Dealing of Data Pre-processing with Computational Intelligence for Sampling and Analysis of Soil-Case Bidar District -Karnataka.

Sangayya Gulledmath¹, Arulkumar V²

¹ Research Scholar, Department of Computer Science and Applications, Reva University, Yelahanka, Bangalore-560064, India
² Assistant Professor, Department of Computer Science and Applications, Reva University, Yelahanka, Bangalore-560064, India

Abstract. Data Mining and Knowledge Discovery becomes buzzwords from last two decades. Data Mining as we all know its trending industry in Data Analytics and knowledge representation. This research article initiates novel approach of how Data Mining can be relatively used in Agricultural domain and their applications more dominant in the specified area of soil classification and soil fertility. Now Data Analysis is becoming part of our life why can’t we consider our own farmers to empower them to use more knowledge-oriented techniques to yield the good crop. I have taken Bidar District in Karnataka for soil Sampling and Analysis of Soil. Data pre-processing is challenging, important and critical step in the data mining process and it has enormous scope and it directly impacts the success of data mining process. There are many pre-processing techniques but which one makes better selection for agricultural data sets is vital task of the computational challenges that needs to address and incorporate more suitable one for the better representation of knowledge in soil sampling and analysis.

Keywords: Data Mining, Knowledge Discovery, Data Pre-processing, Agricultural Data Mining.

1. Introduction
Agriculture is the backbone of our country [1]. Which makes potential contribution of our countries GDP [Gross Domestic Per Capita]. This is the reason where many states put their highest efforts to potentially enhances the agricultural irrigation and good cultivation of the crops. The many of the policies in India either on paper or these policies are not framed by keeping rural farmers in mind. Now if we can’t think of technical incorporation and computation facility for advance agricultural system in India then we may start depend more on import of food products rather than exporting to the other nation. As demand of food products increases then government, researchers, scholars, Institutions and Agricultural scientists put more efforts to encourage use of technology to increase the production of agricultural products to meet the shortages. In this regard use of technology enhances the productivity of the agriculture but which technology is useful and what way it helps
to emerge as the biggest booster to motivate farmers and their morale to walk extra mile for their growth. In this regards this paper helps to find suitable approach of Data Mining techniques [2] as whole process but generally without proper data sets mining becomes fragile and produces the mis leading information to the users. Data Mining can be used for predicting and exploring the new avenues of agricultural domain. Data Pre-Processing [3] initiates the best plan for Data Models and Prediction [4].

The primary intention of the discussion here is that data mining has fascinated many people for a great deal information industry like ICT and Internet of the things [IOT] and allied applications its purely because of large amounts of data availability through applications and plenty of data repositories in government portals and the challenge is converting such large amount of data into useful information and knowledge representation [5]. The information and knowledge gained from various applications or databases can be used for applying in real time applications ranging from business management, production control, and market analysis, to engineering design technology and science exploration. Data mining effectively leads the way, as a result of the natural and present trend & growth for information technology.

This natural phenomenon of Data Mining has greatly witnessed in the database industry in the development of the following main functional areas like data collection and creation of database, Data base management techniques and other various objectives of the cause.

1.1 Introduction to Data Pre-Processing
As we know that huge data sets and various collection of data repositories always leads to disaster or mis managing the data and which in turn leads to wrong information if we can’t handle the data properly. Data Pre-Processing is often either neglected or ignored mechanism in the process of Data Mining applications. Many forget that the Pre-Processing in Data Mining is major step. It’s because data collection is usually process of loosely coupled or less controlled this may result huge disaster in producing the results. In some cases, analysing data that is never been considered for various reasons or lazy attitude. Let’s assume what happens if we can’t consider either model produces the misleading information or the accuracy of algorithm may be not be accurate for either predicting the data behaviour or exploring the correct possibility of data exploration and knowledge representation. If the data consist of unwanted and repetitive information or noisy with unreliable then we can’t produce the good Data representation model. To overcome all these the best method is to deal the Data Pre-Processing [4] this has answer for many reasons.

