SI-E

ISSN - 2229-3620
APPROVED UGC CARE

SHODH SANCHAR BULLETIN
Vol. 10, Issue 39, July-September, 2020
Page Nos. 74-81

AN INTERNATIONAL BILINGUAL PEER REVIEWED REFEREED RESEARCH JOURNAL

# AUGMENTING E-GOVERNANCE TECHNOLOGY IN INDIA USING DATA MINING TOOLS AND TECHNIQUES

❒ **Dr. C. Sunitha***
**Dhanya. V. A****

## ABSTRACT

*With the arrival of digital India association, our country is operational through study data communication. In these days state of affairs, India government provides e-services to any or all peoples at their door steps through e-governance. Sensible organization of India government services are often increased through the utilization of ICT. The entrustment of e-governance comes is notified from central to varied state level so gram panchayat. Fund distributed and therefore the fund maintained totally different e-governance comes and gifts in digital format at each level. Beyond deploying the comes, there's a necessity to look at the record hold on a spread of levels. The effectiveness of state comes are often examined by applying numerous data processing tools and techniques. Data processing techniques are won't to associate degree analyse the information noninheritable from NeGP (National e-Governance plan) is an initiative of Government of India for making all governmental services are available to citizens of India through electronic media of e-health and SBM(Swachh Bharat Mission) projects. The projected approach of information mining is often useful for the government to more analyse alternative public beneficiary schemes.*

***Keywords:*** *ICT- Information and Communication Technology, SVM from databases, E-Health Plan(MRMBS),SBM-Swachh Bharat Mission*

## I. INTRODUCTION

"A Dataset is alleged to be open if anyone is free use, reuse and spread it – Open information shall be code and it ought to even be simply accessible" National information sharing and accessible policy(NDSAP) came within the gazette of India on March 17th ,2012 that permits the researchers to access the information from numerous section enforced policies. This information comes within the class of open information and is accessible for more analysis.Many states of India facilitate and encourage public awareness and participation by making information widely available. This information can be used to help the government in decision making process. Data mining is one of the techniques that can be explored to get detailed insight into the data. A. Information Mining It is method of exploring, identify or observe the patterns hidden in the large data sets ICT based methods. These techniques can be applied on any form of huge quantities of information for

*HOD, Department of MSc.SS, Sri Krishna college of Arts and Science, Kuniyamuthur, Coimbatore
**Department of MSc.SS, Sri Krishna college of Arts and Science, Kuniyamuthur, Coimbatore

information processing. B. Knowledge Discovery Process (KDD) Knowledge Discovery (KDD) is that the way of finding useful data and patterns inside the data. They work with a numerous of databases and use various data mining methodologies like fuzzy logic, neural networks etc. to retrieve useful hidden information. They have been used in various applications in fields like banking, automobile, agriculture, medicine etc. The main goal of KDD is to seek out any high-level data from the low-level data. It uses training sets to seek out the patterns, knowledge or helpful hidden information hidden from the databases. It focuses on the overall process of knowledge discovery from storing data to accessing data, finding necessary algorithms to run capably and produce results, and how to provide the results to the end-user in a comprehensible way.

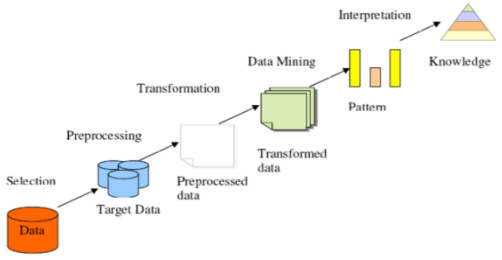



Fig.1 KDD Life Cycle Process

## II. DRAWBACK DEFINITION

E-Governance projects are attractive a boon for the states similarly as for union government of India for delivering a variety of schemes of governments in the welfare of common people of our country with the help of ICT. There are many E-Governance projects running at various levels in our country to make delivery of services fast and reliable to the public. If I analyse the outcomes with the use of data mining techniques, can be evaluate the results that are beneficial to the government but there are several issues in implementing this strategy. There is lack of internet connectivity to rural areas or local government level Unawareness about new technologies in the government authorities is also a problem. Many social, economic and technical issues also exist The outcomes after implementations of various government schemes aren't accurately analysed The managerial is done purely on the idea of results bestowed by government departments.

