SPICE: A Perspective Approach for Gap Analysis to Improve Student Performance Index

Satish Thatavarti
Research Scholar, Computer Science and Engineering, Gandhi Institute of Technology and Management University
Gandhi Nagar, Rushikonda, Visakhapatnam, Andhra Pradesh 530045, Visakhapatnam, India.

Orcid Id: orcid.org/0000-0002-5646-3448

Prof. K. Thammi Reddy
Professor, Department of Computer Science and Engineering, Gandhi Institute of Technology and Management University
Gandhi Nagar, Rushikonda, Visakhapatnam, Andhra Pradesh 530045, Visakhapatnam, India.

Abstract
Indian education system being the second leading in the world, following the United States often cited as one of the major contributors to the monetary rise of India. The main objective of every educational institution is to identify strengths and gaps of the students before they complete their education and offer employment opportunities to them who are in the finishing stage of completing the course. It is the responsibility of the institution to craft the students equipped on all aspects of their career growth and assist them in getting positioned in their dream companies. To meet the objective, improving the quality of delivery is one of the biggest challenges of Educational Institutions. The idea proposed in this paper is to perform correlation analysis considering number of parameters for the derivation of performance prediction indicators needed for student performance assessment, supervising and evaluation. It aims to predict the quality, efficiency and latent i.e., strengths and weaknesses of students across various disciplines which will enable higher level authorities to take decisions and understand certain patterns of student motivation, satisfaction, and growth.

Keywords: Data Mining, Student Performance Index Classification Enumeration, Predictive Analytics, Education Data Mining

INTRODUCTION
One of the main objective for premier educational institution in India are not only disseminating knowledge but also creating promising career opportunities for students in reputed multinational companies. Prediction of student’s achievement in higher learning institution is one means to reach the highest level of quality in higher education system [7][15][43]. The data mining methodology used for mining useful patterns from the student database is able to extract certain unidentified trends in student when assessed across several factors. Assessment as a dynamic process generates data, that acts as a performance indicator for an individual[39][41].

In Educational system data protection and processing by the universities and engineering colleges is very hard in the aspects of student’s results, to offer the facilities and provision of career guidance. In this regard, educational system needs an analytical approach to comprehend the data on educational realm to provide better solutions. Educational domain problems are addressed through various data mining techniques[12][21]. In educational data mining, among various techniques Predictive analytics is widely used to examine on past statistics and identify the future values to generate accurate results[16].

Educational data mining helps in forecasting the future patterns to make the organizations or institutions provide quality based education to the learners[23][48].

Many prediction models with diverse approaches in student performance were reported by researcher, but there is no certainty if there are any predictors that can conclude accurately whether a student will be an academic intellect, a drop out, or an normal performer[30].

This paper talks about the way the data mining technologies are used to grab out uncertain students and evaluate these learners to implement custom-made intervention strategies to diminish failure rate in higher education[17].

LITERATURE SURVEY
Survey has uncovered that data mining can be adopted to determine at-risk learners and facilitate institutes become much more practical in recognizing and responding to those learners. Luan applied data mining as a means to forecast the list of students going to dive out of school, and then revisit to institution later on. Classification and regression trees are applied to educational statistics to estimate the students who are not likely to get nearer to school[27].

Data mining has its lineage in machine learning, artificial intelligence, computer science, and statistics[13][25]. Clustering, classification, and association rule mining are
distinguished data mining techniques and approaches are being used [44]. All of these approaches help to quantitatively analyze huge data sets to reveal hidden meaning and patterns. Despite has not acknowledged much consideration in educational frameworks being the fact that data mining has been functional in various areas like banking, government, military, retail, and industries [36][40].

Blikstein, Dringus and Ellis [3] explored behavior of a learner using data extracting techniques. Specific mining methods are used with a particular data set. It is mainly crucial to discover techniques of applying data mining to study behavior of a learner in a broad way, relatively scrutinizing a single feature within the CMS.

Data mining helps in calculating academic-wise performance and achievement of a student [42]. Researchers put into practice data mining to categorize learners into three groups in the early phase of academic year[33] [46]. The three classes included learners with low, medium and high risk. They applied these mining techniques counting neural networks, random forests and decision trees. Group of high-risk students had a elevated chance of deteriorating or dropping out of school. This classification aid the faculty and team, a means to discover the susceptible learners[18].

Investigators inside EDM focus on studying and applying the data mining practices to develop institutional effectiveness and student learning methods[28][35]. Campbell and Oblinger portrayed that statistical and data mining techniques as academic analytics and described their role statistical techniques and data mining in ways that in helping faculty and advisors to become more practical in classifying at risk students and reacting accordingly[9].

Baker and Yacef stated EDM as “an promising discipline, concerned with developing methods to explore the exclusive types of data that arrive from educational settings, and using these schemes to better recognize students, and environment from which they learn in”[1].

A special approach adopted by Yeats et al. is to analyze student attainment in that it prepared the relation amid writing center attendance and student grades. It doesn't help student withholding concerns. Future study can help in inspecting the relationship involving three perceptions: writing center attendance, grade of a learner as well as retention[49].

