**So sánh kết quả 2 thuật toán :**

**Test 1:**

**KNN :**

**Text

Description automatically generated**

DECISION TREE :

**Text

Description automatically generated**

**Test 2:**

**KNN:**

**Text, letter

Description automatically generated**

DECISION TREE :

**Text

Description automatically generated**

* **Kết luận :**
* Trong 1 số trường hợp 2 thuật toán có dự đoán giống nhau
* Trong 1 số trường hợp khác thì hai thuật toán dự đoán kết quả chênh lệch

**Ưu điểm và nhược điểm của thuật toán KNN:**

* **Ưu điểm của KNN**

1. Độ phức tạp tính toán của quá trình training là bằng 0.
2. Việc dự đoán kết quả của dữ liệu mới rất đơn giản.
3. Không cần giả sử gì về phân phối của các class.

* **Nhược điểm của KNN**

1. KNN rất nhạy cảm với nhiễu khi K nhỏ.
2. Như đã nói, KNN là một thuật toán mà mọi tính toán đều nằm ở khâu test. Trong đó việc tính khoảng cách tới *từng* điểm dữ liệu trong training set sẽ tốn rất nhiều thời gian, đặc biệt là với các cơ sở dữ liệu có số chiều lớn và có nhiều điểm dữ liệu. Với K càng lớn thì độ phức tạp cũng sẽ tăng lên. Ngoài ra, việc lưu toàn bộ dữ liệu trong bộ nhớ cũng ảnh hưởng tới hiệu năng của KNN.

**Ưu điểm và nhược điểm của thuật toán** DECISION TREE**:**

* **Ưu điểm** DECISION TREE:

So với các phương pháp khai phá dữ liệu khác, cây quyết định là phương pháp có một số ưu điểm:

1. **Cây quyết định dễ hiểu**. Người ta có thể hiểu mô hình cây quyết định sau khi được giải thích ngắn.
2. **Việc chuẩn bị dữ liệu cho một cây quyết định là cơ bản hoặc không cần thiết**. Các kỹ thuật khác thường đòi hỏi chuẩn hóa dữ liệu, cần tạo các biến phụ (dummy variable) và loại bỏ các giá trị rỗng.
3. **Cây quyết định có thể xử lý cả dữ liệu có giá trị bằng số và dữ liệu có giá trị là tên thể loại.** Các kỹ thuật khác thường chuyên để phân tích các bộ dữ liệu chỉ gồm một loại biến. Chẳng hạn, các luật quan hệ chỉ có thể dùng cho các biến tên, trong khi mạng nơ-ron chỉ có thể dùng cho các biến có giá trị bằng số.
4. **Cây quyết định là một mô hình hộp trắng.** Nếu có thể quan sát một tình huống cho trước trong một mô hình, thì có thể dễ dàng giải thích điều kiện đó bằng logic Boolean. Mạng nơ-ron là một ví dụ về mô hình hộp đen, do lời giải thích cho kết quả quá phức tạp để có thể hiểu được.
5. **Có thể thẩm định một mô hình bằng các kiểm tra thống kê.** Điều này làm cho ta có thể tin tưởng vào mô hình.
6. **Cây quyết định có thể xử lý tốt một lượng dữ liệu lớn trong thời gian ngắn**. Có thể dùng máy tính cá nhân để phân tích các lượng dữ liệu lớn trong một thời gian đủ ngắn để cho phép các nhà chiến lược đưa ra quyết định dựa trên phân tích của cây quyết định.

* **Nhược điểm** DECISION TREE:

1. Khó giải quyết được những vấn đề có dữ liệu phụ thuộc thời gian liên tục
2. Dễ xảy ra lỗi khi có quá nhiều lớp chi phí tính toán để xây dựng mô hình cây quyết định CAO.

Ứng dụng KNN và DECISION TREE :

KNN:

* KNN là một mô hình đơn giản và trực quan nhưng vẫn có hiệu quả cao vì nó không tham số; mô hình không đưa ra giả định nào về việc phân phối dữ liệu. Hơn nữa, nó có thể được sử dụng trực tiếp để phân loại đa lớp.
* Thuật toán KNN có nhiều ứng dụng trong ngành đầu tư, bao gồm dự đoán phá sản, dự đoán giá cổ phiếu, phân bổ xếp hạng tín dụng trái phiếu doanh nghiệp, tạo ra chỉ số vốn và trái phiếu tùy chỉnh.

DECISION TREE:

* Cây quyết định (DECISION TREE) là sơ đồ trong phân tích tài chính thể hiện mối tương tác của những tỷ số tài chính, và một tỷ số ảnh hưởng đến tỷ số khác như thế nào, cho phép nhà phân tích tài chính hoặc người kiểm tra ngân hàng, xem xét mối quan hệ nguyên nhân và kết quả trong những tỷ số khác nhau.
* Mô hình tính điểm tín dụng kiểm tra tương tác giữa các yếu tố tín dụng khác nhau, ví dụ, tuổi tác, thu nhập, loại nhà ở của bên vay nợ, bằng phương tiện là biểu đồ giống hình cây. Khác với phân tích tính điểm được chấp nhận rộng rãi, là gán điểm cho từng yếu tố tín dụng được xem xét khi chấp nhận hoặc bác bỏ đơn xin cấp tín dụng.
* Cây quyết định là một trong những mô hình có khả năng diễn giải cao và có thể thực hiện cả nhiệm vụ phân loại và hồi quy.