A Study of Application of Data Mining in Rural Development of Kaithal Region

Arvind Kumar
Research Scholar, Department of Computer Science
Baba Mast Nath University, Rohtak,

Dr. Satpal
Professor, Department of Computer Science
Baba Mast Nath University, Rohtak,

Abstract

This paper focuses on the application of Data mining Techniques to efficiently manage the challenges faced by rural areas of Kaithal district in analysing and predicting development level and better manage their economy. This technology can offer vast opportunities and capabilities to rural areas which will in turn aid in effectively solving many problems associated with such areas. Information on rural development is not only crucial but the maintenance and analysis of data is considered one of the main subjects in future decision making and improvement in rural life. Strengthening data collection and analysis provides perhaps the greatest opportunity for researchers and policy makers for rural areas. Utilization of such data is considered a driving force that improves the economic dynamism while creating a new type of knowledge-based economy. Many countries are adopting Data mining applications to Rural Development. The aim of this paper is to study applications of data mining to rural areas for rural development. Classification, prediction, class description and exploratory analysis are the most often used methods in literature.

Keywords— Rural Development (RD), Classification, Clustering, Data Mining (DM); Data Warehouse (DWH);

1. INTRODUCTION

Rural development: Rural development is conceived as strategy aimed at finding ways to improve the rural lives with participation of the rural people themselves so as to meet the required need of the rural area. According to World Bank (2005), rural development is the process of rural modernization and the monetization of the rural society leading to its transition from traditional isolation to integration with the national economy. Also, rural development is perceived as a process of not only increasing the level of per capital income in the rural areas but also the standard of living of the rural population measured by food and nutrition level, health education, housing, recreation and security. Haryana is a leading contributor to the country's production of food grain and milk.. Greening rural development can stimulate rural economies, create jobs and help maintain critical ecosystem services and strengthen climate resilience of the rural people. Conversely, environmental challenges can limit the attainment of development goal. This will also assist in minimizing the rural-urban gap in terms of basic infrastructure facilities essential for 'Sustainable Development' of a settlement [16]. Rural Development Programme plays a significant role in the planning process.

Data Warehouse (DWH): A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site. Data warehouses are constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing. To facilitate decision making, the data in a data warehouse are organized around major subjects, such as customer, item, supplier, and activity. The data are stored to provide information from a historical perspective (such as from the past 5-10 years) and are typically summarized.

Data mining (DM): DM sometimes called as a data or knowledge discovery is a computer assisted process of digging and analyzing enormous set of data and then extracting the meaning of data. It automatically searches large volume of data for models & patterns using computational techniques from statistics and machine learning and information theory. It is the ideal tool for extracting knowledge from hidden data. It can predict future trends & behavior which can be used for decision-making [37].

2. LITERATURE SURVEY

This paper [1] discusses the application of Data Mining in Rural Planning and Management. Analytical Model of Data mining was used for model building and then the results are interpreted and Visualized for management and policy makers use.

Information technology is one the world's most important development aspects and many countries consider the development of information technology as one of the most vital aspects of infrastructure development. This technology can offer vast opportunities and capabilities to rural areas which will in turn aid in effectively solving many problems associated with such areas.

Information in the agricultural field and rural development is not only crucial but the maintenance and analysis of data is considered one of the main subjects in efficiency and capital improvement. Strengthening data collection and analysis provides perhaps the greatest opportunity for researchers and policy makers for rural areas. Utilization of such data is considered a driving force that improves the economic dynamism while creating a new type of knowledge-based economy. The main purpose of this paper is to assess the application of data mining in rural planning and management with the goal of achieving sustainable agricultural and rural development while promoting productivity in the agricultural field.

This paper [2] discusses the application of Data Mining in Rural Development. Here the Remote Sensing Technique was used for data collection. Based on Satellite Image analysis various villages were targeted for Rural Development policy implementation. Satellite imagery is a form of big data that can be harnessed for many social good applications, especially those focusing on rural areas. In this article, we describe the common problem of selecting sites for and planning rural development activities as informed by remote sensing and satellite image analysis. Effective planning in poor rural areas benefits from information that is not available and is difficult to obtain at any appreciable scale by any means other than algorithms for estimation and inference from remotely sensed images. We discuss two cases in depth: the targeting of unconditional cash transfers to extremely poor villages in sub-Saharan Africa and the siting and planning of solar-powered microgrids in remote villages in India. From these cases, we draw out some common lessons broadly applicable to informed rural development.

