

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

NGUYỄN PHÚC HIẾU

NGHIÊN CỨU ỨNG DỤNG PHƯƠNG PHÁP
HỌC MÁY VÀ HỌC SÂU DỰ BÁO NỒNG ĐỘ
BỤI PM_{2.5} TRONG KHÔNG KHÍ XUNG QUANH –
TRƯỜNG HỢP NGHIÊN CỨU CHO THÀNH PHỐ
HỒ CHÍ MINH

Ngành: Quản lý tài nguyên và môi trường

Mã số ngành: 9850101

TÓM TẮT LUẬN ÁN TIẾN SĨ

TP. Hồ Chí Minh - năm 2025

Công trình được hoàn thành tại: Khoa Môi trường, trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Tp. Hồ Chí Minh

Người hướng dẫn khoa học:

1. HDC: **PGS.TS. Đào Nguyên Khôi**
2. HDP: TS. Nguyễn Lý Sỹ Phú

Phản biện 1: PGS.TS. Tô Thị Hiền

Phản biện 2: PGS.TS. Hồ Quốc Bằng

Phản biện 3: GS.TS. Lê Hoài Bắc

Phản biện độc lập 1: GS.TS. Nguyễn Kim Lợi

Phản biện độc lập 2: PGS.TS. Nguyễn Lữ Phương

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án cấp cơ sở đào tạo họp tại Trường Đại học Khoa học Tự nhiên, vào hồi ... giờ ngày tháng ... năm

Có thể tìm hiểu luận án tại thư viện:

- Thư viện Khoa học Tổng hợp TP.HCM
- Thư viện Đại học Quốc gia TP.HCM
- Thư viện Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CÓ LIÊN QUAN ĐẾN LUẬN ÁN

1. Hiếu, N.P.; Vân, Đ.D.H.; Khôi, Đ.N. Xây dựng mô hình mô phỏng chỉ số chất lượng không khí cho thành phố Hồ Chí Minh bằng phương pháp học máy. Tạp chí Khí tượng Thủy văn 2024, 761, 13-24.
2. Hiếu, N.P.; Dương, N.N.; Linh, Đ.Q.; Khôi, Đ.N. Mô phỏng nồng độ bụi PM_{2.5} tại khu vực trung tâm Thành phố Hồ Chí Minh bằng thuật toán học máy và học sâu. Tạp chí Khí tượng Thủy văn 2024, 762, 62-72.
3. Nguyen, P.H.; Dao, N.K.; Nguyen, L.S.P. Development of Machine Learning and Deep Learning Prediction Models for PM_{2.5} in Ho Chi Minh City, Vietnam. Atmosphere 2024, 15, 1163. <https://doi.org/10.3390/atmos151011>

CHƯƠNG 1. MỞ ĐẦU

1.1 ĐẶT VẤN ĐỀ

Ô nhiễm không khí đang trở thành một trong những vấn đề thách thức nhất mà các quốc gia trên thế giới đang phải đối mặt. Ô nhiễm không khí không chỉ ảnh hưởng đến sức khỏe, mà còn ảnh hưởng đến kinh tế [1]. Bụi mịn $PM_{2.5}$ là một trong những chất gây ô nhiễm không khí đặc trưng, xuất hiện nhiều ở các thành phố lớn và cũng là tiêu chí đánh giá chất lượng không khí phổ biến hiện nay trên toàn thế giới. Thành phố Hồ Chí Minh (TP.HCM) là trung tâm kinh tế, chính trị, văn hóa và giáo dục quan trọng của cả nước. Tuy nhiên, cùng với sự phát triển về kinh tế thì TP.HCM cũng đang đối mặt với tình trạng ô nhiễm, trong đó có tình trạng ô nhiễm không khí, đặc biệt là ô nhiễm bụi mịn $PM_{2.5}$, dẫn đến nhu cầu ngày càng cao trong công bố thông tin về ô nhiễm bụi mịn hàng ngày, giúp cung cấp thông tin cảnh báo cho chính quyền thành phố và người dân.

Hiện nay, có nhiều phương pháp dự báo chất lượng không khí khác nhau. Trong những năm gần đây, cùng với sự phát triển của cuộc cách mạng công nghiệp 4.0, mô hình trí tuệ nhân tạo (Artificial Intelligence – AI) đã được sử dụng rộng rãi cho dự báo chất lượng không khí. Học máy (Machine Learning) và học sâu (Deep Learning) là các phương pháp tiên tiến trong AI, đặc biệt hiệu quả trong phân tích và dự báo các hiện tượng phức tạp, bao gồm chất lượng không khí. Với khả năng phân tích dữ liệu lớn từ nhiều nguồn như khí tượng, giao thông, và hoạt động công nghiệp, các mô hình học máy và học sâu có thể xác định các mối liên hệ phi tuyến tính và phức tạp giữa các yếu tố ảnh hưởng đến nồng độ $PM_{2.5}$. Các mô hình này không chỉ có khả năng dự báo chính xác nồng độ $PM_{2.5}$ trong ngắn hạn và dài hạn mà còn hỗ trợ thiết lập hệ thống cảnh báo sớm, giúp chính quyền và cộng đồng có các biện pháp ứng phó kịp thời, bảo vệ sức khỏe cộng đồng. Ngoài ra, việc

ứng dụng học máy và học sâu còn giúp tối ưu hóa chi phí, giảm thiểu sự phụ thuộc vào mạng lưới quan trắc vật lý, đặc biệt là trong điều kiện dữ liệu quan trắc không đầy đủ. Tuy nhiên, hiện nay việc xây dựng và ứng dụng mô hình AI dự báo chất lượng không khí tại TP.HCM vẫn còn hạn chế.

Xuất phát từ những thực trạng đã phân tích, việc ứng dụng phương pháp học máy và học sâu trong dự báo nồng độ bụi $PM_{2.5}$ cho TP.HCM là cần thiết. Do đó, tác giả đã thực hiện đề tài “Nghiên cứu ứng dụng phương pháp học máy và học sâu dự báo nồng độ bụi $PM_{2.5}$ trong không khí xung quanh – Trường hợp nghiên cứu cho thành phố Hồ Chí Minh”.

1.2 MỤC TIÊU

Mục tiêu chung của đề tài là xác định mô hình học máy và học sâu phù hợp cho nghiên cứu mô phỏng và dự báo nồng độ $PM_{2.5}$ tại TP.HCM dựa trên các điều kiện khí tượng cụ thể.

Mục tiêu chi tiết của đề tài gồm các mục tiêu sau:

- Thiết lập và tối ưu hóa các kịch bản đầu vào dựa trên các yếu tố khí tượng cụ thể tại TP.HCM cho mô hình mô phỏng và dự báo nồng độ bụi $PM_{2.5}$.
- Xây dựng, kiểm định, đánh giá các mô hình học máy và học sâu phù hợp cho nghiên cứu mô phỏng và dự báo nồng độ bụi $PM_{2.5}$ tại TP.HCM dựa trên các điều kiện khí tượng cụ thể.
- Ứng dụng kết quả của nghiên cứu trong mô phỏng và dự báo nồng độ bụi $PM_{2.5}$, hỗ trợ quản lý ô nhiễm không khí tại TP.HCM.

1.3 PHẠM VI VÀ GIỚI HẠN NGHIÊN CỨU

1.3.1. Đối tượng nghiên cứu

- Nồng độ bụi $PM_{2.5}$: Dữ liệu về nồng độ bụi $PM_{2.5}$ tại TP.HCM
- Các thông số khí tượng: Nhiệt độ, tốc độ gió, độ ẩm, số giờ nắng, lượng mưa, độ bốc hơi tại TP.HCM.

- Thuật toán học máy và học sâu: Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), General Regression Neural Network (GRNN), Artificial Neural Network (ANN), CNN (Convolutional Neural Network), Long Short-Term Memory (LSTM).

1.3.2. Phạm vi nghiên cứu

- Không gian: Thành phố Hồ Chí Minh.
- Thời gian: Dữ liệu được thu thập trong giai đoạn 2021 đến 2023.

1.3.3. Giới hạn

Do hạn chế về dữ liệu tự động liên tục và có độ chính xác cao, nghiên cứu sử dụng dữ liệu nồng độ bụi $PM_{2.5}$ từ trạm quan trắc chất lượng không khí tại tranh Lãnh sự quán Mỹ, cùng dữ liệu khí tượng từ trạm Tân Sơn Hòa.

