

ỨNG DỤNG MÔ HÌNH CHUỖI THỜI GIAN SARIMA VÀ MẠNG THẦN KINH NHÂN TẠO ANN DỰ BÁO LƯỢNG KHÁCH QUỐC TẾ ĐẾN VIỆT NAM

SV: Lê Huy Hiệp Lóp, Lâm Văn Hạnh
Trường Đại học Bà Rịa – Vũng Tàu
GVHD: ThS. Nghiêm Phúc Hiếu

TÓM TẮT

Du lịch Việt Nam đã và đang trở thành ngành kinh tế mũi nhọn của đất nước, được xem là “Ngành công nghiệp không khói” và đóng góp khoản thu lớn vào ngân sách nhà nước. Tuy nhiên, tiềm năng phát triển du lịch còn rất to lớn trong khi chúng ta chưa thể khai thác và tận dụng hết. Khi cách mạng công nghiệp 4.0 sắp tới, việc dự báo lượng khách quốc tế đến nước ta thực sự có ý nghĩa đối với các nhà quản lý, các nhà đầu tư để có kế hoạch phát triển bền vững. Nghiên cứu sử dụng phương pháp luận Box-Jenkins để xây dựng mô hình ARIMA theo mùa (hay còn gọi là SARIMA) và mô hình mạng thần kinh nhân tạo cho dự báo lượng khách quốc tế đến Việt Nam dựa trên số liệu công bố hàng tháng của Tổng cục Du lịch Việt Nam với 101 quan sát. Kết quả cho thấy mô hình SARIMA (12,1,1) (1,1,3)₁₂ và ANN-12-29-1 là phù hợp nhất cho dự báo. Khi so sánh dự báo trong mẫu hai mô hình này với nhau thì mô hình ANN có kết quả tốt hơn.

Từ khóa: dự báo, khách quốc tế, SARIMA, ANN, công nghiệp 4.0.

1. Giới thiệu

Việt Nam với những lợi thế đặc biệt về vị trí địa lý kinh tế và chính trị, Việt Nam có rất nhiều thuận lợi để phát triển du lịch. Nằm ở trung tâm Đông Nam Á, lãnh thổ Việt Nam vừa gắn liền với lục địa vừa thông ra đại dương nên vì thế Việt Nam hiển nhiên có rất nhiều cảnh đẹp và hùng vĩ, có vị trí giao lưu quốc tế thuận lợi cả về đường biển, đường sông, đường sắt, đường bộ và hàng không. Đây là tiền đề rất quan trọng trong việc mở rộng và phát triển du lịch quốc tế

Đặc biệt hơn, hiện nay Việt Nam có hệ thống di sản thế giới rất phong phú, đa dạng. Theo thống kê năm 2016, cả nước có 26 di sản thế giới được UNESCO công nhận thuộc các lĩnh vực như Di sản thiên nhiên, di sản văn hóa vật thể, di sản văn hóa phi vật thể, di sản địa chất toàn cầu, di sản tư liệu, di sản hỗn hợp. Trong đó có 3 di sản thiên nhiên thế giới; 12 di sản văn hóa thế giới; 11 di sản văn hóa phi vật thể thế giới.

Nhận thức về tầm quan trọng của Du lịch, và để phát triển Du lịch phù hợp với yêu cầu và xu thế của thời đại nên Công tác quảng bá được đẩy mạnh. Thời gian gần đây công tác quảng bá du lịch được chú trọng. Hình ảnh du lịch ngày càng được nâng cao trên thị trường thế giới. Bản đồ thế giới dần hiểu rõ hơn về Việt Nam thông qua các hoạt động giao lưu, hợp tác quốc tế. Các website, tổ chức bầu chọn Việt Nam là một trong những điểm đến hấp dẫn du khách. Với định hướng du lịch trở thành ngành kinh tế mũi nhọn. Nhận thức được tầm quan trọng nên công tác xúc tiến cũng được chú trọng làm chất lượng của việc xúc tiến cũng được nâng cao đáng kể. Gần đây nhất là Việt Nam được Hollywood chọn làm phim trường cho bộ phim Kong - Đảo đầu lâu - đây cũng là cơ hội để hình ảnh Việt Nam được quảng bá rộng rãi trên thị trường thế giới.

Thêm vào đó, cuộc cách mạng công nghiệp 4.0 có tác động tích cực tới sự phát triển du lịch của quốc gia. Các hình thức thanh toán tuyến nhanh chóng, hiệu quả giúp cho việc đáp ứng nhu cầu của mọi đối tượng ở bất kỳ mọi nơi trên thế giới, không bị ảnh hưởng bởi rào cản địa lý tạo thuận lợi tối đa cho du khách. Quá trình ứng dụng công nghệ giúp giảm chi phí nhân lực, giảm chi phí bán hàng từ đó giá thành dịch vụ giảm đồng thời nâng cao chất lượng sản phẩm du lịch. Công nghệ 4.0 giúp đưa thông tin, hình ảnh nhanh chóng đến du khách trên mọi vùng thế giới. Các địa điểm tham quan đẹp hấp dẫn sẽ dễ dàng lan tỏa nhanh chóng trên công đồng thế giới thông qua các công nghệ hiện đại. Với những điều kiện thuận lợi trên ngành du lịch Việt Nam sẽ bức phá dẫn đầu. Mục tiêu đến năm 2030, du lịch Việt Nam sẽ thực sự trở thành ngành kinh tế mũi nhọn và Việt Nam là nước có ngành du lịch phát triển hàng đầu khu vực Đông Nam Á.

Để đạt được mục tiêu đó, công tác dự báo về lượng khách quốc tế đến là điều hết sức cần thiết giúp quốc gia chủ động chuẩn bị các nguồn lực như cơ sở hạ tầng, công nghệ, nhân lực, tài chính... trong bối cảnh phải có tầm nhìn xa trông rộng, áp dụng công nghệ hiện đại vào công tác dự báo, quốc gia mới có khả năng tận dụng được các điều kiện thuận lợi của mình để giúp ngành du lịch cất cánh”. Trên phương diện nghiên cứu mô hình chuỗi thời gian SARIMA và mạng thần kinh nhân tạo ANN cùng những nguyên lý hoạt động của chúng để từ đó ứng dụng vào lĩnh vực dự báo vẫn còn là một nội dung chưa được quan tâm nhiều, tác giả sẽ tiến hành thực hiện nghiên cứu của mình dựa trên cơ sở khoa học này và kết quả mang lại góp phần nào đó giúp ngành du lịch nước nhà bền vững.

2. Tổng quan nghiên cứu

Dự báo khách du lịch là một quá trình thách thức vì nó chịu ảnh hưởng đáng kể bởi tính xu hướng, các nhân tố mùa vụ, các sự kiện không lường trước như thảm họa thiên nhiên hoặc chiến tranh. Trong tổng quan nghiên cứu tác giả đưa ra hai hướng nghiên cứu của các tác giả trước đây bao gồm hướng tiếp cận chuỗi thời gian nhằm ứng dụng mô hình SARIMA và hướng tiếp cận mạng thần kinh nhân tạo ANN.

2.1. Hướng tiếp cận chuỗi thời gian

Theo hướng chuỗi thời gian, mô hình ARIMA giản đơn hoặc SARIMA (ARIMA theo mùa) được sử dụng một cách khá thường xuyên. Tuy nhiên, hầu hết các nghiên cứu gần đây đều cho rằng mô hình SARIMA phổ biến hơn mô hình ARIMA giản đơn bởi vì yếu tố mùa vụ trong dữ liệu khách du lịch là một nhân tố chủ yếu luôn được xem xét bởi các nhà quản lý trong ngành du lịch. Lý do khác liên quan tới sự quan tâm dành cho hướng nghiên cứu này đó là chi phí thu thập dữ liệu mà ước lượng mô hình khá thấp, bởi vì chỉ cần dữ liệu quá khứ của các biến số (Kodituwakku, Wijesundara, & Hettiarachchi, 2015).

