

ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ





# BÁO CÁO

## ĐỒ ÁN (TIỂU LUẬN, BÀI TẬP LỚN)

**Học kỳ I, năm học 2021 - 2022**

**Học phần:**

**PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

***Đề tài:*** ***Xây dựng mô hình dự báo về tình hình covid-19 tỉnh Thừa Thiên Huế***

**Số phách**

*(Do hội đồng chấm thi ghi)*

Thừa Thiên Huế, ngày 18 tháng 1 năm 2022



ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ





# BÁO CÁO

## ĐỒ ÁN (TIỂU LUẬN, BÀI TẬP LỚN)

**Học kỳ I, năm học 2021 - 2022**

**Học phần:**

### **PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

***Đề tài: Xây dựng mô hình dự báo về tình hình covid-19 tỉnh Thừa Thiên Huế***

###### Giảng viên hướng dẫn: Nguyễn Quang Lịch

###### Sinh viên thực hiện: Nguyễn Trịnh Tấn Đạt

###### Lớp: Khoa học dữ liệu & Trí tuệ nhân tạo

**Số phách**

*(Do hội đồng chấm thi ghi)*

Thừa Thiên Huế, ngày 18 tháng 1 năm 2022

ĐẠI HỌC HUẾ

KHOA KỸ THUẬT VÀ CÔNG NGHỆ



PHIẾU ĐÁNH GIÁ ĐỒ ÁN/TIỂU LUẬN/BÀI TẬP LỚN

**Học kỳ I, năm học 2021 - 2022**

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1** | **Cán bộ chấm thi 2** |
| **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT1:**  Bằng số: .........................................  Bằng chữ: ....................................... | **Nhận xét:**  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  ..............................................................  **Điểm đánh giá của CBCT2:**  Bằng số: .........................................  Bằng chữ: ....................................... |

Điểm kết luận: ...........................................................................................................

Bằng số:.....................................................................................................................

Bằng chữ: .................................................................................................................

*Thừa Thiên Huế, ngày tháng năm 2021*

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1**  *(Ký và ghi rõ họ và tên)* | **Cán bộ chấm thi 2**  *(Ký và ghi rõ họ và tên)* |

**MỤC LỤC**

[CHƯƠNG I. MỞ ĐẦU 6](#_Toc93416135)

[CHƯƠNG II. TỔNG QUAN NGHIÊN CỨU 6](#_Toc93416136)

[I. Tính cấp thiết của đề tài 6](#_Toc93416137)

[II. Mục tiêu nghiên cứu 7](#_Toc93416138)

[III. Đối tượng nghiên cứu 7](#_Toc93416139)

[IV. Phạm vi nghiên cứu 7](#_Toc93416140)

[1. Phạm vi nội dung 7](#_Toc93416141)

[2. Phạm vi không gian 7](#_Toc93416142)

[3. Phạm vi thời gian 7](#_Toc93416143)

[V. Phương pháp nghiên cứu 7](#_Toc93416144)

[1. Phương pháp thu thập tài liệu: 7](#_Toc93416145)

[2. Phương pháp thu thập và xử lý số liệu: 7](#_Toc93416146)

[VI. Bố cục báo cáo nghiên cứu 7](#_Toc93416147)

[CHƯƠNG III. TỔNG QUAN NGHIÊN CỨU 7](#_Toc93416148)

[VII. Tham khảo tài liệu 7](#_Toc93416149)

[VIII. Khoảng trống nghiên cứu 8](#_Toc93416150)

[IX. Dữ liệu và phương pháp 8](#_Toc93416151)

[1. Mô tả dữ liệu 8](#_Toc93416152)

[2. Mô tả về mô hình 8](#_Toc93416153)

[3. Mô hình Trung bình trượt Tích hợp Tự động Điều chỉnh (ARIMA) 9](#_Toc93416154)

[4. Quy trình trung bình động (MA) 9](#_Toc93416155)

[5. Quy trình trung bình động tích hợp tự động phục hồi (ARIMA) 9](#_Toc93416156)