### Challenges of E-Governance Projects

- High digital illiteracy level within the rural areas.
- Non accessibility of user friendly interfaces
- In adequate power provide in rural areas.
- Low broadband signal within the rural areas
- Lack of awareness of e-governance comes.

### 2.1 E-HEALTH

Hospital Management System(HMS) is useful for patients, doctors and administrators.

**For Patients:**

1. Patients Identification Number(PIN)-distinctive recognize.
2. On next visit, needn't wait in long queue for registering again
3. Prescription & lab reports printouts are given to the patient.
4. Previous records available on-line safe for a extended amount of your time.
5. Patient will visit any secondary care hospital across with PIN.

**For Doctors:**

1. Saves a great deal of your time.
2. Drugs/Lab investigations are often classified into packages for prescription.

3. Will read the previous clinical history.
4. Specialty op-doctors with a single click can repeat the previous description.
5. In certain cases the doctor can follow the standard treatment guidelines (Master Data).

**For Administrators:**

1. This predominately functions as decision support system.
2. Hospital-level for chief medical officers.
3. District level for joint administrators / Dean.
4. Employee level for HOD'/administrators & Policy Manufacturers.

**CASH ASSISTANCE TO PREGNANT AND DELIVERED MOTHERS(MRMBS) SCHEME.**

It is another category of e-health program within the government of India. To implement a mega welfare scheme through the online, with over , with extra , with too much , with in excess of , with more than Thirty five lakhs bank accounts opened every year within the names of pregnant / Delivered mothers.

1. Dr. Muthulakshmi Reddy Maternity Benefit Scheme (MRMBS) is implemented with an objective of providing assistance to poor pregnant women to ensure the access to nutritional food and compensating the wage losses during pregnancy.
2. Cash Assistance of Rs.18,000 under this Scheme is disbursed in three instalments(RS:6000) to poor pregnant women directly by ECS from government treasury bank to beneficiary bank account. Eligibility Criteria
3. First Instalment- Pregnant lady should have the mandatory antenatal care check-ups including TT Immunization up to 7th month of pregnancy in government institution.
4. Second Instalment- Immediately after delivery in a government health facility.
5. Third Instalment- After third dose Penta valent immunization of the child during fourth month. Specific data to be entered for eligibility of the mother under MRMBS.

Minimum of 3 Antenatal check-up details are to be entered on or before 7th month of pregnancy to become eligible to receive 1st instalment. Details of delivery are to be entered to become eligible to receive 2nd instalment. The details of Immunization for the child are to be entered up to III rd dose of pentavalent (DPT+ Hepatitis B+ Hib)

F. SBM Swachh Bharat Mission (Garmin) Scheme is providing public funds to the eligible beneficiaries for construction of integrated house hold toilets/Latrines (IHHL) in the rural areas of various gram panchayats.

**2.2 SUPPPORT VECTOR MACHINE CLASSIFIER**

SVM is a powerful supervised classifier and correct learning approach. From the statistical idea, it was derived and advanced by using Vapnick in 1982. It yields a success classification results in various utility domains, as an example, medical diagnosis, text categorization, face popularity and Bioinformatics. SVM is based totally on the structural chance minimization precept from his statistical s.22tudying principle [Cris2000]. The kernel controls the empirical risk and type capacity with a purpose to maximize the margin among the training and decrease the proper charges. The unique premier hyper plane set of rules counselled with the aid of Vapnik in 1963 turned into a linear classifier for fixing linearly non-separable problems

Boser, Guyon and Vapnik in 1992 proposed an approach to create nonlinear classifiers with the aid of making use of kernel features, which lets in the algorithm to healthy the maximum margin hyper aircraft in a transformed characteristic space. Some common place kernels encompass linear, polynomial, Gaussian or Radial Basis Function (RBF) and sigmoid. In this studies work the RBF kernel is followed for several motives:

(i) The linear kernel is not able to handle nonlinearly separable type problems and it's far in truth a special shape of the RBF kernel;

