Integrated Learning Framework towards attaining focused outcomes

P. G. Sunitha Hiremath
ISE Department, BVBCET
Hubli, India
pgshiremath@bvb.edu

Shankar Setty
ISE Department, BVBCET
Hubli, India
shankar@bvb.edu

Abstract— **With the changing demands and the needs of the industry, integrated learning becomes a major necessity today. Academicians need to play a vital role in developing such skills in students. To inculcate integrated learning into students we need to attain focused outcomes. In this paper, we propose an integrated learning framework towards attaining focused outcomes. We considered two major core courses - data mining and web technologies - in computer science domain at undergraduate level to develop integrated learning framework. The framework is a three stage process comprising of problem definition phase, knowledge discovery from database phase involving pre and post data processing, and finally validating results phase. In this paper, we discuss prototype of combining the courses data mining and web technologies to come out with an objective of "developing a web based data mining application" as a course project to cater the needs and to explore the technologies and concepts that could be glued together to develop rich internet applications. The experimental results show that the applications are full-fledged and addresses the Accreditation Board for Engineering and Technology (ABET) focused outcomes 'c' (An ability to design a computer based system, component or process to meet the desired needs within realistic constraints) and 'k' (An ability to use the techniques, skills, and tools necessary for computer engineering practice). The key focus was on integrated learning need so as to develop additional skills in students such as presentation skills, team-work, self-learning and analyzing the real world problems which in-turn improves the students' confidence level. This is reported in outcome assessment.**

Keywords — Integrated learning, data mining, web technologies, focused outcomes.

I. INTRODUCTION

Today with booming market where Information Technology is trending towards Big Data and Analytics, Internet Of Things, Smart Machines, 3D Printing, and Digital Business as predicted in Gartner [2]. The information technology trends are majorly focusing primarily on integrating many fields (a.k.a. courses in academics). For example, Big Data and Analytics is interconnection of fields (courses) like Database Management Systems, Data mining, Web technologies and Business Intelligence for collecting, organizing, analyzing and reproducing information in decision making. This clearly implies that industries are expecting students with integrated learning experience. However, the challenge in today's engineering education at undergraduate level is to identify the courses, needed to be integrated towards meeting the industry needs. In this paper we consider the major focus towards "development of data mining web application" using the concepts / techniques acquired from data mining course and web application techniques adopted from web technologies course.

Data mining and web technologies are the two major core courses in Information Science and Engineering undergraduate program at sixth semester level. The course Data Mining [3], also popularly referred to as Knowledge Discovery from Database (KDD), is the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the web and other massive information repositories or data streams. The field of data mining has evolved from the disciplines of statistics and artificial intelligence. Data mining finds its applications in areas such as recommender systems, credit rating, fraud detection, database marketing and stock market investments. The course Web Technologies [6] aim to build proficiency among students for developing web applications and web sites providing a practical exposure of various client side and server side technologies. In this paper we propose to integrate both data mining and web technology courses for integrated course project development which leads into integrated learning. To facilitate integrated learning a dynamic learning environment is required which is provided by a practice-oriented approach [1].

Inculcating of integrated learning into students is attainable only by meeting the focused outcomes. ABET accreditation [7] is a body which accredits engineering institutes to attain criteria 'a' to 'k' as program outcomes. In this paper we focus on outcomes 'c' (An ability to design a computer based system, component or process to meet the desired needs within realistic constraints) and 'k' (An ability to use the techniques, skills, and tools necessary for computer engineering practice) of ABET criteria.

The main objective of carrying out integrated course project is to motivate students to develop an application focusing on data mining concepts, utilizing the knowledge gained from web technologies course in a synchronized manner. In this paper we developed integrated learning framework involving following milestones to be achieved.

- **Problem Identification:** Identify, analyze and collect relevant real time data.
- **Algorithm Identification:** Analyze and choose appropriate algorithm to demonstrate the usage of data mining tools in KDD process.

