



UFABC

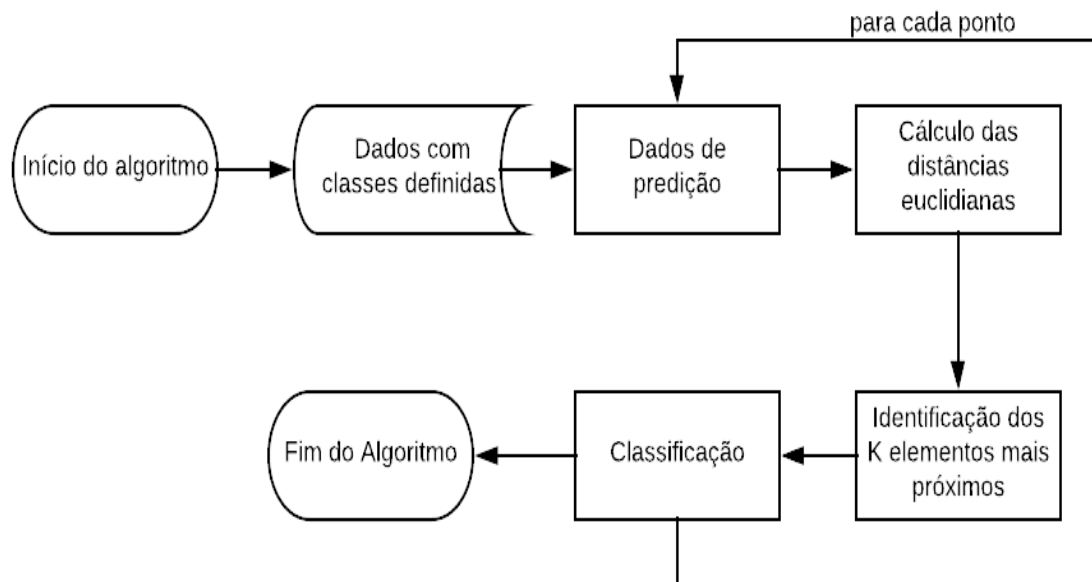
Sistemas Inteligentes

Estudo comparativo entre algoritmos implementados
de classificação e clusterização

Dimitri Leandro de Oliveira Silva	11201720576
Daniel Vieira Batista	11106614
Gabriel Fernandes	11048816
Yan Podkorytoff Ike Chícharo	21068214

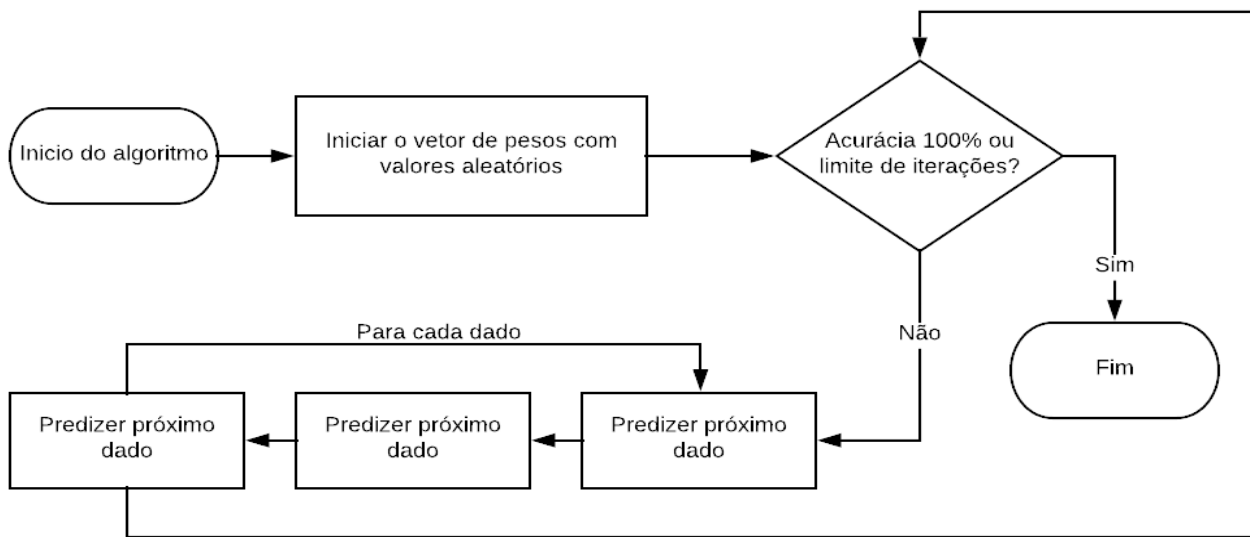
- Implementar algoritmos vistos no curso do zero;
- Implementar métricas para gerar resultados;
- Comparar com os algoritmos e as métricas já implementadas pela biblioteca Scikit Learn;

