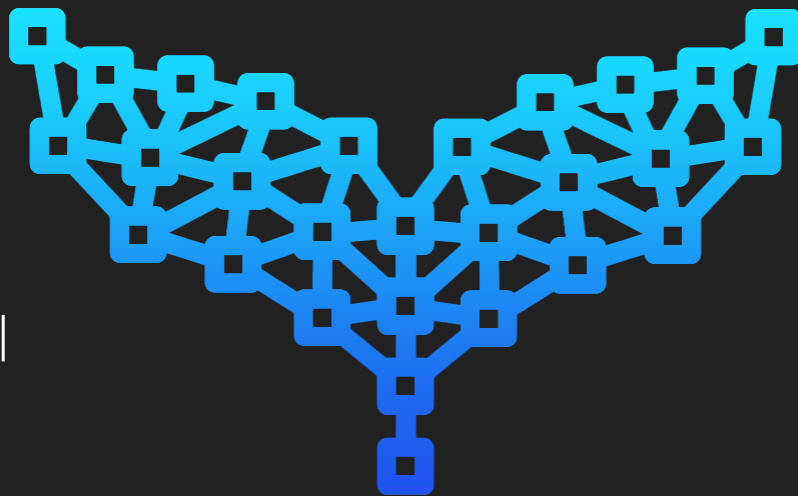
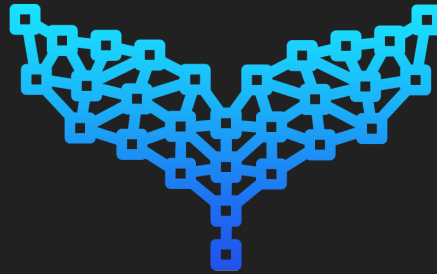


TAIL

Technology and Artificial
Intelligence League



RANDOM FOREST





Sumário

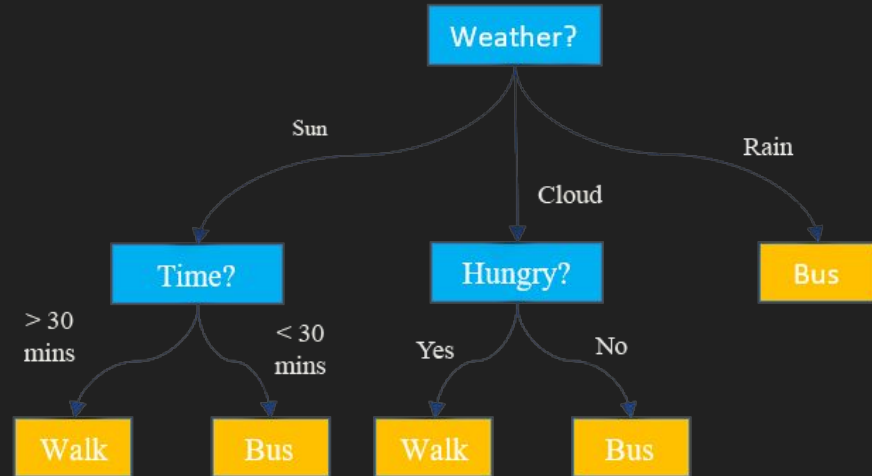
1. Árvore de Decisão	4
2. Passo a passo de uma Random Forest.....	5
3. Árvore de Decisão vs. Random Forest.....	6
4. A eficiência da baixa correlação.....	8
5. Como garantir a baixa correlação?.	9
a. Bagging (Bootstrap Aggregation)	
b. Feature Randomness	
6. Validando o random forest com OOB.....	12
7. Referências Bibliográficas.....	13

Árvore de Decisão



Clique aqui para
acessar a aula de
Árvore de Decisão da
TAIL!

- Grau de Impureza (Entropia) e Ganho de Informação;
- Fácil de visualizar e entender;
- Inflexível.