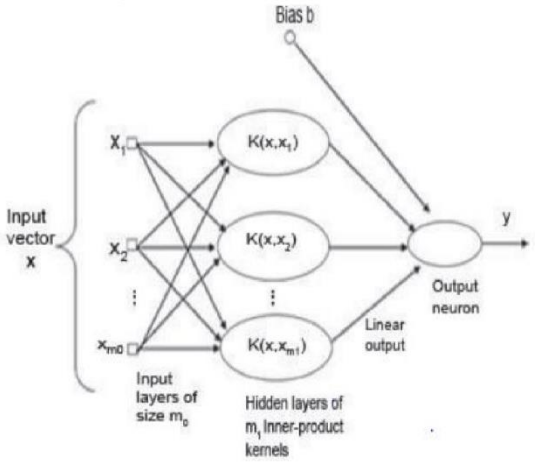
(ii) Computation of the RBF kernel is extra strong than that of the polynomial kernel, which might also go back values of 0 or infinity insure instances;

(iii) The sigmoid kernel is handiest legitimate for positive parameters;

(iv) The RBF kernel has fewer parameters to be calculated whilst in comparison to the polynomial and sigmoid kernels.

The remarkable benefit of the SVM technique over conventional strategies, except the properly-set up theoretical definition, is its capacity of operating with high dimensional characteristic vectors without dropping the generalization performance. It has another benefit compared to the famous method ANN. The variety of hidden units is automatically obtained for SVM, in assessment to ANN. Compared to ANN, it's miles faster, may be used with larger range of capabilities, less complicated to interpret and deterministic. SVM is capable of take care of the simple linear category project, as well as extra complex, i.e. nonlinear category problems. One of the principle advantages of SVM is that they are extended Fig.2.2.1 Architecture of SVM classifier insignificant classifiers. This asset makes SVM very appropriate for category of MR and CT pictures. In lots of these areas SVM has out finished properly mounted techniques along with ANN, RBF and nonparametric cluster category. SVM has been very famous and shown to be green in pc primarily based brain illnesses diagnostic studies within the literature [Sand2006, Zhon2010, Padm2011, Ahme2010, and Ei2010].

Fig.2.2.1 Architecture of SVM classifier




## 2.3 CLASSIFICATION USING SUPPORT VECTOR MACHINES

SVM classifier for real time records is constructed the use of actual time SNMP-MIB records which has 7 attributes. The first stage SVM classifier is built for the actual time data which separates the Normal and Attack training. The Model document is created in schooling the usage of the default SVM parameters. Three styles of kernel capabilities which include Radial Basis Functions, Linear and Exponential are evaluated. The attack classes of first level SVM are given to 2d level SVM. The 2d level SVM classifier is constructed which classifies the assault into 8 training of Network, Transport and Application Layer attacks and an everyday elegance. A hyper plane is a line that splits the enter variable space. In SVM, a hyper plane is

chosen to first-rate separate the points inside the input variable area with the aid of their class, both class 0 or magnificence 1. In -dimensions, you could visualize this as a line and let's expect that each one of our enter factors may be completely separated by this line. For example: B0 + (B1 * X1) + (B2 * X2) = 0 Where the coefficients (B1 and B2) that determine the slope of the line and the intercept (B0) are located through the learning set of rules, and X1 and X2 are the 2 input variables. User preserve create classifications the use of this line. By plugging in enter values into the road equation; you may calculate whether a brand new point is higher than or beneath the line. 1. Significance near the line precedes a cost close to zero and the factor may be hard to categorize. 2. If the significance of the fee is big, the model may have more self-assurance within the prediction. The detachment among the procession and the nearest facts factors is known as the boundary. The first-class or most reliable line that can separate the 2 instructions is the line that as the biggest margin. This is called the Maximal-Margin hyper plane. The margin is calculated because the perpendicular distances from the line to most effective the nearest factors. Only those points are applicable in defining the road and within the creation of the classifier. These points are called the guide vectors. They support or outline the hyper plane. The hyper plane is discovered from training statistics with an optimization technique that maximize the margin.

**Soft Margin Classifier**

Exercise, authentic statistics is messy furthermore cannot be estranged flawlessly hyper plane. Their limitation extended the margin of the procession that separate the lessons needs to be comfortable. This trade permits a few point in the instruction data to contravene the unscrambling column. Supplementary positions of coefficients are brought with the intention of deliver the margin twist room in every measurement. These coefficients are now and again called loose variables. A modification limitation is delivered referred to as significance throughout all proportions. A C=zero isn't any violation and we are again to the rigid MaximalMargin Classifier defined exceeding For the duration of the getting to know of hyper plane commencing proceedings, each and every one teaching times with the purpose of recline within the distance of the boundary will concern the location to hyper plane known as help vectors, influences the variety of instance which can be permissible to reduce inside the scope of version.

