

NGHIÊN CỨU HỆ THỐNG DỰ BÁO CHẤT LƯỢNG KHÔNG KHÍ

AIR QUALITY FORECAST SYSTEM

SVTH: Nguyễn Chí Thành, Nguyễn Tuấn Vương, Nguyễn Thanh Hà

Lớp 18CE, Khoa Kỹ thuật máy tính và điện tử, Trường Đại học Công nghệ Thông tin và Truyền thông Việt - Hàn;

Email: ncthanh.18ce@vku.udn.vn, ntvuong.18ce@vku.udn.vn, nthao.18ce@vku.udn.vn

GVHD: ThS. Nguyễn Anh Tuấn

Khoa Kỹ thuật máy tính và điện tử, Trường Đại học Công nghệ Thông Tin và Truyền thông Việt - Hàn; Email: natua@vku.udn.vn

Tóm tắt - Hiện nay, vấn đề ô nhiễm môi trường không khí, đặc biệt tại các đô thị không vấn đề riêng lẻ của một quốc gia hay một khu vực mà nó đã trở thành vấn đề toàn cầu. Thực trạng phát triển kinh tế - xã hội của các quốc gia trên thế giới trong thời gian qua đã có những tác động lớn đến môi trường, và đã làm cho môi trường sống của con người bị thay đổi và ngày càng trở nên tồi tệ hơn. Những năm gần đây nhân loại đã phải quan tâm nhiều đến vấn đề ô nhiễm môi trường không khí đó là: sự biến đổi của khí hậu – nóng lên toàn cầu, sự suy giảm tầng ôzôn và mưa axit.

Ở Việt Nam ô nhiễm môi trường không khí đang là một vấn đề bức xúc đối với môi trường đô thị, công nghiệp và các làng nghề. Ô nhiễm môi trường không khí không chỉ tác động xấu đối với sức khỏe con người (đặc biệt là gây ra các bệnh đường hô hấp) mà còn ảnh hưởng đến các hệ sinh thái và biến đổi khí hậu như: hiệu ứng nhà kính, mưa axit và suy giảm tầng ôzôn,... Công nghiệp hóa càng mạnh, đô thị hóa càng phát triển thì nguồn thải gây ô nhiễm môi trường không khí càng nhiều, áp lực làm biến đổi chất lượng không khí theo chiều hướng xấu càng lớn. Ở Việt Nam, tại các khu công nghiệp, các trục đường giao thông lớn đều bị ô nhiễm với các cấp độ khác nhau, nồng độ các chất ô nhiễm đều vượt quá tiêu chuẩn cho phép. Và sự gia tăng dân số, gia tăng đột biến của các phương tiện giao tiếp trong khi cơ sở hạ tầng còn thấp làm cho tình hình ô nhiễm trở nên trầm trọng.

Xuất phát từ vấn đề trên nhóm chúng tôi đã đi vào nghiên cứu hệ thống dự báo chất lượng không khí nhằm đưa ra giải pháp kịp thời, giảm thiểu tối đa ô nhiễm không khí.

Từ khóa – Lora; gateway; deep learning; machine learning; chất lượng không khí.

1. Đặt vấn đề

1.1. Trên thế giới

Theo thống kê 12/1952, trên thế giới có 1600 người chết do ô nhiễm không khí và khí hậu. Năm 1930 tại Bỉ, đã phát hiện 63 người chết và 600 người mắc bệnh do sự ô nhiễm không khí quá nặng. Trong một thông báo mới đưa ra ngày 25/03/2014, tổ chức y tế thế giới WHO cho biết: năm 2012 có khoảng 7 triệu người tử vong liên quan đến ô nhiễm không khí. Cứ 8 ca tử vong trên toàn cầu thì có một ca tử vong do ô nhiễm không khí. Số liệu phân tích thống kê này cho thấy con số tử vong do ô nhiễm không khí tăng hơn gấp 2 lần so với ước tính trước đây và xác nhận rằng ô nhiễm không khí là nguy cơ môi trường đơn lẻ lớn nhất thế giới hiện nay đối với sức khỏe. Giảm ô nhiễm môi trường có thể cứu số được hàng triệu người.

1.1.1. Ô nhiễm ở Trung Quốc

Mỗi năm, nền kinh tế Trung Quốc thiệt hại 38 tỷ USD vì ô nhiễm không khí, South China Morning Post dẫn báo

Abstract – Currently, the problem of air pollution, especially in urban areas is becoming a global problem. The current situation of socio-economic development of the countries in the world in recent years has made the living environment of people changed and become worse day by day. Some of the most notable air pollution problems are global warming, ozone depletion and acid rain.

In Vietnam, the air pollution is a pressing problem for urban, industrial and craft villages. This not only has a negative impact on human health, especially causing respiratory diseases, but also affects the ecosystems. The more urbanization develops, the more waste sources cause air pollution. The greater the pressure that changes the air quality in a bad way. Besides, the sudden increase of means of transport while the infrastructure is still poor makes the pollution situation more serious.

Stemming from the above problem, our team has researched the air quality forecasting system in order to provide timely warnings and minimize air pollution.

Key words - Lora; gateway; deep learning; machine learning; Air quality.

cáo mới nhất từ Đại học Trung văn Hong Kong (CUHK). Các nhà nghiên cứu đưa ra kết luận này dựa trên số tiền mà chính quyền Trung Quốc phải chi trả để khắc phục hậu quả của ô nhiễm không khí đến sức khỏe người dân và đến hoạt động sản xuất nông nghiệp.

"Con số này khá lớn, chiếm đến 0,7% GDP quốc gia", Phó giáo sư Steve Yim Hung-lam thuộc trường CUHK cho biết. Nhóm nghiên cứu do ông Yim dẫn dắt phân tích số liệu từ các lĩnh vực như công nghiệp, nông nghiệp, thương mại, năng lượng, giao thông vận chuyển,...

1.1.2. Ở Việt Nam

Theo thống kê của Tổ chức Y tế Thế giới WHO mỗi năm có khoảng gần 1,5 triệu người tử vong do viêm phổi, chiếm 21% các ca tử vong do ô nhiễm không khí. Khoảng 1,4 triệu người tử vong do đột quỵ, chiếm 20% các ca tử vong do ô nhiễm không khí. Gần 2,5 triệu người tử vong do bệnh tim thiếu máu cục bộ, chiếm khoảng 34% tổng số ca tử vong do ô nhiễm không khí. Hơn 1,3 triệu người tử

vong do bệnh phổi tắc nghẽn mạn tính, chiếm khoảng 19% tổng số ca tử vong do ô nhiễm không khí. Khoảng gần 500.000 người tử vong do ung thư phổi, chiếm 7% tổng số ca tử vong do ô nhiễm không khí.

1.2. Nguyên nhân gây ô nhiễm không khí

Trong thực tế có rất nhiều nguyên nhân dẫn đến ô nhiễm không khí. Nhưng chúng ta có thể chia thành 2 nguồn: Tự nhiên và nhân tạo.

1.2.1. Nguồn tự nhiên

- + Núi lửa: núi lửa phun trào những nham thạch nóng và nhiều khói bụi giàu sulfua, metan và nhiều loại khí khác. Không khí chứa bụi được lan đi rất xa.
- + Cháy rừng: các đám cháy rừng trong tự nhiên thải ra rất nhiều khói bụi vào không khí
- + Các quá trình phân hủy xác động vật chết và thực vật trong tự nhiên cũng thải ra rất nhiều khí sulfua nitri ...

1.2.2. Nguồn nhân tạo

Hoạt động giao thông vận tải: sự phát sinh các chất gây ô nhiễm từ loại nguồn này chủ yếu từ quá trình đốt nhiên liệu động cơ đốt trong. Các chất gây ô nhiễm chủ yếu sinh ra từ loại nguồn này là: bụi, cát, đất đá trên đường, Pb (chất rắn); CH_4 , CO, CO_2 , NO_x , SO_2 (chất khí). Đây là loại nguồn gây ô nhiễm không khí lớn, đặc biệt tại các khu đô thị đông dân cư, mật độ dân số lớn. Nếu chúng ta chỉ tính toán trên từng phương tiện giao thông thì nồng độ các chất gây ô nhiễm tương đối nhỏ, nhưng nếu mật độ giao thông lớn và công tác quy hoạch địa hình, đường xá không tốt thì sẽ gây ra sự ô nhiễm nặng.

Hoạt động sản xuất công nghiệp: sự phát sinh các chất gây ô nhiễm từ loại nguồn này chủ yếu từ quá trình đốt các loại nhiên liệu hóa thạch phổ biến: dầu mỏ, khí đốt, than đá. Thành phần các chất gây ô nhiễm từ loại nguồn này rất đa dạng và đặc trưng theo ngành nghề. Các chất gây ô nhiễm chủ yếu sinh ra từ loại nguồn này là: bụi, muội than (chất rắn); CO, CO_2 , NO_x , SO_2 (chất khí). Đây là loại nguồn gây ô nhiễm không khí rất nghiêm trọng, đặc biệt là tại các trung tâm công nghiệp.