This paper [6] discusses the Data Mining for cluster naming of the villages at Java Island. Here clustering technique was used. Clustering of query based data mining to identify the meaning of the naming of the village in Java island, done by exploring the database village with three categories namely: prefix in the naming of the village, syllables contained in the naming of the village, and full word naming of the village which is actually used. While syllables contained in the naming of the village are classified by the behavior of the culture and character of each province that describes the business, feelings, circumstances, places, nature, respect, plants, fruits, and animals. Sources of data used for the clustering of the naming of the village on the island of Java was obtained from Geospatial Information Agency (BIG) in the form of a complete village name data

with the coordinates in six provinces in Java, which is arranged in a hierarchy of provinces, districts / cities, districts and villages. The research method using KDD (Knowledge Discovery in Database) through the process of preprocessing, data mining and post processing to obtain knowledge. In this study, data mining applications to facilitate the search query based on the name of the village, using Java software. While the contours of a map is processed using ArcGIS software. The results of the research can give recommendations to stakeholders such as the Department of Tourism to describe the meaning of the classification of naming the village according to the character in each province at Java island.

This paper [7] discusses the application of Data Mining for prospective analysis and cluster analysis as a tool to aid in the design of Rural development Policies. The EU has promoted initiatives to develop its rural areas. In this research, we apply the last of those initiatives to Andalusia (southern Spain). These are the main aims pursued in this study: first, to present the methodology selected to identify the most influential variables that act as catalysts for development in each rural county of Andalusia; second, to identify rural county groups that show basic similarities in terms of the variables selected and third, to contribute to the integration of development strategies. To reach those objectives, it was applied the Prospective Structural Analysis as well as Cluster Analysis along with Discriminant Analysis to validate it. The results show the existence of four clearly differentiated clusters in the region.

This paper [11] discusses the application of Data Mining as Decision Support System for Village Economy Development Planning. A number of Decision Support Systems has been designed & developed successfully in many areas like medical, business agriculture, production, marketing and etc. Firstly, I will propose a general methodology for designing a DSS. Data mining techniques may be applied to large data warehouses (DWH) to draw useful information or results, which may be used to improve decision making process. Secondly, I have proposed a Decision support system for Village Economy Development Planning (VEDP-DSS) which may be used by District development planning officer (DDPO), Block development officer (BDO) & Village Sarpanch for decision making at appropriate level. The decision regarding village's developments may be further improved using VEDP-DSS. The performance of purposed system is compared with current systems to come out at a conclusion that, decision making can be improved & made more effective with the help of

computer based Decision support systems. Moreover, Decision-making may be made more effective, efficient & systematic by using Decision support system. In Future, VEDP-DSS may be further improved by considering other decision factors like quality of the facilities, demographic factors, relative importance of the facilities etc. Some other DSSs for e-governance like population control, pension planning, pay revision planning and etc. may also be designed and developed.

This paper [12] discusses the application of Data Mining for Classification of rural areas in Serbia. The goal of this paper is to present a method to establish the typology of rural areas in Serbia. Initially the OECD rurality criterion was applied to define the rural areas in Serbia. Subsequently, relevant indicators were selected (demographic, geographic, economic, employment-related, human capital, agricultural, tourism and infrastructure) and used to define and distinguish relatively homogeneous rural regions, based on correlation analysis, factor analysis (VARIMAX method) and cluster analysis. Cluster analysis revealed six regions of different sizes and characteristics. Practical considerations reduced this to four types, resulting in a robust scheme which accurately reflects the heterogeneous nature of rural Serbia.

This paper [13] discusses the application of Data Mining for Mapping the Indonesian territory, based on pollution, social demography and geographical data, using self organizing feature map. This research aims at mapping the 33 proviences in Indonesia based on pollution, social demography and geographical data.