[6. Nhận dạng mô hình 9](#_Toc93416157)

[7. Ước tính tham số và xác thực mô hình 10](#_Toc93416158)

[CHƯƠNG IV. KẾT QUẢ 10](#_Toc93416159)

1. MỞ ĐẦU

Các triệu chứng như viêm phổi, sốt, khó thở và nhiễm trùng phổi thường được hiển thị bởi coronavirus thuộc một họ vi rút. Tổ chức Y tế Thế giới (WHO) vào ngày 11 tháng 2 năm 2020 đã công bố tên chính thức của loại vi rút này là COVID-19, có dạng đầy đủ là Bệnh truyền nhiễm do vi rút Corona và 19 chỉ ra rằng nó có nguồn gốc từ năm 2019. Vụ bùng phát COVID-19 diễn ra đầu tiên tại thành phố Vũ Hán của Trung Quốc vào ngày 29 tháng 12 năm 2019 là nguyên nhân bùng phát dịch bệnh đường hô hấp. Tổng Giám đốc WHO đã tuyên bố sự bùng phát của COVID-19 là Tình trạng Khẩn cấp Y tế Công cộng cần được Quốc tế Quan tâm vào ngày 30 tháng 1 năm 2020. Chẩn đoán nhanh đóng vai trò rất quan trọng trong quản lý dịch bệnh và ổ dịch. Việc chẩn đoán nhanh chóng và chính xác của một loại thuốc đặc hiệu cho phép các biện pháp giám sát, phòng ngừa và kiểm soát sức khỏe cộng đồng nhanh chóng và chính xác. Chẩn đoán nhanh, vắc xin và điều trị là những can thiệp dược phẩm quan trọng để hạn chế lây truyền các bệnh truyền nhiễm qua đường hô hấp. COVID -19 là một trường hợp khẩn cấp về sức khỏe cộng đồng được quốc tế quan tâm nhưng cho đến nay. COVID-19 là một loại vi-rút mới đang ảnh hưởng xấu đến toàn thế giới vì nó lây lan chủ yếu qua tiếp xúc với người đó. Nó lây từ người này sang người khác giữa những người tiếp xúc gần trong vòng 6 feet. Hầu hết các quốc gia đã chậm lại trong việc sản xuất các sản phẩm của họ. Một số vắc xin COVID-19 đã được phê duyệt và phân phối ở nhiều quốc gia khác nhau, những quốc gia này đã bắt đầu các chiến dịch tiêm chủng hàng loạt . Các biện pháp phòng ngừa khác bao gồm cách xa về mặt thể chất hoặc xã hội , cách ly , thông gió cho các không gian trong nhà, che những vết ho và hắt hơi, rửa tay và giữ tay chưa rửa tránh xa mặt. Việc sử dụng khẩu trang hoặc khăn che mặt đã được khuyến nghị ở những nơi công cộng để giảm thiểu rủi ro lây truyền. Trong khi nghiên cứu đang được tiến hành để phát triển các loại thuốc ức chế vi rút, thì phương pháp điều trị chính là điều trị triệu chứng. Quản lý liên quan đến việc điều trị các triệu chứng, chăm sóc hỗ trợ , cách ly và các biện pháp thử nghiệm .

1. TỔNG QUAN NGHIÊN CỨU
   1. Tính cấp thiết của đề tài

Trong sự ảnh hưởng của Covid-19 đến đời sống và sự phát triển kinh tế của nhà nước nói chung và địa bàn thị xã Hương Trà nói riêng vẫn chưa có phương pháp giải quyết triệt để, số ca nhiễm Covid-19 vẫn tăng lên hằng ngày. Người dân được yêu cầu ở nhà và chỉ ra ngoài cho những nhu cầu cực kỳ cần thiết, và tất cả các trường học và đại học đều bị đóng cửa. Để kiểm soát sự lây lan của căn bệnh này, học sinh trung học cơ sở và trung học phổ thông không được phép đến trường, đồng nghĩa với việc các em không có cơ hội gặp bạn cùng lớp và giáo viên. An toàn và hạnh phúc vì các triệu chứng COVID19 ở họ ít nghiêm trọng hơn so với người lớn . Thanh thiếu niên không thể ra ngoài để gặp gỡ bạn bè, bạn học và các thành viên trong gia đình. Vì vậy, họ bị cô lập tại nhà của họ, phát sinh những vấn đề mới như thất nghiệp. Ảnh hưởng đến nền kinh tế thông qua các hoạt động doanh nghiệp bị đình trệ.