- **Developing Application:** Apply client and server side technological concepts to devise a solution for the KDD process.

The main contribution of this paper is (i) proposing integrated learning framework for data mining and web technologies courses, (ii) evaluating the attainment of focused outcomes 'c' and 'k' and (iii) assessment of focused outcomes.

The rest of the paper is organized as follows: Section II describes the integrated learning framework, plan which includes the initiation, problem identification, appropriate algorithm selection, development of web application, validation and evaluation. Section III shows the outcome assessment of carrying out integrated learning. Section IV describes one of the implemented projects towards the complete process and Section V ends with conclusion.

II. INTEGRATED LEARNING FRAMEWORK

The integrated learning framework corresponding to data mining and web technology courses is shown in Fig. 1. The planned schedule to carry out the integrated course project was displayed to students before the commencement of the semester. Integrated course project extends for 12 weeks of the semester. There are two reviews: (i) review, which comprises of demonstration and presentation which carries certain weightage of marks and (ii) self-review, which basically deals with clarification of issues, problems or other discussions pertaining to course project with instructors. The self-review does not carry any marks. Three reviews were planned at the 4th, 9th and 12th week of the semester and deliverables expected are clearly mentioned. Two self-reviews are planned during 6th and 8th week. The steps involved in integrated learning framework are:

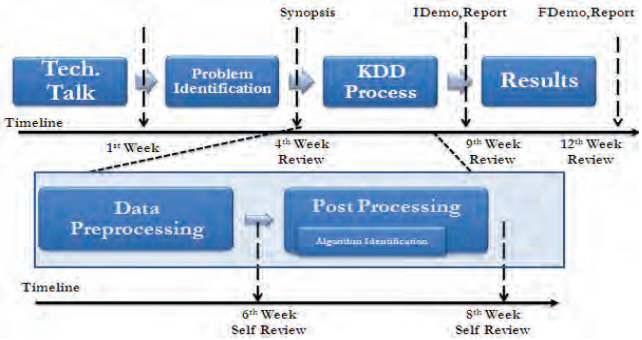

Figure 1: Integrated Learning Framework

A. The Initiation and problem identification

During the commencement of the semester it is observed that it is difficult to give crisp idea over the data mining and web technologies courses. The first step in this framework was to give an introduction to integrating data mining and web technologies courses and make the students understand the importance of integrated learning from the perspective of industry and higher education. We also found that it was largely difficult to convince students on application development by integrating two courses. To overcome this difficulty we adopted the strategy of inviting alumni students as resource persons for technical talk. To motivate students a technical talk was organized on "Significance of Data Mining and Web Technologies in Industry". We invited the resource persons from industries and also those who were pursuing their higher studies abroad. The second step in the process was the student group formation and problem identification. This leads into achieving the first milestone of integrated learning framework. A time line of four weeks was allotted to analyze and collect relevant real time data for the identified problem. The deliverable at this stage was the submission of the synopsis containing the problem statement, real time data, description and methodology about the project. A review was conducted and guidelines were given to carry out further work. The challenges faced by students in this phase were getting the real time data, understanding the data, and finally deciding on what analysis they must carry out. Students were encouraged to collect real time data by communicating with organization like ward voter department, tourism department, police department, college placement cell, competitive exams and from other internet sources. This has lead into development of problem analysis skills among the students. The evaluation weightage given for this phase was 25% with respect to the outcome 'c' as shown in Table 1.

B. Knowledge Discovery from Database (KDD) Process

The KDD process is the most important phase for analyzing and selecting the most appropriate data mining algorithm. This leads into achieving the second and third milestone of integrated learning framework. This phase involves two stages data pre-processing and post-processing and lasts for 5 weeks, from 5th week to 9th week.

