PREDICTING ACADEMIC PERFORMANCE BASED ON SOCIAL ACTIVITIES

MAGESH[1], MOHAMMED ARSATH[2], MANIKANDAN[3], SUNITHA NANDHINI[4]

[1,2,3]BE Students, Dept. of Computer Science and Engineering "Sri Krishna College of Technology",
Coimbatore, Tamilnadu, India.
[4]Asst.Prof.Dept. of Computer Science and Engineering "Sri Krishna College of Technology"
Coimbatore, Tamilnadu, India.

---***---

Abstract - *Prediction modeling is very important a part of data learning analytics, whose main objective is to estimate student success, in terms of performance, knowledge, score or grade. The knowledge used for the prediction model may be either state-based data (e.g., demographics, psychological traits, past performance)or event-driven knowledge (i.e., based on student activity). The latter are often derived from students' interactions with instructional systems and resources; learning management systems area unit a wide analyzed knowledge supply, whereas social media-based learning environments area unit scarcely explored. Data is collected from a Web Applications Design course, in which students use wiki, blog and micro blogging tools, for communication and collaboration activities in a project-based learning scenario .In addition to the novel settings and performance indicators, innovative regression algorithmic rule is employed for grade prediction. Very good correlation coefficients are obtained and EIGHTY FIVE percentages of predictions are within one point of the actual grade, outperforming classic regression algorithms.*

1. INTRODUCTION

Learning analytics (LA) is a growing research area, which aims at selecting, analyzing and reporting student data (in their interaction with the online learning environment), finding patterns in student behavior, displaying relevant information in suggestive formats; the end goal is the prediction of student performance, the optimization of the educational platform and the implementation of personalized interventions . According to the Society of Learning Analytics Research1, LA can be defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and also the environments during which it occurs" . The topic is highly interdisciplinary, including machine learning techniques, educational data mining, statistical analysis, social network analysis, natural language processing, but also knowledge from learning sciences, pedagogy and sociology ; up-to-date overviews of are the world |the realm} are provided in.Various educational tasks can be supported by learning analytics, as identified in analysis and visualization of data; providing feedback for supporting instructors; providing recommendations for students; predicting student's performance; student modeling; police work undesirable student behaviors; grouping students; social network analysis; developing idea maps; constructing courseware; designing and programming. Similarly, seven main objectives of learning analytics are summarized in monitoring and analysis; prediction and intervention; tutoring and mentoring; assessment and feedback; adaptation; personalization and recommendation; reflection. The prediction of students' performance is one of the most popular goals of LA , which aims to estimate future learning outcomes and identify indicators for learning success more specifically, the objective is to develop model which can infer the students' academic performance (i.e., the predicted variable, generally in the form of grades or scores) from a combination of various indicators (i.e., predictor variables) from the educational dataset. The predictive information is highly valuable, as it can offer instructors the ability to monitor the learning progress and provide students with personalized feedback and interventions in particular, the instructor may be suggested regarding students at- risk, who are in need of more assistance. In addition, personalized ways for rising participation might also be steered. Furthermore, the automatic prediction mechanism may be used for a formative assessment tool, which has the potential to decrease the instructors' assessment loads. Finally, providing prediction results and personalized feedback can foster students' awareness. Performance prediction has been extensively studied in web-based academic systems and, in particular, in Learning Management Systems (LMS). This is thanks to the provision of enormous amounts of student behavioural knowledge, automatically logged by these systems, such as: visits and session times, accessed resources, assessment results, online activity and involvement in chats and forums, etc. Thus, student performance prediction models supported Moodle log knowledge are projected in multiple previous studies. Additionally, log knowledge from intelligent tutoring systems (ITS) have additionally been used for performance prediction. In distinction, the students' engagement with social media tools in rising social learning environments has been less investigated as a possible performance predictor.

