

VŨ DUY LINH - HUỖNH XUÂN HIỆP

**GIÁO TRÌNH**  
**KỸ THUẬT DỰ BÁO**



**NHÀ XUẤT BẢN ĐẠI HỌC CẦN THƠ**  
**2019**

**BIÊN MỤC TRƯỚC XUẤT BẢN THỰC HIỆN BỞI  
TRUNG TÂM HỌC LIỆU TRƯỜNG ĐẠI HỌC CẦN THƠ**

---

Vũ, Duy Linh

Giáo trình kỹ thuật dự báo / Vũ Duy Linh, Huỳnh Xuân Hiệp .– Cần Thơ : Nxb. Đại học Cần  
Thơ, 2019.

132 tr. : minh họa ; 24 cm

Sách có danh mục tài liệu tham khảo

ISBN: 9786049651410

1. Mathematical statistics 2. Thống kê dự báo 3. Toán thống kê

I. Nhan đề. II. Huỳnh, Xuân Hiệp

519.5 – DDC 23

MFN 230876

L312

## LỜI GIỚI THIỆU

Nhằm góp phần làm phong phú nguồn tư liệu phục vụ nghiên cứu, học tập cho bạn đọc, sinh viên, học viên và nghiên cứu sinh ngành Công nghệ Thông tin và Truyền thông, Nhà xuất bản Đại học Cần Thơ ấn hành và giới thiệu cùng bạn đọc giáo trình “Kỹ thuật dự báo” do ThS. Vũ Duy Linh và PGS. TS. Huỳnh Xuân Hiệp biên soạn.

Giáo trình gồm 6 chương, nội dung giới thiệu tổng quan về dự báo, phân tích hồi quy tuyến tính, phân tích chuỗi thời gian đơn biến dừng. Thêm vào đó, cuối mỗi chương còn có nhiều bài tập ôn tập với công cụ là R sẽ hỗ trợ đắc lực cho bạn đọc trong việc giải bài tập. Giáo trình là tài liệu học tập có giá trị liên quan đến kỹ thuật dự báo.

Nhà xuất bản Đại học Cần Thơ chân thành cảm ơn các tác giả và sự đóng góp ý kiến của quý thầy cô trong Hội đồng thẩm định trường Đại học Cần Thơ để giáo trình “Kỹ thuật dự báo” được ra mắt bạn đọc.

Nhà xuất bản Đại học Cần Thơ trân trọng giới thiệu đến sinh viên, giảng viên và bạn đọc giáo trình này.

NHÀ XUẤT BẢN ĐẠI HỌC CẦN THƠ



## LỜI NÓI ĐẦU

Giáo trình **Kỹ thuật dự báo** được biên soạn cho đối tượng sinh viên ngành Công nghệ thông tin, trường Đại học Cần thơ. Mục tiêu của học phần này là cung cấp cho người học đủ các kiến thức từ cơ bản đến nâng cao về dự báo, quy trình xác định mô hình, kiểm định cũng như đánh giá mô hình dự báo tối ưu. Cách xác định phương pháp phù hợp vào từng bài toán cụ thể, nhất là có khả năng xây dựng được những mô hình dự báo mà thực tiễn đề ra.

Sau nhiều năm học phần này được giảng dạy, với thời lượng 03 tín chỉ và học phần tiên quyết là Thống kê, giáo trình đã xác định được cấu trúc hợp lý và chặt chẽ với kỳ vọng hướng người học nắm được các kiến thức quan trọng như: Việc quan trắc thu thập và tiền xử lý dữ liệu, các phương pháp dự báo phổ biến như hồi quy tuyến tính, phi tuyến và Logistics, phân tích chuỗi thời gian, và phương pháp lai giữa mạng Bayes và hồi quy, để có thể xây dựng mô hình và diễn dịch kết quả dự báo cho những vấn đề thuộc lĩnh vực kinh tế, xã hội.