Nồng độ bụi $PM_{2.5}$ còn bị ảnh hưởng bởi nhiều yếu tố khác như nguồn thải (khu công nghiệp, công trường xây dựng, khu vực giao thông đông đúc), nồng độ các chất ô nhiễm không khí khác (như NO_x , SO_x , CO_2 , H_2S) cũng có thể tương tác với nồng độ bụi $PM_{2.5}$. Tuy nhiên, các thông số kể trên chưa được nghiên cứu, tích hợp vào mô hình mô phỏng/dự báo nồng độ bụi $PM_{2.5}$.

1.4 Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN

1.4.1 Ý nghĩa khoa học

- Nghiên cứu góp phần ứng dụng phương pháp học máy và học sâu cho vấn đề mô phỏng và dự báo chất lượng không khí, mở ra những tiềm năng nghiên cứu mới trong lĩnh vực quản lý môi trường ứng dụng trí tuệ nhân tạo.
- Kết quả nghiên cứu có thể áp dụng cho các thông số chất lượng không khí khác tại thành phố Hồ Chí Minh, đồng thời có thể là cơ sở khoa học nhằm nhân rộng và áp dụng cho các khu vực khác.

1.4.2 Ý nghĩa thực tiễn

- Kết quả của đề tài phục vụ cho công tác nghiên cứu và quản lý chất lượng không khí, cụ thể trong hoạt động mô phỏng và dự báo nồng độ bụi $PM_{2.5}$ tại TP.HCM.

1.5 TÍNH MỚI

- Việc xây dựng mô hình cho cả mô phỏng và dự báo nồng độ bụi $PM_{2.5}$ TP.HCM là một điểm nổi bật của nghiên cứu này. Nghiên cứu cung cấp một cách tiếp cận đa chiều cho việc hiểu và dự đoán mức độ ô nhiễm bụi $PM_{2.5}$, cho phép kiểm tra tính linh hoạt và độ chính xác của các mô hình trong cả hai tác vụ. Điều này làm tăng giá trị và ứng dụng của nghiên cứu, đặc biệt là trong việc đưa ra các giải pháp và quyết định có thể cải thiện chất lượng không khí ở TP.HCM.
- Trong nghiên cứu này, các yếu tố đầu vào được lựa chọn dựa trên sự phù hợp với mô hình dự báo nồng độ $PM_{2.5}$ và tính khả dụng của dữ liệu. Các yếu tố này được sử dụng trong mô hình theo hướng xây dựng trên dữ liệu thực tế để tối ưu khả năng dự báo mà không đi sâu vào phân tích mối quan hệ giữa các yếu tố với nồng độ bụi $PM_{2.5}$.
- Việc ứng dụng nhiều thuật toán khác nhau giúp đánh giá và so sánh hiệu suất của chúng trong việc dự báo nồng độ bụi $PM_{2.5}$. Nghiên cứu áp dụng các thuật toán học máy và học sâu, từ các mô hình cơ bản đến phức tạp, giúp đánh giá hiệu quả từng mô hình. Điều này giúp xác định thuật toán nào hoạt động tốt nhất trong ngữ cảnh cụ thể của nghiên cứu, từ đó cung cấp thông tin quan trọng để lựa chọn mô hình phù hợp nhất cho các ứng dụng thực tế và phát triển các chiến lược dự báo và quản lý ô nhiễm không khí.

CHƯƠNG 2. TỔNG QUAN NGHIÊN CỨU

2.1. TỔNG QUAN VỀ BỤI $PM_{2.5}$

Bụi là một hỗn hợp chứa các hạt vô cơ và hữu cơ ở dạng lỏng hoặc rắn bay lơ lửng trong không khí, bao gồm sulfate, nitrat, amoniac, natri clorua, cacbon đen, bụi khoáng và nước [2]. Bụi mịn $PM_{2.5}$ là bụi có kích thước đường kính khí động học $\leq 2.5 \mu m$. Bụi trong không khí có thể được chia theo nguồn phát sinh gồm nguồn tự nhiên và nguồn nhân tạo; hoặc bụi sơ cấp và bụi thứ cấp. Bụi $PM_{2.5}$ có khả năng xâm nhập sâu vào cơ thể, gây bệnh hô hấp, và thâm nhập vào các túi phổi và hệ tuần hoàn. Tiếp xúc ngắn hạn với $PM_{2.5}$ có thể kích ứng mắt, mũi, họng, và phổi, dẫn đến ho, hắt hơi, sổ mũi, và khó thở. Bên cạnh đó, bụi mịn này cũng làm suy giảm chức năng phổi và nghiêm trọng hóa các bệnh như hen suyễn và tim mạch [3].

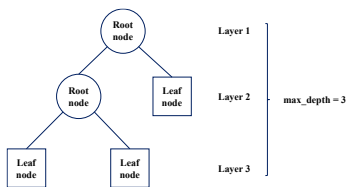
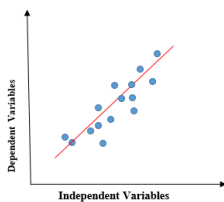
2.2. TỔNG QUAN DỰ BÁO CHẤT LƯỢNG KHÔNG KHÍ

2.2.1. Các mô hình dự báo chất lượng không khí

Các mô hình dự báo truyền thống dựa trên các quá trình khí tượng, vật lý và hóa học để mô tả sự hình thành và tích tụ của chất ô nhiễm trong không khí, yêu cầu hiểu biết về nguồn gây ô nhiễm và các quá trình biến đổi của chúng. Tuy nhiên, việc thu thập dữ liệu đáng tin cậy cho các mô hình này thường gặp khó khăn và phức tạp, ảnh hưởng đến độ chính xác của dự báo [4]. Các mô hình thống kê, dựa trên nguyên lý toán thống kê và xác suất, chủ yếu phản ánh mối liên hệ thống kê giữa các tác động khác nhau và các chất ô nhiễm trong chuỗi thời gian, áp dụng các dữ liệu quá khứ để dự báo chất lượng không khí thay vì dựa vào các quy trình vật lý, hóa học và sinh học. Với sự phát triển của trí tuệ nhân tạo trong những năm gần đây, các mô hình dự báo dựa vào trí tuệ nhân tạo đạt được các kết quả chính xác cao hơn và hiệu suất làm việc tốt hơn đáng kể trong việc dự báo chất lượng không khí.

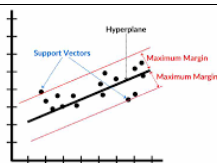
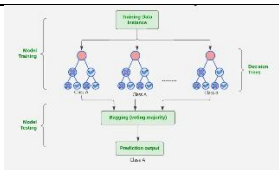
2.2.2. Cơ sở lý thuyết thuật toán học máy

Cấu trúc các thuật toán được trình bày tại Hình 2.1.



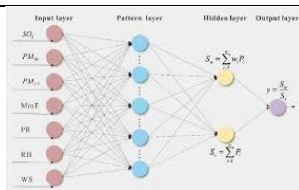
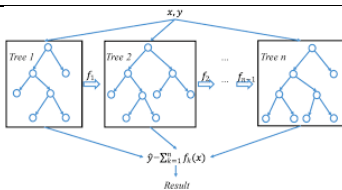
LR

DT



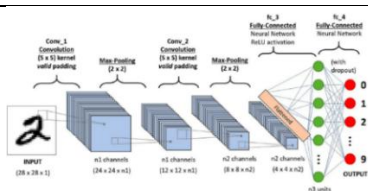
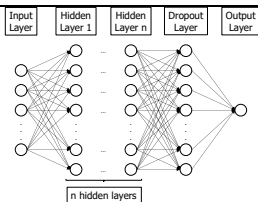
RF

SVR



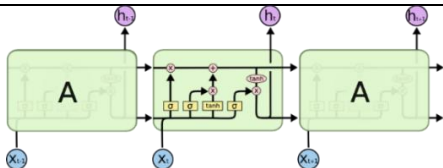
XGboost

GRNN



ANN

CNN



LSTM

Hình 2.1: Sơ đồ cấu trúc các thuật toán học máy và học sâu

2.3. TỔNG QUAN NGHIÊN CỨU TRONG NƯỚC VÀ NGOÀI NƯỚC

2.3.1. Ứng dụng học máy và học sâu trong mô phỏng và dự báo nồng độ bụi PM_{2.5} trên thế giới

