IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Excavating Educational Statistics to Investigate Scholars Performance

V. Maniraj

Department of Computer Science, A.V.V.M Sri Pushpam College, Poondi, Thanjavur, India

Available online at: www.ijcseonline.org

Received: 14/Apr/2018, Revised: 20/Apr/2018, Accepted: 25/Apr/2018, Published: 30/Apr/2018

Abstract**—** The fundamental target of advanced education foundations is to give quality instruction to its understudies. One approach to accomplish most elevated amount of value in advanced education framework is by finding learning for expectation with respect to enrolment of understudies in a specific course, distance of conventional classroom showing model, discovery of out of line implies utilized as a part of online examination, identification of irregular qualities in the outcome sheets of the understudies, forecast about understudies' execution et cetera. The information is covered up among the instructive data set and it is extractable through data mining procedures. Show paper is intended to legitimize the capacities of data mining strategies in setting of advanced education by offering a data mining model for advanced education framework in the college. In this examination, the order undertaking is utilized to assess understudy's execution and as there are numerous methodologies that are utilized for data characterization, the choice tree strategy is utilized here. By this assignment we extricate learning that portrays understudies' execution in end semester examination. It helps prior in recognizing the dropouts and understudies who require exceptional consideration and enable the educator to give proper prompting/guiding.

Keywords**—** Educational Data Mining (EDM); Classification; Knowledge Discovery in Database (KDD); ID3 Algorithm.

I. INTRODUCTION

The coming of data innovation in different fields has lead the vast volumes of data stockpiling in different configurations like records, documents, reports, pictures, sound, recordings, logical data and numerous new data groups. The data gathered from various applications require appropriate strategy for extricating information from extensive vaults for better basic leadership. Learning revelation in databases (KDD), frequently called data mining, goes for the disclosure of valuable data from extensive accumulations of data. The primary elements of data mining are applying different strategies and calculations keeping in mind the end goal to find and concentrate examples of put away data. Data mining and information revelation applications have a rich concentration because of its noteworthiness in basic leadership and it has turned into a basic part in different associations. Data mining procedures have been brought into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capacities and so forth.

There are expanding research interests in utilizing data mining in training. This new rising field, called Educational Data Mining, worries with creating strategies that find learning from data starting from instructive conditions. Instructive Data Mining utilizes numerous strategies, for example, Decision Trees, Neural Networks, Naïve Bayes, K-Nearest neighbor, and numerous others.

Utilizing these systems numerous sorts of information can be found, for example, association rules, orders and clustering. The found information can be utilized for forecast with respect to enrolment of understudies in a specific course, distance of customary classroom showing model, discovery of out of line implies utilized as a part of online examination, location of strange qualities in the outcome sheets of the understudies, expectation about students‟ execution et cetera.

The primary target of this paper is to utilize data mining techniques to ponder students‟ execution in the courses. Data mining gives numerous errands that could be utilized to examine the understudy execution. In this examination, the arrangement errand is utilized to assess student's execution and as there are numerous methodologies that are utilized for data characterization, the choice tree strategy is utilized here. Information's like Attendance, Class test, Seminar and Assignment marks were gathered from the student's administration framework, to anticipate the execution toward the finish of the semester. This paper examines the exactness of Decision tree strategies for anticipating understudy execution.

II. DATA MINING DEFINITION AND TECHNIQUES

Data mining, additionally prominently known as Knowledge Discovery in Database, alludes to extricating or "mining" learning from a lot of data. Data mining strategies are utilized to work on vast volumes of data to find concealed examples and connections supportive in basic leadership. While data mining and learning disclosure in database are as often as possible regarded as equivalent words, data mining is entirely of the information revelation process. The groupings of steps distinguished in removing learning from data are appeared in Figure 1.

