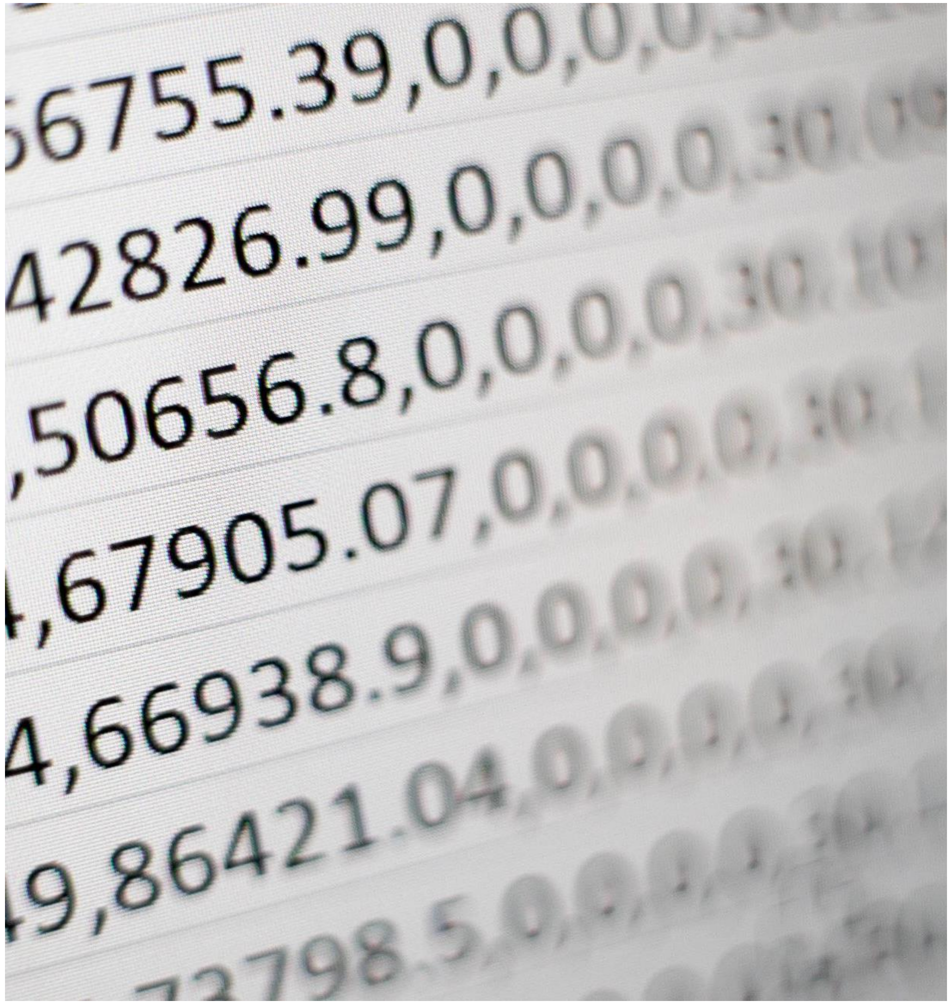
Taiwan Credit Card Defaults

Classification Analysis

Analysis Aim

1. Use data analysis to gain insights about card users who are likely to default
2. Determine what features are more important when determining default prediction
3. Apply multiple algorithms to best model default classification

The Data



- 30000 customers in the original dataset
- 29945 after cleaning data
- New features: monthly credit usage, successive payments & average payment, always paid, always delayed added.

Explorative Data Analysis

What does the data tell us ?

