

# INFLUÊNCIA DA DISCRETIZAÇÃO DE ATRIBUTOS NA ACURÁCIA DA CLASSIFICAÇÃO BAYESIANA

Gabriela Silva de Oliveira<sup>1</sup>, Silvio do Lago Pereira<sup>2</sup>

<sup>1</sup>Ex-aluna do Curso de Análise e Desenvolvimento de Sistemas – DTI/FATEC-SP

<sup>2</sup>Prof. Dr. do Departamento de Tecnologia da Informação – FATEC-SP

gsilvadoliveira@gmail.com, slago@fatecsp.br

## Resumo

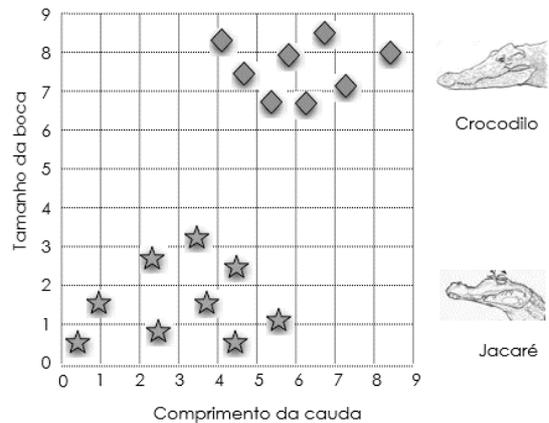
Classificação é uma técnica de aprendizado de máquina que sintetiza modelos para prever a classe de um objeto, a partir dos valores de seus atributos (que podem ser *discretos* ou *contínuos*). Particularmente, classificadores bayesianos são algoritmos que classificam objetos calculando a probabilidade de sua pertinência em cada classe possível e escolhendo a mais provável. Quando um atributo é discreto, o classificador calcula sua probabilidade usando frequência relativa; por outro lado, quando um atributo é contínuo, o classificador calcula sua probabilidade supondo que ele segue uma distribuição *normal*. O problema é que, na prática, essa suposição nem sempre se aplica e pode causar uma diminuição da acurácia da classificação. Uma forma de resolver esse problema é discretizar atributos contínuos, antes de classificar os objetos. Assim, o objetivo deste trabalho foi investigar como a discretização de atributos afeta a acurácia de um classificador bayesiano. Para isso, foram implementados os três principais métodos de discretização descritos na literatura. Para compará-los, foram usados conjuntos de exemplos previamente classificados, disponíveis no repositório de aprendizado de máquina da *Universidade da Califórnia*. Os resultados mostraram que, de fato, a discretização de atributos pode aumentar a acurácia dos modelos preditivos sintetizados por um classificador bayesiano; mas, isso dependerá de uma escolha criteriosa dos atributos a serem discretizados, bem como do número de intervalos usados na discretização, e não do método de discretização usado.

## 1. Introdução

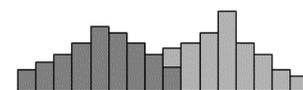
Classificação é uma das tarefas estudadas na área de *Aprendizagem de Máquina* [1], cujo objetivo é produzir algoritmos que sejam capazes de aprender uma tarefa e melhorar seu desempenho à medida que são usados.

Dado um conjunto de objetos previamente classificados por um especialista do domínio de aplicação (i.e., um conjunto de *exemplos*), um algoritmo de classificação deve sintetizar um modelo capaz de prever a classe de um novo objeto.

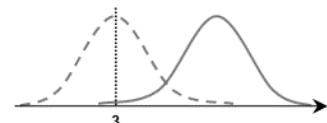
Por exemplo, a **Figura 1** mostra como é possível prever a classe de um animal desconhecido, a partir dos atributos (i.e., *comprimento da cauda* e *tamanho da boca*) e das classes (i.e., *Jacaré* ou *Crocodilo*) de um conjunto de animais conhecidos. Mais especificamente, a **Figura 1-a** mostra a distribuição dos valores dos atributos; e as **Figuras 1-c e 1-d** mostram, respectivamente, que um animal com comprimento de cauda igual a 3 tem maior probabilidade de ser um *Jacaré* (curva pontilhada), enquanto um objeto com comprimento de cauda igual a 5 tem maior probabilidade de ser um *Crocodilo* (curva contínua).