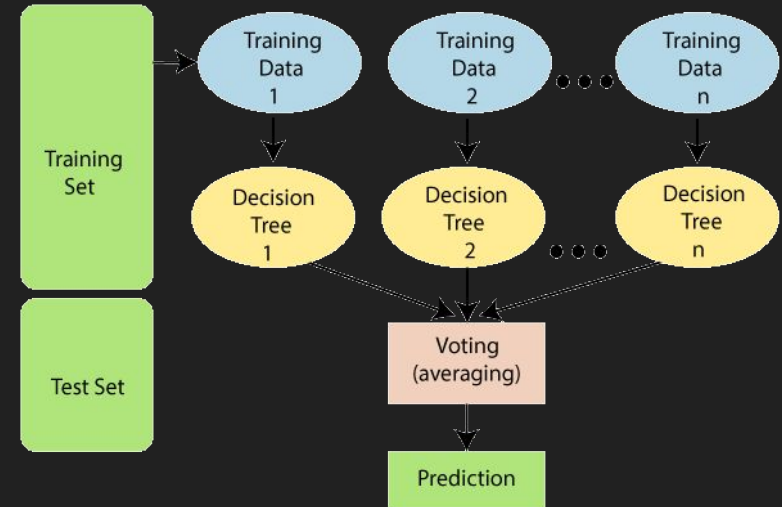


Fonte: <https://www.sqlshack.com/microsoft-decision-trees-in-sql-server/>



Passo a passo de uma Random Forest

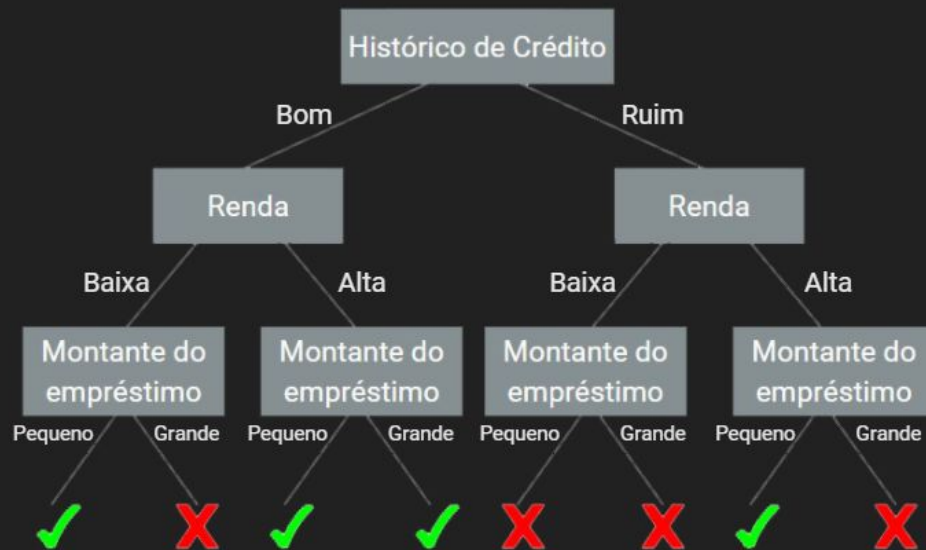
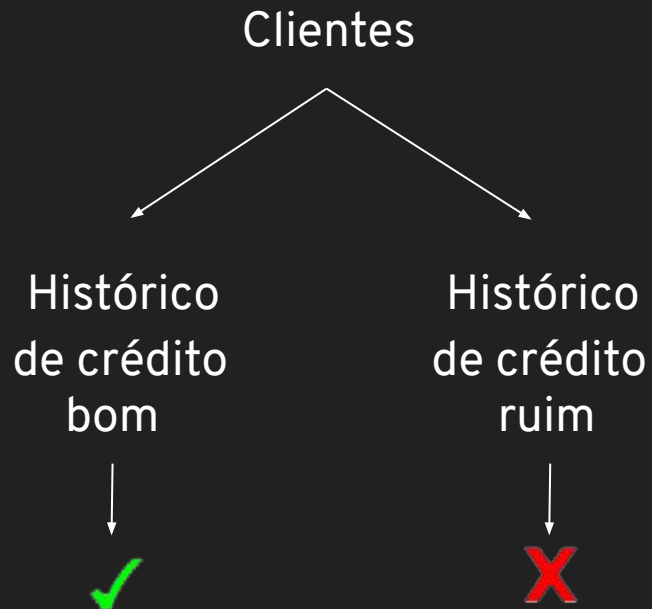
- Criação do Bootstrap Dataset;
- Seleção aleatória dos atributos para a composição da árvore;
- Repetição dos dois primeiros passos várias vezes;
- Verificação dos resultados retornados pelas árvores.



