

Prediction of Results in Education By Data Mining

Dr. Gulnar Sharma

JDBIMS, SNDT Mumbai

drgulnarsharma@gmail.com

Prof. Sushopti Gawade

PIIT, Panvel

sushopti@gmail.com

Prof. Pragati Unde

YTIT, Karjat

pragatiunde@gmail.com

Abstract:

Now a days, a lot of information is available everywhere, but you need to use it smartly and apply it wisely. Author here wants to focus on the basic terminology of data mining can be used to predict the final year results by using students' academic information since 1st standard to graduation. There are many factors which are needed to be considered for analysis. Data is available with all educational institutes during admission. But nobody uses it for improvement or development of the student, institute, university. Finally, author suggests to use the information to get knowledge about student, about institute and ultimately about university.

By using this small exercise, you can predict last year results, rate your products in terms of students, also rate your institute among others.

Keywords:

Datamining, results, knowledge, rating, prediction.

I. Introduction

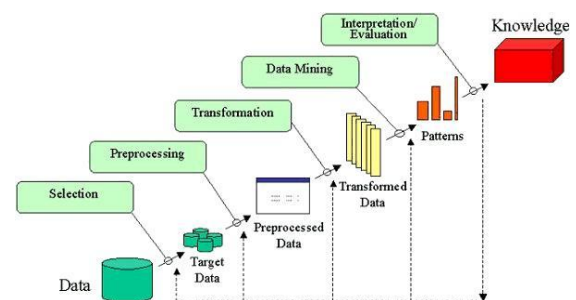
Data mining means discovering hidden value in your data warehouse. Data mining, the extraction of hidden predictive information from large databases. Data mining tools predict future trends and behaviours,

allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time-consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

A. Data Mining And KDD Process:

The term *Knowledge Discovery in Databases*, or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems, and data visualization.

The unifying goal of the KDD process is to extract knowledge from data in the context of large databases.



B. Steps to perform KDD process:

The knowledge discovery process contains five or more steps. Each step of knowledge discovery is briefly discussed below:

1. **Database and flat files:**
Database and flat files are the repository of data. Large volume of data is stored in huge and numerous databases and flat files.
2. **Data cleaning:**
Data cleaning deals with noisy, missy and inconsistent data.
3. **Data integration:**
The aim of data integration is to combine different data sources i.e. database, data warehouse and flat files.
4. **Data selection:**
The task of selecting and retrieving related and relevant data from massive data is done by data selection.
5. **Data transformation:**
Data transformation transformed data from

different format into a unified format.

6. **Data mining:**
Data mining is necessary step in which different data mining techniques are applied to search valuable data, information and knowledge.
7. **Pattern Evaluation:**
The job of evaluating the data provided by the data mining is carried out by the pattern evaluation to discover the patterns, behaviours, data trends and associations.

II. Case Study

The goal of this case study is divided into twopart;

1. **Transformation or prediction among different parameters.**
We can transform the result data in various parameters. As shown in the table the result can be created by year wise, semester wise or further we go for the subject in detailed. Each parameter can be divided into classes like distinction, first, second, pass and fail etc.

International Journal of Computer Architecture and Mobility
(ISSN 2319-9229) Volume 1-Issue 7, May 2013

