**Câu 1. Phân biệt các loại AI khác nhau? Trình bày được định nghĩa của các loại học máy (Học giám sát, không giám sát, bán giám sát, học tăng cường)?**

- Weak or Narrow AI: là AI được tạo ra cho một mục đích cụ thể nào đó. Sử dụng những AI này để hoàn thành 1 công việc nhất định và nó không thể học những cái mới, AI sẽ đưa ra quyết định dựa trên thuật toán được lập trình và tập huấn luyện. Ứng dụng trong web search, recommendation engines, virtual assistant,…

- Strong or Generalized AI: AI có thể tương tác và vận hành với nhiều công việc độc lập và không liên quan đến nhau. Nó có thể học những công việc mới để giải quyết các vấn đề mới xuất hiện, nó cũng có thể tự dạy nó bằng những chiến lược mới. Nó có thể thực hiện bất cứ nhiệm vụ trí tuệ nhân tạo nào mà con người có thể thực hiện.

- Super or Conscious AI: là một khái niệm lý thuyết đề cập đến loại AI ngoài việc giải quyết các nhiệm vụ trí tuệ nhân tạo thì nó còn có khả năng tự nhận thức,, tự suy nghĩ và trải nghiệm cảm xúc giống như con người. Đây là mục tiêu dài hạn và thách thức trong nghiên cứu trí tuệ nhân tạo.

- Học giám sát là phương pháp học máy mà trong đó mô hình được huấn luyện dựa trên một tập dữ liệu đã được gán nhãn. Mỗi mẫu đầu vào có một nhãn tương ứng, giúp mô hình học cách ánh xạ từ đầu vào đến đầu ra. VD: Email spam, giá nhà

- Học không giám sát: là phương pháp học máy mà trong đó mô hình được huấn luyện dựa trên một tập dữ liệu không có nhãn. Mục tiêu là tìm ra cấu trúc hoặc mẫu ẩn trong dữ liệu. VD: Phân cụm khách hàng thành các nhóm khác nhau dựa trên hành vi mua sắm, giảm số chiều dữ liệu

- Học bán giám sát: là kết hợp cả học giám sát và không giám sát. Mô hình được huấn luyện trên một tập dữ liệu mà chỉ có một phần nhỏ dữ liệu được gán nhãn, phần còn lại không có nhãn. VD: Nhận diện khuôn mặt trong các bức ảnh, 1 số ít gán nhãn, còn lại là ảnh không gán

- Học tăng cường: là phương pháp học máy mà trong đó một tác nhân (agent) học cách hành động trong một môi trường sao cho tối ưu hóa một hàm phần thưởng nhất định. Tác nhân sẽ thử nghiệm các hành động khác nhau và nhận phần thưởng hoặc hình phạt tương ứng. VD: cờ vua, các trò chơi điện tử

**Câu 2. Các tiêu chí để đánh giá mô hình học máy? Cách chọn tiêu chí đánh giá phù hợp trong một số ngữ cảnh bài toán An toàn thông tin.**

- Accuracy: thông số đo lường tỷ lệ phần trăm các dự đoán đúng trên tổng số dự đoán. Nó cho biết mô hình của bạn đúng bao nhiêu phần trăm trong tổng số các trường hợp

- Recall: tỉ lệ số lượng mẫu dự đoán đúng so với tổng các mẫu trong dữ liệu kiểm tra . Thông số này trả lời câu hỏi trong tất cả các mẫu dương tính thật, mô hình đã phát hiện được bao nhiêu.

- Precision: tỷ lệ các mẫu được dự đoán là dương tính mà thực sự là dương tính. Nó trả lời câu hỏi: trong tất cả các mẫu mà mô hình dự đoán là dương tính, có bao nhiêu mẫu thật sự là dương tính.

- F1 score: là trung bình điều hòa của Precision và Recall. Nó cung cấp một thước đo cân bằng giữa hai tiêu chí này, đặc biệt hữu ích khi bạn cần cân đối giữa precision và recall của mẫu dương tính.

- AUC-ROC: đo lường khả năng của mô hình trong việc phân biệt giữa các lớp dương tính và âm tính. Giá trị AUC cao (gần 1) cho thấy mô hình có khả năng phân biệt tốt giữa hai lớp, trong khi giá trị gần 0.5 cho thấy mô hình hoạt động ngẫu nhiên.

**1. Phát hiện xâm nhập**:

* **Recall**: Quan trọng để đảm bảo rằng không bỏ sót bất kỳ xâm nhập nào.
* **Precision**: Để giảm thiểu cảnh báo sai, tránh làm phiền người dùng với các cảnh báo không cần thiết.
* **F1 Score**: Cân bằng giữa Recall và Precision.

