Fraudly™

Barak, Demetria, Gina, Jarrod, Joanna

Data

- Found 9% of the requests were Frauds

Feature engineering

Numerical

- Email domains
- Found countries with higher rates of fraud
- Number of total tickets to sell
- Previous history of the users
- Event country = posting country?

Feature engineering

NLP on 'description'

- Topic analysis:

  Topics found: dinner party, education, social networks, logistics, business, university, club logistics, workshop, club amenities.
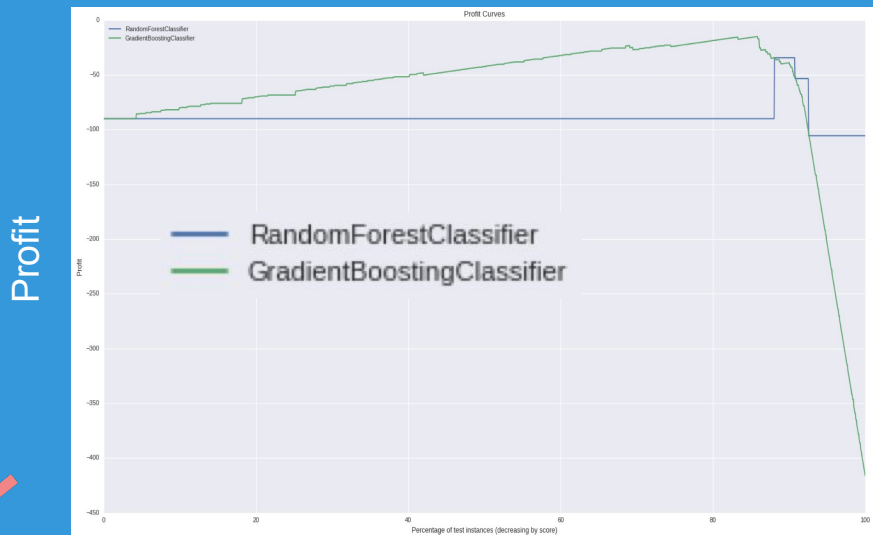
- Exclamation points
- Caps proportion

Profit curve for our models

The cost of false positives was chosen as the median the possible losses in fraud cases = 4225. For example, an fraudulent event with 1000 tickets with $30/each represents a loss of $3000.

Profit

	Was Fraud	Was Not-Fraud
Predicted Fraud	0	- $100
Predicted No-Fraud	- $4225	0

The Fraudly Model

In order to predict fraud, we ran several models and decided to use **Gradient Boosting Classifier** (GBC) model with 22 features. The most important ones are shown here:

1) Previous activity of the client
2) Has previous payouts
3) Number of tickets sold for the event
4) Sale duration
5) Length of the description

Welcome to Fraudly

At Fraudly, we care about our clients. We know that detecting fraudulent activity is important to your business and we are dedicated to providing you the best tools on the market.

Our Team Consists Of

Joanna
Demetria
Barack
Gina
Jarrod

Welcome to Fraudly

Screen-shots

Welcome to Fraudly

Event id number 1275909 is not fraud

Welcome to Fraudly

Event id number 8092379 is fraud

Fraudly v.2.1

Fraudly 2.1 will have the ability to offer batched events for ease of identification.

It will also include the top reasons why we classify each event as fraud.
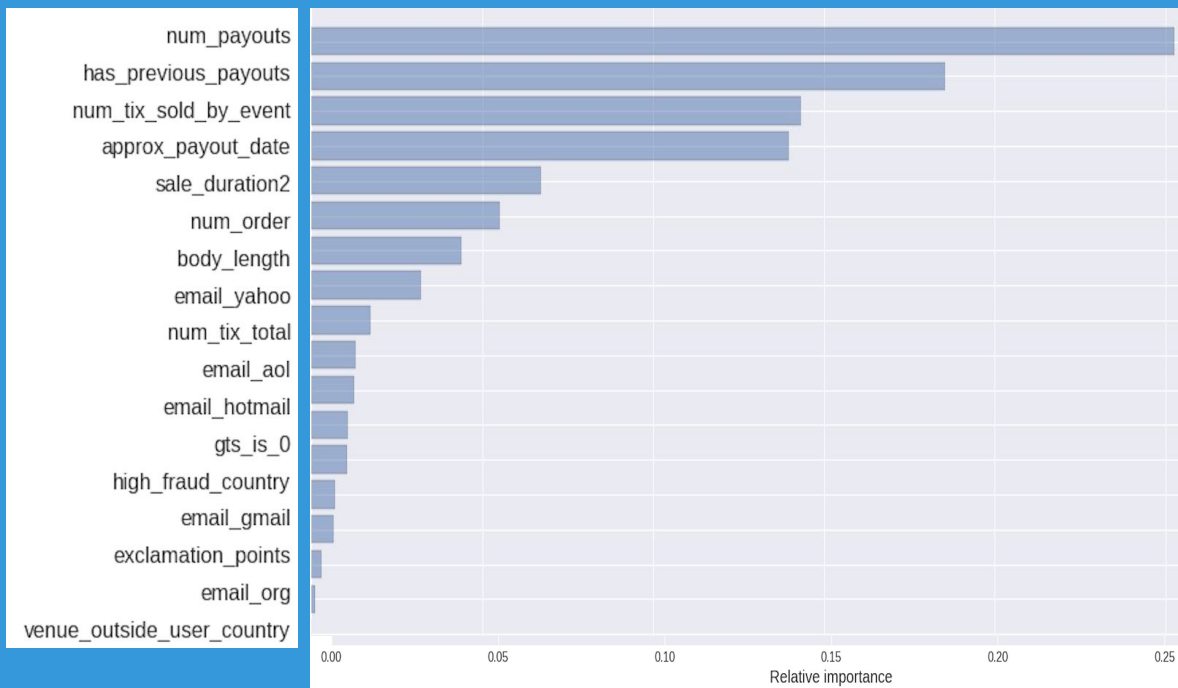
Add topic features.

The Fraudly Model - Important Features

Gradient Boosting Classifier model with 21 features. The most important ones are shown here:

Confusion matrix for test data

Recall = 0.929

Precision = 0.898

f1_score = 0.9135

Accuracy = 0.984

TP = 354	FP = 27
FN = 40	TN = 3881

App

Created a MongoDb database pulling data points from the website.

The app right now pulls a data from this database and makes a prediction.