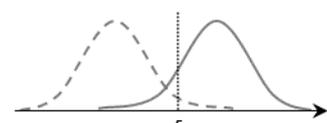
(a) Exemplos previamente classificados



(b) Histograma



(c) Objeto com comprimento de cauda 3



(d) Objeto com comprimento de cauda 5

**Figura 1** – Classificação de animais, adaptado de [2].

Evidentemente, todo modelo preditivo está sujeito a erros e o objetivo da classificação é sintetizar modelos preditivos que errem o mínimo possível (ou que tenham alta acurácia). Na prática, porém, é impossível avaliar a acurácia *real* de um modelo preditivo e, sendo assim, a acurácia costuma ser avaliada com relação a exemplos também previamente classificados por um especialista, mas que não foram usados para a síntese do modelo.

Assim, o objetivo deste artigo é descrever a implementação de um classificador bayesiano, bem como os métodos de discretização de atributos contínuos que foram usados nessa implementação, e avaliar a acurácia dos modelos preditivos sintetizados por esse classificador.

O restante do artigo está organizado do seguinte modo: a Seção 2 introduz os fundamentos teóricos do trabalho; a Seção 3 descreve as principais características do classificador bayesiano implementado; a Seção 4 discute resultados de experimentos feitos com esse classificador; e a Seção 5 apresenta as conclusões finais do artigo.

## 2. Fundamentos teóricos

Os fundamentos teóricos deste artigo são o Teorema de Bayes, a Classificação Bayesiana e a Discretização de Atributos Contínuos.

### 2.1. Teorema de Bayes

O Teorema de Bayes [3] estabelece a relação entre uma probabilidade condicional e sua inversa, isto é, entre a probabilidade de uma hipótese  $H$ , dada uma evidência  $E$ , e a probabilidade de uma evidência  $E$ , dada uma hipótese  $H$  (Fórmula 1):

$$P(H | E) = \frac{P(E | H)P(H)}{P(E)}, \quad (1)$$

sendo:

- $P(H) \neq 0$  e  $P(E) \neq 0$  as probabilidades independentes da hipótese  $H$  e da evidência  $E$ ;
- $P(H | E)$  a probabilidade condicional da hipótese  $H$ , dada a evidência  $E$ ;
- $P(E | H)$  a probabilidade condicional da evidência  $E$ , dada a hipótese  $H$ .

Interpretando *atributos* como evidências e *classes* como hipóteses, é possível aplicar o Teorema de Bayes para prever a classe de um objeto, a partir das probabilidades observadas em um conjunto de exemplos dado.

### 2.2. Classificação Bayesiana

O algoritmo de classificação bayesiana implementado neste artigo, denominado *Naïve-Bayes* [2], usa o Teorema de Bayes para prever a classe de um objeto, supondo que seus atributos são independentes. Apesar de essa suposição nem sempre ser verdadeira, resultados empíricos descritos na literatura mostram que o algoritmo *Naïve-Bayes* tem bons resultados e custo computacional baixo [2,4,5].

O algoritmo pode classificar objetos descritos por atributos discretos ou contínuos. No caso de atributos discretos, as probabilidades são dadas pelas frequências relativas (contagem) dos valores dos atributos. Já no caso de atributos contínuos, as probabilidades são dadas pela função de distribuição normal (Fórmula 2).

$$P(E_i = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

Evidentemente, a suposição de que todos os atributos contínuos seguem uma distribuição normal pode afetar negativamente o desempenho do classificador (pois, na realidade, pode ser que um determinado atributo contínuo não tenha uma distribuição normal). Assim, esse artigo investiga como a substituição dessa suposição pela discretização de atributos afeta a acurácia do classificador.

### 2.3. Discretização de Atributos

Essencialmente, um método de *discretização* transforma um atributo contínuo em um atributo discreto, mapeando intervalos contínuos em valores discretos.

Por exemplo, se um atributo contínuo, em um determinado conjunto de exemplos, tem os valores 64, 65, 68, 69, 70, 71, 72, 75, 80, 81, 83 e 85, então esses valores poderiam ser divididos como mostra a Figura 2.

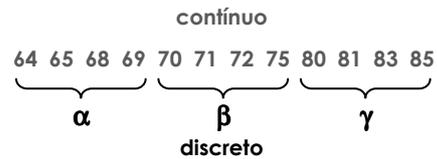


Figura 2 – Exemplo de discretização.