**2. Phân loại email spam**:

* **Precision**: Để tránh việc các email quan trọng bị phân loại nhầm là spam.
* **F1 Score**: Cân bằng giữa Precision và Recall.

**Câu 3. Trình bày cách phân chia dữ liệu và tiền xử lí dữ liệu phổ biến?**

- **Train-Test Split**: Dữ liệu được chia thành hai phần: một phần để huấn luyện mô hình (training set) và phần còn lại để kiểm tra hiệu suất của mô hình (testing set) theo các tủ lệ như 70-30, 80-20, 90-10

- **Cross-Validation (K-Fold Cross Validation):** Dữ liệu được chia thành K phần (folds), mỗi phần lần lượt được sử dụng làm tập kiểm tra trong khi K-1 phần còn lại được sử dụng làm tập huấn luyện. Quá trình này được lặp lại K lần, mỗi lần một phần khác được dùng làm tập kiểm tra. Thông thường K sẽ được chọn là 5 hoặc 10

- **Stratified Sampling:** Dữ liệu được chia sao cho tỉ lệ của các lớp trong tập dữ liệu huấn luyện và tập kiểm tra giống với tỉ lệ của các lớp trong toàn bộ dữ liệu. Thường được sử dụng khi dữ liệu không cân bằng nhằm đảm bảo rằng mỗi lớp được đại diện đầy đủ trong cả tập huấn luyện và tập kiểm tra.

**Tiền xử lý dữ liệu:**

- Xử lý dữ liệu thiếu: có thể loại bỏ dữ liệu giá trị thiếu nếu nó nhỏ hoặc điền giá trị thiếu bằng giá trị trung bình (mean), trung vị (median) hoặc giá trị phổ biến (mode)

- Chuẩn hóa dữ liệu:

+ **Normalization**: Chuyển đổi giá trị của các đặc trưng về một phạm vi chung, thường là [0, 1]. Phù hợp cho các mô hình như KNN hoặc neural networks.

+ **Standardization**: Chuyển đổi giá trị của các đặc trưng sao cho chúng có trung bình bằng 0 và độ lệch chuẩn bằng 1. Phù hợp cho các mô hình như linear regression hoặc SVM

- Mã hóa các biến phân loại

**+ One-Hot Encoding**: Biến một đặc trưng phân loại thành nhiều đặc trưng nhị phân. Mỗi giá trị của biến phân loại được biểu diễn bởi một cột riêng biệt.

+ **Label Encoding**: Gán một giá trị số duy nhất cho mỗi giá trị của biến phân loại.

- Giảm số chiều: Sử dụng PCA để chọn lọc các đặc trưng nhằm giữ lại các thông tin quan trọng nhất, tránh gây nhiễu cho mô hình

- Cân bằng dữ liệu: tăng số lượng mẫu bằng cách sao chép hoặc tạo thêm mẫu mới (over-sampling), giảm số lượng mẫu của lớp thừa (under-sampling)

- Phát hiện và xử lý ngoại lệ: Dùng Z-Score, IQR để phát hiện các giá trị ngoại lệ, sau đó có thể điều chỉnh hoặc loại bỏ các giá trị đó.

**Câu 4: Tác dụng của hàm kích hoạt (activation function) trong các mạng nơ-ron? Một số hàm activation phổ biến?**

Hàm kích hoạt (activation function) là một phần quan trọng trong các mạng nơ-ron nhân tạo. Hàm kích hoạt đưa vào tính phi tuyến tính cho mô hình, giúp mạng nơ-ron có khả năng học và biểu diễn các mối quan hệ phức tạp giữa đầu vào và đầu ra. Nếu không có hàm kích hoạt, mạng nơ-ron chỉ có thể biểu diễn các mối quan hệ tuyến tính

Một số hàm activation phổ biến:

**Hàm Sigmoid**

A white background with black text

Description automatically generated

A white background with black text

Description automatically generated

A white background with black text

Description automatically generated

A white background with black text

Description automatically generated

**Câu 5: Phân biệt sự khác nhau của các dạng thuật toán học máy như Phân cụm (clustering), phát hiện sự dị thường (anomaly detection), phân loại (classification).**

**Phân Cụm (Clustering)**

**Mục tiêu**: Phân cụm là một phương pháp học không giám sát nhằm nhóm các đối tượng dữ liệu thành các cụm sao cho các đối tượng trong cùng một cụm có sự tương đồng cao hơn với nhau so với các đối tượng trong cụm khác.

**Đặc điểm**:

* **Không giám sát**: Không có nhãn trước, dữ liệu được nhóm dựa trên sự tương đồng.
* **Tìm mẫu**: Tìm các cấu trúc hoặc mẫu ẩn trong dữ liệu.