Nghiên cứu của Bingyue Pan (2018) sử dụng mô hình XGBoost (XGB) dự báo nồng độ PM_{2.5} ở thành phố Thiên Tân, Trung Quốc [7]. Tại Tehran, Iran, nghiên cứu của Mehdi Zamani Joharestani và cộng sự (2019) sử dụng các mô hình Random Forest (RF), XGBoost (XGB) và Deep Learning (DL) để dự báo nồng độ PM_{2.5} tại thủ đô [11]. Ngoài ra, nghiên cứu của Laura Goulier và cộng sự (2020) sử dụng mô hình Multilayer Perceptron (MLP) để mô hình hóa ô nhiễm không khí đô thị. Trong nghiên cứu của nhóm tác giả Zaman và cộng sự (2021) đã tập trung vào đánh giá các mô hình học máy để ước tính nồng độ PM_{2.5} trên toàn bộ Malaysia [12]. Nghiên cứu của Lei và cộng sự (2022) sử dụng các phương pháp học máy để dự báo chất lượng không khí ở Macao, nhằm giúp quản lý môi trường và bảo vệ sức khỏe cộng đồng [13]. Một nghiên cứu của Doreswamy và cộng sự (2020) [15] sử dụng các thuật toán hồi quy để dự báo nồng độ PM_{2.5} tại huyện Bình Đông, Đài Loan. Một nghiên cứu của tác giả Kumar và cộng sự (2024) [18] đã thực hiện nghiên cứu dự đoán nồng độ PM_{2.5} tại Delhi, Ấn Độ, bằng các thuật toán máy học dựa trên các yếu tố khí tượng như tốc độ gió, nhiệt độ, độ ẩm và tầm nhìn. Nhóm nghiên cứu đã thử nghiệm các mô hình hồi quy tuyến tính, hồi quy cây quyết định, hồi quy rừng ngẫu nhiên (RF), hồi quy K-nearest neighbors (KNN), và hồi quy Lasso. Tại Afghanistan, Salehie và cộng sự (2024) [19] đã tiến hành nghiên cứu nhằm dự đoán nồng độ PM_{2.5}, sử dụng ba mô hình máy học: AdaBoost, Random Forest (RF) và Support Vector Machine (SVM), dựa trên dữ liệu khí tượng ERA5 từ 1998 đến 2020 và dữ liệu PM_{2.5} từ vệ tinh.

2.3.2. Ứng dụng học máy và học sâu trong mô phỏng và dự báo nồng độ bụi PM_{2.5} tại Việt Nam

Điển hình là một nghiên cứu của Mạc Duy Hưng và cộng sự (2017) sử dụng ANN để xây dựng mô hình dự báo nồng độ SO₂ cực đại ngày cho thành phố Hà Nội. Nghiên cứu của nhóm tác giả Ly và cộng sự (2021) [21] đã sử dụng các cảm biến giá rẻ để theo dõi nồng độ PM_{2.5} hàng giờ từ tháng 10 năm 2017 đến tháng 4 năm 2018 gồm hai địa điểm ở Hà Nội và một địa điểm ở Thái Nguyên. Một nghiên cứu khác là nghiên cứu của Vo và nhóm cộng sự được xuất bản năm 2021 [22] về dự đoán nồng độ bụi PM_{2.5} sử dụng học máy và mô hình WRF, trường hợp điển hình tại thành phố Hồ Chí Minh. Nghiên cứu của nhóm tác giả Rakholia và cộng sự (2022) [26] nhằm mục đích phát triển các mô hình dựa trên AI để dự báo nồng độ NO₂, SO₂, O₃, CO với các mức độ chi tiết hàng giờ và giá trị trung bình luân phiên trong 24 giờ. Hai và cộng sự (2023) [23] đã phát triển phương pháp sử dụng các mô hình máy học tiên tiến để dự đoán nồng độ PM_{2.5} tại tỉnh Bắc Ninh, Việt Nam. Trong nghiên cứu của mình, Minh Thanh Vo và cộng sự (2023) [24] đã phát triển mô hình PM25-CBL kết hợp mạng nơ-ron tích chập (CNN) và mô hình Bidirectional Long Short-Term Memory (Bi-LSTM) để dự đoán nồng độ PM_{2.5} tại TP.HCM, Việt Nam. Sử dụng dữ liệu quan trắc không khí từ ba trạm tự động trong giai đoạn 2020-2022, Dung và Than (2024) [25] đã phát triển mô hình mạng nơ-ron truyền thẳng (Feedforward neural network-FFNN) để dự đoán nồng độ PM_{2.5} tại tỉnh Bà Rịa - Vũng Tàu, Việt Nam. Gần đây, Rakholia và cộng sự (2024) [26] đã phát triển mô hình học máy đa biến dựa trên kỹ thuật ensemble để dự đoán nồng độ PM_{2.5} tại TP.HCM (TPHCM), Việt Nam. Trong nghiên cứu này, mô hình LightGBM, đã đạt kết quả tốt nhất khi dự báo nồng độ PM_{2.5} tại TPHCM, vượt trội hơn các mô hình khác như XGBoost, LSTM, và mô hình kết hợp CNN-LSTM.

2.4. TỔNG QUAN KHU VỰC NGHIÊN CỨU

TP.HCM nằm trong tọa độ địa lý khoảng $10^{\circ}10' - 10^{\circ}38'$ vĩ độ Bắc và $106^{\circ}22' - 106^{\circ}54'$ kinh độ Đông. Diện tích hơn 2.095 km^2 và được chia thành 16 quận, 1 thành phố và 5 huyện¹. TP.HCM nằm trong vùng nhiệt đới gió mùa cận xích đạo. Đặc điểm chung của nền khí hậu và thời tiết là nhiệt độ cao đều trong năm và có hai mùa mưa và mùa khô rõ ràng. TP.HCM là đô thị đặc biệt, một trung tâm lớn về kinh tế, tài chính, thương mại, văn hóa, khoa học công nghệ, đổi mới sáng tạo, giáo dục đào tạo của cả nước. Hàng năm, TP.HCM đóng góp gần 25% nguồn thu ngân sách và gần 22% GRDP của quốc gia.

Trong năm 2021, 100% các quận/huyện thành phố Hồ Chí Minh dưới ngưỡng quy chuẩn QCVN 05:2013/BTNMT. Nồng độ trung bình năm 2021 cấp quận/huyện biến thiên từ 12,9 đến 21,3 $\mu\text{g}/\text{m}^3$, giảm từ 7,3% đến 10,1% so với năm 2019 và tăng từ 0,4% đến 2,8% so với năm 2020. Nồng độ bụi $\text{PM}_{2.5}$ trung bình năm 2021 ở TP. Hồ Chí Minh cao ở các quận trung tâm, thấp ở các huyện xa như Củ Chi, Cần Giờ. Nồng độ bụi $\text{PM}_{2.5}$ ở TP.HCM năm 2021 cao trong các tháng 11, 12, 1, 2, 3 và thấp trong các tháng còn lại.

Theo kết quả kiểm kê phát thải năm 2017 cho TP. Hồ Chí Minh, tổng lượng phát thải bụi $\text{PM}_{2.5}$ là 4.029 nghìn tấn/năm [27]. Trong đó, nguồn đường (giao thông) đóng góp bụi $\text{PM}_{2.5}$ cao nhất tương ứng với 1.813 tấn/năm chiếm 45% (giao thông đường bộ là nguồn phát thải chủ yếu chiếm 75,13% tổng lượng thải $\text{PM}_{2.5}$ của nguồn đường). Nguồn điểm xếp thứ hai với 1.289 tấn/năm, chiếm 32% (dệt may, thực phẩm là hai ngành đóng góp nhiều). Ngoài ra, nguồn diện đóng góp bụi $\text{PM}_{2.5}$ là thấp nhất (927 tấn/năm) chiếm 23% với hộ gia đình, nhà hàng, quán ăn là nguồn chính [28].

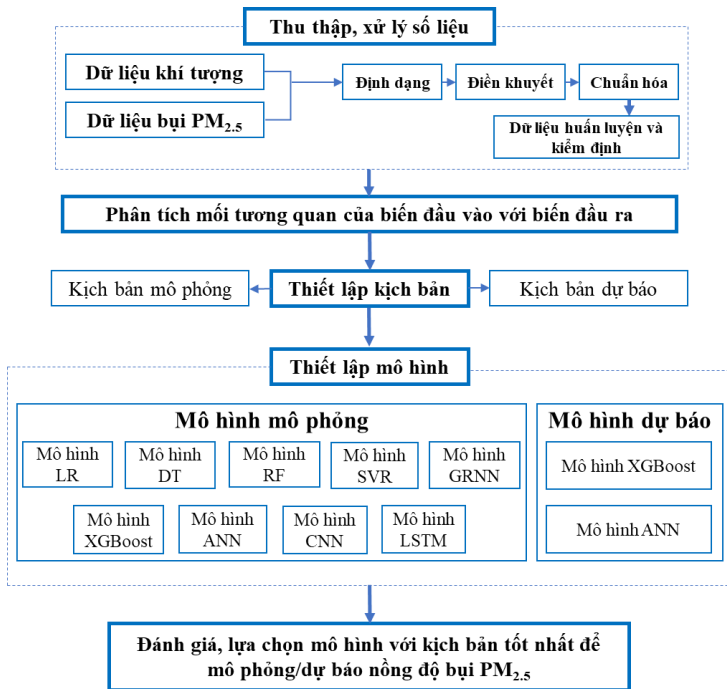
¹ <https://vpub.hochiminhcity.gov.vn/portal/KenhTin/Gioi-thieu-ve-thanh-pho.aspx>.