In a further study, dissimilar data mining methods were adopted to conclude predictions on learner retention. Y u, DiGangi et al applied tree classification, multivariate adaptive regression splines and neural networks to institutional data that helped in finding hours transferred, residency, and traditions as serious essentials in retention efforts [50].

Mining on educational information(EDM) being an upcoming discipline focuses on applying tools, techniques to educationally associated data[1][34]. It focuses on analyzing institutional or academic data to build models to improve learning experiences and institutional efficiencies [38].

The extent of educational data mining comprises of areas that shows impact straight on students. Applying definite data mining procedures like classification, association rule mining, web mining and statistics are the key practices applied to educationally associated data [5][8]. The techniques can be applied to model individual dissimilarities in students and offer a way to counter them thus perk up student learning [11].

Data mining utilizes statistics, machine learning, and artificial intelligence methods. Research on EDM focuses mainly on quantitative analyses [19]. The dominant research standard is quantitative, with results approaching in the form of predictions, clusters, classifications or associations. Existing research studies results are not necessarily generalizable to further institutions as they are highly related to a specific institution for a specific period. Research in EDM ought to scrutinize ways for data mining results to be more generalizable.

As a means to improve student retention hardwork, Lin applied data mining. Lin produced predictive models depending on inward data of students. The models were capable to offer short-term accuracy for calculating which category of students would benefit from student retention programs. Research studies initiated that definite machine learning algorithms can afford valuable predictions of student preservation [24][26].

A system that supports and advances retention was developed by researchers at Bowie State University [10]. This system aids the institution spot and react to helpless students. Their research adds significantly to the EDM literature as it exhibits a victorious implementation and employ data mining. Their effort is highly representative of the discipline in that it go after a severe data mining process and is quantitative. Chacon et al.’s [10] research supports the work made in applying data mining to student retention concerns all with flourishing results.

RESEARCH METHODOLOGY

Problem Statement

A great profusion of information is contained in every educational institution. Many academic, demographic, and attitudinal variables are collected for each student who steps on campus. Regardless of all this information, colleges still struggle with graduation and employment toll. This is an appropriate demonstration of a “surplus of information but a starvation of knowledge”[2].
Proposed Methodology

It is essential to identify the prospective talent by predicting their performance by means of past performance and knowledge to ensure the student start the career and move ahead in the right path for better quality living[6][22].

The idea proposed in this paper is to perform correlation analysis considering number of parameters for the derivation of performance prediction indicators needed for student performance assessment, supervising and evaluation[6][14]. The aim is to predict the quality, efficiency and latent i.e., strengths and weaknesses of students across various disciplines which will enable higher level authorities to take decisions and understand certain patterns of student motivation, satisfaction, and growth[20][47]. The analysis depends on many factors, encompassing student's academic performance, consistent attendance, internal and external assessment, performance assessment in placement training etc. This knowledge can be extorted from other educational data that derive from the assessment processes[29][31].

For this, a novel algorithm, SPICE has been proposed. The aim of this study is to calculate Student Performance Index (SPI) by relating the Correlation method as data mining technique to analyze the relationships between students behavioral and their success to develop the model of student performance predictors’. This is done by using classification and regression techniques. The results of this study reported a model of student academic performance predictors by employing psychometric factors as variables predictors.

SPICE - Student Performance Index Classification Enumeration

Algorithm:

Step1: To read the volumetric data and find the performance attributes of each student.

Step2: To generate the correlation of each attribute of a student with the placements input, placed (1) or unplaced (0) based on linear relationship between two variables. Thus, Pearson’s correlation coefficient helps to predict the strengths and weaknesses of each student.

Step3: To find the Student Performance Index – SPI for each student using the following formula

$$\frac{\sum_{i=1}^{n}(A_i W_i)}{N} \times 100$$

Where $A_i$ – ith Attribute and $W_i$ is the weight assigned to ith attribute

Step4: To classify the group-wise attributes of each student into 3 classes

TP: Trained and Placed, TNP: Trained but not Placed, WTP: Without Training Placed

Step5: To find out the gap between placed and non-placed students from both trained and untrained students sets of strengths and weaknesses using set theory.

At the Initial Stage, the performance attributes of each student taken into consideration are: Assignments, MID Marks, Attendance, Final Exam and Placement Test.
The initial phase revolves around reading the student primary data collected previously. The group-wise performance attributes are then extracted as a list followed by the next phase that produces pair-wise correlation coefficients between each attribute in the generated attribute list and placement data value. The accomplishment of the stage is the prognosis of strengths and weaknesses of each student. Thus, the process flow is applied to every student as he/she is unique in his/her strengths and challenges. This juncture also helps to calculate their performance index that shows their potentiality in major areas. The final stage organizes the attributes into three different classes of students who are trained and successfully placed, trained but not placed and the students who are placed without any coaching. Thus the categorization assists in addressing student learning gaps and suggests the areas to create promising career opportunities in view of the fact that, “Employment is the driving force behind economic growth of a Nation”.