Giáo trình bao gồm 06 chương và phần phụ lục, mỗi chương đều có ví dụ minh họa với công cụ hỗ trợ là ngôn ngữ R và các bài tập yêu cầu, nhằm giúp sinh viên có các bài thực hành trong phòng máy và tự rèn luyện thêm ở nhà. Số ví dụ và bài tập này được đề cập từ đơn giản đến nâng cao mà đích đến là các bài tập lớn theo dạng case study cho mỗi phương pháp dự báo, điều hữu ích cần được trải nghiệm bởi sinh viên.

Với mong muốn có một tài liệu hữu ích cho sinh viên học tập nghiên cứu, giáo trình đã được biên soạn dựa vào kiến thức và các ví dụ minh họa từ nhiều nguồn ở trong và ngoài nước, cũng như từ những kinh nghiệm giảng dạy của bản thân người viết. Tác giả xin cảm ơn tất cả các nguồn dữ liệu được chia sẻ trên các website mà không thể liệt kê hết ở đây. Do sự hiểu biết là hữu hạn nên những thiếu sót chắc hẳn sẽ khó tránh khỏi, giáo trình rất mong nhận được sự chia sẻ và góp ý của bạn đọc cũng như sinh viên qua địa chỉ email: vdlinh@ctu.edu.vn, để hoàn thiện và phục vụ cho việc học và giảng dạy tốt hơn.

*Người biên soạn,*

**Vũ Duy Linh**



# MỤC LỤC

<b>Chương 1. TỔNG QUAN VỀ DỰ BÁO</b>	<b>1</b>
1.1 GIỚI THIỆU	1
1.2 Ý NGHĨA VÀ VAI TRÒ CỦA DỰ BÁO	2
1.3 CÁC PHƯƠNG PHÁP DỰ BÁO	3
1.3.1 Phương pháp dự báo định tính	3
1.3.2 Phương pháp dự báo định lượng	3
1.4 QUY TRÌNH DỰ BÁO	4
BÀI TẬP	4
<b>Chương 2. PHÂN TÍCH HỒI QUY TUYẾN TÍNH</b>	<b>5</b>
2.1 GIỚI THIỆU	5
2.2 CÁC TẬP DỮ LIỆU MẪU	7
2.2.1 Tập dữ liệu Sửa máy tính ComputerRepair	7
2.2.2 Tập dữ liệu công nghiệp ô tô	8
2.2.3 Tập dữ liệu Ducan	8
2.3 CÁC BƯỚC PHÂN TÍCH HỒI QUY	9
2.4 PHÂN TÍCH HỒI QUY TUYẾN TÍNH ĐƠN	10
2.4.1 Giới thiệu	10
2.4.2 Ước lượng tham số hồi quy	12
2.4.3 Kiểm định giả thuyết (hypotheses) các hệ số ước lượng	15
2.4.4 Khoảng tin cậy và khoảng dự đoán	20
2.4.5 Đo lường độ phù hợp của mô hình	21
2.4.6 Phân tích phân dư	22
2.5 PHÂN TÍCH HỒI QUY TUYẾN TÍNH ĐA BIẾN	25
2.5.1 Mô hình hồi quy đa biến	25
2.5.2 Đánh giá mô hình	29
2.5.3 Rút gọn mô hình	32
2.5.4 Khoảng tin cậy và khoảng dự báo	34
2.6 PHÂN TÍCH HỒI QUY THEO BIẾN GIÁ	35
BÀI TẬP	38
<b>Chương 3. PHÂN TÍCH HỒI QUY LOGISTICS</b>	<b>40</b>
3.1 MÔ HÌNH HỒI QUY LOGISTIC	40
3.2 MỘT SỐ BÀI TOÁN PHÂN TÍCH HỒI QUY LOGIT	44
3.2.1 Bài toán 1 - Nguy cơ lây nhiễm bệnh tình dục	45