Assim, após a discretização, o atributo discreto correspondente passaria a ter apenas 3 valores possíveis:  $\alpha$ ,  $\beta$  e  $\gamma$ , e conseqüentemente, as probabilidades poderiam ser dadas por frequências relativas.

Há três métodos principais de discretização que são descritos na literatura [6,7], como mostra a Figura 3.

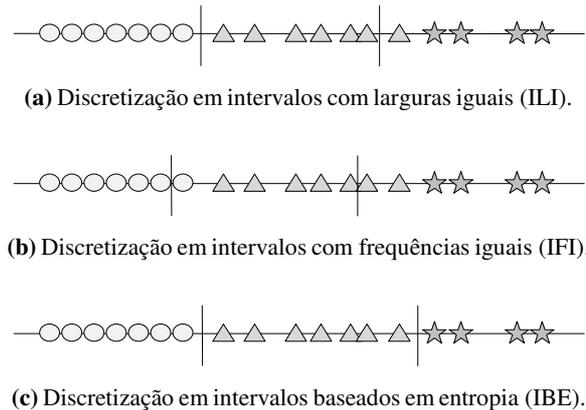


Figura 3 – Principais métodos de discretização.

Na discretização em *Intervalos com Larguras Iguais* (ILI), cada intervalo pode ter um número distinto de valores, mas a amplitude dos intervalos deve ser constante. Na discretização em *Intervalos com Frequências Iguais* (IFI), todos os intervalos devem ter o mesmo número de valores e, portanto, a amplitude deles pode variar. Por fim, na discretização em *Intervalos Baseados em Entropia* (IBE), os intervalos são divididos de modo que valores relacionados a uma mesma classe fiquem num mesmo intervalo.

## 3. O Classificador Bayesiano

O classificador bayesiano usado neste trabalho foi desenvolvido em linguagem *Python* [8], na fase inicial deste projeto [9,10]. Esse classificador apresenta uma interface gráfica de usuário bastante simples e funcional.

### 3.1. Conjuntos de Exemplos

Os conjuntos de exemplos, a partir dos quais os modelos preditivos são sintetizados, devem ser armazenados em arquivos ARFF [5].

Um arquivo ARFF (*Attribute-Relation File Format*) contém o nome do conjunto de exemplos, os tipos e os possíveis valores dos atributos usados na descrição dos objetos a serem classificados e uma sequência de tuplas (cada uma delas representando um exemplo de objeto previamente classificado por um especialista do domínio da aplicação). Em cada tupla, os valores dos atributos são separados por vírgulas e seguem a mesma ordem em que os atributos foram declarados no arquivo. Um exemplo de arquivo ARFF é apresentado na Figura 4.

```

@relation golf
@attribute outlook {sunny, overcast, rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {true, false}
@attribute play {yes, no}

@data
sunny, 85, 85, false, no
sunny, 80, 90, true, no
overcast, 83, 86, false, yes
rainy, 70, 96, false, yes
rainy, 68, 80, false, yes
rainy, 65, 70, true, no
overcast, 64, 65, true, yes
sunny, 72, 95, false, no
sunny, 69, 70, false, yes
rainy, 75, 80, false, yes
sunny, 75, 70, true, yes
overcast, 72, 90, true, yes
overcast, 81, 75, false, yes
rainy, 71, 91, true, no

```

Figura 4 – Arquivo ARFF para a relação *golf* [5].

### 3.2. Síntese do Modelo Preditivo

Para sintetizar um modelo preditivo, o usuário deve clicar o botão *Choose File*, na janela principal do classificador, e selecionar o arquivo ARFF contendo o conjunto de exemplos desejado, como mostra a Figura 5.

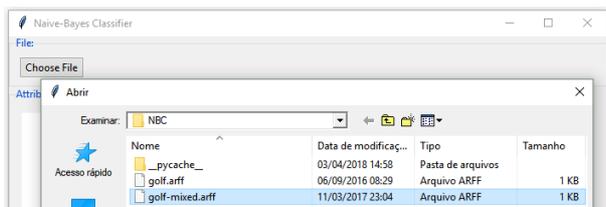


Figura 5 – Janela para seleção do conjunto de exemplos.

Após a seleção do arquivo, o classificador sintetiza o modelo preditivo correspondente e exibe algumas informações relevantes sobre ele. Por exemplo, na Figura 6 podemos ver o tipo do atributo *outlook*, os valores que ele assume, e ainda a probabilidade de ocorrência de cada um deles no conjunto de exemplos selecionado. Esse modelo preditivo inicial é sintetizado *sem* discretização.

