

MBA⁺

Artificial Intelligence &
Machine Learning





Ensembles

Bagging & Boosting

Ensembles



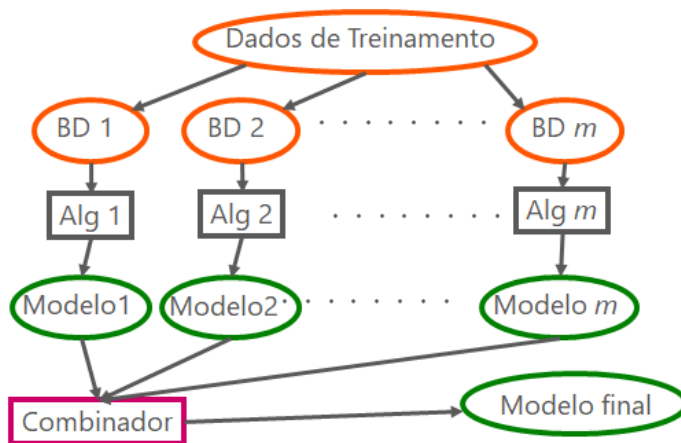
- Para a previsão do tempo, considere que cada coluna representa um dia da semana:

Realidade							
Modelo 1							
Modelo 2							
Modelo 3							
Modelo 4							
Modelo 5							
Combinação							



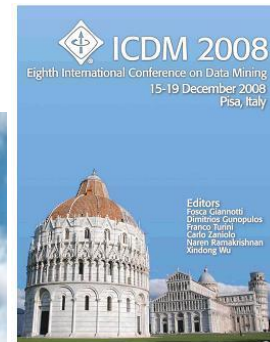
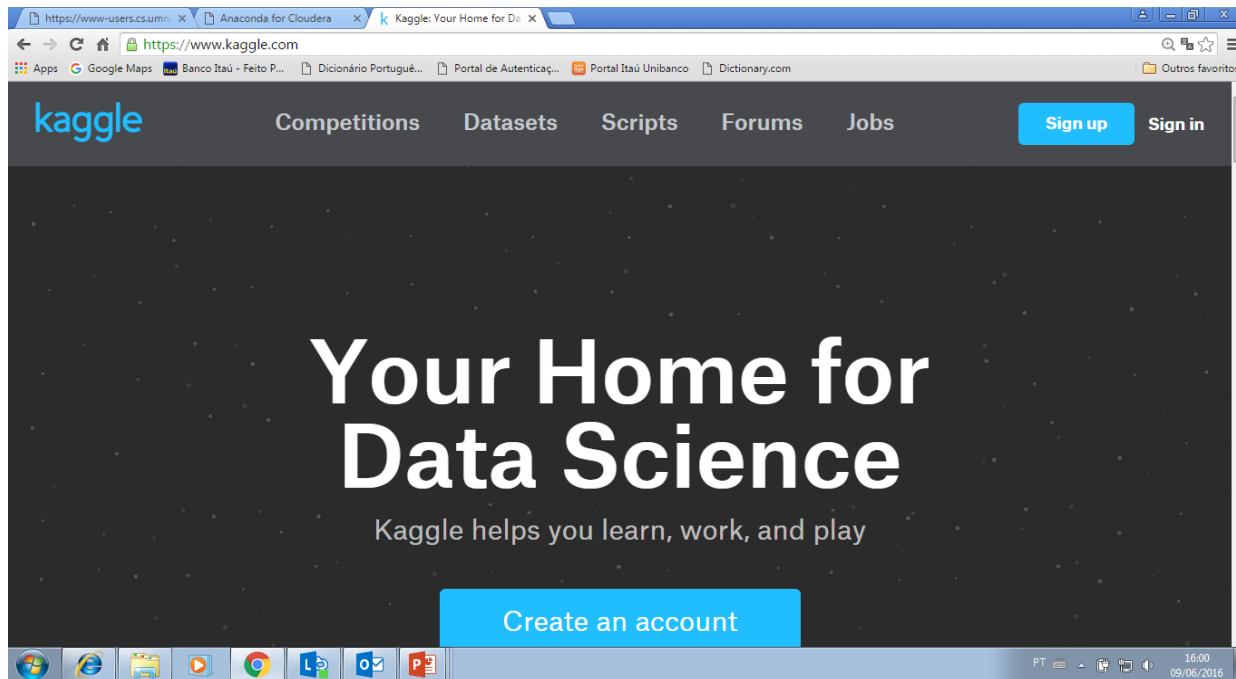
Ensembles

