

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN ĐỨC CHÍ

**“NGHIÊN CỨU PHƯƠNG PHÁP KHAI PHÁ DỮ LIỆU
VÀ ỨNG DỤNG DỰ BÁO NGUỒN GỠ NGUYÊN LIỆU
CHO SẢN XUẤT TẠI NHÀ MÁY GIẤY BÃI BÀNG”**

Chuyên ngành : Khoa học máy tính

Mã số : 60 48 01

LUẬN VĂN THẠC SỸ

NGƯỜI HƯỚNG DẪN KHOA HỌC: GS.TS NGUYỄN THANH THỦY

GS.TS Nguyễn Thanh Thủy

Thái nguyên, tháng 12/2013



LỜI CAM ĐOAN

Tôi xin cam đoan:

Những nội dung trong luận văn này là do tôi thực hiện dưới sự hướng dẫn trực tiếp của thầy giáo Giáo sư, Tiến sĩ Nguyễn Thanh Thủy.

Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng tác giả, tên công trình, thời gian, địa điểm công bố.

Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo hay gian lận tôi xin hoàn toàn chịu trách nhiệm./.

Thái Nguyên, tháng 12 năm 2013

Tác giả luận văn

Nguyễn Đức Chí



LỜI CẢM ƠN

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc tới tập thể các Giáo sư, Tiến sĩ, giảng viên thuộc Viện Công nghệ Thông tin – Viện Hàn lâm Khoa học và Công nghệ Việt Nam; Trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên đã tận tình giảng dạy, truyền đạt kiến thức cho tôi trong suốt thời gian học tập vừa qua. Tôi xin bày tỏ lòng biết ơn tới GS.TS Nguyễn Thanh Thủy người đã tận tình hướng dẫn, chỉ bảo và chia sẻ những tài liệu rất hữu ích để tôi hoàn thành luận văn.

Xin chân thành cảm ơn lãnh đạo trường Đại học Công nghệ thông tin và Truyền thông – Đại học Thái Nguyên đã tạo điều kiện giúp đỡ tôi về mọi mặt trong suốt thời gian học tập tại trường cũng như trong thời gian thực hiện luận văn.

Tôi xin bày tỏ lòng biết ơn sâu sắc đến gia đình, người thân, bạn bè, đồng nghiệp những người luôn động viên, khuyến khích và giúp đỡ để tôi có thể hoàn thành tốt nội dung nghiên cứu./.

Thái Nguyên, tháng 12 năm 2013

Tác giả luận văn

Nguyễn Đức Chí

MỤC LỤC

I. CHƯƠNG I: TỔNG QUAN VỀ DỰ BÁO DỮ LIỆU	3
1.1 Khái niệm về dự báo.	3
1.2 Một số đặc điểm của dự báo.	3
1.3 Các phương pháp dự báo.	4
1.3.1 Phương pháp dự báo định tính (phán đoán).....	5
1.3.2 Phương pháp dự báo định lượng.	8
1.3.3 Quy trình dự báo.	11
1.4 Kết luận:	11
II. CHƯƠNG 2: KHAI PHÁ DỮ LIỆU VÀ CÁC ỨNG DỤNG CHO DỮ LIỆU	12
2.1. Khai phá dữ liệu:	12
2.1.1 Phát hiện tri thức và khai phá dữ liệu.	12
2.1.2 Quá trình phát hiện tri thức từ cơ sở dữ liệu.	12
2.2. Các phương pháp dự báo dữ liệu:	14
2.2.1 Các phương pháp trực quan	14
2.3.1 Chuỗi thời gian thực	19
2.3.2. Thành phần xu hướng dài hạn	20
2.3.3 Thành phần mùa	21
2.3.4. Thành phần chu kỳ	21
2.3.5. Thành phần bất thường.....	21
2.3.6 Khai phá tri thức trên cơ sở dữ liệu chuỗi thời gian.	21
2.4 Các ứng dụng cho dự báo dữ liệu:	22
III. CHƯƠNG 3: MÔ HÌNH ARIMA, PHẦN MỀM EViews	24
3.1 Mô hình ARIMA.....	24
3.1.1 Lập mô hình AR, MA và ARIMA với dữ liệu chuỗi thời gian. ...	24
3.1.2 Xem xét tính dừng của chuỗi quan sát.....	27
3.1.3 Các bước lập trong phương pháp luận Box-Jenkins (BJ).....	28
3.2 Phần mềm Eviews	28
3.3 Áp dụng cho bài toán dự báo gốc nguyên liệu	30
3.3.1 Dữ liệu cho dự báo.....	30
3.3.2 Mô hình ARIMA cho dự báo nguyên liệu gốc.....	31
3.4 Các đánh giá sau dự báo.....	36
KẾT LUẬN	37
HƯỚNG NGHIÊN CỨU TIẾP THEO	39
DANH MỤC TÀI LIỆU THAM KHẢO.....	40

DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT

Các ký hiệu, chữ viết tắt	Nội dung
KPDL	Khai phá dữ liệu
CNTT	Công nghệ thông tin
ARIMA	Autoregresssive Intergrated Moving Avegage
AR	Autoregresssive
MA	Moving Avegage

DANH MỤC CÁC BẢNG

Bảng	Nội dung	Trang
1.1	Một số phương pháp dự báo phổ biến	4
3.1	Các dữ liệu đầu vào cho dự báo	26
		22

DANH MỤC CÁC HÌNH (HÌNH VẼ, ẢNH CHỤP, ĐỒ THỊ...)

Hình	Nội dung	Trang
2.1	Quy trình phát hiện tri thức từ cơ sở dữ liệu	9
2.2	Minh họa chuỗi thời gian	14
2.3	Đồ thị minh họa thành phần xu hướng dài hạn	16
3.1	Giao diện chương trình Eviews	25
3.2	Giao diện làm việc với workfile khởi tạo các biến	25
3.3	Giao diện nhập số liệu vào chương trình eviews	27
3.4	Tương quan lượng gỗ tự cấp và mua ngoài	28
3.5	Ước lượng ARIMA(1,0,1) với tucap và muangoai	29
3.6		
3.7	Biểu đồ tương quan lượng giấy và bột giấy sản xuất được	29
3.8	So sánh lượng nguyên liệu cho sản xuất và xuất khẩu	30
3.9	So sánh lượng giấy và bột giấy đã sản xuất	30
3.10	Ước lượng hồi quy lượng giấy và bột giấy đã sản xuất	31

MỞ ĐẦU

Tổng công ty giấy Việt Nam hiện nay là một trong những đơn vị lớn trong cả nước về sản xuất bột giấy và giấy. Đặc biệt sản phẩm giấy Bãi Bằng (sản xuất tại Nhà máy giấy Bãi Bằng) của Tổng công ty là thương hiệu nổi tiếng đã được người tiêu dùng trong nước biết đến từ lâu và tin dùng.

Trong quá trình phát triển và hoạt động, giấy Bãi Bằng đã nhiều lần nâng cấp thiết bị và mở rộng sản xuất. Năm 2003 đã đầu tư thành công Dự án mở rộng công ty giấy Bãi Bằng giai đoạn I, nâng công suất sản xuất bột giấy từ 48.000 tấn/năm lên 71.000 tấn/năm và công suất sản xuất giấy từ 50.000 tấn/năm lên 100.000 tấn/năm. Tiếp theo thành công của dự án đầu tư giai đoạn I, Tổng công ty giấy Việt Nam đã lập dự án đầu tư giai đoạn II - dây chuyền bột giấy tẩy trắng 250.000 tấn/năm, dự án giấy nhà máy giấy Thanh Hoá - dây chuyền sản xuất giấy in/viết 100.000 tấn/năm, năm 2010 Tổng công ty đã đầu tư nâng công suất sản xuất giấy lên 120.000 tấn/năm, tiếp nhận dự án nhà máy bột giấy Phương Nam – tỉnh Long An và một số dự án khác. Tuy nhiên, trong đó có các dự án hoặc là đã dừng lại, hoặc là chưa thành công hay triển vọng sinh lời khi đưa vào sản xuất còn hạn chế. Nguyên nhân là thiếu sự nghiên cứu, phân tích đầy đủ về thị trường đầu vào và đầu ra của các sản phẩm trong các dự án đầu tư, cũng như công nghệ phù hợp với điều kiện sản xuất ở Việt Nam.

Trong khi Tổng công ty chưa thành công trong các dự án đầu tư ngành giấy thì Việt Nam lại đang được coi là nước thu hút đầu tư ngành giấy, trong đó các công ty giấy của Nhật Bản, Trung Quốc, Thái Lan... đã và đang đầu tư lớn vào Việt Nam. Các công ty trong nước cũng đang mở rộng sản xuất, xúc tiến đầu tư. Tổng công ty giấy Việt Nam đang đứng trước nguy cơ tụt hậu so với các công ty sản xuất giấy khác ngay tại Việt Nam.

