**Nghiên Cứu và Phát Triển Các Mô Hình AI Để Giảm Thiểu Thời Gian Ngừng Hoạt Động Của Máy Móc Công Nghiệp**

1. Nội dung
2. Giới thiệu về đề tài:
   * Vấn đề nghiên cứu: Phát hiện bất thường trong các máy móc công nghiệp để giảm thiểu chi phí bảo trì và ảnh hưởng đến quá trình sản xuất.
   * Input và Output: Dữ liệu đầu vào là tập dữ liệu từ cảm biến máy móc; output là nhận dạng trạng thái bất thường và cảnh báo cần thiết.
   * Lí do chọn đề tài: Giảm thiểu chi phí và thời gian bảo trì thông qua phát hiện sớm các dấu hiệu hỏng hóc, từ đó lên kế hoạch bảo trì kịp thời và hiệu quả.
   * Khả năng ứng dụng thực tế: Cải thiện hiệu quả hoạt động của các nhà máy và giảm thiểu thời gian ngừng máy không lên kế hoạch.
   * Tính thời sự: Áp dụng công nghệ AI để đối phó với những thách thức về bảo trì trong công nghiệp nặng.
3. Mục tiêu của đề tài:
   * Phát triển một mô hình có khả năng phát hiện bất thường chính xác và kịp thời.
   * Đánh giá hiệu quả của các thuật toán Isolation Forest, VAE, và GAN trong việc phát hiện bất thường.
   * Đề xuất một hệ thống cảnh báo tự động để hỗ trợ quyết định bảo trì và sửa chữa.
4. Nội dung nghiên cứu của đề tài:
   * Nghiên cứu các thuật toán để xử lý và phân tích dữ liệu từ cảm biến máy móc.
   * Thử nghiệm và so sánh các phương pháp phát hiện bất thường.
   * Tích hợp và phát triển hệ thống cảnh báo dựa trên kết quả phân tích.
5. Phương pháp thực hiện:
   * Thu thập dữ liệu từ các máy móc trong nhà máy.
   * Phát triển và huấn luyện mô hình phát hiện bất thường.
   * Xây dựng và thử nghiệm hệ thống cảnh báo tự động.
6. Kết quả, sản phẩm dự kiến:
   * Một hệ thống thông minh có khả năng cảnh báo bất thường và hỗ trợ bảo trì máy móc.
   * Báo cáo đánh giá hiệu quả các mô hình và khuyến nghị các biện pháp tối ưu.
7. Tham khảo
   *  **Barnett, W. & Lewis, T. (1994). Outliers in Statistical Data. Wiley.**
   * Cuốn sách này cung cấp một cái nhìn tổng quan về phát hiện và xử lý ngoại lệ trong dữ liệu thống kê, có thể ứng dụng trong phân tích dữ liệu máy móc.
   *  **Chalapathy, R. & Chawla, S. (2019). Deep Learning for Anomaly Detection: A Survey. Artificial Intelligence Review.**
   * Bài báo này đánh giá các kỹ thuật học sâu áp dụng trong phát hiện bất thường, bao gồm VAE và GAN, cung cấp cái nhìn sâu sắc về phương pháp hiện đại.
   *  **Liu, F.T., Ting, K.M., & Zhou, Z. (2008). Isolation Forest. Proceedings of the 2008 Eighth IEEE International Conference on Data Mining.**
   * Bài báo này giới thiệu thuật toán Isolation Forest, một phương pháp mới và hiệu quả trong việc phát hiện bất thường.
   *  **Kingma, D.P. & Welling, M. (2014). Auto-Encoding Variational Bayes. Proceedings of the 2nd International Conference on Learning Representations, ICLR.**
   * Công trình nghiên cứu này trình bày về mô hình VAE, một kỹ thuật quan trọng trong việc học biểu diễn dữ liệu và phát hiện bất thường.
   *  **Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. Proceedings of the 27th International Conference on Neural Information Processing Systems.**
   * Giới thiệu về mạng GAN, một cách tiếp cận sáng tạo cho việc phát hiện bất thường thông qua học máy đối kháng.
8. **Kế hoạch thực hiện**