* 1. Mục tiêu nghiên cứu

Mục tiêu nghiên cứu nhằm đưa ra mô hình dự báo tình hình covid-19 trên đại bàn thị xã Hương Trà, Thừa Thiên Huế. . Bên cạnh đó, tìm hiểu, đánh giá và phân tích tác động của đại dịch Covid- 19 đến đời sống. Từ vấn đề đó đề xuất các dự báo kịp thời, giải pháp, kiến nghị đẩy lùi đại dịch Covid-19.

* 1. Đối tượng nghiên cứu

Số liệu covid-19 tỉnh Thừa Thiên Huế.

* 1. Phạm vi nghiên cứu
     1. Phạm vi nội dung

Tập trung nghiên cứu số liệu tỉnh Thừa Thiên Huế trước tác động khó lường của đại dịch covid-19.

* + 1. Phạm vi không gian

Tập trung nghiên cứu trong phạm vi tỉnh Thừa Thiên Huế.

* + 1. Phạm vi thời gian

Nghiên cứu thực trạng covid-19 tỉnh Thừa Thiên Huế qua số liệu từ ngày 24/12/2022 đến 15/1/2022, dự báo ngày 16/1/2022 đến ngày 28/1/2022.

* 1. Phương pháp nghiên cứu
     1. Phương pháp thu thập tài liệu:

Bài nghiên cứu tham khảo nguồn tài liệu từ nhiều bài báo, thông thông tin điện tử, ... cùng các tạp chí, luận văn , tài liệu nghiên cứu của các tỉnh thành khác.

* + 1. Phương pháp thu thập và xử lý số liệu:

Thông tin và số liệu về số ca nhiễm, điều trị khôi phục và tử vong từ nhiều nguồn khác nhau, tiến hành chắt lọc thông tin và xử lý để đánh giá quy mô, bản chất, sự khác nhau của đối tượng nghiên cứu theo thời gian và không gian.

* 1. Bố cục báo cáo nghiên cứu

Tham khảo tài liệu

Đánh giá biểu đồ

Dự báo

1. TỔNG QUAN NGHIÊN CỨU
   1. Tham khảo tài liệu

Đại dịch Covid- 19 đã và đang diễn ra một cách khó lường tại hầu khắp các nước trên thế giới trong đó có Việt Nam. Đây cũng được coi là vấn đề nhận được nhiều sự quan tâm và nhiều công trình nghiên cứu được khai thác.

Bài nghiên cứu “**mô hình SEIR-C19”** do **tác giả GS. TS. Joseph T Wu, TS. Kathy Leung, GS. TS. BS. Gabriel M Leung (Trung tâm Hợp tác của WHO về Nghiên cứu dịch bệnh và Kiểm soát Bệnh truyền nhiễm, Khoa Y tế công cộng, Trường Y Li Ka Shing, Đại học Hồng Kông, Hồng Kông, Trung Quốc) thực hiện. Bài nghiên cứu này** chẩn đoán bệnh do coronavirus ở những người bệnh nghi nhiễm, về tiên lượng những người bệnh mắc Covid-19 và để phát hiện người trong cộng đồng chung có nguy cao nhiễm Covid-19 hoặc phải nhập viện vì căn bệnh này.