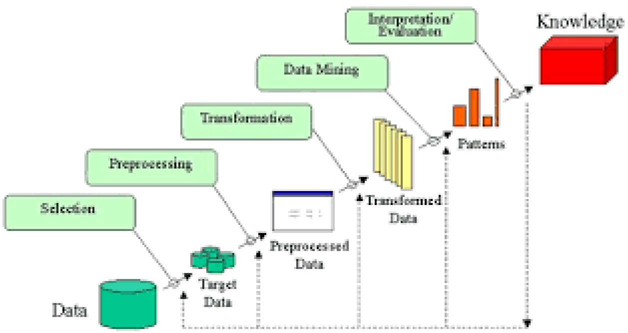

Figure 1: The steps of extracting knowledge from data

Different calculations and procedures like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor strategy and so on, are utilized for information revelation from databases. These procedures and techniques in data mining need brief say to have better understanding.

A. Grouping

Grouping is the most generally connected data mining method, which utilizes an arrangement of pre-characterized cases to build up a model that can group the number of inhabitants in records on the loose. This approach oftentimes utilizes choice tree or neural system based grouping calculations. The data characterization process includes learning and order. In Learning the preparation data are broke down by characterization calculation. In arrangement test data are utilized to gauge the exactness of the grouping rules. On the off chance that the exactness is satisfactory the rules can be connected to the new data tuples. The classifier-preparing calculation utilizes these pre-arranged cases to decide the arrangement of parameters required for legitimate segregation. The calculation at that point encodes these parameters into a model called a classifier.

B. Clustering

Clustering can be said as distinguishing proof of comparative classes of articles. By utilizing clustering methods we can additionally recognize thick and inadequate locales in question space and can find general appropriation example and connections among data characteristics. Order approach can likewise be utilized for powerful methods for recognizing gatherings or classes of question yet it turns out to be expensive so clustering can be utilized as preprocessing approach for trait subset choice and characterization.

C. Predication

Relapse method can be adjusted for predication. Relapse examination can be utilized to display the connection between at least one autonomous factors and ward factors. In data mining autonomous factors are properties definitely known and reaction factors are what we need to foresee. Shockingly, some genuine issues are not just forecast. Thusly, more mind boggling systems (e.g., strategic relapse, choice trees, or neural nets) might be important to conjecture future qualities. A similar model writes can frequently be utilized for both relapse and grouping. For instance, the CART (Classification and Regression Trees) choice tree calculation can be utilized to assemble both arrangement trees (to order clear cut reaction factors) and relapse trees (to figure persistent reaction factors). Neural systems also can make both arrangement and relapse models.

D. Association rule

Association and connection is more often than not to discover visit thing set discoveries among substantial data sets. This sort of discovering causes organizations to settle on specific choices, for example, index configuration, cross showcasing and client shopping conduct examination. Association Rule calculations should have the capacity to create rules with certainty esteems short of what one. However the quantity of conceivable Association Rules for a given dataset is for the most part expansive and a high extent of the rules are more often than not of little (assuming any) esteem.

E. Neural systems

Neural system is an arrangement of associated input/yield units and every association has a weight give it. Amid the learning stage, organize learns by altering weights to have the capacity to foresee the right class marks of the info tuples. Neural systems have the astounding capacity to get significance from muddled or uncertain data and can be utilized to extricate designs and identify patterns that are too mind boggling to ever be seen by either people or other PC procedures. These are appropriate for consistent esteemed data sources and yields. Neural systems are best at recognizing examples or patterns in data and appropriate for expectation or estimating needs.

F. Decision Trees

Choice tree will be tree-molded structures that speak to sets of choices. These choices create rules for the characterization of a dataset. Particular choice tree strategies incorporate Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

G. Nearest Neighbor Method

A method that arranges each record in a dataset in view of a mix of the classes of the k record(s) most like it in a chronicled dataset (where k is more prominent than or equivalent to 1). Once in a while called the k-closest neighbor procedure.