This paper [14] discusses the application of Data Mining for A Conceptual Analysis on Development of Rural Villages with Scarcely Resources in Prakasam District, Andhra Pradesh State. Most of the people (nearly 70%) lived in rural villages. Generally we can't imagine the world without rural villages. So the study explains, development of rural villages with scarcely resources in prakasam dist Andhra Pradesh and explains how to utilization of scarcely resources effectively in rural villages. Rural Development means it is process in all round development of rural area based on scientific utilization of India's natural resources. With this aim, several efforts were made at rural development to better the socio-economic lot of villagers from ancient times to British period. The main objective of the study is to develop the rural villages with available limited resources in prakasam dist Andhra Pradesh state. The purpose of the study is to find out and evaluate the scarcely resources in each every backward rural villages in prakasam dist, AP

and work towards to effective utilization of these resources for better development of rural areas. It might be caused for economic growth and economic development of rural villages in the country. Further the author says after completion of the study we known how to develop rural villages with scarcely resources in prakasam dist, Andhra Pradesh and we get fully awareness and knowledge about scarcely resources in rural areas.

This paper [15] discusses the study of Data Mining for development of village as a smart village. This paper deals with study and development of village as a smart village. Here smart village is defined as bundle of services of which are delivered to its residence and businesses in an effective and efficient manner. " Smart Village" is that modern energy access acts as a catalyst for development in education , health, security, productive enterprise, environment that in turns support further improvement in energy access. In this research paper the author focuses on improved resource use efficiency, local self-governance, access to assure basic amenities and responsible individual and community behavior to build happy society. Here the author says that making smart village by taking smart decisions using smart technologies and services.

This paper [19] discusses the application of Data Mining for Analysing Soil Data using Classification Techniques. Soil is an essential key factor of agriculture. The objective of the work is to predict soil type using data mining classification techniques. Methods/Analysis: Soil type is predicted using data mining classification techniques such as JRip, J48 and Naive Bayes. These classifier algorithms are applied to extract the knowledge from soil data and two types of soil are considered such as Red and Black. Findings: In this paper, Data Mining and agricultural Data Mining are summarized. The JRip model can produce more reliable results of this data and the Kappa Statistics in the forecast were increased. Application/Improvement: For solving the issues in Big Data, efficient methods can be created that utilize Data Mining to enhance the exactness of classification of huge soil data sets.

This paper [22] discusses the Application of Data-mining Technique and Intelligent System on Demography Analysis in Nigeria. Controversy over Nigeria's census figures is nothing new. Accusations that the country's official population figures had been rigged date back to the 1950s and have continued unabated under military and civilian regimes. Furthermore, the demographics of Nigeria have undergone several changes over the past few decades as a result of migration and settlement from

the far off countries. At the present, demographic figures in Nigeria mires in controversy. This paper is an attempt to straighten Nigerian demographic analysis through data mining techniques and intelligent system to properly identify patterns and trends in Nigerian demographic figures. The study will first focus on Uli community and subsequently applied for Nigeria at large. Here in this research paper the author used Decision-Tree-Based classification model to extract from rich demographic data hidden information that can be used for the investigation of national conditions and national power. The author also utilized the power of intelligent system for complex processing and data analysis. Here the author proposed an intelligent system developed on Microsoft .NET platform. Since the intelligent system is an object oriented system, author used the Object Oriented Analysis and Design Methodology (OOADM).

This paper [26] discusses the study of Data Mining for Clustering (Information from Rural Villages of Sivagangai District). Research work is aimed to mining the rural villages of sivagangai district. Key factors to incorporated for mining information useful to village peoples and government are number of villages, number of families, number of schools based on government and private, number of colleges based on government and private, number of universities, educated level of study up to elementary schools, Secondary Schools, Higher Secondary Schools, Under graduation, Post graduation, Research, Drought hit area in villages, Availability of waste land, frequent causes of diseases, etc… From this data, useful information is proposed to generate the benefit to the peoples, as well as provided useful reporting to the government for sanction various beneficial schemes for rural village peoples in sivagangai district.