CHƯƠNG 3. PHƯƠNG PHÁP NGHIÊN CỨU

3.1. KHUNG NGHIÊN CỨU

Quá trình xây dựng mô hình mô phỏng và dự báo nồng độ bụi $PM_{2.5}$ của đề tài bao gồm năm (05) bước chính: (1) Thu thập và xử lý dữ liệu, (2) Phân tích mối tương quan của các biến đầu vào với biến đầu ra ($PM_{2.5}$), (3) Xây dựng các kịch bản cho bài toán mô phỏng và dự báo nồng độ bụi $PM_{2.5}$, (4) Xây dựng mô hình thuật toán học máy và học sâu để mô phỏng/dự báo $PM_{2.5}$, và (5) Kiểm định và lựa chọn mô hình với kịch bản tốt nhất để mô phỏng/dự báo nồng độ $PM_{2.5}$ từ những mô hình học máy/học sâu và các kịch bản đã được phát triển.

Hình 3.1 mô tả chi tiết các bước để xây dựng mô hình dự đoán nồng độ $PM_{2.5}$.



Hình 3.1: Sơ đồ tiến trình nghiên cứu

3.2. ĐÁNH GIÁ HIỆU QUẢ MÔ HÌNH

Bước tiếp theo là đánh giá, kiểm định tính chính xác và hiệu quả của mô hình trong mô phỏng và dự báo nồng độ bụi PM_{2.5}. Các công cụ thống kê phổ biến được sử dụng để đánh giá hiệu suất của mô hình bao gồm sai số bình phương trung bình của căn (RMSE - Root Mean Square Error), sai số tương phản tuyệt đối (MAPE - Mean Absolute Percentage Error), chỉ số thống nhất (IOA – Index of Agreement) và độ lệch trung bình chuẩn hóa (NMB – Normalized Mean Bias).

CHƯƠNG 4. KẾT QUẢ VÀ THẢO LUẬN

4.1. ĐÁNH GIÁ NỒNG ĐỘ BỤI $PM_{2.5}$ GIAI ĐOẠN 2021 – 2023 VÀ PHÂN TÍCH ẢNH HƯỞNG CỦA YẾU TỐ KHÍ TƯỢNG

4.1.1. Đánh giá dữ liệu khí tượng và nồng độ bụi $PM_{2.5}$

Dữ liệu về khí tượng và nồng độ bụi $PM_{2.5}$ ở TP.HCM từ ngày 1/1/2021 đến 30/6/2023 được thể hiện trong Hình 4.1.

Bảng 4.1: Tóm tắt các giá trị thấp, cao và trung bình của dữ liệu

Thông số	Giá trị thấp nhất	Giá trị cao nhất	Trung bình
Nhiệt độ (°C)	24,0	32,2	28,5
Tốc độ gió (m/s)	0,0	9,0	2,3
Độ ẩm (%)	56,0	93,0	75,3
Số giờ nắng (h)	0,0	9,9	5,9
Lượng mưa (mm)	0,0	101,5	5,7
Độ bốc hơi (mm/day)	0,8	6,3	3,4
$PM_{2.5}$, ($\mu g/m^3$)	6,5	90,3	22,4

4.1.2. Kịch bản mô phỏng

Nhằm xây dựng các kịch bản mô phỏng phù hợp và tối ưu, nghiên cứu đã sử dụng tương quan Pearson (r) xác định mức độ tương quan giữa các biến đầu vào với nồng độ $PM_{2.5}$, tìm ra mối liên hệ giữa các thông số với nhau từ đó đề xuất các kịch bản dựa trên kết quả phân tích mối tương quan.

Bảng 4.2: Hệ số tương quan r giữa các yếu tố khí tượng và bụi $PM_{2.5}$

Độ ẩm	Nhiệt độ	Tốc độ gió	Mưa	Độ bốc hơi	Số giờ nắng
- 0,293	- 0,280	- 0,227	- 0,111	0,107	- 0,037

Bảng 4.3: Các kịch bản mô phỏng

Kịch bản mô phỏng	Biến đầu vào	Biến đầu ra
Kịch bản 1	Độ ẩm	$PM_{2.5}$
Kịch bản 2	Độ ẩm, Nhiệt độ	$PM_{2.5}$
Kịch bản 3	Độ ẩm, Nhiệt độ, Tốc độ gió	$PM_{2.5}$
Kịch bản 4	Độ ẩm, Nhiệt độ, Tốc độ gió, Lượng mưa	$PM_{2.5}$
Kịch bản 5	Độ ẩm, Nhiệt độ, Tốc độ gió, Lượng mưa, Độ bốc hơi	$PM_{2.5}$
Kịch bản 6	Độ ẩm, Nhiệt độ, Tốc độ gió, Lượng mưa, Độ bốc hơi, Số giờ nắng	$PM_{2.5}$

4.1.3. Kịch bản dự báo

Tương tự, nghiên cứu đã phân tích mối tương quan giữa các biến khí tượng đầu vào và $PM_{2.5}$ ở các thời điểm 1 ngày, 3 ngày, 5 ngày, và 7 ngày, từ đó đề xuất các kịch bản dựa trên kết quả phân tích mối tương quan.

Bảng 4.4: Hệ số tương quan r giữa các biến đầu vào với bụi $PM_{2.5}$

Dự báo $PM_{2.5}$	Độ ẩm	Nhiệt độ	Độ bốc hơi	Tốc độ gió	Mưa	Số giờ nắng
1 ngày	-0,364	-0,297	0,182	-0,197	-0,109	-0,006
3 ngày	-0,350	-0,260	0,220	-0,121	-0,131	0,053
5 ngày	-0,313	-0,244	0,199	-0,086	-0,144	0,083
7 ngày	-0,293	-0,284	0,165	-0,103	-0,080	-0,023

Bảng 4.5: Các kịch bản dự báo 1 ngày

Kịch bản dự báo	Biến đầu vào	Biến đầu ra
Kịch bản 1	Độ ẩm	$PM_{2.5}$
Kịch bản 2	Độ ẩm, Nhiệt độ	$PM_{2.5}$
Kịch bản 3	Độ ẩm, Nhiệt độ, Tốc độ gió	$PM_{2.5}$
Kịch bản 4	Độ ẩm, Nhiệt độ, Tốc độ gió, Độ bốc hơi	$PM_{2.5}$
Kịch bản 5	Độ ẩm, Nhiệt độ, Tốc độ gió, Độ bốc hơi, Lượng mưa	$PM_{2.5}$
Kịch bản 6	Độ ẩm, Nhiệt độ, Tốc độ gió, Độ bốc hơi, Lượng mưa, Số giờ nắng (Nắng)	$PM_{2.5}$

Bảng 4.6: Các kịch bản dự báo 3 ngày

Kịch bản dự báo	Biến đầu vào	Biến đầu ra
Kịch bản 1	Độ ẩm	$PM_{2.5}$
Kịch bản 2	Độ ẩm, Nhiệt độ	$PM_{2.5}$
Kịch bản 3	Độ ẩm, Nhiệt độ, Độ bốc hơi	$PM_{2.5}$
Kịch bản 4	Độ ẩm, Nhiệt độ, Độ bốc hơi, Lượng mưa	$PM_{2.5}$
Kịch bản 5	Độ ẩm, Nhiệt độ, Độ bốc hơi, Lượng mưa, Tốc độ gió,	$PM_{2.5}$
Kịch bản 6	Độ ẩm, Nhiệt độ, Độ bốc hơi, Lượng mưa, Tốc độ gió, Nắng	$PM_{2.5}$

Bảng 4.7: Các kịch bản dự báo 5 và 7 ngày

Kịch bản dự báo	Biến đầu vào	Biến đầu ra
Kịch bản 1	Độ ẩm	$PM_{2.5}$
Kịch bản 2	Độ ẩm, Nhiệt độ	$PM_{2.5}$
Kịch bản 3	Độ ẩm, Nhiệt độ, Độ bốc hơi	$PM_{2.5}$
Kịch bản 4	Độ ẩm, Nhiệt độ, Độ bốc hơi, Tốc độ gió	$PM_{2.5}$
Kịch bản 5	Độ ẩm, Nhiệt độ, Độ bốc hơi, Tốc độ gió, Lượng mưa	$PM_{2.5}$
Kịch bản 6	Độ ẩm, Nhiệt độ, Độ bốc hơi, Tốc độ gió, Lượng mưa, Nắng	$PM_{2.5}$

4.2. MÔ PHỎNG NỒNG ĐỘ BỤI PM_{2.5} BẰNG CÁC THUẬT TOÁN HỌC MÁY VÀ HỌC SÂU

4.2.1. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 1

Kết quả cho thấy các mô hình đều có hiệu suất thấp, được đánh giá trên một loạt các chỉ số đánh giá hiệu suất như RMSE, MAPE, IOA và NMB.

