TRƯỜNG ĐẠI HỌC SÀI GÒN

**KHOA CÔNG NGHỆ THÔNG TIN**

****----- 🙡 🕮 🙣 -----

**MÔN PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

**Tên Bài: Khảo sát bài báo cho dữ liệu Ames Housing**

Giảng viên hướng dẫn: **TS. Đỗ Như Tài**

Sinh viên thực hiện:

Tạ Hồng Quí – 3122410348

TP. THCM

Nội dung

[1. Paper của cơ sở dữ liệu ở đâu? 3](#_Toc192752858)

[2. Dữ liệu ở đâu? 3](#_Toc192752859)

[3. Từ điển dữ liệu? 3](#_Toc192752860)

[4. Các bài toán liên quan? 3](#_Toc192752861)

[5. Kết quả đạt được là gì? Độ đo? 3](#_Toc192752862)

[6. Có bài survey về nó không? 4](#_Toc192752863)

[Tài liệu tham khảo: 5](#_Toc192752864)

# Paper của cơ sở dữ liệu ở đâu?

* Paper chính thức là:

Dean De Cock (2011). "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project," *Journal of Statistics Education*, Volume 19, Number 3.

* Đường link của paper là: <https://doi.org/10.1080/10691898.2011.11889627>​Ames Iowa Alternative….

# Dữ liệu ở đâu?

* Dữ liệu Ames Housing được công khai và có thể truy cập tại các đường link:
  + Excel:<http://www.amstat.org/publications/jse/v19n3/decock/AmesHousing.xls>
  + Text:<http://www.amstat.org/publications/jse/v19n3/decock/AmesHousing.txt>
  + Mô tả dữ liệu: <http://www.amstat.org/publications/jse/v19n3/decock/DataDocumentation.txt>​Ames Iowa Alternative….

# Từ điển dữ liệu?

* Từ điển dữ liệu chi tiết có sẵn ở đường dẫn: [Data Documentation](http://www.amstat.org/publications/jse/v19n3/decock/DataDocumentation.txt)
* Nó mô tả cụ thể từng biến trong bộ dữ liệu, cách thức đo lường, và ý nghĩa của từng trường​Ames Iowa Alternative….

# Các bài toán liên quan?

* Chủ yếu là bài toán hồi quy dự đoán giá nhà dựa vào các biến đầu vào như kích thước lô đất, diện tích nhà, số phòng ngủ, garage, chất lượng vật liệu, khu vực nhà (neighborhood),…
* Các biến gồm 80 biến đầu vào, bao gồm các biến liên tục, định danh, và thứ tự​Ames Iowa Alternative….
* Bộ dữ liệu này dùng thay thế cho bộ dữ liệu Boston Housing vốn nổi tiếng nhưng đã cũ, giúp sinh viên trải nghiệm với dữ liệu hiện đại và thực tế hơn.

# Kết quả đạt được là gì? Độ đo?

* Một số kết quả cụ thể trong bài báo đã đạt được:
  + Dùng một mô hình đơn giản với biến khu vực và tổng diện tích nhà có thể giải thích tới khoảng 80% biến thiên giá bán.
  + Một mô hình phức tạp hơn với 36 biến (bao gồm cả các biến tương tác, biến chuyển đổi, và recoding) đã giải thích được tới 92% biến thiên giá nhà.
* Các độ đo đánh giá mô hình trong bài báo gồm:
  + Bias: trung bình sai lệch (Yhat - Y)
  + Maximum Deviation: sai số lớn nhất
  + Mean Absolute Deviation (MAD): sai số tuyệt đối trung bình
  + Mean Square Error (MSE): trung bình bình phương sai số​Ames Iowa Alternative….

# Có bài survey về nó không?

* Bài báo này không phải dạng survey, mà là dạng bài báo giới thiệu và đề xuất một bộ dữ liệu thay thế (thay thế bộ dữ liệu cũ Boston Housing Data Set), nhằm mục đích giảng dạy môn học về hồi quy thống kê. Nó cũng cung cấp một số hướng dẫn sử dụng dữ liệu trong dạy học và nghiên cứu.
* Tuy nhiên, paper này được trích dẫn khá nhiều (đến thời điểm công bố đã có 24 trích dẫn), chứng tỏ bộ dữ liệu này được nhiều nơi sử dụng trong giảng dạy và nghiên cứu​Ames Iowa Alternative…

# Tài liệu tham khảo:

**[1]:** Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project