Bài nghiên cứu “**Factors associated with viral clearance periods from patients with COVID-19: A retrospective observational cohort study” do tác giả** Nobuyasu Hirai, Yuichi Nishioka, Kei Kasahara thực hiện. Sau cùng, kết quả ngụ ý rằng thời kỳ khai báo virus từ bệnh nhân với COVID - 19 có thể tùy vào tuổi ; do đó, phân loại các giai đoạn cách ly dựa trên tuổi có thể được được xem xét. Một khoảng thời gian thanh thải virus đồng đều cho những bệnh nhân lớn tuổi có thể khó xác định vì những thành kiến ​​chẳng hạn như điều kiện y tế. Theo dõi thêm và nghiên cứu phân tử, như cũng như các nghiên cứu thuần tập lớn hơn, được khuyến nghị để xác định thời gian thanh thải virus và thời gian cách ly tối ưu.

* 1. Khoảng trống nghiên cứu

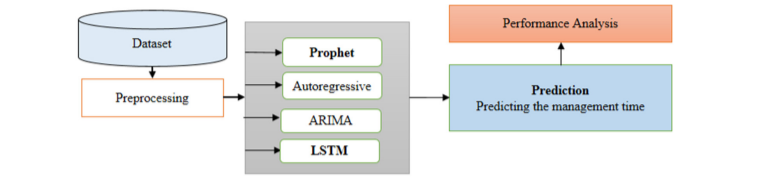
Từ tổng quan tài liệu nghiên cứu trong và ngoài nước, có thể thấy rằng đề tài nghiên cứu về mô hình dự báo đại dịch Covid-19 thu hút rất nhiều sự quan tâm của những nhà nghiên cứu. Tuy nhiên, một phần vì vấn đề này còn nhiều sự mới mẻ, các “ngõ ngách” của vấn đề vẫn chưa được khai thác hết.

* 1. Dữ liệu và phương pháp
     1. Mô tả dữ liệu

Thông tin cập nhật thường xuyên về các trường hợp COVID-19 được xác nhận chính thức được thu thập từ trang web chính thức của tỉnh Thừa Thiên Huế.

Bảng số liệu : bao gồm số ca nhiễm hằng ngày, đang điều trị, hồi phục từ ngày 24/12/2021 đến ngày 11/1/2022.

* + 1. Mô tả về mô hình



* + 1. Mô hình Trung bình trượt Tích hợp Tự động Điều chỉnh (ARIMA)

 Là phương pháp dự báo yếu tố nghiên cứu một cách độc lập ( dự báo theo chuỗi thời gian). Bằng các thuật toán sử dụng độ trễ sẽ đưa ra mô hình dự báo thích hợp.  Mô hình đã chọn sẽ được kiểm tra dựa trên dữ liệu lịch sử để xem trình tự có được trình bày chính xác hay không.

* + 1. Quy trình trung bình động (MA)

Mô hình này sử dụng các lỗi trong quá khứ như một biến phụ thuộc. là một quá trình nhiễu trắng, một chuỗi các biến ngẫu nhiên được phân phối độc lập và giống hệt nhau (iid): và sau đó đặt hàng mô hình MA được đưa ra là:

Mô hình này được mô tả theo các lỗi trong quá khứ và do đó, chúng tôi ước tính các hệ số . Do đó, chỉ có q lỗi sẽ ảnh hưởng đến cấp, nhưng lỗi đơn đặt hàng cao hơn không ảnh hưởng . Điều này chỉ ra rằng nó là một mô hình bộ nhớ ngắn.

#### Tự động hồi quy (AR)

Theo mô hình tự hồi quy có bậc p , AR ( p ) có thể được biểu thị bằng:

Mô hình được mô tả theo các giá trị trong quá khứ và do đó chúng tôi muốn ước tính các hệ số và sử dụng mô hình để dự báo. Tất cả các giá trị trước đó sẽ có tác động tích lũy đối với giá trị hiện cócấp, là một mô hình bộ nhớ dài hạn.