- Usar vários modelos para melhorar a acurácia preditiva;
- A combinação de modelos diversificados e independentes nos permite tomar melhores decisões;

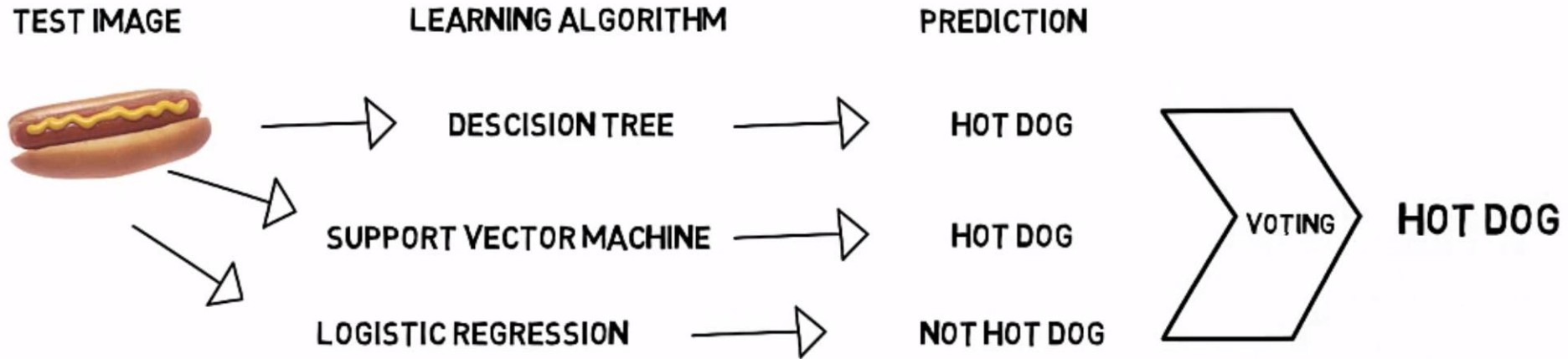


Nota: BDs podem ser iguais ou diferentes.

Ensembles – sucesso na prática



Ensemble



O que fazer na prática?



- Mesmos algoritmos com parâmetros diferentes (e.g., k-NN, redes neurais etc.); **Árvores são melhores que k-NN???**
- Bases de dados com diferentes atributos;
- Subconjuntos da mesma base de dados (*bagging*);
- Reponderar a base de treinamento (*boosting*);

