BỘ CÔNG THƯƠNG
CỤC TMĐT & KINH TẾ SỐ

TRƯỜNG ĐẠI HỌC THƯƠNG MẠI
KHOA HTTT KINH TẾ & TMĐT

HỘI THẢO QUỐC GIA
PHÁT TRIỂN THƯƠNG MẠI ĐIỆN TỬ VIỆT NAM
TRONG KỶ NGUYÊN SỐ

GIẢI PHÁP PHÁT HIỆN NGƯỜI BÁN HÀNG GIAN LẬN TRONG THỊ TRƯỜNG TRỰC TUYẾN BẰNG PHƯƠNG PHÁP HỌC MÁY SVM

Th.S Nguyễn Thị Vân Trang
Khoa Hệ thống thông tin kinh tế và thương mại điện tử

Hà Nội, 2020

NỘI DUNG

1. Gian lận thương mại
2. Giải pháp máy véc tơ hỗ trợ (Support vector machine – SVM)
3. Quy trình phát hiện người bán hàng gian lận
4. Kết luận

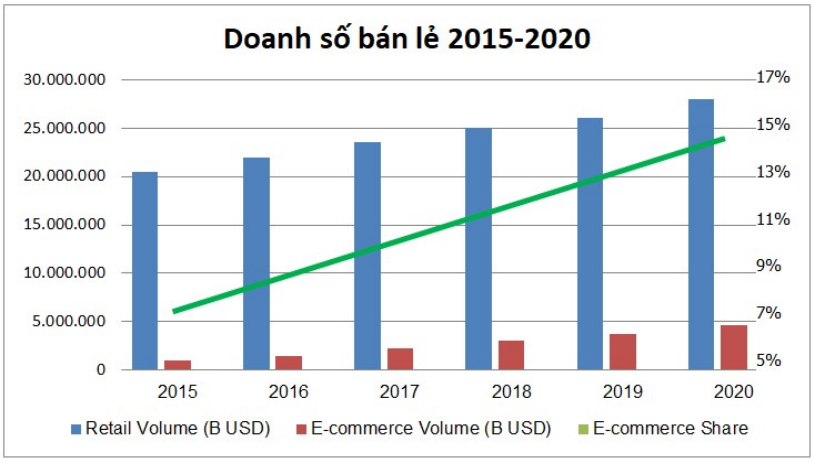
1

GIAN LẬN THƯƠNG MẠI

- Hoạt động kinh doanh trực tuyến trên nền tảng thương mại điện tử như shopee, mạng xã hội…ngày càng thu hút nhiều người tham gia, nhất là trong thời điểm dịch Covid – 19 diễn ra phức tạp.
- Tương tác trên mạng, trong đó nhu cầu mua bán được người dùng tiếp cận ngày càng nhiều hơn khi mọi thứ được thực hiện nhanh chóng, thuận tiện.

2

GIAN LẬN THƯƠNG MẠI



Hình 1: Sự phát triển của thương mại điện tử trong ngành bán lẻ
Nguồn: eMarketer. com

3

GIAN LẬN THƯƠNG MẠI

- Gian lận thương mại là hành vi dối trá, mánh khóe, lừa lọc trong lĩnh vực thương mại thông qua hoạt động mua, bán, kinh doanh, xuất nhập khẩu hàng hóa, dịch vụ nhằm mục đích thu lợi bất chính.

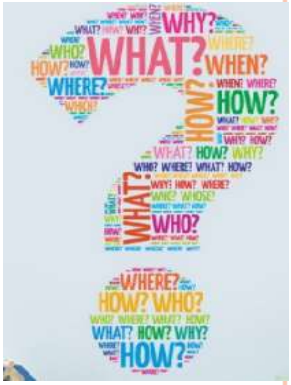
→ **Tập trung nghiên cứu và đưa giải pháp phát hiện gian lận bên phía người bán**

4

THỰC TẾ???

- Khuôn khổ pháp lý
- Thông tin chủ tài khoản
- Phân loại các trang thương mại điện tử ảo

→ Bị hạn chế, khó kiểm soát



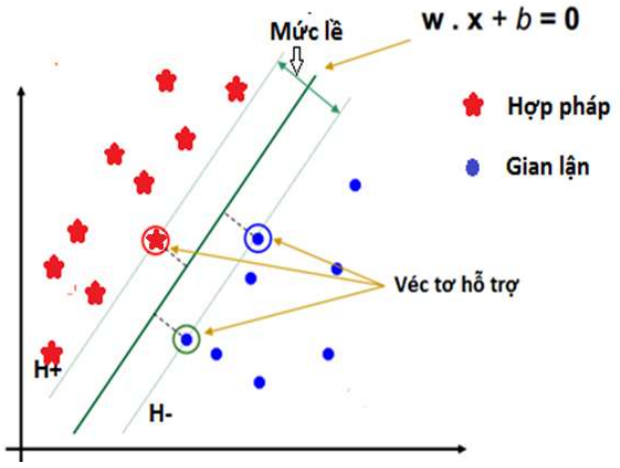
5

GIẢI PHÁP SVM

- SVM là phương pháp phân lớp tuyến tính với mục đích xác định một siêu phẳng để phân tách 2 lớp của dữ liệu (lớp âm và dương) sao cho khoảng cách từ siêu phẳng phân tách này đến các điểm gần nó nhất (gọi là lề) cực đại.
- Hàm phân tách tuyến tính có dạng f(x) = <w.x> + b. Trong đó, w là véc tơ trọng số các thuộc tính, b là một giá trị số thực.

