Xác định gian lận giao dịch là gì: là các gian lận trong quá trình giao dịch trong một ứng dụng ví điện tử dựa trên các yếu tố như tần suất giao dịch, số tiền giao dịch, địa chỉ IP(location), thời gian giao dịch, khoảng cách giữa các giao dịch( quá nhanh, lâu rồi mới giao dịch), blacklist.

Giao dịch hợp lệ:

* Tiền giao dịch không quá số dư
* Ip thiết bị quen thuộc
* Thực hiện giao dịch lớn với địa chỉ quen thuộc
* Giao dịch lớn, nhỏ không thuộc diện từng trong blacklist

Các cách thức đánh giá gian lận giao dịch:

* Địa chỉ ip khác lạ so với các lần giao dịch trước
* Thực hiện nhiều giao dịch trong vòng thời gian ngắn
* Có trong danh sách đen
* Lâu chưa giao dịch nhưng đột nhiên giao dịch trở lại
* Không có lịch sử giao dịch với số tiền lớn trước đó
* Rút hết tiền
* Rút/chuyển tiền theo kiểu nhỏ giọt
* Thời gian giao dịch quá nhanh giữa các lần giao dịch (tính từ gửi thông tin)
* Chuyển tiền cho nhiều ví khác nhau trong thời gian ngắn
* Giao dịch amount lớn balance
* Tần suất giao dịch bất thường
* 1 Giao dịch bị thực hiện 2 lần
* Sử dụng nhiều thẻ trên 1 địa chỉ IP
* Nếu User thực hiện giao dịch với Ví Của User khác

Đặc trưng dataset: TransactonID, WalletID, UserID, Amount, TransactionDate, Transaction Type, IP address, Frequency, transactionDuration, PreviousTransactionDate, Balance, isFraud

ID: có thể là kiểu UUID hoặc tùy thuộc theo yêu cầu hay nghiệp vụ cơ bản của hệ thống

TransactionID: Mỗi giao dịch phải có 1 id cố định

WalletID: có thể suất hiện nhiều lần, mỗi người dùng có thể có nhiều ví tiền, nhưng 1 ví chỉ thuộc về 1 user nếu ngược lại thì đó là giao dịch gian lận.

UserIDI: Có thể suất hiện nhiều lần, mỗi người dùng có transaction với walletID tương ứng với vi của họ sở hữu.

IP address: có thể có nhiều địa chỉ giao dịch khác nhau.

Tài liệu tham khảo:  
Phân loại các thuật toán ML:

* <https://machinelearningcoban.com/2016/12/27/categories/>

Over fitting:

* <https://machinelearningcoban.com/2017/03/04/overfitting/>

Decision tree(ID3):

* [https://scikit-learn.org/stable/modules/tree.html#classification](https://scikit-learn.org/stable/modules/tree.html%23classification)
* <https://machinelearningcoban.com/2018/01/14/id3/>

Random Forest:

* [Random Forest algorithm — Machine Learning cho dữ liệu dạng bảng](https://machinelearningcoban.com/tabml_book/ch_model/random_forest.html)
* [https://scikit-learn.org/stable/modules/ensemble.html#random-forests-and-other-randomized-tree-ensembles](https://scikit-learn.org/stable/modules/ensemble.html%23random-forests-and-other-randomized-tree-ensembles)

Ensemble learing:

* <https://viblo.asia/p/ensemble-learning-va-cac-bien-the-p1-WAyK80AkKxX>