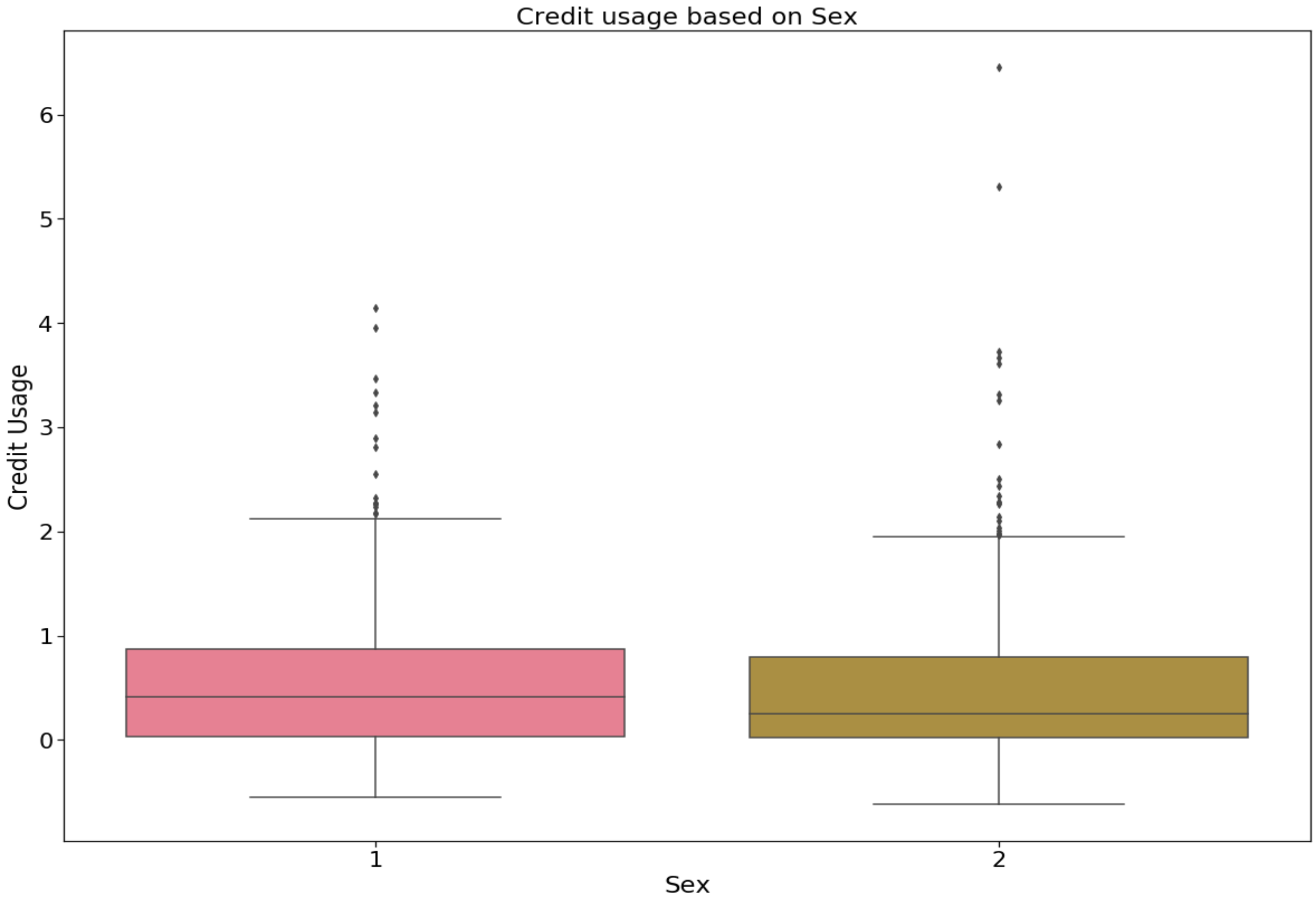
Class Imbalance

78% pay on time

vs

