**4.Đánh giá mô hình hồi quy tuyến tính dựa trên các chỉ số:**

1. **Tập Train**:
   * **MAE**: 2.22 - Mô hình dự đoán sai lệch trung bình khoảng 2.22 đơn vị so với giá trị thực.
   * **MSE**: 9.26 - Bình quân của các sai số bình phương là 9.26, cho thấy mô hình có độ chính xác cao.
   * **R-squared (R²)**: 0.89 - Khoảng 89% biến thiên của giá trị mục tiêu có thể được giải thích bởi mô hình, cho thấy mô hình hoạt động tốt.
2. **Tập Test**:
   * **MAE**: 2.45 - Sai lệch trung bình lớn hơn một chút so với tập train, cho thấy mô hình có thể không tổng quát tốt như trên tập train.
   * **MSE**: 15.28 - Sai số bình phương cao hơn so với tập train, cho thấy có một số dự đoán không chính xác hơn trên tập test.
   * **R-squared (R²)**: 0.79 - Khoảng 79% biến thiên có thể giải thích, thấp hơn so với tập train, cho thấy mô hình có thể bị ảnh hưởng bởi hiện tượng underfitting.

**6.Viết nhận xét về hiệu quả của hai mô hình và đề xuất cải tiến:**

**6.1 Nhận xét về hiệu quả của hai mô hình:**

1. **Hồi quy tuyến tính**:
   * **Ưu điểm**: Dễ hiểu, giải thích đơn giản, tính toán nhanh.
   * **Nhược điểm**: Không phù hợp với dữ liệu phi tuyến tính, nhạy cảm với giá trị ngoại lai.
2. **Cây quyết định**:
   * **Ưu điểm**: Nắm bắt mối quan hệ phi tuyến tính tốt, dễ hình dung.
   * **Nhược điểm**: Có thể bị overfitting, không ổn định với dữ liệu nhỏ.

**6. 2Đề xuất cải tiến**

1. **Thêm mô hình khác**: Thử nghiệm với Random Forest hoặc Gradient Boosting.
2. **Tiền xử lý dữ liệu**: Lọc đặc trưng và xử lý giá trị ngoại lai để cải thiện hiệu suất.
3. **Phân tích sai số**: Đánh giá sai số dự đoán để điều chỉnh mô hình.