

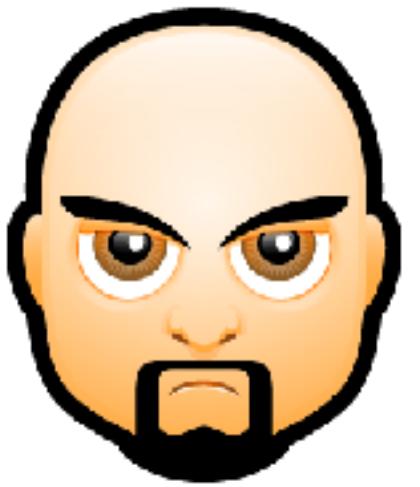
# 初探消費者信用風險大數據

信用評等、詐欺偵測、金融科技 - Roger Kuo

# Outline

- 信用風險介紹
- 台灣信評生態
- 美國信評生態
- FICO介紹
- 大數據風險的Fintech公司
- Kaggle上的風險分析題目
- 詐欺偵測與風險分析大數據架構

**怎麼定義信用風險？**



指交易對手未能履行約定

契約中的義務而造成經濟損失的風險

- Wikipedia



**那什麼是信用評等系統？**

指根據借款方的信用相關紀錄與狀態給與分數，用來決定是否核准新貸款或信用額度的統計與評估系統

# **信用評等方法基本類型**

小

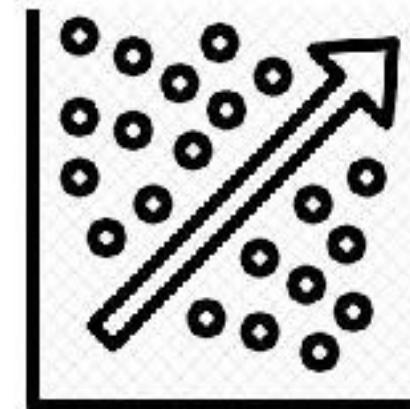


# 專家判斷法

人工檢視

低

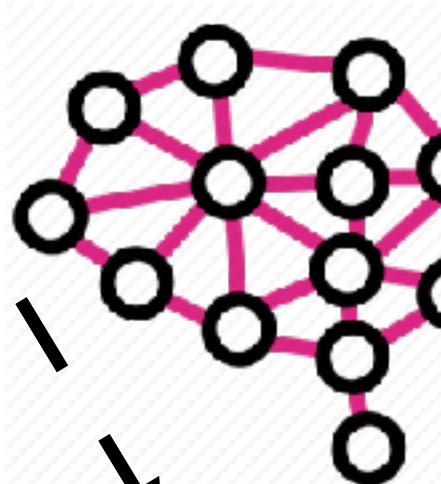
觸及顧客規模



# 統計分析法

20 ~ 100變數 + 回歸

數據多元程度



大

# 大數據徵信

1萬變數up + 多個機器學習模型

高

**信用評等 vs 詐欺偵測**

# 信用評等



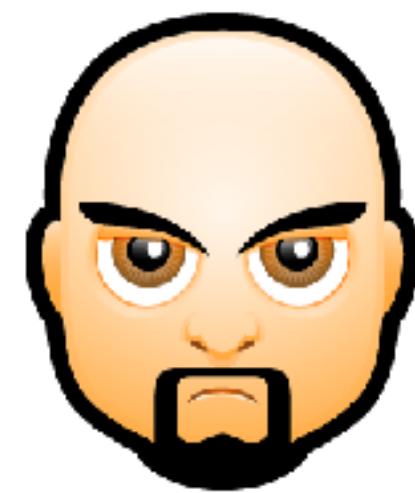
這人會違約嗎？

# 詐欺偵測

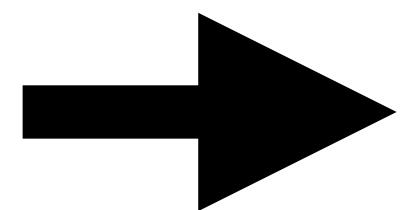


異常行為出現！

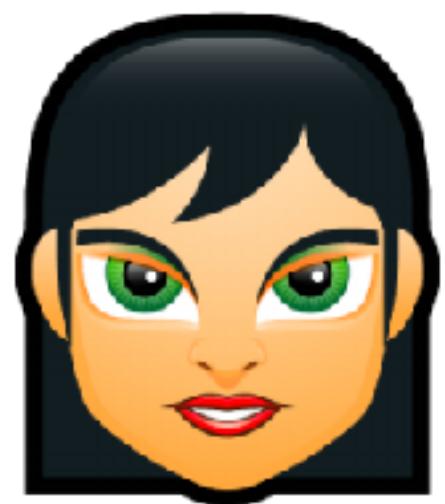
**先來了解台灣信評的生態吧**



上繳



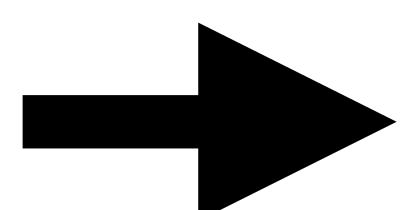
消費者個人信用評分(J10)



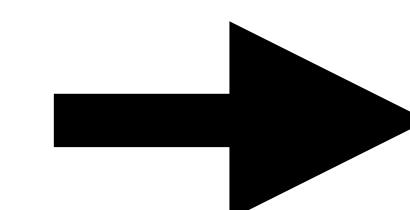
他行互動



上繳



上繳



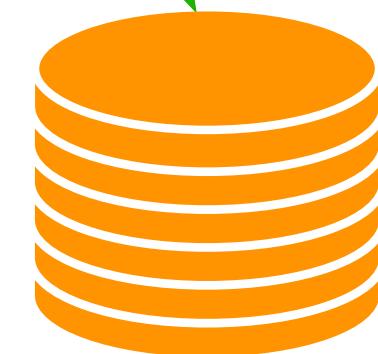
統計報告



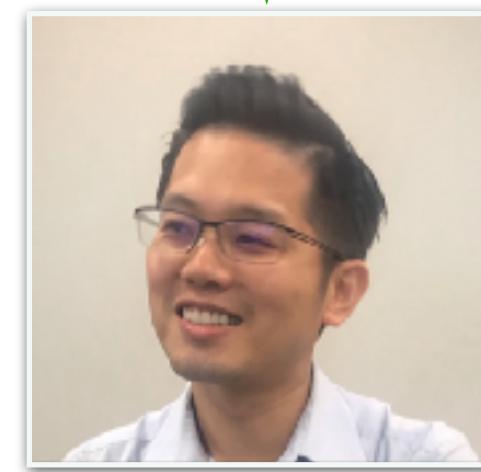
財團法人金融聯合徵信中心



