

Dropout Prediction: A Systematic Literature Review

SLR Dropout

Pedro Sobreiro, José Garcia Alonso, Javier Berrocal

Webinar, 26 June 2020

University of Extremadura

pdealexa@alumnos.unex.es

Summary

Research goals

Introduction

Methodology

Results

Conclusion

Research goals

What is the current state of machine learning research studies to predict dropout in contractual settings?

Introduction

Data mining

- Customer analysis is fundamental to develop business and marketing intelligence (Sheth, Mittal, & Newman, 1998), supporting the understanding of historical data identifying trends and patterns (Berry & Linoff, 2004)
- This process is also known as data mining, the extraction of knowledge from data (Han & Kamber, 2006)

Allen, I. E., & Seaman, J. (2011). Going the Distance: Online Education in the United States, 2011. Em Sloan Consortium (NJ1). Obtido de <https://eric.ed.gov/?id=ED529948>

Martinho, D. S. (2014). O Ensino Online nas Instituições de Ensino Superior Privado. As perspectivas: Docente e discente e as implicações na tomada de decisão institucional.

McCarthy, S. A. (2009). Online Learning as a Strategic Asset. Volume I: A Resource for Campus Leaders. A Report on the Online Education Benchmarking Study Conducted by the APLU-Sloan National Commission on Online Learning. Em Association of Public and Land-grant Universities. Obtido de <https://eric.ed.gov/?id=ED517308>

Methodology

- Was developed a Systematic Literature Review (SLR) developed in three stages (Kitchenham & Charters, 2007): Plan, Conduct and Report;
- Plan: definition of the research need, identification of the research questions and the development of the review protocol;
- Conduct: research identification, study selections, quality assessment, data extraction, finishing with the data synthesis;
- Report: stage that develops the activity report review;

Kitchenham, B., & Charters, S. (2007). Guidelines for performing structural literature reviews in software engineering (pp. 1–26) [Joint technical report]. Australia: Keele Univ., and Empirical Software Eng., Nat'l ICT.

Systematic Literature Review Phases

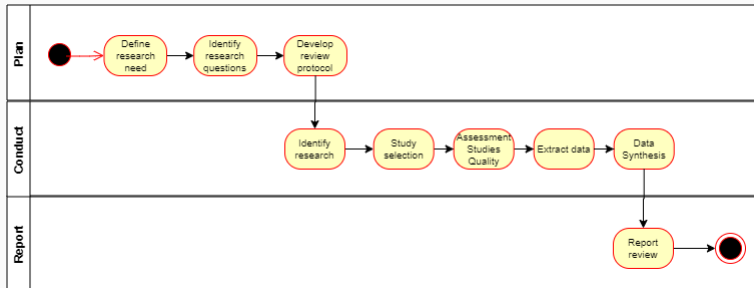


Figure 1: SLR phases based on Kitchenham and Charter (2007)

What is the current state of machine learning research studies to predict dropout in contractual settings? Based in this question were identified the following questions:

- RQ1: What studies have been published?;
- RQ2: Which algorithms have been used to predict the dropout?
- RQ3: What are the more relevant features related to predicting customer dropout?
- RQ4: When the dropout occurs?
- RQ5: What is the accuracy of the machine learning algorithms to predict dropout?

Population, Intervention, Comparison, Outcomes and Context

Table 1: PICOC criteria

PICOC	Description
Population	Research papers about dropout with contractual settings
Intervention	Machine learning algorithms to predict dropout
Comparison	Studies addressing machine learning algorithms to predict dropout
Outcome	Synthesis identifying research questions, gaps in the research domain and also best practices identified
Context	Academia and industry

Note: Context (PICOC) as suggested Kitchenham and Charters (2007) and proposed by Petticrew and Roberts (Petticrew & Roberts, 2006) to support the development of the search string.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing structural literature reviews in software engineering (pp. 1–26) [Joint technical report]. Australia: Keele Univ., and Empirical Software Eng., Nat I ICT.

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Malden, MA ; Oxford: Blackwell Pub.

- Search string: ((“customer dropout”) OR (“customer churn”) AND “machine learning” AND (“contractual” OR “membership”));
- Applied to the title, abstract, and keywords in the search period between January 2000 and June 2020
- The exclusion criteria were Books, Non-English articles, patents, and thesis
- The selection process was developed using ASReview (ASReview Core Development Team, 2019) creating a dataset of the identified articles, providing five relevant papers and five irrelevant papers to train Machine Learning model Naïve Bayes;

Kitchenham, B., & Charters, S. (2007). Guidelines for performing structural literature reviews in software engineering (pp. 1–26) [Joint technical report]. Australia: Keele Univ., and Empirical Software Eng., Nat I ICT.

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Malden, MA ; Oxford: Blackwell Pub.

Results

Results

- Os resultados revelam que a maioria dos professores que responderam ao questionário são do sexo masculino (56%)
- Questionados sobre a avaliação da qualidade do ensino online versus ensino presencial 24% considera que é inferior, 44% refere que não existem diferenças e 26% por vezes é superior
- Quando confrontados com gosto pelo ensino online em relação ao seu gosto pelo ensino presencial 6% considera inferior, 29% por vezes inferior, 21% não apresenta preferência, 38% acha que é por vezes superior e 6% considera superior
- Em relação à disponibilidade para o ensino online em relação à sua disponibilidade para o ensino presencial 18% apresenta mais disponibilidade para o ensino online, 41% considera a sua disponibilidade por vezes superior, 35% é indiferente e 6% considera ter menos disponibilidade.

- Quando questionados sobre os aspectos tecnológicos, 100% apresenta experiência de mais de 6 anos na utilização processadores de texto e email. 97% PowerPoint e motores de busca;
- Em relação a ferramentas de videoconferência verifica-se que 56% dos respondentes têm mais de 6 anos de experiência com o Zoom, 21% de 3 a 6 anos e 24% até 3 anos de experiência
- Na utilização de ferramentas para chat 76% mais de 6 anos, 15% de 3 a 6 anos, 6% menos de 3 e 3% não apresentava nenhuma experiência
- Na utilização do Moodle 53% apresentava mais de 6 anos de experiência, 29% 3 a 6 anos e 18% menos de 3 anos

- A análise de correlação mostrou relações com elevado significado estatístico entre a qualidade do ensino online versus ensino presencial* e: (1) gosto em ensinar presencialmente ($r_s=.51, p<.01$); (2) disponibilidade para o ensino online ($r_s=.44, p<.01$); (3) preferência por testes online ($r_s=.46, p<.01$); (4) preferência com apresentações orais ($r_s=.40, p<.05$); (5) experiência com chat ($r_s=.36, p<.05$);
- Não se identificaram relações significativas entre qualidade do ensino online e: (1) idade, (2) anos docência, (2) satisfação com Moodle, Zoom e Google Docs;

* Escala de likert com as opções: 1-Inferior; 2- Por vezes inferior; 3 – Sem diferenças significativas ; 4-Por vezes superior; 5-Superior

Conclusion

- A dimensão da amostra não permite a generalização dos resultados obtidos pelo que o trabalho de campo vai continuar com o alargamento da amostra dos respondentes para outras instituições do ensino superior;
- Pretende-se ainda investigar outros aspetos, nomeadamente como é que os professores perspetivam o ensino online para o desenvolvimento dos estudantes, da sua carreira profissional e das instituições de ensino superior;

Obrigado!

Pedro Sobreiro - IPSantarém

Domingos Martinho - ISLA Santarém

Marco Tereso - ISLA Santarém