1. Data preprocessing

The real world data is mostly noisy and impure, that is it may be incomplete, noisy and inconsistent, hence preprocessing of the data is required. To enhance students' knowledge on data preprocessing, students were encouraged to use a data mining tool in the data mining course (precisely in theory class) to explore the data mining processes, algorithms, and functions. The first warm-up hands-on project was for students to get exposure to the data mining tool such as Knime[4]. The Tool based approach for teaching data mining as proposed in [5] was used to provide students an opportunity for applying basic theoretical concepts taught in the class. The data preprocessing stage involved using open source data mining tools. Students identified suitable data mining tool for their respective project and the real data set was cleansed and fed to the data mining tool. Preprocessing techniques are explored, applied to the dataset and stored. For the identified preprocessing techniques web technology based Javascript frameworks such as Angular.js, Backbone.js, Node.js, etc were used for implementation. Javascript frameworks such as Angular.js, Backbone.js, Node.js are the leading and demanding frameworks for many application developments in industry today. This phase led into selecting the appropriate pre-processing techniques by using chosen data mining tool and then implementing the pre-processing module using javascript

frameworks. The obtained results were checked in the second review.

Table 1: Mapping of evaluation criteria as per data mining and web technology courses with focused outcomes and blooms taxonomy.

Stages (Figure 1)	Evaluation criteria as per Data mining	Evaluation criteria as per Web technologies	Weightage (%)		Blooms level	Time line
			Outcome 'c'	Outcome 'k'		
Problem Identification	**Identify, analyze and collect relevant real time data.**	**-**	**25**	**-**	**L4**	**Review 1 (Week 4)**
KDD process development **Data Pre-processing** **Data Post Processing**	**Tool Utilization** **[i] Appropriate preprocessing technique** **[ii] Appropriate algorithm identification**	**Pre-processing implementation using Javascript Framework.** **Implementation of the mining algorithm using PHP Frameworks**	**10** **20**	**15** **15**	**L5**	**Review 2 (Week 9) Intermediate Demo**
Results	**Validation of results**	**Demonstration of the Project**	**15**	**-**	**L3**	**Review 3 (Week 12) Final Demo**

2. Data post processing

For the identified data mining functionality in the pre-processing stage, suitable data mining algorithms need to be chosen during the post-processing stage. To facilitate this, student teams explored the algorithms supported by the data mining tool for the chosen functionality. Certain projects took the challenge of determining new data mining algorithms appropriate for their project, which was not a part of their syllabus that they had learnt. In such cases students were exposed to self-learning which is one of the outcomes of ABET. Each student team explored various data mining algorithms and tried them using the tool on pre-processed dataset and interpreted the results. By comparing the obtained results for various algorithms, each team came up with a conclusion of the best algorithm suited for their dataset. The post processing module of the project was implemented using PHP Frameworks such as CodeIgniter, Symfony, CakePHP, PHPFusion, Yii, Content Management System, etc. The web technology PHP frameworks adopted for implementation highly correlates to the industry needs. This phase has led into effective utilization of data mining tool for choosing appropriate algorithm and implementing using the appropriate PHP framework. The obtained results were checked in the second review.

Self-review was the highlight of both the data pre-processing and post-processing phase as it helped students to discuss problems or issues openly with all other teams and reviewers for inputs. Self-review does not carry any marks. The second review was evaluated in the 9th week. Focus was not just on pre or post processing tool and implementation techniques but also on demonstration, oral and written presentation. The evaluation weightage given for data pre-processing phase was 10% with respect to the outcome 'c' and 15% with respect to outcome 'k'. The evaluation weightage given for data post-processing phase was 20% with respect to the outcome 'c' and 15% with respect to outcome 'k' as shown in Table 1.

C. Validation of results

The third phase is the validation phase in which each team had to rework on the project based on the reviewer comments. The teams had to do a thorough testing and validation over the developed data mining web application. Finally students demonstrated working of the complete project in which more emphasis was on validation of the results using metrics which was implemented using PHP frameworks. The efficiency or the accuracy was represented using appropriate visualization tools. The third review was carried at the 12th week and evaluated against the outcome 'c' for 15% of the weightage.