Trong những năm qua, có nhiều nghiên cứu được tiến hành để dự báo lượng khách quốc tế sử dụng mô hình SARIMA dựa theo phương pháp chuẩn Box-Jenkins. Kim và Moosa (2005) sử dụng mô hình dự báo hồi quy tuyến tính, mô hình Structural Harvey và mô hình SARIMA để dự báo lượng khách du lịch và nhận thấy rằng phương pháp tốt nhất đó là phương pháp gián tiếp so với phương pháp trực tiếp khi dự báo lượng khách quốc tế tới Úc hàng tháng. Phương pháp trực tiếp không kiểm định được các yếu tố chuỗi thời gian trong khi phương pháp gián tiếp có thể kiểm định được. Cho (2003) sử dụng mô hình SARIMA và ANN để dự báo lượng khách từ 6 khu vực tới Hồng Kông khi sử dụng dữ liệu trong thời gian từ tháng 01/1999 tới tháng 12/2000.

Chaitip và cộng sự (2008) áp dụng SARIMA, ARIMA, các mô hình Holt-Winters, mạng thần kinh, VAR, GMM, ARCH-GARCH-M, ARCH-GARCH, TAR, PAR, EGARCH nhằm dự báo lượng khách du lịch tới Thái Lan. Mô hình SARIMA đưa ra kết quả tốt nhất. Tương tự như vậy, Suhartono (2011) cũng thực hiện những phương pháp mới với dữ liệu khách theo đường hàng không tới Bali. Một lần nữa mô hình SARIMA là mô hình tốt nhất dùng để dự báo.

Kodituwakku và cộng sự (2015) sử dụng 3 mô hình để dự báo lượng khách du lịch quốc tế tới Sri Lanka trong giai đoạn từ tháng 1/2010 tới tháng 8/2014. Mô hình mùa Holt - Winters được xem là mô hình tốt nhất dùng dự báo so với SARIMA, và mạng thần kinh. Nghiên cứu này không kiểm định yếu tố mùa vụ trong mô hình SARIMA.

Không chỉ dự báo lượng khách du lịch mà SARIMA còn là kỹ thuật dự báo phù hợp được vận dụng trong những lĩnh vực khác sử dụng dữ liệu chuỗi thời gian. Savas (2013) sử dụng mô hình SARIMA và Kalman để dự báo lạm phát hàng tháng ở Luxembourg, Mexico, Bồ Đào Nha và Thụy Sĩ. Reininger và Fingerlos (2007) sử dụng dữ liệu chuỗi GDP thực trong thời gian từ quý 1/1980 tới quý 4/2006 và tìm ra được mô hình phù hợp với mục đích giai thích được các đặc điểm mùa vụ của chuỗi thời gian. Kết quả cho thấy mô hình SARIMA được xem là tốt nhất trong nghiên cứu này.

2.2. Hướng tiếp cận ANN

Trong những năm gần đây, nghiên cứu về mạng thần kinh nhân tạo ANN đã mang lại nhiều quan tâm trong các lĩnh vực như sinh học, tâm lý học, y học, kinh tế, toán học và máy tính. Lý do cho sự quan tâm này là do ANN là kỹ thuật ước lượng toàn diện có thể sử dụng cho bất kỳ mô

hình tuyến tính hoặc phi tuyến (Cybenko, 1989; Funahashi, 1989; Hornik, Stinchcombe & White, 1989; Wasserman, 1989). Vì tính linh động trong quá trình ước lượng, ANN là phương pháp mạnh mẽ thực hiện các nhiệm vụ liên quan đến phân loại mẫu, hồi quy và dự báo biến liên tục (Kaastra & Boyd, 1996). Trong thập kỷ trước, mạng thần kinh chiếm ưu thế so với các phương pháp khác – trong đó có mô hình chuỗi thời gian ARIMA khi giải quyết được vấn đề về dữ liệu phi tuyến và bất thường (Hansen và cộng sự, 1999).

Phương pháp ANN được giới thiệu lần đầu về dự báo lượng cầu khách du lịch trong cuối những năm 1990. Bằng chứng thực nghiệm cho rằng ANN thực sự có kết quả dự báo lượng khách tốt hơn so với các mô hình chuỗi thời gian và đa biến. Ví dụ như, Pattie và Snyder (1996) ứng dụng mô hình mạng thần kinh lan truyền ngược (BNP) với 2 lớp ẩn để dự báo theo tháng lượng khách nghỉ qua đêm tại công viên quốc gia Mỹ. Dự báo của họ chính xác hơn so với các mô hình chuỗi thời gian truyền thống với mẫu nghiên cứu lớn.

Law và Au (1999) giới thiệu mạng thần kinh truyền thẳng với 6 nơ-ron đầu vào và một nơ-ron đầu ra để dự báo lượng khách tới Hồng Kông. Tác giả cho rằng có 6 yếu tố tác động tới lượng khách du lịch. Kết quả cho thấy, mô hình mạng thần kinh có kết quả tốt hơn so với các mô hình đa biến khác.

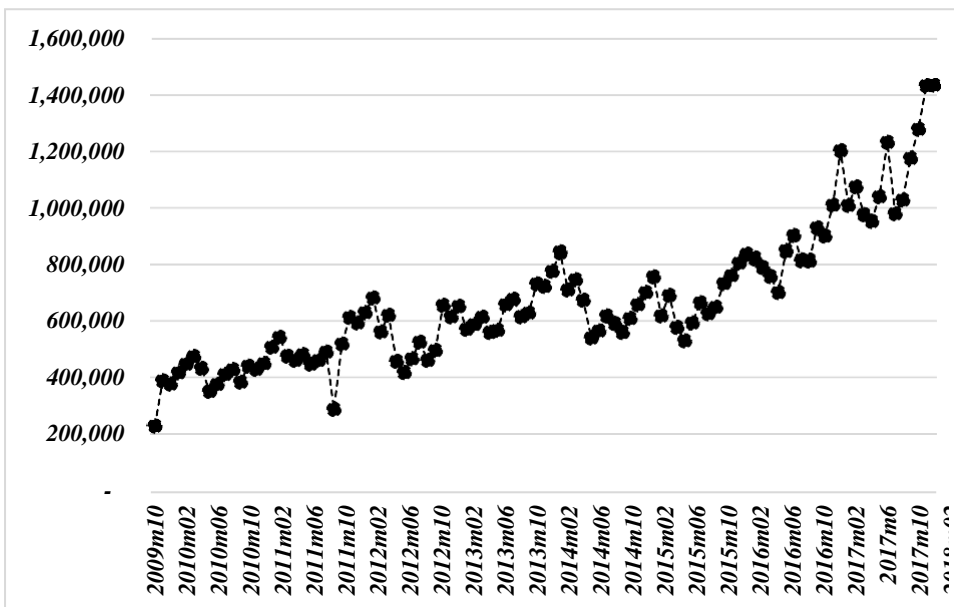
Law (2000) thử nghiệm mô hình mô hình BNP để dự báo lượng khách quốc tế Đài Loan tới Hồng Kông. Mô hình mạng thần kinh được đề nghị có 6 biến độc lập tương ứng với số nơ-ron đầu vào và có một nơ-ron đầu ra. Law sử dụng một hàm phi tuyến để tách dữ liệu ngẫu nhiên thành tập các dữ liệu để huấn luyện và kiểm tra. Khả năng dự báo của mô hình chính xác và bền vững.

