USE OF DEEP LEARNING TECHNIQUES ON ONLINE LEARNING DATA

Author: Celia González Nespereira.
Thesis Advisors: Rebeca P. Díaz Redondo and Ana Fernández Vilas.
Affiliation: AtlantTIC Research Centre, Department of Telematics Engineering (University of Vigo)

MOTIVATION

- The online learning platforms are increasingly important in education. These platforms not only provide contents to students, they also store a lot of information about the students behavior.
- Data mining and analysis techniques allow studying large amount of data, with the objective of obtain more information about the users behavior and preferences.
- Our main purpose is to use these techniques to analyze the data coming from the e-learning platforms, with the objective of improve:
  - The learning process, improving the way in which students learn and detecting students learning problems in an early stage.
  - The teaching process, improving the teaching techniques and the course organization.

THESIS OBJECTIVES

- Studying of the state of the art of the different data mining algorithms, focusing on machine learning and deep learning techniques.
- Using data mining and analysis techniques to:
  - Study the different courses types and their characteristics, in order to detect what could be improved in the courses organization to increase the number of successful students.
  - Study the different types of students and their habits, in order to predict students' behavior and success.
- Analyzing one or more datasets coming fromfrom e-learning platforms. The data analysis is divided into two parts:
  - Collecting and processing the data to retrieve only the information that is useful in the study.
  - Application of deep learning techniques or any other type of data mining algorithms to analyzing the data.

RESEARCH PLAN

➢ Analyzing the state of the art in learning analytics.
➢ Analyzing the state of the art in data mining, focusing on deep learning and machine learning.
➢ Performing different studies to:
  ➢ Improve the learning process:
    ➢ Performing an exploratory analysis of a blended course, studying the relationship between the students' interaction with the e-learning platform and their final mark.
    ➢ Performing a prediction algorithm to detect the students that are in risk of failing a course.
  ➢ Improve the teaching process:
    ➢ Studying and classifying the different type of courses in base of their characteristics.
    ➢ Studying how the variation of some course characteristics (number of professors, variation in the number, dates and types of assignments, etc.) affects students' successful.
➢ Validating the studies explained above with real data coming from the University of Vigo. Specifically, we use a dataset from the e-learning platform faiTIC.
➢ Developing some plugins of the e-learning platform to put into practice the studies and algorithms developed in order to:
  ➢ Allow teachers obtain information about how to improve their lessons.
  ➢ Allow students know their performance in the courses and what things they should improve.

RESULTS

Academic year 14/15

Exploratory analysis of a blended course

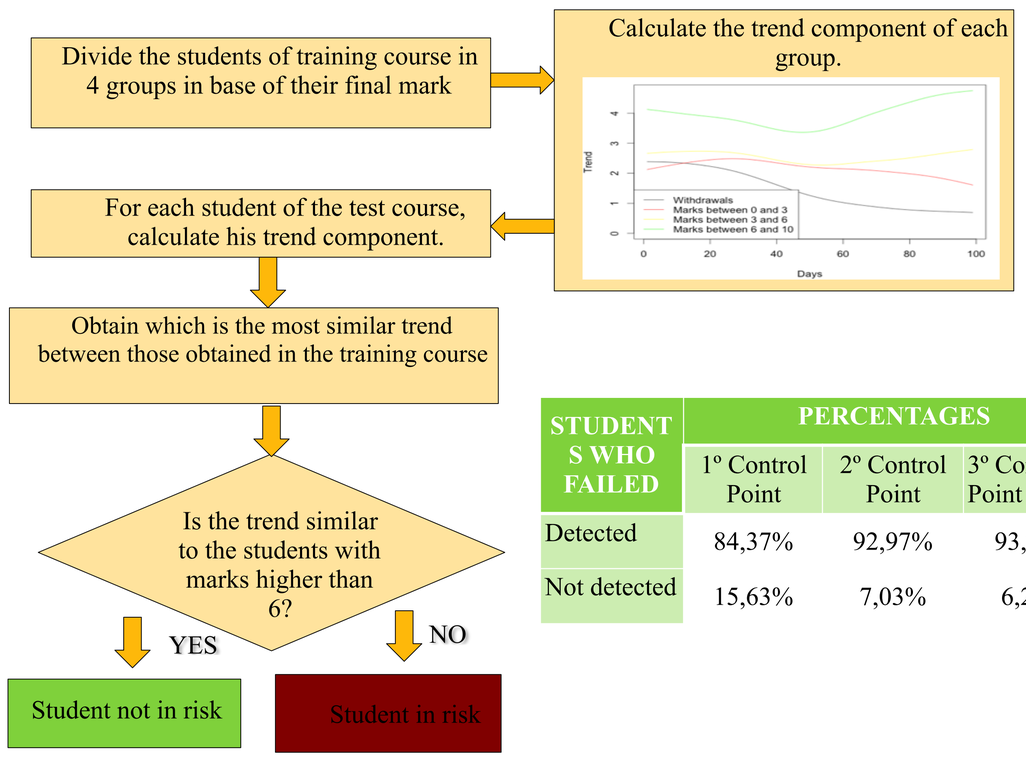
- We studied the data coming from the e-learning platform of one course of the Telecommunications Engineering of the UPublication:Publication:niversity of Vigo.
- Three main analysis:
  - **Correlation**: Use Pearson correlation [1] to look for relations between students' final grade and their different types of interactions with the e-learning platform → Positive values are obtained
  - **Clustering**: Divide the students in clusters [2] in base of their interactions with the e-learning platforms and represent the final mark of each cluster → The 1st cluster correspond to students that pass the course (in mean) and the 2nd to students that fail the course.
  - **Time series decomposition** [3]: To study the behavior of the students with highest and lowest grades → The trend component is a good indicator of students' success
- Publication:

"Is the LMS Access Frequency a Sign of Students' Success in Face-to-Face Higher Education?", published in the Technological Ecosystem for Enhancing Multiculturality 2014 (TEEM'14)

Published

Assessing the risk of failing

- We developed an algorithm to detect the students in risk of failing a course using time series and warn them and the professors about this situation.