i) Support Vector Machines -SVM

The SVM algorithm is implementing in exercise the use of operating system. The studying hyper plane in least support vector machines is executed via remodelling to solve. A commanding perception internal manufactured from any two specified interpretation, as opposed to the clarification them. For instance, the internal fabricated from the vectors [2, 3] and [5, 6] is 2*5 + three*6 or 28. The equation for creating a prediction for a new enter the use of the dot product between the input (x) and each help vector (xi) is calculated as follows: F (x) = B0 + sum (ai * (x, xi)) This is an equation that entail manipulative the internal goods of a brand innovative enter vector (x) with all aid vectors in preparation records. The coefficients B0 and ai (for each input) have to be predictable commencing the training information by using the mastering set of rules. Linear Kernel SVM The dot-product is known as the most important part and can be re-written as: K(x, xi) = sum(x * xi) Polynomial Kernel SVM

K(x, xi) = 1 + sum(x * xi)^d Wherever the diploma of the polynomial need to be precise with the aid of hand to the getting to know algorithm. When d=1 this is similar to the least kernel. The polynomial kernel lets in for rounded traces within the input space.

ii) Radial Kernel SVM

K(x, xi) = exp (-gamma * sum((x – xi^2))

Where gamma is a parameter that must be specific to the mastering set of rules. An exact default fee for gamma is 0.1, wherein gamma is regularly zero < gamma < 1. The radial kernel could be very local and can create complicated areas within the feature area, like closed polygons in two-dimensional area.

## 2.4 DATA PREPARATION FOR SVM

The SVM set of rules operates natively on numeric attributes. The algorithm mechanically "explodes" express statistics into a fixed of binary attributes, one per category cost. For example, a man or woman editorial for marital attractiveness with values marital or single would be transformed to two numeric attributes: married and single. The new attributes ought to have the value 1 (authentic) or 0 (false). When there are lacking values in columns with easy statistics sorts (not nested), SVM interprets them as lacking at random. The set of rules automatically replaces lacking express values with the mode and missing numerical values with the mean. At what time in attendance are missing principles in nested columns, SVM interprets them as sparse. The set of rules automatically replaces sparse numerical records with zeros and sparse categorical facts with 0 vectors. This phase lists some tips for a way to excellent prepare your schooling records whilst getting to know an SVM model.

Normalization SVM requires the normalization of numeric input. Normalization places the values of numeric attributes on the equal scale and stops attributes with a huge authentic scale from biasing the answer. Normalization in addition minimizes the probability of overflows and underflows. Furthermore, normalization brings the numerical attributes to the identical scale (0, 1) as the exploded categorical facts. SVM and Automatic Data Preparation The SVM algorithm regularly handle misplaced fee handling and the alteration of specific statistics; however normalization and outlier recognition should be dealt with by way of ADP or organized physically. SVM Classification

A choice aircraft is one which separates among a hard and fast of gadgets having special elegance memberships. SVM finds the vectors ("assist vectors") that outline the separators giving the widest separation of instructions. Class Weights

In SVM categorization, weights are a biasing system for specifying the experienced consequence of objective standards (lessons).Routinely initialized to attain the outstanding standard forecast throughout all training. Users amplify the burden for a class, the percentage for that magnificence must boom. If user calculation is one, the case is taken into consideration normal. If user prediction is zero, values are taken into consideration inconsistent.

You can specify the share of the facts which you anticipate to be anomalous with the SVMS_OUTLIER_RATE construct placing. The model will pick out approximately that many ÒrareÓ instances when implemented to the overall population. The default is ten percent which is probably excessive for plenty inconsistency recognition difficulty.

SVM Regression SVM makes use of Associate in nursing epsilon-in responsive thrashing purpose to clear up deterioration difficulty. SVM deterioration attempt to notice continuous characteristic such as the utmost number of facts factors lies in the epsilon-huge insensitivity tube.

## III. PROJECTED RESOLUTION

An interactive plat form needs to be developed by the organizations for performing analysis data. This paper suggests that for the proper decision making process on the data of various government schemes there must be a data analysis wing or department available at every region level.

In the departments experts will be recruited by the government. This department provides the results of various schemes and their draw backs why a particular scheme is not successful. It is also helpful in the anomaly detection in public fund utilization.