本行數據



行內分析大師



調聯徵



JCIC

## 財團法人金融聯合徵信中心

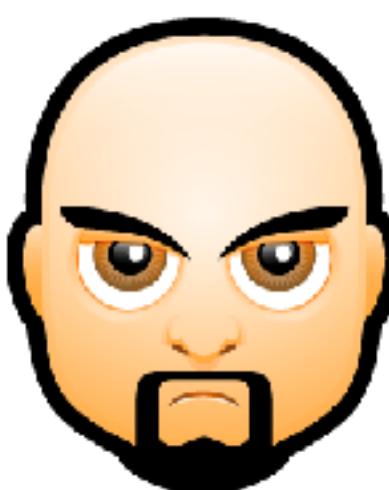
買FICO系統



750

針對本行客戶  
在不同情境評分  
進而決策

↑ 本行互動



壞



壞



好



好

報表編號: PLR1-1	財團法人金融聯合徵信中心 當事人綜合信用報告	103/02/24
		Page. 2 / 2
謹慎使用信用報告 保障良好信用		
102/09/	全額繳清 無遲延	無 0
	全額繳清 無遲延	無
	全額繳清 無遲延	無
	全額繳清 無遲延	無
	全額繳清 無遲延	無 0
	全額繳清 無遲延	無 1707
102/03/	全額繳清 無遲延	無 3414
【信用卡債權轉讓及清債資訊】查資料庫中無原債權銀行將一台灣尚未在揭露期限內之債權轉讓與資產管理公司或第三人及清債資訊		
【被查詢紀錄】最近三個月內(不含查詢當日)無金融機構向本中心查詢 台灣信用資料		
【當事人查詢紀錄】查資料庫中無一台灣最近三個月(不含查詢當日)向本中心申請查詢紀錄		
【附加訊息】查資料庫中無一台灣加註訊息資訊		
※當事人之中英文姓名、出生日期或戶籍地址等若需更改，請來函並檢具戶政所相關證明文件，掛號郵寄至本中心辦理。		
※上述信用資料若有發現金融機構報送的資料有誤，您除可以向原來提報的金融機構反映，請其向本中心通知改正外；也可以用書面的方式檢具身分證明文件（如有證據請一併檢附），直接向本中心反映，本中心會主動協助將您的資料與報送的金融機構聯繫，如果查明屬實，便會把您的資料更正正確。		
※信用卡資訊與信用卡繳款資訊之更新，乃收件隨即處理上線，各發卡機構報送頻率原則上每週一次，惟各發卡機構報送週期不一，故信用卡類資訊無單一資料截止日期。		
※本中心信用報告所示信用資料僅供申請人參考，不能等同或證明資料當事人於全體金融機構實際存在之所有金融負責（含保證）情形。申請人為任何決定時，請勿以本信用報告作為唯一依據。		
※本中心所提供之信用報告之各項資料，為報經金融監督管理委員會核備之資料揭露期限內之信用資訊。 << 電子密章 : >>		
*** 報表結束 ***		

