### Câu 3:

**How do the authors substantiate their claims:**

Tác giả đã chứng minh được các công bố của họ :

Tác giả quan sát và thấy rằng số lượng các vụ lừa đảo tăng lên, dẫn đến thiệt hại hàng tỉ đô mỗi năm trên toàn thế giới. Vì thế tác giả đã phát triển và áp dụng các kĩ thuật phát hiện gian lận. Băng việc áp dụng và đánh giá các thuật toán thông minh như khoa học dữ liệu và máy học để xác định các gian lận trong giao dịch điện tử, cụ thể hơn là các hoạt động thẻ tín dụng được thực hiển ở các cổng thanh toán trong các trang web. Để đánh giá mức độ chính xác các kĩ thuật đó, họ đã áp dụng và đánh giá chúng, bằng bộ dữ liệu thực tế của dịch vụ thanh toán trực tuyến phổ biến nhất ở Brazil. Kết quả cho thấy hiệu suất tốt trong phát hiện gian lận, đạt 43%

**What is the methodology adopted to substantiate the claims**

Phương pháp được sử dụng để chứng minh bài báo là :

Tác giả đã dùng chung một phương pháp cho tất cả các kĩ thuật, bắt đầu bằng việc mô tả đặc đặc điểm của bộ dữ liệu, sau đó họ lựa chọn các thuộc tính cần thiết cho quá trình phân tích. Sử dụng InfoGain hiện thị sự tương quan của từng biến, thông qua phần mềm Weka.Tiếp theo, tác xác định các tập huấn luyện và kiểm tra các thuật toán.Để đánh giá các kĩ thuật phát hiện gian lận, họ đã sử dụng những môi trường khác nhau ứng với mỗi mẫu có tham số riêng. Việc hoàn thiện các tham số được thực hiện bằng việc sử dụng các giá trị thực nghiệm khác nhau cho tưng kĩ thuật. Mô tả chi tiết về các kĩ thuật sử dụng thực nghiệm chẳng hạng những tham số được sử dụng cho từng kĩ thuật. Sau thi thực hiện các kĩ thuật, họ xây dựng mức độ tin cậy đối với các gian lận được giao dịch bằng việc xây dựng bảng xếp hạng. Đứng đầu bảng xếp hạng sẽ có xác xuất lừa đảo cao.

**What is the argument of the paper? What are the major theorems:**

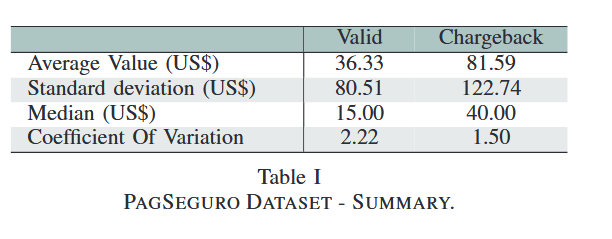
Lập luận của bài báo là gì?

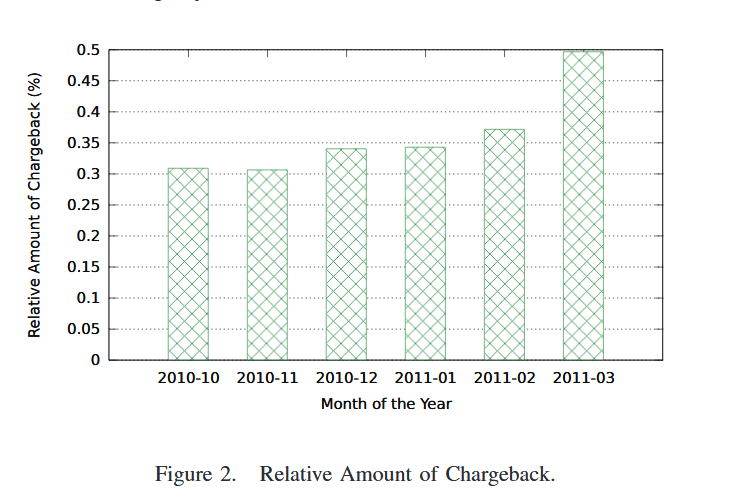
* Sử dụng thuật toán khai thác dữ liệu để xử lí dữ liệu lớn, trong công việc này, họ áp dụng kĩ thuật thuật thông minh tính toán để xác định gian lận đặt biệt trong hoạt động thẻ tín dụng
* Các kết quả được áp dụng trong bộ dữ liệu thực tế của dịch vụ thanh toán điện tử ở Brazil.
* Phần II, mô tả các công việc liên quan,
* Phần III, trình bày các công nghệ thông minh được áp dụng trong phân tích: Bayesian Networks, Logistic Regression, Neural Networks và Random forest.
* Phần IV: Mô tả các trường hợp, sử dụng bộ dữ liệu trong đó họ đã trình bày về tổng quan dữ liệu, các thử nghiệm và công việc tương lại.

**What experimentsare conducted? Data analyses ? Simulations? Bench-marks ? User studies ? Case studies ? Examples ?**

Các thí nghiệm đã được thực hiện là gì? Phân tích dữ liệu ? Mô phỏng ? Nghiên cứu người dùng ? Nghiên cứu trường hợp? Ví dụ?