**Thuật toán phổ biến**:

* **K-means**: Chia dữ liệu thành K cụm, tối thiểu hóa khoảng cách giữa các điểm và tâm cụm.
* **Hierarchical Clustering**: Xây dựng cây phân cấp của các cụm.

**Phát Hiện Sự Dị Thường (Anomaly Detection)**

**Mục tiêu**: Phát hiện các điểm dữ liệu không bình thường hoặc bất thường trong tập dữ liệu. Các điểm bất thường thường có sự khác biệt rõ rệt so với phần còn lại của dữ liệu.

**Đặc điểm**:

* **Không giám sát hoặc bán giám sát**: Thường không có nhãn hoặc chỉ có một số ít nhãn cho các điểm bất thường.
* **Phát hiện điểm ngoại lệ**: Tìm ra các điểm dữ liệu không tuân theo mẫu chung của tập dữ liệu.

**Thuật toán phổ biến**:

* **Isolation Forest**: Xây dựng cây để cô lập các điểm dữ liệu, điểm bất thường thường được cô lập nhanh hơn.
* **One-Class SVM**: Học mô hình từ dữ liệu bình thường để phát hiện các điểm bất thường.
* **Autoencoders**: Mô hình mạng nơ-ron học cách mã hóa và giải mã dữ liệu, các điểm bất thường sẽ có lỗi tái tạo cao.

**Phân Loại (Classification)**

**Mục tiêu**: Phân loại là một phương pháp học có giám sát nhằm gán nhãn cho các đối tượng dữ liệu vào các lớp hoặc nhóm đã biết.

**Đặc điểm**:

* **Có giám sát**: Dữ liệu được huấn luyện với các nhãn trước.
* **Dự đoán nhãn**: Gán nhãn cho dữ liệu đầu vào mới dựa trên mô hình đã học.

**Thuật toán phổ biến**:

* **Logistic Regression**: Mô hình tuyến tính cho các bài toán phân loại nhị phân.
* **Support Vector Machines (SVM)**: Tìm siêu phẳng tốt nhất phân tách các lớp dữ liệu.
* **Decision Trees và Random Forests**: Sử dụng các cây quyết định hoặc tập hợp các cây quyết định để phân loại dữ liệu.
* **Neural Networks**: Sử dụng các lớp nơ-ron để học các mẫu phức tạp trong dữ liệu.

**Câu 6. Học sâu (DL) là gì? Xác định được các lớp, đầu ra, đầu vào trong một số mô hình mạng nơ-ron nhân tạo, học sâu (DL)? Một số ứng dụng cơ bản trong lĩnh vực ATTT.**

Học sâu (Deep Learning) là một nhánh của học máy (Machine Learning) tập trung vào việc sử dụng các mạng nơ-ron nhân tạo với nhiều lớp (deep neural networks) để học từ dữ liệu. Deep Learning đã trở nên phổ biến nhờ khả năng xử lý và học các mẫu phức tạp từ lượng dữ liệu lớn, đặc biệt là trong các lĩnh vực như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, và nhận dạng giọng nói.

Ứng dụng cơ bản: phát hiện gian lận, nhận diện mã độc, IDS, lỗ hổng trong mã nguồn

A screenshot of a computer

Description automatically generated

A white text with black text

Description automatically generated

**Câu 7: Nguyên tắc hoạt động của mạng sinh đối kháng? Một số ứng dụng của mô hình mạng sinh đối kháng (GANs) trong ATTT.**

Mạng Sinh Đối Kháng là một loại mô hình học sâu (Deep Learning) đặc biệt bao gồm hai mạng nơ-ron: **Mạng Sinh (Generator)** và **Mạng Phân Biệt (Discriminator)**. Cả hai mạng này cạnh tranh với nhau bằng việc đối kháng, với mục tiêu cuối cùng là tạo ra các dữ liệu tổng hợp chân thực đến mức không thể phân biệt được với dữ liệu thực.

Ứng dụng trong ATTT: phát hiện và tạo mã độc, phát hiện tấn công mạng, phát hiện gian lận, bảo mật dữ liệu

**Câu 8: Điều gì xảy ra khi kẻ tấn công sử dụng kỹ thuật tấn công đối kháng để đánh lừa mô hình ML? Các biện pháp phòng ngừa là gì?**

Qua mặt các trình AV, IDS, làm giảm độ chính xác của mô hình phân loại. Trong các lĩnh vực nhạy cảm như y tế, tài chính, hoặc an ninh, việc mô hình bị đánh lừa có thể dẫn đến các quyết định sai lầm với hậu quả nghiêm trọng.