Data Pre-Processing includes the loosely coupled data can be prepared known as data preparation, included with tightly coupled data integration. Which further includes like Data Cleaning, Data Normalization and transformation of data [6]. Data reduction is a major task which includes techniques of feature selection, collective reference of instance selection, applied discretization etc [7].

1.2 Data Integration
Our biggest problem in this soil sampling and soil analysis is to collect the data from various sources and pick up the best data attributes which are useful for us. As we know technically speaking integration not so easy in this critical situation we reached Gandhi Krishi Vignyan Kendra Bangalore [GKVK Bangalore-www.uasbangalore.com] for various information to find the exact physical traces of soil and its properties. They further directed us to University of Agricultural Sciences Dharwad. After collecting various non-structured data now, the job is to integrate them in synchronising with other possible outcomes keeping in mind. If the data is not properly integrated then soon redundancies and inconsistencies will appear this will further hamper the process. The main objective of the integration is to build a data map that establishes how each instance should
be arranged in a readable structure to deal with real example like this. Now creating schema for the known attributes is required so that we can avoid the redundancies. More no of sources are flat files and converting them into table structure and later make them comma separated files for running pre-process algorithms.

Figure 1.1 Data Flow Illustration for integrating the data elements from various sources

1.3 Finding the Redundant Attributes

Redundancy is a problem that can be most addressable issue in data pre-processing and should be avoided as much as possible. It will generally disturb the outcomes on model sampling or while running algorithm. The great deal for data redundancy is either overfitting in selection or auto generation of data while integrating them or it can some time derived from the other attribute. In some cases, if attribute is itself redundant then it provokes the wrong signal for model. In our cases following attributes are prefiltered with future selection techniques. Then we thought let us apply certain practical aspects of avoiding redundancies then immediate thought for us to diagnosis the redundancies with detecting the factor of redundant values with correlation analysis. This means we wanted to check the how one attribute makes the strong presence into another by selecting non-required attribute.

Figure 1.2 Appearance of Redundancy Data in CSV file

Data Sampling Credit: National Council of Soil Survey and Land Usage Planning Bangalore
In above table if we refer there is place called Aurad one of the talukas in Bidar District and Village name is also Aurad. Now condition which Aurad is Either Barally Aurad or Aurad district. Then we went for impact of farmer no is table. Here intentional redundancies also make more important but repetitive can be filtered out. If the data is nominal and set of relevant values are finite its always suggest to go for $x^2$ (Chi-Squared) test is commonly applied. In numeric values using the techniques correlation coefficient and covariance is typical.

1.4  $X^2$ -Chi Squared Test

In General, suppose if we selecting the two nominal attributes A and B containing values of c and r distinct values each in notion having a1, a2, a3........ac and b1, b2, b3........ br. we can check the correlation between both of them using x2 test. In order to do so general contingency table with joint events to be created $(A_j, B_j)$ in which attribute A takes the values ai and attribute b takes the value of bj is created every possible joint event has its own entry in the table which we create then the statistic is compared as [8]

$$(x)^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} (o_{ij} - e_{ij})^2/e_{ij}$$

In above generalised mathematical equation where $a_{ij}$ is noted frequency of the joint event $(A_i, B_j)$ and $e_{ij}$ is the expected frequency of $(A_i, B_j)$ computed as in following equation.

$$e_{ij} = \frac{\text{count}(A = a_i)\times \text{count}(B = b_j) }{m}$$

Where m is the number of instances in the data set, count $(A=ai)$ is the number of instances with the value of ai for the given attribute A and Count $(B=bj)$ is the number of instances having the noted value bj for attribute B. The Chi Squared checks the hypothesis of given event are independent with $(r-1)C(c-1)$ degree of freedom. This statistical approach is obtained from equation 1 is compared with any either chi-squared values of event with given suitable degrees of freedom or any other available statistical tool. In case significance level of such table is more realistic then assumption of the table then null hypothesis is rejected and therefore A and B are statistically correlated.

