

Predição de evasão de cursos

Abordado por aprendizado de máquina

Apresentando Problema

- **A evasão de cursos é algo que preocupa a gestão de organizações de ensino, prejudica o desenvolvimento social e tecnológico de um país e fragiliza a economia a um longo prazo [1,2].**
- **Desta forma, desenvolver-se ferramentas de auxílio ao ensino e prevenção de evasão é de extrema importância e benéfico a todos, com impactos a curto e longo prazo.**

Evasão: USP

INSTITUTO	PORCENTAGEM DE EVASÃO
Univesp*	32%
EESC	12,20%
ICMC	26,95%
IFSC	42,86%
IME	62,87%

Tabela 1: Apresentação da porcentagem de evasão em alguns institutos de ensino da Universidade de São Paulo (USP) - 2009.

* Universidade Virtual do Estado de São Paulo EAD

Introdução à Abordagem Proposta

- Foi-se adaptado e aplicado um processo de mineração de dados para abordar o problema.
- Uma análise de dados relevante para otimização do processo de aprendizado foi desenvolvido.
- O método foi testado simulando um cenário real, considerando fluxo de dados e entradas diversas para o método de aprendizado.

Diagrama do processo



Sobre o banco de dados

O curso *Big Data Analytics da Future Learn* disponibiliza um banco de dados sobre as submissões de atividades de alunos cadastrados em seus cursos. Mais especificamente, o curso de análise compreende quatro módulos online, cada um apresentando e analisando um elemento diferente de aplicação e teoria de big data, sendo ministrado pela Universidade de Tecnologia de Queensland. Os criadores pensando na mesma análise exposta anteriormente, geram logs de acessos dos alunos matriculados ao longo das 12 semanas do programa. Este banco de logs é de domínio público desde que não proporcione artigos acadêmicos ou de caráter lucrativo no formato CSV.