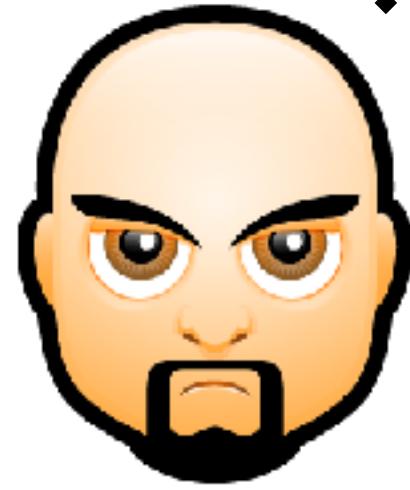


每年每人免費一次  
申請一份聯徵報告



## 財團法人金融聯合徵信中心

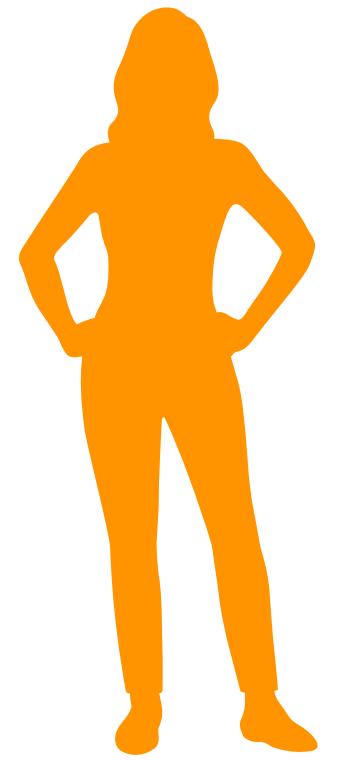
自己查



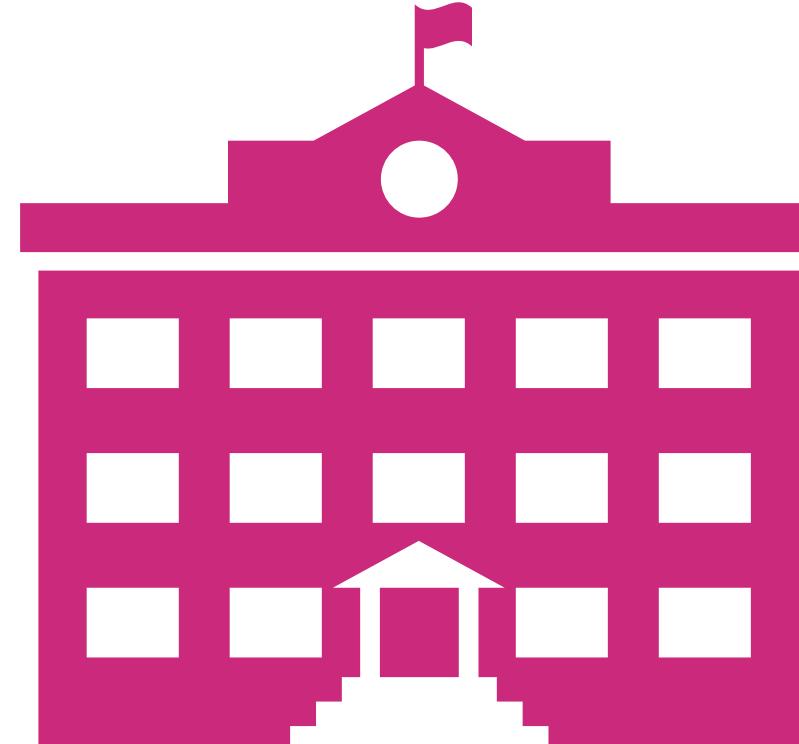
覺得簡陋...

