

Após a obtenção da acurácia do classificador J48 para cada um dos subconjuntos de atributos, foram desprezados os atributos selecionados em que a acurácia não obteve significância estatística, quando comparados ao melhor resultado obtido.

Para se obter os melhores atributos foi utilizado o seguinte procedimento: 1) ordenou-se de forma decrescente a frequência em que o atributo foi selecionado pelos algoritmos WrapperSubsetEval e CfsSubsetEval; 2) ordenou-se de forma crescente pela posição média que o atributo foi ordenado pelos algoritmos que utilizam o ranqueamento. O resultado dessa seleção está indicado na Tabela 5.

4.5 Análise dos resultados

Preliminarmente à aplicação dos algoritmos de mineração, foi verificado que aproximadamente 80% da evasão de curso ocorre até o 3º período, independente se o curso possui duração de 6, 8 ou 10 períodos.

Nos três *datasets* utilizados a abordagem *wrapper* obteve a melhor acurácia, com resultados entre 83 e 87%.

Com os resultados apresentados na Tabela 5 pode-se concluir que a criação de atributos contribuiu para a tarefa de mineração de dados: dos dez melhores atributos classificados, cinco deles são novos atributos. Estes novos atributos podem facilitar a tarefa de análise da evasão com o objetivo de alertar os gestores para o problema e para a necessidade de buscar soluções pra reduzir a evasão. O atributo de dificuldade média das disciplinas cursadas pelo aluno revelou-se uma boa medida de prognóstico de desempenho do aluno, uma vez que possui um componente coletivo em sua avaliação.

5 Conclusão e Trabalhos Futuros

Esta pesquisa apresentou um método de seleção dos melhores atributos aplicados em algoritmos de classificação para a previsão da evasão em cursos de graduação, utilizando a criação de novos atributos e a seleção dos melhores atributos previsores. O algoritmo de seleção de atributos que apresentou os melhores

resultados para a acurácia foi o WrapperSubsetEval, que utiliza a abordagem *wrapper*, empregando o classificador de árvore de decisão J48. Este resultado é consistente com o indicado em (HALL; HOLMES, 2003), em que a abordagem *wrapper* também aparece com os melhores resultados.

Dos seis melhores atributos selecionados para a tarefa de mineração, cinco deles foram novos atributos, indicando a sua contribuição na tarefa de previsão da evasão. A criação do atributo “dificuldade média das disciplinas cursadas pelo aluno” melhorou a acurácia dos algoritmos de classificação, agregando um componente coletivo (percentual dos alunos aprovados na disciplina) no desempenho individual do aluno.

Este trabalho apresenta uma extensão do artigo (JÚNIOR; KAESTNER; NORONHA, 2016), e buscou fazer um melhor detalhamento da etapa de seleção dos melhores atributos, subdividindo-a em 4 etapas: Seleção do Subconjunto de Atributos, Execução do Classificador em Cada Subconjunto de Atributos, Exclusão dos Subconjuntos sem Significância Estatística e Ordenação e Escolha dos Melhores Atributos.

Com o método proposto, espera-se proporcionar aos gestores educacionais indicadores e/ou um conjunto de regras que permitam avaliar a possibilidade da evasão de cada aluno. Como trabalhos futuros pretende-se aplicar o método em outras amostras e também avaliar a aplicação da classificação sensível ao custo.

Referências

- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 19, n. 02, p. 03, 2011.
- BORGES, V. A.; NOGUEIRA, B. M.; BARBOSA, E. F. Uma análise exploratória de tópicos de pesquisa emergentes em informática na educação. *Revista Brasileira de Informática na Educação*, v. 23, n. 01, p. 85, 2015.

- BRAGA, M. M.; PEIXOTO, M. D. C. L.; BOGUTCHI, T. F. A evasão no ensino superior brasileiro: O caso da ufmg. *Avaliação*, Unicamp, v. 8, n. 3, p. 161–189, 2003.
- CHAU, V. T. N.; PHUNG, N. H. Imbalanced educational data classification: An effective approach with resampling and random forest. In: IEEE. *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2013 IEEE RIVF International Conference on*. [S.l.], 2013. p. 135–140.
- CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002.
- DEKKER, G. W.; PECHENIZKIY, M.; VLESHOUWERS, J. M. Predicting students drop out: A case study. *International Working Group on Educational Data Mining*, ERIC, 2009.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996.
- GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. In: *Anais do Simpósio Brasileiro de Informática na Educação*. [S.l.: s.n.], 2012. v. 23, n. 1.
- GOTTARDO, E.; KAESTNER, C. A. A.; NORONHA, R. V. Estimativa de desempenho acadêmico de estudantes: Análise da aplicação de técnicas de mineração de dados em cursos a distância. *Revista Brasileira de Informática na Educação*, v. 22, n. 01, p. 45, 2014.
- HALL, M. et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009.
- HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. *Knowledge and Data Engineering, IEEE Transactions on*, IEEE, v. 15, n. 6, p. 1437–1447, 2003.
- HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3rd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123814790, 9780123814791.
- JOHN, G. H. et al. Irrelevant features and the subset selection problem. In: *ICML*. [S.l.: s.n.], 1994. v. 94, p. 121–129.
- JÚNIOR, J. G. O.; KAESTNER, C.; NORONHA, R. V. Criação e seleção de atributos aplicados na previsão da evasão de curso em alunos de graduação. *Anais do Computer on the Beach*, p. 061–070, 2016.
- KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. *Artificial intelligence*, Elsevier, v. 97, n. 1, p. 273–324, 1997.
- KOTSIANTIS, S. B.; PIERRAKEAS, C.; PINTELAS, P. E. Preventing student dropout in distance learning using machine learning techniques. In: SPRINGER. *Knowledge-Based Intelligent Information and Engineering Systems*. [S.l.], 2003. p. 267–274.
- MANHÃES, L. M. B. et al. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Anais do Simpósio Brasileiro de Informática na Educação*, 2011.
- MÁRQUEZ-VERA, C.; MORALES, C. R.; SOTO, S. V. Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, IEEE, v. 8, n. 1, p. 7–14, 2013.
- RIGO, S. J.; CAZELLA, S. C.; CAMBRUZZI, W. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In: *Anais do Workshop de Desafios da Computação Aplicada à Educação*. [S.l.: s.n.], 2012. p. 168–177.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. ed. San Francisco,

CA, USA: Morgan Kaufmann Publishers Inc.,
2011. ISBN 0123748569, 9780123748560.