**SILE 26:**

**True Positive (TP):** Số lượng trường hợp mô hình dự đoán là Positive và thực tế cũng là Positive.

* **False Positive (FP):** Số lượng trường hợp mô hình dự đoán là Positive nhưng thực tế là Negative (nhầm lẫn).
* **True Negative (TN):** Số lượng trường hợp mô hình dự đoán là Negative và thực tế cũng là Negative.
* **False Negative (FN):** Số lượng trường hợp mô hình dự đoán là Negative nhưng thực tế là Positive (nhầm lẫn).

**SILE 28+29**

• True Positive (TP)

• False Positive (FP)

• True Negative (TN)

• False Negative (FN)

**SILE 30:**

Một giá trị AUC cụ thể có thể diễn đạt như sau:

* AUC = 0.5 : Mô hình không có khả năng phân loại tốt hơn một cách ngẫu nhiên.
* AUC > 0.5 và gần 1: Mô hình có khả năng phân loại tốt và tốt hơn càng gần 1.
* AUC = 1: Mô hình phân loại tuyệt vời, hoàn hảo.

**SILE 37:**

**Numpy: Gói thư viện xử lý dãy số và ma trận nhiều chiều**

**Matplotlib: Biểu diễn dữ liệu dưới dạng đồ thị 2 chiều, 3 chiều**

**Pandas: Xử lý, phân tích dữ liệu dưới dạng bảng**

**Scikit-learn (Sklearn) là công cụ tính toán , training**

**SILE 38:**

**std : độ lệch chuẩn**

**count : đếm số lượng**

**bộ tứ phân vị**

**bộ tứ phân vị thứ nhất**

**trung vị**

**SILE 39:**

Từ đồ thị và hình vẽ ta thấy không có mối tương quan tuyến tính nào giữa rain và các feature khác

Nên rất khó để mô hình hồi quy tuyến tính đạt được hiệu quả cao. Vì thế em dứa trên ý tưởng ban đầu em quyết định chọn mô hình phân loại cho bài .

**SILE 42:**

đối với các feature với kiểu dữ liệu số ta tiến hành giảm phương sai bằng kỹ thuật min max scaler để đưa giá trị về đoạn 0-1 nhằm đảm bảo hiệu suất của mô hình k bị ảnh hưởng

Tập huấn luyện (train) và tập kiểm tra (test) với tỷ lệ dữ liệu kiểm tra là 30% (test\_size = 0.3). Điều này có nghĩa là chúng ta lấy 70% dữ liệu để huấn luyện mô hình và 30% dữ liệu được sử dụng để đánh giá và kiểm tra hiệu suất của mô hình. Gía trị ngẫu nhiên random\_state để đảm bảo quá trình chia dữ liệu được diễn ra một cách ngẫu nhiên.

Sau khi thực hiện quá trình train, test, mô hình đạt được những kết quả sau đây:

**SILE 45:**

True positive (TP): Số dự đoán đúng thời tiết sẽ mưa là 9137

True negative (TN): Số dự đoán đúng thời tiết sẽ không mưa là 40722

False positive (FP): Số dự đoán thời tiết sẽ không mưa nhưng thực tết là mưa là 3019

False negative (FN): Số dự đoán thời tiết sẽ mưa nhưng thực tết là không mưa là 1710 Từ bảng Confusion Matrix ta tính được các giá trị F1, Precision, Recall xấp xỉ 91.34%. Mô hình dự đoán rất tốt