Burger và cộng sự (2001) phát triển 8 cách tiếp cận để dự báo lượng khách du lịch từ Mỹ tới Durban ở Nam Phi trong giai đoạn 1992-1998. Tác giả cho rằng phương pháp ANN là mô hình giải thích tốt nhất so với 7 mô hình còn lại. Cũng tương tự như vậy, Cho (2003) kết luận rằng mô hình ANN thể hiện tốt nhất so với mô hình san bằng số mũ và ARIMA trong việc mô hình hóa và dự báo lượng cầu khách du lịch Hồng Kông. Kon và Turner (2005) cũng kết luận mô hình ANN thể hiện tốt hơn mô hình chuỗi thời gian cấu trúc, naïve 1 và Holt-Winterstrong quá trình dự khách lượng khách du lịch nước ngoài tới Singapore.

3. Cơ sở lý luận và phương pháp nghiên cứu

3.1. Mẫu nghiên cứu

Nhằm xây dựng mô hình dự báo lượng khách quốc tế đến Việt Nam hàng tháng cho mục tiêu dự báo mang tính ngắn hạn, tác giả sử dụng dữ liệu lượng khách quốc tế hàng tháng tới Việt Nam được đăng tải công khai trên website <http://vietnamtourism.gov.vn> của Tổng cục Du lịch Việt Nam từ tháng 10/2009 tới tháng 2/2018 gồm 101 quan sát. Trước thời gian đó lượng khách quốc tế có biến động mạnh liên quan tới bất ổn tình hình tài chính thế giới (khủng hoảng kinh tế) nên tác giả không sử dụng giai đoạn này. Ngoài ra, để tập trung cho mục đích du lịch, đáng lẽ tác giả sử dụng số liệu lượng khách quốc tế đến Việt Nam theo mục đích du lịch nhưng do đã bỏ tờ khai khi nhập cảnh vào Việt Nam nên từ tháng 01/2015 báo cáo khách quốc tế đến Việt Nam hàng tháng sẽ không có số liệu về lượng khách phân theo các mục đích. Trong nghiên cứu này, tác giả chỉ sử dụng dữ liệu lượng khách quốc tế đến Việt Nam trong suốt giai đoạn nghiên cứu. Trong đó, 92 quan sát đầu tiên (hay còn được gọi là dữ liệu trong mẫu) từ tháng 10/2009 tới tháng 5/2017 được dùng để xây dựng hàm dự báo và 9 quan sát còn lại - từ tháng 6/2017 tới tháng 2/2018 được dùng để kiểm tra tính chính xác dự báo. Đặc biệt đối với mô hình ANN, 92 quan sát đầu tiên tiếp tục chia thành 64 quan sát để huấn luyện và 28 quan sát để kiểm tra (tỷ lệ 70%:30%). Dữ liệu được thu thập từ một nguồn công bố thông tin duy nhất nên sẽ đảm bảo nguyên tắc nhất quán trong suốt quá trình phân tích.



Nguồn: Tổng cục Du lịch Việt Nam

Hình 1. Lượng khách quốc tế đến Việt Nam từ tháng 10/2009 đến tháng 02/2018

3.2. Mô hình SARIMA

Hai tác giả George Box và Gwilym Jenkins (1976) đã nghiên cứu mô hình tự hồi qui tích hợp trung bình trượt (Autoregressive Integrated Moving Average), viết tắt là ARIMA. Tên của họ (Box-Jenkins) được dùng để gọi cho các quá trình ARIMA tổng quát áp dụng vào phân tích và dự báo các chuỗi thời gian. Mô hình tích hợp trung bình trượt tự hồi quy (mô hình ARIMA) được tích hợp từ 3 quá trình: Tự hồi quy (AR) đối với các số liệu kinh tế - xã hội, trung bình trượt (MA) đối với phần sai số của số liệu (hay phần nhiễu ngẫu nhiên) và quá trình tích hợp hay Sai phân (I) nhằm biến đổi chuỗi không dừng thành chuỗi dừng trước khi thực hiện các thao tác phân tích và dự báo khác. Bản chất của mô hình ARIMA là dự báo giá trị tương lai của một biến số (biểu thị theo chuỗi thời gian) dựa trên giá trị quá khứ và các sai số ngẫu nhiên. Tuy nhiên, mô hình ARIMA chỉ thích hợp cho việc phân tích, dự báo dữ liệu chuỗi thời gian không có yếu tố mùa vụ.

Để có những dự báo chính xác các chỉ tiêu kinh tế - xã hội, người ta đã đi sâu tìm hiểu, nghiên cứu và đề xuất một số phương pháp dự báo dữ liệu chuỗi thời gian có yếu tố mùa vụ. Trong đó, những mô hình được phát triển tiếp từ mô hình ARIMA được xem là thích hợp hơn cả, đó là những mô hình như X11ARIMA, X12ARIMA hoặc SARIMA. Mô hình SARIMA ra đời muộn hơn, vào nửa cuối thập niên 1990. Nó được phát triển từ mô hình ARIMA và giải quyết yếu tố mùa vụ đồng thời cho cả 3 quá trình: tự hồi quy, tích hợp và trung bình trượt. Vì thế, SARIMA phức tạp hơn X11ARIMA và X12ARIMA - những mô hình được phát triển từ ARIMA chỉ ở quá trình trung bình trượt. Mô hình SARIMA được phát triển tiếp từ mô hình ARIMA phù hợp với bất kỳ dữ liệu chuỗi thời gian mùa vụ nào có thể là 4 quý trong năm; 7 ngày trong tuần; 11, 12 tháng trong một năm... Nếu chuỗi dữ liệu quan sát có tính mùa vụ, thì mô hình ARIMA tổng quát lúc này là SARIMA(p,d,q)(P,D,Q)L (với P và Q lần lượt là bậc của thành phần mùa AR và MA, D là bậc sai phân có tính mùa, L là số thời đoạn trong một vòng chu kỳ).

Bài viết ứng dụng mô hình SARIMA trong phân tích và dự báo lượng khách quốc tế đến Việt Nam, được thực hiện theo 4 bước sau đây:

Bước 1 - Nhận dạng mô hình: Xác định các giá trị (D, d, p, P, q, Q). Trong đó, trước hết cần xác định bậc sai phân theo mùa vụ D, sai phân thường d và thực hiện biến đổi chuỗi thành chuỗi dừng. Thành phần d của mô hình được nhận dạng thông qua kiểm định tính dừng của chuỗi thời gian. Nếu chuỗi thời gian dừng ở bậc không ta có I(d=0), nếu sai phân bậc 1 của chuỗi dừng ta có I(d=1), nếu sai phân bậc 2 của chuỗi dừng ta có I(d=2). Phương pháp kiểm định tính dừng thường được áp dụng là kiểm định Augmented Dickey-Fuller (ADF) và Perron-Phillips (PP). Sau đó, kiểm tra biểu đồ của hàm tự tương quan (Autocorrelation Function - ACF), và hàm tự

tương quan riêng phần (Partial Autocorrelation Function - PACF) tại các trễ mùa vụ và trễ thường để xác định bậc tự hồi quy p và tự hồi quy mùa vụ P , bậc trung bình trượt q và trung bình trượt mùa vụ Q .

Nếu biểu đồ của hàm tự tương quan chỉ có q giá trị đầu tiên là khác và các giá trị của biểu đồ tương quan riêng phần giảm từ từ ta có thể tiên đoán có một $MA(q)$.