III. OUTCOME ASSESMENT

Integrated learning has developed the programming, analytical (outcome "c") and tool utilization (outcome "k") skills in students. These outcomes are measured by conducting the continuous internal evaluation and semester end exams (as show in Fig. 3). However, the key focus was integrated learning need, to develop additional skills in students' such as presentation skills, team-work, self-learning, analyzing the real world problems which in-turn improves students' confidence level. Our major focus was on outcome assessment of these skills which directly has an impact on students' placement.

Assessment 1 – Conducting workshop

A team of six students organized a one day workshop on "Developing web-based data mining applications" for computer science and information science students from other colleges. The topics covered were usage of data mining tools in KDD process and hands-on session to get exposure on projects development. Hands-on-session comprised of exposure to data mining tools and web-based data mining application development using five case studies. One of the case studies is described as a sample project in section IV. The assessment was done by collecting the feedbacks from the workshop audience to

measure their satisfaction as shown in Fig. 2. The feedback analysis clearly shows that audience agree that integrated learning is essential in engineering education. The pre-processing and post-processing techniques using tools seemed to strongly influence understanding of KDD process concepts. Though on an average many students agree that developing data mining application using tools and web technologies is clear. But, when we analyzed why it was not strongly agreed, we came across two reasons Firstly, students expressed that they needed more time in understanding and developing. Secondly, as the workshop was conducted in the first month of the commencement of semester not all concepts were known to them. However, students felt that overall the workshop was effective. This analysis shows that the expected outcomes of the integrated learning were achieved.

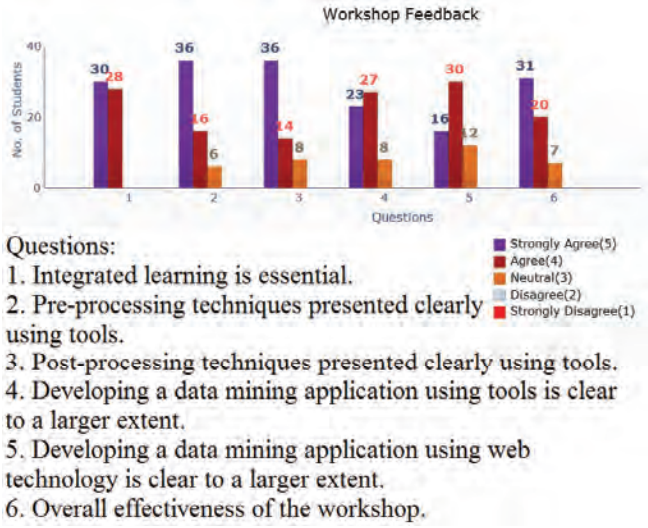

Figure 2: Audience satisfaction based on the students organized workshop.

Assessment 2 – Results and placement

Fig. 3 shows courses when offered independently perform marginally better in comparison with integrated course project. However, the objectives, expectations, and learning of integrated course project have lead into presentation skills, team-work, self-learning, analyzing the real world problems which in-turn improves students' confidence level. Though the passing percentage is bit low but we found that the campus placement due to integrated learning framework had lead into 10% raise over the previous years (2013) placement on campus (as of writing this paper, currently on campus placement is under way more results are likely).

Assessment 3 - Integrated course project feedback

Students who got placed in placement were questioned with a set of feedback questionnaires. Fig. 4 shows the analysis of integrated course project. A sample study was done by taking feedback questionnaires from 50 students only. Most of the students either agreed or strongly agreed towards influence of integrated course project towards campus placement. Applying integrated learning framework we attained the set milestones and the focused outcomes 'c' and k'.