Hoạt động phục vụ cho nhu cầu sinh hoạt: sự phát sinh các chất gây ô nhiễm từ loại nguồn này chủ yếu từ hoạt động đun nấu, sưởi ấm, sử dụng nhiên liệu hóa thạch phục vụ cho mục đích sinh hoạt. Các chất gây ô nhiễm chủ yếu sinh ra từ loại nguồn này là: bụi, muội than (chất rắn); CO, CO_2 (chất khí). Đặc điểm của loại nguồn này là gây ô nhiễm không khí ở mức độ tương đối nhỏ, tuy nhiên nếu chúng ta xem xét trong phạm vi hộ gia đình hay vài hộ xung quanh thì mức độ ô nhiễm không khí cục bộ là khá lớn.

1.3. Các chất gây ô nhiễm không khí

1.3.1. Bụi mịn

Bụi siêu mịn hay bụi PM1.0 là những hạt dạng lỏng, hoặc rắn trôi nổi ngoài không khí. Các nhà khoa học thường sử dụng chỉ số PM10, PM2.5, PM1.0 để thể hiện hàm lượng tiêu chuẩn của các hạt dạng lỏng, rắn trôi nổi trong 1 m³ không khí. Và bụi siêu mịn PM 1.0 hiểu như sau.

Chữ PM là viết tắt của chữ tiếng Anh - Particulate Matter, có nghĩa là chất dạng hạt (rắn hoặc lỏng).

Kí hiệu 1.0 tức là kích thước các hạt có đường kính nhỏ hơn hoặc bằng 1 micromet (micromet viết tắt là μm , bằng 1 phần triệu mét).

Kí hiệu PM1.0, PM2.5 là chỉ những hạt có đường kính nhỏ hơn hoặc bằng 10 μm , 2,5 μm .

Bụi càng nhỏ hay càng mịn càng dễ đi sâu vào hệ hô hấp và có thể ảnh hưởng đến cấu trúc DNA, bởi sự mất cân bằng oxy khiến các tế bào khỏe mạnh bị hủy hoại và ảnh hưởng đến sự chuyển hóa các chất hữu cơ của DNA. Các kim loại chuyển tiếp trong thành phần bụi như Cr, Cd, Ni và chất aldehyde có thể cản trở cơ chế sửa lỗi của DNA, gây ung thư phổi.

Loại bụi mịn PM 1.0 (có kích thước 1 μm), thậm chí cả bụi nano ($\leq 0,1\mu\text{m}$) có thể vượt qua tất cả các hàng rào ngăn bụi của hệ hô hấp, bít các lỗ trao đổi oxy ở phế nang, tác động đến cấu trúc DNA.

1.3.2. Carbon monoxide (CO)

+ Tính chất vật lý:

Carbon monoxide (CO) là một chất khí không màu, không mùi, không vị đặc trưng, có khối lượng riêng nhẹ hơn không khí. CO có khối lượng phân tử là 28,010 g/mol, $d = 0,967$, nhiệt độ hóa lỏng ở -191°C . Tại nhiệt độ 0°C , 1 lít CO nặng 1,254 g. CO có đặc tính là ít tan trong nước: ở trạng thái 0°C và 1 atm, 100 ml nước sẽ hòa tan được 3,54 ml CO; ở trạng thái 25°C và 1 atm, 100 ml nước sẽ hòa tan được 2,14 ml CO. CO có một đặc điểm nổi bật là không bị than hoạt tính hấp phụ.

+ Tính chất hóa học

Khí cháy, ngọn lửa của CO có màu xanh và sản phẩm tạo thành là CO_2 . CO trở về mặt hóa học khi ở điều kiện bình thường về nhiệt độ và áp suất, nhưng khi đưa nhiệt độ lên cao thì CO lại trở thành một chất khử mạnh. Chính vì tính chất này mà người ta thường dùng CO làm chất khử trong công nghệ và phân tích. Ngoài ra, người ta còn nhận thấy quá trình oxi hóa CO thành CO_2 sẽ tăng tốc nếu sử dụng các chất xúc tác như Pd trên gel silic hay hỗn hợp oxide Mn và Cu.

+ Nguồn phát thải CO:

CO được tạo ra từ các nguồn phát thải chủ yếu sau:

- + Quá trình đốt nhiên liệu của các loại động cơ đốt trong.
- + Khí thải tạo ra từ nhà máy nhiệt điện.
- + Các loại chất hữu cơ bị đốt cháy không hoàn toàn.
- + Sử dụng than cốc trong công nghiệp sản xuất gang – thép.
- + Sản xuất khí đốt từ than đá.
- + Sản xuất đất đèn (hay còn gọi là gió đá) làm nguyên liệu tạo ra acetylene (C_2H_2).
- + Tác hại của khí CO:
- + Nhiễm độc cấp tính:

Trường hợp siêu cấp tính: nạn nhân bị nhiễm độc siêu cấp tính CO có thể bị hôn mê và ngất xỉu ngay tại chỗ dẫn đến tử vong. Nếu được cấp cứu kịp thời, nạn nhân có thể sống được, nhưng vẫn sẽ xuất hiện một số triệu chứng như co giật cơ, nhức đầu, chóng mặt.

Trường hợp cấp tính thể nặng: nạn nhân bị nhiễm độc

cấp tính thể nặng CO có thể bị suy giảm khả năng phán đoán, rối loạn hệ vận động và hệ hô hấp, mất phản xạ co giật, hôn mê, liệt hô hấp dẫn đến tử vong. Nếu được cấp cứu kịp thời, nạn nhân có thể sống được, nhưng vẫn sẽ xuất hiện một số triệu chứng như nhức đầu, rối loạn tâm thần, khó phát âm, nói ngọng, rối loạn hệ thị giác.

Trường hợp cấp tính thể nhẹ: nạn nhân bị nhiễm độc cấp tính thể nhẹ CO có thể bị nhức đầu, buồn nôn, mệt mỏi, rối loạn hệ thị giác, nếu không tiếp xúc với khí CO nữa thì sẽ hết các triệu chứng này.

Trường hợp cấp tính thời gian kéo dài: nạn nhân bị nhiễm độc cấp tính CO trong khoảng thời gian dài sẽ bị các di chứng như tổn thương tim (rối loạn điện tim, nhồi máu cơ tim), tổn thương hệ thần kinh (giảm trí nhớ), tổn thương da (ban, trứng cá, hoại thư).

+ Nhiễm độc mãn tính:

Nạn nhân bị nhiễm độc mãn tính CO có thể gặp phải các triệu chứng sau: nhức đầu, chóng mặt, mệt mỏi, suy nhược, khó thở, dễ cáu gắt, cảm giác buồn ngủ khi tập trung, rối loạn hệ thị giác và hệ tiêu hóa.

1.3.3. Ozone (O_3)

+ Tính chất vật lý:

Ozone, hay còn gọi là trioxigen (O_3), là một chất khí tạo thành lớp lá chắn bảo vệ trái đất khỏi tác động của tia cực tím từ mặt trời. O_3 có khối lượng phân tử là 48,0 g/mol, nặng hơn không khí, không có màu hay ở điều kiện tiêu chuẩn có màu xanh nhạt, mùi hăng mạnh. O_3 có điểm sôi là $-112^\circ C$, điểm chảy là $-193^\circ C$ [3]. Ở $0^\circ C$, O_3 tan trong nước theo tỉ lệ 49 ml/100 ml nước.

+ Tính chất hóa học:

O_3 có thể tạo ra các peroxide gây nổ với các alkene. Trong trường hợp bị làm nóng, O_3 có thể gây cháy hay nổ mạnh. O_3 cũng có thể bị phân hủy thành O_2 khi gặp tác nhân nhiệt độ. O_3 là một chất oxi hóa – khử mạnh. O_3 có thể phản ứng với các alkene, các aromatic, ether và cao su. O_3 có thể ăn mòn các chất kim loại, trừ vàng và bạch kim.

+ Nguồn phát thải:

O_3 được tạo ra từ các nguồn phát thải chủ yếu sau:

- + Quá trình khử trùng nước uống, khử mùi thực phẩm trong ngành công nghiệp thực phẩm.
- + Quá trình tẩy trắng hàng dệt may, bột giấy, sáp, tinh bột.
- + Quá trình lọc dầu khoáng và dẫn xuất của dầu.
- + Quá trình chế tạo nước hoa, long não.
- + Quá trình sản xuất có hồ quang điện, hàn điện.
- + Các loại đèn tia X, đèn tia cực tím, đèn hơi thủy ngân, máy photocopy cũng tạo ra O_3 .
- + Tác hại:
- + Nhiễm độc cấp tính:

Nồng độ 0,05-0,1ppm O_3 : nạn nhân có thể bị kích thích mũi, miệng, khó thở.

