



O C4.5 constrói árvores de decisão a partir de um conjunto de dados de treinamento da mesma forma que o algoritmo ID3, utilizando o conceito de Entropia, apresentando uma série de melhorias ao algoritmo ID3. Algumas dessas melhorias são:

- Lidar com atributos contínuos e discretos. No intuito de lidar com atributos contínuos, o algoritmo C4.5 cria um valor limiar e então particiona o conjunto de dados em dois subconjuntos dos quais um contém valores de atributos maiores do que aquele valor limiar e o outro conjunto contém valores menores ou iguais aquele valor limiar.
- Lidar com dados de treinamento com atributos incompletos. O algoritmo C4.5 permite que atributos sejam rotulados como ? para casos onde os valores não estejam presentes. Valores de atributos que não estejam presentes não são utilizados em cálculos de entropia ou ganho de informação.
- Lidar com atributos com diferentes custos.
- Poda de árvores após a criação. O algoritmo C4.5 retrocede pela árvore quando esta é criada e tenta remover ramificações que não ajudam no processo de decisão e substitui estes ramos por nós folha.