Nếu biểu đồ của hàm tự tương quan riêng phần chỉ có p giá trị đầu tiên là khác 0 và các giá trị của biểu đồ tương quan đơn giản từ từ ta có thể tiên đoán có một $AR(p)$.

Nếu biểu đồ của hàm tự tương quan và biểu đồ của hàm tự tương quan riêng phần không có sự cắt ngắn như hai trường hợp trên, ta sẽ có một quá trình $ARMA$ và các thông số của nó tùy thuộc vào dạng cụ thể của các biểu đồ tương quan.

Bước 2 – Ước lượng mô hình: Ước lượng các tham số, sử dụng phương pháp ước lượng bình phương nhỏ nhất (OLS-Ordinary Least Square) để ước lượng giá trị các tham số này. Người ta thường sử dụng phần mềm như SPSS, Minitab, Eviews để ước tính tham số mô hình.

Bước 3 – Kiểm định: Kiểm định tính hợp lý của mô hình $SARIMA$ được lựa chọn, bao gồm kiểm định các tham số và kiểm định phần dư. Nếu kiểm định mô hình được lựa chọn không thỏa mãn thì quay lại từ giai đoạn nhận dạng để lựa chọn mô hình khác hợp lý hơn. Chuỗi giá trị thặng dư là một nhiễu trắng. Nếu nó không phải là một nhiễu trắng ta kết luận mô hình là không hoàn chỉnh và ta phải thêm vào mô hình các bậc bổ sung cần thiết.

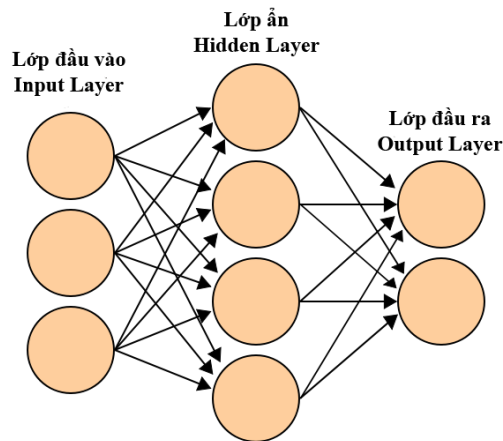
Bước 4 - Dự báo: Dựa trên mô hình được lựa chọn thực hiện dự báo giá trị tương lai của dữ liệu chuỗi mùa vụ, cũng như đưa ra khoảng tin cậy của dự báo. Giá trị tương lai có thể được dự báo cho thời điểm kế tiếp hoặc mùa vụ kế tiếp. Khi có dữ liệu quan sát mới, nên đưa điểm quan sát này vào mô hình để làm ngắn thời đoạn dự báo. Nếu đặc điểm, hành vi của chuỗi có thể bị thay đổi bởi dữ liệu mới, có thể phải ước lượng lại các tham số mô hình, hoặc xây dựng mô hình mới.

3.3. Mô hình ANN

3.3.1. Giới thiệu

Mạng thần kinh nhân tạo ANN là một mô phỏng xử lý thông tin, được nghiên cứu ra từ hệ thống thần kinh của sinh vật, giống như bộ não để xử lý thông tin. Nó bao gồm số lượng lớn các mối gắn kết cấp cao để xử lý các yếu tố làm việc trong mối liên hệ giải quyết vấn đề rõ ràng. ANN được giới thiệu đầu tiên vào năm 1943 bởi nhà thần kinh học Warren McCulloch và nhà logic học Walter Pitts. Những năm gần đây, mô phỏng ANN xuất hiện và phát triển nhiều, các nghiên cứu ứng dụng đã được thực hiện trong các ngành điện, điện tử, kỹ thuật chế tạo, y học, quân sự, kinh tế... và mới nhất là các nghiên cứu ứng dụng trong lĩnh vực quản lý dự án xây dựng. Tại Việt Nam, việc nghiên cứu ứng dụng ANN vào quản lý chỉ mới bắt đầu trong vài năm gần đây nhưng nhanh chóng thu hút được sự quan tâm và cần được phát triển. Trong lĩnh vực kinh tế, mặc dù đã được thừa nhận và ứng dụng từ nhiều năm trước nhưng nó vẫn còn khá mới do sự phức tạp của nó. Nhiều nhà kinh tế đã tiến hành hàng loạt các nghiên cứu tập trung vào việc ứng dụng mạng thần kinh nhân tạo trong việc xác định mối quan hệ và dự báo các nhân tố, biến số kinh tế, chẳng hạn như tốc độ tăng trưởng Tổng sản phẩm quốc nội, lạm phát, tỷ giá hối đoái, giá chứng khoán, ... ở nhiều nước và đã đạt được nhiều kết quả khả quan.

Mạng thần kinh nhân tạo là một hệ thống xử lý thông tin được phỏng theo cách thức xử lý thông tin của nơ-ron sinh học, bao gồm rất nhiều các nơ-ron nhân tạo hoạt động song song. Một nhóm các nơ-ron được sắp xếp sao cho tất cả chúng đều nhận được các tín hiệu đầu vào tại cùng một thời điểm, sau đó xử lý và cho ra các tín hiệu đầu ra cùng một lúc, được gọi là một lớp mạng (Layer). Mạng thần kinh nhân tạo đơn giản nhất có hai lớp: một lớp đầu vào (Input layer) nhận các tín hiệu đầu vào và một lớp đầu ra (Output layer) cho các tín hiệu đầu ra của mạng. Lớp đầu vào thực chất không phải là các nơ-ron thực, các nơ-ron hoạt động đơn giản chỉ nhằm giới thiệu các tín hiệu, thông tin vào (giá trị của biến đầu vào). Chúng ta cũng có thể kết hợp nhiều lớp mạng tạo thành mạng đa lớp, các lớp nằm giữa lớp đầu vào và lớp đầu ra gọi là các lớp ẩn (Hidden layers).



Nguồn: Xử lý của tác giả

Hình 2. Cấu trúc mạng thần kinh nhân tạo

3.3.2. Quá trình huấn luyện mạng

Quá trình học của Mạng thần kinh

Mạng thần kinh được huấn luyện bằng cách đưa các cặp tín hiệu đầu vào và tín hiệu đầu ra vào mạng. Các cặp số liệu này liên tục được đưa vào mạng và mạng nhanh chóng học mối quan hệ giữa đầu vào và đầu ra, quá trình này được gọi là quá trình mạng được huấn luyện. Sau đó, khi ta đưa tín hiệu vào mạng, mạng sẽ xử lý dựa trên mối quan hệ giữa biến vào và biến ra đã học để cho ra tín hiệu ra tương ứng. Quá trình huấn luyện thực chất là việc điều chỉnh các trọng số kết nối của mạng. Trong quá trình huấn luyện, các trọng số được điều chỉnh đến khi đạt được các giá trị sao cho với mỗi tín hiệu đầu vào thì mạng sẽ cho ra tín hiệu đầu ra gần với giá trị đầu ra mục tiêu nhất; nói cách khác, sai số giữa giá trị đầu ra của mạng và giá trị đầu ra mục tiêu là nhỏ nhất.

Hàm truyền

Hàm truyền hay còn gọi là hàm kích hoạt là một thành phần không thể thiếu trong mô hình mạng thần kinh. Hàm truyền giúp cho thông tin được truyền từ nơ-ron này đến các nơ-ron khác, và kết quả của hàm truyền là thông tin đầu ra của mỗi lớp ẩn và lớp đầu ra.