Nồng độ 0,3-1ppm O_3 : nạn nhân có thể bị giảm thị lực, nhức đầu, ho, khó thở, co thắt ngực, rối loạn hệ hô hấp nhẹ.

Nồng độ 1,5-2ppm O_3 : nạn nhân có thể bị suy giảm khả năng bão hòa oxi, tổn thương hình thái hồng cầu, rối loạn

hệ thần kinh, rối loạn khả năng phối hợp, khó khăn khi đi lại.

Nồng độ 4-5ppm O_3 : nạn nhân có thể bị phù phổi, tổn thương phế quản.

Nồng độ 50ppm O_3 : nạn nhân có thể bị ngưng thở cấp tính dẫn đến tử vong trong vài phút.

+ Nhiễm độc mãn tính:

Nạn nhân nếu tiếp xúc dài hạn hay liên tục với không khí có nồng độ O_3 khoảng 1 ppm có thể bị nhức đầu, mệt mỏi, khó hô hấp và rối loạn chức năng hô hấp.

1.3.4. Sulfur dioxide (SO_2)

+ Tính chất vật lý của SO_2 :

Sulfur dioxide (SO_2) là một chất khí không màu, có mùi hơi hắc và cay khi nồng độ ở 1 ppm, gây phản xạ ngạt thở và có vị chua như acid. SO_2 có khối lượng phân tử là 64,066 g/mol, nặng hơn không khí, $d = 2,279$. SO_2 dễ hóa lỏng dưới áp suất cao hay bị làm lạnh ở $-15^\circ C$, 1 lít SO_2 lỏng sẽ cho 500 lít SO_2 khí. Khí SO_2 khi bốc hơi sẽ thu nhiều nhiệt, người ta áp dụng tính chất này của SO_2 để làm các máy lạnh.

+ Tính chất hóa học của SO_2 :

SO_2 phản ứng với nước sẽ trở thành acid H_2SO_3 , sau đó do sự có mặt của O_2 hòa tan trong nước, acid H_2SO_3 sẽ bị oxi hóa từ từ thành H_2SO_4 . SO_2 có tính khử tương đối mạnh, chính vì tính chất này mà người ta thường dùng SO_2 làm chất tẩy màu và chất khử trong hóa phân tích.

+ Nguồn phát thải SO_2 :

SO_2 được tạo ra từ các nguồn phát thải chủ yếu sau:

- + Quá trình sản xuất H_2SO_4 từ quặng pirit sắt.
- + Quá trình sản xuất hóa chất có lưu huỳnh.
- + Quá trình đốt nhiên liệu chứa lưu huỳnh trong các lò hơi đốt than và dầu, trong ngành công nghiệp nhiệt điện.
- + Quá trình chiết xuất các sản phẩm từ dầu mỏ.
- + Quá trình tẩy trắng trong ngành công nghiệp giấy, đường, sợi, da.
- + Các thiết bị làm lạnh, kho lạnh trong công nghiệp cũng là một nguồn phát sinh SO_2 đáng kể.
- + Tác hại của SO_2 :
- + Nhiễm độc cấp tính:

Nạn nhân bị nhiễm độc cấp tính SO_2 ở nồng độ cao có thể bị kích thích mắt dữ dội, kích thích niêm mạc đường hô hấp trên, gây nên tình trạng khó thở, tím tái, rối loạn tri giác, tuần hoàn phổi ngừng đột ngột, có thể dẫn đến tử vong.

+ Nhiễm độc mãn tính:

Nạn nhân bị nhiễm độc mãn tính SO_2 có thể bị các triệu chứng sau: kích ứng cục bộ niêm mạc miệng, cảm giác nóng rát, mũi họng sưng đau, tăng tiết dịch, ho, đau ngực, khó thở, cay mắt, buồn nôn; niêm mạc mũi, thanh quản bị phù nề, xung huyết, tổn thương răng và da; ở người nữ có thể làm rối loạn tuyến giáp và kinh nguyệt.

Trong trường hợp tiếp xúc với SO_2 nồng độ cao trong một khoảng thời gian dài, nạn nhân có thể bị xơ cứng phổi, khí thũng, tổn thương hệ hô hấp.

2. Nghiên cứu và khảo sát

2.1. Chỉ số chất lượng không khí(AQI)

Chỉ số chất lượng không khí Việt Nam (viết tắt là VN_AQI) là chỉ số được tính toán từ các thông số quan trắc các chất ô nhiễm trong không khí ở Việt Nam được quy định tại quyết định số 1459/QĐ TCMT, nhằm cho biết tình trạng chất lượng không khí và mức độ ảnh hưởng đến sức khỏe con người, được biểu diễn qua một thang điểm.

Chỉ số chất lượng không khí được tính theo thang điểm (khoảng giá trị AQI) tương ứng với biểu tượng và các màu sắc để cảnh báo chất lượng không khí và mức độ ảnh hưởng tới sức khỏe con người, cụ thể như sau:

Khoảng giá trị AQI	Chất lượng không khí	Màu sắc	Mã màu RGB
0-50	Tốt	Xanh	0;228;0
51-100	Trung bình	Vàng	255;255;0
101 -150	Kém	Da cam	255;126;0
151-200	Xấu	Đỏ	255;0;0
201 – 300	Rất xấu	Tím	143;63;151
301 - 500	Nguy hại	Nâu	126;0;35

Bảng 2.1: Khoảng giá trị AQI và đánh giá chất lượng không khí

2.1.1. Yêu cầu đối với việc tính toán VN_AQI

VN_AQI được tính toán bao gồm AQI giờ và AQI ngày. Số liệu sử dụng để tính toán VN_AQI là giá trị quan trắc trung bình 1 giờ, trung bình 8 giờ và trung bình 24 giờ.

VN_AQI được tính toán cho dữ liệu của từng trạm quan trắc không khí tự động liên tục đối với môi trường không khí xung quanh.

Đối với mỗi trạm quan trắc, AQIx được tính toán cho từng thông số quan trắc, giá trị AQI cuối cùng là giá trị lớn nhất trong các giá trị AQIx của mỗi thông số.

Các thông số được sử dụng để tính VN_AQI bao gồm: SO₂, CO, NO₂, O₃, PM10, và PM2.5.

Phương pháp tính toán VN_AQI yêu cầu bắt buộc phải có tối thiểu 01 trong 02 thông số PM10, PM2.5 trong công thức tính.

2.1.2. Tính toán giá trị VN_AQI

- Tính toán giá trị AQI giờ (AQIh)

Số liệu để tính toán AQI giờ là giá trị quan trắc trung bình 1 giờ.

- + Tính giá trị Nowcast đối với thông số PM2.5 và PM10

Gọi c₁, c₂, ...c₁₂ là 12 giá trị quan trắc trung bình 1 giờ (với c₁ là giá trị quan trắc trung bình 1 giờ hiện tại, c₁₂ là giá trị quan trắc trung

bình 1 giờ cách 12 giờ so với hiện tại).

Tính giá trị trọng số: $W^* = \frac{C_{min}}{C_{max}}$

Trong đó: C_{min} là giá trị nhỏ nhất trong số 12 giá trị trung bình 1 giờ.

C_{max} là giá trị lớn nhất trong số 12 giá trị trung bình 1 giờ

Nếu: $W^* \leq \frac{1}{2}$ thì $W = \frac{1}{2}$

Nếu: $W^* > \frac{1}{2}$ thì $W = W^*$

Trong trường hợp $W > \frac{1}{2}$ thì lấy giá trị $nowcast = \frac{\sum_{i=1}^{12} w^{i-1} c_i}{\sum_{i=1}^{12} w^{i-1}}$

Trong trường hợp $W = \frac{1}{2}$ thì lấy giá trị:

$nowcast = \frac{1}{2} c_1 + \left(\frac{1}{2}\right)^2 c_2 + \dots + \left(\frac{1}{2}\right)^{12} c_{12}$

Chú ý:

- Nếu có ít nhất 2 trong 3 giá trị c₁ c₂ c₃ thì mới tính được giá trị nowcast, và nếu không có giá trị thì không tính được
- Nếu c_i không có giá trị thì lấy $w^{i-1} = 0$

Tính giá trị AQIh của từng thông số (AQIx)

Giá trị AQIh của các thông số SO₂, CO, NO₂, O₃ được tính toán theo công thức 1, giá trị AQIh của các thông số PM10, PM2.5 được tính toán theo công thức 2:

$AQI_x = \frac{I_{i+1} - I_i}{BP_{i+1} - BP_i} (C_x - BP_i) + I_i$ (Công thức 1)

$AQI_x = \frac{I_{i+1} - I_i}{BP_{i+1} - BP_i} (Nowcast_x - BP_i) + I_i$ (Công thức 2)

Trong đó:

AQIx: Giá trị AQI thông số của thông số x.