Bảng 4.8: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 1

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	9,130	36,230	0,345	0,000	10,012	43,152	0,384	0,073
DT	9,098	35,793	0,366	0,000	9,998	43,054	0,40	0,077
RF	9,100	35,852	0,361	0,000	9,986	42,880	0,396	0,075
ANN	9,398	32,372	0,357	-0,086	10,088	38,670	0,328	-0,011
CNN	9,198	38,084	0,396	0,039	10,176	46,157	0,437	0,119
SVR	9,134	36,027	0,322	-0,004	10,027	43,033	0,361	0,068
GRNN	9,089	35,886	0,344	-0,001	10,053	43,199	0,372	0,073
XGBoost	9,093	35,897	0,302	0,000	10,167	43,741	0,320	0,073
LSTM	9,341	32,504	0,347	-0,076	9,341	38,371	0,320	-0,013

4.2.2. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 2

Bảng 4.9: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 2

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	8,536	33,625	0,541	0,000	9,505	41,704	0,541	0,090
DT	8,249	32,117	0,619	0,000	9,400	40,843	0,589	0,085
RF	8,104	31,642	0,627	0,000	9,244	39,944	0,596	0,083
ANN	8,275	31,564	0,614	-0,017	9,348	39,361	0,589	0,070
CNN	8,389	33,914	0,573	0,039	9,455	42,403	0,567	0,106
SVR	8,467	33,935	0,54	0,005	9,451	41,674	0,543	0,093
GRNN	8,097	31,997	0,601	-0,006	9,337	39,941	0,545	0,076
XGBoost	7,599	29,984	0,647	-0,001	9,454	40,089	0,517	0,075
LSTM	8,260	29,991	0,613	-0,044	9,388	38,301	0,575	0,041

Kết quả cho thấy các mô hình có sự cải thiện so với kịch bản 1, tuy nhiên hiệu suất dự báo vẫn chưa cao thông qua các chỉ số đánh giá hiệu suất như RMSE, MAPE, IOA và NMB.

4.2.3. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 3

Kết quả mô phỏng cho kịch bản 3 được thể hiện tại Bảng 4.10, cho thấy sự cải thiện đáng kể so với hai kịch bản trước đó. Hiệu suất của các mô

hình được cải thiện khi đánh giá qua các chỉ số như RMSE, MAPE, IOA và NMB, nhưng vẫn còn hạn chế cần khắc phục.

Bảng 4.10: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 3

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	8,375	32,879	0,581	0,000	9,079	39,378	0,599	0,082
DT	8,192	31,940	0,625	0,000	9,118	38,952	0,625	0,080
RF	7,732	30,139	0,653	0,000	9,005	38,666	0,599	0,078
ANN	7,948	29,219	0,680	-0,035	8,891	34,948	0,642	0,042
CNN	8,376	32,519	0,596	-0,013	9,303	39,820	0,589	0,075
SVR	8,285	32,762	0,585	0,001	8,930	38,450	0,609	0,080
GRNN	7,882	30,716	0,631	-0,008	9,031	37,955	0,582	0,006
XGBoost	7,608	29,757	0,684	-0,001	9,120	38,811	0,600	0,082
LSTM	8,283	30,627	0,642	-0,035	9,119	36,911	0,620	0,042

4.2.4. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 4

Trong các mô hình, ANN thể hiện hiệu suất tốt nhất với các chỉ số kiểm định: RMSE = 8,740 μ g/m³, MAPE = 33,449%, IOA = 0,672 và NMB = 0,034.

Bảng 4.11: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 4

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	8,370	32,923	0,582	0,000	9,061	39,266	0,602	0,082
DT	8,045	31,181	0,651	0,000	9,090	38,390	0,617	0,075
RF	7,738	30,106	0,662	0,000	8,845	37,840	0,628	0,078
ANN	8,120	29,116	0,676	-0,047	8,740	33,449	0,672	0,034
CNN	8,356	33,549	0,561	0,010	9,152	40,394	0,579	0,092
SVR	8,306	33,042	0,571	0,005	9,020	39,210	0,593	0,085
GRNN	7,854	30,603	0,635	-0,007	9,002	38,016	0,585	0,068
XGBoost	7,830	30,800	0,625	-0,001	8,962	38,410	0,585	0,076
LSTM	8,548	30,962	0,612	-0,037	9,108	35,747	0,603	0,027

4.2.5. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 5

Thuật toán SVR và LSTM có hiệu suất tốt nhất với các chỉ số kiểm định của SVR là RMSE = 8,391 μ g/m³, MAPE = 34,055%, IOA = 0,709 và NMB = 0,041, trong khi LSTM có RMSE = 8,514 μ g/m³, MAPE = 34,315%, IOA = 0,705 và NMB = 0,053. Các kết quả này cho thấy SVR và LSTM phản ánh khá tốt mối quan hệ giữa các yếu tố đầu vào và nồng độ PM_{2.5}.

Bảng 4.12: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 5

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	8,325	32,635	0,591	0,000	8,990	38,805	0,613	0,079
DT	8,073	31,046	0,645	0,000	9,184	38,223	0,607	0,065
RF	7,307	28,282	0,709	0,000	8,631	37,148	0,654	0,076
ANN	7,761	28,685	0,595	-0,014	8,390	35,026	0,697	0,060
CNN	8,453	33,670	0,515	0,004	9,199	40,123	0,540	0,080
SVR	7,665	26,917	0,720	-0,026	8,391	34,055	0,709	0,041
GRNN	7,225	27,579	0,781	-0,008	8,584	37,050	0,652	0,073
XGBoost	7,134	27,654	0,740	-0,001	8,416	36,195	0,687	0,072
LSTM	8,217	30,270	0,670	-0,019	8,514	34,315	0,705	0,053

4.2.6. Mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 6

Mô hình ANN có hiệu suất cao nhất với các chỉ số: RMSE = 7,978 $\mu\text{g}/\text{m}^3$, MAPE = 32,452%, IOA = 0,736 và NMB = 0,032. Là mô hình có sai số kiểm định thấp nhất và đạt chỉ số IOA cao nhất, cho thấy ANN có khả năng dự đoán chính xác và phản ánh tốt mối quan hệ giữa các yếu tố khí tượng và nồng độ PM_{2.5}.

Bảng 4.13: Kết quả mô phỏng nồng độ bụi PM_{2.5} cho kịch bản 6

Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
	RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
LR	8,309	32,411	0,595	0,000	8,944	38,571	0,618	0,078
DT	8,026	31,008	0,655	0,000	8,952	38,016	0,649	0,074
RF	6,464	24,577	0,789	0,001	8,510	36,721	0,670	0,079
ANN	7,675	26,919	0,713	-0,036	7,978	32,452	0,736	0,032
CNN	8,345	34,000	0,581	0,022	9,083	40,819	0,607	0,104
SVR	8,214	32,539	0,580	0,003	8,856	38,458	0,607	0,079
GRNN	6,605	24,772	0,785	-0,009	8,306	36,339	0,695	0,068
XGBoost	7,021	27,082	0,747	0,000	8,397	36,374	0,685	0,076
LSTM	7,456	25,063	0,734	-0,055	8,270	32,085	0,713	0,023

4.2.7. Đánh giá chung

Thông qua việc áp dụng 9 thuật toán và 6 kịch bản trong mô phỏng nồng độ bụi PM_{2.5} tại TP.HCM, ta có thể nhận thấy rằng:

- Các thuật toán ANN, LSTM, và SVR đã đạt hiệu suất cao nhất trong việc dự đoán nồng độ bụi PM_{2.5}, với khả năng dự báo khá chính xác và ổn định so với các phương pháp khác.