1.5. Soils of the Study Area

Soil is the interface between atmosphere and lithosphere. It also interfaces with bodies of fresh and salt water generally called as hydrosphere. The soil sustains the growth of many plants and animals and which makes the biosphere. It is a combination of four main components like air, weathered rock, water and organic matter. The weathered rock [9] can be in the form of sand, silt, clay, pebbles, cobbles, boulders, etc. Organic matter can be addressed like it may generate from old leaves, even non-leaving animals and plants, or small living things. Presence of organic matter in a soil increases the water preserving capacity of soil and uses mechanism of infiltration. Soil acts as a reservoir of 14 mineral nutrients which are essential and fundamental important for normal plant development and crop production in any region of the earth. Soils are mainly derived from parent rocks. Soil is the residual result of physical disintegration and also identified as a chemical decomposition of the underlying rocks. In due course of time they are transformed into different colours and textures. The main factors influencing the formation of the soil are rock material, climatic conditions and slope of land, which reflects the geomorphic history of the area. Soils have characteristic drainage conditions that depend on surface runoff, soil permeability and internal soil drainage.

The porosity and permeability of soil influences and impact the groundwater potential and quality of water in any area. Soils are an important media promoting infiltration of rain or surface water to the underground. The infiltration of water through soils is dependent on texture, structure,
surface conditions, storage capacity and transmissibility of soils. A portion of rainfall infiltrates through the pores in the soil due to the gravity. The water is absorbed by the soil due to surface tension, hydroscope, etc.

The major soil types found in sub-basin of (Watersheds) Dakshina Pinakini River Basin are clayey, clayey mixed, loamy skeletal, fine loamy and rocky land (NBSS and LUP, 2001) (Credit source:Map 3.3-Reference of Geo Maps or Google earth maps).

a) Clayey Soil
The clayey soil generally found in southern watersheds (SW1, SW2 and SW3). It occurs on gently sloping upper pediments with slope ranging from 2% to 5%. This series represents very deep, gravely texture (Plate 3.2), moderate available moisture retaining capacity and brown to dark brown and red in colour. Clayey soils have good potential for cultivated crops and respond to fertiliser management, but require soil conservation measures (Map 3.3) [10].

b) Loamy Skeletal
The loamy skeletal soil found in northern, eastern and south-eastern (NW1, NW2, NW3, SEW and EW) watersheds of the study area. Soils of this association occur on undulating to rolling upper pediments with slopes of 5% to 10%. These soil series are very deep, red coloured and gravely in nature. The available moisture retaining capacity is 10.5 for 90 cm depth (Map 3.3).

c) Red Loamy Soil
Red loamy soils exhibit fine texture, found in the gently upper Pedi plains, which is found in northern, eastern and south-eastern (NW1, NW2, NW3, SEW and EW) watersheds of the study area. These soils have good water-holding capacity but poor retention. This type of soils is noticed in parts of Chikabalapura, Bagaluru, White field, Devanahalli, Hosakote taluks of Bangalore rural district.

d) Rocky Land
This type of soil occurs on undulating summits with outcrops of granite in large proportion and found in undulating uplands with 5% to 10% slope. These gravelly textured soils are resting on hard substrata, which have very low available moisture retaining capacity, very low nutrient status and are susceptible to erosion. This type of soil found in all watersheds of sub-basin.

e) Red Soils
Red soil in the study area is derived from the residual products of granites and gneisses. They are light textured, varying from sand or gravel to loams and are highly leached. The red colour soil is technical detailed its presence of iron, less inorganic content and plant nutrients. Red soils have good water-storage capacity. These are relatively more permeable than black soil and have good infiltration capacity. On the basis of texture, red soils further classified as follows:

f) Red Sandy Soils
These are poor in clay content and hence have a fairly good moisture-holding capacity. They exhibit coarse to medium texture, found in the upper slopes of pediments and upper Pedi plains, and thickness shows shallow to moderate depth. This type of soil found in most of the study area.