Fonte: <https://www.javatpoint.com/machine-learning-random-forest-algorithm>

Árvore de Decisão vs. Random Forest

Exemplo: Análise de Crédito de



Árvore de Decisão vs. Random Forest

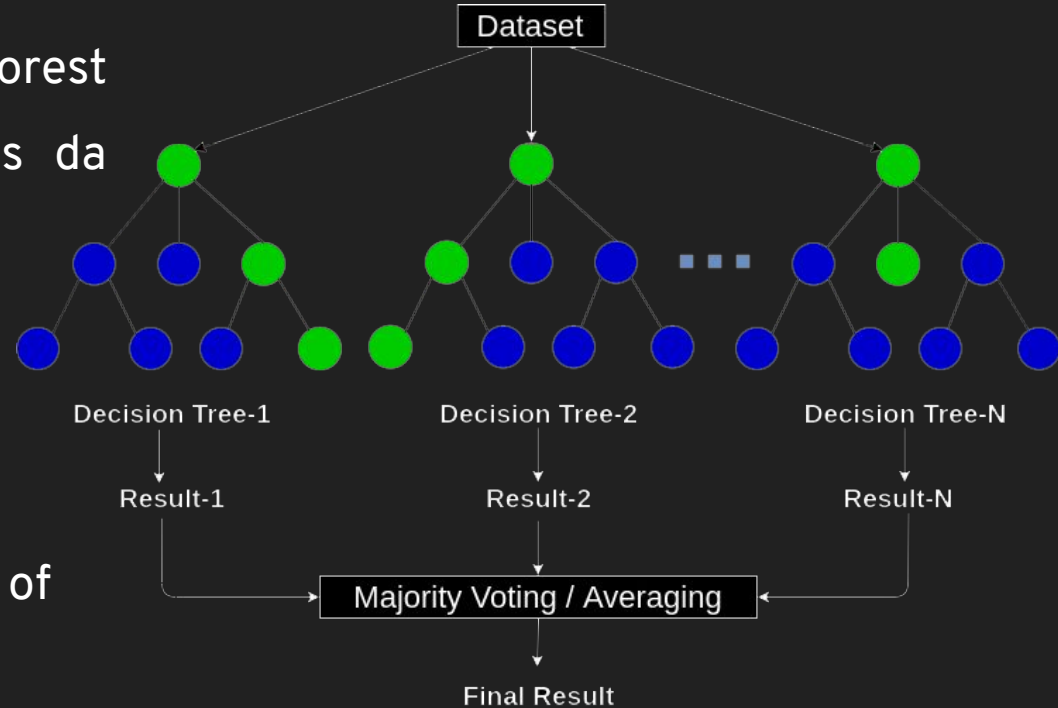
- Perguntas:
 - Por que, na árvore de decisão, um feature é analisado antes do outro?
 - O que faz o modelo de Random Forest ser mais indicado em alguns casos ao invés da Árvore de Decisão?