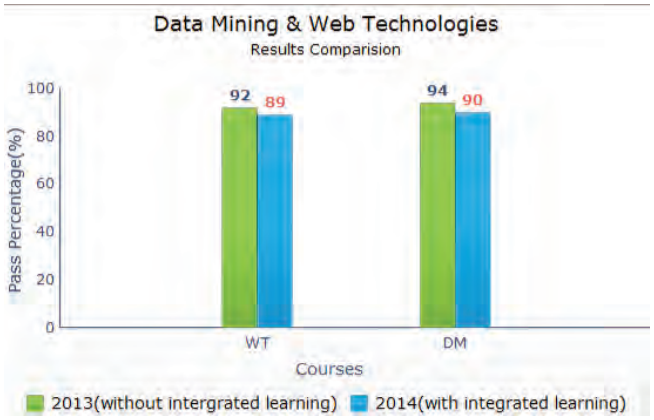

Figure 3: Results comparison of courses (WT, DM) and pass percentage without (2013 year) and with (2014 year) integrated learning framework.

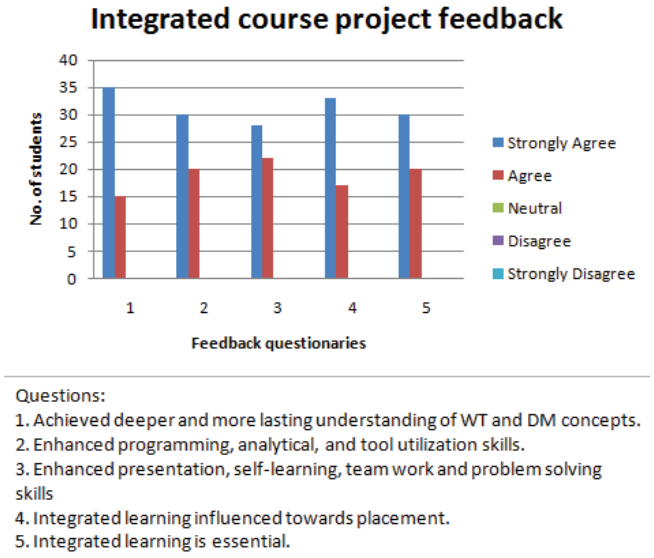

Figure 4: Integrated course project feedback analysis.

IV. A SAMPLE PROJECT - *GATE QUESTION AND ANSWER ANALYZER*

A. Problem Identification

Design a web application to predict the number of GATE(Graduate Aptitude and Test in Engineering)[8] questions that may appear in the syllabus of a particular domain (such as DBMS, Computer Networks, Operating System in Computer Science) for the current year. GATE is a competitive exam for qualifying for higher studies. Students willing to take GATE exam needs some kind of strategy to know how many questions would be asked in each domain for preparation. As questions in GATE follow certain pattern of similarity, it is necessary to find the pattern of repetition. This

application will classify the questions into suitable category based on past 13 years of GATE question papers (2000 – 2013). This classification leads into prediction of the number of questions of a particular domain in current year. In this project, the challenge was to collect the previous year GATE question paper (2000 to 2013) which was available in UTF encoded form and was difficult to process. The data needed to be converted to tool specific format. Fig. 5 depicts the system model of GATE question prediction. The identified problem was presented during review one.

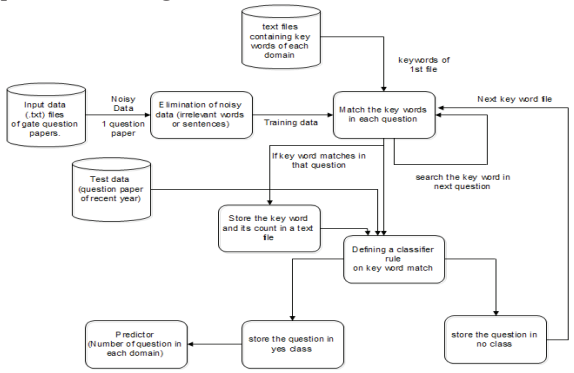

Figure 5: System model for GATE questions prediction

B. KDD process

KDD process comprises of two steps, data preprocessing and data post processing.