Therefore, our objective is to address academic performance prediction based on social media traces, in the novel context of social learning environments. More specifically, we focus on our eMUSE platform, which integrates three social media tools (wiki, blog and micro- blogging tool). These tools were used by Computer Science students enrolled in a Web Applications Design course, to support communication and collaboration activities in a project-based learning (PBL) scenario. Data was collected from six consecutive course installments (unfolding over six years), with a total of 343 students, leading to a relatively large educational dataset. A further novelty of our approach consists in the use of an innovative regression algorithm called "Large Margin Nearest Neighbor Regression" (LMNNR) for grade prediction, based on students' activity on wiki, blog and micro-blogging tool. Very good results are obtained, outperforming commonly used regression algorithms. The rest of the paper is structured as follows: in section II we provide an overview of related work on performance prediction. Subsequently, in section III we present a short technical background, including a description of the LMMNR algorithm and an overview of the algorithms used for comparison. The results obtained by applying these algorithms in our social learning environment context (which is described in more detail in section IV) are reported and discussed in section V. We end the paper with some conclusions and future research directions.

2. PRELIMINARIES

The purpose of this paper is to report and analyse the student marks and to predict the future score of the student. The basic block diagram contains the following modules like Fig1.Our focus is on the third and subsequent blocks.

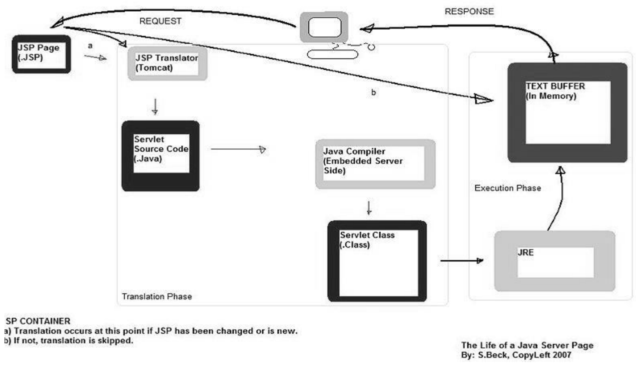
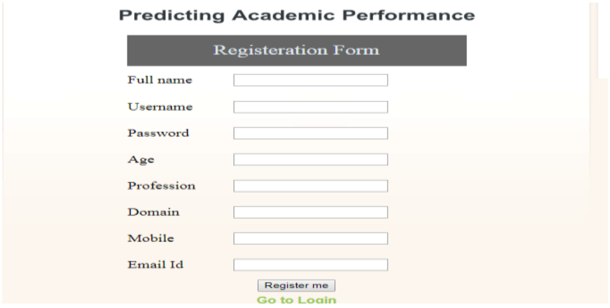


Fig 1 Basic block diagram

2.1 REGISTRATION AND LOGIN

Our study took place in the context of an undergraduate course for Computer Science students, on Web Applications Design (WAD). The instructional approach was project based learning (PBL) in which the students had to work collaboratively on various complex, challenging and authentic tasks, over extended periods of time; learning was organized around team projects, while the teacher played the role of a facilitator. More specifically, the students collaborated in teams of around 4 peers in order to build a complex web application of their choice. The project spanned over the whole semester and the evaluation took into account both the final product and the continuous collaborative work. Since PBL has a strong social component, the increasingly popular social media tools appear suitable for communication and collaboration support in PBL framework. Hence, we implemented our PBL scenario with the help of several social media tools (wiki, blog, and microblogging tool) integrated in our social learning environment, called eMUSE. More specifically, a blended learning approach was used consisting of weekly face-to-face meetings between each team and the instructor (for checking the project progress, providing feedback and answering questions), while students had to rely on the social media tools for the rest of the time, as a support for their communication and collaboration activities. In particular, Media Wiki was used for collaborative writing tasks, for gathering and organizing the team knowledge-base and resources, and for documenting the project. Blogger was used for reporting the progress of each project, similar to a "learning diary" in terms of publishing ideas and resources, as well as for providing feedback and solutions to peer problems. Each team had its own blog, butinter-team cooperation was encouraged as well. Twitter was meant to foster additional connections between peers and to encourage the posting of short news, announcements, questions, and status updates regarding each project. The eMUSE social learning platform provides an integration point for the social media tools, together with additional support for both students and teachers: basic administrative services, learner tracking and data visualizations, as well as evaluation and grading functionalities. eMUSE also offers data collection mechanisms, as detailed in the next subsection. Of course, students could choose to use additional communication channels for working on their projects, including face to face meetings, phone calls, chats, email, document sharing or other social media tools. Obviously, these could not be monitored by eMUSE; this means that a part of learner data may not be collected, which is a general limitation of learning analytics approaches based on student activity indicators. In order to mitigate this problem in our PBL scenario, we provided specific instructions to the learners at the beginning of the semester: students were clearly informed that their collaborative learning activity needs to be documented on the social media tools integrated in eMUSE, so that it can be assessed by the instructor. We therefore expect that a large part of the students' communication and collaboration activities indeed took place on the three recommended social media tools.