g) Red Gravelly Soils
These are derived from weathering of grey granites and gneisses. These soils occur on residual hills within pediments and in the severely eroded area of upper Pedi plains. The soil thickness varies from place to place. It is found to be very shallow on steeps having no vegetation. It is moderately thick on gentle slopes. These soils supporting vegetation of different types and density. Water and nutrients get drained quickly in this type of soils. This type of soil found in southern watersheds of the study area (SW1, SW2 and SW3). Red soils with different shades of red colour are found in parts of Bidar, Bhalki, Humnabad, Basavakalyan, Aurad and other parts of the study area.

In agriculture, a soil test common technique which refers to the analysis of soil sample to determine various nutrient contents, general composition and other characteristics such as the
acidity or PH [10] level in common. The analysis of data clearly indicates that the cropping pattern in India has undergone major changes over time.

The stark observation made in the economic survey of 2015-16 that “Indian Agriculture, is in way, a victim of its own past success especially the green revolution”. The green revolution, which is often characterized by introduction of high-productive variety of seeds and use of minimum fertilizers, without any second thought it increases the productivity of land considerably.

In India 70%-80% of water resources are used for agriculture, due to many factors like increase urbanization structures in city and Industrialization land acquiring [11] and also forced use of agricultural chemicals and fertilizers, problem of rampant water pollution results ground water depletion, water logging, salinity and desertification increases in many cases. The testing Labs situated in Bangalore and Dharwad taking initiative to educate the farmers about various scientific tools and mechanisms for identifying best soil and water for Agricultural purposes. They analyse the soil and water sample from different farms and regions across Karnataka and assess the quality of irrigation water by providing appropriate and corrective recommendation for using regionally dominant quality of waters to eliminate many soil related issues such as alkalinity, salinity, acidity etc., which surely result in increased agricultural output within short span of time [12].

1.6 Architecture of the System

Building model-based approach and simple mechanism of how effective agriculture data can be used and incorporated in real world of machine learning is major task in this research paper. Identifying noisy data is major concerned and can be addressed with either robust learning mechanism, Data polishing methods or readily available noise filters can be used to test. As we never dealt with any robust learning mechanism because noise level is relatively low. However common practice of machine learning mechanism like bit of polishing and some readily available filters are tested in this scenario and built testing soil models on zonal wise or call it as taluka wise data sets. Simple architect of our system is to select the two aspect of noise like either class noise or check with attribute noise. We initiated some class noise which is generally called as label noise. There are some data sets which duplicates having different class labels like pody number and farmers id of pody number and survey. Some of the data labels different from their original labels. Some of data which we polished like attribute either or missing or incomplete like soil type is too general or no comment data attributes are just ignored and considered for the sampling.

2.Method of Soil Sampling

A soil surveyor generally determines texture by moistening a soil sample until it glistens and kneading between the cross fingers and thump then aggregates are broken down the soil grains thoroughly wetted. The proportion quantity of sand, loussy silt and clay are observed and recorded according to the following qualitative criteria.

1) Coarse sand grains are large enough to grate against the each other and can be detected either by expert sand surveyor or sight and feel
2) Fine sand grains are much less obvious but when more than about 10-15% sample detected by feel and touch
3) Silt grains generally collected and detected with just feel. However, their presence makes soil feel smooth and soapy with little sticky features.
4) Clay is self-explanatory of nature like sticky or sometime moistening and kneading before they develop the maximum stickiness.
5) Allocation of soil sampling sites which is determined by the places.
3. Methods of Soil Analysis

