

Identificação Automática de Abandono em Cursos de Longa Duração

Direitos Humanos, Multiculturalismo e C&T

Daniel Verdelli - 92387

Ricardo Manhães Savii - 92482

Instituto de Ciência e Tecnologia - Universidade Federal de São Paulo

10/10/2016

Tema

Aplicação de conhecimentos sobre Aprendizado de Máquina (Machine Learning) para a criação de um sistema para suporte ao setor do NAE (Núcleo de Apoio ao Estudante) com a finalidade de identificação automática de alunos com possibilidade de abandono de curso.

Resumo

Aprendizado de máquina (AM), em linhas gerais, é uma etapa de mineração de dados que consiste na criação de técnicas computacionais que aprendem com o tempo.

Com a aplicação dessas técnicas, é possível desenvolver um algoritmo que analisa dados referentes a evasão de alunos em instituições de ensino superior, com o objetivo de oferecer estratégias melhores para a retenção desses alunos.

Através de uma parceria com o Núcleo de apoio ao estudante (NAE), o projeto visa a implementação de uma aplicação baseada em AM que possa auxiliar o NAE no oferecimento de maior suporte aos possíveis alunos em condições de abandono do curso.

Palavras-chaves: Direitos humanos, apoio ao estudante, aprendizado de máquina

Introdução a técnicas de Aprendizado de Máquinas:

Aprendizado de Máquina (AM) é uma vasta classe de algoritmos que podem ser utilizados para “aprender”. AM é utilizado na indústria em usos como detecção de fraude, resultados em buscadores web, marketing em tempo real para sites e aplicativos, análise de sentidos em textos, nível de crédito e ofertas, predição em falhas de equipamentos, modelos de precificação, detecção de invasão em redes, reconhecimento de padrões em imagens, filtros de spam, etc

São divididas quanto a sua forma de aprender (aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço):

- **Aprendizado supervisionado:** Fundamentado na indução, é dado causas, efeitos e o algoritmo induz uma regra.
- **Aprendizado não supervisionado:** Relacionado a descrição dos dados, podendo complementar uma técnica de aprendizado supervisionado.
- **Aprendizado por reforço:** Através da recompensa tardia (ex: Jogo de Xadrez, onde o algoritmo faz diversas decisões para só ao final saber o resultado do jogo).

Introdução a técnicas de Aprendizado de Máquinas:

Também são classificadas em dois tipos, dependendo do problema a ser resolvido:

- **Problemas de classificação:** Prever uma classe para certas combinações de valores (ex: Classes alunos que abandonaram e que não abandonaram o curso, notas e informações pessoais são as combinações de valores);
- **Problemas de regressão:** São problemas que prevêm valores contínuos (ex: prever a altura de uma pessoa através das medidas de suas mãos).

Data

```
In [13]: # TODO: Calculate number of students
n_students = student_data.shape[0]

# TODO: Calculate number of features
n_features = student_data.shape[1]

# TODO: Calculate passing students
n_passed = student_data[student_data.passed == 'yes'].shape[0]

# TODO: Calculate failing students
n_failed = student_data[student_data.passed == 'no'].shape[0]

# TODO: Calculate graduation rate
grad_rate = n_passed/n_students

# Print the results
print ("Total number of students: {}".format(n_students))
print ("Number of features: {}".format(n_features))
print ("Number of students who passed: {}".format(n_passed))
print ("Number of students who failed: {}".format(n_failed))
print ("Graduation rate of the class: {:.2f}%".format(grad_rate))
```

```
Total number of students: 395
Number of features: 31
Number of students who passed: 265
Number of students who failed: 130
Graduation rate of the class: 0.67%
```