Pearson r

The way to measure strength between variables and relationships is to use Pearson correlation coefficient (r)[51]. The formula to compute the Pearson correlation coefficient, r is

\[ r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \]

Here, \( r \) = Pearson correlation coefficient
X = Values in the first set of data
Y = Values in the second set of data
N = Total number of values.

This formula produces the coefficient value to determine how strong the relationship is between two variables. The value of r ranges between +1 and -1:
- \( r > 0 \) indicates a positive relationship of X and Y
- \( r < 0 \) indicates a negative relationship
- \( r = 0 \) indicates no relationship

<table>
<thead>
<tr>
<th>Quant-Average1</th>
<th>Quant-Average2</th>
<th>Tech-Eng-1</th>
<th>Tech-Eng-2</th>
<th>Tech-Eng-3</th>
<th>AMCAT Logical-1</th>
<th>AMCAT Logical-2</th>
<th>AMCAT Programming</th>
<th>AMCAT Quant-1</th>
<th>AMCAT Quant-2</th>
<th>Placed</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>12</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>16</td>
<td>8</td>
<td>11</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>29</td>
<td>36</td>
<td>39</td>
<td>0</td>
<td>14</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>16</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>25</td>
<td>51</td>
<td>39</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>18</td>
<td>48</td>
<td>43</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>27</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>6</td>
<td>30</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>19</td>
<td>40</td>
<td>47</td>
<td>0</td>
<td>21</td>
<td>10</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>44</td>
<td>35</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td>56</td>
<td>42</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>13</td>
<td>51</td>
<td>52</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>30</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>22</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>23</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>23</td>
<td>45</td>
<td>40</td>
<td>0</td>
<td>25</td>
<td>11</td>
<td>0</td>
<td>30</td>
<td>28</td>
<td>1</td>
</tr>
</tbody>
</table>
Here Placement test Data set contains 10 attributes each of which are correlated with Placed attribute. The Pearson correlation coefficient, $r$, can take a range of values from +1 to -1. If the coefficient is '0', there is no association between attributes and if it is nearly +1 there is a strong association between the attributes. Apply the similar formula to each pair of one among ten attributes and placed data to find the positive and negative aspect of each student.

Table 3.2.2 Example to show the calculation of Pearson correlation between two attributes: Quant- Average1 and Placed

<table>
<thead>
<tr>
<th>Quant-Average1 (x)</th>
<th>Placed (y)</th>
<th>$X^2$</th>
<th>$Y^2$</th>
<th>$xy$</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>0</td>
<td>676</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>900</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1024</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>484</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>361</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>196</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>576</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>900</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>676</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>729</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>400</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>324</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>400</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>841</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>841</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>676</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>1089</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sum(x)=425</td>
<td>Sum(y)=13</td>
<td>Sum($x^2$)=11093</td>
<td>Sum($y^2$)=13</td>
<td>Sum(xy)=269</td>
</tr>
</tbody>
</table>

Here, N= number of pairs of attributes =18

\[ \Sigma XY=\text{Sum of products of paired attributes } = 269 \]

Sum of X attributes $\Sigma X= 425$ ;

Sum of Y attributes $\Sigma Y=13$

Sum of squared X attributes $\Sigma X^2=11093$ ;

Sum of squared Y attributes $\Sigma Y^2= 13$

Based on these correlation factors, the attributes are organized into three groups: Trained and Placed, Trained but unplaced, Placed without Training to find the strengths and weaknesses of each student, results as shown follows.

**Figure 3.2.2:** the positive and negative attributes of the above categories to find the gaps between placed students

**Figure 3.2.3:** the positive and negative attributes of the above categories to find the gaps between non placed students
EXPERIMENTAL RESULTS

This system is built and implemented using C# language built on .NET framework; Live Charts - a real time charting tool, Visual Studio 2013- an IDE and SQL Server, a SQL-based relational database management system and Win Forms, a graphical GUI technology.

CONCLUSION

The idea of the gap analysis and calculation for SPI is to give knowledge, skill and meet the manpower necessities of the Industry and to direct the students opt right career[32]. Understanding, predicting and avoiding the academic failure are intricate and incessant processes anchored in past and present information collected from pedagogic situations and students performance[4]. In the proposed system, SPICE the experiments were carried out in the educational area, based on correlation of attributes and classification in order to depict the student’s performance index and to find the gaps to suggest improvements in the weak areas for academic success.

ACKNOWLEDGEMENT:

This paper publication is a part of major research project funded by University Grants Commission (UGC), MHRD, titled: “Usage of Data mining techniques in design and...
development of Academic Audit System for effective Teaching Learning Process”.

REFERENCES


© Research India Publications.  http://www.ripublication.com


[27] Luan, J. (2002). Data Mining and Knowledge Management in Higher Education Applications. Paper presented at the Annual Forum for the Association for Institutional Research, Toronto, Ontario, Canada


[31] Pratiwi, O.N.(2013) "Predicting Student Placement Class using Data Mining” IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Pp:618 – 621


Improving quality of educational processes providing new knowledge using data mining techniques.