Phần mềm cho mô hình ANN đều có sẵn các loại hàm truyền, hoặc là dạng hàm truyền tuyến tính, hoặc là dạng hàm phi tuyến để xây dựng mô hình nhưng được phân hóa thành các dạng cụ thể như hàm HyperTanh, Tanh, TanhAxon, LinearTanh, LinearSigmoid, Arctan, Arcotan, sin, cos...

Thuật toán lan truyền ngược

Thuật toán lan truyền ngược (BackPropagation) là một thuật toán điều chỉnh trọng số được sử dụng rất phổ biến. Thuật ngữ truyền ngược đề cập đến chiều truyền của sai số. Phương pháp lan truyền ngược sử dụng một tập hợp các giá trị đầu vào và đầu ra để tìm ra mạng nơ-ron thần kinh mong muốn. Một tập hợp đầu vào được đưa vào một hệ thống giả định trước nào đó để tính ra giá trị đầu ra, sau đó giá trị đầu ra này được so sánh với giá trị giá trị thực. Nếu không có sự khác biệt nào, thì không cần thực hiện một quá trình kiểm tra nào, ngược lại các trọng số sẽ được thay đổi trong quá trình lan truyền ngược trong mạng thần kinh để giảm sự khác biệt đó.

3.4. Phần mềm sử dụng

3.4.1 Eviews

EViews (Econometric Views) phiên bản 6.0 là phần mềm chuyên về kinh tế lượng, nghiên cứu với dữ liệu chuỗi thời gian, dữ liệu chéo, dữ liệu mảng... Với khả năng linh hoạt trong thao tác, quản lý dữ liệu dễ dàng, kết quả hiển thị nhanh và dễ dàng hiểu được đã giúp Eviews đang trở thành một trong những phần mềm thống kê và phân tích dữ liệu được các nhà nghiên cứu sử dụng phổ biến nhất. Phần mềm này sẽ được sử dụng dùng để phân tích và hồi quy mô hình SARIMA trong nghiên cứu

3.4.2. SPICE-MLP

Spice-MLP phiên bản 2.2 là phần mềm mạng nơ ron 3 lớp, với nhiều đầu vào và nhiều đầu ra. Spice-MLP được viết với mục đích hướng dẫn sinh viên và nghiên cứu sinh học tập và sử dụng mạng nơ ron để mô hình hóa nhiều loại dữ liệu khác nhau. Hiện tại Spice-MLP đang được nhiều bạn trên thế giới sử dụng. Spice-MLP có giao diện với tiếng Việt, tiếng Anh và tiếng Nhật. Spice-MLP được viết bởi khi tác giả làm việc tại Soft Intelligence Laboratory, Ritsumeikan University, Japan, 2003-2007 và thường xuyên được cập nhật theo yêu cầu của người sử dụng.

4. Kết quả nghiên cứu

4.1. Phương pháp chuỗi thời gian SARIMA

4.1.1. Nhận dạng mô hình

Chuỗi số liệu lượng khách quốc tế đến Việt Nam sử dụng trong mô hình SARIMA theo phương pháp Box-Jenkins được giả định là chuỗi dừng, vì vậy để dự báo lượng khách quốc tế bằng mô hình này cần phải xem xét chuỗi dữ liệu nghiên cứu có dừng hay chưa. Trước tiên, dựa vào việc quan sát đồ thị của chuỗi số liệu, sau đó tiến hành kiểm tra tính dừng này thông qua hai kiểm định phổ biến: Augmented Dickey-Fuller (ADF) và Perron-Phillips (PP) được gọi là kiểm định nghiệm đơn vị (unit root test).

Hình 1 trong cho thấy, chuỗi dữ liệu nghiên cứu chưa dừng, ta cần lấy sai phân bậc 1 chuỗi dữ liệu và tiến hành hai kiểm định ADF và PP theo kết quả Bảng 1.

Bảng 1: Kết quả kiểm định ADF và PP

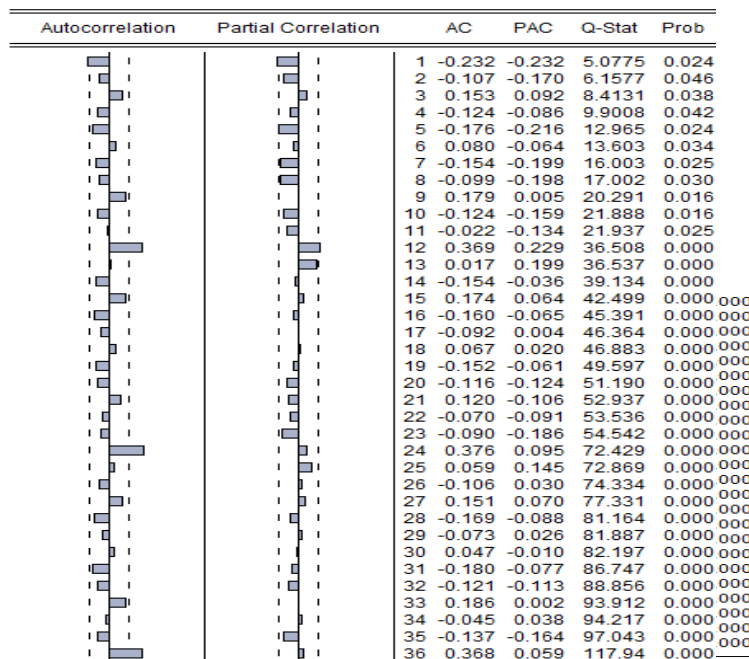
Kiểm định	Giá trị t	Xác suất
ADF	- 12.0821	0.0001
PP	- 12.1572	0.0001

Các giá trị tới hạn ở mức ý nghĩa thống kê 1%, 5%, 10% tương ứng là: -3.505, -2.894, -2.584

Nguồn: Xử lý số liệu của tác giả

Kết quả của cả hai kiểm định ADF và PP đều cho phép ta bác bỏ giả thuyết H_0 về tính dừng của dữ liệu ở mức ý nghĩa 1% tức là dữ liệu sau khi lấy sai phân bậc 1 đã dừng. Tại đây ta xác định được $d=1$

Tiếp đó, để xác định giá trị p, q của mô hình SARIMA, ta cần dựa vào biểu đồ hàm tự tương quan ACF và tự tương quan từng phần PACF. Trong biểu đồ PACF ở Hình 3, các hệ số tương quan riêng phần khác không có ý nghĩa ở các độ trễ 1, 5 và 12 sau đó tắt dần về 0.



Nguồn: Xử lý số liệu của tác giả

Hình 3. Biểu đồ ACF VÀ PACF

Còn đối với biểu đồ ACF, ta có các hệ số tương quan khác không có ý nghĩa ở các độ trễ 1, sau đó tắt dần về 0. Như vậy, biểu đồ chỉ ra rằng ta nên chọn p (1, 5, 12) và q (1) cho thành phần không có tính mùa.

Ngoài ra, Hình 3 cũng cho thấy có những đỉnh nhọn ở các độ trễ 12, 24 và 36 trên ACF sau đó tắt hết về 0, gợi ý rằng thành phần MA có tính mùa cần được xem xét trong mô hình. Trên PACF tồn tại những đỉnh nhọn ở độ trễ 12 sau đó tắt hết về 0, do đó thành phần AR có tính mùa cũng phải cần được bao gồm. Và điều này cũng có nghĩa nên chọn P = 1, Q = 3 và L = 12 cho thành phần có tính mùa. Sai phân theo mùa D = 1 giống với sai phân thường.

Tóm lại, các dạng mô hình SARIMA được nhận diện bao gồm SARIMA (1,1,1) (1,1,3)₁₂, SARIMA (5,1,1) (1,1,3)₁₂, SARIMA (12,1,1) (1,1,3)₁₂.