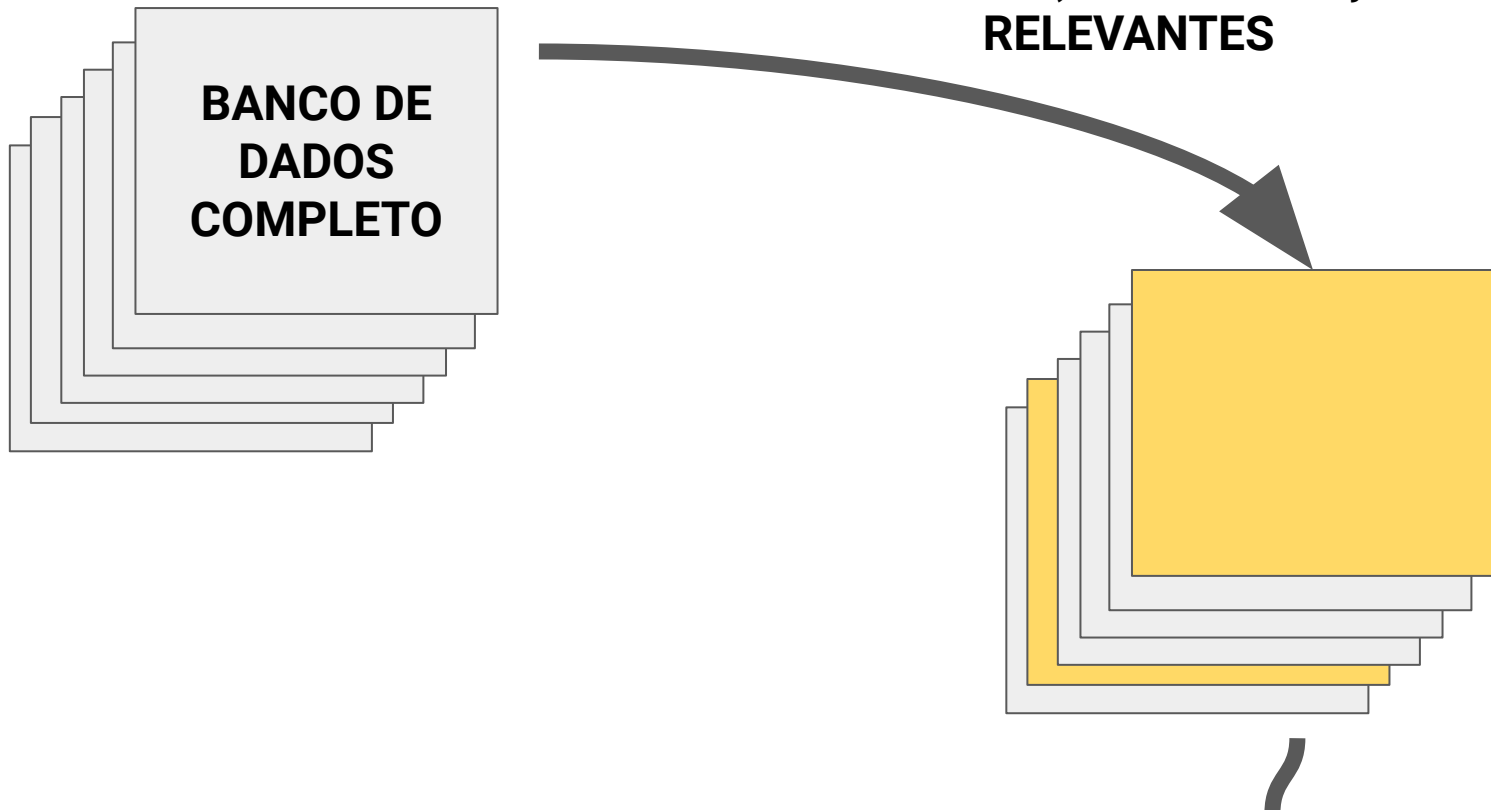
Pré-Processamento

O Pré-Processamento definido é composto por 6 passos principais:

1. Identificação dos atributos importantes para análise
2. Recolher informações de cadastro importantes
3. Marcar alunos desistentes e não desistentes
4. Criar relacionamento Cadastro-Submissão
5. Separar alunos desistentes e não desistentes
6. Transformar dados em experiência