BP_i : Nồng độ giới hạn dưới của giá trị thông số quan trắc được quy định trong Bảng 2 tương ứng với mức i.

BP_{i+1}: Nồng độ giới hạn trên của giá trị thông số quan trắc được quy định trong Bảng 2 tương ứng với mức i+1.

I_i : Giá trị AQI ở mức i đã cho trong bảng tương ứng với giá trị BP_i.

I_{i+1}: Giá trị AQI ở mức i+1 cho trong bảng tương ứng với giá trị BP_{i+1}.

C_x: Giá trị quan trắc trung bình 1 giờ của thông số x.

Nowcast_x: Giá trị Nowcast được tính toán ở phần a

i	I _i	Giá trị BP _i quy định đối với từng thông số (Đơn vị: $\mu\text{g}/\text{m}^3$)					
		O 3 (1 h)	O3 (8h)	CO	SO 2	NO 2	PM1 0

1	0	0	0	0	0	0	0	0
2	5 0	16 0	100	10. 000	125	100	50	25
3	1 0 0	20 0	120	30. 000	350	200	150	50
4	1 5 0	30 0	170	45. 000	550	700	250	80
5	2 0 0	40 0	210	60. 000	800	1.2 00	350	150
6	3 0 0	80 0	400	90. 000	1.6 00	2.3 50	420	250
7	4 0 0	1. 00 0	-	120 .00 0	2.1 00	3.1 00	500	350
8	5 0 0	1. 20 0	≥ -	150 .00 0	≥ 2.6 30	≥ 3.8 50	≥6 00	≥5 00

Bảng 2.2: Các giá trị BPi đối với các thông số

+ Giá trị AQI giờ tổng hợp

Sau khi đã có giá trị AQIx của mỗi thông số, chọn giá trị AQI lớn nhất của các thông số để lấy làm giá trị AQI giờ tổng hợp. $AQI_h = \max(AQIx)$ Ghi chú: Giá trị AQI giờ được làm tròn thành số nguyên.

- Tính toán giá trị AQI ngày (AQId)

Giá trị AQI ngày được tính toán dựa trên các giá trị như sau:

- + Thông số PM2.5 và PM10: giá trị trung bình 24 giờ.
- + Thông số O₃: giá trị trung bình 1 giờ lớn nhất trong ngày và giá trị trung bình 8 giờ lớn nhất trong ngày.
- + Thông số SO₂, NO₂ và CO: giá trị trung bình 1 giờ lớn nhất trong ngày.

Xác định giá trị trung bình 1 giờ lớn nhất trong ngày, giá trị trung bình 8 giờ lớn nhất trong ngày và giá trị trung bình 24 giờ

- + Giá trị trung bình 1 giờ lớn nhất trong ngày là giá trị lớn nhất trong số các giá trị quan trắc trung bình 1 giờ.
- + Giá trị quan trắc trung bình 8 giờ lớn nhất trong ngày là giá trị lớn nhất trong số các giá trị trung bình 8 giờ. Giá trị trung bình 8 giờ là trung bình cộng các giá trị trung bình 1 giờ trong 8 giờ liên tiếp.

Tính giá trị AQId của từng thông số (AQIx)

Giá trị AQI ngày của các thông số SO₂, CO, NO₂, O₃, PM10, PM2.5 được tính toán theo công thức 1 như sau:

$$AQI_x = \frac{I_{i+1} - I_i}{BP_{i+1} - BP_i} (C_x - BP_i) + I_i \quad (\text{Công thức 1})$$

Trong đó:

Bảng giá trị BPi và Ii lấy trong bảng 2.

AQIx: Giá trị AQId thông số của thông số x.

BPi : Nồng độ giới hạn dưới của giá trị thông số quan trắc được

quy định trong Bảng 2 tương ứng với mức i.

BPi+1: Nồng độ giới hạn trên của giá trị thông số quan trắc được quy định trong Bảng 2 tương ứng với mức i+1.

Ii : Giá trị AQI ở mức i đã cho trong bảng tương ứng với giá trị BPi.

Ii+1: Giá trị AQI ở mức i+1 cho trong bảng tương ứng với giá trị BPi+1.

Cx: được quy định cụ thể như sau:

- Đối với thông số PM2.5 và PM10: Cx là giá trị trung bình 24 giờ.
- Đối với thông số O₃: Cx là giá trị lớn nhất trong giá trị trung bình 1 giờ lớn nhất trong ngày và giá trị trung bình 8 giờ lớn nhất trong ngày.

Giá trị AQI ngày tổng hợp

Sau khi đã có giá trị AQIx ngày của mỗi thông số, chọn giá trị AQI lớn nhất của các thông số để lấy làm giá trị AQI ngày tổng hợp.

$$AQId = \max(AQIx)$$

Ghi chú: Giá trị AQI ngày được làm tròn thành số nguyên

Kh oảng giá trị AQI	Chấ t lượng không khí	Ảnh hưởng tới sức khỏe con người
0-50	Tốt	Chất lượng không khí tốt, Không ảnh hưởng tới sức khỏe
51-100	Trung bình	Chất lượng không khí ở mức chấp nhận được. Tuy nhiên đối với những người nhạy cảm (Người già trẻ em người mắc bệnh hô hấp, tim mạch, ...) có thể chịu những tác động nhất định tới sức khỏe.
101 - 150	Kém	Những người nhạy cảm gặp phải các vấn đề về sức khỏe, những người bình thường ít ảnh hưởng
151- 200	Xấu	Những người bình thường bắt đầu có cả ảnh hưởng tới sức khỏe, nhóm người nhạy cảm có thể gặp những vấn đề nghiêm trọng hơn

201 – 300	Rất xấu	Cảnh báo tới sức khỏe: mọi người bị ảnh hưởng tới sức khỏe nghiêm trọng hơn
301 - 500	Nguy hại	Cảnh báo khẩn cấp về sức khỏe toàn bộ dân số bị ảnh hưởng tới sức khỏe nghiêm trọng

Bảng 2.3: Các mức VN_AQI tương ứng ảnh hưởng tới sức khỏe

2.2. Xu hướng ứng dụng IOT trong quan trắc chất lượng không khí hiện nay

Theo tài liệu “Phát triển và ứng dụng mạng vạn vật kết nối vào hệ thống quan trắc môi trường” của 2 tác giả Lê Hoàng Anh và Dương Hoàng Nam, trong quan trắc môi trường nói chung, các thiết bị kết nối mạng thường liên kết theo giao thức máy móc - máy móc (M2M). Các hệ thống quan trắc tự động đa phần có trang bị cảm biến nhằm đo đạc và thông báo một số thông số môi trường. Tuy nhiên, những cảm biến này thường chỉ cung cấp thông tin trực tiếp cho PLC (thiết bị điều khiển lập trình), hoặc bộ điều khiển nội bộ, do vậy, chúng hoạt động riêng lẻ và không kết nối trong hệ thống điều phối chung của doanh nghiệp (DN). M2M nếu được sử dụng trong những hệ thống này cũng thường liên quan tới hạ tầng kết nối riêng của hệ thống. Không như giao thức M2M hiện tại, IoT sẽ cung cấp giao tiếp dữ liệu ở 11 mức hệ thống thông qua Ethernet (một công nghệ mạng cục bộ - LAN) và các chuẩn của nó, kiến trúc mạng mở thay cho mạng đóng trong các giao thức M2M.

Và cũng theo 2 tác giả này, IoT gồm 3 loại hình kết nối: máy móc - máy móc (M2M), con người - máy móc (P2M) và con người - con người (P2P). Trong đó, kết nối M2M đóng vai trò quan trọng trong hoạt động của IoT. Các thiết bị, máy móc trong IoT sẽ “phản ứng” dựa vào các sự kiện diễn ra trong lúc chúng hoạt động theo thời gian thực. Giải pháp IoT cho phép thực hiện việc đo lường, thu thập và truyền nhận dữ liệu từ hệ thống các cảm biến/đầu đo về trung tâm tích hợp dữ liệu để phân tích, xử lý trên nền điện toán đám mây. Các ứng dụng IoT được phát triển trên nền điện toán đám mây cho phép phân tích xử lý và chuyển đổi khối lượng dữ liệu lớn từ vô số các cảm biến đo lường.