This paper [27] discusses the study of Data Mining for analysis of Female Work Participation in Villages of India. This paper attempts to identify patterns of female work participation in more than 578 thousand villages of India using the data mining approach. The analysis is based on an index of participation that has been developed for the purpose and takes into the consideration both the extent and the intensity of participation in productive activities.

The analysis reveals that Indian villages can be grouped into 10 clusters with different level of female participation and with distinct village characteristics and there are distinct regional patterns. An interesting finding of the analysis is that participation of females in productive activities at the village level is relatively lower in villages higher level of female education as compared to villages with lower levels of female education. It appears that appropriate opportunities of participation for educate females are not available in the villages of India. Creating these opportunities at the village level is necessary not only for the transformation of village economy but also for women's empowerment.

This paper [33] discusses the study of Data Mining for Performance Analysis of MGNREG Scheme using Classification. The Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS) has been implemented in Andhra Pradesh since 2006. The scheme focuses on works relating to water conservation and water harvesting, flood control, rural connectivity, repair of water bodies, drought proofing, irrigation canals and land development. There have been a number of studies that look at the implementation of the scheme in terms of employment created as well as issues of wages, processes of implementation, etc. This paper gives the analysis of the performance of MGNREG scheme in villages of Visakhapatnam district, using distance weighted k-nearest neighbor classification technique. The paper also gives the comparison of previous year statistical data provided by the government.

This paper [36] discusses the study of Data Mining for Classification of soil data Using the Compact Rule Generation. Various attributes within a dataset relate to each other and with the class attribute. The relationship between the different attributes with class attribute may improve the classification accuracy. The paper introduces CCSA algorithm that performs the clustering that is cascaded by classification based on association. The Clustering process generates a group of various instances within the dataset. These clustered instances are classified by using the association. This paper uses the Apriori association to generate the rules for classification. The technique is analyzed by using the soil data set and various other online available datasets using WEKA. The simulation result using the WEKA shows that reduced rules with the improved classification accuracy as compared to the existing association with classification algorithms.

3. DATA MINING TASKS

Data mining is the process of finding correlations or patterns among dozens of fields in large database. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Technically, Data mining as a term used for the specific classes of six activities or tasks as follows:

a) Classification

b) Estimation
c) Prediction
d) Association rules
e) Clustering

a) Classification
Classification is a set of techniques which are aimed at recognizing categories with new data points. In contrast to cluster analysis, a classification technique uses training data sets to discover predictive relationships.

b) Estimation
Estimation deals with continuously valued outcomes. Given some input data, we use estimation to come up with a value for some unknown continuous variables such as income, height or credit card balance.

c) Prediction
It's a statement about the way things will happen in the future, often but not always based on experience or knowledge. Prediction may be a statement in which some outcome is expected.

d) Association Rules
Association rule learning is set of techniques designed to detect valuable relationships or association rules among variables in databases.

e) Clustering
Cluster analysis is based on principles of similarities to classify objects. This technique belongs to unsupervised learning where training data is used [37].

4. STUDY AREA
The present study has been carried out in the Kaithal district. It is one of the 26 districts of Haryana, state in northern India. Kaithal, the north eastern district of Haryana State with a total geographical area of 2317 sq. km (approx 228000 hac). Kaithal was previously a part of Karnal District and later, Kurukshetra District. Kaithal came in to existence as a district of Haryana in 1 November 1989. The boundaries of Kaithal district are touching the three district of Haryana namely Karnal, Jind and Kurukshetra and the Patiala district of Punjab. The district is under control of Ambala division. The Kaithal city, occupies an area of 43.76 sq. km within the municipal council.

It is having a total population of 1072861 as per 2011 census. Out of this population 78 percent people live in the rural areas. Sex ratio in the district is 880 females per 1000 males as compared to 877 in the state and the population density is 463 per sq.km as compared to 573 of the state. The literacy rate is 76.4 percent as compared to 76.6 of the state. The district is having six blocks and 263 villages. There are seven blocks in Kaithal district namely Kaithal , Kalayat, Rajound, Pundari, Dhand, Guhla and Siwan that has been selected for the present study. Ten percent of villages from each block have randomly been selected for the study. In

this district rural poor are primarily reliant on agriculture and animal husbandry. 4% of households are selected from each randomly selected village. The distribution of the sample, List of Blocks, List of villages in each block is given in appendix.