Predicting Academic Performance

Registeration Form

Full name

Username

Password

Age

Profession

Domain

Mobile

Email Id

Fig 2. Screenshot of registration page

TABLE I.

DATABASE SKELETON OF USER DETAILS

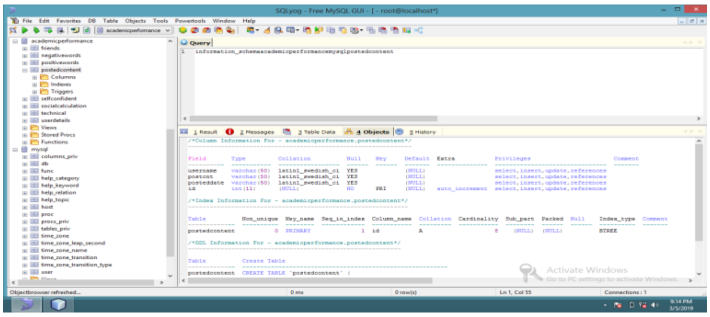
Field Name	Data type	Size
Full Name	Varchar	50
Username	Varchar	50
Password	Varchar	50
Profession	Varchar	50
Age	Number	10
Domain	Varchar	50
Mobile	Number	10
Email id	Varchar	50

2.2 DATA COLLECTION AND PREPROCESSING

The instructional scenario described above has been applied over 6 consecutive winter semesters (Year 1: 2010/2011 – Year 6: 2015/2016), with 4th year undergraduate students in Computer Science from the University of Craiova, Romania. Small improvements and refinements were made from one year to the consecutive one, based on students' feedback and instructor experience. A total of three forty three students, enrolled in the WAD course, participated in this study. All student actions on the three social media tools were monitored and recorded in the eMUSE platform. The system retrieves learner actions from each of the disparate Web 2.0 tools (by means of open APIs or Atom/RSS feeds) and stores them in a local database, together with a description and an associated timestamp. Thus, a total of almost 19000 social media contributions were recorded: 2609 blog posts and comments, 5470 tweets, 10895 wiki page revisions and file uploads.

Based on these actions, a set of fourteen numeric features were computed for each student:

- NO_BLOG_POSTS(the number of blog posts)
- NO_BLOG_COM(the number of blog comments)
- NO_ACTIVE_DAYS_BLOG(the number of days in which a student was active on the blog, i.e., wrote at least a post or a comment)
- NO_ACTIVE_DAYS_BLOG_POST (the number of days in which a student wrote at least a post on the blog)
- NO_ACTIVE_DAYS_BLOG_COM (the number of days in which a student wrote at least a comment on the blog)
- NO_TWEETS(the number of tweets)
- NO_ACTIVE_DAYS_TWITTER (the number of days in which a student was active on Twitter, i.e., posted at least a tweet)
- NO_WIKI_REV(the number of wiki page revisions)
- NO_WIKI_FILES(the number of files uploaded on the wiki)
- NO_ACTIVE_DAYS_WIKI (the number of days in which a student was active on the wiki, i.e., revised at least a page or uploaded at least a file)
- NO_ACTIVE_DAYS_WIKI_REV (the number of days in which a student revised at least a wiki page)
- NO_ACTIVE_DAYS_WIKI_FILES (the number of days in which a student uploaded at least a file on the wiki)



```
1  information_schemaacademicperformancemysqlpostedcontent
```