III. Data Mining Process

In display day's instructive framework, a student's execution is dictated by the inward appraisal and end semester examination. The inside evaluation is done by the instructor in view of students‟ execution in instructive exercises, for example, class test, workshop, assignments, general capability, participation and lab work. The end semester examination is one that is scored by the understudy in semester examination. Every understudy needs to get least stamps to pass a semester in inner and in addition end semester examination.

A. Data Preparations

The data set utilized as a part of this investigation was gotten from KNGAC Bharathidasan University, Thanjavur on the inspecting technique for PC Applications division obviously M.Sc (Master of Computer Science) from session 2007 to 2010. At first size of the data is 50. In this progression data put away in various tables was participated in a solitary table subsequent to joining process mistakes were expelled.

B. Data determination and change

In this progression just those fields were chosen which were required for data mining. A couple of determined factors were chosen. While a portion of the data for the factors was separated from the database. All the indicator and reaction factors which were gotten from the database are given in Table I for reference.

Table I. Understudy RELATED VARIABLES

Variable	Description	Possible Values
PSM	Previous Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36%}
CTG	Class Test Grade	{Poor , Average, Good}
SEM	Seminar Performance	{Poor , Average, Good}
ASS	Assignment	{Yes, No}
GP	General Proficiency	{Yes, No}
ATT	Attendance	{Poor , Average, Good}
LW	Lab Work	{Yes, No}
ESM	End Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36%}

The space esteems for a portion of the factors were characterized for the present examination as takes after:

- PSM – Previous Semester Marks/Grade acquired in MCA course. It is part into five class esteems: First – >60%, Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%.
- CTG – Class test grade got. Here in every semester two class tests are directed and normal of two class test are utilized to compute sessional marks. CTG is part into three classes: Poor – < 40%, Average – > 40% and < 60%, Good – >60%.
- SEM – Seminar Performance acquired. In every semester workshop are composed to check the execution of understudies. Workshop execution is assessed into three classes: Poor – Presentation and correspondence ability is low, Average – Either introduction is fine or Communication aptitude is fine, Good – Both introduction and Communication expertise is fine.
- ASS – Assignment execution. In every semester two assignments are given to understudies by every educator. Task execution is partitioned into two classes: Yes – understudy submitted task, No – Student not submitted task.
- GP - General Proficiency execution. Like workshop, in every semester general capability tests are sorted out. General Proficiency test is partitioned into two classes: Yes – understudy took an interest by and large capability, No – Student not partook when all is said in done capability.
- ATT – Attendance of Student. Least 70% participation is necessary to take part in End Semester Examination. Be that as it may, even through in extraordinary cases low participation

understudies additionally partake in End Semester Examination on certifiable reason. Participation is partitioned into three classes: Poor - <60%, Average - > 60% and <80%, Good - >80%.

- LW – Lab Work. Lab work is partitioned into two classes: Yes – understudy finished lab work, No – understudy not finished lab work.
- ESM - End semester Marks acquired in MCA semester and it is announced as reaction variable. It is part into five class esteems: First – >60% , Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%.

C. Decision Tree

A choice tree is a tree in which each branch hub speaks to a decision between various choices, and each leaf hub speaks to a choice.

Choice tree are regularly utilized for picking up data with the end goal of choice - making. Choice tree begins with a root hub on which it is for clients to take activities. From this hub, clients split every hub recursively as indicated by choice tree learning calculation. The last outcome is a choice tree in which each branch speaks to a conceivable situation of choice and its result.

The three generally utilized choice tree learning calculations are: ID3, ASSISTANT and C4.5.

D. The ID3 Decision Tree

ID3 is a basic choice tree learning calculation created by Ross Quinlan. The essential thought of ID3 calculation is to develop the choice tree by utilizing a best down, voracious hunt through the offered sets to test each trait at each tree hub. Keeping in mind the end goal to choose the quality that is most valuable for arranging a given sets, we present a metric - data pick up.