Biện pháp phòng ngữa:

+ Huấn Luyện Với Dữ Liệu Đối Kháng (Adversarial Training): tạo và đưa mẫu dữ liệu được tạo mới vào quá trình huấn luyện để mô hình học và nhận diện tốt hơn

+ Sử Dụng Các Mạng Nơ-ron Bền Vững (Robust Networks): thiết kế mạng nơ-ron có khả năng chịu đựng tốt hơn với các biến đổi nhỏ trong dữ liệu đầu vào

+ Phát Hiện Tấn Công Đối Kháng (Adversarial Detection): sử dụng các mô hình khác hỗ trợ phát hiện các mẫu đối kháng trước khi đưa vào mô hình chính

+ Tiền xử lý dữ liệu: xử lý dữ liệu đầu vào để loại bỏ hoặc giảm thiểu các yếu tố đối kháng trước khi đưa vào mô hình

Ensemble Learning: sử dụng nhiều mô hình để tăng cường khả năng tổn quát hóa và giảm xác suất mô hình bị đánh lừa

**Câu 9: Học tăng cường là gì? Một số ứng dụng của Reinforcement Learning (RL) trong An toàn thông tin? Xác định được các khái niệm cơ bản, định nghĩa của RL trong các bài toán điển hình của ATTT.**

Học Tăng Cường (Reinforcement Learning - RL) là một lĩnh vực của học máy, trong đó một tác nhân (agent) học cách hành động trong một môi trường (environment) để tối đa hóa một phần thưởng tổng hợp (cumulative reward) – hay một chiến lược hoàn hảo thông qua việc thử và sai. Các thành phần cơ bản trong 1 mô hình học tăng cường bao gồm:

 **Tác nhân (Agent)**: Thực hiện các hành động trong môi trường.

 **Môi trường (Environment)**: Nơi tác nhân thực hiện các hành động và nhận được phản hồi (feedback).

 **Trạng thái (State)**: Biểu diễn tình trạng hiện tại của môi trường.

 **Hành động (Action)**: Các hành động mà tác nhân có thể thực hiện trong môi trường.

 **Phần thưởng (Reward)**: Phản hồi từ môi trường đối với hành động của tác nhân, giúp tác nhân đánh giá hiệu quả của hành động đó.

Ứng dụng trong ATTT:

+ Phát Hiện và Phòng Chống Tấn Công Mạng: hành vi bất thường

+ Quản Lý và Phân Tích Lỗ Hổng: tối ưu hóa, đề xuất biện pháp khắc phục

+ Tối Ưu Hóa Quản Lý Sự Cố: phản ứng, đối phó với sự cố với các tình huống bảo mật khác nhau

Pentest tool sử dụng RL: DeepExploit (có sử dụng Metasploit)

**Câu 10: Một số phương pháp phát hiện mã độc sử dụng các thuật toán học máy, học sâu**

**+** Phân Loại Sử Dụng Các Thuật Toán Cổ Điển: Decision Tree, Random Forest, SVM, ..

+ Ensemble Learning: Kết hợp nhiều mô hình để tạo ra 1 mô hình mạnh hơn như AdaBoost hoặc Gradient Boosting

+ CNN

+ RNN và LSTM

+ Autoencoders

+ GANs

**Câu 11: Ảnh hưởng của tốc độ học, batch size và epoch lên quá trình huấn luyện mô hình học máy nói chung**

**Tốc độ học** là một siêu tham số xác định mức độ điều chỉnh các trọng số của mô hình trong mỗi lần cập nhật.

**Ảnh Hưởng**

**- Tốc độ học quá cao**:

* + Mô hình có thể không hội tụ và dao động xung quanh giá trị tối ưu mà không đạt được tối ưu thực sự.
  + Có thể gây ra sự mất ổn định và dẫn đến các cập nhật trọng số quá lớn, khiến mô hình học sai lệch.
* **Tốc độ học quá thấp**:
  + Quá trình huấn luyện sẽ diễn ra rất chậm, mất nhiều thời gian để hội tụ.
  + Có thể dẫn đến việc mô hình bị kẹt trong cực tiểu cục bộ và không đạt được hiệu quả cao nhất.
* **Batch size** là số lượng mẫu dữ liệu được sử dụng trong mỗi lần cập nhật trọng số của mô hình.

**Ảnh Hưởng**

* **Batch size nhỏ**:
  + Cập nhật trọng số thường xuyên hơn, điều này có thể giúp mô hình tìm ra các cực tiểu tốt hơn trong không gian tham số.
  + Nhưng có thể dẫn đến việc các cập nhật không ổn định và có nhiều biến động hơn.
  + Yêu cầu ít bộ nhớ hơn, phù hợp với các hệ thống có tài nguyên hạn chế.
* **Batch size lớn**:
  + Cập nhật trọng số ít thường xuyên hơn, điều này có thể làm cho việc hội tụ ổn định hơn.
  + Nhưng có thể bỏ qua các đặc điểm quan trọng trong dữ liệu do việc làm mịn quá mức (over-smoothing).
  + Cần nhiều bộ nhớ hơn, đòi hỏi hệ thống có tài nguyên lớn.

