

# Análise do Desempenho de Alunos da UnB Usando um Sistema de Inteligência Artificial

Gabriel Silva<sup>1</sup>

Orientador: Prof. Dr. Marcelo Ladeira<sup>1</sup>

Colaboradores: Prof<sup>a</sup>. Dr<sup>a</sup>. Maristela Holanda<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação, Universidade de Brasília



## Introdução

- ▶ Evasão em universidades brasileiras traz desperdícios acadêmicos, sociais e econômicos.
- ▶ UnB teve prejuízo estimado em 95,6 milhões em 2014.

## Abordagem da UnB e Seus Problemas

- A UnB separa os alunos em condição dos demais, e tem os alunos em condição orientados.
- Problemas dessa abordagem:
- ▶ Alunos são separados em apenas dois grupos.
  - ▶ UnB age apenas quando aluno já está em condição.

## Proposta de Solução

- ▶ Criação de um sistema preditor capaz de identificar alunos em risco de serem desligados.
- ▶ Sistema preditor fornece uma tripla  $(v_1, v_2, v_3)$  de valores que indicam a chance do aluno graduar, evadir ou migrar.
- ▶ Vantagens:
  - ▷ Ações podem ser aplicadas antes de um aluno entrar em condição.
  - ▷ Ações podem ser aplicadas de acordo com o risco apresentado pelo aluno.

## Metodologia

- ▶ Utilizaram-se dados descaracterizados de alunos de graduação da UnB. Considerou-se apenas alunos que ingressaram a partir de 2000 e saíram até 2016.
- ▶ Cursos estudados: Ciência da Computação, Engenharia de Computação, Engenharia Mecatrônica, Engenharia de Redes, Engenharia de Software, Licenciatura em Computação.
- ▶ Consideraram-se tanto atributos pessoais quanto atributos acadêmicos.
- ▶ Pode-se constatar que os atributos variavam significativamente de acordo com o curso e a idade do aluno. Assim, realizou-se uma partição na base de dados original:
  - ▷ Alunos Seniores
  - ▷ Alunos Jovens da FT
  - ▷ Alunos Jovens de Licenciatura
  - ▷ Alunos Jovens de Computação
- ▶ Algoritmos de aprendizagem de máquina usados: ANN, *Naive Bayes*, *Random Forest*, Regressor Linear, SVR e ZeroR.
- ▶ Avaliação de Desempenho:
  - ▷ Alunos que ingressaram de 2000 até 2009 compõem os dados de treino. Alunos que ingressaram de 2010 até 2016 compõem os dados de teste.
  - ▷ Modelos induzidos e avaliados separadamente para cada semestre. Utilizou-se a métrica *F-measure*.
  - ▷ Para sumarizar o desempenho do algoritmo, calculou-se a média das *F-measures* de cada semestre.

