TÓM TẮT BÀI BÁO

## Giới thiệu:

* Các công ty ngày càng phải đối mặt với nhu cầu thay đổi về số lượng sản phẩm và thông số kỹ thuật do sản phẩm tiêu dùng ngày càng có thể tùy chỉnh được.
* Sự biến động này tác động đáng kể đến chi phí sản xuất, buộc các công ty phải quản lý chuỗi cung ứng của mình theo cách tích hợp và năng động để đạt được mức hiệu suất mong muốn.
* Trong môi trường như vậy, việc ra quyết định trở nên phức tạp. Sự phối hợp giữa các mắt xích khác nhau của chuỗi cung ứng vẫn là một nhiệm vụ khó khăn, vì mỗi mắt xích đều tìm kiếm lợi ích tối ưu cục bộ của riêng mình, điều này không nhất thiết tương ứng với lợi ích tối ưu toàn cầu của chuỗi.
* Do đó, các công ty cần trang bị cho mình một hệ thống ra quyết định thông minh có khả năng dự đoán và tiên đoán các trở ngại và vấn đề có thể xảy ra để đảm bảo giao tiếp cả nội bộ và bên ngoài.
* Sự ra đời của Công nghiệp 4.0 đã biến đổi sâu sắc quy trình sản xuất, mở đường cho khái niệm **sản xuất thông minh** (intelligent manufacturing). Sản xuất thông minh hứa hẹn các quy trình sản xuất **tự chủ** (self-sufficient) thông qua việc sử dụng máy móc và thiết bị có khả năng **giao tiếp với nhau qua kết nối kỹ thuật số**. Nhờ đó, các **hệ thống tích hợp** đã trở nên **thông minh hơn** khi chúng **học hỏi từ kinh nghiệm trong quá khứ**, làm cho việc **ra quyết định trở nên đáng tin cậy hơn**.
* Dự đoán rủi ro giao hàng chậm từ trước có thể **thay đổi cuộc chơi**. Các nghiên cứu đã chỉ ra rằng khách hàng dễ chấp nhận việc chậm trễ hơn khi họ được thông báo trước. Họ **ít thất vọng hơn** và **có nhiều khả năng tha thứ** cho vấn đề.
* Vấn đề nghiên cứu trong bài báo này là xác định một mô hình học máy thông minh, hiệu suất cao có khả năng dự đoán và dự báo các đơn hàng giao hàng chậm.

## Chuỗi cung ứng trong thời đại học máy: chuỗi cung ứng 4.0

**Định nghĩa và Bối cảnh:**

* Chuỗi cung ứng 4.0 được mô tả là mối quan hệ giữa Industry 4.0 và chuỗi cung ứng. Đây là hậu quả trực tiếp của sự ra đời của Industry 4.0, bắt nguồn từ ngành công nghiệp Đức.

**Đặc điểm chính:**

* Sử dụng dữ liệu để đưa ra các quyết định sáng suốt và dự đoán các kịch bản tương lai. Thách thức là cung cấp thông tin có thể hành động cho người dùng kịp thời.
* Không chỉ đơn thuần là áp dụng công nghệ; đòi hỏi sự hiểu biết về các khả năng cần thiết (cơ sở hạ tầng, kỹ năng con người, phối hợp) để triển khai hiệu quả công nghệ Industry 4.0 và tạo tác động lên tiêu chí hiệu suất (tính minh bạch, khả năng phản ứng, hiệu quả, tính linh hoạt) và mục tiêu chiến lược của chuỗi cung ứng.
* Các mạng lưới cung ứng kỹ thuật số có tính động, tích hợp và đặc trưng bởi luồng thông tin và phân tích liên tục, tốc độ cao. Là sự kết hợp sử dụng những đổi mới mới nhất trong lĩnh vực internet, robot và công nghệ dữ liệu.
* Các bước ban đầu hướng tới chuỗi cung ứng mới này chủ yếu liên quan đến việc tạo ra các phiên bản kỹ thuật số và/hoặc tự động hóa các nhiệm vụ đơn giản, đơn điệu nhưng tốn công sức.

**So sánh với Chuỗi cung ứng truyền thống:**

* Chuỗi cung ứng truyền thống đi theo một đường thẳng từ nhà cung cấp đến khách hàng.
* Chuỗi cung ứng 4.0 là phiên bản đồng nhất hơn, trong đó thông tin lưu chuyển theo mọi hướng.