Ensembles

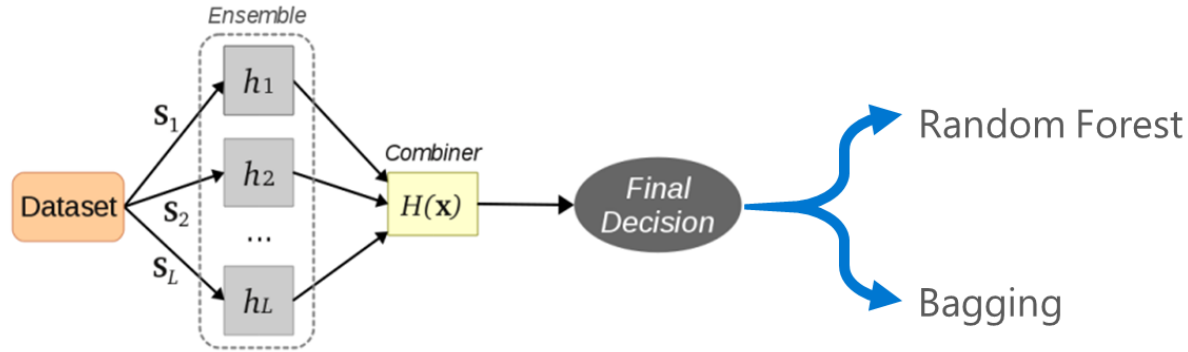


Fig. 1. Parallel architecture

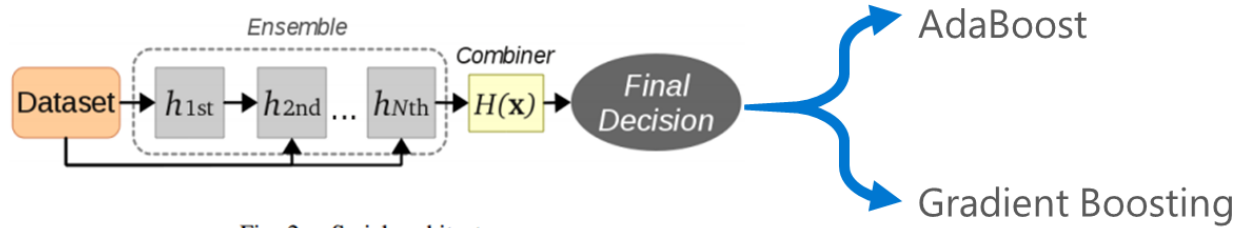


Fig. 2. Serial architecture

Ponti Jr, Moacir P. "Combining Classifiers: from the creation of ensembles to the decision fusion."



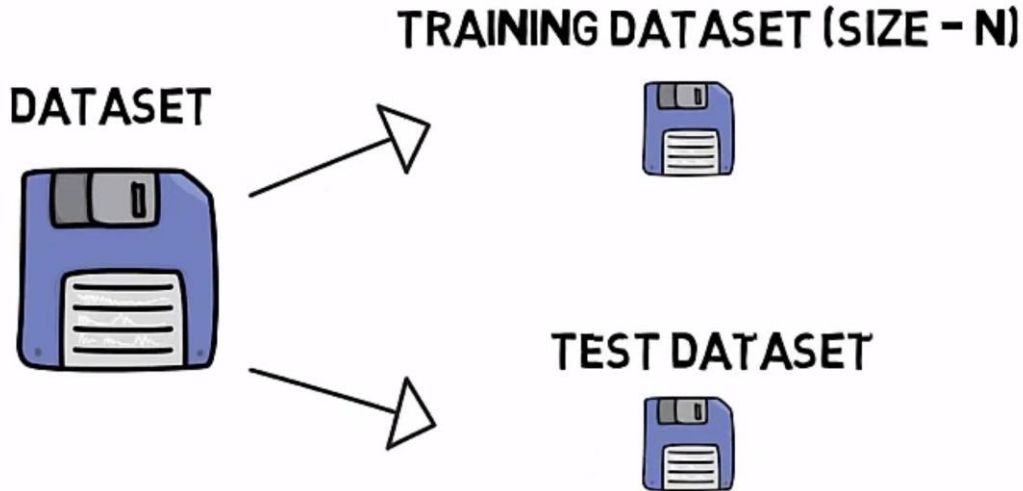
Bagging

- **Algoritmo de Bootstrap Aggregation** (Brieman, 1996):
 - 1) Amostrar, M vezes, N exemplos da base de dados (com reposição);
 - 2) Treinar M classificadores (um para cada amostra);
 - 3) Combinar os classificadores via voto majoritário.
- Passo 1) insere variância nas bases de treinamento dos componentes, aumentando a estabilidade do *ensemble*.
- Espera-se que em cada amostra existam 63,2% de tuplas não repetidas;
- *Random Forests* são baseadas nessa ideia.