STUDENTS WHO FAILED	PERCENTAGES		
	1º Control Point	2º Control Point	3º Control Point
Detected	84,37%	92,97%	93,75%
Not detected	15,63%	7,03%	6,25%

- Publication:

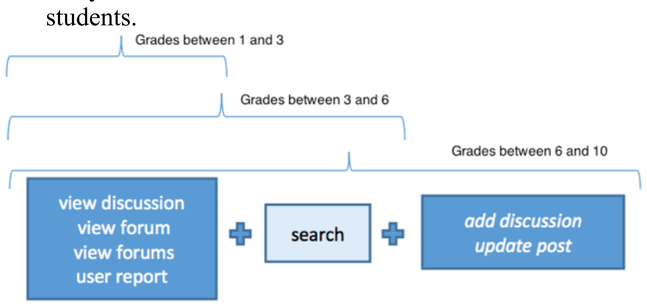
"Am I failing this course? Risk prediction for learning platforms", published in the Technological Ecosystem for Enhancing Multiculturality 2015 (TEEM'15)

Published

Academic year 15/16

Assessing the risk of failing

- Extension and improvement of the past year study.
  - Study of the relation between the different actions in the e-learning platform (faiTIC) and the final grade of the students.



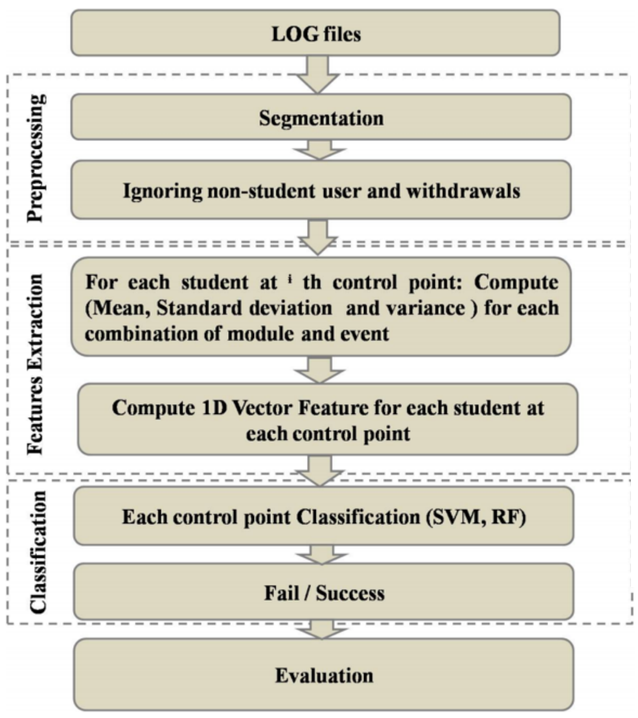
- Publication:

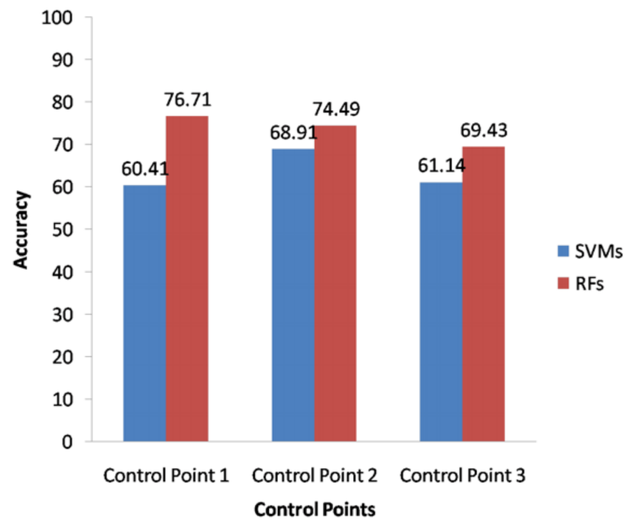
"Is it possible to predict if you will approve the academic year?"

Looking for a journal

Applying machine learning techniques to students classification

- This study compares the performance of two different machine learning techniques to predict if the students will pass/ fail the subject in base of their interactions with the e-learning platforms.
- Machine learning techniques used:
  - Support Vector Machines (SVMs) [4]
  - Random Forest (RF) [5]





- Publication:

"Machine Learning Classification Approach for Predicting Students Performance in Blended Learning" published in International Conference on Advanced Intelligent Systems and Informatics 2015 (AISI 2015)

Published

NEXT YEAR PLANNING

- Improving our risk detection algorithm:
  - Using other machine learning or deep learning techniques.
  - Detecting the control points and the grade thresholds dynamically.
- Analyzing the different courses characteristics, trying to detect what could professors improve to obtain better students' results.
- Publishing our approaches in an international journal.

REFERENCES

[1] Hauke, J. and Kossowski, T. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones geographicae, 30, 2.
[2] Murtagh, F. and Legendre, P. 2011. Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. *arXiv preprint arXiv:1111.6285*.
[3] E. B. Dagum, "Time series modeling and decomposition," Statistica, no. 4, 2010.
[4] Q. Wu and D.-X. Zhou, "Analysis of support vector machine classification", *J. Comput. Anal. Appl.*, vol. 8, pp. 99–119, 2006.
[5] Vrushali Y Kulkarni and Pradeep K Sinha, "Efficient Learning of Random Forest Classifier using Disjoint Partitioning Approach", *Proceedings of the World Congress on Engineering*. Vol. 2. 2013.