

COMPARATIVO ENTRE ALGORITMOS DE CLASSIFICAÇÃO EM IMAGENS DE POSSÍVEIS METEOROS





Contexto



Can I make a wish? Detecting shooting stars

Classify images into one of two classes: meteor vs non-meteor.

Contexto

- Conjunto de dados baseado em uma série de fotos que foram tiradas do céu pelo Observatório de Astronomia da UNIVAP (Universidade do Vale do Paraíba) com a intenção de identificar meteoros.





Introdução

- Proposta:
 - Classificar as imagens em meteoros e não-meteoros.
 - Estudar algoritmos de classificação e ver qual obtém as melhores métricas.



Introdução

- Problemas:
 - Imagens podem conter: insetos, pássaros, aviões, raios, chuva e os próprios ruídos da foto.

Exemplos



Meteoro



Não-Meteoro



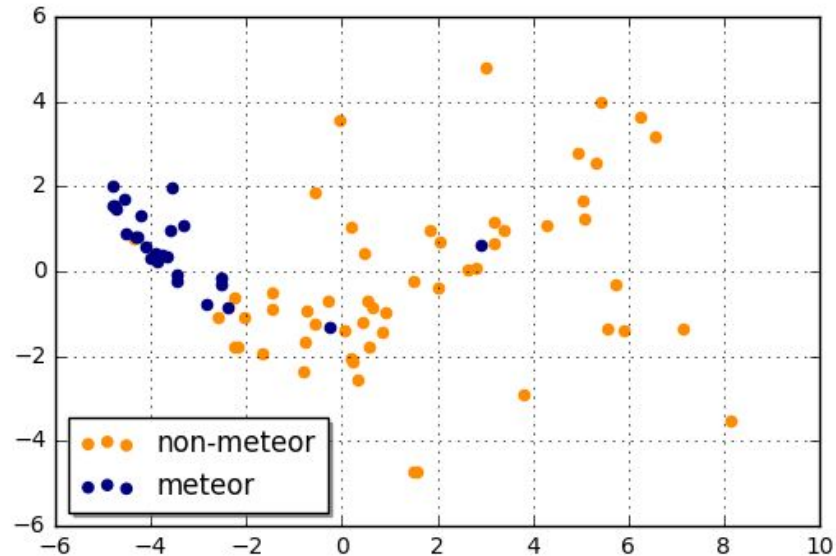
Conjunto de dados

CONJUNTO DE DADOS BRUTO

Conjunto	Linhas	Atributos
Treino	80	3453
Teste	42	3453

Conjunto	Meteoro	Não-Meteoro	Total
Treino	26	54	80
Teste	15	27	42

Conjunto de dados





Pré-Processamento

- Remoção de dados com variação nula

CONJUNTO DE DADOS APÓS REMOÇÃO DE DADOS COM VARIAÇÃO NULA

Conjunto	Linhas	Atributos
Treino	80	1 306
Teste	42	1 306



Pré-Processamento

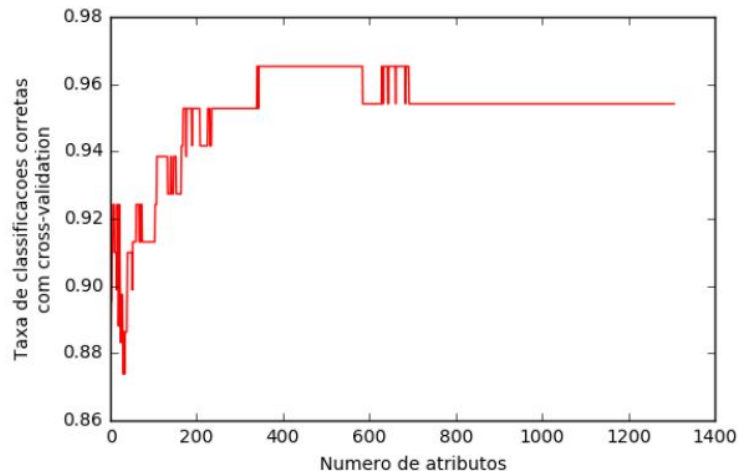
- Seleção de atributos
 - Árvores de decisão

CONJUNTO DE DADOS APÓS TFS

Conjunto	Linhas	Atributos
Treino	80	96
Teste	42	96

Pré-Processamento

- Seleção de atributos
 - Recursão com validação cruzada



CONJUNTO DE DADOS APÓS **RFE**

Conjunto	Linhas	Atributos
Treino	80	691
Teste	42	691



Algoritmos analisados

- K-Vizinhos mais próximos
- Redes Bayesianas
- Árvores de Decisão
- Perceptron Multicamadas



Métricas de análise

- Acurácia
- Log Loss
- Precisão
- Média
- Mediana
- Desvio Padrão



Análises

- Conjunto de dados brutos

CONJUNTO DE DADOS **BRUTO**: ACURÁCIA, LOG LOSS E PRECISÃO

Resultados	Acurácia	Log Loss	Precisão
KNN	0,9286	2,4671	1,0000
MLP	0,9048	3,2894	1,0000
DECISION TREE	0,8810	4,1118	1,0000
REDES BAYESIANAS	0,7381	9,0459	0,83333



Análises

- Conjunto de dados brutos

CONJUNTO DE DADOS **BRUTO**: ANÁLISE DA ACURÁCIA DE TODOS OS ALGORITMOS

Resultados	Acurácia
MÉDIA	0,8631
MEDIANA	0,8929
DESVIO PADRÃO	0,0741



Análises

- Conjunto de dados após TFS

CONJUNTO DE DADOS APÓS **TFS**: ACURÁCIA, LOG LOSS E PRECISÃO

Resultados	Acurácia	Log loss	Precisão
KNN	0,8571	4,9341	0,90909
MLP	0,8810	4,1118	0,91667
DECISION TREE	0,9047	3,2894	1,0000
REDES BAYESIANAS	0,7619	8,2236	0,77778



Análises

- Conjunto de dados após TFS

CONJUNTO DE DADOS APÓS **TFS**: ANÁLISE DA ACURÁCIA DE TODOS OS ALGORITMOS

Resultados	Acurácia
MÉDIA	0,8571
MEDIANA	0,8810
DESVIO PADRÃO	0,0558



Análises

- Conjunto de dados após RFE

CONJUNTO DE DADOS APÓS **RFE**: ACURÁCIA, LOG LOSS E PRECISÃO

Resultados	Acurácia	Log loss	Precisão
KNN	0,88095	4,11178	0,91667
MLP	0,9048	3,2894	1,0000
DECISION TREE	0,8810	4,1118	1,0000
REDES BAYESIANAS	0,8333	5,7565	1,0000



Análises

- Conjunto de dados após RFE

CONJUNTO DE DADOS APÓS **RFE**: ANÁLISE DA ACURÁCIA DE TODOS OS ALGORITMOS

Resultados	Acurácia
MÉDIA	0,8750
MEDIANA	0,8810
DESVIO PADRÃO	0,0259



Conclusões

- O algoritmo KNN obteve a melhor acurácia entre todos os outros, com o conjunto de dados bruto.
- Na média, os algoritmos se saíram melhor com o conjunto de dados após a seleção de atributos recursiva.
- No geral, todos obtiveram métricas satisfatórias, porém isso não pode ser conclusivo pelo fato do conjunto ser pequeno.