To locate an ideal method to order a learning set, what we have to do is to limit the inquiries asked (i.e. limiting the profundity of the tree). In this way, we require some capacity which can gauge which questions give the most adjusted part. The data increase metric is such a capacity.

E. Measuring Impurity

Given a data table that contains characteristics and class of the traits, we can quantify homogeneity (or heterogeneity) of the table in light of the classes. We say a table is unadulterated or homogenous on the off chance that it contains just a solitary class. On the off chance that a data table contains a few classes, at that point we say that the table is polluted or heterogeneous. There are a few files to gauge level of contamination quantitatively. Most notable files to gauge level of contamination are entropy, gini list, and order mistake.

$$Entropy = \sum_j -p_j log_2 p_j$$

Entropy of an unadulterated table (comprise of single class) is zero on the grounds that the likelihood is 1 and log (1) = 0. Entropy achieves greatest esteem when all classes in the table have break even with likelihood.

$$Gini\ Index = 1 - \sum_j p_j^2$$

Gini file of an unadulterated table comprise of single class is zero in light of the fact that the likelihood is 1 and 1-12 = 0. Like Entropy, Gini record additionally achieves greatest esteem when all classes in the table have level with likelihood.

$$Classification\ Error = 1 - \max\{p_j\}$$

Like Entropy and Gini Index, Classification blunder file of an unadulterated table (comprise of single class) is zero in light of the fact that the likelihood is 1 and 1-max (1) = 0. The estimation of characterization mistake record is dependably in the vicinity of 0 and 1. Actually the greatest Gini list for a given number of classes is constantly equivalent to the most extreme of characterization mistake file in light of the fact that for various classes n, we set likelihood is equivalent to $p = \frac{1}{n}$ furthermore, greatest Gini record occurs at $1 - n\frac{1}{n^2} = 1 - \frac{1}{n}$ while most extreme characterization blunder file likewise occurs at $1 - max\left\{\frac{1}{n}\right\} = 1 - \frac{1}{n}$.

F. Part Criteria

To decide the best quality for a specific hub in the tree we utilize the measure called Information Gain. The data pick up, Gain (S, An) of a property An, in respect to a gathering of illustrations S, is characterized as

$$Gain(S, A) = Entropy(S) - \sum_{V \epsilon Values(A)} \frac{|S_V|}{|S|} Entropy(S_V)$$

Where Values (An) is the arrangement of every single conceivable incentive for property An, and S_V is the subset of

S for which trait A has esteem v (i.e., $S_v = \{s \in S | A(s) = v\}$). The main term in the condition for Gain is only the entropy of the first gathering S and the second term is the normal estimation of the entropy after S is divided utilizing quality A. The normal entropy portrayed by this second term is essentially the aggregate of the entropies of each $|S_v|$that subset, weighted by the division of illustrations $\frac{|S_v|}{|S|}$have a place with Gain (S, A) is along these lines the normal diminishment in entropy caused by knowing the estimation of quality A.

$$Split\ information\ (S, A) = -\sum_{1=1}^{n} \frac{|S_i|}{|S|} log_2 \frac{S_i}{S}$$

also,

$$Gain\ Ratio\ (S, A) = \frac{Gain(S, A)}{Split\ Information\ (S, A)}$$

The way toward choosing another quality and dividing the preparation cases is presently rehashed for each non terminal relative hub. Qualities that have been consolidated higher in the tree are prohibited, with the goal that any given trait can show up at most once along any way through the tree. This procedure proceeds for each new leaf hub until the point when both of two conditions is met:

1. Every property has just been incorporated along this way through the tree, or
2. The preparing illustrations related with this leaf hub all have a similar target characteristic esteem (i.e., their entropy is zero).