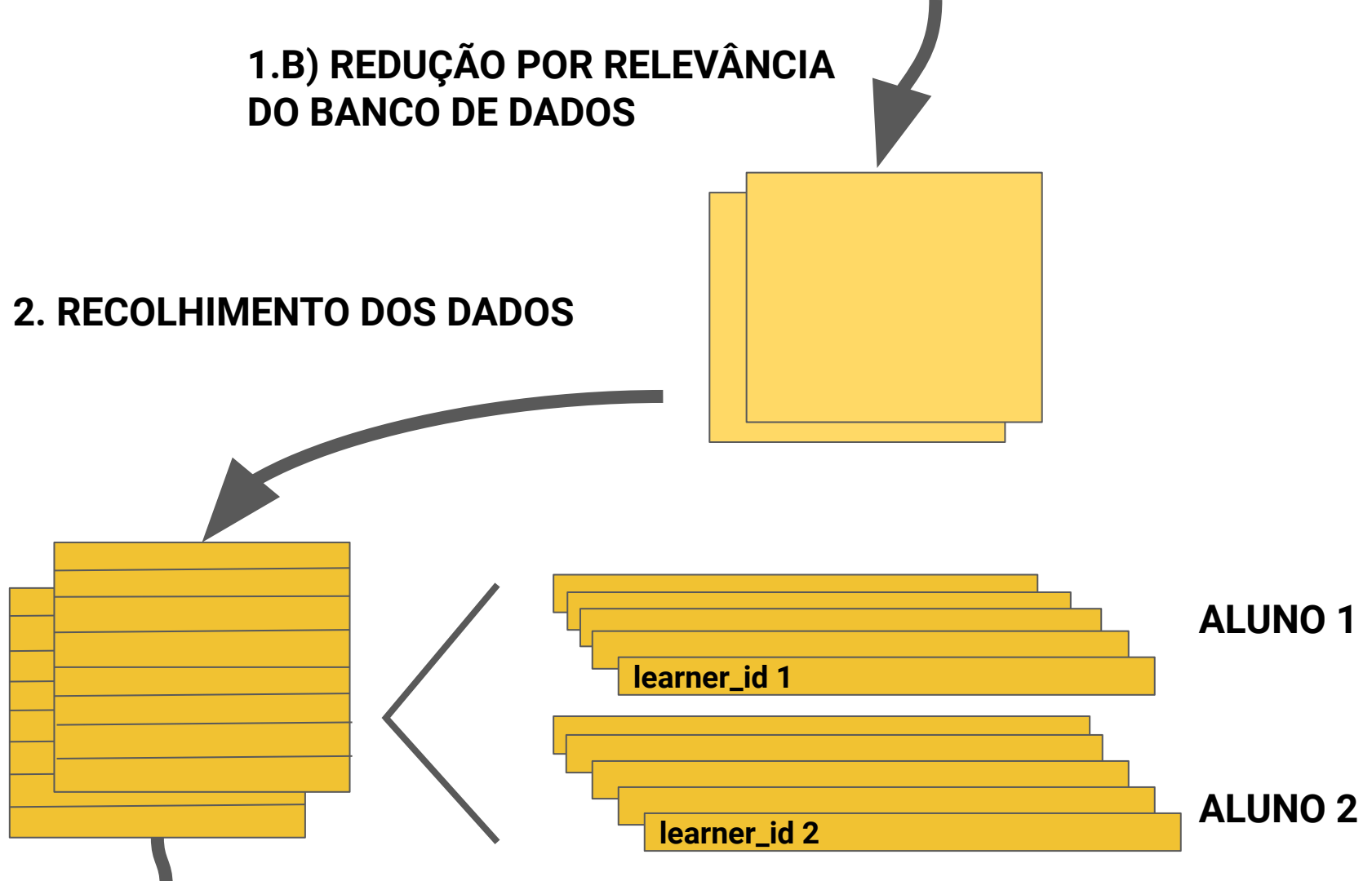
Pré-Processamento

1.A) IDENTIFICAÇÃO DE ATRIBUTOS RELEVANTES

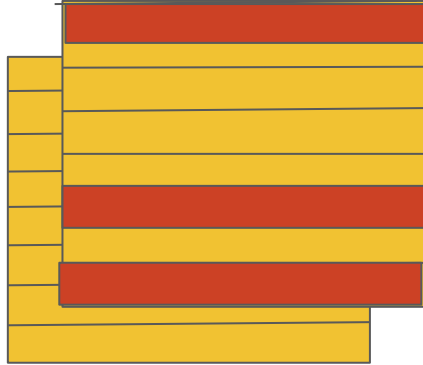


1.B) REDUÇÃO POR RELEVÂNCIA DO BANCO DE DADOS

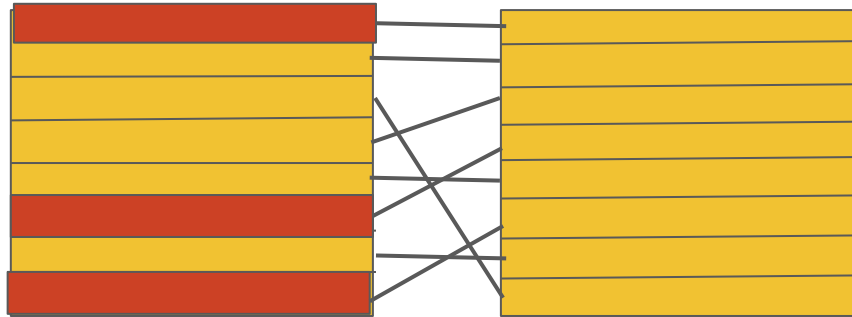
2. RECOLHIMENTO DOS DADOS



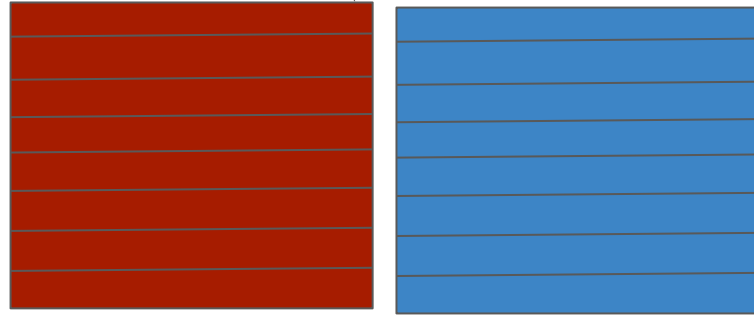
3. MARCAÇÃO DE ALUNOS DESISTENTES E NÃO DESISTENTES



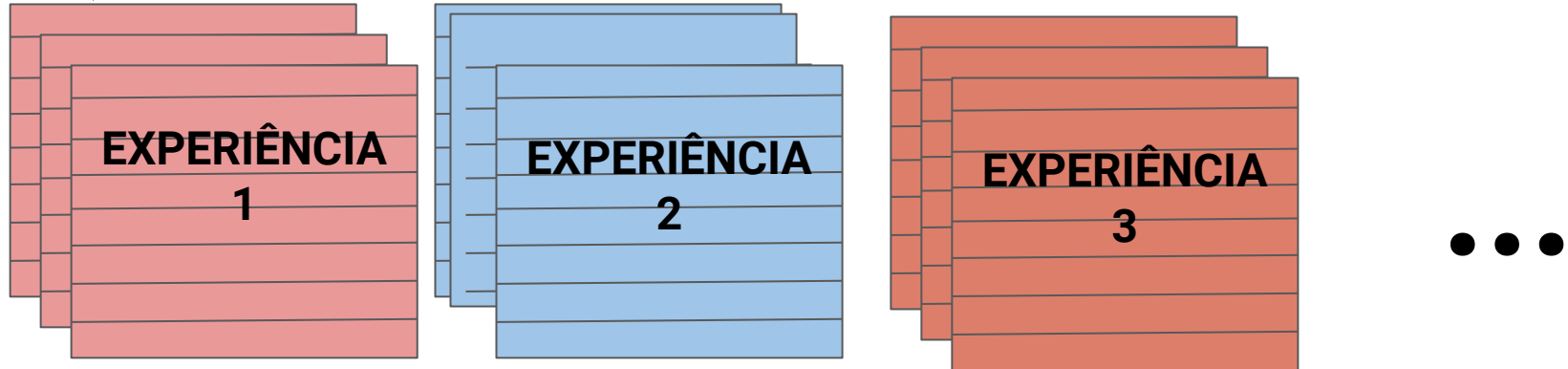
4. CRIAÇÃO DA RELAÇÃO CADASTRO-SUBMISSÃO



**5. SEPARAÇÃO DE ALUNOS
DESISTENTES E NÃO
DESISTENTES**



**6. TRANSFORMAÇÃO EM
EXPERIÊNCIA**



Aprendizado de Máquina

- Divisão aleatória do banco de dados em conjunto de testes e treinamento;
- Treinamento da rede neural com os dados de treinamento por aluno em batch (dados de todas as semanas do aluno inseridas de uma só vez);
- Avaliação dos resultados obtidos.