## Thuật toán

### Dữ liệu

|  |  |  |
| --- | --- | --- |
| # | Thuộc tính (tên mã) | Ý nghĩa lâm sàng |
| 1 | Order number | Số đơn hàng được gán trong hệ thống ERP |
| 2 | Client ID | Mã định danh của khách hàng trong ERP |
| 3 | Order cost | Giá trị của dòng đơn hàng |
| 4 | Order quantity | Số lượng mặt hàng trong đơn hàng |
| 5 | Due date | Ngày đến hạn giao hàng |
| 6 | Delivery date | Ngày giao hàng thực tế |
| 7 | Delay in days | Số ngày giao trễ (delivery date - due date) |
| 8 | Non-compliance quantity | Số lượng hàng không đạt tiêu chuẩn |
| 9 | Non-compliance level | Tỷ lệ giữa hàng giao và hàng không đạt tiêu chuẩn |
| 10 | Delivered quantity | Số lượng hàng đã được giao thực tế |
| 11 | Remaining quantity | Số lượng còn lại (Order quantity - Delivered quantity) |
| 12 | Administrator ID | Mã định danh của quản trị viên xử lý đơn hàng |
| 13 | Delivery ID | Mã định danh tài liệu/lệnh giao hàng |
| 14 | Delivery Status | Trạng thái giao hàng (sớm, đúng hạn, hoặc trễ) – **biến mục tiêu** |

### Phương pháp nghiên cứu

* **Thu thập dữ liệu:** Dữ liệu được lấy từ hệ thống ERP của một doanh nghiệp sản xuất hàng không vũ trụ ở Morocco. Dữ liệu lịch sử đơn hàng trong bốn năm (2018-2021) được trích xuất, bao gồm các thông tin như số đơn hàng, ID khách hàng, chi phí, số lượng, ngày đến hạn, ngày giao hàng, số ngày chậm trễ, số lượng/mức độ không tuân thủ, số lượng đã giao/còn lại, ID quản trị viên/giao hàng, và tình trạng giao hàng (sớm, đúng hạn, chậm).
* **Tiền xử lý dữ liệu:** Dữ liệu đã được làm sạch trong bước này.
* **Xây dựng mô hình huấn luyện và kiểm tra:** Dựa trên các trường hợp giao hàng chậm trong quá khứ, một mô hình huấn luyện đã được xây dựng. Kết quả từ mô hình huấn luyện được dùng để phát triển mô hình kiểm tra.
* **Triển khai thuật toán Học máy ban đầu:** Thuật toán Random Forest Classifier được sử dụng để dự đoán liệu một đơn hàng có bị giao chậm hay không. Thuật toán này được chọn vì khả năng xử lý dữ liệu chiều cao, đa cộng tuyến, nhanh và không nhạy cảm với overfitting.
* **Đánh giá hiệu suất ban đầu:** Hiệu suất mô hình ban đầu được đánh giá bằng độ chính xác (accuracy). Độ chính xác ban đầu đạt được là 97.9%.
* **Áp dụng Lựa chọn đặc trưng (Feature Selection):** Kỹ thuật này được triển khai sau đó nhằm giảm kích thước tập dữ liệu và tăng độ tin cậy thông tin. Phương pháp sử dụng là ANOVA f-test trong hàm f\_classif() của thư viện scikit-learn.
* **Thử nghiệm với các kịch bản Lựa chọn đặc trưng:** Số lượng đặc trưng được lựa chọn (K) được điều chỉnh để thử nghiệm ba kịch bản: K=5, K=7, và K=10.
* **Đánh giá hiệu suất sau Lựa chọn đặc trưng**: Mô hình được kiểm tra lại với tập dữ liệu đã giảm chiều sau khi áp dụng lựa chọn đặc trưng cho từng kịch bản K.
* **So sánh kết quả và Kết luận:** Kết quả của ba kịch bản lựa chọn đặc trưng được so sánh với mô hình ban đầu. Độ chính xác tốt nhất đạt được là với K=7, đạt 99.3%. Kết quả thực nghiệm cho thấy lựa chọn đặc trưng đã giúp tăng độ chính xác của mô hình dự đoán, mặc dù việc loại bỏ quá nhiều đặc trưng có thể gây tác động ngược lại.