3.2.2 Bài toán 2 - Nguy cơ bệnh động mạch vành	49
<b>BÀI TẬP</b>	<b>52</b>
<b>Chương 4. PHÂN TÍCH CHUỖI THỜI GIAN ĐƠN BIẾN DỪNG</b>	<b>55</b>
4.1 KHÁI NIỆM VỀ CHUỖI THỜI GIAN	55
4.2 CÁC THÀNH PHẦN CỦA CHUỖI THỜI GIAN	56
4.2.1 Thành phần xu hướng	57
4.2.2 Thành phần mùa	58
4.2.3 Thành phần chu kỳ	59
4.2.4 Thành phần bất thường	60
4.3 CÁC HÀM CƠ BẢN TRÊN CHUỖI THỜI GIAN	60
4.3.1 Nhiều trắng	60
4.3.2 Hàm tự tương quan	61
4.3.3 Hàm tự tương quan riêng phần	62
4.3.4 Tính dừng/ ổn định	63
4.3.5 Phép toán sai phân	67
4.4 MÔ HÌNH TỰ HỒI QUY BẬC P	69
4.5 MÔ HÌNH TRUNG BÌNH TRƯỢT BẬC Q	75
4.6 MÔ HÌNH TRUNG BÌNH TRƯỢT TỰ HỒI QUY	76
<b>BÀI TẬP</b>	<b>82</b>
<b>Chương 5. PHÂN TÍCH CHUỖI THỜI GIAN ĐƠN BIẾN CHƯA DỪNG</b>	<b>83</b>
5.1 MÔ HÌNH TỰ HỒI QUY TÍCH HỢP TRUNG BÌNH TRƯỢT	83
5.2 MÔ HÌNH CHUỖI THỜI GIAN ĐƠN BIẾN CÓ TÍNH MÙA	91
<b>BÀI TẬP</b>	<b>102</b>
<b>Chương 6. MẠNG BAYES</b>	<b>103</b>
6.1 MẠNG BAYES	103
6.1.1 Các khái niệm xác suất	103
6.1.2 Các khái niệm mạng Bayes	104
6.1.3 Giải thuật suy diễn Junction tree trong mạng Bayes	107
6.2 MẠNG BAYES TRONG GÓI GRAIN	107
6.2.1 Giới thiệu	107
6.2.2 Xây dựng mạng và suy diễn xác suất để chuẩn đoán bệnh khó thở	109
<b>BÀI TẬP</b>	<b>117</b>
<b>PHỤ LỤC</b>	<b>121</b>
<b>TÀI LIỆU THAM KHẢO</b>	<b>123</b>



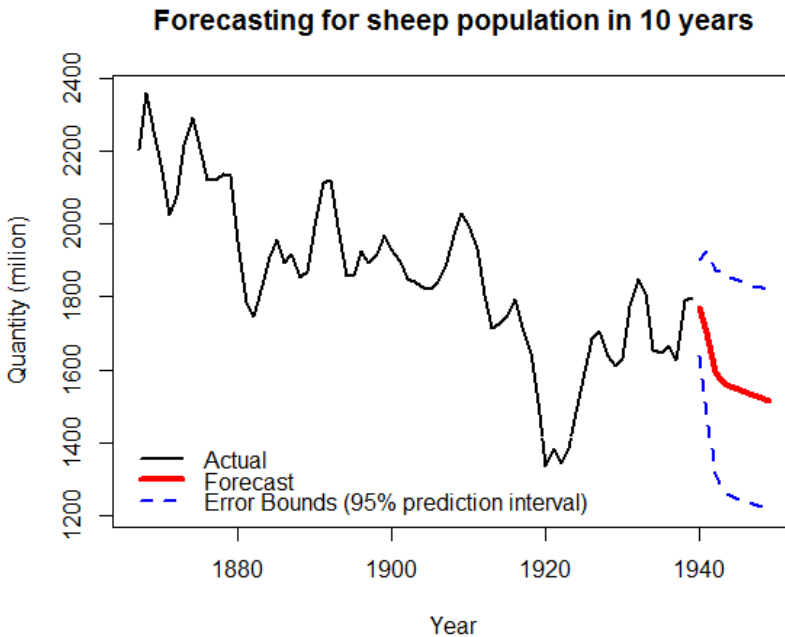
# Chương 1

## TỔNG QUAN VỀ DỰ BÁO

Chương này sẽ cung cấp cho người học các khái niệm cần thiết để có cái nhìn tổng quát và đầy đủ về những thành phần của mô hình dự báo, và vai trò, ý nghĩa của việc dự báo.