**SILE 46:**

* + Biểu đồ thể hiện tổng lượng mưa của các tỉnh dễ nhìn tỉnh có lượng mưa nhiều nhất ở đây là Đà Lạt nơi có khí hậu vô cùng mát mẻ.
  + Lượng mưa ít nhất ở tỉnh Phan Thiết
  + Qua 2 biểu đồ trên ta biết được lượng mưa , và thời tiết của các tỉnh như thế nào từ đó có những biện pháp phòng chóng mưa và lũ lụt. Nếu lượng mưa nhiều quá thì mọi người ở các tỉnh này hay đi du lịch sẽ cần phải trang bị dụng cụ , đồ dung thích hợp để tránh những trường hợp xấu xảy ra.
* **SILE 47:**
* Để ứng phó với nhiệt độ cao, có một số biện pháp quan trọng cần được thực hiện. Để đảm bảo an toàn và sức khỏe, việc duy trì cơ thể được giữ trong tình trạng mát mẻ là quan trọng. Đầu tiên, hãy đảm bảo uống đủ nước suốt ngày để tránh mất nước do mồ hôi. Thứ hai, hạn chế hoạt động ngoài trời vào thời gian nhiệt độ cao nhất, thường là giữa buổi trưa. Sử dụng áo mát mẻ và nón bảo vệ đầu cũng là cách bảo vệ cơ thể khỏi tác động của nhiệt độ cao. Đồng thời, tăng cường việc sử dụng điều hòa không khí hoặc quạt gió trong những nơi nhiệt độ tăng cao. Quan trọng nhất, hãy lắng nghe cơ thể và nếu có dấu hiệu của căng thẳng nhiệt độ, như sốt, hoặc chói loá, cần tìm nơi mát mẻ và sẵn sàng cần cấp cứu nếu cần.
* **SILE 48:**
* Nhận xét: Ba tháng có nhiệt độ cao nhất ở đây là tháng 4,5,6 với nhiệt độ trung bình cao nhất lên đến 32,726 °C .
* Biện Pháp: Để giúp ứng phó với các tháng có nhiệt độ cao, có thể áp dụng một số biện pháp sau đây:
* Điều chỉnh hoạt động hàng ngày: Tránh thực hiện các hoạt động vật lý mạnh vào thời gian nhiệt độ cao nhất trong ngày. Chuyển các hoạt động ra ngoài buổi sáng sớm hoặc buổi tối khi nhiệt độ thấp hơn.
* Duy trì cân bằng nước: Uống đủ nước để duy trì cân bằng nước cơ thể trong thời tiết nhiệt đới. Tránh thức uống có chứa cafein và cồn, vì chúng có thể gây mất nước.
* Sử dụng quần áo mát mẻ: Mặc quần áo thoải mái, mỏng và mát mẻ để giúp cơ thể thoát nhiệt dễ dàng hơn.
* Bảo vệ da khỏi tác động của nhiệt độ cao: Sử dụng kem chống nắng và đội nón để bảo vệ da khỏi tác động của tia UV và nhiệt độ cao.
* Tạo bóng mát và thông gió: Sử dụng rèm cửa, màn cửa hoặc cửa sổ để che nắng và tạo bóng mát cho ngôi nhà. Mở cửa sổ để tạo thông gió và làm ngôi nhà mát mẻ hơn.
* Sử dụng máy lạnh và quạt: Sử dụng máy lạnh và quạt để làm giảm nhiệt độ bên trong nhà, tạo môi trường thoải mái cho gia đình.
* Chăm sóc đặc biệt cho người già và trẻ em: Đặc biệt chú trọng đến việc chăm sóc sức khỏe cho người già và trẻ em, bởi vì họ thường dễ bị ảnh hưởng bởi nhiệt độ cao.
* Giữ gìn vệ sinh: Thường xuyên tắm rửa và giữ vệ sinh cá nhân để tránh bị nhiễm trùng hoặc các vấn đề sức khỏe khác do nhiệt độ cao.
* Tìm nơi trú ẩn mát mẻ: Khi cần thiết, tìm nơi trú ẩn mát mẻ như thư viện, trung tâm thể dục hoặc cửa hàng để tránh tác động của nhiệt độ cao.
* Chế độ ăn uống hợp lý: Bổ sung thêm các loại thực phẩm có chứa nước, như rau cải và trái cây tươi, để giữ cho cơ thể luôn đủ nước và dưỡng chất trong thời tiết nhiệt đới.

Kết hợp những biện pháp trên sẽ giúp bạn và gia đình ứng phó tốt hơn với tháng có nhiệt độ cao, bảo vệ sức khỏe và tạo môi trường sống thoải mái hơn.