A eficiência da baixa correlação

- Como uma Random Forest consegue ser flexível através da baixa correlação?
- Cada árvore de decisão é um evento independente
- Teoria da probabilidade:
Lei dos grandes números (Law of Large Numbers)

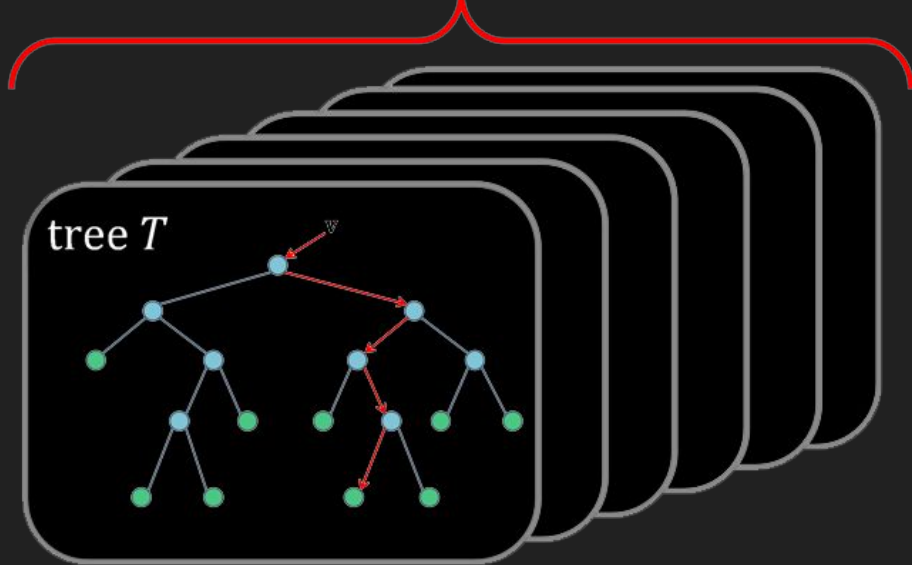


Fonte: <https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/>



Como garantir a baixa correlação?

Random Forest



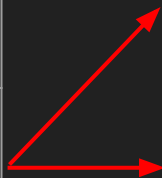
- A resposta está na construção de cada árvore do Random Forest
- Transformando o dataset:
Bagging (Bootstrap Aggregation)
- Transformando o subconjunto de nós:
Feature Randomness

Fonte: <https://dimensionless.in/introduction-to-random-forest/>



Bagging (Bootstrap Aggregation)

Idade	Peso	Altura	Obeso
22	70	150	Sim
30	130	180	Sim
18	55	155	Não
45	80	165	Sim



Idade	Peso	Altura	Obeso
22	70	150	Sim
18	55	155	Não
18	55	155	Não
45	80	165	Sim



Feature Randomness

Idade	Peso	Altura	Obeso
22	70	150	Sim
30	130	180	Sim
18	55	155	Não
45	80	165	Sim

Idade	Peso	Altura	Obeso
22	70	150	Sim
30	130	180	Sim
18	55	155	Não
45	80	165	Sim



Validando o Random Forest com OOB

- Uma das maneiras de estimar a acurácia do Random Forest é com o *OOB score (out-of-bag score)*
- Um novo *dataset* é criado com os dados que foram deixados de lado em cada árvore de decisão na etapa de *Bagging*
- O nome deste novo *dataset* é *Out-of-bag sample*
- Utilizando o *Out-of-bag sample* nas árvores de decisão que não possuem estes dados, para verificar cada resposta
- O *OOB score* será medido pelas previsões corretas de cada *Out-of-bag sample*



Referências Bibliográficas

SHARMA, Abhishek. Decision Tree vs. Random Forest – Which Algorithm Should you Use?. Disponível em: <<https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/>>. Acesso em 15 de Dezembro de 2020;

YIU, Tony. Understanding Random Forest. Disponível em: <<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>>. Acesso em 15 de Dezembro de 2020;

O que é e como funciona o algoritmo RandomForest. Disponível em: <<https://didatica.tech/o-que-e-e-como-funciona-o-algoritmo-randomforest/>>. Acesso em 15 de Dezembro de 2020;

STARMER, Josh. StatQuest: Random Forests Parte 1 - Construindo, Usando e Avaliando. Disponível em: <https://www.youtube.com/watch?v=J4Wdy0Wc_xQ&t=490s&ab_channel=StatQuestwithJoshStarmer>. Acesso em 15 de Dezembro de 2020;