### 1.1 GIỚI THIỆU

Từ thời xưa cho đến ngày nay, nhu cầu dự báo về một tình huống, sự kiện quan trọng nào đó có thể xảy ra trong tương lai là điều quan trọng trong hầu hết lĩnh vực của đời sống xã hội, nhất là ở tầm quản trị vĩ mô. Một số ví dụ điển hình như: Dự báo nhu cầu sử dụng điện để xây dựng nhà máy có công suất phù hợp cho một khu dân cư trong một khoảng thời gian xác định, dự báo lịch xuống giống né rầy nâu trên cánh đồng mẫu lớn, dự báo mức độ phát sinh khí thải CO<sub>2</sub>, dự đoán về bệnh động mạch vành (CHD),...

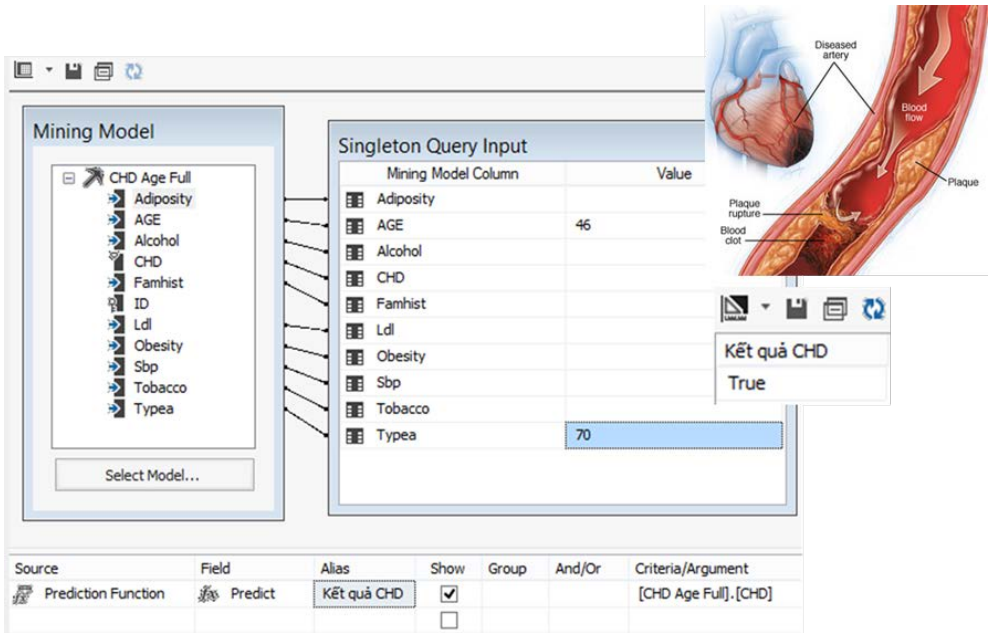


**Hình 1.1** Dự báo số lượng cừu trong 10 năm tới

Dự báo là ước tính của một sự kiện có thể xảy ra trong một khoảng thời gian ở tương lai với độ chính xác cao nhất. Điều này đạt được bằng cách xây dựng một mô hình từ các thông tin có sẵn, bao gồm dữ liệu lịch sử và các tri thức ở hiện tại tác động đến kết quả dự báo. Trong giáo trình này, chúng tôi

không nhằm phân biệt rạch ròi thuật ngữ forecast và predict mà sử dụng thống nhất chung một thuật ngữ là dự báo (forecast).

Khoảng thời gian dự báo (forecast horizon) thường ở các mức: Tức thì (immediate) độ dưới 01 tháng; ngắn hạn (short-term) khoảng 01 đến 03 tháng; Trung hạn (medium-term) khoảng từ 03 tháng đến 02 năm; và dài hạn (long-term) từ 2 năm trở lên.



**Hình 1.2** Dự đoán bệnh mạch vành

Việc xác định khoảng thời gian dự báo chỉ mang tính tương đối và phụ thuộc vào vấn đề và lĩnh vực cần dự báo cũng như dữ liệu, phương pháp và độ chính xác của dự báo. Chú ý rằng thang thời gian và các đơn vị đo trong từng lĩnh vực dự báo là có thể khác nhau. Ví dụ như trong lĩnh vực kinh tế, thì người ta thường sử dụng thang thời gian (unit time) theo tháng, quý hoặc năm, còn trong dự báo thời tiết hoặc sâu bệnh (hại lúa/hoa màu) thì thường sử dụng là ngày.