**Epoch** là một vòng lặp hoàn chỉnh qua toàn bộ tập dữ liệu huấn luyện.

**Ảnh Hưởng**

* **Số lượng epoch quá ít**:
  + Mô hình có thể chưa học được đầy đủ từ dữ liệu, dẫn đến underfitting (không khớp).
  + Chưa tối ưu được trọng số, mô hình không đạt được hiệu suất tốt nhất.
* **Số lượng epoch quá nhiều**:
  + Mô hình có thể học quá mức từ dữ liệu huấn luyện, dẫn đến overfitting (quá khớp).
  + Lãng phí tài nguyên tính toán và thời gian huấn luyện, mà không cải thiện thêm hiệu suất.

Câu 12: Khi xây dựng các ứng dụng dựa trên học máy, cần làm gì khi dữ liệu bị mất cân bằng? Tiêu chí nào có thể dùng để đánh giá mô hình trong trường hợp dữ liệu bị mất cân bằng?

- Oversampling: Tăng số lượng mẫu của lớp thiểu số bằng cách sao chép hoặc tạo ra các mẫu mới. Kỹ thuật phổ biến bao gồm:

* **SMOTE (Synthetic Minority Over-sampling Technique)**: Tạo các mẫu mới bằng cách nội suy giữa các mẫu hiện có.
* **ADASYN (Adaptive Synthetic Sampling)**: Tương tự SMOTE, nhưng tập trung vào các mẫu thiểu số khó phân loại.

- Undersampling: Giảm số lượng mẫu của lớp đa số để cân bằng với lớp thiểu số. Phương pháp này có thể gây mất mát thông tin nếu số lượng mẫu của lớp đa số giảm quá nhiều

- Ensemble: Mẫu của lớp đa số sẽ được chia thành từng phần, sau đó kết hợp với mẫu lớp thiểu số và lần lượt đưa vào mô hình, lần lượt kết hợp cho đến khi kết hợp hết tất cả mẫu của lớp đa số. Kết quả cuối cùng của tất cả các mô hình sẽ được kết hợp lại và đưa ra kết quả cuối cùng.

Các tiêu chí đánh giá:

F1, Precision, Recall, AUC Score, G-Mean, Average precision score (AP)

**Câu 13: Học liên kết (Federated Learning - FL) là gì? Ưu và nhược điểm của nó? Công thức tổng hợp mô hình trong FL (thuật toán tổng hợp FedAvg)?**

Học liên kết (Federated Learning) là một phương pháp học máy mà trong đó, các mô hình học tập trên nhiều thiết bị hoặc máy chủ cục bộ và chỉ trao đổi các tham số mô hình (ví dụ: gradient hoặc trọng số) thay vì dữ liệu thô. Điều này cho phép việc học máy được thực hiện một cách phân tán mà không cần di chuyển dữ liệu từ các thiết bị cá nhân lên một máy chủ trung tâm, đảm bảo tính riêng tư và bảo mật của dữ liệu.