Thus the ultimate dataset consists of 4521 samples from 28randomly selected villages. The data has been collected from primary as well as secondary sources. Secondary data has been taken up because of the easy availability. Primary data has been collected from the borrowers in the field by personnel interview method through an interview schedule designed for the purpose. To collect the primary data the we have visited the villages of the study area and has interacted with the peoples extensively by putting the questions in the local language. Secondary data has been collected from various sources including Department of Economic and Statistical Analysis, Haryana and District Rural Development Agency.

5. DATA MINING FOR STUDY AREA
The aim of this study is to develop a system for data mining application in rural development. The system perform clustering of data points given as input to this system. The system takes inputs, data points to be clustered and number of clusters(k) to be made of these input data points. The data points are prepared into data preparation phase of data mining process before using this system. The raw data is pre-processed, normalized and then data points are clustered using this system (k-means technique). The input data points into the system are in the form of instances $\{x_i, y_i\}$, where x_i, y_i are the set of attributes of the instances. The selection of the attributes is done carefully. As an example education, standard of living, agriculture and livestock, and economic status are used as input attributes to make the data point which are given to k-mean clustering algorithm as input. Clustering of these attributes is successfully performed. In this way this paper shows the applicability of data mining tools in rural development.

6. STEPS OF DATA MINING ACTIVITIES
The steps of data mining provide an overview of the life cycle of a data mining project. It contains the corresponding phases of a project, their respective tasks, and relationships between these tasks. Data mining steps need to take place in a certain order. Figure 4.1 shows set of activities that can take place in order. The following list describes data mining activities concisely.

1) Understanding Problem Domain: This initial phase focuses on understanding the project objectives and requirements. In this step the domain expert defines the data mining problem, learn about current solution to the problem and identify the

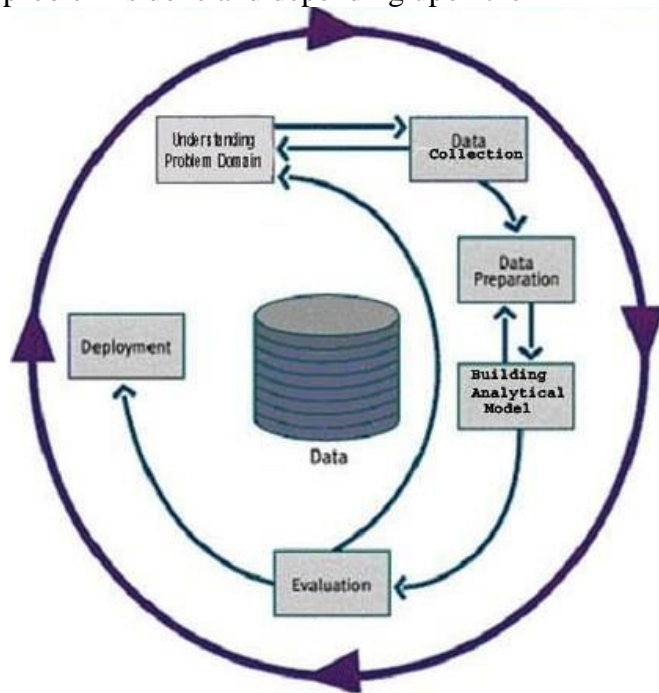requirements of key peoples. A description of the problem is done and depending upon the

Figure 4.1: Phases data mining Process Model [38]

problem selection of the data mining tools may also be done. A preliminary plan is designed to achieve the project objectives. Cost and time investments are needed to achieve each of these goals. This advance planning can save time and money spent in data mining efforts. Here in this paper we have chosen rural development as our problem domain. Our goal is to apply the data mining tools to analyze rural data.