* + 1. Quy trình trung bình động tích hợp tự động phục hồi (ARIMA)

Các phương pháp mô hình hóa ARIMA được sử dụng trong nghiên cứu này dựa trên một phương pháp phổ biến có sẵn để mô hình hóa và dự báo dữ liệu chuỗi thời gian. ARIMA là loại mô hình chuỗi thời gian phổ biến nhất có thể được tạo ra "cố định" bằng cách sai phân (nếu cần), có thể kết hợp với các phép biến đổi phi tuyến tính như ghi nhật ký hoặc giảm phát (nếu cần)

ARIMA (p, d, q) là mô hình ARIMA tổng quát không theo mùa: trong đó p là số điều khoản tự hồi quy, d là số lượng chênh lệch và q là số lượng điều khoản trung bình động. Mô hình tiếng ồn trắng được phân loại là ARIMA (0, 0, 0) vì không có phần AR vì không phụ thuộc vào , không có sự khác biệt liên quan và cũng không có phần MA vì  không dựa vào . Ví dụ, nếu không cố định, chúng tôi coi trọng điểm khác biệt đầu tiên là  để có thể  trở nên đứng yên.  (d = 1 ngụ ý sai lệch một lần)

là một mô hình ARIMA (p, 1, q). Mô hình đi bộ ngẫu nhiên được phân loại là ARIMA (0, 1, 0) vì không có phần AR và MA tham gia và chỉ tồn tại một điểm khác biệt.

* + 1. Nhận dạng mô hình

Dữ liệu cần thiết phải cố định để phát triển các mô hình chuỗi thời gian. Nếu dữ liệu không cố định được sử dụng trong một mô hình, kết quả có thể cho thấy một mối quan hệ bị sai lệch. Do đó, dữ liệu chuỗi thời gian phải được kiểm tra xem có tĩnh không trước khi xác định mô hình.

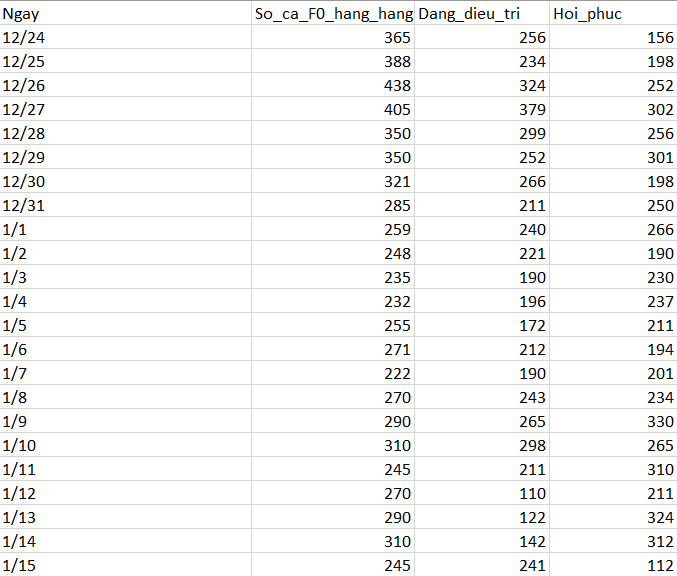
Nói chung, một chuỗi thời gian là đứng yên nếu nó được mô tả bằng giá trị trung bình và phương sai không đổi, và phương sai tự động không phụ thuộc vào thời gian. Nếu bất kỳ yêu cầu nào trong số này không được đáp ứng, dữ liệu sẽ được coi là không ổn định. Hàm tự tương quan (ACF) sẽ được sử dụng để xác định vấn đề này, và nếu biểu đồ ACF là dương và hiển thị một mẫu phân rã tuyến tính rất chậm, dữ liệu là không cố định. Vấn đề không ổn định có thể được giải quyết bằng cách phân biệt dữ liệu thích hợp nếu nó là do biến đổi trung bình hoặc mô hình gây ra bởi phương sai. Tự tương quan một phần (PACF) được đặc trưng như một mối tương quan tuyến tính giữa Y t và Y (tk), kiểm soát các tác động có thể có của mối quan hệ tuyến tính giữa các giá trị độ trễ trung gian. Tiếp theo là xác định các giá trị ban đầu cho các đơn hàng theo mùa và không theo mùa (P và q).