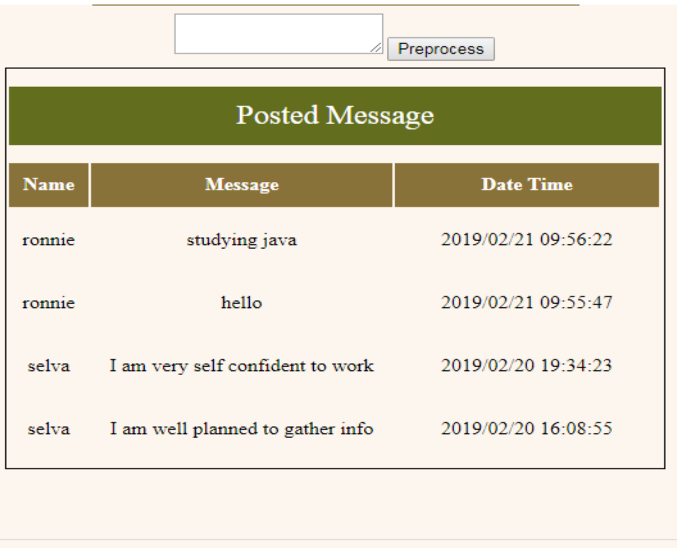
```
/*Column Information For - academicperformance.postedcontent*/
---------------------------------------------------------------

Field       Type         Collation          Null    Key     Default  Extra            Privileges                       Comment
----------  -----------  -----------------  ------  ------  -------  ---------------  -------------------------------  -------
username    varchar(50)  latin1_swedish_ci  YES             (NULL)                    select,insert,update,references
postcnt     varchar(50)  latin1_swedish_ci  YES             (NULL)                    select,insert,update,references
posteddate  varchar(50)  latin1_swedish_ci  YES             (NULL)                    select,insert,update,references
id          int(11)      (NULL)             NO      PRI     (NULL)   auto_increment   select,insert,update,references

/*Index Information For - academicperformance.postedcontent*/
-------------------------------------------------------------

Table          Non_unique  Key_name  Seq_in_index  Column_name  Collation  Cardinality  Sub_part  Packed  Null    Index_type  Comment
-------------  ----------  --------  ------------  -----------  ---------  -----------  --------  ------  ------  ----------  -------
postedcontent           0  PRIMARY              1  id           A                    8    (NULL)  (NULL)          BTREE

/*DDL Information For - academicperformance.postedcontent*/
-----------------------------------------------------------

Table          Create Table
-------------  ---------------------------------------
postedcontent  CREATE TABLE `postedcontent` (
```

Fig 4. Screenshot of the database

Students' performance at the end of the semester was assessed on a 1 to 10 scale, with 5 being the minimum passing grade; the evaluation took into consideration both the final project, as well as each student's continuous collaborative work. Our aim is to predict this final grade based on the above set of features, using the LMNNR algorithm, as described in the next section.

2.3 DATA ANALYSIS

In this section, we present the results obtained with the LMNNR algorithm, in comparison with those of the algorithms that provided the best results in, namely Random Forest with 100 trees and k-Nearest Neighbors, with k obtained by cross-validation and inverse-distance weighting of the neighbors. In, an additional setting for unimplemented in Weak was that mean absolute error was used when doing cross-validation. In this paper, we use mean squared error (MSE) instead, as we observed that it slightly improves the accuracy of the kNNmodel. For the analysis of individual years, we chose 3 neighbors and 1 prototype for LMNNR, because it is a simple model that also provides good results. A drawback of LMNNR training is that it sometimes converges into local optima. Several examples of convergence, plotting the value of the obtained objective function F against the corresponding MSE. It can be seen that the lowest value of F does lead to the lowest value of the MSE, but there are also many situations when the obtained MSE is not so good, even for low values of F. That is why we run the algorithm several times and retain the best results. Even if this requires additional training time, we consider that the quality of the obtained results, which are clearly better than those found by the other models, compensate for this inconvenience. In a previous work, an evolutionary algorithm was used for training, but the gradient-based method used here is much faster, although it needs multiple starting points. One of the main motivations for analyzing learning data is the ability to predict student behavior early in the course, therefore we also assessed how useful the trained model is for predicting student behavior in future years. Thus, we trained the algorithm on data from earlier years and tested its prediction performance on data from later years.