**Ưu Điểm**

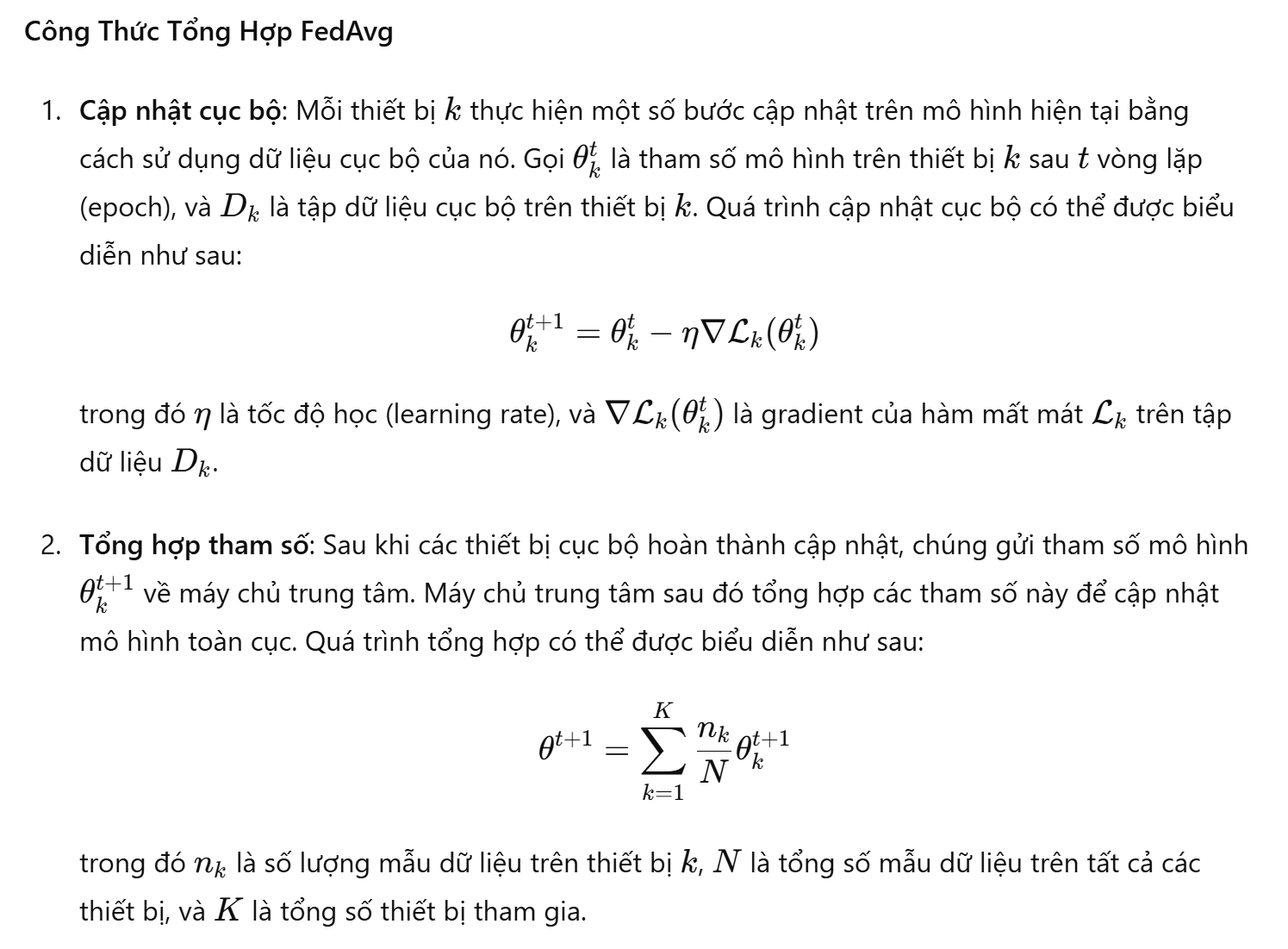
1. **Bảo vệ quyền riêng tư và dữ liệu**: Dữ liệu cá nhân không cần phải rời khỏi thiết bị cá nhân, giảm nguy cơ lộ thông tin cá nhân.

**2. Hiệu quả sử dụng băng thông**: Chỉ các tham số mô hình được trao đổi thay vì toàn bộ dữ liệu thô, tiết kiệm băng thông mạng.

**3. Khả năng mở rộng**: Học liên kết có thể mở rộng đến nhiều thiết bị khác nhau, từ điện thoại di động đến các thiết bị IoT, giúp thu thập và học từ dữ liệu phân tán rộng rãi.

**Nhược Điểm**

1. **Đồng bộ hóa khó khăn**: Đảm bảo các thiết bị đồng bộ và cùng tham gia vào quá trình huấn luyện có thể gặp nhiều khó khăn.
2. **Tính không đồng nhất của dữ liệu**: Dữ liệu trên các thiết bị có thể không đồng nhất và không đồng đều, ảnh hưởng đến chất lượng mô hình.
3. **Hiệu suất mô hình**: Hiệu suất mô hình có thể bị ảnh hưởng do các thiết bị có tài nguyên tính toán hạn chế và việc truyền tham số mô hình qua mạng không ổn định.
4. **Phức tạp trong triển khai**: Cần thiết lập cơ chế bảo mật và đồng bộ phức tạp, gây khó khăn trong triển khai và quản lý.



**Câu 14: Xử lý đặc trưng (feature engineering) là gì? Mục đích áp dụng? Có những kỹ thuật nào để thực hiện?**

Xử lý đặc trưng (Feature Engineering) là quá trình sử dụng dữ liệu thô để tạo ra các đặc trưng (features) mới hoặc cải thiện các đặc trưng hiện có để giúp các mô hình học máy hoạt động tốt hơn. Quá trình này bao gồm các bước như chọn lọc, biến đổi, tạo mới các đặc trưng và xử lý các đặc trưng hiện có.