IMPLEMENTAÇÕES



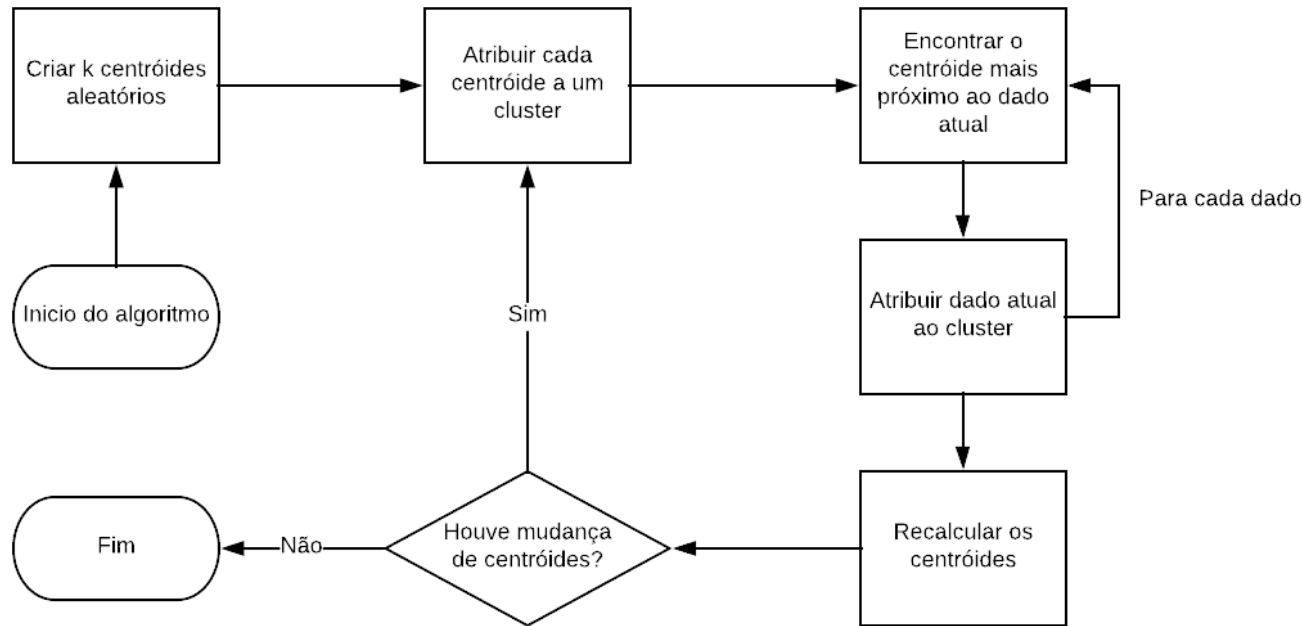
- Classificação;
- K vizinhos mais próximos;
- Bases de dados utilizadas: Íris dataset e Olivetti Faces.

Perceptron



- Classificação;
- Rede neural mais simples;
- Base de dados utilizada: Breast cancer wisconsin dataset.