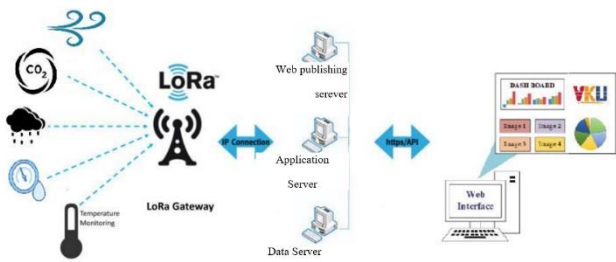
3. Phân tích và thiết kế hệ thống

3.1. Giải pháp

3.1.1 Giải Pháp IOT

Mô hình quan trắc môi trường truyền thống thường được dựa vào các trạm quan trắc cố định. Các trạm quan trắc này được xây dựng với nhiều thiết bị phân tích dữ liệu hiện đại, có độ chính xác cao và phải có người điều hành. Trạm quan trắc cố định có kích thước lớn (cỡ 1 căn nhà) và rất tốn chi phí do nó phải có khả năng giám sát môi trường xung quanh với phạm vi đủ rộng. Tuy nhiên, những đặc trưng cơ bản của trạm quan trắc cố định như kích thước lớn, nặng và đặc biệt là rất tốn kém, khiến nó không thể triển khai trong phạm vi thành phố, nơi mật độ dân cư thường rất đông và có nhiều vật cản, làm cho việc đo đạc không chính xác và khách quan. Trạm quan trắc truyền thống thường phải đặt ở các khu vực biệt lập, cách xa khu

dân cư.



Hình 3.1: Mô hình triển khai

Cách tiếp cận hiện nay cho các ứng dụng quan trắc hiện đại đa số được dựa trên nền tảng Internet of Things (Internet vạn vật). Cụ thể, các điểm quan trắc có kích thước nhỏ, mỗi điểm quan trắc được gắn các cảm biến cần thiết cho việc lấy thông tin về môi trường được sử dụng. Một số lượng lớn các điểm quan trắc này được phân bố rộng khắp môi trường cần giám sát, và gửi thông tin (bằng giao tiếp không dây) về 1 trạm chủ. Máy chủ sẽ chọn lọc, xử lý dữ liệu từ các điểm quan trắc để đưa ra kết luận về chất lượng môi trường. Với mô hình này, thông tin về môi trường có thể liên tục được cập nhật từng phút, hoặc thậm chí là từng giây. Yêu cầu này là không thể đối với các trạm quan trắc cố định truyền thống. Thêm nữa, các điểm quan trắc thường có kích thước nhỏ và có giá thành thấp, nên rất thuận tiện cho việc mở rộng ứng dụng. Người dùng có thể truy xuất được thông tin môi trường xung quanh mình bằng cách truy vấn dữ liệu từ các điểm cảm biến gần vị trí của mình nhất. Chính vì thế, các hệ thống quan trắc môi trường dựa trên IoT được xem là thế hệ kế tiếp trong quan trắc môi trường (viết tắt là TNGAPMS – The Next Generation Air Pollution Monitoring System).

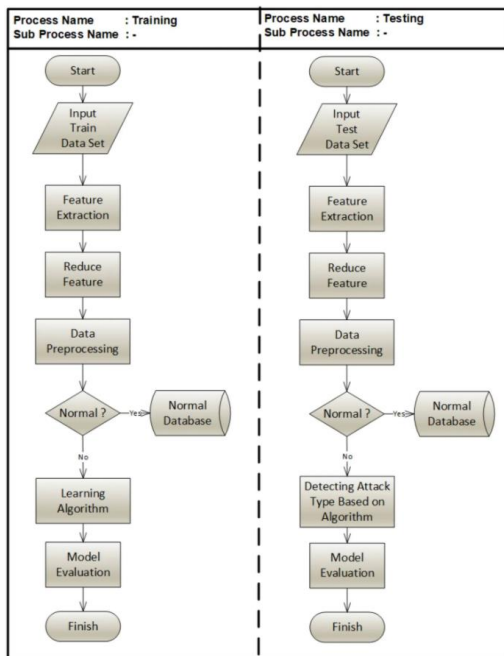
1.1.2 Giải Pháp Machine Learning

Học máy (machine learning, ML) là một tập con của trí tuệ nhân tạo. Machine learning là một lĩnh vực nhỏ trong khoa học máy tính, có khả năng tự học hỏi dựa trên dữ liệu được đưa vào mà không cần phải được lập trình cụ thể (Machine Learning is the subfield of computer science, that “gives computers the ability to learn without being explicitly programmed” – Wikipedia).

Những năm gần đây, sự phát triển của các hệ thống tính toán cùng lượng dữ liệu khổng lồ được thu thập bởi các hãng công nghệ lớn đã giúp machine learning tiến thêm một bước dài. Một lĩnh vực mới được ra đời được gọi là học sâu (deep learning, DL). Deep learning đã giúp máy tính thực thi những việc vào mười năm trước tưởng chừng là không thể: phân loại cả ngàn vật thể khác nhau trong các bức ảnh, tự tạo chú thích cho ảnh, bắt chước giọng nói và chữ viết, giao tiếp với con người, chuyển đổi ngôn ngữ, hay thậm chí cả sáng tác văn thơ và âm nhạc.

3.2. Thiết kế hệ thống

Thuật toán xử lý số liệu



Hình 3.2: Lưu đồ thuật toán

R-IOT EXPERIMENT

Hình 3.3: Bước 1 Chọn file data set

Chọn file csv nhập dữ liệu đầu vào bao gồm 3600 bộ dữ liệu cảm biến được thu thập từ 2017-2020, bao gồm các dữ liệu được thu thập từ các cảm biến, mức độ cảnh báo được gán nhãn bằng luật chuyên gia, dựa trên công thức quy ước và kinh nghiệm.

R-IOT EXPERIMENT

Hình 3.4: Bước 2 Chọn nhãn dữ liệu đầu vào

Tiếp theo ta quy ước cho bộ dữ liệu training, chọn các nhãn dữ liệu đầu vào (xs) ở đây là các cảm biến PM2.5, PM10, NO₂, CO, SO₂, O₃ của bộ dữ liệu data_training, sau đó quy ước nhãn đầu ra là AQI_Bucket trong file training.

Sau đó ta dùng thư viện tensor là đơn vị tính cơ bản trong deep learning phân chia nhóm dữ liệu cảm biến và kết quả vào 2 nhóm xs(các cảm biến) và ys(các mức độ cảnh báo).

Để cấu hình hàm thuật toán xử lý cho bộ dữ liệu chúng ta cần 2 functions để làm việc này, đó là loss và optimizer. loss: dùng để đo độ chính xác của phương trình dự đoán (ví dụ $y = 4x + 1$).

optimizer: đưa ra phương trình dự đoán, dựa trên giá trị

mà function loss đưa ra, function optimizer sẽ đưa ra câu trả lời sao cho giá trị của function loss nhỏ nhất.

Kèm theo đó, để đánh giá độ chính xác của model qua từng vòng lặp training sử dụng thêm hàm metrics để thể hiện ra độ chính xác tăng dần từ vòng lặp epoch đầu tiên đến vòng cuối cùng.

R-IOT EXPERIMENT

Hình 3.5 Bước 3 chỉnh các thông số training

Với các tham số xử lý training :

Epoch:200

Số lần học, training lặp lại của bộ dữ liệu, số lần càng nhiều độ chính xác càng cao, sau khi test trên số lần học, chúng tôi thấy 200 là số lần học cho kết quả chính xác và tốt nhất.

Learning rate:0,005

Tốc độ học của thuật toán, tốc độ càng lớn, yêu cầu, cấu hình máy càng mạnh và tốc độ training nhanh hơn.

Batch size:60

Chia bộ dữ liệu data thành 60 phần, để tránh epoch cao sẽ gây ra lỗi khi training, batch size càng cao training sẽ càng ổn định nhưng, tốc độ training sẽ lâu hơn.

R-IOT EXPERIMENT

Hình 3.6: Quá Trình training

Ta quy ước cho quá trình training gồm có 6 dữ liệu đầu vào và 6 mức dữ liệu kết quả cho ra, mỗi mức độ các giá trị cho ra sẽ nằm trong khoảng 0->1 tương ứng mức phần trăm từ 0 -> 100% dự báo.

Sau khi quá trình training hoàn thành ta tạo một hàm predictSample() gồm nhả dữ liệu đầu vào và model mô hình, để test đối với từng giá trị cụ thể và thu lại kết quả, có thể dùng để sử dụng kỹ thuật Cross-validation với bộ dữ liệu data_test và thu về kết quả độ chính xác của mô hình

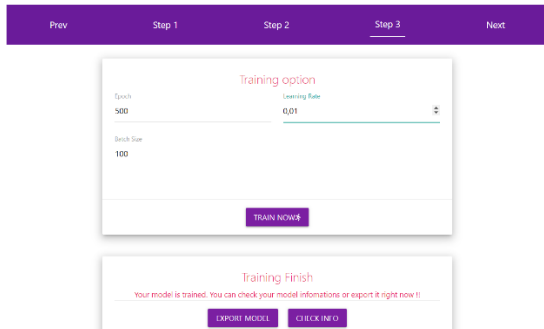
Sau khi training xong một model được tạo ra bao gồm

2 file.