6

GIẢI PHÁP SVM

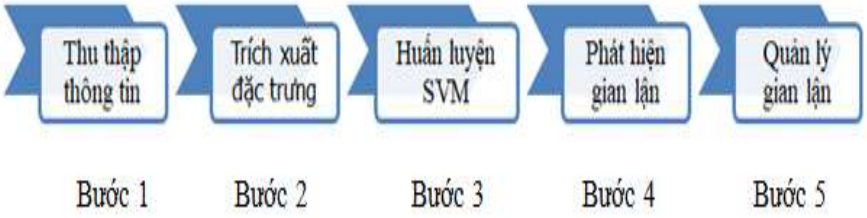


Hình 2: Phát hiện gian lận bằng SVM

7

QUY TRÌNH PHÁT HIỆN NGƯỜI BÁN HÀNG GIAN LẬN

- Bài toán gian lận thương mại phía người bán hàng áp dụng giải thuật SVM được thực hiện qua 5 bước:



Hình 3: Quy trình phát hiện gian lận áp dụng SVM

8

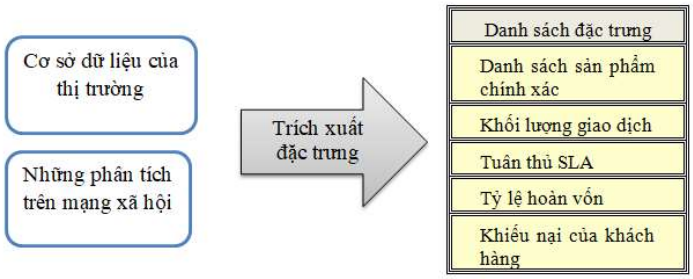
THU THẬP THÔNG TIN

- Thị trường trực tuyến lưu lịch sử mọi hoạt động của người bán diễn ra trên nền tảng công nghệ của họ như Facebook, trang web,..
- Mục đích: Nắm bắt thông tin người bán → Đưa ra nhận định đúng

9

TRÍCH XUẤT ĐẶC TRƯNG

- Liên quan đến việc chuyển đổi thông tin người bán thành tập hợp các đặc trưng riêng.
- Chuẩn hóa, khử trùng lặp và loại bỏ sai lệch → thu được dữ liệu tốt nhất cho quá trình học.

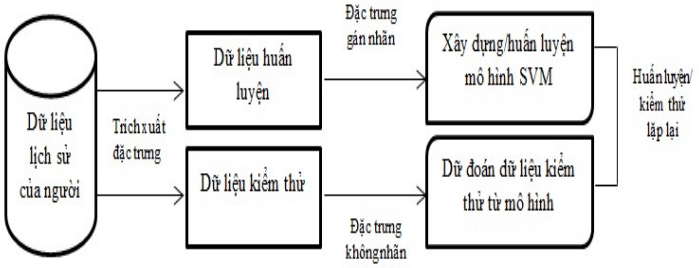


10

Hình 4: Trích xuất đặc trưng

HUẤN LUYỆN SVM

- Thuật toán huấn luyện SVM học hỏi từ tập hợp dữ liệu huấn luyện trong đó mỗi người bán được gán cho một trong hai nhãn (hợp pháp hoặc gian lận) để xây dựng mô hình phân lớp.

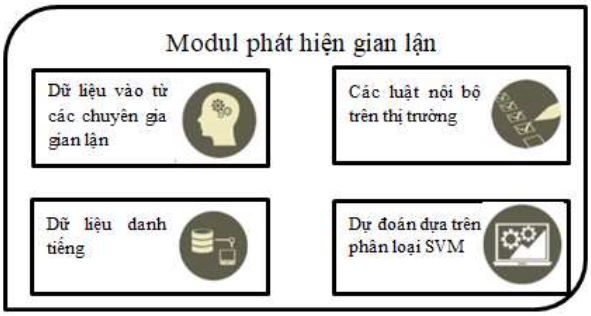


11

Hình 5: Huấn luyện SVM

PHÁT HIỆN GIAN LẬN

- Là tập hợp của nhiều quy trình hoặc chiến lược hoạt động độc lập → kết quả đầu ra giúp hỗ trợ quyết định cuối cùng về việc phân loại người bán có gian lận hay không.



12

Hình 6: Phát hiện gian lận

QUẢN LÝ GIAN LẬN

- Lưu trữ và tập hợp lại danh tính những người bán hàng gian lận khi kết thúc việc dự đoán gán nhãn là gian lận.
- Sau đó thực hiện hành động thích hợp đối với người bán gian lận để xác định chắc chắn đối tượng đó dựa trên những chính sách và luật pháp của thị trường cụ thể.

13

KẾT LUẬN

- Rà soát, phân loại, nâng cao ý thức... → Đánh giá đúng thực trạng, kịp thời phát hiện gian lận thương mại.
- SVM là một phương pháp tốt (phù hợp) đối với những bài toán phân lớp có không gian nhiều chiều (các đối tượng cần phân lớp được biểu diễn bởi một tập rất lớn các thuộc tính) như bài toán phát hiện gian lận người bán hàng.

14

Chân thành cảm ơn!

15