K-Means

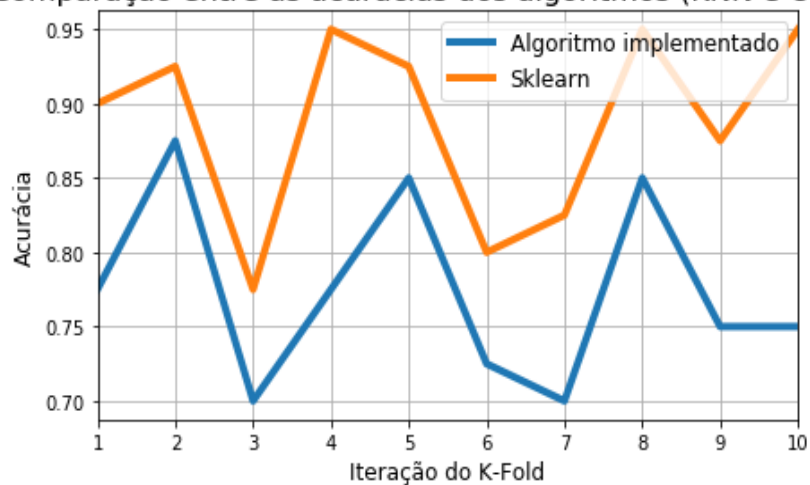


- Clusterização;
- Calcular k clusters a partir da média de distâncias;
- Base de dados utilizada: Wine dataset.

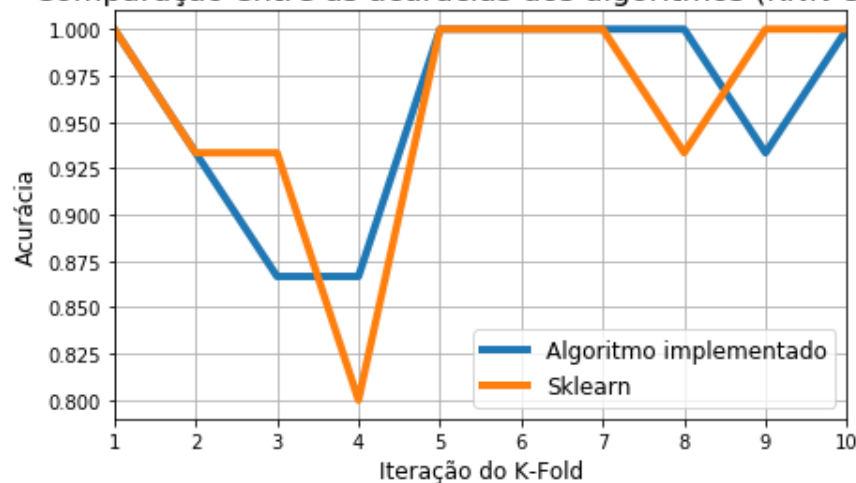
RESULTADOS E COMPARAÇÕES

Resultados KNN

Comparação entre as acurácias dos algoritmos (KNN e Olivetti)



Comparação entre as acurácias dos algoritmos (KNN e Iris)