Type	Domain	Count.	Prob.
Discrete	sunny	5	0.36
	rainy	5	0.36
	overcast	4	0.29

Figura 6 – Informações sobre o modelo preditivo sintetizado.

### 3.3. Classificação de uma Nova Instância

Após a síntese do modelo preditivo, o classificador habilita duas outras funcionalidades: classificação de um objeto desconhecido, do mesmo tipo daqueles descritos no conjunto de exemplos (botão *Classify New Instance*) e teste de acurácia (botão *Evaluate Classification Model*).

Caso o usuário clique no botão *Classify New Instance* e informe os valores dos atributos da nova instância, o classificador abre uma janela informando a probabilidade de essa instância pertencer a cada uma das possíveis classes consideradas no modelo preditivo (Figura 7).

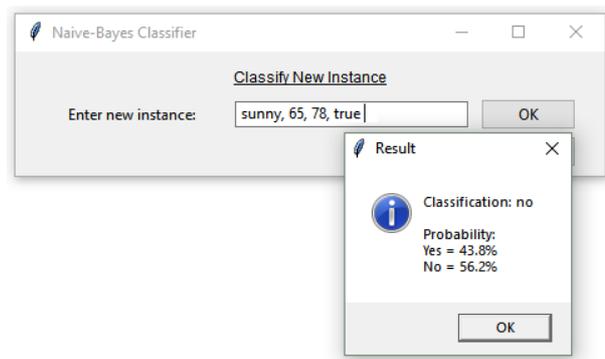


Figura 7 – Resultado da classificação de uma instância.

Caso o usuário clique o botão *Evaluate Classification Model*, o sistema exibe a *acurácia* (i.e., taxa de acerto) e a *g-measure* (i.e., média geométrica das taxas de acerto e erro) para o modelo preditivo sintetizado, bem como o desvio padrão para cada medida (Figura 8).

Accuracy Mean = 0.5533333333333333
Accuracy Standard Deviation = 0.2145296914316224
G-measure Mean = 0.4917706146613318
G-measure Standard Deviation = 0.10872531698246676

Figura 8 – Resultado da avaliação do modelo preditivo.

### 3.4. Seleção de um Método de Discretização

Caso o usuário queira discretizar um atributo contínuo, basta selecionar o atributo, o método de discretização desejado e o número de intervalos e, depois, clicar o botão *Discretize*. O classificador então exibirá as modificações feitas naquele atributo, substituindo os valores numéricos pelos valores discretos correspondentes, gerados pelo método de discretização escolhido (Figura 9).

Attributes	Type	Mean	Std. Dev.
outlook			
temperature	Numeric	73.57	6.57
humidity			
windy			
play			

(a) Antes da discretização do atributo.

Attributes	Type	Domain	Count.	Prob.
outlook				
temperature	Same Width	2	4	0.29
humidity		0	5	0.36
windy		1	5	0.36
play				

(b) Após a discretização do atributo.

Figura 9 – Discretização do atributo *temperature*.

#### 4. Resultados Empíricos

Todos os experimentos com o classificador bayesiano foram feitos com conjuntos de exemplos previamente classificados e disponíveis no repositório de aprendizado de máquina da UCI (*University of California at Irvine*) [11].

Cada conjunto de exemplos foi dividido em duas partes, sendo que uma delas foi usada para a síntese do modelo preditivo (i.e., *conjunto de treinamento*) e a outra foi usada para avaliação da acurácia do modelo preditivo sintetizado (i.e., *conjunto de validação*). Esse método de avaliação, denominado *holdout* [12], especifica que uma boa proporção para a divisão do conjunto de exemplos é usar 2/3 dos exemplos para treinamento e 1/3 dos exemplos para validação. Porém, para evitar que a ordenação original dos dados nos conjuntos de exemplos interferisse nos resultados dos experimentos (por exemplo, em alguns conjuntos de dados, as classes ocorrem agrupadas e a mera divisão desses conjuntos em 2/3 e 1/3 pode resultar em conjuntos de treinamento ou validação em que todos os exemplos sejam da mesma classe), antes de serem divididos, os conjuntos de exemplos foram *embaralhados*. Ademais, para manter a proporção da distribuição de classes observada no conjunto de exemplos original, os exemplos para compor os conjuntos de treinamento e validação foram criteriosamente selecionados. Por exemplo, para um conjunto de dados original contendo 60% de exemplos de uma classe e 40% de outra, foram selecionados 2/3 dos exemplos, de modo que 60% deles fossem de uma classe e os demais 40% fossem da outra.