**Mục Đích Áp Dụng**

1. **Cải thiện hiệu suất mô hình**: Các đặc trưng tốt hơn có thể giúp mô hình học máy học được nhiều thông tin hơn từ dữ liệu, từ đó cải thiện độ chính xác và hiệu suất của mô hình.
2. **Tăng tốc độ huấn luyện và dự đoán**: Việc chọn lọc và biến đổi các đặc trưng giúp giảm số lượng đặc trưng không cần thiết, làm cho quá trình huấn luyện và dự đoán nhanh hơn.
3. **Giảm độ phức tạp của mô hình**: Các đặc trưng tốt hơn có thể giúp đơn giản hóa mô hình, làm cho mô hình dễ hiểu và dễ bảo trì hơn.
4. **Xử lý dữ liệu mất cân bằng**: Tạo ra các đặc trưng mới hoặc điều chỉnh các đặc trưng hiện có có thể giúp mô hình phân loại tốt hơn trong các trường hợp dữ liệu bị mất cân bằng.

Những kỹ thuật để thực hiện:

- Feature Selection: supervise gồm filters, embeddef, hydrid, wrapper

- Feature Transformation: chuẩn hóa và tiêu chuẩn hóa

- Handling Missing Values: Drop or replace NA

- Handling Imbalanced data: undersampling, oversampling (duplicate or SMOTE)

- Handling Outliers: Loại bỏ, thay thế, capping và rời rạc hóa

- Binning: đổi các biến liên tục thành rời rạc

- Encoding: label, one-hot

**Câu 15: MLOps là gì? Nó giải quyết vấn đề gì trong ngữ cảnh các ứng dụng có sử dụng Học máy? Đánh giá tiềm năng và cơ hội của MLOps trong lĩnh vực An toàn thông tin?**

MLOps (Machine Learning Operations) là một phương pháp và thực tiễn nhằm cải thiện việc triển khai, quản lý và vận hành các mô hình học máy (ML) trong môi trường sản xuất. MLOps kết hợp các nguyên tắc của DevOps (phát triển và vận hành phần mềm) với các yêu cầu cụ thể của học máy để đảm bảo rằng các mô hình ML có thể được phát triển, triển khai, và duy trì một cách hiệu quả và nhất quán.

Các vấn đề MLOps giải quyết:

+ Tích Hợp và Triển Khai Liên Tục (CI/CD)

+ Quản Lý Phiên Bản và Theo Dõi Mô Hình

+ Giám Sát và Bảo Trì Mô Hình

+ Tự Động Hóa và Tái Sử Dụng: tạo điều kiện cho việc tự động hóa và tái sử dụng các thành phần trong quy trình ML, từ đó giảm chi phí và thời gian phát triển

Tiềm năng và cơ hội  
+ Nâng cao hiệu quả phát hiện và ngăn chặn mối đe dọa

+ Giám Sát Liên Tục và Cảnh Báo Sớm

+ Tối ưu hóa chi phí cho hoạt động an ninh mạng

+ Tăng cường khả năng phục hồi sau sự cố

**Câu 16: Reinforcement Learning có thể được áp dụng để tự động hóa các quyết định an ninh mạng như thế nào? Cho ví dụ**

Reinforcement Learning (RL) có thể được áp dụng để tự động hóa các quyết định an ninh mạng thông qua việc học hỏi từ môi trường và tối ưu hóa hành động nhằm đạt được mục tiêu bảo mật.

**Ví Dụ Cụ Thể**: Một agent RL có thể được huấn luyện để phát hiện các hành vi bất thường của phần mềm trên máy tính. Khi phát hiện một tập tin có hành vi lạ, agent có thể quyết định cách ly tập tin này và thực hiện các kiểm tra chi tiết hơn. Phần thưởng được định nghĩa dựa trên khả năng phát hiện chính xác và nhanh chóng mã độc.

**Câu 17. Qui trình kiểm thử xâm nhập hiện tại có những vấn đề nào mà các phương pháp học máy có thể giúp giải quyết? Trình bày phương pháp ứng dụng Học tăng cường trong bài toán Kiểm thử xâm nhập (Pentest/Network Exploitation)?**

 **Tính thủ công và tốn thời gian**: Kiểm thử xâm nhập thường yêu cầu kỹ thuật viên có kỹ năng cao thực hiện các bước thủ công, dẫn đến việc tốn nhiều thời gian và chi phí.

 **Phạm vi hạn chế**: Con người có thể bỏ sót một số lỗ hổng hoặc không kiểm tra hết tất cả các khả năng tấn công do giới hạn về thời gian và tài nguyên.

 **Khả năng tái tạo kém**: Mỗi lần kiểm thử xâm nhập có thể có kết quả khác nhau do phụ thuộc vào người thực hiện và phương pháp sử dụng.