**-lượng mưa**

* Nhận xét: Ba tháng có lượng mưa cao nhất là 8,9, 10 với lượng mưa trung bình cao nhất lên đến 11,234 mm
* Biện Pháp: Để ứng phó với các tháng có lượng mưa cao và tăng cường khả năng quản lý dòng chảy nước, có thể áp dụng một số biện pháp sau đây:
* Xây dựng hệ thống thoát nước hiệu quả: Đảm bảo hệ thống thoát nước trong khu vực làm việc tốt, bao gồm các cống thoát nước, cống rãnh và hố chứa nước. Thực hiện công tác bảo dưỡng định kỳ để tránh tắc nghẽn và ngập úng.
* Rừng ngập mặn và cây xanh: Xây dựng và bảo vệ các khu vực rừng ngập mặn, cảnh quan cây xanh để hấp thụ lượng nước dư thừa và giảm thiểu tác động của lũ lụt.
* Xây dựng hệ thống chứa nước: Xây dựng các hồ chứa nước dự trữ để giữ lại lượng nước lớn trong thời gian mưa lớn. Các hồ này có thể giúp kiểm soát dòng chảy nước ra các con sông và hạn chế nguy cơ lũ lụt.
* Hệ thống dẫn nước: Xây dựng hệ thống đường ống dẫn nước từ các vùng có lượng mưa lớn đến các vùng cần nước để tận dụng nguồn nước dư thừa.
* Quản lý cảnh quan và xây dựng bền vững: Thực hiện quản lý cảnh quan thông minh, sử dụng các biện pháp bảo vệ đất đai như xây dựng bậc cảnh quan, cải tạo đất, và trồng cây trồng phủ.
* Hệ thống cảnh báo sớm: Phát triển hệ thống cảnh báo sớm về lượng mưa dự kiến, để cung cấp thông tin kịp thời cho cư dân và cơ quan chính trị để chuẩn bị ứng phó.
* Kế hoạch ứng phó dự lựa: Xây dựng kế hoạch ứng phó dự lựa cho các trường hợp lượng mưa cao đột ngột. Điều này bao gồm việc chuẩn bị các biện pháp sơ tán dân cư và đảm bảo nguồn cung cấp nước sạch.
* Hợp tác vùng địa phương: Hợp tác giữa các tổ chức chính trị, cơ quan khí tượng, chính phủ địa phương và cộng đồng dân cư để thực hiện các biện pháp ứng phó hiệu quả.
* Tầm nhìn bền vững: Thực hiện tầm nhìn bền vững trong quản lý tài nguyên nước và đất đai để giảm thiểu tác động của thay đổi khí hậu và tăng cường khả năng chịu đựng.
* Kết hợp những biện pháp trên sẽ giúp cải thiện khả năng quản lý lượng mưa cao, hạn chế tác động của lũ lụt và đảm bảo an toàn cho cư dân trong thời kỳ mưa lớn.
* **SILE 49:**

Việc sử dụng thông tin thời tiết trong cuộc sống hàng ngày mang lại nhiều lợi ích quan trọng. Thông tin về thời tiết giúp chúng ta dự đoán và chuẩn bị cho những điều kiện môi trường thay đổi, từ đó đảm bảo an toàn và tiện lợi hơn. Khi biết trước về thời tiết, chúng ta có thể lên kế hoạch cho các hoạt động ngoài trời, chuẩn bị trang phục và trang thiết bị phù hợp. Thông tin này còn hỗ trợ quyết định về việc di chuyển, lựa chọn phương tiện giao thông và đối phó với tình huống khẩn cấp. Khuyến khích sử dụng thông tin thời tiết không chỉ giúp tiết kiệm thời gian và năng lượng, mà còn tạo điều kiện tốt hơn cho cuộc sống hàng ngày.

1. Logistic Regression: Mô hình hồi quy logistic dùng để dự đoán xác suất của hai lớp.
2. K-nearest Neighbors (KNN): Dự đoán dựa trên các điểm gần nhất trong không gian đặc trưng.
3. Extra Trees: Một biến thể của Random Forest, xây dựng nhiều cây quyết định để dự đoán.
4. Naive Bayes Classification: Dự đoán dựa trên giả định độc lập giữa các đặc trưng và sử dụng Bayes' theorem.
5. Decision Tree: Cây quyết định phân chia dựa trên các câu hỏi liên quan đến dữ liệu.
6. Random Forest: Tập hợp các cây quyết định để cải thiện độ chính xác và tránh overfitting.
7. Multilayer Perceptron: Mạng neural nhân tạo với nhiều tầng ẩn, thường được sử dụng cho dự đoán phức tạp.
8. XGBoost: Thuật toán boosting ensemble mạnh mẽ, kết hợp nhiều cây quyết định.