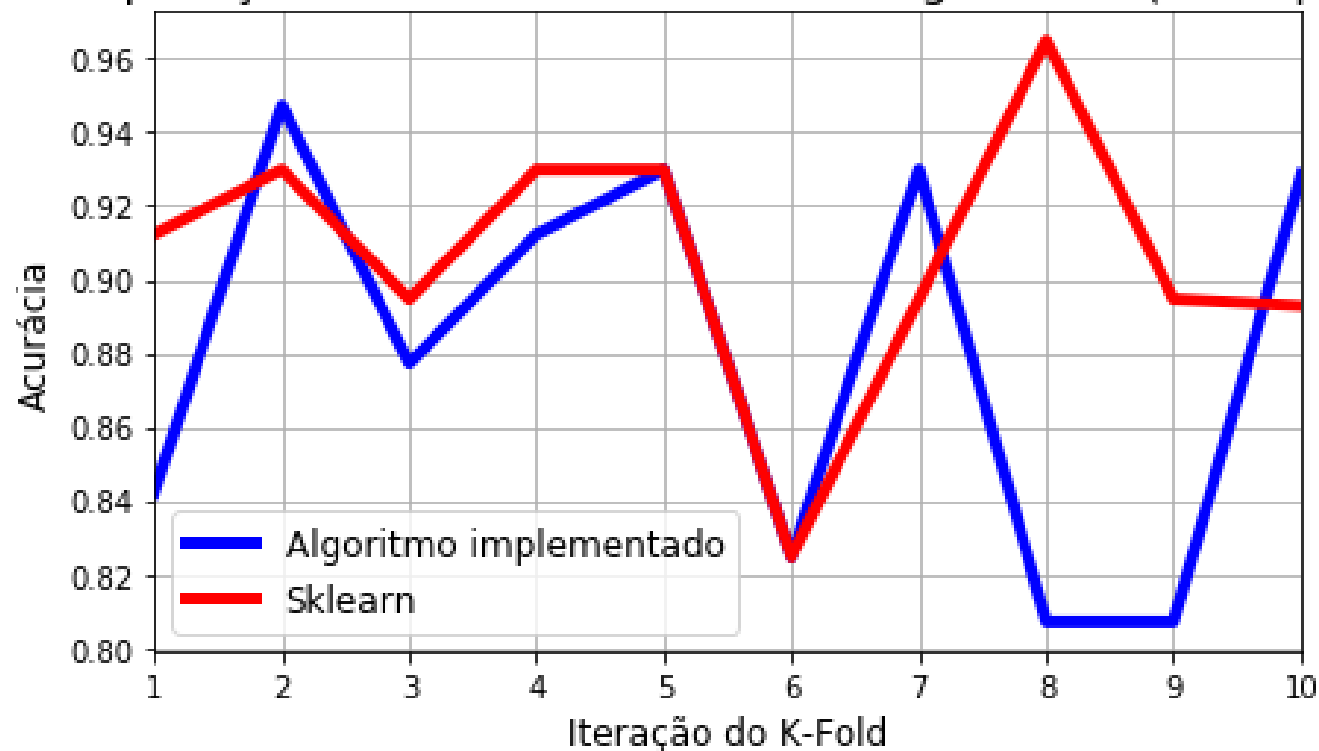
Resultados KNN

KNN					
Média das métricas utilizadas com K-Fold Cross Validation					
	Acurácia	F1 Score	Precisão	Recall	Dataset
Implementado	0,780	0,696	0,702	0,724	Olivetti Faces
Sklearn	0,870	0,826	0,831	0,840	
Implementado	0,960	0,960	0,966	0,964	Load_Iris
Sklearn	0,960	0,959	0,959	0,963	

- Duas bases de dados: Olivetti Faces e Iris;
- Diferenças de resultados entre bases de dados com mais classes e dimensões diferentes;
- Resultados superiores em algumas instâncias.

Resultados Perceptron

Comparação entre as acurácias dos algoritmos (Perceptron)