Collected samples as we required (preferably not less than 500 gm fresh soil) should be stored and sent to lab in UAS Dharwad. All samples have to be prepared according to standard methodology in order to extract original properties in chemical and mineral. It should be air dried at 40 °C. They then can be stored until analysis. Keeping all particles intact with microscopic roots by protecting mineral and organic with diameter larger than 2mm, should be used to remove samples by dry sieving. The remaining soil particles which not passing the 2mm sieve may be kept for weighing separately for the determination of the coarse fragments contents. The fraction smaller than 2 mm is used for soil analysis. Samples may be milled or ground immediately before analysis for the next level. The procedure following techniques like Drying, Seperation of fraction<2mm, milling and sub sampling.
Soil Data Sheets - Table 1.

| Determination of Soil Moisture content |
|----------------------------------------|
| Method Sheet                           | Summative code: SAM01 |
| Reference Methods                      | ISO 11465             |
| Method Suitable for                    | Organic Layer, Mineral Layer |
| Procedure:                              | Analytical Balance Accuracy: 0.001 g |
| Tightly coupled tins or flasks with    | Mineral Layer: Dry at 105±/− 5 °C |
| drying oven with closely fitting lead  | Here Lid is removed until constant mass is reached. |
|                                        |                        |
| Calculation                             |                        |
| The moisture content in weight is      |                        |
| calculated by                      |
| Moist%=(A-B)/(B-tare time X 100)       |                        |
| A: Weight of tared moisture tin and    |                        |
| air-dried soil.                       |                        |
| B: Weight of tared moisture tin and    |                        |
| oven dried soil sample.               |                        |
| Correction Factor: 100+%-moist/100     |                        |

Applied Pre-Processing Techniques in Weka: In this research paper we used open source tool for the data Pre-processing technique, which is to be applies on various data sources collected from different source files and use the filter technique.

Table 2. Data Sets collected for Farmer profile of each piece of land. (Data sets credit: NCSR)