# **再來介紹美國信評生態**

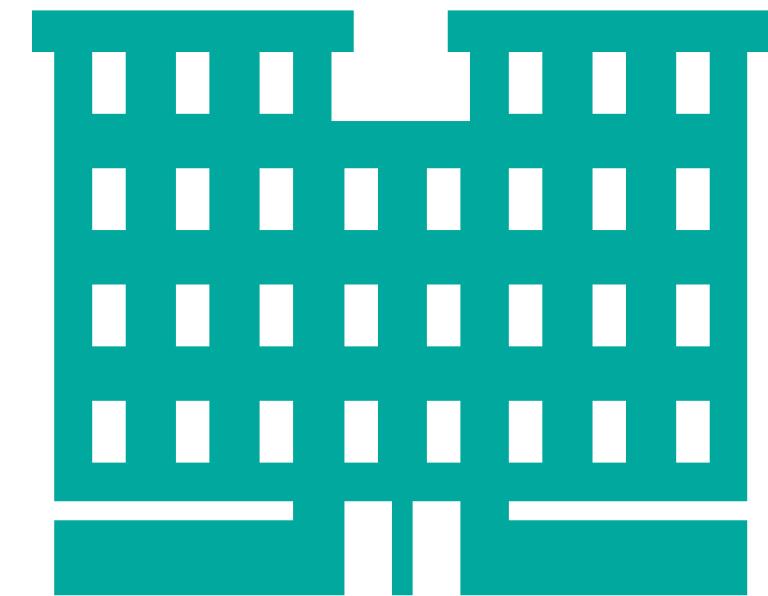
**- 最成熟**



消費者



貸方

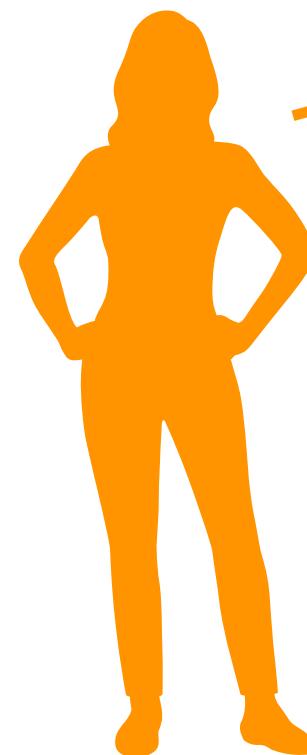


信用機構

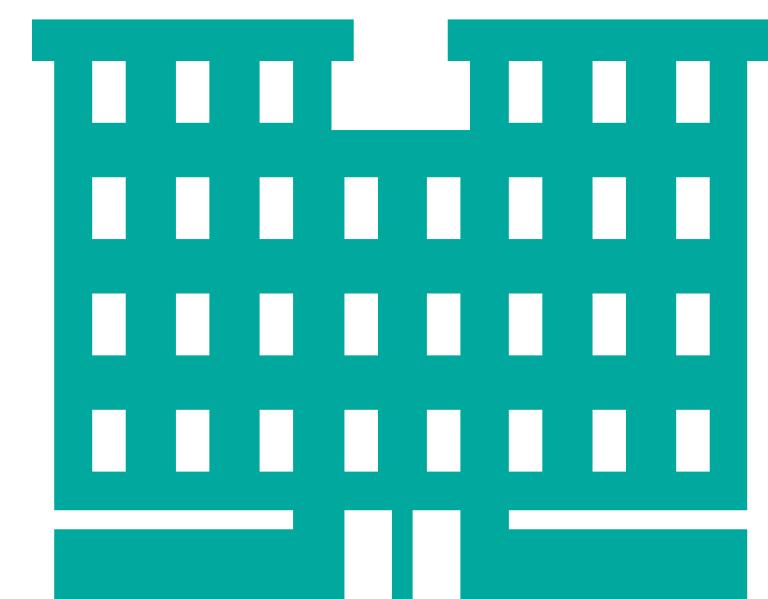
FICO<sup>TM</sup>