Bootstrap Aggregating (Bagging)

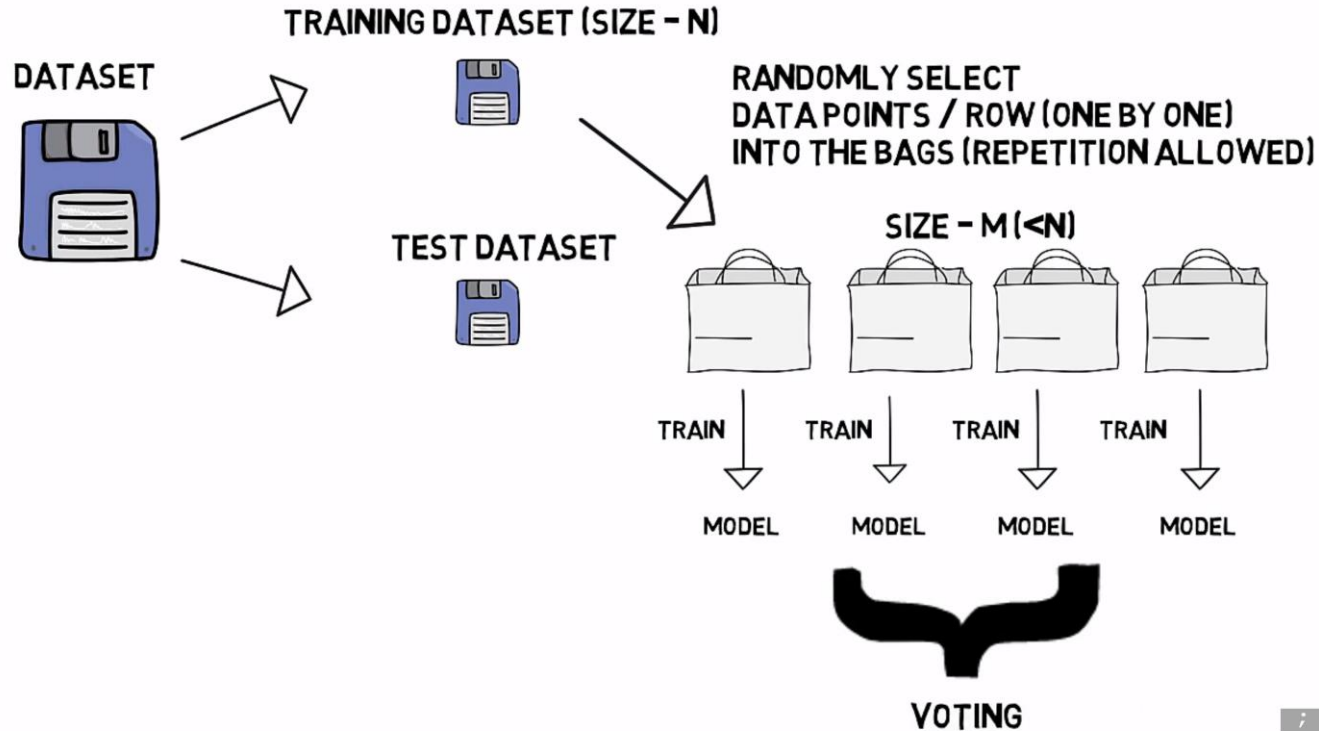


MULTIPLE MODELS OF SAME LEARNING ALGORITHM TRAINED WITH SUBSETS OF DATASET RANDOMLY PICKED FROM THE TRAINING DATASET.