G. The ID3Algoritm

ID3 (Examples, Target_Attribute, Attributes)

- Create a root hub for the tree
- If all cases are sure, Return the single-hub tree Root, with mark = +.
- If all cases are negative, Return the single-hub tree Root, with name = - .
- If number of anticipating traits is void, at that point Return the single hub tree Root, with mark = most basic estimation of the objective characteristic in the illustrations.
- Otherwise Begin
  - A = The Attribute that best orders illustrations.
  - Decision Tree characteristic for Root = A.
  - For every conceivable esteem, vi, of A,
    - Add another tree limb underneath Root, relating to the test A = vi.
    - Let Examples (vi) be the subset of cases that have the esteem vi for A
    - If Examples(v_i) is unfilled
    - Then beneath this new branch include a leaf hub with name = generally normal target an incentive in the illustrations
    - Else beneath this new branch include the subtree ID3 (Examples (v_i), Target_Attribute, Attributes – {A})
- End
- Return Root

IV. Results And Discussion

The data set of 50 students used in this study was obtained from KNGAC Bharathidasan University, Thanjavur, Computer Science department of course M.Sc (Master of Computer Science) from session 2007 to 2010.

S. No.	PSM	CTG	SEM	ASS	GP	ATT	LW	ESM
1.	First	Good	Good	Yes	Yes	Good	Yes	First
2.	First	Good	Average	Yes	No	Good	Yes	First
3.	First	Good	Average	No	No	Average	No	First
4.	First	Average	Good	No	No	Good	Yes	First
5.	First	Average	Average	No	Yes	Good	Yes	First
6.	First	Poor	Average	No	No	Average	Yes	First
7.	First	Poor	Average	No	No	Poor	Yes	Second
8.	First	Average	Poor	Yes	Yes	Average	No	First
9.	First	Poor	Poor	No	No	Poor	No	Third
10.	First	Average	Average	Yes	Yes	Good	No	First
11.	Second	Good	Good	Yes	Yes	Good	Yes	First
12.	Second	Good	Average	Yes	Yes	Good	Yes	First
13.	Second	Good	Average	Yes	No	Good	No	First
14.	Second	Average	Good	Yes	Yes	Good	No	First
15.	Second	Good	Average	Yes	Yes	Average	Yes	First
16.	Second	Good	Average	Yes	Yes	Poor	Yes	Second
17.	Second	Average	Average	Yes	Yes	Good	Yes	Second
18.	Second	Average	Average	Yes	Yes	Poor	Yes	Second
19.	Second	Poor	Average	No	Yes	Good	Yes	Second
20.	Second	Average	Poor	Yes	No	Average	Yes	Second
21.	Second	Poor	Average	No	Yes	Poor	No	Third
22.	Second	Poor	Poor	Yes	Yes	Average	Yes	Third
23.	Second	Poor	Poor	No	No	Average	Yes	Third
24.	Second	Poor	Poor	Yes	Yes	Good	Yes	Second
25.	Second	Poor	Poor	Yes	Yes	Poor	Yes	Third
26.	Second	Poor	Poor	No	No	Poor	Yes	Fail
27.	Third	Good	Good	Yes	Yes	Good	Yes	First
28.	Third	Average	Good	Yes	Yes	Good	Yes	Second
29.	Third	Good	Average	Yes	Yes	Good	Yes	Second
30.	Third	Good	Good	Yes	Yes	Average	Yes	Second
31.	Third	Good	Good	No	No	Good	Yes	Second
32.	Third	Average	Average	Yes	Yes	Good	Yes	Second
33.	Third	Average	Average	No	Yes	Average	Yes	Third
34.	Third	Average	Good	No	No	Good	Yes	Third