1. Data Preprocessing

In data preprocessing step, students explored different data mining tools such as Knime [4], Weka [9], Orange [10] and Rapid miner [11]. For this project Knime tool was chosen. The raw data (question papers of GATE exam) which was in PDF format was converted to tool specific format. The question papers also had noise such as web address, instructions, etc. Noise was removed using options like rule based row filtering and string replacer which was part of Knime tool. Fig. 6 shows the sample data of GATE 2005 computer science question paper. Knime tool displays data before and after pre-processing. Once the selected pre-processing methods were found to be successful using Knime tool, next coding of these methods was essential. The pre-processing methods were implemented using javascripts framework. These results were presented during the second review.

2. Data post processing

The data mining functionality chosen for the post processing of this data was classification and prediction. The suitable algorithms for these functionalities were explored using the Knime tool. The Knime tool supports the classification algorithms ID3, C4.5 and cart. The ID3 algorithm was chosen and applied on the training data for classification to build the model. Fig. 7 represents the output of ID3 algorithm in Knime tool. The identified classification algorithm was implemented using PHP framework. This project also required keyword matching which was not supported by Knime tool. However, coding of this was essential for frequency of occurrence of the question along with the corresponding year. This module was also implemented using PHP. Final output of this module was stored in CSV formatted file.

Next module, to predict the number of questions for the year 2014, linear regression learner and predictor was chosen. The CSV file obtained from the classification model was applied as the input to this predictor model. This model predicted the number of questions for computer networks domain as 7 for the year 2014. Fig. 8 shows the number of questions appeared for the computer networks domain in each year and also the predicted number of questions for the year 2014.

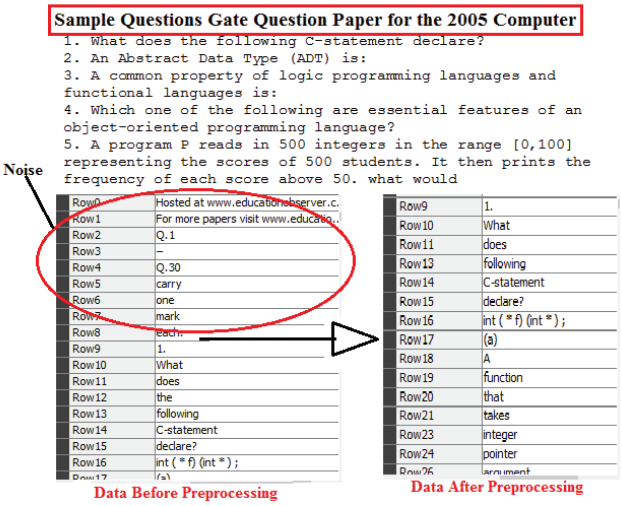

Figure 6: KDD Process - The top row shows the sample set of GATE 2005 computer science questions. Tool display of sample data before pre-processing (bottom left) and tool display of sample data after the pre-processing techniques adopted eliminated noise (bottom right).

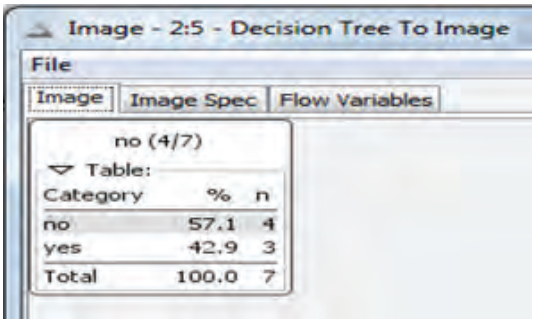

Figure 7: Classification (ID3 algorithm) of data using Knime tool.

The post processing of the data (classification and prediction) was implemented using PHP frameworks, string matching and regular expressions. Prediction of number of questions for the year 2014 for computer networks domain was implemented using linear regression formula written in PHP. The number of questions predicted for the computer network domain was 7 which were same as obtained using the data mining tool. During both pre-processing and post-processing phases the student teams discussed with other teams for issues,

suggestions, inputs. It was like a self review for all the teams. The reviewers accompanied the discussions during self-reviews and provided comments when expected. The output of this KDD process was reviewed during review two.