Como o embaralhamento dos conjuntos de exemplos afeta os conjuntos de treinamento e validação obtidos, todos os resultados reportados nessa seção são valores médios obtidos a partir de 30 execuções do procedimento de avaliação, para cada conjunto de exemplos considerado.

Em trabalho publicado anteriormente [9], descrevemos detalhes da implementação da versão preliminar do classificador bayesiano (que não tinha a função de discretização de atributos contínuos) e relatamos resultados de sua avaliação com 12 conjuntos de exemplos distintos. Os resultados mostram que, para a maioria dos conjuntos de exemplos (75%), a acurácia média dos modelos preditivos sintetizados pelo classificador é alta (acima de 85%). Portanto, o classificador desenvolvido mostrou-se capaz de atingir resultados satisfatórios o suficiente para que pudéssemos realizar os testes de discretização.

Nesta segunda etapa do trabalho, o objetivo foi implementar os métodos de discretização e avaliar a sua influência na acurácia dos modelos preditivos sintetizados.

##### 4.1. Experimento I: Discretização de Atributos

O primeiro experimento feito neste artigo teve como objetivo investigar a influência individual de dois fatores envolvidos na discretização: a escolha do atributo a ser discretizado e a escolha do número de intervalos usado na discretização. Neste experimento, usamos o conjunto de exemplos *Íris*, cujos objetos são descritos por 4 atributos.

Os resultados apresentados na **Tabela I** foram obtidos com a discretização de um único atributo de cada vez, com um número de intervalos fixo. Eles mostram que a influência da discretização do atributo *petalwidth* é maior que influência da discretização dos demais atributos. Isso indica que, na prática, os atributos a serem discretizados

devem ser cuidadosamente escolhidos por um especialista do domínio de aplicação.

**Tabela I** – Discretização de apenas um atributo.

Atributo Discretizado	Acurácia	G-Measure
sepalength	0.8150	0.7610
sepalwidth	0.8712	0.8288
petallength	0.9431	0.9274
petalwidth	0.9569	0.9426

Em seguida, apenas o atributo *petalwidth* foi discretizado (essa escolha é justificada pelos resultados obtidos na **Tabela I**) em 2, 4, 5 e 10 intervalos, e novamente a avaliação dos modelos preditivos correspondentes foi feita. Os resultados dessa avaliação, apresentados na **Tabela II**, mostram que a escolha do número de intervalos também é importante. De fato, o conjunto de exemplos *Íris* possui 150 objetos e dividir esses valores em apenas 2 grupos diminui a precisão do resultado, pois valores muito distantes são inseridos num mesmo intervalo.

**Tabela II** – Discretização do atributo *petalwidth*.

Número de Intervalos	Acurácia	G-Measure
2	0.8608	0.8219
4	0.9209	0.8840
5	0.9444	0.9250
10	0.9660	0.9549

##### 4.2. Experimento II: Comparação dos Métodos

Para investigar como os métodos de discretização de atributos considerados neste artigo afetam a acurácia do classificador bayesiano, foram feitos experimentos com 21 conjuntos de exemplos. A **Tabela III** mostra os percentuais de melhoria nas acurácias dos modelos obtidos.