35.	Third	Good	Average	No	Yes	Average	Yes	Third
36.	Third	Average	Poor	No	No	Average	Yes	Third
37.	Third	Poor	Average	Yes	No	Average	Yes	Third
38.	Third	Poor	Average	No	Yes	Poor	Yes	Fail
39.	Third	Average	Average	No	Yes	Poor	Yes	Third
40.	Third	Poor	Poor	No	No	Good	No	Third
41.	Third	Poor	Poor	No	Yes	Poor	Yes	Fail
42.	Third	Poor	Poor	No	No	Poor	No	Fail
43.	Fail	Good	Good	Yes	Yes	Good	Yes	Second
44.	Fail	Good	Good	Yes	Yes	Average	Yes	Second
45.	Fail	Average	Good	Yes	Yes	Average	Yes	Third
46.	Fail	Poor	Poor	Yes	Yes	Average	No	Fail
47.	Fail	Good	Poor	No	Yes	Poor	Yes	Fail
48.	Fail	Poor	Poor	No	No	Poor	Yes	Fail
49.	Fail	Average	Average	Yes	Yes	Good	Yes	Second
50.	Fail	Poor	Good	No	No	Poor	No	Fail

To work out the data pick up for A with respect to S, we first need to compute the entropy of S. Here S is an arrangement of 50 cases are 14 "Initial", 15 "Second", 13 "Third" and 8 "Fizzle".

$$Entropy(s) = \begin{matrix} -P_{First} log_2(P_{First}) - P_{Second} log_2(P_{Second}) \\ -P_{Third} log_2(P_{Third}) - P_{Fail} log_2(P_{Fail}) \end{matrix}$$

$$= \begin{matrix} -\left(\frac{14}{50}\right) log_2\left(\frac{14}{50}\right) - \left(\frac{15}{50}\right) log_2\left(\frac{15}{50}\right) \\ -\left(\frac{13}{50}\right) log_2\left(\frac{13}{50}\right) - \left(\frac{8}{50}\right) log_2\left(\frac{8}{50}\right) \end{matrix}$$

$$= 1.964$$

To decide the best trait for a specific hub in the tree we utilize the measure called Information Gain. The data pick up, Gain (S, An) of a quality an, in respect to a gathering of cases S,

$$Gain\ (S, PSM) = Entropy(S) - \frac{|S_{First}|}{|S|} Entropy(S_{First})$$

$$- \frac{|S_{Second}|}{|S|} Entropy(S_{Second}) - \frac{|S_{Third}|}{|S|} Entropy(S_{Third})$$
$$\frac{|S_{Fail}|}{|S|} Entropy(S_{Fail})$$

Table III. Pick up VALUES

Gain	Value
Gain(S, PSM)	0.577036
Gain(S, CTG)	0.515173
Gain(S, SEM)	0.365881
Gain(S, ASS)	0.218628
Gain (S, GP)	0.043936
Gain(S, ATT)	0.451942
Gain(S, LW)	0.453513

PSM has the most astounding increase, consequently it is utilized as the root hub as appeared in figure 2.

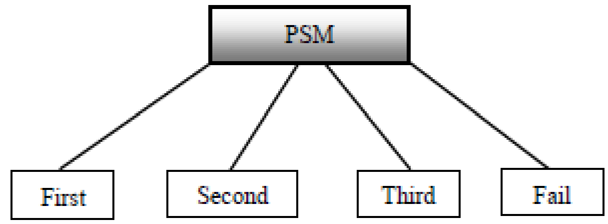

Figure 2. PSM as root hub

Pick up Ratio can be utilized for property determination, before ascertaining Gain proportion Split Information is appeared in table IV.

Table IV. Split Information

Split Information	Value
Split(S, PSM)	1.386579
Split (S, CTG)	1.448442
Split (S, SEM)	1.597734
Split (S, ASS)	1.744987
Split (S, GP)	1.91968
Split (S, ATT)	1.511673
Split (S, LW)	1.510102

Gain Ratio is shown in table V.