Configuration được lưu trữ dưới dạng JSON file.

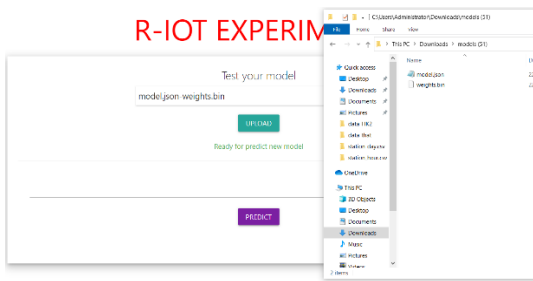
Trọng số (weights) được lưu dưới dạng binary file.

Ta trích xuất model để dùng đánh giá dự đoán



Hình 3.7: Trích xuất model

Ta tiến hành cấu hình dự báo cho model, bằng cách thêm model vừa được training xong là file bin và json và sau đó mô hình đã sẵn sàng để dự đoán mức độ ô nhiễm không khí



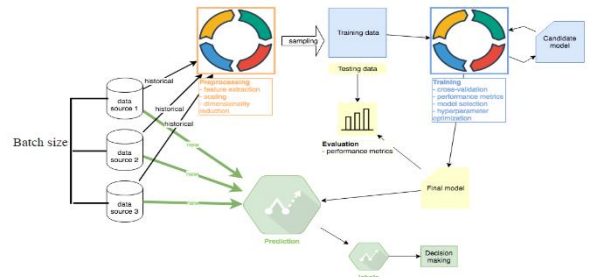
Hình 3.8: Nhập file model

Getway lora gửi dữ liệu đến server, máy chủ server nhận dữ liệu, phân tích giá trị và đưa đến hàm dự đoán

Server sau khi đã được cấu hình model từ quá trình training bộ dữ liệu, nhận giá trị từ getway và tiến hành xử lý dự đoán kết quả đưa ra 6 mức đánh giá của từng mức độ cảnh báo theo quy ước, sau đó gửi dữ liệu cho các thiết bị đầu cuối phát cảnh báo.

1.1.3 Đánh giá kết quả training

Áp dụng các kiến thức đã học về trí tuệ nhân tạo nói chung và học máy nói riêng. Ứng dụng được công nghệ thư viện học máy tensorflow js mạnh mẽ của google, chúng em đưa ra giải pháp ứng dụng học máy vào mô hình dự đoán chất lượng không khí, bằng cách này chúng ta có thể tái sử dụng model, xây dựng trạm quan trắc ở nhiều vị trí, nhiều nơi khác nhau, dễ dàng tiện lợi triển khai, cập nhập dữ liệu đánh giá quá khứ liên tục, phù hợp với tình trạng của môi trường hiện tại.



Hình 3.9: Sơ đồ mô hình trainer trên tensorflow

Sau khi đã training dữ liệu bộ dataset, chúng ta tiến hành sử dụng kỹ thuật Cross – validation để ước tính độ chính xác hay lỗi của thuật toán, mục đích của kỹ thuật nhằm chia tập dữ liệu ban đầu thành training Data được sử dụng để huấn luyện mô hình và một tập dữ liệu độc lập được sử dụng để đánh giá.

Sau khi khai báo bộ dữ liệu DATA TEST là array_datatest_x và array_datatest_y ta tiến hành gọi đến hàm predictSample() để lấy kết quả thu được từ mô hình. Ta thu được bảng kết quả 1000 dữ liệu, so sánh giữa kết quả dự đoán và kết quả training.

Việc ước tính độ chính xác hay lỗi được tính trung bình trên các thử nghiệm thực tế để đánh giá mức độ hiệu quả của mô hình.

Các yếu tố quan trọng để đánh giá độ chính xác của 1 mô hình học máy là:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Trong đó :

TP (True Positive): Số lượng dự đoán chính xác.

TN (True Negative): Số lượng dự đoán chính xác một cách gián tiếp

FP (False Positive - Type 1 Error): Số lượng các dự đoán sai lệch

FN (False Negative - Type 2 Error): Số lượng các dự đoán sai lệch một cách gián tiếp

Thuật toán ở đây chúng em sử dụng là Adam là bộ thuật toán xử lý dữ liệu số, tìm kiếm độ dốc học có giám sát mạnh nhất của thư viện tensorflow

Adam là một thuật toán trong feedforward neural network(FFNN) thuộc lĩnh vực deep learning, là thuật toán phân lớp nhị phân, các hàm kích hoạt trong thuật toán (activation functions) đóng vai trò quan trọng trong việc điều chỉnh độ dốc của đạo hàm và có thể được sử dụng thay cho thủ tục giảm độ dốc ngẫu nhiên cổ điển để cập nhật lặp đi lặp lại trọng số mạng dựa trên dữ liệu huấn luyện(nguồn:<https://bitly.com.vn/gCxme>)

Các thuật toán liên quan khác:

Thuật toán Gradient Thích ứng (AdaGrad) duy trì tốc độ học tập trên mỗi tham số giúp cải thiện hiệu suất đối với các vấn đề có gradient thưa thớt (ví dụ: các vấn đề về ngôn ngữ tự nhiên và thị giác máy tính).

Root Mean Square Communication (RMSProp) cũng

duy trì tốc độ học trên mỗi tham số được điều chỉnh dựa trên mức trung bình của các cường độ gần đây của các gradient đối với trọng số (ví dụ như tốc độ thay đổi của nó). Điều này có nghĩa là thuật toán giải quyết tốt các vấn đề trực tuyến và không cố định (ví dụ: nhiễu).

Sử dụng các mô hình và bộ dữ liệu lớn, chúng em chứng minh Adam có thể giải quyết các vấn đề học sâu thực tế một cách hiệu quả.

Với các tham số xử lý training

Epoch: 400

Số lần học, training lặp lại của bộ dữ liệu, số lần càng nhiều độ chính xác càng cao, sau khi test trên số lần học, chúng tôi thấy 200 là số lần học cho kết quả chính xác và tốt nhất

Learning rate: 0,005

Tốc độ học của thuật toán, tốc độ càng lớn, yêu cầu, cấu hình máy càng mạnh và tốc độ training nhanh hơn

Batch size: 60

Chia bộ dữ liệu data thành 60 phần, để tránh epoch cao sẽ gây ra lỗi khi training, batch size càng cao training sẽ càng ổn định nhưng, tốc độ training sẽ lâu hơn

a) Độ chính xác(ACC)

Cách đánh giá độ chính xác cao và đơn giản nhất của 1 mô hình học máy, là dựa vào công thức tính ACC

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Trong đó : ta sẽ chia ra 2 mức cảnh báo, là có thể chấp nhận được và nguy hiểm, với Threshold=0

TP (True Positive): Số lượng dự đoán chính xác, và có mức độ dự đoán >3

TN (True Negative): Số lượng dự đoán chính xác, và có mức độ dự đoán <=3

FP (False Positive): Số lượng dự đoán sai, và có mức độ dự đoán >3

FN (False Negative): Số lượng dự đoán sai, và có mức độ dự đoán <=3

Với bộ dữ liệu TEST ,ta thu được 1000 kết quả, trong đó

- TP= 529
- TN=410
- FP=21
- FN=40

$$\text{ACC} = \frac{529+410}{529+410+21+40} * 100 = 93\%$$

Kết quả thu được cho thấy độ chính xác (ACC) đối với một mô hình học máy là khá cao, đến 93%.

Biểu đồ ROC

Tiếp đến để đánh giá về một mô hình học máy ta tiến hành đánh giá biểu đồ ROC: thể hiện mối quan hệ, sự đánh đổi và ý nghĩa lựa chọn một model phù hợp của độ nhạy và tỷ lệ cảnh báo sai. Việc model đưa ra cảnh báo có nguy cơ cháy rừng hoặc không có nguy cơ cháy rừng đúng so với thực tế thì model có độ chính xác cực kỳ cao, nhưng không hoàn toàn là 100%. Nếu như model dự báo là không có

nguy cơ cháy rừng nhưng thực tế là có nguy cơ cháy rừng sẽ vô cùng nguy hiểm hơn so với việc không có nguy cơ cháy rừng mà dự báo thành có nguy cơ cháy rừng. Do đó ta sẽ phân loại có nguy cơ cháy rừng vào nhóm positive còn không có nguy cơ cháy rừng vào nhóm negative (tiêu cực).

