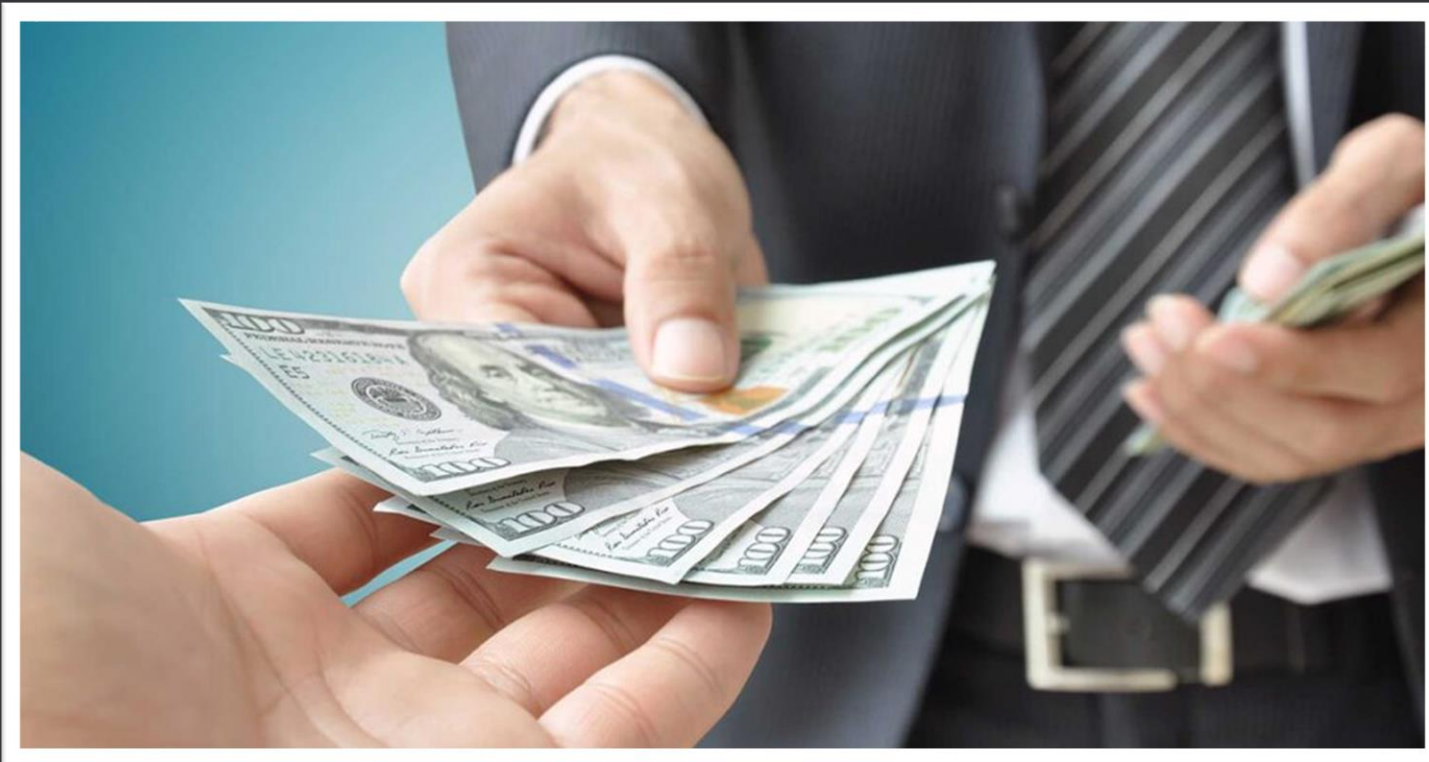


Classificação de assinatura eletrônica de empréstimos com base no histórico financeiro

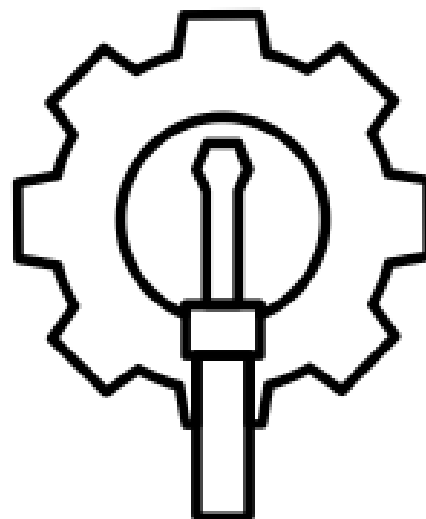
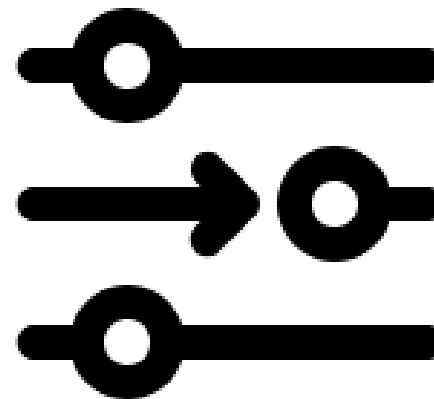