信用模型公司

借錢、申請信用卡  
付帳單

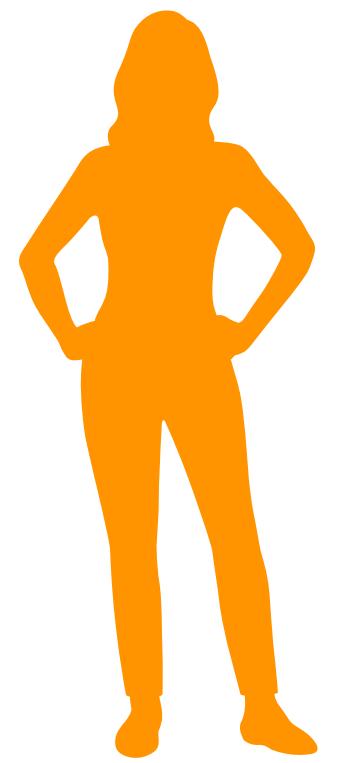


消費者



FICO<sup>TM</sup>

信用模型公司



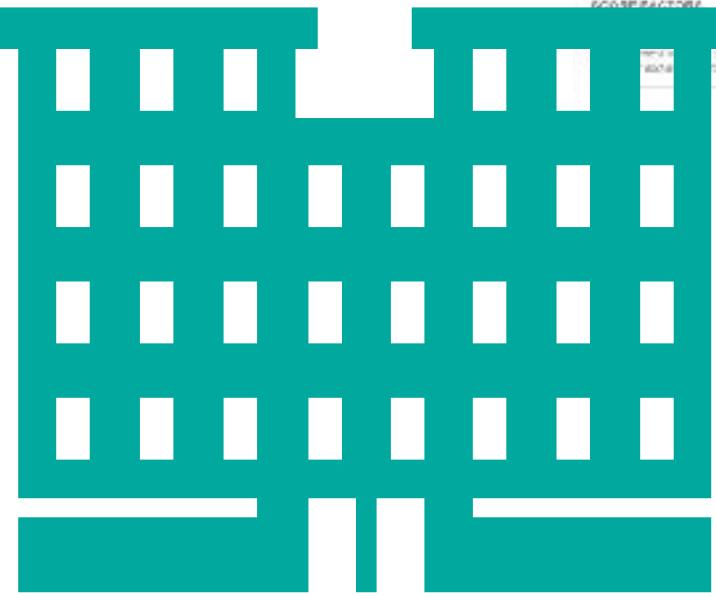
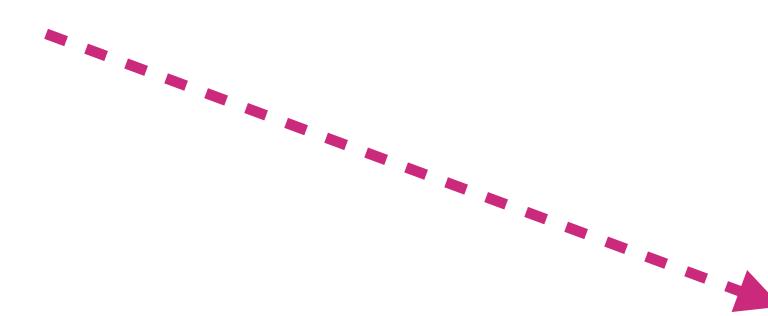
消費者