## IV. EXPERIMENTAL AND OUTCOMES

After the text edit has been completed, organized for the template. Duplicate template file by using the save as command, and use the naming convention prescribed by your conference for the name of your paper. In this recently created file, high light of all the contents and import your prepared text file. In data mining including the pre-processing of the data and the build up to the predictive model. It includes algorithms and tools for cluster analysis, classification, regression analysis, visualization, and feature selection or selecting attributes/variables. It contains a group of visualization tools and algorithms for data analysis and predictive modelling, together with graphical computer program for easy access to these functions.

## V. CONCLUSION

This work will facilitate the government to frame better policies focusing on public benefit. Our approach is to apply all possible data mining techniques to find out patterns and hidden results. Our focuses find out the flaws in the E-governance system. The interface for performing the data mining is created. The process of data mining leads us to desired outcomes in the form of analysed data. The interface design is helpful to the government local bodies such as municipalities, Gram panchayats (GPs), and at the blocks as well as district level local governing bodies.

## VI. FUTUURE WORK

This work are useful to the government to investigate several alternate public beneficiary schemes of MRMBS and SBM etc... This may prevent corruption at the local level and will be beneficial for the economics and financial growth of the country. This can also be helpful in increasing development and more public, awareness about the frauds happening in our villages.

## REFERENCES


- E-Governance In India- Problems and acceptability: Sanjay Kumar Dwivedi & Ajay kumar bharti (2010), Journal of Theoretical and applied information technology Pp37-437-437437-437-43.
- Knowledge discovery and Data Mining: Towards a Unifying framework, From: KDD-96 proceedings.copyright@1996,AAAI.
- Joarder, M. M. M. (2003, December). Brain gain network: a proposal e-governance model to convert the brain drain into brain gain. In Multi Topic Conference, 2003. INMIC 2003. 7th International (pp. 424-427). IEEE.

- Murthy, D. N., & Kumar, R. V. P. (2003, October). Software architectural design model for e-governance systems. In TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region (Vol. 1, pp. 183-187). IEEE.
- Nkomo, P. T., Terzoli, A., Muyingi, H., & Rao, G. K. (2006, February). Smart card initiative for South African egovernance-a study. In 2006 8th International Conference Advanced Communication Technology (Vol. 3, pp. 2-pp). IEEE.
- Marjit, U., Roy, R., Santra, S., & Biswas, U. (2009, August). A semantic web service based approach to EGovernance. In Applications of Digital Information and Web Technologies, 2009. ICADIWT'09. Second International Conference on the (pp. 232-237). IEEE.
- Chakravarti, B., & Varma, V. (2008, November). An enterprise architecture framework for building service oriented e-governance portal. In TENCON 2008-2008 IEEE Region 10 Conference (pp. 1-6). IEEE.
- Islam, G. Z., Khan, M. I., &Mazady, M. A. B. (2008, December). Developing a Model of E-governance for Urban and Rural Areas of Bangladesh. In Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on (pp. 587-592). IEEE.
- Costake, N. (2001). E-Governance and the Judicial System. A point of view. In Database and Expert Systems Applications, 2001. Proceedings. 12th International Workshop on (pp. 423427). IEEE.
- Shackleton, P., Fisher, J., & Dawson, L. (2004, January). Evolution of local government e-services: the applicability of ebusiness maturity models. In System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on (pp. 9-pp). IEEE.
- Hashemi, S. M., Razzazi, M., &Teshnehlab, M. (2008, April). Streamlining the global village grid through unifying EGovernments, E-Businesses, and E-Commerce services. In Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on (pp. 1-4). IEEE.
- Gopakumar, K. (2006, May). E-governance services through Telecentres-Role of Human Intermediary and issues of Trust. In 2006 International Conference on Information and Communication Technologies and Development (pp. 131-142). IEEE.
- Allan, J. D. O., Rambajun, N., Sood, S. P., Mbarika, V., Agrawal, R., &Saquib, Z. (2006, November). The e Government Concept: A Systematic Review of Research and Practitioner Literature. In 2006 Innovations in Information Technology (pp. 1-5). IEEE.
- Joarder, M. M. M. (2003, December). Brain gain network: a proposal e-governance model to convert the brain drain into brain gain. In Multi Topic Conference, 2003. INMIC 2003. 7th International (pp. 424-427). IEEE. [15] Murthy, D. N., & Kumar, R. V. P. (2003, October). Software architectural design model for e-governance systems. In TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region (Vol. 1, pp. 183-187). IEEE