4.1.2. Ước lượng mô hình

Các mô hình đã nhận diện được so sánh tính phù hợp dựa trên các thông số kiểm định sau hồi quy bao gồm: R² điều chỉnh, công cụ thông tin Akaike (AIC), công cụ Schwarz (SC) để lựa chọn mô hình phù hợp nhất. Thông số R² điều chỉnh phải càng lớn, trong khi đó AIC và SC phải càng nhỏ thì càng tốt, mô hình sẽ càng phù hợp.

Từ Bảng 2 ta thấy, mô hình SARIMA (12,1,1) (1,1,3)₁₂ là mô hình thỏa mãn nhiều nhất các tiêu chuẩn sử dụng vì có thông số R² lớn nhất, thông số AIC và SC nhỏ nhất do đó đây là mô hình được vận dụng vào việc dự báo.

Bảng 2: Các mô hình SARIMA (p,d,q) (P,D,Q)_L thử nghiệm

Mô hình	R ² điều chỉnh	AIC	SC
SARIMA (1,1,1) (1,1,3) ₁₂	0.717	24.283	24.494
SARIMA (5,1,1) (1,1,3) ₁₂	0.718	24.319	24.537
SARIMA (12,1,1) (1,1,3) ₁₂	0.814	23.798	24.029

Nguồn: Xử lý số liệu của tác giả

Kết quả ước lượng Bảng 3 cho thấy, có 6 hệ số có ý nghĩa ở mức 1%. Cũng trong Bảng 3, SAR (thể hiện điều kiện chạy mô hình mang tính thời vụ) được thêm vào mô hình khi ACF ở khoảng thời gian mùa vụ (12 tháng) là dương và SMA (thể hiện điều kiện chạy mô hình mang tính thời vụ) cũng được thêm vào nếu như ACF ở khoảng thời gian mùa vụ (12 tháng) là âm.

Mô hình sau đó được kiểm tra mức độ phù hợp với chuỗi dữ liệu nghiên cứu bằng cách phân tích phần dư.

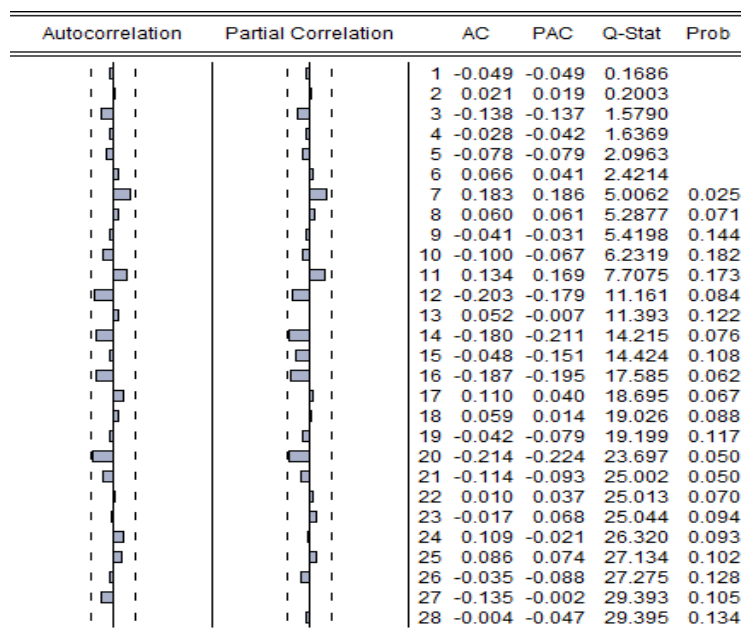
Bảng 3: Kết quả ước lượng của mô hình SARIMA (12,1,1) (1,1,3)₁₂

Biến	Hệ số	Sai số chuẩn	Thống kê t	Xác suất
C	-81221.500	105270	-0.771	0.443
AR (12)	-0.372	0.124	-3.006	0.004
SAR (12)	1.052	0.059	17.762	0.000
MA (1)	-0.480	0.124	-3.878	0.000
SMA (12)	-0.336	0.055	-6.110	0.000
SMA (24)	-0.308	0.000	-232529.8	0.000
SMA (36)	0.905	0.058	15.662	0.000

Nguồn: Xử lý số liệu của tác giả

4.1.3. Kiểm định phần dư

Biểu đồ ACF của phần dư ở Hình 4 cho thấy, không có thanh nào vượt quá 2 đường biên cho thấy sai số là một nhiễu trắng. Ngoài ra, kết quả kiểm định Breusch-Godfrey ở mức ý nghĩa 1% cũng cho thấy không tồn tại hiện tượng tự tương quan bậc 2.



Nguồn: Xử lý số liệu của tác giả

Hình 4. Biểu đồ ACF VÀ PACF của phân dư

Kết quả kiểm tra mô hình SARIMA (12,1,1) (1,1,3)₁₂ bằng kiểm định Breusch – Godfrey là thích hợp và có thể sử dụng để dự báo (Bảng 4).

Bảng 4: Kết quả kiểm định Breusch-Godfrey

Kiểm định	Giá trị t	Xác suất
Breusch-Godfrey	2.168	0.373

Nguồn: Xử lý số liệu của tác giả

4.2. Phương pháp mạng thần kinh nhân tạo ANN

4.2.1. Chuẩn hóa dữ liệu

Chuẩn hóa dữ liệu đến việc phân tích và chuyển đổi các biến đầu vào và đầu ra để tối thiểu độ nhiễu, loại bỏ tính xu hướng. Có khá nhiều cách để xử lý dữ liệu, trong đó được sử dụng phổ biến đó là lấy sai phân bậc nhất và lấy logarit tự nhiên. Theo Lê Đạt Chí (2010), dữ liệu thô sẽ được lấy logarit tự nhiên đối với đầu vào và đầu ra. Phép biến đổi logarit tự nhiên hữu ích đối với dữ liệu mà trong đó có thể xuất hiện đồng thời cả giá trị rất lớn, rất bé. Tác giả quyết định sử dụng phép biến đổi logarit tự nhiên đối với dữ liệu khách quốc tế (đơn vị tính: triệu người) trong nghiên cứu này nhằm đạt được hiệu quả cao.

4.2.2. Phân chia dữ liệu

Trong giai đoạn huấn luyện ANN, nghiên cứu chia chuỗi dữ liệu lượng khách quốc tế đến Việt Nam được thu thập thành ba tập dữ liệu lần lượt được gọi là: Tập hợp dữ liệu huấn luyện, kiểm tra và dự báo. Quá trình phân chia này được thực hiện một cách ngẫu nhiên.

Tập hợp dữ liệu huấn luyện là tập hợp có số lượng mẫu lớn nhất trong số 3 tập hợp vừa nêu, sử dụng 70% số quan sát để huấn luyện mạng ANN trong tổng số 92 quan sát dùng để huấn luyện và kiểm tra.

Tập hợp dữ liệu kiểm tra, nghiên cứu sử dụng số lượng 30% số quan sát ngẫu nhiên trong tổng số 92 quan sát dùng để huấn luyện và kiểm tra, quá trình này được dùng để ước lượng khả năng khái quát hóa trong phân tích mà ANN đã học tập được sau khi trải qua bước huấn luyện trước đó.

Tập hợp dữ liệu đánh giá kết quả dự báo. Bước cuối cùng trong quá trình huấn luyện là kiểm tra hiệu quả hoạt động của ANN đã trải qua tập hợp huấn luyện và kiểm tra là xác minh hiệu quả dự báo của ANN trên một tập hợp dữ liệu mà ANN chưa từng nhìn thấy trước đó. Tập hợp dữ liệu này gồm 9 quan sát từ tháng 6/2017 tới tháng 2/2018.