Posted Message

Name	Message	Date Time
ronnie	studying java	2019/02/21 09:56:22
ronnie	hello	2019/02/21 09:55:47
selva	I am very self confident to work	2019/02/20 19:34:23
selva	I am well planned to gather info	2019/02/20 16:08:55

Fig 5. Screenshot of the posted messages

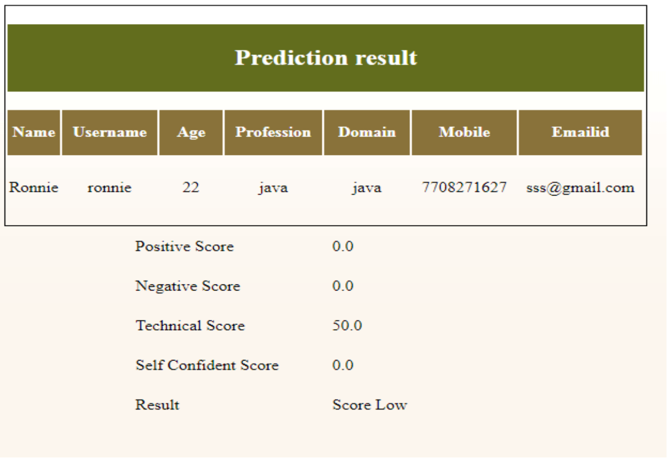
2.4 DATA PREDICTION

Academic Performance:

The function of the predictive system is to make predictions of the students' marks at the end of the term, based on the data collected from the interactive web. These data, selected and organized into features, are normalized and given to a machine learning algorithm as input. The predictive system classifies the student expected performance (measured as a mark in percentage) in three possible classes: high performance (expected mark > 80.5%), medium performance (57.5% < expected mark < 80.5%) or low performance (expected mark < 57.5%).The reason to split the output in three classes is to get an adequate performance out of the classification algorithm, adjusted to the size of the sample (the sample has only 336 students, so three equilibrated classes of 112 individuals are proposed). The balanced distribution of the students in the mark range explains the selected intervals for each class.

Social activities:

Only a number of previous studies have investigated the impact of social media activity on educational performance, despite the growing availability of such data and undisputed presence of these media in our daily lives. The majority of existing studies found a decrease in educational performance with increasing time spent on social media. However, not all studies confirm this result. In some studies, time spent on social media was found to be unrelated to academic performance or even a had positive effect on performance. There is a growing interest within the relationship between social interactions and educational performance. In the relevant literature there exist 2 dominant approaches. The first approach focuses on the relation between own performance and that of peers, based on a hypothesis of similarity in peer achievement. The similarity between pairs of individuals connected via social ties are attributed to various aspects: selection into friendships by similarity (i.e., homophily); influence by social peers (also known as peer effect); and correlate shocks (e.g., being exposed to the same teacher). As noted by the issue of separating these effects is inherently difficult. The second approach emphasizes the positive influence of getting a central position within the social network between students. The majority of results in the existing research which measure social networks are, however, based on self-reports and therefore subject to various biases that are in many ways mitigated by using smartphones to measure the social network. However, it ought to be noted that surveys and empiric studies usually live terribly totally different aspects of reality. For instance, in the case of assessing tie strengths, observational studies may be more accurate in quantifying duration and frequency variables of a relationship, while surveys can provide qualitative insights into depth and intimacy.



