

Projeto:

Criação de um SAD para prever para cada aluno um índice de possível evasão

Motivo para criação do SAD:

Coordenadores pedagógicos, sabendo previamente qual aluno tem mais chance de evadir, pode mitigar essa previsão com ações de acompanhamento mais específicas.

Abaixo um sistema SAD para previsão de doenças baseado em sintomas.

ao colocar a intensidade dos sintomas o SAD abaixo retorna um índice para cada doença

<https://marconunes.shinyapps.io/diagnostico/>

No SAD para índice de evasão dos cursos, funciona de forma parecida, mas neste caso não é medido intensidade e sim algumas variáveis chaves que influenciam na evasão.

Estudo da Evasão Escolar dos Cursos Profissionalizantes em uma Unidade do Serviço Nacional de Aprendizagem Industrial de Santa Catarina - SENAI

<https://www.aedb.br/seget/arquivos/artigos13/31818288.pdf>

neste estudo são apresentada variaveis como:

Qual é seu nível de escolaridade?

Qual a instrução do chefe da família?

Atualmente você está trabalhando/estagiando de forma remunerada?

Por que você NÃO está trabalhando atualmente?

Possui Filhos?

Estas são algumas das variáveis altamente relacionadas à evasão apresentada no estudo

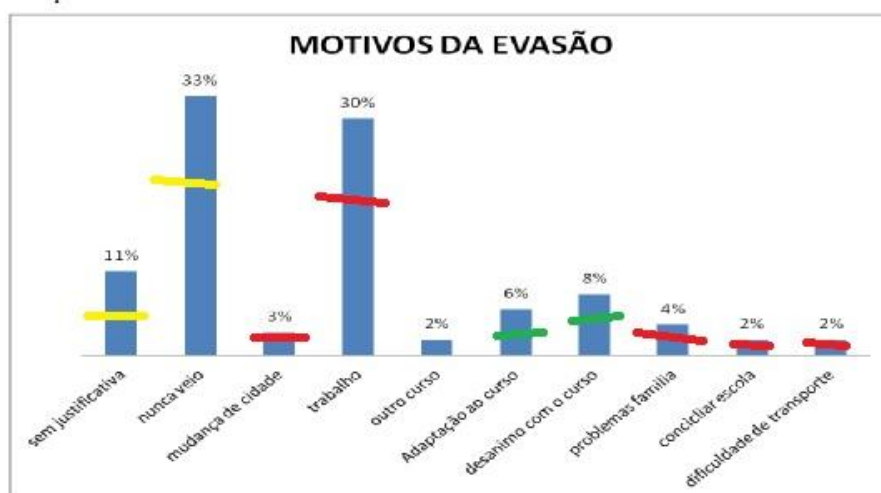


Gráfico 3: Motivos da Evasão (primeiro semestre de 2011)

Ressalta-se ainda a dificuldade de adaptação do aluno ao curso com 6% e desinteresse pelo curso com 8%, somados representam 14%, ou seja, 14 que abandonaram o curso em andamento.

Analisando o gráfico

Perceba que os índices em vermelho podem ser previstos por um modelo com as perguntas acima, os marcados em verde podem ser previsto no decorrer do curso e apenas as marcadas em amarelo tem previsão indefinida.

Devido aos diversos tipos de curso do SENAC PE e os perfis de egressos as variáveis de maior peso na evasão podem ser diferentes.

Partimos primeiramente da criação desta base com o cruzamento de dados das nossas bases para analisar se temos informações relevantes para criação de um modelo de treinamento para o SAD.

Cenário ideal para treinar a inteligencia:

10 mil alunos, 7 mil aprovados e 3 mil evadidos.

usamos 90% para treinamento do modelo e 10% usamos para avaliar sua efetividade.

Caso validado, passamos a observar o comportamento deste índice em futuros alunos por 1 ano.

validando, passamos a usar o índice como parametro de acompanhamento na busca de mitigar evasões.

abaixo um modelo simples de treinamento para os dados mencionados criado rapidamente no chat gpt

```
# Carregar a base de dados
```

```
data <- read.csv("alunos.csv")
```

```
# Dividir a base de dados em conjunto de treinamento e teste
```

```
set.seed(123)
```

```
trainingIndex <- sample(1:nrow(data), 0.7 * nrow(data))
```

```
trainingData <- data[trainingIndex, ]
```

```
testData <- data[-trainingIndex, ]
```

```
# Treinar o modelo usando Árvores de Decisão
```

```
library(rpart)
```

```
model <- rpart(Evasao ~ Escolaridade + EscolaridadeChefe + Empregado + Filhos,  
data=trainingData, method="class")
```

```
# Avaliar o modelo usando a acurácia
```

```
predictions <- predict(model, testData, type="class")
```

```
accuracy <- mean(predictions == testData$Evasao)
```

```
print(paste("Acurácia:", accuracy))
```

```
# Fazer previsões
```

```
newData <- data.frame(Escolaridade = c(10, 12, 14), EscolaridadeChefe = c(8, 10, 12),  
Empregado = c(1, 1, 0), Filhos = c(0, 2, 1))  
predictions <- predict(model, newData, type="class")  
print(predictions)
```