4.2.3. Thiết kế mạng nơ-ron

Sự phức tạp của mô hình mạng thần kinh nhân tạo là khi xây dựng mô hình không những phải chọn bộ dữ liệu cho mô hình mà còn phải lựa chọn cấu trúc mạng phù hợp. Bên cạnh một số kinh nghiệm rút ra từ thực nghiệm của các nhà nghiên cứu thì việc xây dựng được một mô hình mạng thần kinh tốt nhất với các thông số được lựa chọn thích hợp cần được thực hiện dựa trên phương pháp thử và sai.

Tác giả lựa chọn thủ tục đưa dần biến vào để xây dựng mô hình, và do số lượng cấu trúc cần thực hiện thử và sai rất lớn nên mỗi cấu trúc tác giả thực hiện huấn luyện một lần, sau đó chọn ra 10 cấu trúc có MSE của dữ liệu kiểm tra nhỏ nhất để thực hiện thêm 10 lần cho từng cấu trúc. Cuối cùng, tác giả tìm ra mô hình Mạng thần kinh với MSE dữ liệu kiểm tra nhỏ nhất.

Số nơ-ron đầu vào

Số nơ-ron đầu vào của mô hình chính là số biến độc lập trong mô hình tuyến tính truyền thống. Dựa theo nghiên cứu của Lin và công sự (2011), số nơ-ron lớp đầu vào sẽ là 12 tương ứng với lượng khách du lịch có độ trễ 12 tháng trước đó so với tháng dự báo. Điều này gần tương tự như tính mùa trong mô hình SARIMA đã phân tích trước đây.

Số lớp ẩn

Việc sử dụng càng nhiều lớp ẩn khiến cho việc huấn luyện trở nên chậm càng lớn. Một số nghiên cứu đã chỉ ra rằng đối với phần lớn các bài toán cụ thể, chỉ cần sử dụng một lớp ẩn cho mạng là đủ. Tăng số lượng lớp ẩn có thể làm gia tăng thời gian tính toán và gặp phải vấn đề khít quá mức, điều này sẽ dẫn tới dự báo ngoài mẫu kém hiệu quả. Với số lượng quan sát không nhiều nên trong phạm vi nghiên cứu này, tác giả chỉ sử dụng một lớp ẩn, một lớp đầu vào và một lớp đầu ra.

Hàm truyền

Trong phạm vi phần mềm SPICE-MLP, tác giả sử dụng hai hàm truyền phi tuyến phổ biến nhất trong kinh tế đó là hàm HyperTanh và Sigmoid, chọn ra hàm truyền có MSE nhỏ nhất để tiến hành dự báo. Hàm kích hoạt trong mạng MLP được chia thành hai loại là hàm kích hoạt đầu vào lớp ẩn và hàm kích hoạt đầu ra lớp ẩn. Kết quả thực hiện chạy mô hình cho thấy hàm kích hoạt đầu vào lớp ẩn nên là hàm HyperTanh và hàm kích hoạt đầu ra lớp ẩn nên là hàm HyperTanh vì có MSE nhỏ nhất.

Bảng 5: Kết quả so sánh hàm kích hoạt

Hàm kích hoạt đầu vào lớp ẩn	Hàm kích hoạt đầu ra lớp ẩn	MSE nhỏ nhất
HyperTanh	HyperTanh	0.01018
HyperTanh	Sigmoid	0.19265
Sigmoid	Sigmoid	0.19336
Sigmoid	HyperTanh	0.01332

Nguồn: Xử lý số liệu của tác giả

4.2.3.4. Số nơ-ron trong lớp ẩn

Quá trình xây dựng các lớp ẩn cho mạng ANN là một quá trình thử và sai. Ta sẽ thay đổi số lớp ẩn và số nơ-ron trong mỗi lớp ẩn. Để tìm ra mô hình ANN tốt nhất, tác giả lựa chọn 7 mô hình sau đây để đưa vào phân tích:

Bảng 6: Kết quả so sánh mô hình với các số nơ-ron lớp ẩn

Số biến vào	Số Nơ-ron lớp ẩn	Số biến ra	Ký hiệu	MSE của tập dữ liệu kiểm tra
12	25	1	12-25-1	0.01040
12	26	1	12-26-1	0.01031
12	27	1	12-27-1	0.01029
12	28	1	12-28-1	0.01045
12	29	1	12-29-1	0.01018
12	30	1	12-30-1	0.01036
12	31	1	12-31-1	0.01033
12	32	1	12-32-1	0.01025

Nguồn: Xử lý số liệu của tác giả

Để tìm ra mô hình tối ưu với số nơ-ron lớp ẩn cần phải huấn luyện nhiều lần trên mỗi mô hình – số lần lặp mặc định là 100.000 lần theo Lin và cộng sự (2011). Mỗi 10.000 lần lặp, tác giả sẽ kiểm tra lại MSE của tập dữ liệu kiểm tra để kiểm tra MSE đã thấp nhất hay chưa. Kết quả cuối cùng nghiên cứu chỉ ra mô hình ANN-12-29-1 tối ưu nhất với 29 nơ-ron lớp ẩn.

Các thông số được sử dụng để huấn luyện Mạng thần kinh được thể hiện ở bảng 7:

Bảng 7: Thông số huấn luyện

Tên thông số	Các lựa chọn
Mạng thần kinh sử dụng	Mạng thần kinh truyền thẳng đa lớp MLP (Multi Layer Perceptron)
Phần mềm sử dụng	Spice-MLP
Phương pháp chuẩn hóa dữ liệu	Logarit tự nhiên
Số lượng nơ-ron đầu vào	12
Số lượng nơ-ron lớp ẩn	1
Số lượng nơ-ron đầu ra	29
Hàm kích hoạt đầu vào lớp ẩn	Tanh Hyperbolic
Hàm kích hoạt đầu ra lớp ẩn	Tanh Hyperbolic
MSE yêu cầu	0.01
Số lần lặp	100.000

Nguồn: Xử lý số liệu của tác giả

4.3. So sánh kết quả dự báo hai phương pháp

Sử dụng các chỉ tiêu so sánh hiệu quả dự báo như MSE, RMSE, MAE và hệ số MAD đã được trình trong tổng quan nghiên cứu để thực hiện so sánh hoạt động dự báo của hai mô hình: Mô hình chuỗi thời gian SARIMA và mô hình mạng thần kinh nhân tạo (ANN) có kết quả theo bảng 8:

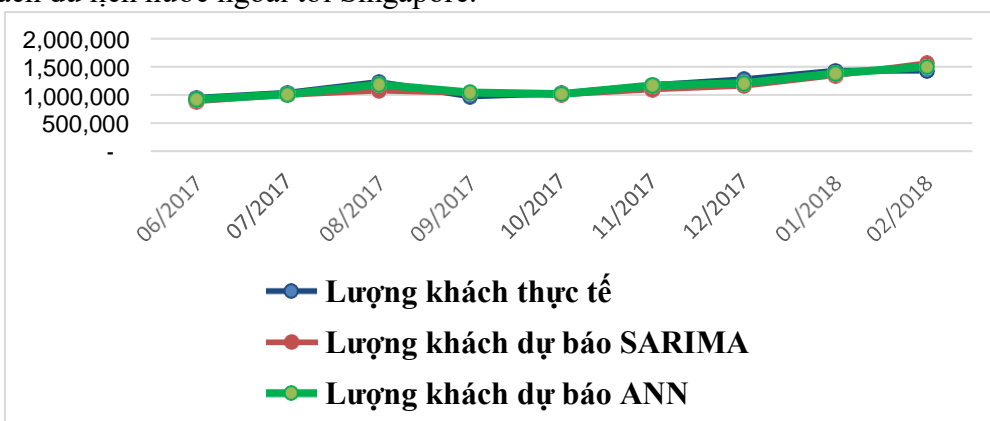
Bảng 8: Kết quả dự báo hai phương pháp

Thời gian	Lượng khách thực tế	Lượng khách dự báo SARIMA	Chênh lệch lượng khách	Lượng khách dự báo ANN	Chênh lệch lượng khách
06/2017	949,362	889,549	-6.30%	923,584	-2.72%
07/2017	1,036,880	1,005,056	-3.07%	1,011,563	-2.44%
08/2017	1,229,163	1,076,734	-12.40%	1,188,292	-3.33%
09/2017	975,952	1,037,479	6.30%	1,044,403	7.01%
10/2017	1,024,899	1,007,445	-1.70%	1,014,770	-0.99%
11/2017	1,172,568	1,099,246	-6.25%	1,169,029	-0.30%
12/2017	1,276,353	1,163,207	-8.86%	1,206,737	-5.45%
01/2018	1,430,242	1,353,802	-5.34%	1,387,074	-3.02%
02/2018	1,431,845	1,563,050	9.16%	1,501,122	4.84%
	RMSE	90154.89		46291.01	
	MAE	0.0839		0.0417	
	MAPE	0.066004		0.033441	
	MAD	79684.44		39571.78	

Nguồn: Xử lý số liệu của tác giả

Từ bảng trên, ta thấy rằng các chỉ số đánh giá hiệu quả của mô hình dự báo MSE, RMSE, MAE và MAD của 2 mô hình khác nhau và các hệ số của mô hình ANN nhỏ hơn (khoảng gần 2 lần) so với mô hình SARIMA. Có thể nói rằng khả năng tổng quát hóa của mô hình mạng thần kinh nhân tạo ANN tốt hơn so với mô hình chuỗi thời gian SARIMA; hay nói cách khác, mô hình mạng thần kinh nhân tạo ANN đã được xây dựng trong nghiên cứu này cho kết quả dự báo lượng khách quốc tế đến Việt Nam từ tháng 6/2017 tới tháng 2/2018 tốt hơn mô hình SARIMA. Kết quả này phù hợp với các nghiên cứu trước đây của các tác giả Pattie và Snyder (1996) dự

báo theo tháng lượng khách nghỉ qua đêm tại công viên quốc gia Mỹ, Law và Au (1999) dự báo lượng khách các khu vực tới Hồng Kông, Law (2000) dự báo lượng khách quốc tế Đài Loan tới Hồng Kông, Burger và cộng sự (2001) dự báo lượng khách du lịch từ Mỹ tới Durban ở Nam Phi, Cho (2003) dự báo lượng cầu khách du lịch Hồng Kông, Kon và Turner (2005) dự báo lượng khách du lịch nước ngoài tới Singapore.



Hình 5. Đồ thị dự báo lượng khách quốc tế đến Việt Nam từ tháng 6/2017 tới tháng 2/2018

5. Kết luận

Nghiên cứu sử dụng phương pháp Box-Jenkins (1976) và mạng thần kinh nhân tạo ANN để lập các mô hình và dự báo lượng khách quốc tế đến Việt Nam. Mô hình SARIMA là một mô hình chuỗi thời gian tuân theo quá trình tự hồi quy tích hợp với trung bình trượt với độ trễ thời gian là 12 hay SARIMA (12,1,1) (1,1,3)12. Mô hình ANN-12-29-1 được lựa chọn có 12 nơ-ron đầu vào, 29 nơ-ron lớp ẩn và 1 nơ-ron đầu ra cùng các thông số khác. So sánh kết quả với dự báo giữa hai mô hình theo các tiêu chuẩn báo MSE, RMSE, MAE và MAD kết quả cho thấy rằng mô hình ANN dự báo tốt hơn so với SARIMA. Điều này phù hợp với nhiều kết quả các tác giả như Burger và cộng sự (2001), Cho (2003) ...

Nghiên cứu còn có một số hạn chế như sau:

Thứ nhất, độ dài của dữ liệu ngắn chỉ có 101 tháng dẫn tới số quan sát đưa vào mô hình còn hạn chế, nên quá trình dự báo lượng khách ngoài mẫu trong tương lai dựa trên hai mô hình nghiên cứu chỉ trong thời gian ngắn, nếu dự báo trong thời gian dài sẽ dẫn tới sai số lớn. Dữ liệu có nhiều quan sát hơn có thể áp dụng trên một lớp ẩn có thể dự báo sẽ chính xác hơn.

Thứ hai, biến độc lập hay nơ-ron lớp đầu vào chỉ xét tới dữ liệu quá khứ của khách quốc tế mà thiếu đi các biến kiểm soát có tác động như GDP, lạm phát, tỷ giá hối đoái, sự kiện bất thường... dẫn tới mức độ giải thích của mô hình còn chưa cao.

Thứ ba, phần mềm sử dụng dự báo ANN là SPICE-MLP còn khá nhiều hạn chế do đây là phần mềm dùng cho học tập và nghiên cứu đơn giản, thiếu một số tính năng, chưa mang tính chuyên nghiệp và thương mại cao như Neuro Solutions.

Hướng nghiên cứu trong tương lai tác giả cố gắng hoàn thiện mô hình ANN hơn bằng cách sử dụng phần mềm chuyên nghiệp, có độ tin cậy cao dùng để áp dụng dự báo không chỉ trong lĩnh vực du lịch mà còn trong những lĩnh vực khác được quan tâm như tài chính, y tế, giáo dục... Ngoài ra, trong thời gian tới tác giả có thể nâng cao nghiên cứu của mình bằng cách so sánh thêm nhiều mô hình tuyến tính lẫn phi tuyến khác, mở rộng mẫu nghiên cứu với các quốc gia trong khu vực và trên thế giới.

TÀI LIỆU THAM KHẢO

- [1] Box, G.E.P., and G.M. Jenkins (1976), Time Series Analysis: Forecasting and Control, Revised Edition, Holden Day, San Francisco.
- [2] Burger, C., Dohnal, M., Kathrada, M., & Law, R. (2001), A practitioners guide to time-series methods for tourism demand forecasting: A case study of Durban, South Africa, *Tourism Management*, 22, 403–409.

- [3] Chaitip, P., Chaiboonsri and R. Mukhjang (2008), Time Series Models for Forecasting International Visitor Arrivals to Thailand, International Conference on Applied Economics, 2008, pp 159-163.
- [4] Cho, V. (2003), A comparison of three different approaches to tourist arrival forecasting, Tourism Management, 24(3), 323-330.
- [5] Cybenko, G. (1989), Approximation by superpositions of a sigmoidal function. Mathematical Control, Signal and Systems, 2, 303–314.
- [6] Kodituwakku, W., Wijesundara, W., & Hettiarachchi, C. (2015), Modelling and Forecasting Tourism Demand for Sri Lanka, Colombo: University of Colombo School of Computing.
- [7] Funahashi, K. (1989), On the approximate realization of continuous mappings by neural networks, Neural Networks, 2, 183–192.
- [8] Hansen, J. V., McDonald, J. B., & Nelson, R. D. (1999), Time series prediction with genetic-algorithm designed neural networks: An empirical comparison with modern statistical models, Computational Intelligence, 15(3), 171–184.
- [9] Hornik, K., Stinchcombe, M., & White, H. (1989), Multilayer feedforward networks are universal approximators, Neural Networks, 2(5), 359–366.