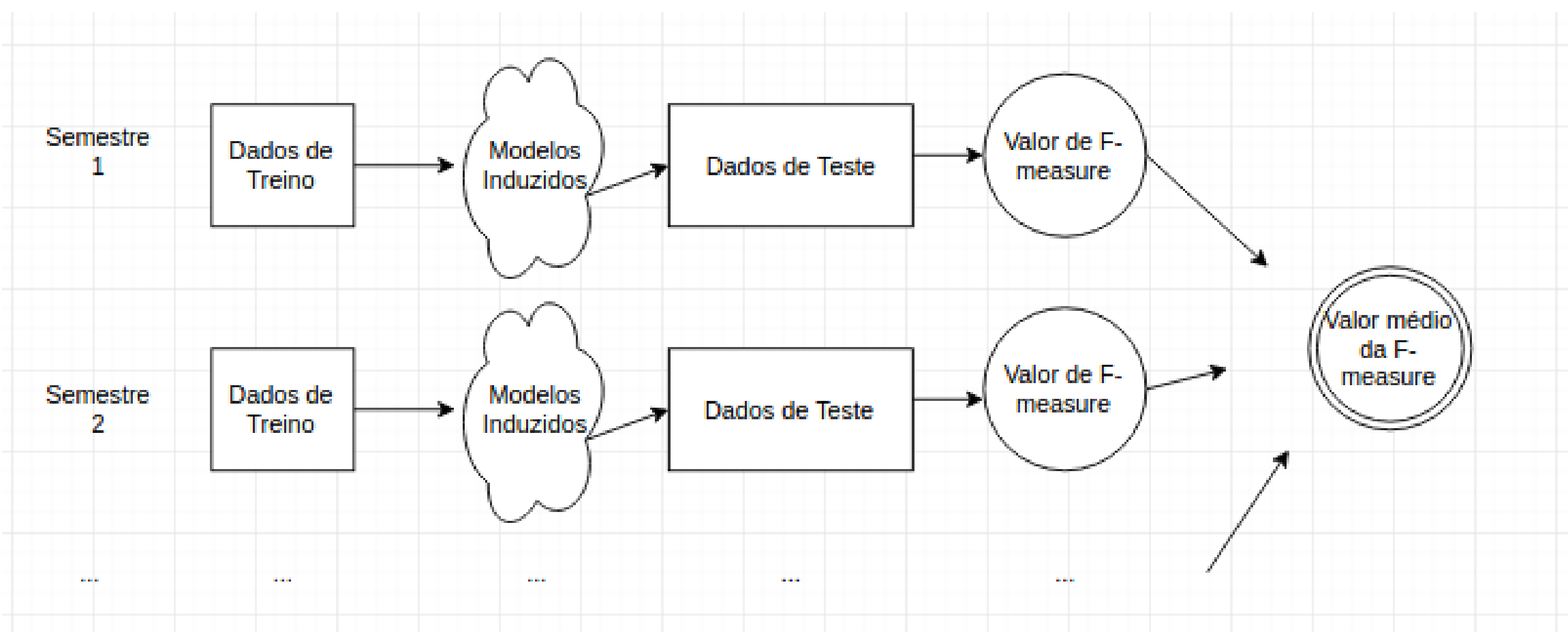


Figura 1: Diagrama para Avaliação do Desempenho

## Exemplo - Distribuição do Atributo Forma de Saída

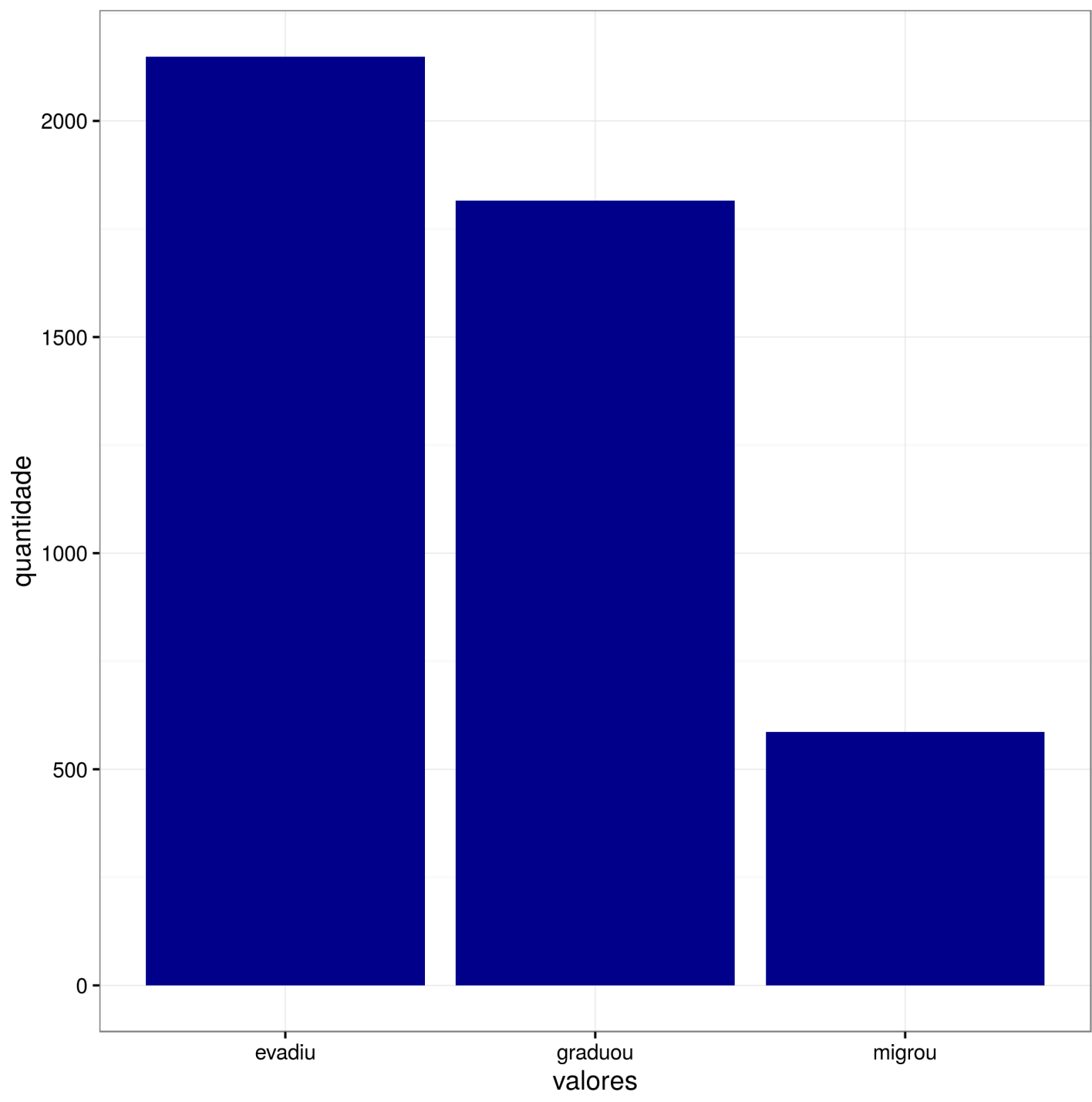


Figura 2: Forma de Saída para a Base de Dados Original

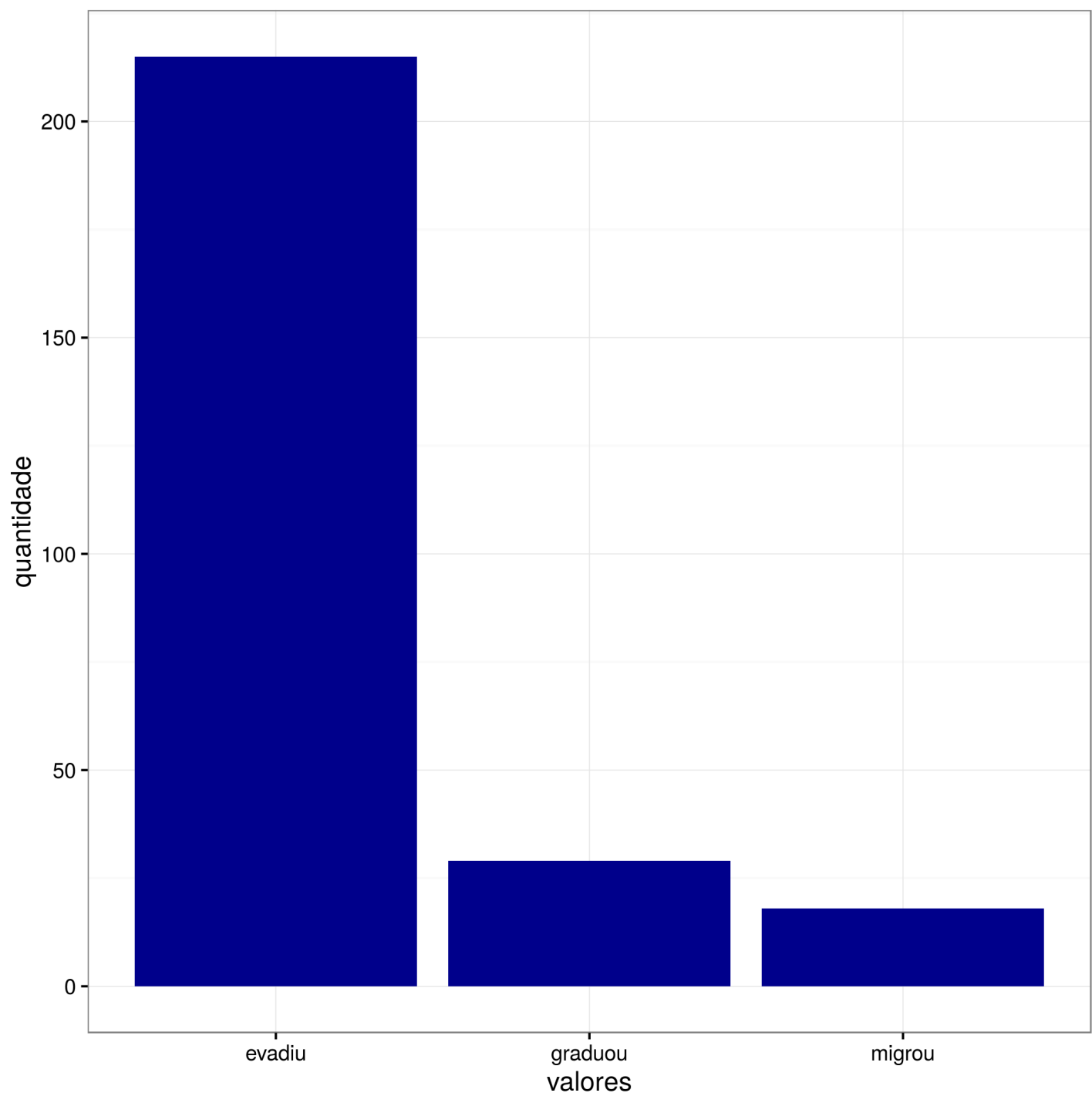


Figura 3: Forma de Saída para a Base de Dados dos Alunos Seniores

## Resultados

Tabela 1: Desempenho dos Algoritmos de acordo com a Base de Dados

Algoritmo	Sen	J - FT	J - Lic	J - Comp
ANN	0.62	0.76	0.85	0.74
Naive Bayes	0.28	0.56	0.76	0.65
Random Forest	0.70	0.73	0.85	0.76
Regressor Linear	0.75	<b>0.80</b>	<b>0.86</b>	<b>0.77</b>
SVR	<b>0.79</b>	0.76	0.82	0.70
ZeroR	0.61	0.64	0.70	0.60

- ▶ De forma geral, algoritmos conseguiram ter melhor desempenho do que o ZeroR.
- ▶ Regressor Linear teve um bom desempenho, em torno de 0.79.

## Conclusão

- ▶ Pesquisa aponta a viabilidade de usar aprendizagem de máquina para análise preditiva de alunos em risco de evasão na UnB.
- ▶ Metodologia usada na pesquisa pode ser aplicada a outros cursos de graduação da UnB.

## Agradecimentos

Agradecemos à UnB e ao CNPq pelo apoio à pesquisa. Agradecemos à Prof<sup>a</sup>. Dr<sup>a</sup> Maristela Holanda pela apresentação dessa pesquisa na conferência KDMiLe 2017.