 **Độ phức tạp và quy mô lớn**: Các hệ thống hiện đại phức tạp và có quy mô lớn, đòi hỏi các phương pháp kiểm thử phải hiệu quả và có thể mở rộng

Ứng dụng của RL trong Pentest

**Ví Dụ 1: Tấn Công Từ Chối Dịch Vụ (DDoS)**

1. **State**:
   * Lưu lượng mạng hiện tại, thông tin về các thiết bị và dịch vụ đang chạy.
2. **Action**:
   * Quét cổng, phát hiện các dịch vụ mở.
   * Thử các payload tấn công DDoS trên các dịch vụ phát hiện được.
3. **Reward**:
   * Thành công làm gián đoạn dịch vụ: +10 điểm.
   * Thử tấn công nhưng bị phát hiện: -10 điểm.

4. Môi trường: Bao gồm toàn bộ hệ thống lưu lượng mạng sau mỗi hành động của agent

**Câu 18: An ninh mạng thích ứng là gì? Các mô hình học tăng cường có thể được sử dụng để phát triển các hệ thống phòng thủ mạng thích ứng (adaptive cybersecurity) như thế nào?**

An ninh mạng thích ứng là một phương pháp bảo mật mạng có khả năng tự động điều chỉnh và thích nghi với các mối đe dọa mới xuất hiện, thay đổi trong môi trường và hành vi của người dùng. Thay vì dựa vào các biện pháp bảo mật tĩnh và quy tắc cố định, bằng cách triển khai bảo mật thích ứng, ta có thể đánh giá và phân tích các hành vi và sự kiện trong thời gian thực để thực hiện hành động phòng ngừa trước khi có hành vi vi phạm.

RL có thể học hỏi và thích nghi với các mối đe dọa mới, giúp hệ thống phòng thủ mạng luôn cập nhật, ngoài ra RL có thể tự động hóa nhiều nhiệm vụ bảo mật, giúp tăng hiệu quả của hệ thống phòng thủ mạng. Cuối cùng, RL có thể giúp hệ thống phòng thủ mạng phục hồi nhanh chóng sau các cuộc tấn công.

**Câu 19: Nêu một số ưu điểm mà các mô hình ngôn ngữ lớn (Large Language Model – LLM) trong 1 bài toán điển hình của an toàn thông tin? Gợi ý: phát hiện lỗ hổng phần mềm (vulnerability detection), sửa lỗi bảo mật phần mềm (vulnerability fixing code generation), kiểm thử fuzzing, phát hiện mã độc (malware detection)**

- Xác định ngứ cảnh của mã nguồn với độ chính xác cao, giúp phân tích và xác định chính xác vị trí và loại lỗ hổng.

- Có kiến thức chuyên môn về các ngôn ngữ lập trình, lỗi bảo mật phổ biến, nguyên tắc lập trình an toàn và cách vá mã nguồn nhờ được huấn luyện trên big data

- Giảm thời gian sửa lỗi do LLMs có thể tự động hóa đề xuất và tạo mã sửa lỗi nhanh chóng

- Khả năng học hỏi liên tục giúp kịp thời cập nhật các lỗ hổng mới xuất hiện và các biện pháp bảo mật phù hợp

- Cải thiện quy trình xây dựng dự án (CI/CD), giúp phát hiện lỗ hổng bảo mạt ngay trong giai đoạn phát triển

**Câu 20: Generative AI là gì? Phân tích tiềm năng mang lại cũng như rủi ro có thể gây ra vấn đề mất an toàn thông tin, tấn công mạng, khống chế và điều khiển các hệ thống thông tin.**

Generative AI là một nhánh của trí tuệ nhân tạo sử dụng các thuật toán học máy để tạo ra dữ liệu mới, mô phỏng các mẫu và đặc điểm có trong dữ liệu huấn luyện. Nó có khả năng tạo ra văn bản, hình ảnh, âm thanh, video, code.

Tiềm năng: Sáng tạo nội dung, hỗ trợ phát triển phần mềm, tạo văn bản, hỗ trợ trong phân tích y khoa,.. Ở các mô hình học máy đang có lượng dữ liệu hạn chế hoặc mất cân bằng, Generative AI cũng có thể tạo thêm dữ liệu phù hợp để góp phần vào hiệu suất đầu ra của mô hình

Rủi ro:

+ Tấn công lừa đạo hoặc mạo danh: Gen AI có thể tạo ra các nội dung giả mạo như hình ảnh hoặc video deepfake, gây ảnh hưởng đến danh tiếng, tài sản

+ Tấn công phần mềm và hệ thống: có thể tạo các mã độc có khả năng trốn tránh dựa vào các file mã độc đã được huấn luyện, từ đó tấn công

+ Lam dụng dữ liệu cá nhân bằng cachs tạo ra các cuộc tấn công phising, social engineering để lừa đảo -> xâm phâm quyền riêng tư, gây thiệt hại tài chính cho nạn nhân