Các kết quả của model xảy ra sẽ rơi vào 4 nhóm sau:

TP (True Positive): Số lượng dự đoán chính xác, và có mức độ dự đoán > Threshold

TN (True Negative): Số lượng dự đoán chính xác, và có mức độ dự đoán <= Threshold

FP (False Positive): Số lượng dự đoán sai, và có mức độ dự đoán > Threshold

FN (False Negative): Số lượng dự đoán sai, và có mức độ dự đoán <= Threshold

Bây giờ ta coi lớp 1 là lớp Positive, lớp 0 là lớp Negative, câu hỏi đặt ra là làm thế nào để tăng mức độ báo nhầm (FPR) để giảm mức độ bỏ sót (FNR)?

Để mô tả mối liên hệ giữa độ nhạy và đặc hiệu là biểu đồ ROC (receiver operating characteristic). Biểu đồ ROC có nguồn gốc quân sự, vì nó được ứng dụng trong việc phát hiện tàu của địch trên màn hình radar trong Thế chiến thứ II, nhưng trong thời gian 3 thập niên qua, biểu đồ ROC đã được ứng dụng trong chẩn đoán và tiên lượng rất thành công.

Biểu đồ ROC có trục tung (y-axis) là tỉ lệ TPR, và trục hoành (x-axis) là tỉ lệ FPR. Cả hai tỉ lệ dĩ nhiên có giá trị dao động từ 0 đến 100 (hay từ 0 đến 1, nếu dùng xác suất). Hai tỉ lệ này được ước tính cho từng giá trị tham chiếu. Phương pháp xét nghiệm tốt sẽ có những điểm tham chiếu tập trung vào khu vực “Tây Bắc”, tức là những điểm ở góc trái thuộc phía trên của biểu đồ. Những điểm này cho chúng ta biết đó là những giá trị tham chiếu có độ nhạy cao và độ dương tính giả thấp.

Ta có:

True Positive Rate (TPR)

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

False Positive Rate (FPR)

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

Threshold

Bao gồm 6 mức từ 0->6 tương ứng 6 mức độ cảnh báo được quy ước

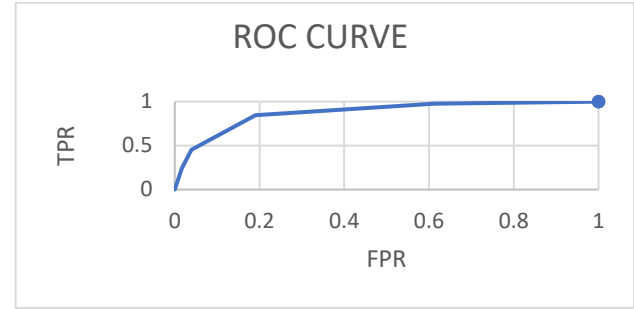
Áp dụng tính mức TPR và FPR đối với từng ngưỡng Threshold ta thu được bảng sau:

Threshold	FPR	TPR
0	1	1
1	1	1
2	0,61265	0,97858
3	0,19037	0,84532
4	0,03913	0,45174

5	0,01641	0,24038
6	0	0

Bảng 3.1: Giá trị ROC

Ta tiến hành vẽ biểu đồ ROC và đánh giá nhận xét:



Hình 3.10: Biểu đồ ROC CUVE

Ứng với từng Threshold từ 0->6 ta có các cặp giá trị TPR và FPR. Bằng cách nối kết các điểm trên biểu đồ ROC, chúng ta sẽ có một biểu đồ ROC liên tục

Ở đây chúng ta có hai chỉ số (FPR và TPR), và chúng biến thiên ngược chiều nhau. Do đó, chúng ta cần một “chỉ số dung hòa” cả hai chỉ số này. Một cách đánh giá đường cong ROC tốt nhất là ước tính diện tích dưới đường biểu diễn ROC (còn gọi là area under the curve – AUC).

AUC	Ý nghĩa
>0.90	Rất tốt (Excellent)
0.80 đến 0.90	Tốt (Good)
0.70 đến 0.80	Trung bình (Fair)
0.60 đến 0.70	Không tốt (Poor)
0.50 đến 0.60	Vô dụng (Fail)

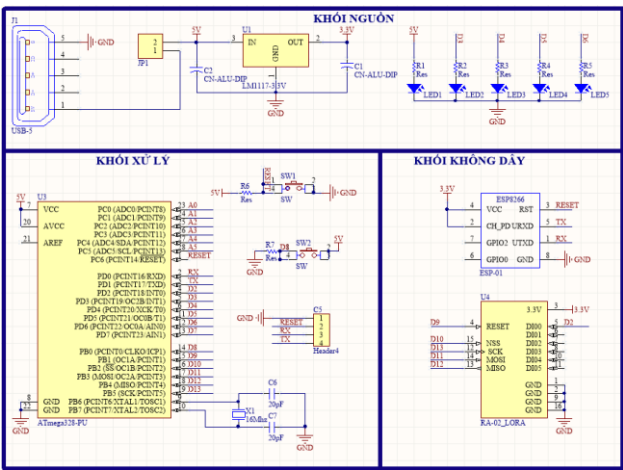
Bảng 3.2: Diễn giải ý nghĩa của diện tích dưới đường biểu diễn ROC (AUC)

So sánh đường cong ROC của chúng ta với biểu đồ và bản diễn giải ý nghĩa diện tích dưới đường cong ROC ta có thể ước lượng thấy AUC của chúng ta thuộc khoảng 0.8 -> lớn hơn 0.9 thuộc trong khoản tốt đến rất tốt

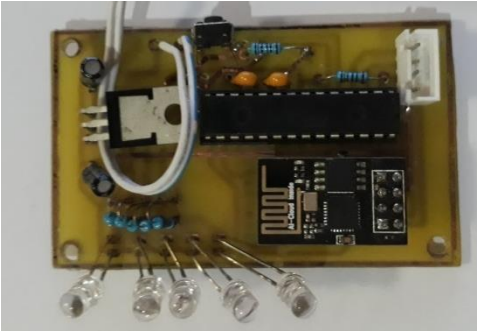
Vậy kết quả đã rõ đánh giá độ chính xác model với đường cong ROC cho về kết quả rất tốt.

3.2.1. Thiết kế phần cứng

Gateway lora được dùng để thu thập dữ liệu từ các cảm biến với tần số 433Mhz, cho phép khoảng cách truyền nhận tín hiệu lên đến 3km. Gateway sử dụng chip lập trình atmega328 kết hợp với esp8266 sử dụng bootloader arduino để lập trình cho phép quản lý truyền nhận gói tin từ các cảm biến, gửi về lưu trữ tại server data. Từ nguồn dữ liệu này ta có thể chúng ta còn có thể sử dụng cho các dự án nghiên cứu khác.



Hình 3.11: Sơ đồ nguyên lý mạch Gateway



Hình 3.12: Mạch Gateway Lora



Hình 3.13: Gateway Lora

Ngay trên gateway chúng tôi có thiết kế 5 led báo hiệu trạng thái hoạt động của gateway. Bao gồm các trạng thái như: Led báo nguồn, kết nối wifi, nhận tín hiệu từ node lora, báo lỗi lora.

Để thuận tiện cho việc kết nối wifi cho gateway, Chúng tôi đã thiết kế nên 1 giao diện web ngay tại gateway.

Kết Nối Wifi

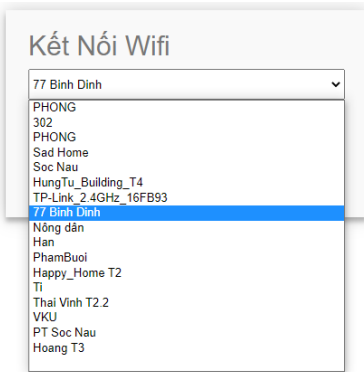
77 Bình Định

Mật khẩu Wifi

Kết nối

Hình 3.14: Form kết nối wifi

Tại đây web sẽ liệt kê tất cả các mạng wifi khả dụng mà nó có thể bắt được. Chúng ta chỉ cần chọn tên wifi để kết nối và nhập mật khẩu wifi. Sau đó gateway sẽ khởi động lại để thay đổi mạng wifi.

