

APRENDIZADO ESTATÍSTICO EM CIÊNCIA DE DADOS

XXXVIII Encontro de Iniciação Científica

Alberto Rodrigues Ferreira, Juvêncio Santos Nobre, Rafael Braz Azevedo Farias

Foram realizadas várias análises referente ao banco de dados dos alunos da Universidade Federal do Ceará referente aos anos de 2010 a 2017, foram utilizadas as seguintes variáveis para o estudo: Raça, Idade, Sexo e Turno para relacionar ou prever a variável Evasão, em questão da variável Evasão foi realizado o seguinte procedimento, foram estudados somente pessoas que ingressaram no ano de 2010, mas temos as suas variáveis até o ano de 2017, no caso se a pessoa evadir de alguma forma até o ano de 2017 essa variável será considerada como Evadiu, caso contrário, Não Evadiu. Primeiramente foi realizado uma análise descritiva a fim de entender o comportamento dos dados. Foram realizados várias análises como regras de associação com o algoritmo Apriori que têm o objetivo de encontrar subgrupos frequentes relacionados à evasão, também foram realizados diversos algoritmos de aprendizagem de máquina a fim de prever se uma pessoa vai evadir, com precisões (excelentes), foram realizados os seguintes algoritmos: Árvore de decisão, Floresta Aleatória, Boosting, K-Vizinhos mais próximos, Máquinas de Vetores Suporte e Redes Neurais. Nesse projeto um dos principais objetivos é prevenir um dos maiores problemas que a Universidade Federal do Ceará enfrenta que é a evasão bem como os fatores mais importantes que levam a este acontecimento, isso ocasiona diversos prejuízos como o não retorno do investimento realizado nesse indivíduo, para que quando identifiquemos essas pessoas serem

Palavras-chave: Aprendizagem de máquina. Regras de associação. Ciência de dados. Estatística.