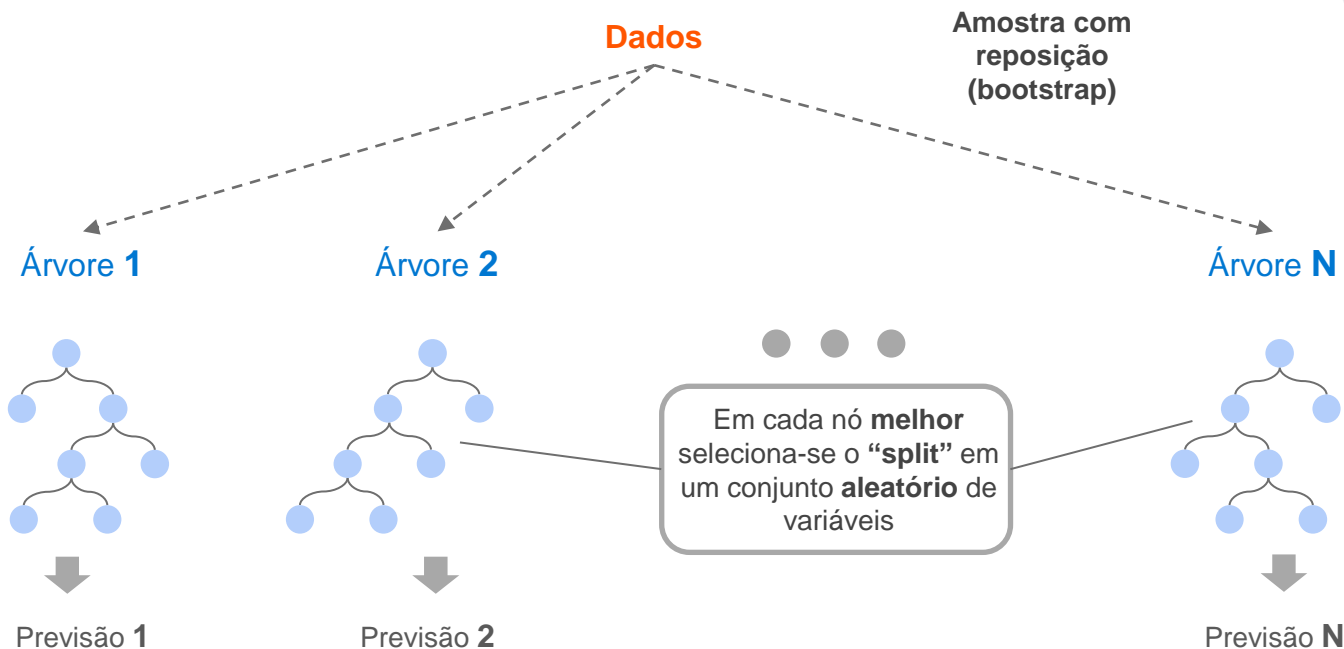
Bootstrap Aggregating (Bagging)



;



Random Forest

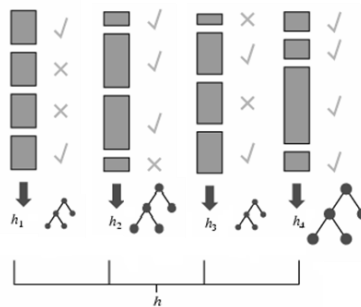


$$\text{Previsão Final} = \frac{\sum_i^n \text{Previsão}_i}{n}$$



Boosting

- Em vez de reamostrar, repondera exemplos;
- Cada iteração induz um classificador e repondera exemplos;
- Ensemble é baseado no voto ponderado (acurácia) dos componentes.

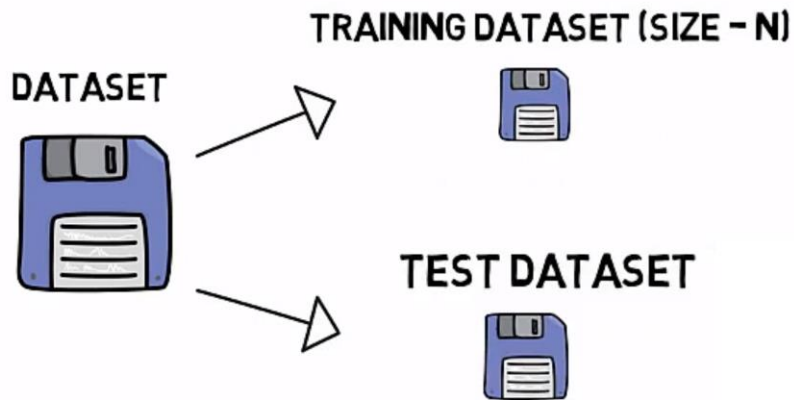


- Cada retângulo representa um exemplo;
- Tamanho da árvore reflete a acurácia e indica seu peso no ensemble.