- Các kịch bản phù hợp nhất là kịch bản 5 và 6 như được minh họa trong Bảng 4.3. Trong đó, thuật toán SVR mô phỏng tốt nhất nồng độ bụi PM_{2.5} dựa trên kịch bản 5, với các chỉ số đánh giá mô phỏng lần lượt là: RMSE = 8,391 $\mu\text{g}/\text{m}^3$, MAPE = 34,055%, IOA = 0,709, và NMB = 0,041. Những kết quả này cho thấy thuật toán SVR có khả năng phản ánh khá chính xác sự biến động của nồng độ bụi PM_{2.5} tại TP.HCM trong kịch bản 5.
- Đối với thuật toán ANN, mô phỏng tốt nhất nồng độ bụi PM_{2.5} trên kịch bản 6, với các chỉ số RMSE = 7,978 $\mu\text{g}/\text{m}^3$, MAPE = 32,452%, IOA = 0,736, và NMB = 0,032. Đây là những chỉ số rất khả quan, cho thấy ANN có thể cung cấp những dự báo chính xác và đáng tin cậy cho nồng độ bụi PM_{2.5} trong điều kiện mô phỏng với kịch bản 6.

4.3. DỰ BÁO SỚM NỒNG ĐỘ BỤI BẰNG CÁC THUẬT TOÁN HỌC MÁY VÀ HỌC SÂU

Trong nghiên cứu này, các mô hình học máy và học sâu bao gồm ANN, SVR, và XGboost được áp dụng để dự báo sớm nồng độ bụi PM_{2.5} 1 ngày, 3 ngày, 5 ngày, và 7 ngày.

4.3.1. Dự báo nồng độ bụi PM_{2.5} 1 ngày

Kết quả dự báo 1 ngày cho thấy mô hình ANN trong kịch bản 3 và 4 cho kết quả dự báo 1 ngày tốt nhất, với RMSE, MAPE, IOA, và NMB lần lượt là 7,091 $\mu\text{g}/\text{m}^3$; 28,643 %; 0,723; và -0,025 cho kịch bản 3; và 6,856 $\mu\text{g}/\text{m}^3$; 29,025 %; 0,765; và 0,019 cho kịch bản 4.

Bảng 4.14: Kết quả dự báo 1 ngày nồng độ bụi PM_{2.5} tại TP.HCM

KB	Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
		RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
1	SVR	9,200	36,552	0,422	0,005	8,458	38,537	0,480	0,076
	XGboost	9,086	35,448	0,482	-0,003	8,455	37,859	0,510	0,072
	ANN	9,619	31,722	0,413	-0,110	8,448	32,314	0,436	-0,043
2	SVR	8,523	34,232	0,601	0,003	7,821	36,122	0,650	0,089
	XGboost	8,323	32,384	0,638	-0,001	7,401	34,024	0,710	0,083
	ANN	8,590	29,612	0,650	-0,086	7,280	30,043	0,712	-0,015
3	SVR	8,426	34,143	0,621	0,008	7,608	34,773	0,676	0,088
	XGboost	7,850	31,398	0,669	-0,001	7,468	33,986	0,654	0,073

	ANN	8,078	27,573	0,706	-0,086	7,091	28,643	0,723	-0,025
4	SVR	8,416	34,022	0,621	0,007	7,587	34,614	0,677	0,085
	XGboost	7,606	30,309	0,708	-0,001	7,109	32,576	0,717	0,075
	ANN	8,257	29,065	0,692	-0,053	6,856	29,025	0,765	0,019
5	SVR	8,416	34,059	0,620	0,007	7,586	34,642	0,676	0,086
	XGboost	7,937	30,566	0,666	0,000	7,099	32,675	0,723	0,075
	ANN	8,211	31,021	0,682	-0,015	7,030	30,602	0,751	0,055
6	SVR	8,300	33,427	0,632	0,006	7,426	34,195	0,689	0,084
	XGboost	7,525	29,330	0,708	0,001	7,127	32,778	0,715	0,076
	ANN	8,468	29,686	0,635	-0,080	7,166	29,962	0,711	-0,005

4.3.2. Dự báo nồng độ bụi PM_{2.5} 3 ngày

Đối với dự báo 3 ngày, ANN trong kịch bản 5 đã đạt kết quả dự báo tốt nhất trên tập kiểm định với RMSE, MAPE, IOA và NMB lần lượt là 8,335 $\mu\text{g}/\text{m}^3$; 34,295 %; 0,654; và -0,005.

Bảng 4.15: Kết quả dự báo 3 ngày nồng độ bụi PM_{2.5} tại TP.HCM

KB	Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
		RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
1	SVR	9,073	34,826	0,420	0,003	9,235	41,435	0,444	0,041
	XGboost	9,001	34,174	0,458	0,001	9,258	41,115	0,493	0,046
	ANN	9,137	32,363	0,478	-0,045	9,262	38,658	0,493	-0,003
2	SVR	8,477	32,963	0,580	0,009	8,700	39,742	0,591	0,058
	XGboost	8,266	31,686	0,612	0,001	8,637	38,723	0,603	0,050
	ANN	8,605	30,368	0,563	-0,058	8,625	35,943	0,577	-0,011
3	SVR	8,493	33,000	0,575	0,007	8,687	39,645	0,588	0,056
	XGboost	8,254	31,624	0,613	0,001	8,609	38,603	0,606	0,049
	ANN	8,564	29,745	0,598	-0,061	8,515	34,595	0,617	-0,017
4	SVR	8,495	33,094	0,571	0,008	8,689	39,683	0,583	0,057
	XGboost	8,045	30,395	0,639	-	8,403	37,736	0,630	0,046
	ANN	8,493	29,787	0,585	-0,046	8,663	35,728	0,587	-0,005
5	SVR	8,494	32,991	0,569	0,007	8,648	39,176	0,586	0,053
	XGboost	7,984	30,230	0,640	-	8,374	37,391	0,630	0,045
	ANN	8,514	30,087	0,619	-0,053	8,335	34,295	0,654	-0,005
6	SVR	8,531	32,971	0,569	0,002	8,715	39,429	0,580	0,051
	XGboost	7,543	28,497	0,693	-	8,443	37,864	0,633	0,051
	ANN	7,965	27,322	0,699	-0,056	8,706	35,091	0,658	0,007

4.3.3. Dự báo nồng độ bụi PM_{2.5} 5 ngày

Mô hình ANN được xây dựng dựa trên kịch bản 6 cho kết quả dự báo tốt nhất cho nồng độ bụi trong 5 ngày tiếp theo.

Bảng 4.16: Kết quả dự báo 5 ngày nồng độ bụi PM_{2.5} tại TP.HCM

KB	Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
		RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
1	SVR	9,319	35,628	0,363	-0,001	9,077	40,817	0,423	0,020
	XGboost	9,241	35,291	0,416	0,000	9,097	41,279	0,464	0,028
	ANN	9,354	33,622	0,405	-0,039	9,113	39,078	0,461	-0,013
2	SVR	8,517	29,861	0,640	-0,043	8,892	37,276	0,629	-0,001

3	XGboost	8,199	31,414	0,630	0,000	8,687	38,377	0,594	0,029
	ANN	8,747	30,772	0,577	-0,056	8,561	35,972	0,615	-0,013
	SVR	8,856	34,133	0,496	-0,001	8,605	39,375	0,545	0,030
	XGboost	8,091	31,129	0,638	0,000	8,561	38,364	0,604	0,032
	ANN	8,511	30,365	0,618	-0,039	8,547	36,314	0,639	0,004
4	SVR	8,838	33,760	0,496	-0,006	8,608	39,143	0,538	0,023
	XGboost	8,078	31,104	0,638	0,000	8,570	38,507	0,601	0,032
	ANN	8,772	31,803	0,564	-0,031	8,569	37,037	0,603	0,001
5	SVR	8,851	33,858	0,497	-0,007	8,582	38,792	0,545	0,022
	XGboost	8,017	30,747	0,646	0,000	8,540	38,138	0,608	0,031
	ANN	8,786	31,651	0,566	-0,035	8,555	37,021	0,609	0,002
6	SVR	8,795	33,714	0,495	-0,004	8,678	39,318	0,527	0,025
	XGboost	7,973	30,595	0,649	0,000	8,528	38,278	0,611	0,035
	ANN	8,466	29,569	0,626	-0,053	8,653	35,344	0,628	-0,018

