	0	11355	42	3	510	1	0
11	SANKAGIRI	3884	5085	0	74	116	0	0	0	0	0	5275	277	1815	354	126	391
12	MAGUDANCHAVADI	454	5181	0	66	265	0	0	0	0	0	5512	90	908	202	317	594
13	EDAPPADI	1268	3792	11	727	382	0	0	0	0	0	4912	97	896	563	130	737
14	KONGANAPURAM	152	5173	5	152	132	0	0	0	0	0	5462	234	625	173	266	212
15	KOLATHUR	1593	2451	35	581	2122	0	0	25	0	0	5214	18	125	456	706	226
16	MECHERI	350	1665	293	4245	15	0	0	0	0	0	6218	129	498	597	885	1486
17	NANGAVALLI	466	1792	22	1225	53	0	0	0	0	0	3092	148	1635	377	901	2043
18	OMALUR	1578	4714	100	1156	99	0	0	0	0	0	6069	236	885	738	600	1366
19	THARAMANGALAM	423	2582	42	782	77	0	0	0	0	0	3483	61	2330	251	395	1298
20	KADAYAMPATTI	470	2451	159	452	59	0	0	0	0	0	3121	120	93	546	758	488
	TOTAL	32574	63409	1202	11085	41698	415	313	480	0	269	118871	2065	10747	8786	6432	11507

Table 2: Areas corresponding crop coverage area in ha units

IRRIGATION SOURCES						
Sl. No	Source	Nos.	Gross Irrigated area (Ha.)	Net Irrigated area (ha.)	Irrigation Intensity	% of contribution
1	Canal	78	2021	1981	1.02	1.89
2	Pond/ Tank	546	200	170	1.18	0.16
3	Tube Well / Bore well	9829	10893	9673	1.13	9.25
4	Open Well	115750	112253	92746	1.21	88.69
	TOTAL		***125366***	***104570***	***1.20***	
						Source: District G return - Fasli - 1419

Table 3: Complete irrigation resources in and around the North West Zone

SL. NO	CROP	ABSTRACT										
				2010-11				2011-12				
		Production (L.Mt.)		Area coverage (Ha.)		Productivity (kg/ha.)	Production (L.Mt.)		Area coverage (Ha.)		Productivity (kg/ha.)	Production (L.Mt.)	
		Target	Achievement	Target	Achievement		Target	Achievement	Target	Achievement		Target	Achievement
1	PADDY	1.634	1.512	43000	36787	6595	1.618	1.659	43000	28433	6958	1.634	1.331
2	MILLETS	1.846	2.009	90100	120479	1515	1.846	3.412	105000	127439	1651	1.846	3.745
3	PULSES	0.386	0.350	75933	59866	603	0.475	0.402	69515	46051	695	0.386	0.322
4	OIL SEEDS	1.034	0.541	64214	31576	1740	1.052	0.567	41900	23097	1605	1.034	0.425
5	COTTON	0.370	0.541	15000	17073	2.537	0.390	0.437	22100	16114	3.080	0.370	0.536
6	SUGARCANE	14.850	14.132	15000	13512	93.268	20.250	13.520	13600	16687	97.867	14.850	19.228
	TOTAL	3.866	3.870	303247	279293		3.934	5.473	295115	257821		3.866	5.399

Table 4: Crop And Yearly

Above illustrated will give the neat understanding of all parametric correlation. Soil type corresponding fertility index associated area coverage. Crop production verses area coverage. Irrigation terminology with area observed. Yearly crop estimation progress all these small relational calculation will leads to the final Total Production Factor (TPF). This TPF is extracted from the YP, which can be extracted from the complex correlation drawn from the simple dependency charts.

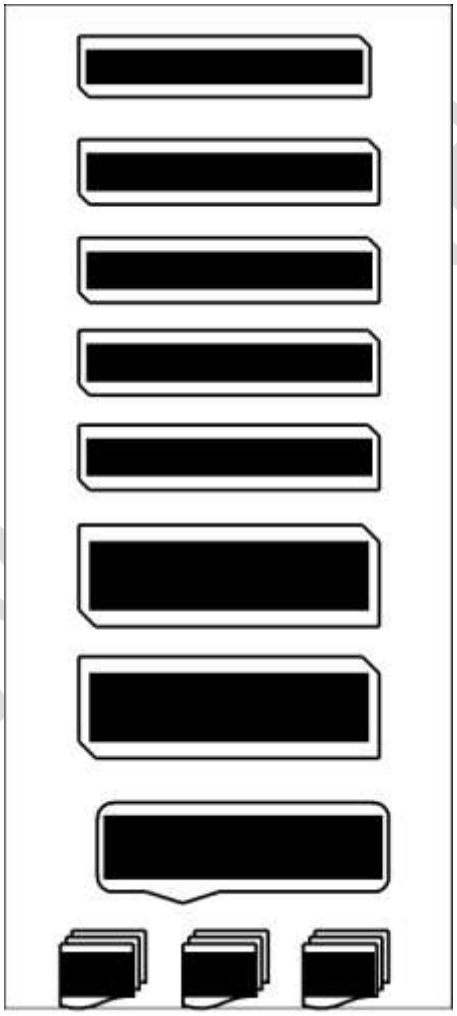
This principle of making complex relation to learn something very big from the simple dependency all over the scattered environment leads to the distributed data mining platform. This platform will ensure maximum winning probability in the battle of knowledge accomplishment and future prediction for the universe of big substances.

