

# **Estudos e Experimentos com Técnicas de Seleção de Atributos Lazy**

Lincoln Pires

Orientador: Luiz Merschmann

# Sumário

1. Introdução
2. Justificativa
3. Objetivos
4. Desenvolvimento
5. Trabalhos Futuros
6. Cronograma

# Introdução

“Estamos afogados em informação mas famintos por conhecimento”

John Naisbitt

# Introdução

## Mineração De Dados



# Introdução

Knowledge Discovery in Database – KDD

Processo de Descoberta de Conhecimento  
em bases de Dados

# Introdução

## Pré-Processamento

Seleção

Limpeza

Transformação

Redução dos Dados

## Pós-Processamento

Visualização / Análise

# Introdução

## Tarefas Descritivas

- Extração de Regras de Associação
- Agrupamentos

## Tarefas Preditivas

- Regressão
- Classificação\*

# Justificativa

Qualidade dos dados é fundamental

Seleção

Limpeza

Transformação

Redução dos Dados

# Justificativa

## Seleção de Atributos

- Abordagem Individual
- Abordagem Subconjuntos

# Justificativa

Problemas da abordagem de Subconjuntos

$$n \Rightarrow 2^n$$

# Justificativa

## Avaliação dos Subconjuntos

- Consistency-based Feature Selection
- Correlation-based Feature Selection

# Justificativa

- Abordagem Eager
- Abordagem Lazy

# Justificativa

Base ordenada por X			Base ordenada por Y		
- X -	- Y -	- C -	- X -	- Y -	- C -
1	2	B	2	1	A
1	3	B	3	1	B
1	4	B	4	1	A
2	1	A	1	2	B
2	2	A	2	2	A
2	3	A	3	2	B
3	1	B	1	3	B
3	2	B	2	3	A
3	4	B	4	3	B
4	1	A	1	4	B
4	3	B	3	4	B
4	4	B	4	4	B

# Objetivos

- Testar abordagem lazy – Métrica de consistência com grande variedade de bases
- Analisar os resultados e identificar as limitações desta técnica
- Avaliar uma abordagem lazy de seleção com base na métrica de correlação

# Desenvolvimento



# Trabalhos Futuros



# Cronograma

1. Estudo sobre Mineração de Dados
2. Estudo do código do software Weka
3. Estudo sobre Seleção de Atributos
4. Revisão bibliográfica para identificação do estado da arte das heurísticas utilizadas pelas técnicas

# Cronograma

5. Seleção de instâncias e realização dos testes computacionais
6. Análise dos Resultados
7. Estudo e implementação do algoritmo baseado na métrica de Correlação
8. Elaboração da monografia