**Tabela III** – Percentuais de melhoria em acurácia.

Conjunto de Exemplos	ILI	IFI	IBE
Abalone	12.23 %	12.08 %	12.87 %
Appendicitis	17.39 %	19.07 %	18.79 %
Balance	<b>-10.16 %</b>	<b>-10.48 %</b>	<b>-9.42 %</b>
Breast Cancer	0.45 %	0.42 %	0.62 %
Bupa	7.34 %	6.08 %	6.71 %
Ecoli	<b>-27.65 %</b>	<b>-26.81 %</b>	<b>-30.22 %</b>
Glass Identification	70.27 %	69.76 %	65.81 %
Haberman	<b>-1.07 %</b>	<b>-0.67 %</b>	<b>-0.76 %</b>
Hayes-Roth	39.87 %	38.71 %	39.61 %
Image Segmentation	11.12 %	11.93 %	9.99 %
Íris	10.05 %	10.21 %	10.36 %
LED Display	75.24 %	75.10 %	77.23 %
Mammographic Mass	<b>-1.11 %</b>	<b>-1.97 %</b>	<b>-1.27 %</b>
Page Blocks	7.48 %	7.46 %	7.32 %
Phoneme	<b>-0.93 %</b>	<b>-0.24 %</b>	<b>-0.33 %</b>
Pima	<b>-3.08 %</b>	<b>-1.53 %</b>	<b>-4.12 %</b>
Spambase	44.83 %	45.32 %	45.12 %
Titanic	1.07 %	1.06 %	1.12 %
Vehicle	15.07 %	16.67 %	15.26 %
Vowel Recognition	286.49 %	<b>289.31 %</b>	285.40 %
Wine	0.86 %	1.60 %	1.35 %

Esses percentuais foram obtidos comparando-se as acurácias dos modelos sintetizados com e sem discretização. Como a acurácia de um modelo preditivo é uma medida calculada com base apenas nas taxas de acerto, para cada conjunto de exemplos, também foram calculadas as *g-measures*, que são médias geométricas das taxas de acertos e erros de predição. Os percentuais de melhoria em relação à *g-measure* são apresentados na **Tabela IV**.

**Tabela IV** – Percentuais de melhoria em *g-measure*.

Conjunto de Exemplos	ILI	IFI	IBE
Abalone	7.38 %	6.82 %	7.77 %
Appendicitis	37.54 %	40.87 %	40.98 %
Balance	5.70 %	18.30 %	<b>-3.75 %</b>
Breast Cancer	0.32 %	0.85 %	0.75 %
Bupa	<b>-2.09 %</b>	<b>-10.65 %</b>	<b>-6.87 %</b>
Ecoli	<b>-13.82 %</b>	<b>-11.05 %</b>	<b>-15.03 %</b>
Glass Identification	76.15 %	66.54 %	67.53 %
Haberman	<b>-33.93 %</b>	<b>-33.92 %</b>	<b>-35.59 %</b>
Hayes-Roth	161.68 %	161.44 %	155.87 %
Image Segmentation	4.09 %	8.22 %	<b>-7.53 %</b>
Iris	11.66 %	12.33 %	11.72 %
LED Display	180.17 %	167.89 %	184.60 %
Mammographic Mass	<b>-0.93 %</b>	<b>-1.15 %</b>	<b>-0.46 %</b>
Page Blocks	<b>-10.32 %</b>	<b>-11.04 %</b>	<b>-11.08 %</b>
Phoneme	<b>-2.09 %</b>	<b>-5.23 %</b>	<b>-4.60 %</b>
Pima	<b>-4.90 %</b>	<b>-2.85 %</b>	<b>-3.63 %</b>
Spambase	45.64 %	50.59 %	53.08 %
Titanic	11.06 %	10.80 %	11.14 %
Vehicle	10.49 %	9.12 %	14.17 %
Vowel Recognition	201.69 %	202.31 %	202.76 %
Wine	0.73 %	0.33 %	0.74 %

Os resultados nas **Tabelas III e IV** mostram que, para a maioria dos conjuntos de exemplos considerados (14 de um total de 21 conjuntos de exemplos), a discretização dos atributos contínuos melhorou o desempenho dos modelos preditivos sintetizados pelo classificador bayesiano (percentuais positivos). Além disso, nenhum dos três métodos de discretização implementados (ILI – *Intervalos com Larguras Iguais*, IFI – *Intervalos com Frequências Iguais*, e IBE – *Intervalos Baseados em Entropia*) apresentou vantagem significativa sobre os outros dois métodos.

É importante ressaltar que, nesses experimentos, todos os atributos contínuos foram discretizados (pois não foi possível definir um critério para escolher quais atributos deveriam ou não ser discretizados). Acreditamos que uma escolha mais criteriosa dos atributos a serem discretizados poderia melhorar bastante os resultados obtidos.

## 5. Conclusões

Os resultados empíricos mostraram que a discretização de atributos contínuos pode melhorar a precisão dos modelos preditivos sintetizados por um classificador bayesiano, mas os níveis de melhoria são muito variados.