4.3.4. Dự báo nồng độ bụi PM_{2.5} 7 ngày

Bảng 4.17: Kết quả dự báo 7 ngày nồng độ bụi PM_{2.5} tại TP.HCM

KB	Mô hình	Kết quả huấn luyện				Kết quả kiểm định			
		RMSE	MAPE	IOA	NMB	RMSE	MAPE	IOA	NMB
1	SVR	9,493	37,591	0,325	0,009	8,645	38,421	0,386	0,027
	XGboost	9,372	36,476	0,391	-0,001	8,609	37,481	0,431	0,018
	ANN	9,772	32,322	0,426	-0,106	8,818	33,703	0,471	-0,085
2	SVR	9,005	36,155	0,489	0,017	8,466	38,850	0,508	0,049
	XGboost	8,441	2,912	0,601	-0,001	8,442	36,686	0,553	0,025
	ANN	8,883	33,634	0,578	-0,009	8,099	35,299	0,637	0,026
3	SVR	9,005	36,155	0,489	0,017	8,466	38,850	0,508	0,049
	XGboost	8,092	31,129	0,638	-	8,561	38,364	0,604	0,032
	ANN	9,229	32,530	0,504	-0,067	8,486	34,491	0,534	-0,033
4	SVR	8,848	35,563	0,017	0,525	8,161	37,400	0,557	0,042
	XGboost	8,146	31,990	0,627	-	8,222	35,788	0,556	0,019
	ANN	9,218	31,494	0,500	-0,094	8,351	33,380	0,533	-0,063
5	SVR	8,831	35,389	0,524	0,013	8,202	37,400	0,551	0,038
	XGboost	8,340	32,621	0,607	-0,001	8,424	36,417	0,549	0,024
	ANN	8,892	32,222	0,556	-0,047	8,087	33,998	0,581	-0,021
6	SVR	8,866	35,373	0,539	0,018	8,231	37,655	0,564	0,047
	XGboost	7,718	30,105	0,676	-0,001	8,145	35,452	0,568	0,019
	ANN	8,768	32,200	0,557	-0,034	8,215	35,191	0,550	-0,010

ANN với kịch bản 6 được đánh giá là mô hình tốt nhất về mặt tổng thể do có RMSE thấp, MAPE hợp lý, IOA cao, và NMB gần bằng 0. Điều này cho thấy ANN có khả năng khai thác tốt các đặc điểm phức tạp trong dữ liệu khi có nhiều biến đầu vào, giúp dự báo chính xác hơn nồng độ PM_{2.5}.

4.3.5. Đánh giá chung

- Mô hình ANN cho thấy hiệu suất dự báo ổn định và chính xác hơn so với các thuật toán khác, đặc biệt khi sử dụng các kịch bản có nhiều biến đầu vào. Các mô hình này có khả năng khai thác hiệu quả các đặc

điểm phi tuyến tính và phức tạp của dữ liệu, giúp nâng cao độ chính xác trong dự báo nồng độ $PM_{2.5}$.

- Mô hình SVR cũng đạt được hiệu quả tốt trong một số kịch bản nhất định, đặc biệt khi số lượng biến đầu vào hạn chế. Tuy nhiên, khi số lượng biến tăng lên, hiệu suất của SVR có xu hướng giảm so với các mô hình phi tuyến như ANN và XGBoost.
- Kết quả cho thấy kịch bản tốt nhất phụ thuộc vào khoảng thời gian dự báo. Khi thời gian dự báo kéo dài, cần sử dụng nhiều biến đầu vào hơn để đảm bảo cung cấp đủ thông tin, giúp mô hình dự báo chính xác hơn.

4.4. THẢO LUẬN VÀ ỨNG DỤNG KẾT QUẢ

4.4.1. Phân tích và so sánh kết quả mô phỏng nồng độ bụi $PM_{2.5}$ giữa thuật toán LR, DT, RF, XGBoost và SVR

Dựa trên các kết quả mô phỏng, có thể thấy rõ sự khác biệt về hiệu suất giữa các thuật toán LR, DT, RF, XGBoost, và SVR trong dự báo nồng độ bụi $PM_{2.5}$. SVR là mô hình tốt nhất trong số các mô hình học máy khác như DT, RF và XGBoost, nhờ vào khả năng dự báo chính xác cao và độ ổn định trong các kịch bản khác nhau. XGBoost, mặc dù có lợi thế trong các bài toán dự báo phức tạp, lại không thể hiện hiệu suất vượt trội hơn SVR trong bài toán dự báo nồng độ $PM_{2.5}$ trên tập dữ liệu hiện tại. RF có độ ổn định nhưng chưa đủ mạnh để cạnh tranh với SVR về độ chính xác, đặc biệt khi số lượng biến đầu vào tăng. DT, dù dễ triển khai và hiểu rõ, lại không đạt hiệu suất dự báo cao do thiếu khả năng học các mối quan hệ phi tuyến.

4.4.2. Phân tích và so sánh kết quả mô phỏng nồng độ bụi $PM_{2.5}$ giữa thuật toán ANN, CNN, GRNN và LSTM

Kết luận, LSTM và ANN là hai mô hình mạnh nhất trong nhóm deep learning cho dự báo nồng độ $PM_{2.5}$, đặc biệt là trong các kịch bản phức tạp với nhiều biến đầu vào. GRNN có thể là một lựa chọn tốt cho các kịch bản đơn giản hơn với ít biến đầu vào, do có tốc độ học nhanh và độ chính xác

tương đối cao. Tuy nhiên, nó gặp khó khăn trong các kịch bản phức tạp. CNN, không đạt hiệu quả cao trong dự báo $PM_{2.5}$.

4.4.3. Phân tích và so sánh kết quả giữa nghiên cứu mô phỏng và dự báo nồng độ bụi $PM_{2.5}$

Phân tích cho thấy:

- Về độ chính xác, các mô hình mô phỏng thường đạt độ chính xác cao hơn dự báo, do có dữ liệu lịch sử đầy đủ để học và tái hiện lại các xu hướng. Ngược lại, dự báo yêu cầu mô hình phải dự đoán trước các biến động mà không có thông tin đầy đủ, dẫn đến RMSE và MAPE cao hơn trong hầu hết các mô hình.
- Về tính ổn định, trong khi mô phỏng cho thấy tính ổn định cao hơn giữa các bộ dữ liệu huấn luyện và kiểm định, dự báo có sự chênh lệch lớn hơn giữa các kết quả kiểm định và dự đoán thực tế do tính bất định của các yếu tố khí tượng trong tương lai.
- Về khả năng ứng dụng, mô phỏng hữu ích cho việc hiểu rõ hơn về mối quan hệ giữa các yếu tố khí tượng và $PM_{2.5}$, trong khi dự báo lại có ý nghĩa thực tiễn hơn trong việc cung cấp thông tin kịp thời để đưa ra biện pháp kiểm soát ô nhiễm. Các mô hình dự báo cần được điều chỉnh liên tục dựa trên dữ liệu mới để đảm bảo độ chính xác và tính hiệu quả trong cảnh báo ô nhiễm không khí.
- Kết quả từ cả hai loại nghiên cứu cho thấy rằng việc kết hợp mô phỏng và dự báo có thể cung cấp một cái nhìn toàn diện và chính xác hơn về sự biến động của nồng độ $PM_{2.5}$, đồng thời giúp đưa ra các biện pháp kiểm soát ô nhiễm một cách kịp thời và hiệu quả hơn.

4.4.4. Ứng dụng kết quả trong quản lý chất lượng không khí

- Các dự báo sớm này cho phép chính quyền cảnh báo người dân về thời điểm nồng độ bụi cao, khuyến nghị các biện pháp bảo vệ như đeo khẩu trang, hạn chế ra ngoài, và sử dụng các thiết bị lọc không khí. Nhờ có

dự báo chính xác, các kế hoạch ứng phó có thể được chuẩn bị tốt, như tăng cường dự trữ khẩu trang và thiết bị lọc không khí nhằm đáp ứng nhu cầu bảo vệ sức khỏe của người dân trong những giai đoạn ô nhiễm cao. Các mô hình dự báo có thể được tích hợp vào hệ thống giám sát chất lượng không khí theo thời gian thực, giúp cung cấp cảnh báo tự động khi nồng độ $PM_{2.5}$ vượt ngưỡng an toàn. Điều này không chỉ hỗ trợ các cơ quan chức năng trong việc giám sát liên tục mà còn giúp nâng cao nhận thức của cộng đồng về tình trạng ô nhiễm và cách bảo vệ sức khỏe bản thân.