貸方



申報你的信用活動與繳款紀錄  
從機構查你的FICO分數



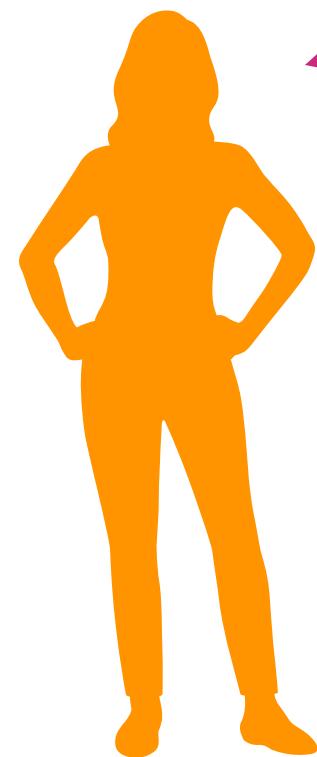
信用機構



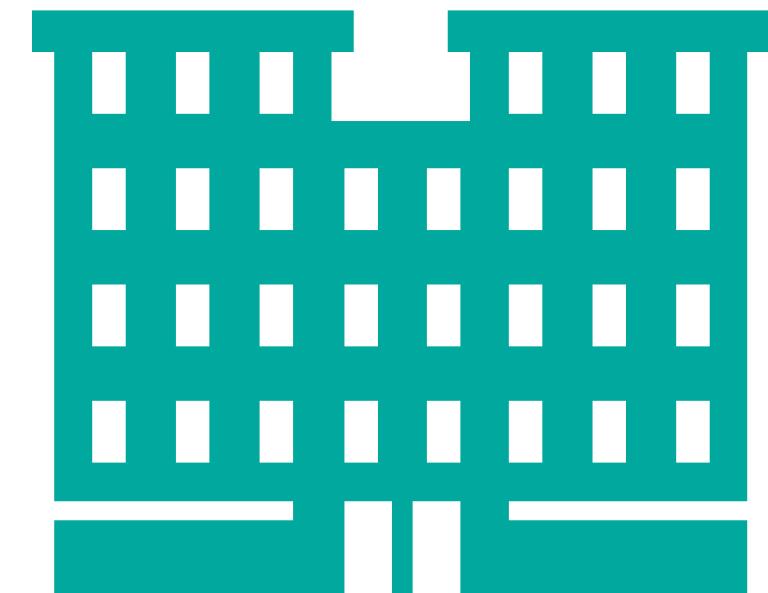
FICO<sup>TM</sup>

信用模型公司

# 核貸、發卡 回覆貸款決策原因



消費者

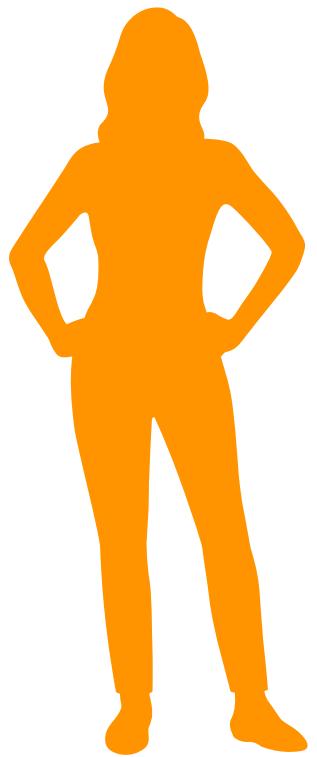


FICO<sup>TM</sup>

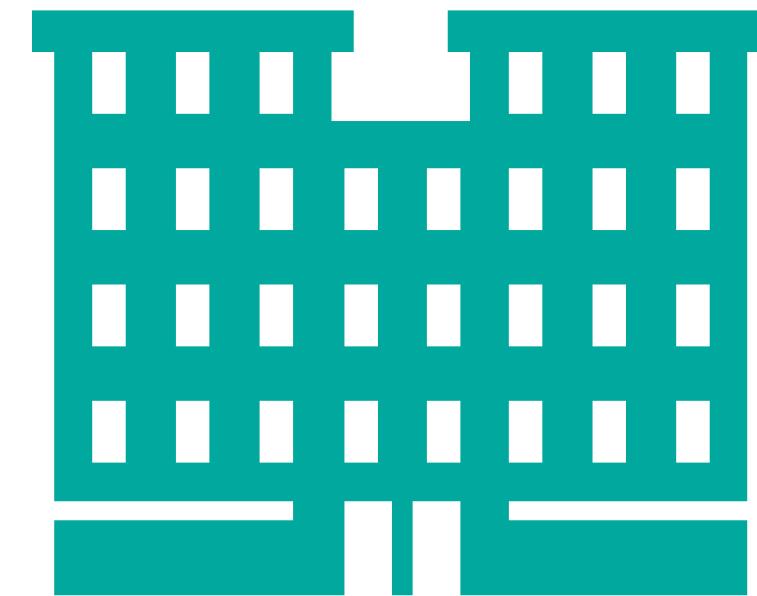
信用模型公司



貸方



消費者

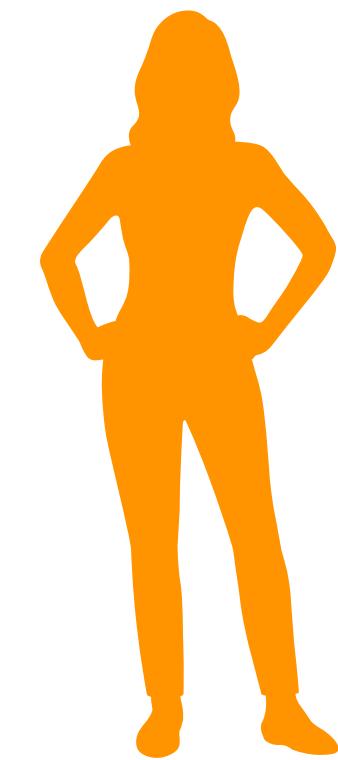


信用機構