Componentes da Equipe:

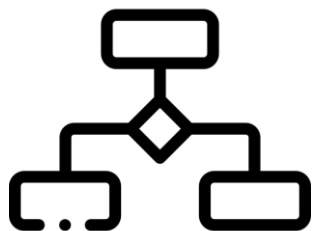
- Carlos Souza
- Rafael Kanda
- Saile Santos

Sprint 03

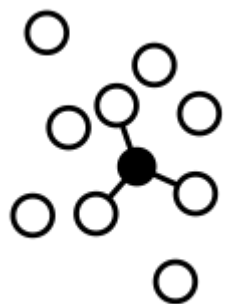
Com base nos resultados obtidos da Sprint anterior usar os 4 melhores modelos ajustando seus hiperparâmetros visando uma melhora nos algoritmos.



Algoritmos alterados:



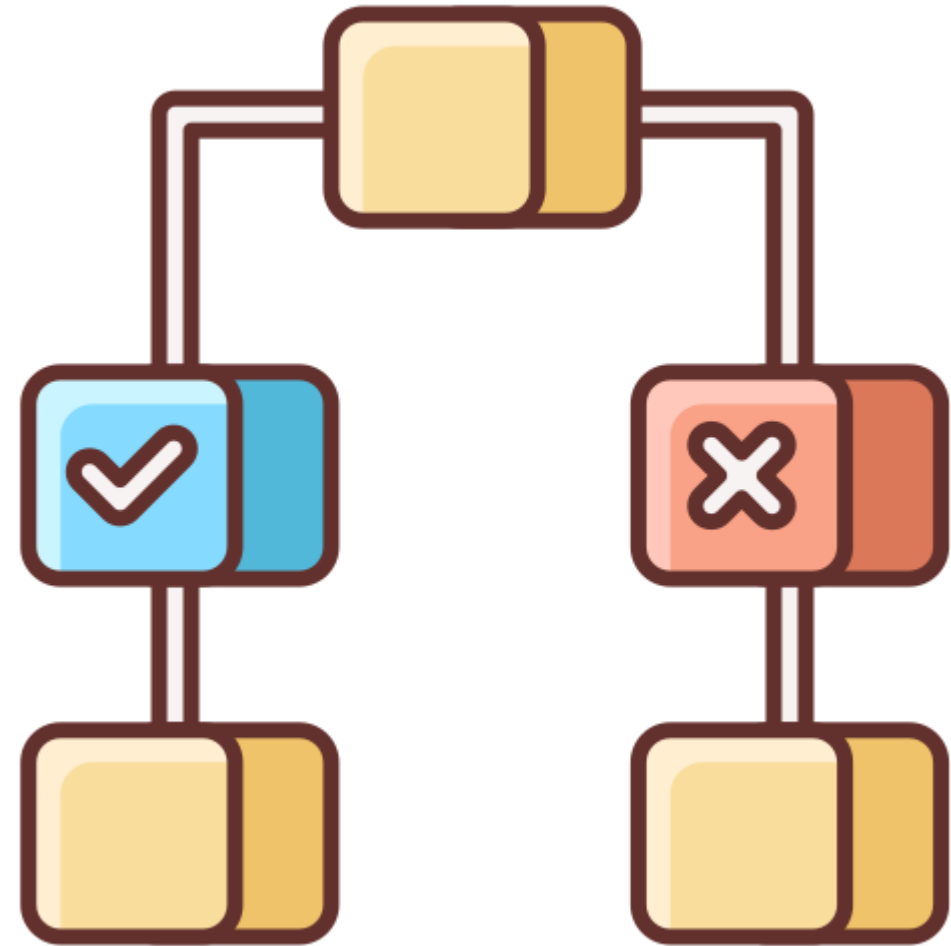
- Árvore de Decisão;
- Floresta Aleatória;
- KNN;
- SVC;



Árvore Binária

Parâmetros testados:

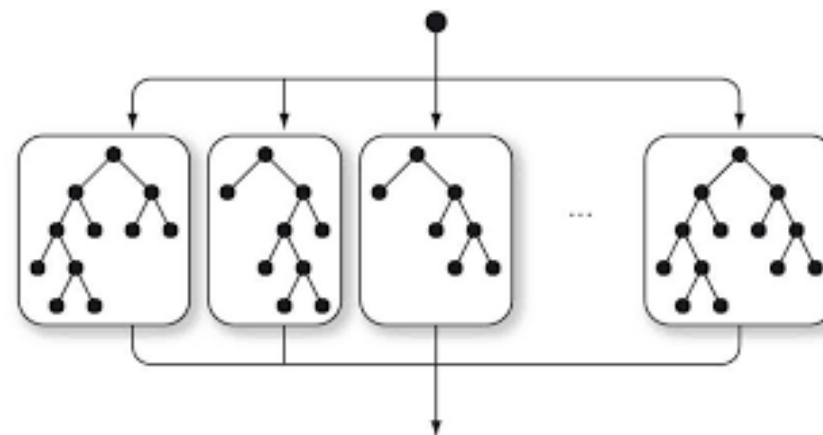
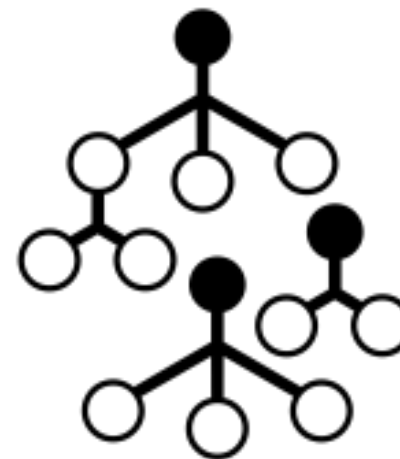
- `criterion = ['gini', 'entropy', 'log_loss'];`
- `max_features = ['auto', 'sqrt', 'log2'].`



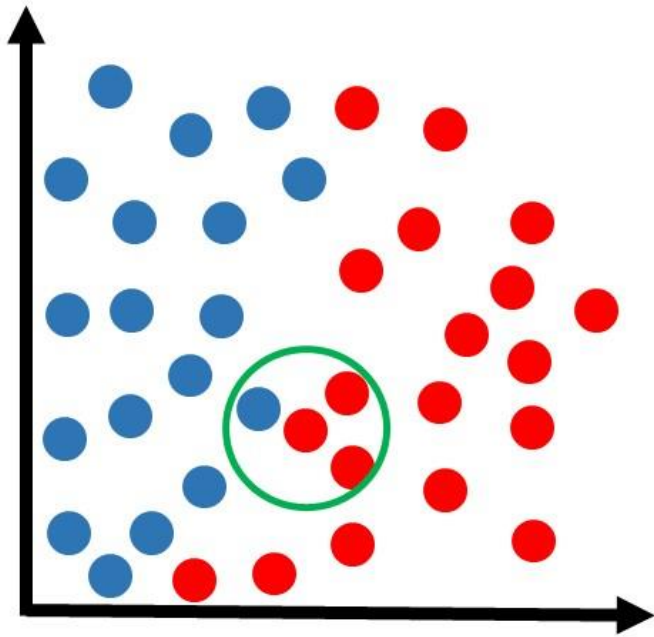
Floresta Aleatória

Parâmetros testados:

- `n_estimators = [10, 50, 100, 150, 200];`
- `criterion = ['gini', 'entropy', 'log_loss'];`
- `max_features = ['sqrt', 'log2'].`