2) Data collection: Just like before actual production company collect raw material similarly for data mining and analysis raw data should be collected for processing. All data mining process revolve around data. Data collection is one of the most time consuming tasks in data mining. In order to do this effectively, all relevant data for analysis should be gathered. This data may come from multiple heterogeneous sources such as Ministry of rural development, rural cooperative companies, stored data in government organizational operating databases, rural services companies, villager's personal information, rural architecture; rural guidance plans, District Rural development office (DRDO), Block Development office (BDO), Gram Panchayat Office, other data warehouses such as GISTNIC, NIC, web resources, other data marts that contain specific information and any other relevant data etc. Primary data can be collected from rural areas. When the data sources are identified, all relevant data elements are extracted.

3) Data preparation: Before creating an analytical model, data should be prepared. The data preparation phase covers all activities (tasks) to construct the final dataset which will be fed into the modelling tools from the initial raw data in other words data preparation is to define and process the mining data to make it fit specific data mining method. Data preparation phase includes six subtasks, data cleaning, data option, data pre-processing, data Integration, data transformation and data representation. A part of data preparation is to classifying variables. Variables may be discrete or continuous, qualitative or quantitative. Uncertain variables must be eliminated or replaced with probable values. Knowing the minimum, maximum, mean, median, and mode of quantitative variables provides adequate insight for data. To simplify the preparation process, data reduction transforms should be applied. The aim of data reduction is to combine a few variables into one variable for controllability analysis. For example, education levels, income, marital status and postcodes are combined.

4) Building Analytical model: One of the most prominent activities in data mining is to create an analytical model. An analytical data model shows an integrated, comprehensive, and time-dependent data structure consisting of internal and external data sources that are pre-processed. Upon implementation, this model should be able to continue the "learning" process such that data mining tools are utilized regularly and assessed by data mining experts. In other words, results of each analysis should be considered as new inputs and used in later analysis.

5) Data mining results interpretation/Evaluation: When data mining is implemented and results are obtained, the main responsibility is interpreting these results. Two points must be considered at this stage: how easily can the results be influenced and can these results be persuasive enough to present to the Ministry of rural development.

6) Results accreditation/Deployment: Results should be compared with other published statistics of rural area. Deviations from these statistics should be determined along with the reasons for such deviations. Up to date statistics of rural area should be used since these statistics vary over time. Data collection criteria should be compared with data collection criteria relevant to these statistics. Time periods for resulting data and rural statistics time periods should be compared. The created model data selection criteria with relevant time period should resemble rural statistics data selection criteria and time periods.

7) Monitoring the analytical data model over time: Rural statistics are usually obtained using very larger samples. Rural statistics change over time. Therefore it is important to accredit the analytical data model with these statistics at regular time intervals.

7. CONCLUSIONS

This paper has reviewed data mining applications in rural development. Readers should be cautious in interpreting the results of the survey, since the findings are based on data collected from the business web sites, journal articles, news web sites and working papers. As previously stated, strengthening data collection and analysis is a fundamental necessity for rural areas policymakers such that efficient utilization of the resulting opportunities will entail society's economic dynamism while creating a knowledge-based economy. In other words, the data mining process of identifying appropriate data, strengthening data collection, building a data warehouse, data analysis and subsequent decision making will have tremendous impact in the progression of societies especially in third world countries. Data collection and adequate data analysis has many applications in various areas including market management, agricultural products, immigration management, rural products distribution methods, fraud Such approach is employed because data mining applications in rural development are still rarely described in journal articles. However, I feel that even such a survey can describe the current state in data mining applications in rural development. The most often used methods are classification and prediction, concept/class description and evolution analysis. It can be concluded that data mining methods and other related techniques of knowledge discovery in databases and intelligent data analysis are indispensable in rural development.

8. REFERENCES

[1] Abolfazl Shahbazi, Maryam Karambeygi, Application of data mining in Rural Planning, ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 5, Issue 1, July 2015

[2] Kush R. Varshney,1,2,* George H. Chen,3 Brian Abelson,1,4 Kendall Nowocin,3 Vivek Sakhrani,5
Ling Xu,6 and Brian L. Spatocco7 ,Targeting Villages for Rural Development Using Satellite Image Analysis, Big Data Volume 3 Number 1, 2015 Mary Ann Liebert, Inc. DOI: 10.1089/big.2014.0061.