## 1.2 Ý NGHĨA VÀ VAI TRÒ CỦA DỰ BÁO

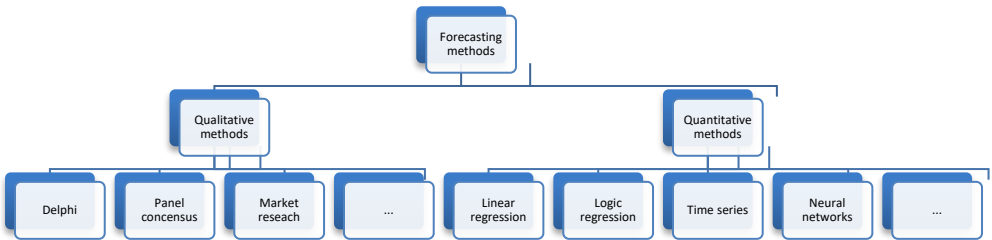
Dự báo là công việc quan trọng, cần có để thực hiện thành công các chức năng hoạt động và có tính chiến lược của các tổ chức trong xã hội để giúp cho các quyết định có độ tin cậy cao và giảm thiểu các rủi ro có thể có khi triển khai ở hiện tại cũng như trong tương lai.

Ngày nay, dự báo để hỗ trợ quyết định/quản lý ở cấp vĩ mô là không thể phủ nhận. Một doanh nghiệp, một tổ chức, một đất nước muốn cạnh tranh/phát triển được thì cần phải có "tầm nhìn". Nếu không có ứng dụng dự báo thì tầm nhìn của họ sẽ bị che khuất, việc chỉ đạo và thực hiện không có kế hoạch rõ ràng và như vậy rất khó để né tránh hậu quả phát sinh.

### 1.3 CÁC PHƯƠNG PHÁP DỰ BÁO

Hiện nay, có nhiều kỹ thuật dự báo khác nhau được phân ra thành hai phương pháp, đó là, phương pháp dự báo định tính (Qualitative methods) và phương pháp định lượng (Quantitative methods).

Trong giáo trình này, chúng tôi chỉ đề cập đến những phương pháp định lượng. Còn phương pháp định tính chỉ được sử dụng trong các bài báo được tham khảo trong phần phụ lục.



**Hình 1.3** Các phương pháp dự báo

#### 1.3.1 Phương pháp dự báo định tính

Các phương pháp này dựa trên cơ sở phân tích các yếu tố trong mô hình dựa vào việc lấy các ý kiến, nhận xét, hoặc tri thức chuyên gia khi được tham vấn về các vấn đề mà họ am hiểu. Ví dụ như độ tuổi nào của lúa và rầy nâu sẽ gây ảnh hưởng lớn nhất đến bệnh cháy rầy, chọn lọc những yếu tố nào quan trọng cần đầu tư để kinh doanh, để tạo môi trường học tập tốt nhất trong một khoảng thời gian xác định nào đó trong tương lai. Các phương pháp này có ưu điểm là thu nhận được các thông tin từ những người hiểu biết sâu rộng về lãnh vực mà mô hình dự báo quan tâm, nhưng chúng lại mang tính chủ quan, đôi khi thiếu tính công bằng, khách quan giữa các chuyên gia khi cho ý kiến.

#### 1.3.2 Phương pháp dự báo định lượng

Các phương pháp dự báo định lượng các mô hình toán, các số liệu lịch sử hoặc/và các yếu tố hiện tại khác để ước lượng ra các hệ số của mô hình dự báo được biểu diễn bởi một hàm toán học tương ứng cho mỗi phương pháp định lượng. Ưu điểm của phương pháp này là khách quan vì kết quả như thế nào là phụ thuộc vào dữ liệu và kỹ thuật được chọn. Tuy nhiên, trong thực