| K | L | M | N | O | P | Q | R | S | T | U | V | W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Card No | District | Ec | Farmer Name | Farmer Nu Oc | Ph | Sau | Soil Type | State | Survey Nu Taluk | Village |
| CP9000 | Bidar | 0.1 | Parwaal_Ballaram | K 21081 | 0.36 | 7.4 | RAU | Red | Karnatako | 145/2 | Aurad | Aurad (B) |
| CP9001 | Bidar | 0.27 | Bandappa_Anangonda | K 21082 | 0.64 | 8 | RAU | Black Soil | Karnatako | 182/A | Aurad | Aurad (B) |
| CP9002 | Bidar | 0.16 | Iaswaan_I_Masamathappa | K 21083 | 0.54 | 8.3 | RAU | Black Soil | Karnatako | 58 | Aurad | Aurad (B) |
| CP9003 | Bidar | 0.21 | Basappa_Shivarayya | K 21084 | 0.58 | 8 | RAU | Black Soil | Karnatako | 221 | Aurad | Aurad (B) |
| CP9004 | Bidar | 0.24 | Bandappa_Margonda | K 21085 | 0.71 | 8.1 | RAU | Black Soil | Karnatako | 150/B | Aurad | Aurad (B) |
| CP9005 | Bidar | 0.23 | Shivamath_Kalappa | K 21086 | 0.65 | 8.2 | RAU | Black Soil | Karnatako | 277 | Aurad | Aurad (B) |
| CP9006 | Bidar | 0.2 | Nagamath_Shankar | K 21087 | 0.59 | 8.2 | RAU | Black Soil | Karnatako | 275 | Aurad | Aurad (B) |
| CP9007 | Bidar | 0.19 | Baswaan_Tribhannao | K 21088 | 0.55 | 8.3 | RAU | Black Soil | Karnatako | 1 | Aurad | Aurad (B) |
| CP9008 | Bidar | 0.22 | Basappa_Sengappa | K 21089 | 0.6 | 8.2 | RAU | Black Soil | Karnatako | 232 | Aurad | Aurad (B) |
| CP9009 | Bidar | 0.17 | Basappa_Siddappa | K 21090 | 0.57 | 8.2 | RAU | Black Soil | Karnatako | 225 | Aurad | Aurad (B) |
| CP9010 | Bidar | 0.12 | Shivamand_Sumukshappa | K 21091 | 0.72 | 7.9 | RAU | Black Soil | Karnatako | 58 | Aurad | Aurad (B) |
| CP9011 | Bidar | 0.12 | Samrubi_Basappa | K 21092 | 0.36 | 7.5 | RAU | Black Soil | Karnatako | 170/6 | Aurad | Aurad (B) |
| CP9012 | Bidar | 0.13 | Donida_Banapatirao | K 21093 | 0.43 | 7.8 | RAU | Black Soil | Karnatako | 275/7 | Aurad | Aurad (B) |
| CP9013 | Bidar | 0.14 | Prabhurao_Shankappa | K 21094 | 0.52 | 7.9 | RAU | Black Soil | Karnatako | 107/1/ | Aurad | Aurad (B) |
| CP9014 | Bidar | 0.16 | Sangayya_Ballangayya | K 21095 | 0.65 | 8 | RAU | Black Soil | Karnatako | 58036 | Aurad | Aurad (B) |
| CP9015 | Bidar | 0.13 | Baswaal_Ganpathi | K 21096 | 0.35 | 7.0 | RAU | Black Soil | Karnatako | 247 | Aurad | Aurad (B) |
| CP9016 | Bidar | 0.07 | Baswaal_Gadgappa | K 21097 | 0.49 | 8.5 | RAU | Black Soil | Karnatako | 231 | Aurad | Aurad (B) |
| CP9017 | Bidar | 0.21 | Bhavamunna_Shivangappa | K 21098 | 0.86 | 8.1 | RAU | Black Soil | Karnatako | 266/1 | Aurad | Aurad (B) |
| CP9018 | Bidar | 0.58 | Vidyaant_Kalappa | K 21099 | 0.58 | 7.6 | RAU | Black Soil | Karnatako | 29/1A | Aurad | Aurad (B) |
| CP9019 | Bidar | 0.26 | Ganpathi_Shankarappa | K 21000 | 0.52 | 7.9 | RAU | Black Soil | Karnatako | 241/1 | Aurad | Aurad (B) |
| CP9020 | Bidar | 0.19 | Shivangappa_Hanmanath | K 23361 | 0.8 | 8.1 | RAU | Black Soil | Karnatako | 73 | Aurad | Ladha |
| CP9021 | Bidar | 0.42 | Homnantaapppa_Shivlinga | K 23362 | 0.74 | 8 | RAU | Black Soil | Karnatako | 86 | Aurad | Ludha |
| CP9022 | Bidar | 0.24 | PramilaBabbaVithalrao | K 23364 | 0.46 | 7.9 | RAU | Black Soil | Karnatako | 68 | Aurad | Ladha |
| CP9023 | Bidar | 0.29 | Vijnanath_Hanmanthrao | K 23365 | 0.7 | 8 | RAU | Black Soil | Karnatako | 87 | Aurad | Ladha |
| CP9024 | Bidar | 0.56 | NagachettyIshodappa | K 23366 | 0.52 | 7.9 | RAU | Black Soil | Karnatako | 8 | Aurad | Ladha |
| CP9025 | Bidar | 0.18 | VijayakumarManikrao | K 23367 | 0.53 | 8.1 | RAU | Black Soil | Karnatako | 14 | Aurad | Ladha |
| CP9026 | Bidar | 0.26 | NagachettiShivangappa | K 23368 | 0.56 | 7.9 | RAU | Black Soil | Karnatako | 5 | Aurad | Ladha |
| CP9027 | Bidar | 0.27 | ShankarappaBallappa | K 23369 | 0.43 | 8.1 | RAU | Black Soil | Karnatako | 4 | Aurad | Ladha |
Table 3. Data Sets collected for mineral profile of each piece of land. (Data Sets Credit: NCSR)

Following figures illustrate soil survey sampling images by applying attribute balance filter and results were displayed through visual images.