Paseguro là 1 dịch vụ Web cung cấp thanh toán trực tuyến, thuộc sử hữu của nhà cung cấp nội dung và Internet lớn nhất ở khu vực Mỹ Laitin, tuy nhiên các trang thái giao dịch có thể bị thay đổi vì thế mục địch của việc này là phân tích tập hợp các giao dịch đang diễn ra trong hệ thống của Paseguro. Họ đã sử dụng các thuộc tính đặc trưng trong các giao dịch để sử dụng các kĩ thuật thông minh như Bayesian Networks, Logistic Regression, Neural Networks và Random forest để phát hiện các gian lận trong giao dịch. Các số liệu được biểu diễn như sau:

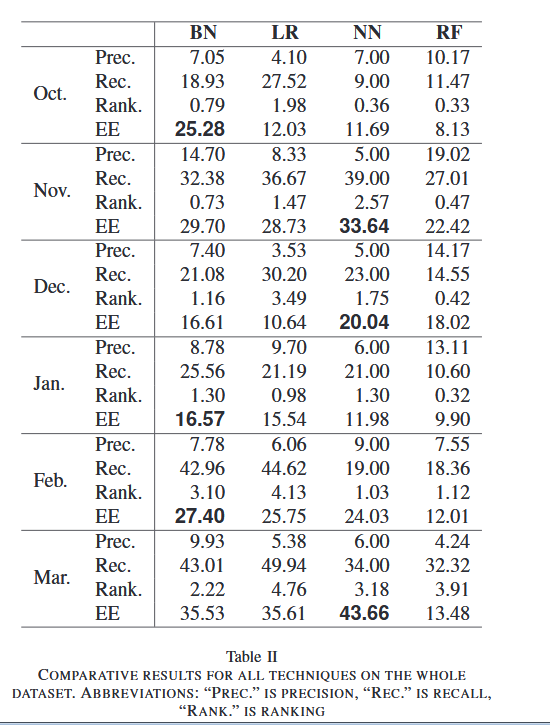


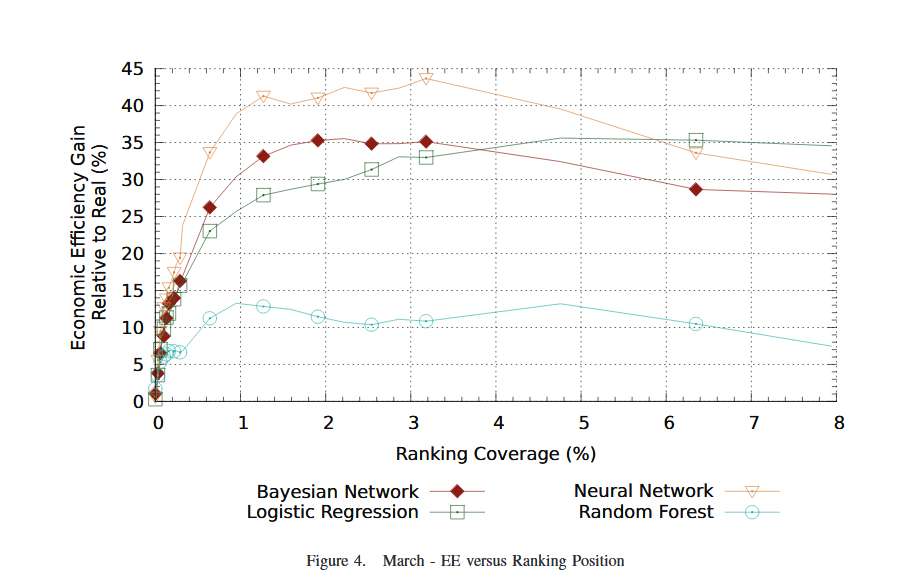


Qua các số liệu thống kê ở trên thì tác giả đã chọn 21 thuộc tính để áp dụng vào các kĩ thuật đó là các thuộc tính:

**Value, Score, Hour, Buyer Type, Buyer Registration Time, Day, Buyer’s Points, Registered, Seller Registration Time, Store Main Category, Credit Card Operator, Credit Card Owner Age, Quantity of installments, Status at the Serasa, Had response from Serasa, CPF, DDD, Federation Unit,**

Tóm tắt kết quả được mô tả ở hình sau, Kết quả tốt nhất là 43,66% EE ở tháng 3. Tháng 10 là tồi tệ nhất với các kĩ thuật. Hiệu quả kinh tế được biểu diễn bằng thứ hạng,





**In short, what makes the claims scientific(as opposed to being mere opinions1)?**

Tác giả sử dụng mô hình phát hiện gian lận khác nhau để dự đoán gian lận trong giao dịch trực tuyến, Áp dụng các kĩ thuật và thuật toán để tính toán để phân tích đánh giá các bộ dữ liệu. Ngoài ra tác giả còn sử dụng bộ dữ liệu thực tế có độ chính xác cao, chứa nhiều giao dịch trong ngày ở dịch vụ thanh toán trực tuyến ở Brazil gọi là Pagseguro. Cuối cùng họ áp dụng chức năng hiệu quả kinh tế (EE) để so sánh các kĩ thuật. Kết quả cho thấy độ chính xác lên đến 43,66% có thể áp dụng để phát hiện các gian lận trong hệ thống thanh toán trực tuyến.