Prediction result

Name	Username	Age	Profession	Domain	Mobile	Emailid
Ronnie	ronnie	22	java	java	7708271627	sss@gmail.com

Positive Score	0.0
Negative Score	0.0
Technical Score	50.0
Self Confident Score	0.0
Result	Score Low

Fig 6. Final screenshot of the result

3. DISCUSSION AND CONCLUSION

The study has shown that students' actions on social media tools are good predictors of academic performance. The innovative LMNNR algorithm proved very suitable for our prediction problem, outperforming classic regression algorithms. Very good correlation coefficients were obtained and 85% of predictions were within only 1 point of the actual grade. From a pedagogical perspective, the results indicate that, as a general rule, a higher engagement with social media tools correlates with a higher final grade. This is in line with several previous studies, which found that online participation is a strong indicator of student performance and improves learning effectiveness. Nevertheless, there are also contradictory studies, which concluded that students learned equally well regardless of their level of online participation. At the same time, the body of literature specifically focused on students' active participation on social media is scarce, hence the novelty and added value of our study. It is worth mentioning that the performance of the generalized predictive model (from all six years combined) is slightly lower than the performance of each individual year model. This is in line with the findings in, which addresses the issue of aggregating trace data from different courses for creating one generalized model for academic success prediction. The differences in instructional conditions and technology use, even in the context of the same discipline, may influence the prediction of academic success; in addition, the individual differences of the students involved in the studies (e.g., metacognitive and motivational factors) may have an impact on the learning analytics results. Hence, the findings obtained in this paper need to be interpreted in the context of our specific instructional scenario; this is the case for most of the studies on academic performance prediction, which rely on data collected from one or few courses in the same discipline. Nevertheless, the model offers interesting insights into the learning process, in particular in the context of a PBL scenario supported by social media tools. We also showed that the model may be used with good results from one year to the other, although the specificities of each year may lead to slightly different comparing courses with different internal and external conditions. Hence, investigating the LMNNR algorithm performance on patterns of feature influence. Any attempt at generalizability needs to carefully consider the pedagogical and disciplinary context of the predictive model. Therefore, further studies are needed for student data collected from different courses and instructional scenarios is an interesting research direction. Furthermore, combining the predictive analytics approach proposed here with our previous work on social network analytics and discourse analytics could lead to a more comprehensive perspective on the social learning process and environment.

4. REFERENCES

1] Abdous, M., He, W., & Yen, C.J. (2012). Using data mining for predicting relationships between online question theme and final grade. Educational Technology & Society, 15(3), 77–88.

[2] Alstete, J.W., & Beutell, N.J. (2004). Performance indicators in online distance learning courses: a case study of management education. Quality Assurance in Education, 12(1), 6–14.

[3] Ardaiz-Villanueva, O., Nicuesa-Chacon, X., Brene-Artazcoz, O., Sanz de Acedo Lizarraga, M.L., & Sanz de Acedo Baquedano, M.T.(2011). Evaluation of computer tools for idea generation and team formation in project-based learning. Computers & Education, 56(3), 700–711.

[4] Assi, K.C., Labelle, H., & Cheriet, F. (2014). Modified large margin nearest neighbor metric learning for regression. IEEE Signal Processing Letters, 21(3), 292-296.

[5] Baker, R.S., & Inventado, P.S. (2014). Educational data mining and learning analytics. In Larusson J., White B. (Eds.), Learning Analytics, Springer, pp. 61-75.

[6] Barber, R., & Sharkey, M. (2012). Course correction: using analytics to predict course success. In Proceedings 2nd Int. Conf. on Learning Analytics and Knowledge (LAK 2012), pp. 259–262, ACM

7] Becheru, A., & Popescu, E. (2017). Using social network analysis to investigate students' collaboration patterns in eMUSE platform. In Proceedings ICSTCC 2017, pp. 266-271.

[8] Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data

BIOGRAPHIES

MAGESH M
“BE Student, Dept. of Computer Science and Engineering,
Sri Krishna College Of Technology-Coimbatore”

MANIKANDAN M
“BE Student, Dept. of Computer Science and Engineering,
Sri Krishna College Of Technology-Coimbatore”

MOHAMMED ARSATH I
“BE Student, Dept. of Computer Science and Engineering,
Sri Krishna College Of Technology-Coimbatore“

SUNITHA NANDHINI A
“Asst. Prof. Dept. Of Computer Science and Engineering, Sri Krishna College Of Technology Coimbatore“