Pós-Processamento

Apesar da boa taxa de acerto percebesse um problema:

- **A rede gerada só consegue fazer predição com base em todas as atividades, não sendo possível predizer em tempo real o abandono do curso.**

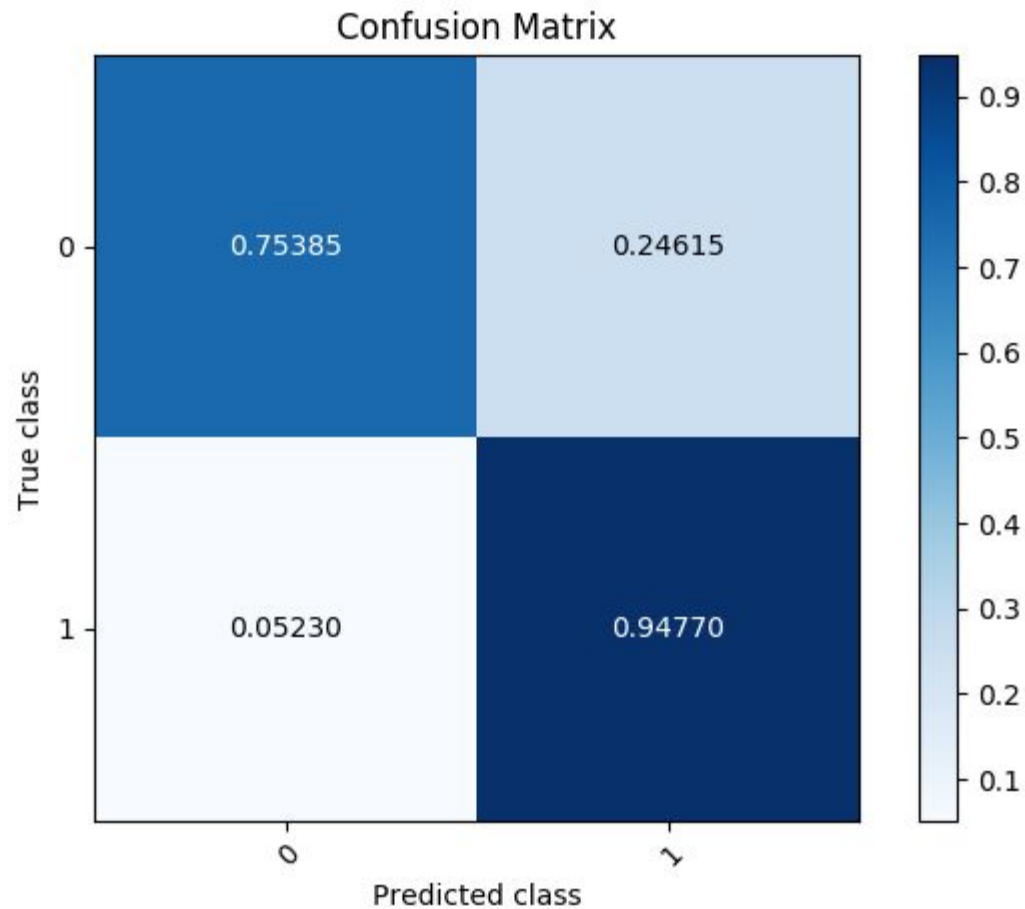
E qual a solução encontrada para tal problema?

- **Aproveitar que o processo é cíclico, voltar para a etapa de pré-processamento e gerar um matriz de entrada para cada semana do curso.**

Qual o resultado?

- **Após chegar pela segunda vez na etapa de pós-processamento notou-se uma melhora no tempo de treinamento de rede além da predição em tempo real não apresentada antes.**

Resultados



REFERÊNCIAS

[1] SILVA FILHO, Roberto Leal Lobo et al. A evasão no ensino superior brasileiro. **Cadernos de pesquisa**, v. 37, n. 132, p. 641-659, 2007.

[2] DE ALMEIDA, Onília Cristina de Souza. Evasão em cursos a distância: análise dos motivos de desistência. In: **14 Congresso Internacional ABED de Educação a Distância. Santos–São Paulo, Brasil. 2008.**