The algorithm procedures follows the following flowchart and steps demonstrated in detail.

1. Fix the problem scenario

2. Locate the problem area
3. List out the parameters in
4. Channel the randomness
5. Verify data strength index
6. Draw dependent and independent factor
7. Generalize the predictor and predictant
8. Predict
9. Check whether it obeys all rules stipulated
10. Calculate TPF – Total Production Factor.

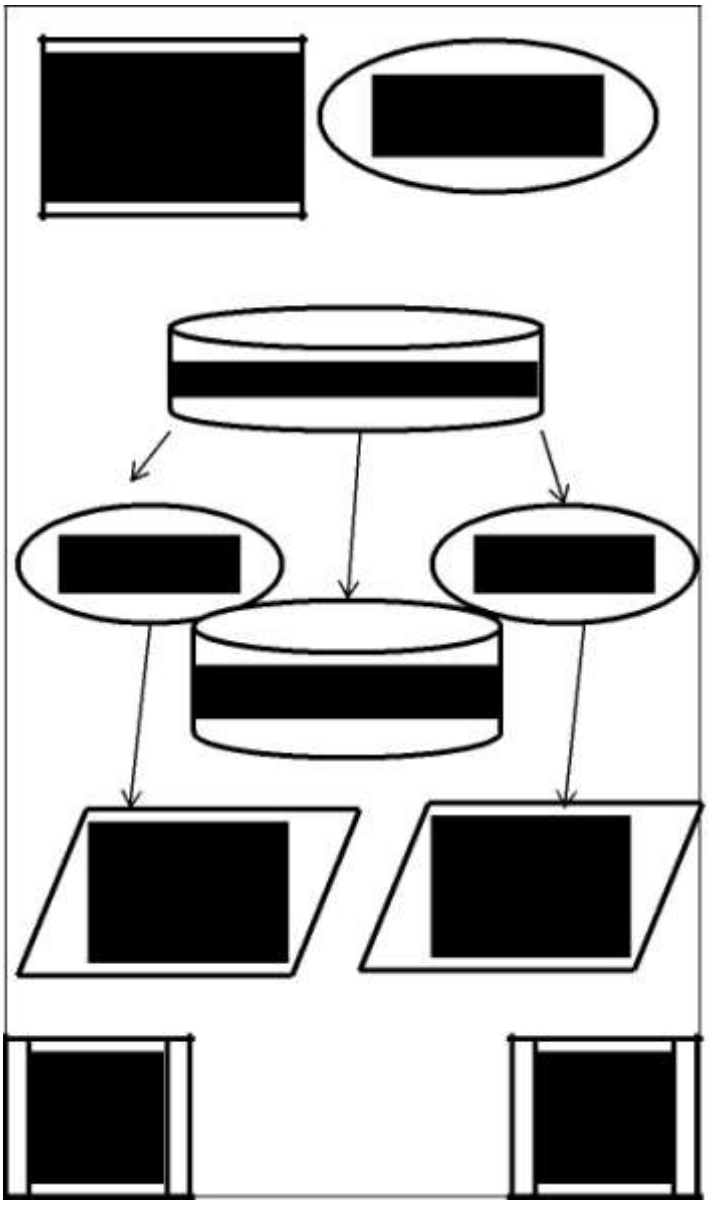
Conclude with the Yield Prediction derivation which should states the universal proven knowledge should be useful for further implementation and usage.



Another factor urging us to find the new knowledge extraction phenomenon is the cost of the algorithm. The algorithm cost factor that decides the cost of distributed knowledge extraction phenomenon are.

1. Storage Cost
2. Communication Cost
3. Computational Cost.
4. Private and Sensitive data intensity in the algorithm.

Above factors determine the algorithm efficiency. According to which the algorithm derived here will give consistent factor index in all directions. The layout of our algorithm is given below.



V. CONCLUSION

Finally, getting attached to the basic algorithms won't provide the possibility to accomplish the battle in the positive way. As stated earlier the Human-Technology relation is accelerating at high speed than the Human-environmental relation, which means the Human-Environment-Technology relation is needed to empower the human community. Human environment and technical related area is Agriculture which is also the new born area in Data mining and knowledge extraction. Hope the given principal analyses made on the existing algorithm and the final concept of Distributed Data Mining will enrich the future scenario. The final statement of assurance will be given that the future Digital World will make use of this algorithm for the knowledge accomplishment in the battle for the universe of big substances.

VI. REFERENCES

i Data Mining Techniques and Applications to Agricultural Yield Data - D Ramesh1, B Vishnu Vardhan Associate Professor of CSE, JNTUH College of Engineering , Karimnagar Dist., Andhra Pradesh.

ii Farming by the Numbers Precision agriculture

iii brings computing's accuracy to the ancient art of cultivation By Ariel Bleicher .

iv Feeding the World With Big Data. Agriculture experts say that open data could lift people out of poverty By Prachi Patel.

v Agricultural Best Management Practices (BMP) Database By - Wright Water Engineers, Inc. Geosyntec Consultants, Inc.

vi Re-focusing Environmental and Health Concerns of Nanotechnology By Dexter Johnson .

vii THE GROWTH OF WORLD AGRICULTURAL PRODUCTION, 1800-1938. Giovanni Federico Running Head: World Agricultural Production .