



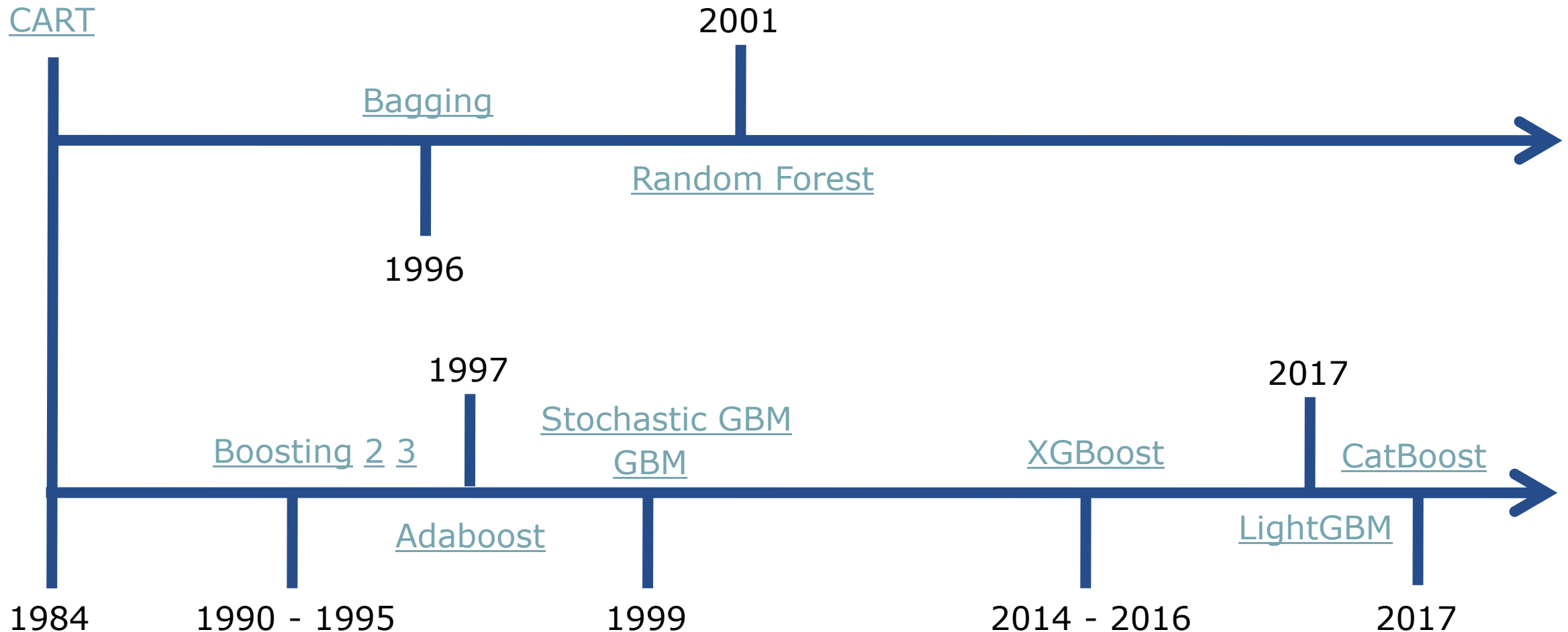
BC  
TEC

escola  
britânica de  
artes criativas  
& tecnologia

**Profissão Cientista de Dados  
eXtreme Gradient  
Boosting - XGBoost**

# Por que estamos aprendendo nessa ordem?

Pois essa foi a ordem das invenções.



# XGBoost – Extreme Gradient Boosting

[XGBoost: A Scalable Tree Boosting System](#)  
[XGBoost Documentation](#)

The impact of the XGBoost has been widely recognized in a number of machine learning and data mining challenges. Take the challenges hosted by the machine learning competition site Kaggle for example. Among the 29 challenge winning solutions published at Kaggle's blog during 2015, 17 solutions used XGBoost. Among these solutions, eight solely used XGBoost to train the model, while most others combined XGBoost with neural nets in ensembles. For comparison, the second most popular method, deep neural nets, was used in 11 solutions. The success of the system was also witnessed in KDDCup 2015, where XGBoost was used by every winning team in the top-10. These results demonstrate that our system gives state-of-the-art results on a wide range of problems. The most important factor behind the success of XGBoost is its scalability in all scenarios. The system runs more than ten times faster than existing popular solutions on a single machine and scales to billions of examples in distributed or memory-limited settings. The scalability of XGBoost is due to several important systems and algorithmic optimizations. These innovations include: a novel tree learning algorithm is for handling sparse data; a theoretically justified weighted quantile sketch procedure enables handling instance weights in approximate tree learning. Parallel and distributed computing makes learning faster which enables quicker model exploration.

**Chen e Guestrin, 2016**

# XGBoost – Extreme Gradient Boosting

[XGBoost: A Scalable Tree Boosting System](#)  
[XGBoost Documentation](#)

A performance do XGBoost foi amplamente percebido em vários desafios de *Machine Learning* e mineração de dados. Por exemplo, os desafios hospedados pelo site Kaggle, de competição de *Machine Learning*. **Entre as 29 soluções vencedoras** do desafio publicadas no Kaggle durante 2015, **17 soluções usaram o XGBoost**. Para comparação, o segundo método mais popular, **redes neurais profundas, foi usado em 11 soluções**. O sucesso do XGBoost também foi testemunhado no **KDDCup 2015**, onde o **XGBoost** foi usado por **todas as equipes** vencedoras do **top 10**. O fator mais importante por trás do sucesso do **XGBoost é sua escalabilidade em todos os cenários. O XGBoost é executado mais de dez vezes mais rápido** do que as soluções populares existentes em uma única máquina e é dimensionado para bilhões de exemplos em configurações distribuídas ou com limitação de memória. **A escalabilidade do XGBoost se deve a vários sistemas importantes e otimizações algorítmicas**. Essas inovações incluem: um novo algoritmo de aprendizado de árvore para lidar com dados esparsos, um procedimento de quartil ponderado e etc. A computação paralela e distribuída torna o aprendizado mais rápido, o que permite uma exploração mais rápida do modelo.

**Chen e Guestrin, 2016**