Para alguns conjuntos de exemplos, a precisão do modelo preditivo diminuiu após a discretização dos atri-

butos contínuos. Esses conjuntos de exemplos, em geral, estão relacionados à área de medicina. Provavelmente, nesses conjuntos de exemplos os valores dos atributos contínuos têm uma distribuição normal e, portanto, a sua discretização reduz a precisão dos modelos preditivos sintetizados pelo classificador bayesiano. Além disso, em outros conjuntos de dados, o aumento de precisão foi muito pequeno. Isso ocorreu nos conjuntos de exemplos que já tinham alta precisão de classificação antes da discretização (logo, não era possível melhorar significativamente a precisão dos modelos preditivos).

Os resultados empíricos também mostraram que uma escolha criteriosa dos atributos a serem discretizados, e do número de intervalos a serem usados na discretização, é um fator muito importante para a precisão da classificação bayesiana.

No geral, para a maioria dos conjuntos de exemplos considerados nos experimentos realizados, pode-se constatar que a discretização de atributos melhorou a acurácia dos modelos preditivos sintetizados pelo classificador bayesiano. Além disso, nenhum dos três métodos de discretização de atributos considerados neste trabalho se mostrou significativamente melhor do que os outros dois.

## Agradecimentos

Ao CNPq pela bolsa de Iniciação Científica<sup>1</sup> (Processo Nº 107647/2017-6).

## Referências Bibliográficas

- [1] A. Blum. **Machine Learn Theory**. Carnegie Mellon University, Department of Computer Science. Disponível em: [www.cs.cmu.edu/afs/cs/user/avrim/www/Talks/mlt.pdf](http://www.cs.cmu.edu/afs/cs/user/avrim/www/Talks/mlt.pdf). Acesso em: 04/04/2018.
- [2] E. Keogh. **Naïve Bayes Classifier**. University of California, Riverside. Disponível em: [www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect\\_examples.pdf](http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf). Acesso em: 04/04/2018.
- [3] **Introduction to Probability Theory, Lecture 6: Bayes' Theorem**. PennState University, Eberly College of Science. Disponível em: <https://online.courses.science.psu.edu/stat414/node/45>. Acesso em: 04/04/2018.
- [4] D. Barber. **Bayesian Reasoning and Machine Learning, draft**, University College London, 2010. Disponível em: [web4.cs.ucl.ac.uk/staff/D.Barber/textbook/090310.pdf](http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/090310.pdf). Acesso em: 04/04/2018.
- [5] I. H. Witten; E. Frank; M. A. Hall. **Data Mining: Practical Machine Learning Tools and Techniques, 3<sup>rd</sup> edition**, Elsevier, 2011.
- [6] F. Muhlenbach; R. Rakotomalala. **Discretization of Continuous Attributes**. Encyclopedia of Data Warehousing and Mining, Idea Group Reference, p.397-402, 2005.
- [7] F. Kaya. **Discretizing Continuous Features for Naive Bayes and C4.5 Classifiers**. University of Maryland. Disponível em: [https://www.cs.umd.edu/sites/default/files/scholarly\\_papers/fatih-kaya\\_1.pdf](https://www.cs.umd.edu/sites/default/files/scholarly_papers/fatih-kaya_1.pdf). Acesso em: 04/04/2018.

- [8] S. F. Lott. **Building Skills in Python**, Creative Commons, USA, 2010.
- [9] G. S. Oliveira; S. L. Pereira. **Implementação de Classificação Bayesiana**. Boletim Técnico da Faculdade de Tecnologia de São Paulo, v. 43, p. 10-15, São Paulo, 2017. Disponível em [http://bt.fatecsp.br/bulletins/show\\_article/1036](http://bt.fatecsp.br/bulletins/show_article/1036). Acesso em 04/04/2018.
- [10] G. S. Oliveira; S. L. Pereira. **Classificação Bayesiana e Discretização de Atributos Contínuos**. Trabalho de Conclusão de Curso, Faculdade de Tecnologia de São Paulo, 2017.
- [11] UCI Machine **Learning Repository**. Disponível em: <http://archive.ics.uci.edu/ml/datasets.html?sort=name&view=list>. Acesso em: 04/04/2018.
- [12] D. Michie et al. **Machine Learning, Neural and Statistical Classification**, 1994. Disponível em: [www1.maths.leeds.ac.uk/~charles/statlog/whole.pdf](http://www1.maths.leeds.ac.uk/~charles/statlog/whole.pdf). Acesso em: 04/04/2018.