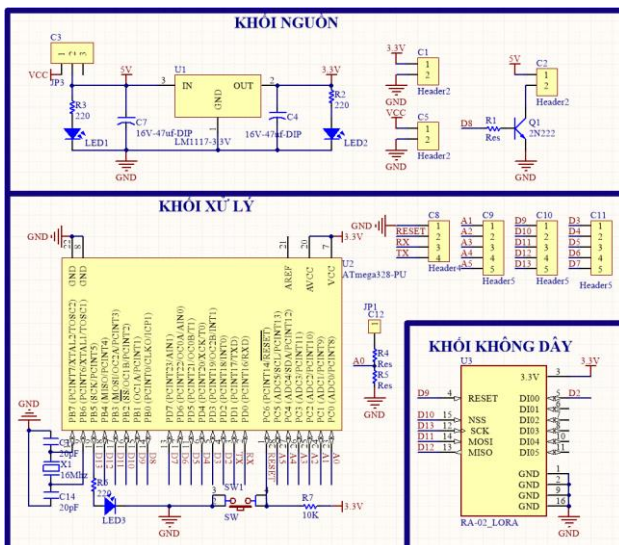


Hình 3.15: Danh Sách Wifi

Node Lora được sử dụng để thu thập dữ liệu gửi về gateway với tần số 433Mhz, cho phép khoảng cách truyền nhận tín hiệu lên đến 3km. Node Lora sử dụng chip lập trình atmega328 kết hợp sử dụng bootloader arduino để lập trình. Với công nghệ truyền gói tin bằng sóng Lora, Node Lora sử dụng rất ít năng lượng để hoạt động do đó tiết kiệm được rất nhiều chi phí.

Với độ phủ sóng rộng như Lora chúng ta có thể triển khai được nhiều cảm biến ở nhiều nơi khác nhau. Cùng với khả năng tiết kiệm năng lượng, ta có thể lắp đặt tại cái nơi khó tiếp cận để bảo dưỡng, hoặc những nơi mạng lưới điện chưa phủ tới để hoạt động trong thời gian dài mà không cần phải thay pin. Ngoài ra chúng tôi đã thiết kế Node lora có thể hoạt động trong dải điện áp rộng từ 5v – 12v. Do đó chúng ta có thể sử dụng nguồn từ pin hoặc acquy kết hợp với pin năng lượng mặt trời cho các dự án lớn cần thu thập dữ liệu ở cường độ cao.

Node Lora chúng tôi đã thiết kế với 6 pin Analog và 6 pin digital có thể linh hoạt trong việc lắp đặt và thay thế các cảm biến cho phù hợp với mục đích sử dụng cho từ dự án khác nhau.



Hình 3.16: Sơ đồ nguyên lý Node Lora



Hình 3.17: Mạch Node Lora

4. Đánh giá quá trình thử nghiệm

Một models không thể hoạt động nếu thiếu data, đối với trường hợp những dự án ứng dụng thực tế IOT trong đời sống xã hội, y học, Có tầm quan trọng rất lớn. Lượng dataset quá nhỏ thì dễ dẫn đến overfitting và model không thể học được đầy đủ các features cho các trường hợp tổng quan hay nói khác đi là model thiếu khả năng tổng quan hóa của một chương trình, điều này khá là nguy hiểm vì trên thực tế có rất nhiều trường hợp có thể xảy ra theo cả 2 hướng tích cực và tiêu cực.

Thu thập dữ liệu xong thì cần tiến hành tiền xử lý vì hầu như dữ liệu thu thập được đều là dữ liệu thô nên không thể đưa thẳng vào Models. Ta sẽ tiến hành đánh nhãn dữ liệu. Đây là bước khá quan trọng vì sẽ đánh giá mô hình chúng ta làm việc tốt hay không! Đánh nhãn sai dữ liệu sẽ làm cho model dự đoán và đánh giá sai dẫn đến tốn nhiều thời gian và công sức bỏ ra cho quá trình training.

Một dự án học máy nói chung và IOT nói riêng sau khi qua các bước: chuẩn bị dữ liệu, mô hình hóa dữ liệu, tinh chỉnh các mô hình thì bước cuối cùng quan trọng nhất là đánh giá độ chính xác, một mô hình có quy mô tầm cỡ, bao gồm nhiều thuật toán chuyên sâu, cấu hình xử lý cực khủng, nhưng nếu không có sự đồng bộ giữa dữ liệu, thuật toán, các mô hình xử lý, cho ra kết quả tỷ lệ rủi ro cao, độ chính xác thấp thì cũng xem như không man lại giá trị gì.

Qua biểu đồ (ROC-AUC) áp dụng cho dự báo cháy rừng, rút ra kết luận sau:

Đối với biểu đồ ROC, bất kỳ dự án IoTs nào nói chung và chất lượng không khí nói riêng thì luôn yêu cầu ngưỡng TPR (độ chính xác thật) càng cao để tỉ lệ dự báo sai sót là thấp nhất và ngưỡng FPR (độ chính xác giả) ở mức trung bình (không thấp cũng không cao). Vì việc cho ra kết quả “sai tích cực” FPR (mức độ an toàn mà cảnh báo có nguy hiểm) vẫn tốt hơn việc dự đoán “sai tiêu cực” FNT(có mức đo lường nguy hiểm mà báo an toàn) .

Để giảm thiểu FNT từ mức thấp đến mức thấp nhất ta buộc phải tăng FPT từ ngưỡng thấp lên trung bình bằng cách thay đổi threshold

Nếu TPR = 1 thì độ chính xác của dự đoán là cao nhất, nhưng model chưa hoàn toàn chính xác bởi vì dự đoán không bao giờ có tính chất hoàn hảo có thể chúng ta bỏ sót một khía cạnh nào đó chưa được đánh giá, cho nên TPR =1 vẫn chưa thể kết luận model là hoàn hảo 100%

Vì dự án này yêu cầu đầu ra bao gồm 2 class (có nguy cơ cháy hoặc không) nên FPR không được quá 0,5

Tùy vào mức độ dự án, số class thì FPR sẽ có một mức nhất định, đối với dự án cháy rừng nếu thấp hơn 0,5 thì sẽ là một model tốt, thuật toán adam đã cho thấy điều đó tốt hơn so với sgd khi cho mức tỉ lệ FTR thấp hơn.

4.1. Kết luận

Một dự án IOT ứng dụng kỹ thuật machine learning có rất nhiều phương pháp, model để đánh giá và tiến hành xử lý cho phù hợp. Chúng ta nên tiến hành đánh giá một cách khách quan, chuẩn xác nhằm đáp ứng đầy đủ nhất những nhu cầu cần thiết của dự án. Khi tiến hành xử lý nên áp dụng nhiều chương trình đánh giá, thuật toán để có cách nhìn khách quan về độ chính xác hay yêu cầu của chương trình mà ta điều chỉnh, tìm ra thuật toán phù hợp. Cần phải xây dựng một mô hình dự án, hoàn chỉnh về cả phần cứng lẫn phần mềm, nên là một mô hình mở để có thể dễ dàng cập nhật, nâng cấp thêm chức năng để ứng dụng mở rộng phạm vi dự án trong tương lai.

4.2. Hướng phát triển trong tương lai

Trong tương lai chúng tôi tiếp tục hoàn thiện sản phẩm, mô hình huấn luyện. Trong tương lai gần, chúng tôi dự định cung cấp thêm kết quả thực nghiệm trên những tập dữ liệu lớn hơn và cải thiện độ chính xác cũng như tốc độ của chương trình. Và thử nghiệm các loại mô hình huấn luyện khác nghiên cứu triển khai tính năng dự báo nồng độ các



chất trong không khí.

Tài liệu tham khảo

- [1] N. A. Tuấn, in *A Predictive System for IoTs Reconfiguration*, Đà Nẵng.
- [2] C. Mai, "zingnews," 02 10 2018. [Online]. Available: <https://zingnews.vn/o-nhiem-khong-khi-o-tq-khien-1-1-trieu-nguoi-chet-tre-moi-nam-post881199.html>.
- [3] vinmec, "Thông tin sức khỏe," vinmec, [Online]. Available: <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/bui-sieu-minh-anh-huong-toi-suc-khoe-nao/>.
- [4] BỘ TÀI NGUYÊN VÀ MÔI TRƯỜNG, "QUYẾT ĐỊNH," 2019. [Trực tuyến]. Available: <https://thuvienphapluat.vn/van-ban/Tai-nguyen-Moi-truong/Quyết-dinh-1459-QĐ-TCMT-2019-ky-thuat-tinh-toan-va-cong-bo-chi-so-chat-luong-khong-khi-Viet-Nam-428215.aspx>.
- [5] "khí, Tiêu luận ô nhiễm không," [Trực tuyến]. Available: <http://thuviendethi.org/tieu-luan-o-nhiem-khong-khi-van-de-toan-cau-thuc-trang-va-giai-phap-4209/>.
- [6] [Online]. Available: <http://timmachhoc.vn/din-gii-nghien-cu-tien-lng-roc-receiver-operating-characteristic/>.

(BBT nhận bài: .../.../2021, phân biện xong: .../.../2021))

Thông tin về tác giả

	Nguyễn Chí Thành: - Điện thoại: 0935530521
	Nguyễn Tuấn Vương: - Điện thoại: 0399460184
	Nguyễn Thanh Hào: - Điện thoại: 0868621096