Data

- **`school`** : student's school (binary: "GP" or "MS")
- **`sex`** : student's sex (binary: "F" - female or "M" - male)
- **`age`** : student's age (numeric: from 15 to 22)
- **`address`** : student's home address type (binary: "U" - urban or "R" - rural)
- **`famsize`** : family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
- **`Pstatus`** : parent's cohabitation status (binary: "T" - living together or "A" - apart)
- **`Medu`** : mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- **`Fedu`** : father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- **`Mjob`** : mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
- **`Fjob`** : father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
- **`reason`** : reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
- **`guardian`** : student's guardian (nominal: "mother", "father" or "other")
- **`traveltime`** : home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
- **`studytime`** : weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
- **`failures`** : number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
- **`schoolsup`** : extra educational support (binary: yes or no)
- **`famsup`** : family educational support (binary: yes or no)
- **`paid`** : extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
- **`activities`** : extra-curricular activities (binary: yes or no)
- **`nursery`** : attended nursery school (binary: yes or no)
- **`higher`** : wants to take higher education (binary: yes or no)
- **`internet`** : Internet access at home (binary: yes or no)
- **`romantic`** : with a romantic relationship (binary: yes or no)
- **`famrel`** : quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
- **`freetime`** : free time after school (numeric: from 1 - very low to 5 - very high)
- **`goout`** : going out with friends (numeric: from 1 - very low to 5 - very high)
- **`Dalc`** : workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
- **`Walc`** : weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
- **`health`** : current health status (numeric: from 1 - very bad to 5 - very good)
- **`absences`** : number of school absences (numeric: from 0 to 93)
- **`passed`** : did the student pass the final exam (binary: yes or no)

Justificativa

Motivações:

- Alunos que não terminam a graduação ocasionam um prejuízo financeiro e emocional para suas famílias, instituições e sociedade;
- Melhoria na eficiência com a finalidade do aperfeiçoamento no oferecimento de um serviço de ensino público de qualidade, procurando auxiliar os alunos com dificuldade.

Obrigado!

Identificação Automática de Abandono em Cursos de Longa Duração

Direitos Humanos, Multiculturalismo e C&T

Daniel Verdelli - 92387

Ricardo Manhães Savii - 92482

Instituto de Ciência e Tecnologia - Universidade Federal de São Paulo

10/10/2016

Referências bibliográficas

- [1] “NAE quem somos,” <http://www.sjc.unifesp.br/nae/quem-somos>, acesso em: 18-09-2016.
- [2] “PRAE Regimento,” <https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwivt-TlJrPAhUJiZAKHe3zBM4QFggeMAA&url=https%3A%2F%2Fwww.unifesp.br%2Ffreitoria%2Fprae%2Finstitucional%2Fdocumentos%2Fregimento-prae%3Fdownload%3D2%3Aregimento-prae&usg=AFQjCNEt6yoWvB1GOG3LpQuQBIJQNyCYIg&sig2=EeRj7fGJvIXvf5Y3pVKaA>, acesso em: 18-09-2016.
- [3] “Diário Oficial da União de 20/07/2010, p. 5 - Decreto nº 7.234 de 19 de julho de 2010,” http://www.planalto.gov.br/ccivil_03/ato2007-2010/2010/decreto/d7234.htm, acesso em: 18-09-2016. [4] “MINISTERIO DA EDUCAC, ´ AO. Programa Nacional de Assist ~ encia Estudantil.” ^ http://portal.mec.gov.br/arquivos/pdf/portaria_pnaes.pdf, acesso em: 18-09-2016. [5] N. Bobbio, Era dos direitos. Elsevier Brasil, 2004.
- [6] F. K. Comparato, A afirmac,ao hist ~ orica dos direitos humanos ´ . Saraiva, 2001.
- [7] B. d. S. Santos, “Uma concepcao multicultural de direitos humanos,” ~ Lua nova, vol. 39, pp. 105–124, 1997.
- [8] B. de Sousa Santos, Reconhecer para libertar: os caminhos do cosmopolitismo multicultural. Civilizacao brasileira, 2003, ~ vol. 3. 7
- [9] N. T. Casals, Group rights as human rights: a liberal approach to multiculturalism. Springer Science & Business Media, 2006, vol. 75.
- [10] E. Alpaydin, Introduction to machine learning. MIT press, 2014.
- [11] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, “API design for machine learning software: experiences from the scikit-learn project,” in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 2013, pp. 108–122.
- [12] D. Delen, “A comparative analysis of machine learning techniques for student retention management,” Decision Support Systems, vol. 49, no. 4, pp. 498–506, 2010.
- [13] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West, “Predicting student dropout in higher education,” arXiv preprint arXiv:1606.06364, 2016.