Roll No	Name	SSC	HSC	1st year	2nd year	3rd year	Sem VII	Personal Problem	Prediction of Sem VIII
1	AkshayKhane	77.6	74.67	899	1011	1105	562	overconfidence	468
2	AlokTiwari	71.7	56.83	649	951	986	478	-	435
3	Amit Mehta	68.5	68.13	640/1300	930	930	560	-	462
4	Anil Nair	83.4	66	902	898	984	546	-	425
5	AnkitaArora	81.8	70.8	963	1000	1125	563	-	488
6	Bharat Khanna	80.5	72.5	930	887	962	496	-	456
7	BhushanDhormale	80.3	76.83	933	1017	1160	565	Illness	439
9	Deepak Yadav	73	71.8	951	1010	1045	521	-	395
10	DilipKolekar	82.7	72	888	943	1003	562	-	455
11	DipanwitaBiswas	74.1	74.33	684/1300	773	1045	538	-	453
12	FaizanFodkar	85.9	NA	NA	1072	1187	592	Disturbed	453
13	JijnasaPatil	85.5	76.83	762	1009	1007	548	-	475
14	Karishma Singh	74.7	NA	NA	983	1085	528	-	450
16	Mahesh Kad	88	78	717	1046	1123	535	-	421
17	MangeshPatil	80.5	70	937	1074	1104	560	-	503
18	MayurMathurkar	65.7	55	654/1300	923	1040	550	-	455
20	NeelamPanse	59	NA	NA	989	1017	563	-	469
22	Nikhil Lad	78.6	84.67	738	935	958	553	-	470
25	Pallavi Singh	77.2	69	960	957	1015	505	-	480
26	PoojaNaik	77.3	60.33	706	988	1005	567	-	495
27	Pornima Maid	75.9	63.17	743	1043	1088	592	-	520
28	PragatiUnde	80.9	72.33	924	1209	1205	599	Disturbed	523
29	PranavPatil	81	66	675	967	966	560	-	460
30	Prasad Walawalkar	73.7	72.83	974	1077	1020	469	-	469
31	Priti Rani Gupta	78.9	73	991	1118	1251	585	-	539
32	RichaRane	87	76.62	762	964	968	502	-	399
33	RupaliKasare	68	68.4	919	986	1077	565	-	489
34	RupaliKathvale	69.7	NA	NA	964	1163	570	-	540
35	Sachin Yewale	75.7	78.5	745	1045	1029	536	-	503
36	Seema Sharma	68.7	68.36	884	942	1032	470	-	445

2. Predictions for future results:

There are two data mining methods- Predictions method and Description method. In prediction method we use some variables to predict unknown or future values of other variables.

Example: classification, regression, deviation detection.

III. Proposed Framework

We use WEKA tool for our case study.

By using dataset, classifiers, weka.filters, weka.classifiers. using this we can predict the future result.

1. Dataset :

A set of data items, the dataset is basic concept of machine learning. A dataset is roughly equivalent to two dimensional spread sheet or database table. In WEKA it is implemented by the weka .core. Instances class. Each instance consist of attributes, each of which can be nominal, numeric or string.

In our example attributes are degree, percentages, result , grade etc. Example.

attribute degree(SSC,HSC,BE, ...)
attribute percen(85,75,74,...) attribute
result (pass, fail)
attribute grade(distinct,first,second)

2. Classifier:

A classifier is an arbitrary complex mapping from all but one dataset attributes to the class attribute. The specific form and creation of this mapping differs from classifiers to classifiers.

3. weka.filters:

theweka.filters package is concerned with classes that transforms dataset by removing or adding attributes, resampling the dataset, removing

examples and so on. This package offers useful support for data pre-processing which is important step in machine learning.

4. weka .classifiers:

Classifiers are core of the weka .Some of them are,

-t specifies the training file -T specifies the test file

-c this parameter specifies the class variable with one based-index.

and so on.

IV. Future Scope

As author has done this exercise on engineering institutes students. The same way it can be applied to other departments .So institutes result can be predicted. The same process can be applied to all the institutes

.This will be helpful to analyse the failure to achieve good academic results. In future author wants to work on employability skills of students. The same process will be followed, but knowledge will be more focussed on analytical ,apitude skills of students as well as skill set required by different industries.

V. Conclusion

Author here wants to focus on the basic terminology of data mining can be used to predict the final year results by using students academic information since 1st standard to graduation. There are many factors which are needed to be considered for analysis. Data is available with all educational institutes during admission

.But nobody uses it for improvement or development of the student, institute, university. This can be considered as a application of data mining which can give analysis of students academics

progress. Also it can be used to enhance and improve the results of students. Ultimately it can be helpful to achieve excellent results of institute, university.

VI. References

Websites:

[1] www.google.com

[2] Open source WEKA tool

[3] WEKA tool manual

[4].Thomas Connolly, Carolyn Begg. (2003)
Database Systems: A practical Approach To
Design ,Implementation
and 4th Ed.

[5]. M. Tamer Ozsu , Patric Vaduriez (2003)
Principal Of Distributed Database Sysytems.
2nd Ed.