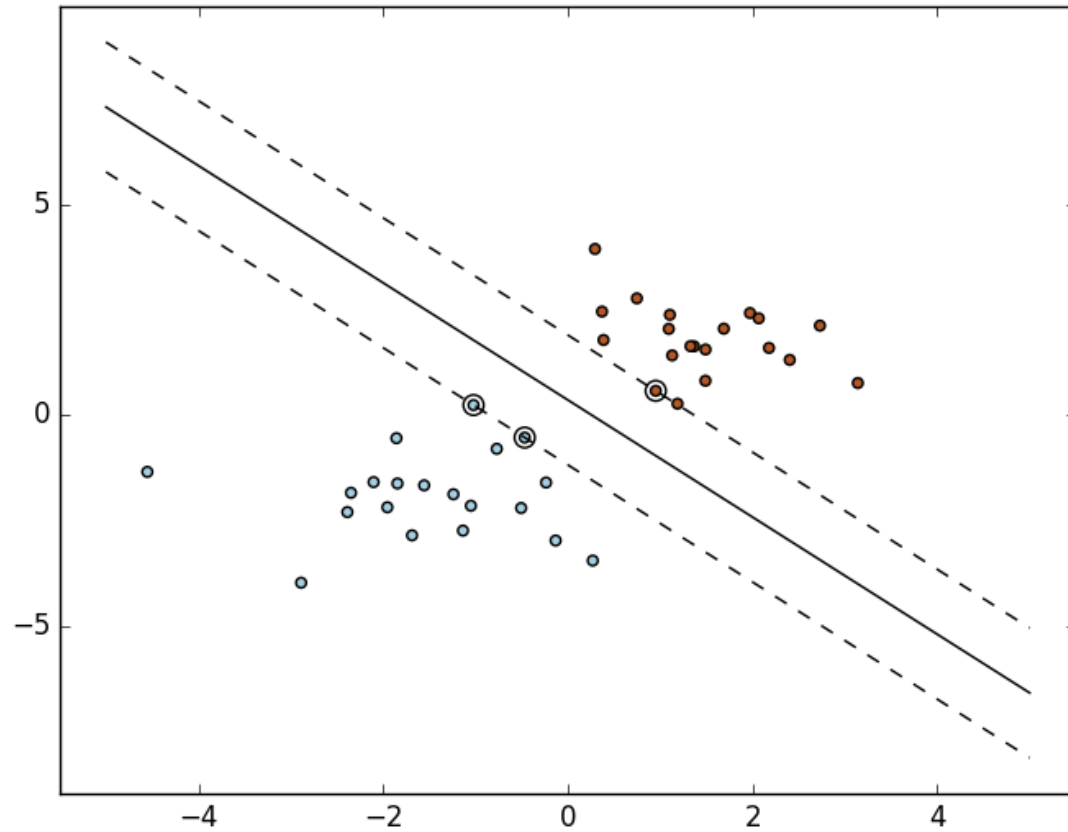
KNN



Paramêtros testados:

- `weights = ['uniform','distance'];`
- `algorithm = ['auto', 'ball_tree', 'kd_tree', 'brute'];`
- `metric = ['cityblock','euclidean','l1','l2','manhattan'].`

SVC



Paramêtros testados:

- $C = [0.1, 1, 10, 100]$;
- `kernel = ['linear', 'poly', 'rbf', 'sigmoid']`.

Resultados – Árvore Binária

Métricas	Algoritmo Sprint 02	Melhor configuração
Accuracy	0.5636	0.5727
F1 Score	0.5900	0.6059
Recall	0.5868	0.6139
Precision	0.5932	0.5981
EQM	0.4364	0.4223

Usando:

- `criterion = 'entropy';`
- `max_features = 'sqrt'.`

Resultados – Floresta Aleatória

Métricas	Algoritmo Sprint 02	Melhor configuração
Accuracy	0.6010	0.6060
F1 Score	0.6370	0.6409
Recall	0.6543	0.6570
Precision	0.6206	0.6255
EQM	0.3990	0.3940

Usando:

- `n_estimators = 200;`
- `criterion = 'gini';`
- `max_features = 'sqrt'.`

Resultados – KNN

Métricas	Algoritmo Sprint 02	Melhor configuração
Accuracy	0.5552	0.5595
F1 Score	0.5968	0.5998
Recall	0.6153	0.6170
Precision	0.5794	0.5836
EQM	0.4448	0.4405

Usando:

- weights = 'distance';
- algorithm = 'brute';
- metric = 'euclidean' or 'l2'.

Resultados – SVC

Métricas	Algoritmo Sprint 02	Melhor configuração
Accuracy	0.5503	0.5362
F1 Score	0.6842	0.6976
Recall	0.9103	1.0000
Precision	0.5481	0.5357
EQM	0.4497	0.4638

Usando:

- $C = '100'$;
- $\text{kernel} = \text{'poly'}$.

Conclusão

O processo de ajuste de parâmetros apresentou alguns resultados positivos, mesmo que pequenos. O KNN foi o algoritmo que as menores melhoras enquanto a Floresta Aleatória e o SVC apresentaram os melhores resultados.

Apesar disso, os algoritmos ainda não estão em uma eficiência adequada, onde o melhor apresentou apenas um F1 Score de quase 70%, o que não é suficiente para a utilização em um produto real.

Link do Trabalho

Acesso pelo GitHub:

<https://github.com/Equipe5-ifam-cmzl/trabalho-final>