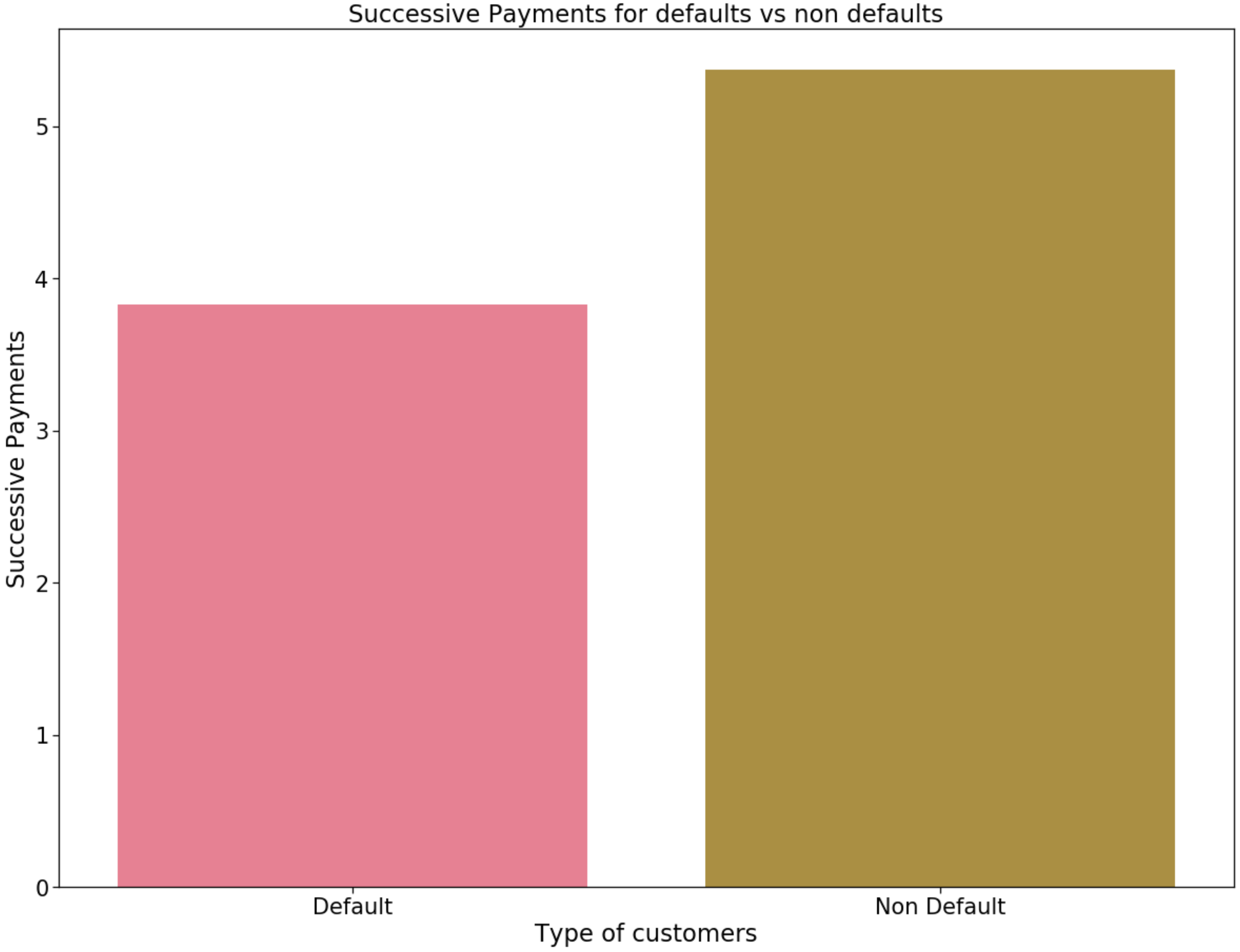
22% who default

Does Sex have anything to do with it?