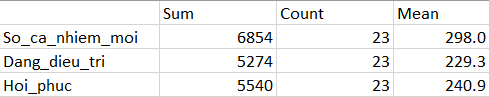
* + 1. Ước tính tham số và xác thực mô hình

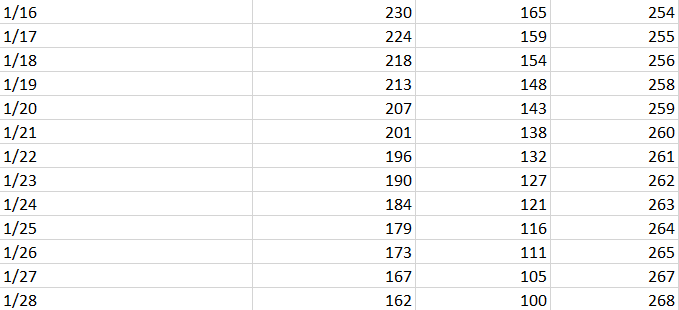
Sau khi xác định thứ tự ARIMA thích hợp (p, d, q), chúng tôi cố gắng tìm các ước lượng chính xác của các tham số mô hình bằng cách sử dụng các bình phương nhỏ nhất như Box và Jenkins mô tả. Các tham số thu được bằng xác suất lớn nhất đối với chuỗi thời gian, chính xác về mặt tiệm cận. Đối với các phân bố Gauss, các ước lượng nói chung là đầy đủ, hiệu quả và nhất quán và tiệm cận bình thường và hiệu quả đối với các phân phối không phải Gauss. Trong nghiên cứu này, phần mềm STATA phiên bản 15 và SPSS phiên bản 25 được sử dụng để phát triển mô hình ARIMA. Mức ý nghĩa thống kê được đặt là 0,05. Các mô hình được chọn trong giai đoạn cuối đã được xác thực bằng cách sử dụng các phương pháp bao gồm lỗi bình phương trung bình gốc (RMSE), lỗi phần trăm tuyệt đối trung bình (MAPE) và chuẩn hóa tiêu chí thông tin Bayes (BIC).

1. KẾT QUẢ

Dữ liệu tổng thể về sự phân bố COVID-19 được thu thập và phân tích từ ngày 24 tháng 12 năm 2021 đến ngày 15 tháng 1 năm 2022.



 Tổng cộng 6854 trường hợp COVID-19 đã được quan sát từ ngày 24 tháng 12 năm 2021 đến ngày 15 tháng 1 năm 2022 và tỷ lệ mắc cho thấy có xu hướng giảm từng ngày, với tỷ lệ tăng vào giữa tháng 1. Giá trị trung bình mỗi ngày số ca nhiễm, đang điều trị, hồi phục từ ngày 24 tháng 12 năm 2021 đến ngày 15 tháng 1 năm 2022 lần lượt là 298, 229, 241.

Dữ liệu nhiễm bệnh hằng ngày từ ngày 16 tháng 1 đến ngày 28 tháng 1 được dự đoán bằng cách sử dụng mô hình ARIMA (0,1,5), số liệu điều trị được dự đoán bằng cách sử dụng mô hình ARIMA (2,1,3) và số liệu hồi phục được dự đoán bằng cách sử dụng mô hình ARIMA (0,1,3) dựa trên mức lan truyền của COVID- 19 từ ngày 24 tháng 12 năm 2021 đến ngày 15 tháng 1 năm 2022. Dự báo hàng ngày là dự báo điểm với giới hạn tin cậy 95% của các giá trị cận trên và cận dưới.

Biểu đồ diễn tả mô hình và dự báo sự lây lan của COVID-19 ở Thừa Thiên Huế bằng các thuật toán ARIMA. Do đó, các dự báo cho thấy sự lây lan của COVID-19 đã được xác nhận và các trường hợp điều trị có sự đồng biến và giảm dần hàng ngày trong những ngày tới, đồng thời số ca hồi phục cũng tăng lên từng ngày.