



# Relatório Mineração de Dados - Classificação semi-supervisionada com aprendizado por auto-treinamento

Claudemir Woche, Francisco Yuri

27 de Junho de 2019

# 1 Tabelas de precisões

Foram executados 3 experimentos para cada valor de k. As tabelas apresentam a média das precisões dos experimentos e foram feitas com precisão de 3 casas decimais nas precisões. Foi utilizado o conjunto de dados *breast\_cancer* da biblioteca *scikit – learn*.

## 1.1 Floresta Aleatória

Classificador	K	Precisão Média
Floresta Aleatória(n_estimators=100)	5	0,954
	10	0,956
	15	0,963
	20	0,965
	25	0,956
	50	0,953
	75	0,963
	100	0,961
	125	0,953
	150	0,960
	200	0,954
	250	0,958

## 1.2 Vizinhos mais próximos

Classificador	K	Precisão Média
Vizinhos mais próximos()	5	0,930
	10	0,933
	15	0,937
	20	0,933
	25	0,935
	50	0,939
	75	0,933
	100	0,942
	125	0,940
	150	0,939
	200	0,937
	250	0,942

## 1.3 Support Vector Machines Linear

Classificador	K	Precisão Média
Support Vector Machines Linear(gamma = "scale")	5	0,900
	10	0,896
	15	0,898
	20	0,898
	25	0,900
	50	0,907
	75	0,909
	100	0,907
	125	0,911
	150	0,914
	200	0,918
	250	0,911

## 1.4 Árvore de Decisão

Classificador	K	Precisão Média
Árvore de Decisão()	5	0,933
	10	0,914
	15	0,925
	20	0,925
	25	0,918
	50	0,926
	75	0,933
	100	0,925
	125	0,923
	150	0,925
	200	0,935
	250	0,925

## 1.5 Rede Neural

Classificador	K	Precisão Média
Rede Neural(max_iter = 1000)	5	0,914
	10	0,900
	15	0,939
	20	0,918
	25	0,926
	50	0,916
	75	0,925
	100	0,937
	125	0,912
	150	0,944
	200	0,737
	250	0,830

## 2 Conclusão

Diferentes valores de K tem baixa influência na precisão média de um classificador. Com exceção do classificador Rede Neural que para  $k = 200$  e  $250$  apresentou precisões médias consideravelmente diferentes das anteriores. O classificador Random Forest com  $n\_estimators = 100$  teve maior precisão para todos os valores de K.