- Nghiên cứu giúp xây dựng các mô hình mô phỏng và dự báo nồng độ $PM_{2.5}$ dựa trên số lượng đầu vào các yếu tố khí tượng tối thiểu nhưng đạt được kết quả tối ưu nhất. Điều này giúp phát huy lợi ích, đặc biệt trong khía cạnh tiết kiệm chi phí, đơn giản hóa các yếu tố quan trắc trong các điều kiện nhất định, nhưng vẫn đảm bảo kết quả dự báo chính xác ở mức tương đối.
- Nghiên cứu đóng góp vào việc phát triển kiến thức và kỹ năng về học máy và học sâu trong lĩnh vực môi trường. Thành công của nghiên cứu có thể khuyến khích các nhà nghiên cứu và chuyên gia môi trường sử dụng các phương pháp dự báo tiên tiến khác để cải thiện độ chính xác và hiệu quả của các công cụ giám sát ô nhiễm.
- Khẳng định tầm quan trọng của công tác quan trắc môi trường trong dự báo. Công tác quan trắc đóng vai trò thiết yếu trong việc xây dựng các hệ thống cảnh báo sớm. Việc đầu tư mở rộng mạng lưới quan trắc tự động và sử dụng dữ liệu thời gian thực sẽ tăng cường khả năng dự báo và quản lý chất lượng không khí một cách hiệu quả hơn.

4.5. XÂY DỰNG ỨNG DỤNG DỰ ĐOÁN NỒNG ĐỘ BỤI $PM_{2.5}$

Ứng dụng dự báo nồng độ bụi $PM_{2.5}$ tại TP.HCM được phát triển từ kết quả nghiên cứu của luận án. Ứng dụng cung cấp khả năng dự báo trong

nhiều khoảng thời gian khác nhau, bao gồm mô phỏng hiện tại và dự báo sớm 1, 3, 5, và 7 ngày, đáp ứng nhu cầu đa dạng của người sử dụng. Ứng dụng có thể được truy cập thông qua đường dẫn sau: <https://hcm-pm25-prediction.streamlit.app/>

Dự báo nồng độ bụi PM_{2.5} tại thành phố Hồ Chí Minh

PM 2.5

Nồng độ PM_{2.5} được dự báo dựa trên các thông số khí tượng

Loại thời tiết hiện tại (đơn vị: km/h): Mây mỏng

Vận tốc gió: 0.00 10.00

Lượng mưa (mm): 0.00 10.00

Ánh sáng mặt trời (%): 0.00 10.00

Độ ẩm tương đối (%): 0.00 10.00

Số giờ nắng (giờ): 0.00 10.00

Dự đoán

Chú ý:

1. Ứng dụng là một phần của luận án của MSc Nguyễn Phước Hữu với đề tài: "Nghiên cứu ứng dụng phương pháp học máy để dự báo nồng độ bụi PM_{2.5}". Trường hợp nghiên cứu của nhóm phát hiện lỗi sai sót.

2. Nguyên nhân của lỗi sai sót: Nguyên nhân của lỗi sai sót có thể do lỗi của dữ liệu đầu vào hoặc do lỗi của mô hình dự báo.

3. Phương pháp nghiên cứu: Phương pháp nghiên cứu được sử dụng là phương pháp học máy để dự báo nồng độ bụi PM_{2.5}.

4. Thông tin liên hệ: MSc Nguyễn Phước Hữu: nguyendep@hcm.edu.vn

Hình 4.1: Giao diện ứng dụng dự đoán nồng độ bụi PM_{2.5}

CHƯƠNG 5. KẾT LUẬN VÀ KIẾN NGHỊ

5.1. KẾT LUẬN

Luận án đã đạt được:

1. Nghiên cứu đã lựa chọn, so sánh và đánh giá kết quả mô phỏng và dự báo nồng độ bụi PM_{2.5} tại TP. Hồ Chí Minh bằng các mô hình học máy và học sâu, bao gồm LR, DT, RF, SVR, XGBoost, GRNN, ANN, CNN, và LSTM. Các kịch bản mô phỏng và dự báo được xây dựng dựa trên mối tương quan giữa các yếu tố đầu vào và đầu ra, trong đó dữ liệu đầu vào của mô hình bao gồm các thông số khí tượng gồm nhiệt độ, tốc độ gió, độ ẩm, số giờ nắng, lượng mưa, và độ bốc hơi; dữ liệu đầu ra là nồng độ bụi PM_{2.5}.
2. Kết quả mô phỏng:

Thông qua 9 thuật toán và 7 kịch bản, nhận thấy rằng:

- Các thuật toán ANN, LSTM, và SVR đã đạt hiệu suất cao nhất trong việc dự đoán nồng độ bụi $PM_{2.5}$, với khả năng dự đoán khá chính xác và ổn định so với các phương pháp khác.
- Các kịch bản phù hợp nhất là kịch bản 5 và 6. Trong đó, thuật toán SVR mô phỏng tốt nhất nồng độ bụi $PM_{2.5}$ dựa trên kịch bản 5, kết quả này cho thấy thuật toán SVR có khả năng phản ánh khá chính xác sự biến động của nồng độ bụi $PM_{2.5}$ tại TP.HCM trong kịch bản 5.
- Đối với thuật toán ANN, mô phỏng tốt nhất nồng độ bụi $PM_{2.5}$ dựa trên kịch bản 6, với các chỉ số $RMSE = 7,978 \mu g/m^3$, $MAPE = 32,452\%$, $IOA = 0,736$, và $NMB = 0,032$. Đây là những chỉ số rất khả quan, cho thấy ANN có thể cung cấp những dự báo chính xác và đáng tin cậy cho nồng độ bụi $PM_{2.5}$ trong điều kiện mô phỏng, kịch bản 6.

3. Kết quả dự báo:

- Mô hình ANN cho thấy hiệu suất dự báo ổn định và chính xác hơn so với các thuật toán khác, đặc biệt khi sử dụng các kịch bản có nhiều biến đầu vào. Các mô hình này có khả năng khai thác hiệu quả các đặc điểm phi tuyến tính và phức tạp của dữ liệu, giúp nâng cao độ chính xác trong dự báo nồng độ $PM_{2.5}$.
- Mô hình SVR cũng đạt được hiệu quả tốt trong một số kịch bản nhất định, đặc biệt khi số lượng biến đầu vào hạn chế. Tuy nhiên, khi số lượng biến tăng lên, hiệu suất của SVR có xu hướng giảm so với các mô hình phi tuyến như ANN và XGBoost.
- Kết quả cho thấy kịch bản tốt nhất phụ thuộc vào khoảng thời gian dự báo. Khi thời gian dự báo kéo dài, cần sử dụng nhiều biến đầu vào hơn để đảm bảo cung cấp đủ thông tin, giúp mô hình dự báo chính xác hơn.

5.2. ĐỀ XUẤT HƯỚNG NGHIÊN CỨU TIẾP THEO

- Mở rộng mạng lưới trạm quan trắc: Để cải thiện độ chính xác của dự báo, cần bổ sung thêm các trạm quan trắc tại nhiều vị trí khác nhau trong TP.HCM và các khu vực lân cận.
- Nâng cao khả năng áp dụng mô hình cho các khu vực khác: Nghiên cứu nên tiếp tục mở rộng để tinh chỉnh mô hình cho các khu vực khác ngoài TP.HCM.
- Mở rộng và cải thiện mô hình cho dự báo dài hạn: Để cải thiện khả năng dự báo trong các khoảng thời gian dài hơn, cần tích hợp thêm các yếu tố bổ sung như dữ liệu giao thông, mật độ dân cư, các nguồn xả thải công nghiệp, và các yếu tố mùa vụ.

DANH MỤC CÔNG TRÌNH

1. Hiếu, N.P.; Vân, Đ.D.H.; Khôi, Đ.N. Xây dựng mô hình mô phỏng chỉ số chất lượng không khí cho thành phố Hồ Chí Minh bằng phương pháp học máy. Tạp chí Khí tượng Thủy văn 2024, 761, 13-24.
2. Hiếu, N.P.; Dương, N.N.; Linh, Đ.Q.; Khôi, Đ.N. Mô phỏng nồng độ bụi $PM_{2.5}$ tại khu vực trung tâm Thành phố Hồ Chí Minh bằng thuật toán học máy và học sâu. Tạp chí Khí tượng Thủy văn 2024, 762, 62-72.
3. Nguyen, P.H.; Dao, N.K.; Nguyen, L.S.P. Development of Machine Learning and Deep Learning Prediction Models for $PM_{2.5}$ in Ho Chi Minh City, Vietnam. Atmosphere 2024, 15, 1163. <https://doi.org/10.3390/atmos151011>