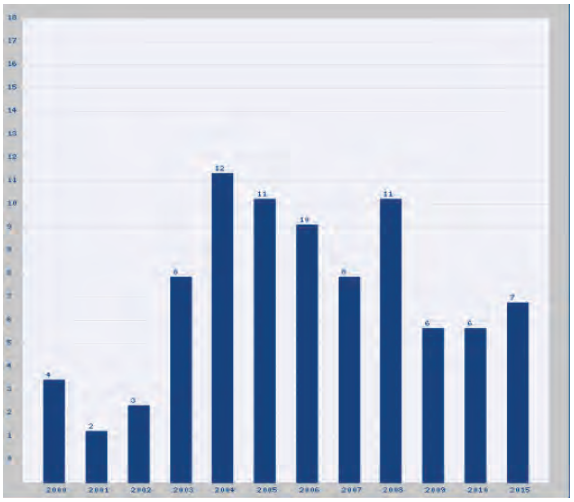

Figure 8: Prediction graph for the domain computer networks for the year 2014.

C. Validation of Results

The validation of the results was done using accuracy measures (accuracy= TP+TN/P+N where TP = True positive TN= True negative P= positive and N= negative tuples). The accuracy metric was implemented using PHP graph concepts.

The actual count of the questions for computer network domain was 6 and the predicted count was 7 for the year 2014 as shown in Fig. 8. Fig. 9 shows the accuracy of the predicted data which was 95%.

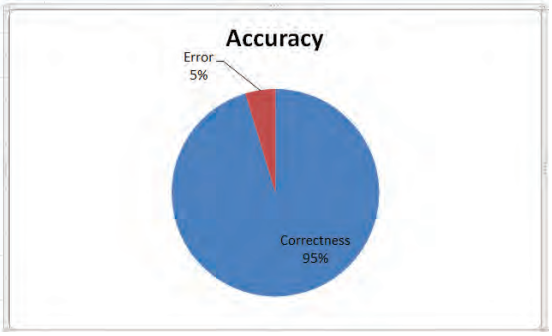

Figure 9: Accuracy for predicted results on computer networks domain for the year 2014.

V. CONCLUSION AND FUTURE WORK

In this paper we proposed integrated learning framework, evaluation process and the outcome assessment from various perspectives. Integrated learning of data mining and web technologies resulted in enhancement of analytical, programming (outcome "c") and tool utilization (outcome "k") skills. The major focus was towards attainment of enhancing presentation, self-learning, and problem solving skills to achieve confidence in students. The integrated learning project has lead into growth in the placement results. This opens doors to other courses which could further influence for better teaching learning process. Future work is to expand this framework to collaborate with other courses such as business intelligence analytics, big data, etc.

REFERENCES

[1] D. Li, "A Practice-Oriented Approach to Teaching Undergraduate Data Mining Course", 2011 American Society for Engineering Education (ASEE) Annual Conference, June 2011.

[2] Gartner Predicts 2014 - http://www.gartner.com/technology/research/predicts/

[3] Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, Elsevier, 2006.

[4] Knime - https://www.knime.org/

[5] Musa Jafar and Russell Anderson, "A Tools-Based Approach To Teaching Data Mining Methods", Journal of Information Technology Education: Innovations in Practice, Volume 9, 2010.

[6] W3Schools - http://www.w3schools.com/

[7] ABET Accreditation - http://www.abet.org/accreditation/

[8] GATE - http://en.wikipedia.org/wiki/Graduate_Aptitude_Test_in_Engineering

[9] Weka - http://www.cs.waikato.ac.nz/ml/weka/

[10] Orange - http://orange.biolab.si/

[11] Rapid Miner - http://rapidminer.com/