[6] Atje Setiawan Abdullah1, Budi Nurani Ruchjana2, Akik Hidayat1, Akmal1, and Deni Setiana, Data mining for clustering naming of the village at Java Island, The Asian Mathematical Conference 2016 (AMC 2016) IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series **893** (2017) 012059 doi :10.1088/1742-6596/893/1/012059.

[7] Pedro-Pablo Perez-Hernández2, José-Manuel Martin-Lozan, Mercedes Torres-Jiménez4, Application of Prospective Analysis and Cluster Analysis as a Tool to Aid in the Design of Rural Development Policies: the Case of Andalusia1, Recibido: 2013-10 -02 Aprobado: 2013-12-15 Disponible en línea: 2014-05-30 doi:10.11144/Javeriana.CDR11-73.apac.

[11] Arvind Selwal , A Decision Support System for Village Economy Development Planning.

[12] *Natalija Bogdanov, David Meredith, Sophia Efstratoglou,* A TYPOLOGY OF RURAL AREAS IN SERBIA, DOI:10.2298/EKA08177007B

[13] Kuswari Hernawati, Nur Insani, Bambang S. H. M., Nur Hadi W., and Sahid, Mapping the Indonesian territory, based on pollution, social demography and geographical data, using self organizing feature map
Citation: AIP Conference Proceedings 1868, 040002 (2017); doi: 10.1063/1.4995117
View online: https://doi.org/10.1063/1.4995117

[14] Khadar Basha Shaik , A Conceptual Analysis on Development of Rural Villages with Scarcely Resources in Prakasam District, Andhra Pradesh State, International Journal of Economics & Management Sciences, Shaik, Int J Econ Manag Sci 2017, 6:4 DOI: 10.4172/2162-6359.1000439

[15] Rutuja Somwanshi, Utkarsha Shindepatil, Deepali Tule, Archana Mankar, Namdev Ingle Guided By- Dr. V. S. Rajamanya, Prof. A. Deshmukh Study and development of village as a smart village, International Journal of Scientific & Engineering Research, Volume 7, Issue 6, June-2016 395 ISSN 2229-5518 IJSER © 2016 http://www.ijser.org.

[16] Kanu Raheja, Rural Development in Haryana, International Journal of Scientific and Research Publications, Volume 5, Issue 6, June 2015 1 ISSN 2250-3153.

[19] V. Rajeswari[*] and K. Arunesh , Analysing Soil Data using Data Mining Classification Techniques , Indian Journal of Science and Technology, *Vol 9(19), DOI: 10.17485/ijst/2016/v9i19/93873, May 2016*

[22] Okeke Ogochukwu C. ,Ezenwegbu Nnamdi C. , Application of Data-mining Technique and Intelligent System on Demography Analysis in Nigeria, The International Journal Of Engineering And Science (IJES) Volume 4 Issue 8 Pages PP -05-11 2015| ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805 www.theijes.com The IJES Page 5

[26] Dr.S.S.Dhenakaran, M.Sathish Kumar, Datamining: Clustering (Information from Rural Villages of Sivagangai District) , *International Journal of P2P Network Trends and Technology (IJPTT) - Volume 3 Issue 4 July to August 2013* ISSN: 2249-2615 http://www.ijpttjournal.org Page 1

[27] Aalok Ranjan Chaurasia MLC Foundation and

'Shyam' Institute, Female Work Participation in Villages of India, 20 January 2018, Online at https://mpra.ub.uni-muenchen.de/84345/ MPRA Paper No. 84345, posted 5 February 2018 14:45 UTC.

[33] P. Sumithra, V. Valli Kumari, Performance Analysis of MGNREG Scheme using Classification, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611.

[36] Navneet and Nasib Singh Gill, Classification Using the Compact Rule Generation, ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY, ISSN: 0974-6471 April 2015,Vol. 8, No. (1):Pgs. 49-58

[37] Supriya Nagarkar, Data mining Applications for Smart city: A Review, I J C T A, 10(8), 2017, pp. 705-710 ISSN: 0974-5572 © International Science Press.

[38].https://www.kdnuggets.com/wpcontent/uploads/crisp-dm-4-problems-fig1.png