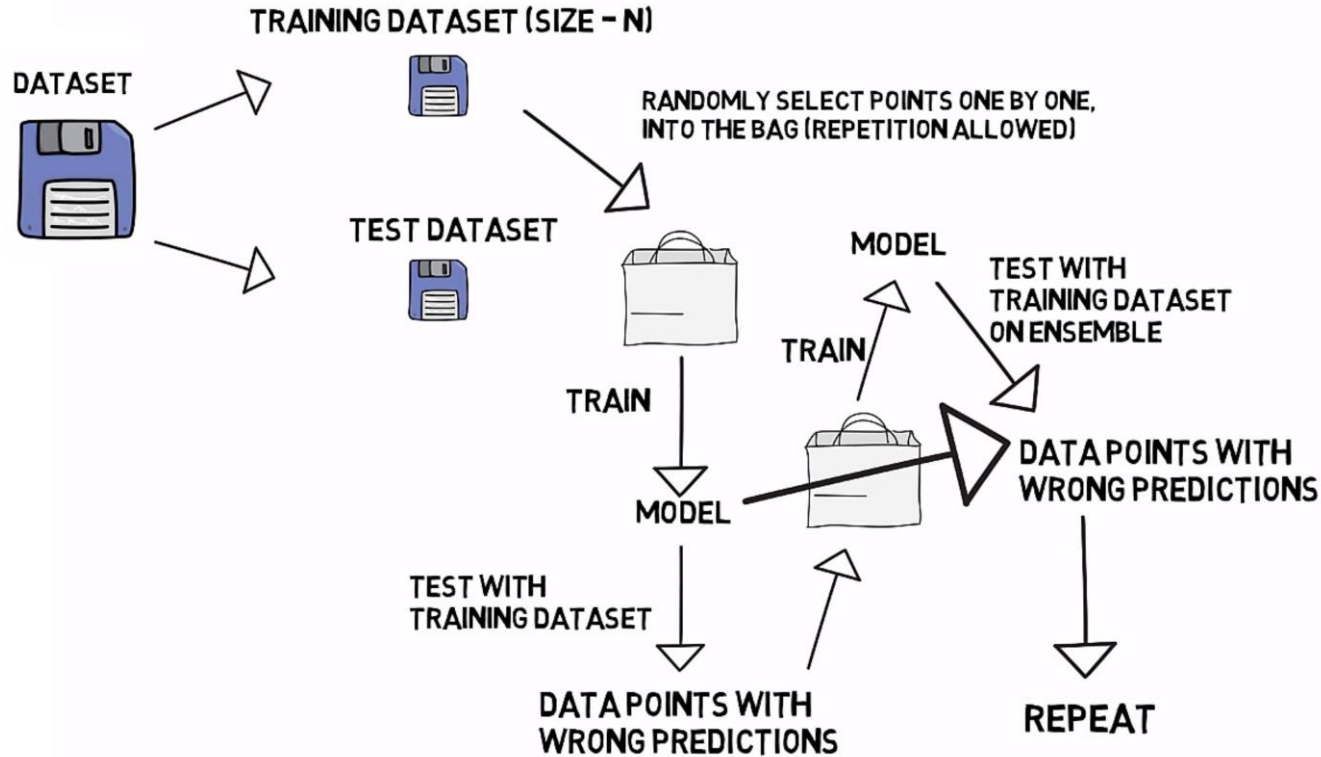
Boosting

- LITTLE VARIATION ON BAGGING
- SELECTING POINTS WHICH GIVE WRONG PREDICTIONS.