Giai đoạn 1: Lên kế hoạch và tìm hiểu (Tháng 2 - Tháng 3)

1. Tuần 1-2:
   1. Xác định mục tiêu cụ thể của khóa luận, định nghĩa các yêu cầu và các tiêu chí đánh giá thành công.
   2. Tìm hiểu về thuật toán Isolation Forest và VAE (Variational Autoencoder), so sánh ưu và nhược điểm của hai thuật toán.
2. Tuần 3-4:
   1. Nghiên cứu các công cụ và thư viện sẽ sử dụng cho việc triển khai thuật toán (Python, Scikit-learn, TensorFlow, Keras, v.v.).
   2. Tìm hiểu về dữ liệu sử dụng (loại dữ liệu, đặc trưng, cấu trúc, xử lý trước khi sử dụng).

Giai đoạn 2: Thu thập và xử lý dữ liệu (Tháng 3 - Tháng 4)

1. Tuần 1-2:
   1. Thu thập dữ liệu từ các nguồn thích hợp hoặc từ các dataset sẵn có (Kaggle, các bộ dữ liệu công khai khác, hoặc dữ liệu của công ty ).
   2. Tiến hành tiền xử lý dữ liệu, loại bỏ các điểm dữ liệu không hợp lệ và làm sạch dữ liệu (missing values, outliers, v.v.).
2. Tuần 3-4:
   1. Chuẩn hóa dữ liệu và tạo các đặc trưng cần thiết cho việc huấn luyện các mô hình.
   2. Phân chia dữ liệu thành các tập training, validation, và testing để đảm bảo đánh giá mô hình chính xác.

Giai đoạn 3: Xây dựng mô hình và huấn luyện (Tháng 4 - Tháng 5)

1. Tuần 1-2:
   1. Cài đặt và huấn luyện mô hình Isolation Forest. Tối ưu hóa các tham số và kiểm tra kết quả.
2. Tuần 3-4:
   1. Cài đặt và huấn luyện mô hình VAE. Tối ưu hóa các tham số, kiểm tra và so sánh với kết quả của Isolation Forest.
   2. Sử dụng các kỹ thuật như Cross-Validation để đánh giá hiệu quả của mô hình.

Giai đoạn 4: Đánh giá, so sánh và tối ưu (Tháng 6 - Tháng 7)

1. Tuần 1-2:
   1. Đánh giá kết quả của hai mô hình dựa trên các tiêu chí: độ chính xác (accuracy), độ nhạy (recall), độ đặc hiệu (specificity), v.v.
   2. So sánh hiệu suất của Isolation Forest và VAE, và lý giải nguyên nhân của sự khác biệt.
2. Tuần 3-4:
   1. Tiến hành tối ưu hóa thêm các mô hình nếu cần thiết (fine-tuning).
   2. Kiểm tra lại các mô hình với các tập dữ liệu khác nhau để đảm bảo tính ổn định và tổng quát.

Giai đoạn 5: Viết báo cáo và chuẩn bị bảo vệ (Tháng 7 - Tháng 8)

1. Tuần 1-2:
   1. Bắt đầu viết báo cáo khóa luận: trình bày lý thuyết nền tảng, phương pháp nghiên cứu, kết quả và đánh giá.
2. Tuần 3-4:
   1. Hoàn thiện báo cáo và chuẩn bị slide thuyết trình cho buổi bảo vệ khóa luận.
   2. Chạy thử các demo và các kịch bản trình bày để đảm bảo sự mạch lạc và chuẩn bị cho các câu hỏi từ hội đồng.

Kết thúc (Cuối tháng 8)

* Nộp báo cáo chính thức và bảo vệ khóa luận trước hội đồng