Mặc dù nhiều dự án đầu tư đã và đang được tiến hành, hầu hết các công ty và các dự án giấy tại Việt Nam mới chỉ đầu tư và sản xuất được giấy in, giấy viết, một phần nhỏ giấy in báo, giấy vàng mã, giấy các tông sóng ở quy mô nhỏ, đa số không có xử lý môi trường, chưa sản xuất được các loại giấy cao cấp như giấy bao bì tráng phủ, giấy kraft tẩy trắng hoặc không tẩy trắng, giấy làm lớp mặt thùng có lớp sóng,...

Mặt khác, tuy các điều kiện về kinh tế và dân số của Việt Nam vẫn tăng trưởng đều đặn, sự phát triển của công nghệ thông tin và các công nghệ mới đã làm cho nhu cầu giấy in báo tại Việt Nam giảm sút rõ rệt, nhu cầu giấy in/viết không còn tăng mạnh và chịu sự cạnh tranh khốc liệt từ giấy ngoại nhập, các loại sách điện tử, những lý do này làm cho các thị trường giấy in báo, giấy in/viết của Tổng công ty Giấy Việt Nam không còn là cơ hội đầu tư nữa.

Trong điều kiện môi trường kinh doanh như vậy, Tổng công ty giấy Việt Nam đã nhận thấy sự cần thiết và cơ hội đầu tư sản xuất sản phẩm giấy cao cấp khác để cạnh tranh. Cùng với việc nghiên cứu thị trường và công nghệ sản xuất giấy trước khi đi đến quyết định đầu tư cần phải nghiên cứu và dự báo rõ ràng nguồn nguyên liệu cho việc sản xuất mang tính ổn định lâu dài. Việc tự lực được nguồn nguyên liệu là quyết định được lợi thế cạnh tranh lớn trên thị trường.

Mục đích bài luận văn là tìm hiểu một số phương pháp khai phá, phân tích dữ liệu từ các số liệu thực tế được tổng hợp và ứng dụng cho bài toán dự báo nguồn nguyên liệu cho sản xuất của nhà máy giấy Bãi Bằng – Tổng công ty Giấy Việt Nam.

NỘI DUNG

I. CHƯƠNG I: TỔNG QUAN VỀ DỰ BÁO DỮ LIỆU

1.1 Khái niệm về dự báo.

Dự báo là biện pháp lập luận khoa học tiên đoán những sự việc sẽ xảy ra trong tương lai, trên cơ sở phân tích, tổng hợp về các dữ liệu đã thu thập được. Khi tiến hành dự báo cần căn cứ vào việc thu thập, xử lý số liệu trong quá khứ và hiện tại để xác định xu hướng vận động của các hiện tượng trong tương lai nhờ vào một số mô hình toán học (Định lượng). Tuy nhiên dự báo cũng có thể là một dự đoán chủ quan hoặc trực giác về tương lai (Định tính) và để dự báo định tính được chính xác hơn, cần phải loại trừ những tính chủ quan của người dự báo.

Dù định nghĩa có sự khác biệt nào đó, nhưng đều thống nhất về cơ bản là dự báo bàn về tương lai, nói về tương lai. Dự báo trước hết là một thuộc tính không thể thiếu của tư duy của con người, con người luôn luôn nghĩ đến ngày mai, hướng về tương lai. Trong thời đại công nghệ thông tin và toàn cầu hóa, dự báo lại đóng vai trò quan trọng hơn khi nhu cầu về thông tin thị trường, tình hình phát triển tại thời điểm nào đó trong tương lai càng cao. Dự báo được sử dụng trong nhiều lĩnh vực khác nhau, mỗi lĩnh vực có một yêu cầu về dự báo riêng nên phương pháp dự báo được sử dụng cũng khác nhau.

1.2 Một số đặc điểm của dự báo.

Tính không chính xác của dự báo: Dù phương pháp chúng ta sử dụng là gì thì luôn tồn tại yếu tố không chắc chắn cho đến khi thực tế diễn ra.

Luôn có điểm mù trong các dự báo: Không phải cái gì cũng có thể dự báo được nếu chúng ta thiếu hiểu biết về vấn đề cần dự báo.