![Table 3 Data Setts collected for mineral profile of each piece of land](image_url)

**Figure 1.7** farmers profile Pre-Process  
**Figure 1.8** Farmers Profile Visualisation
4. Results and Conclusion

In this research paper, we have experimented various pre-processing mechanisms for finding the best data of farmers profile and their soil of identified pedo number for testing their soil and effective of fertility of soil profiles using correctly classified instances and incorrectly classified instances for various soil testing criteria such as separation of the mineral part of the soil into various size fractions and determination of the proportion of these fractions. The analysis in this research includes all soil material including gravel and coarser material of mapped soils using data filters. The main objective is to test the pre-treatment of the sample intern in complete dispersion of the primary particles. Therefore cementing materials such as zink and potassium is not binning for the soil which is available. Further lot of studies can be done for balanced soil profile using pre-processing data methods. Future scope of the paper is many aspirants can work on various soil profiles of their region and fertility factors. This is just beginning how technology can be integrated in agricultural domain to enhance the soil fertility and empowering the farmers about the various possibility to be encashed in betterment of the farmers life and contribution of the national economy.

5. References

[1] M.C.S Geetha, 2015, A Survey on Data Mining Techniques in Agriculture, *International Journal of Innovative Research in Computer and Communication Engineering* Vol 3(Issue 2) pp 887-892

[2] P Grace Sharon 2014 Application of Data Mining in Agriculture *International Journal of Computer Science and Information Technology Research* Vol 2, Issue 3 pp (158-162)

[3] Agarwal, Rakesh and Ramkrishna Srikant, 1994, Fast Algorithms for Mining and Pre-processing Association Rules, *Proceedings of 12th VLDB Conference, Santiago Chilli* pp 456-462
[4] Agarwal, Rakesh and Ramkrishna Srikant, 1995, Active Data Mining and Clustered Models, *Proceedings knowledge Discovery and Data Mining [KDD-95] Menlo park*, pp 385-394

[5] Namita Mirjankar, Smitha Hiremath, May 2016, Applications of Data Mining in Agriculture Field, *International Journal of Computer Engineering and Applications, ICCSTAR-2016, Special Issue*, pp 363-368

[6] S.Garsia et al, 2015, Data Pre-processing in Data Mining, *Intelligent Reference Library 72 DOI 10.1007/978-3-319-10247 Springer International Publishing Switzerland, 213-221

[7] S.B.Kotsiants, D.Kanellopoulos, P.E.Pintelas, 2007, Data Pre-processing for Supervised Learning, *Worlds Academy of Science Engineering and Technology International Journal of Computer and Information Engineering*, Vol 1 No 12 pp 4104-4109

[8] Cohen W.W, 1998, Integration of heterogenous databases without common domain using queries based on textual similarity, *Proceedings of ACM SIGMOD International conference on Management of Data New York*, pp 201-212

[9] R.E.White, 1979, *Introduction to the Principles and Practice of the Soil Science*, Blackwell Scientific Publications, Osney Mead, Oxford 1979 pp:3-8 & 9-23

[10] Handbook and Manual 2003, Methods and Criterial for Harmonized Sampling, *International Cooperative Programme on Assessment and Monitoring of Air Pollution Effects on Soil*, pp 2-30

[11] Survey Report’ etc, 2011-12 Proceedings on Land Usage and Monitoring critical review, National Bureau of Soil Survey and Land Use Planning Hebbal Bangalore pp 02-89

[12] Handbook and Manual 2003, Soil Solution Collection and Analysis, *International Cooperative Programme on Assessment and Monitoring of Air Pollution Effects on Soil*, pp 30-96