

Entendimento dos Dados

24 de novembro de 2016

Coleta de Dados

Dados do SIGRA (Sistema de Graduação), obtidos em parceria com a UnB.

Pesquisas prévias utilizaram o desempenho acadêmico do aluno [1].

Outros estudos apontam a necessidade de considerar também dados sociais [2].

Como os dados contém tais informações, devem ser suficientes.

Divisão dos Dados

Utilizar dados de 2000 até 2015.

Dados de treino: alunos que entraram depois de 2000 e saíram antes de 2010.

Dados de teste: alunos que entraram depois de 2010 e se formaram até 2015.

Dados Brutos a Serem Utilizados

Dados sociais: sexo, idade, residente do local, cotista, tipo da escola, raça, curso, forma de ingresso.

Desempenho acadêmico: IRA.

Dados Derivados a Serem Utilizados

coeficiente de melhora acadêmica (razão entre notas do primeiro semestre e notas do segundo semestre)

razão entre número de disciplinas reprovadas por semestre, trancadas por semestre e aprovadas por semestre frente ao total de disciplinas do semestre

razão entre disciplinas cursadas por semestre e quantidade de disciplinas do curso

razão entre disciplinas obrigatórias cursadas por semestre e quantidade de disciplinas obrigatórias do curso (*)

flag indicando aprovação ou não na disciplina mais difícil do semestre.

Propriedades Gerais dos Dados

São 7 features relacionados às condições sociais dos estudantes
e 11 features relacionados ao desempenho acadêmico

São x entradas na base de dados

Propriedades Gerais dos Dados - Sexo

Propriedades Gerais dos Dados - Idade ao entrar

Propriedades Gerais dos Dados - Residente do Local

Propriedades Gerais dos Dados - Cotista

Propriedades Gerais dos Dados - Tipo da Escola

Propriedades Gerais dos Dados - Raça

Propriedades Gerais dos Dados - Curso

Propriedades Gerais dos Dados - Forma de Ingresso

Referências



da Silva, H. R. B., and Adeodato, P. J. L.

A data mining approach for preventing undergraduate students retention.

In *The 2012 International Joint Conference on Neural Networks (IJCNN)* (2012), IEEE, pp. 1–8.



Yu, C. H., DiGangi, S., Jannasch-Pennell, A., and Kaprolet, C.

A data mining approach for identifying predictors of student retention from sophomore to junior year.

Journal of Data Science 8, 2 (2010), 307–325.