Table V Gain Ratio

Gain Ratio	Value
Gain Ratio (S, PSM)	0.416158
Gain Ratio (S, CTG)	0.355674
Gain Ratio (S, SEM)	0.229
Gain Ratio (S, ASS)	0.125289
Gain Ratio (S, GP)	0.022887
Gain Ratio (S, ATT)	0.298968
Gain Ratio (S, LW)	0.30032

This procedure goes ahead until the point when all data ordered impeccably or come up short on characteristics. The information spoke to by choice tree can be separated and spoken to as IF-THEN rules.

One order rules can be created for every way from every terminal hub to root hub. Pruning procedure was executed by evacuating hubs with not as much as wanted number of articles. On the off chance that THEN rules might be less demanding to comprehend is appeared in figure 3.

V. Conclusion

In this paper, the order assignment is utilized on understudy database to foresee the understudies division based on past database. As there are numerous methodologies that are utilized for data characterization, the choice tree technique is utilized here.

Information's like Attendance, Class test, Seminar and Assignment marks were gathered from the student's past database, to foresee the execution toward the finish of the semester.

This examination will help to the understudies and the educators to enhance the division of the understudy. This investigation will likewise work to recognize those understudies which required uncommon consideration regarding decrease come up short apportion and making proper move for the following semester examination.

REFERENCES

[1] Yongjian Fan, Yanguang Shen, Jianying Mai, "Study of the Model of E-commerce Personalized Recommendation System Based on Data Mining", Electronic Commerce and Security, PP. 647 – 651, 2008.

[2] Xuejun Yang, Hongchun Zeng, Yonghong Huang, "ARTMAP-Based Data Mining Approach and Its Application to Library Book Recommendation", Intelligent Ubiquitous Computing and Education, PP. 26 – 2, 2009.

[3] J.Sheela Jasmine, "Application of Fuzzy Logic in Neural Network Using Data Mining Techniques: A Survey", International Journal of Computer Sciences and Engineering, Vol.4, Issue.4, pp.333-341, 2016.

[4] Lavannya Bhatia, S. S. Prasad, "Building a Distributed Generic Recommender Using Scalable Data Mining Library", Computational Intelligence & Communication Technology (CICT), PP. 98 – 10, 2015.

[5] Liu Guo-rong, Zhang Xi-zheng, "Collaborative Filtering Based Recommendation System for Product Bundling", 2006 International Conference on Management Science and Engineering, PP. 251 – 254, 2006.

[6] Hiral Y. Modi, Meera Narvekar, "Supporting Domain Analysis through Mining and Recommending Features from Online Product Listings", IEEE Transactions on Software Engineering, Vol. 39, No. 12, PP. 1736 – 1752, 2013.

[7] Vincent Chan, Tony Hao Wu, Granfham Pang, "Decision support and data mining for direct consumer-to-consumer trading, Internet Technology and Secured Transactions (ICITST), PP. 205 – 208, 2014.

[8] Anand Shanker Tewari, Tasif Sultan Ansari, Asim Gopal Barman, "Opinion based book recommendation using Naive Bayes classifier", Contemporary Computing and Informatics (IC3I), PP. 139 – 144, 2014.

[9] Silvana Aciar, Debbie Zhang, Simeon Simoff, John Debenham, "Informed Recommender: Basing Recommendations on Consumer Product Reviews", IEEE Intelligent Systems, Vol. 22, No. 3 PP. 39 – 47, 2007.

[10] Suhasini Parvatikar, Bharti Joshi, "Online book recommendation system by using collaborative filtering and association mining", IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), PP. 1 – 4, 2015.

[11] Kumari Priyanka, Anand Shanker Tewari, Asim Gopal Barman, "Personalised book recommendation system based on opinion mining technique", Communication Technologies (GCCT), 2015 Global Conference, PP. 285 – 289, 2015.

[12] Shahab Saquib Sohail, Jamshed Siddiqui, Rashid Ali, "Ordered ranked weighted aggregation based book recommendation technique: A link mining approach", Hybrid Intelligent Systems (HIS), 2014 14th International Conference, PP. 309 – 314, 2014.