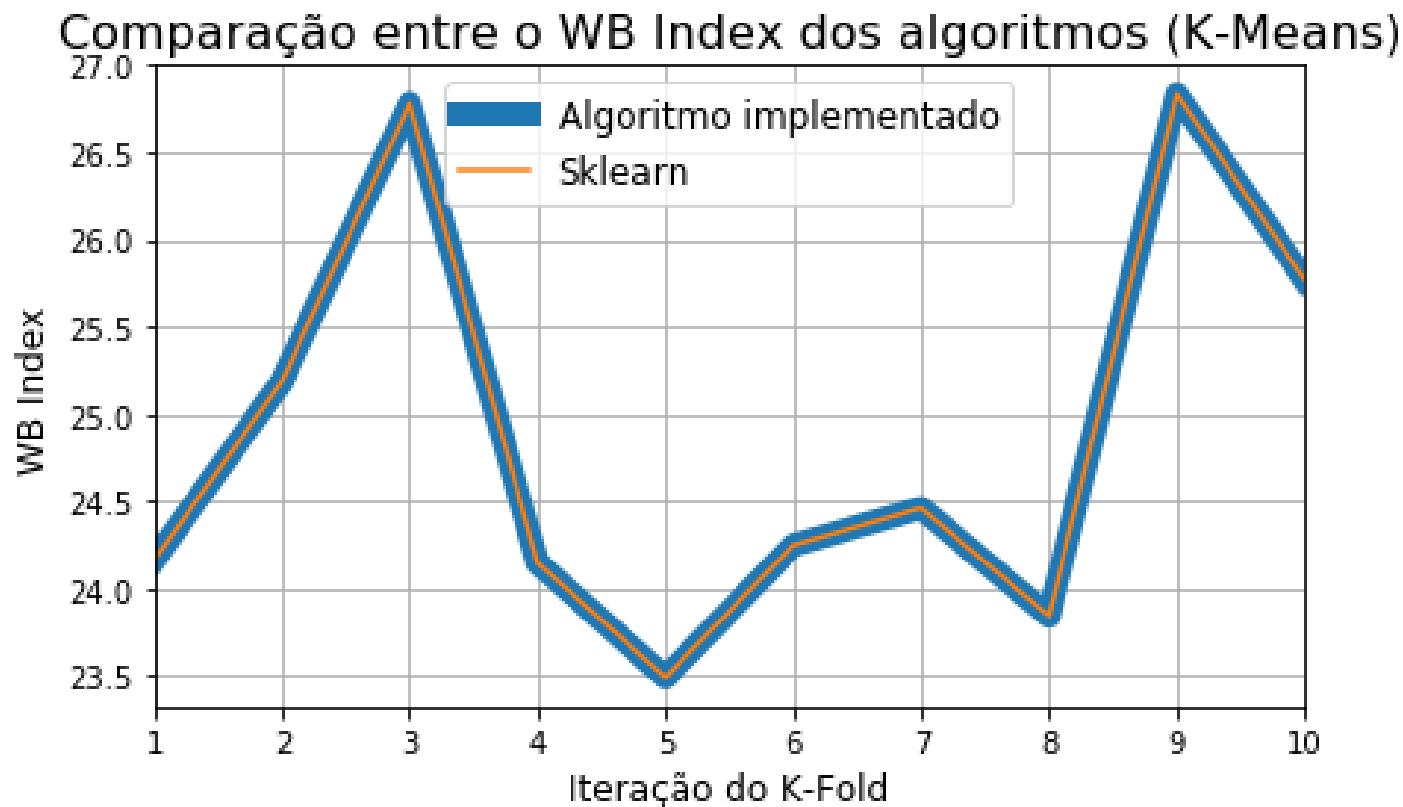


Resultados Perceptron

PERCEPTRON				
Média das métricas utilizadas com K-Fold Cross Validation				
	Acurácia	F1 Score	Precisão	Recall
Implementado	0,905	0,924	0,947	0,909
Sklearn	0,903	0,920	0,927	0,924

- Os códigos foram executados com parâmetros iguais;
- Resultados bastante parecidos com o Scikit Learn;
- O algoritmo implementado obteve resultados superiores em todas as métricas, exceto em Recall.

Resultados K-Means



- Três métricas consideradas: WB Index, Coeficiente de Silhueta e Calinski Score (Variance Ratio Criterion).
- Resultados exatamente iguais em todas instâncias realizadas;

- Os resultados almejados pela equipe foram alcançados;
- Implementações sem tantas funções como o Scikit Learn;
- Tempos de execução maiores;
- Tópicos que poderiam ser melhorados: mais funções, tratamento a grandes bases de dados, otimização, etc.



Obrigado pela atenção!