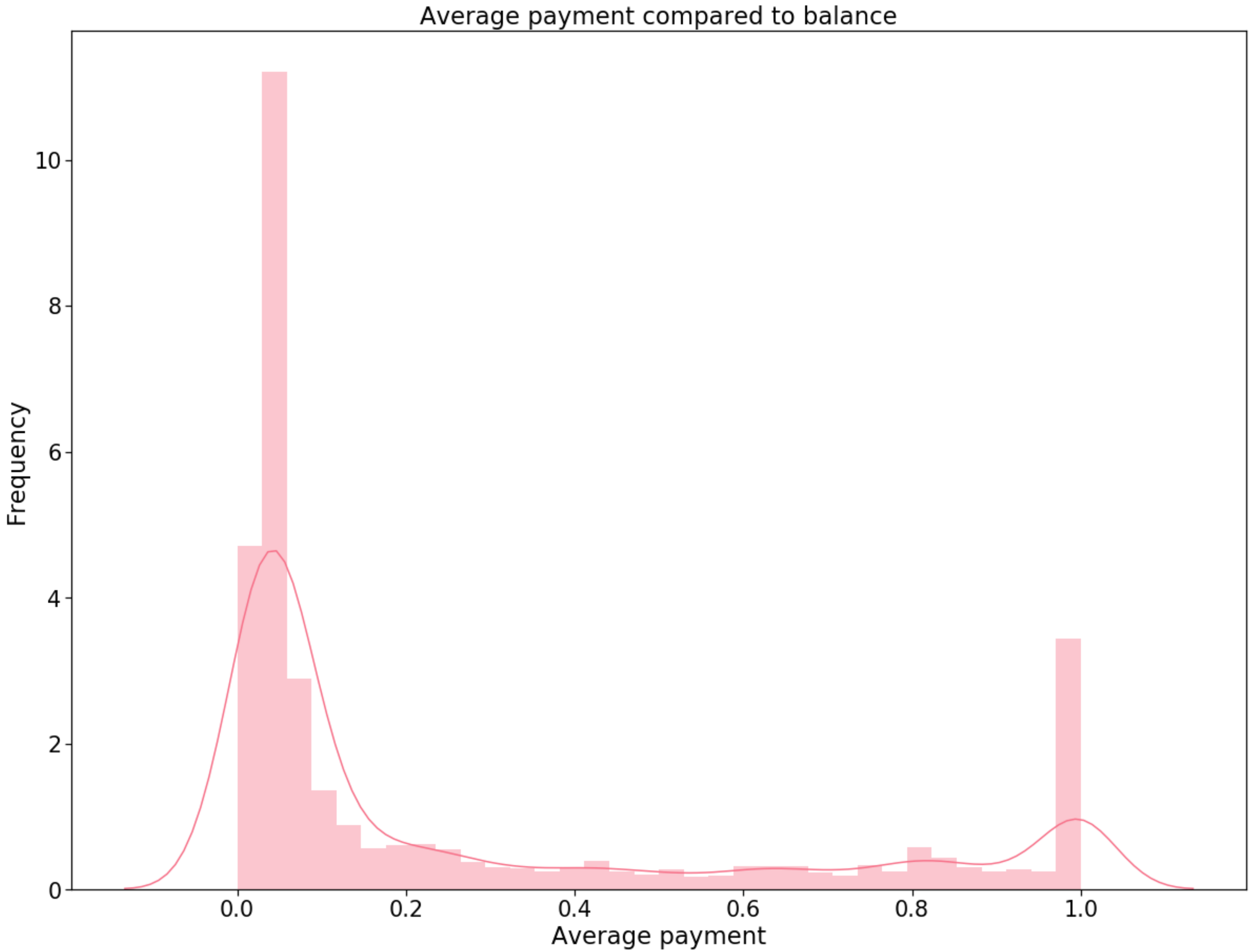


Does consistency matter?

Are people who make more successive payments less likely to default?



How much do people pay towards to card?



Feature importances

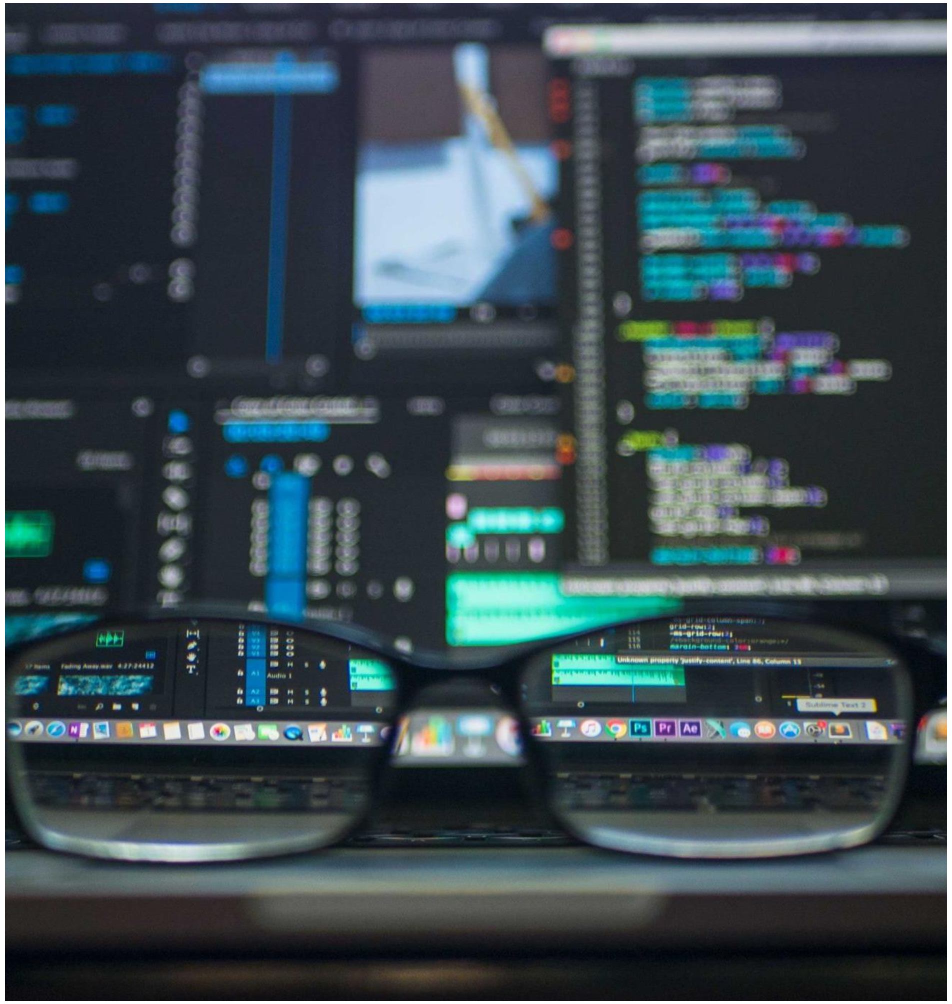
What features have been most influential for classification?

- Successive payments - consistency matters
- Average payments
- Monthly credit usage
- Age

Best model

1. XMboost - weight adjusted
2. 88% recall for default prediction, 76% for non default
3. F1 score of 64% for default, 85% for non default
4. Classification accuracy heavily affected by data imbalance

Future Work



1. Get more than 6 months of data - A multi year dataset with monthly resolution could reveal insights into how different seasons affect the ability to make a payment.

2. A variety of datasets that covers different parts of the world would be a great place to identify new patterns.

3. Incorporate economic datasets for the countries.

4. Apply more advanced models such as neural networks.

5. Deploy model for potential hosts to use


THANK
YOU