**FICO**  
TM

信用模型公司

研發並販售評分模型  
給機構使用



消費者

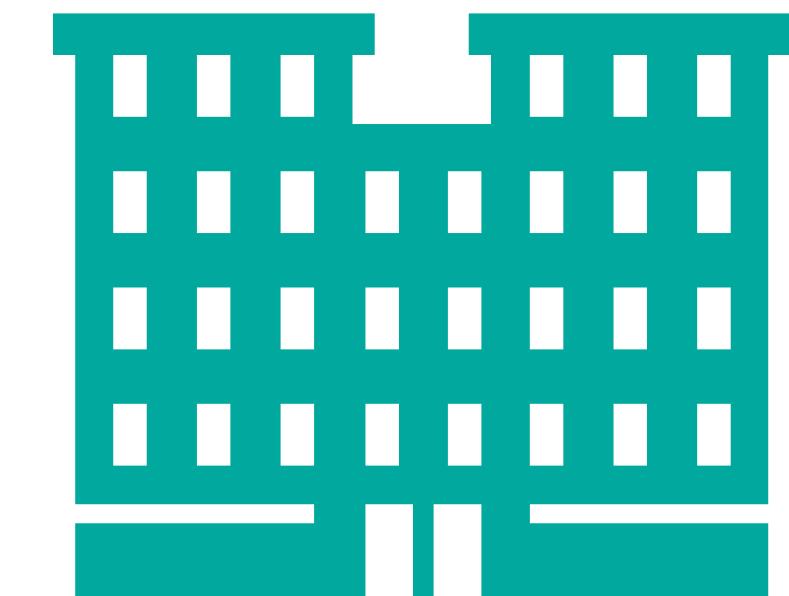


貸方

**FICO**  
TM

信用模型公司

根據貸方資訊  
更新消費者信用報告



信用機構

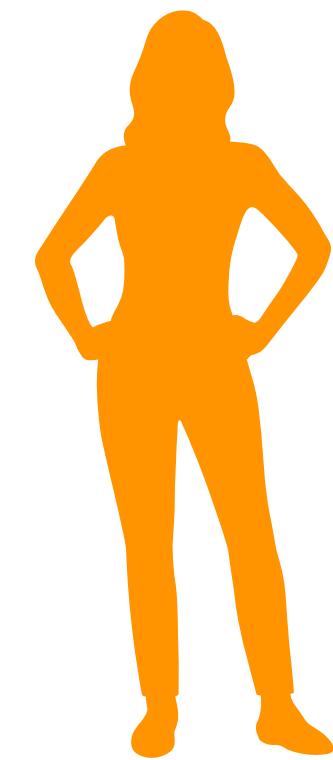
**EQUIFAX®**

 Experian<sup>SM</sup>

 TransUnion<sup>®</sup>

利用FICO模型產出  
消費者信用分數





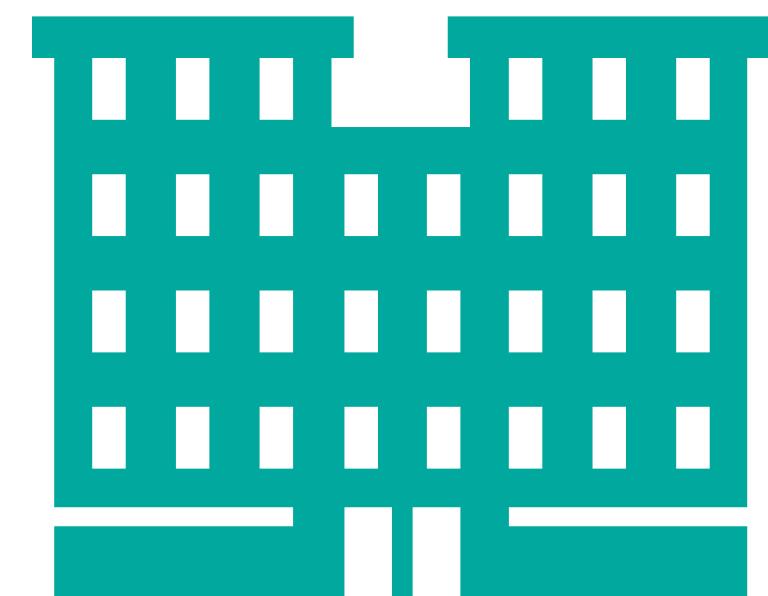
消費者



貸方



回報貸方  
消費者信用分



信用機構