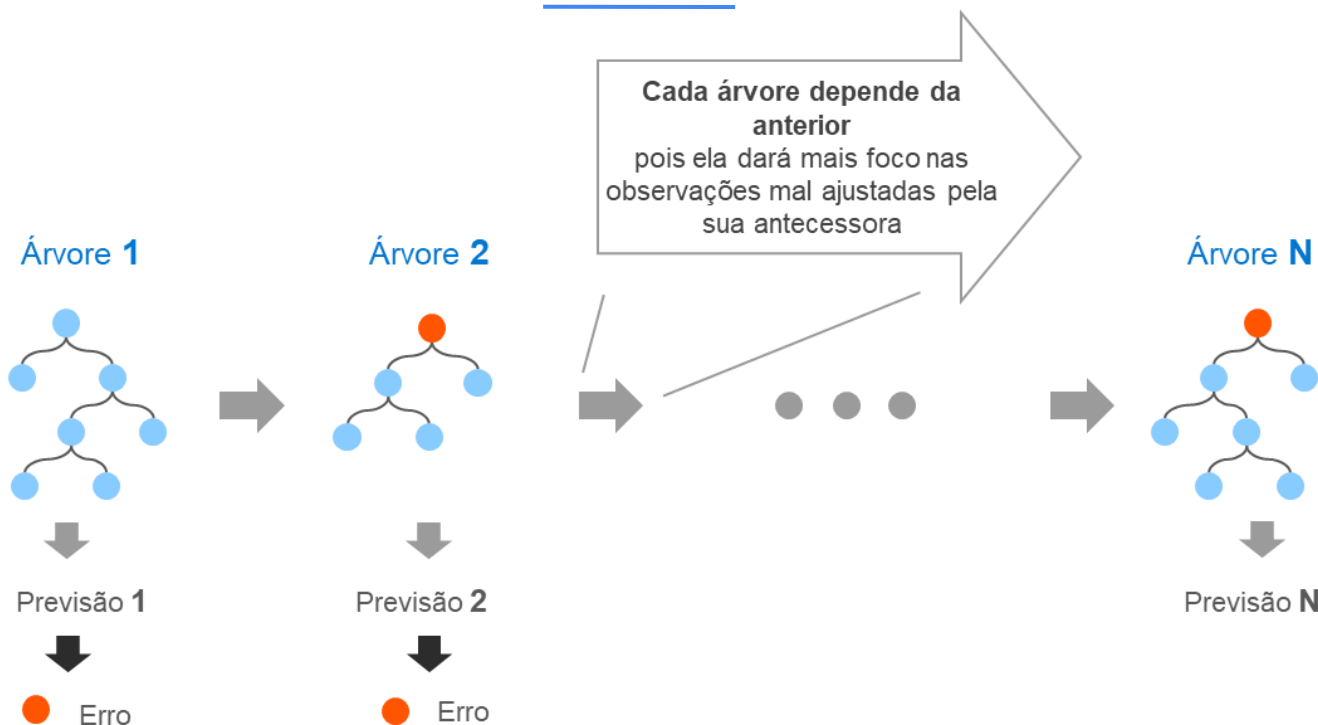


Boosting





Gradient Boosting



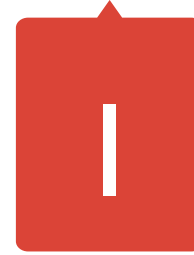
$$\text{Previsão Final} = w_1 \text{Previsão}_1 + \dots + w_N \text{Previsão}_N$$

Questions and Feedback



[Thank you!](#)

Obrigado !



Vinicius Fernandes Caridá
vfcarida@gmail.com



@Vinicius Caridá



@Vinicius Caridá
@machine learning Brasil



@vfcarida

MBA⁺

Copyright © 2018 **Prof. Vinicius Fernandes Caridá**
 Todos direitos reservados. Reprodução ou divulgação
 total ou parcial deste documento é expressamente
 proibido sem o consentimento formal, por escrito, do
 Professor (autor).