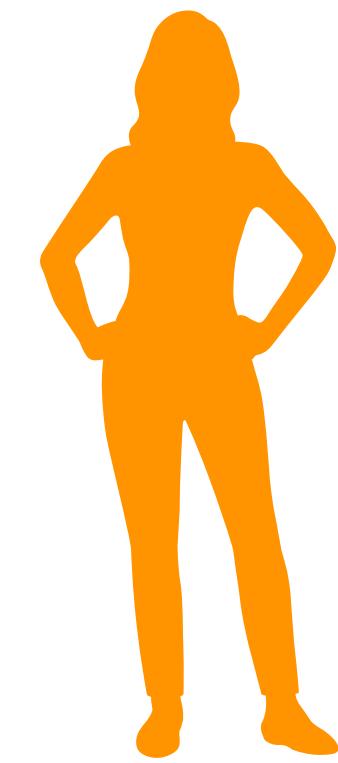
EQUIFAX®

Experian™

TransUnion®

FICO™

信用模型公司

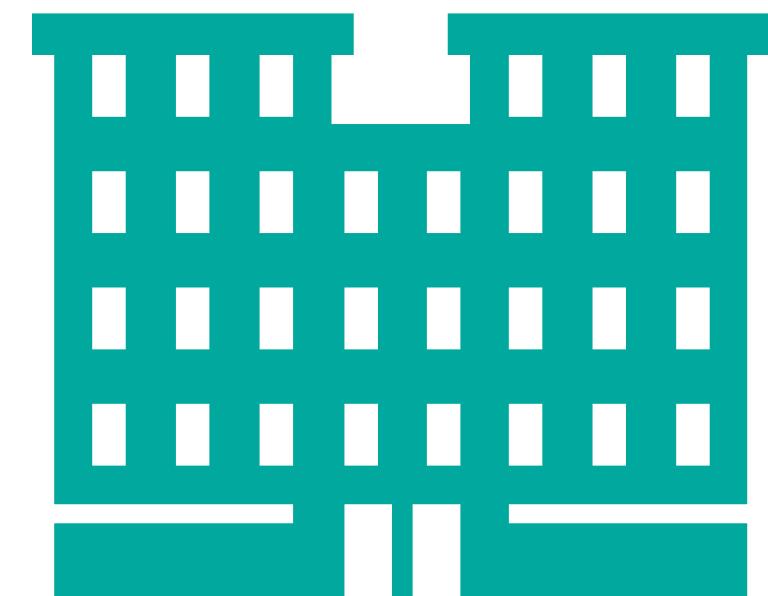


消費者



貸方

提供每年每人一份免費報告



信用機構

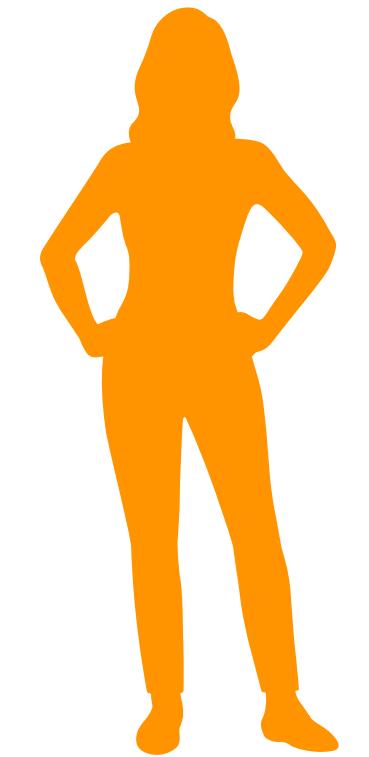


信用模型公司

EQUIFAX®

Experian™

TransUnion.



消費者

# Credit Report

TransUnion

Experian

Equifax

Your Credit Score:

717    706    715

## Personal Information

NAME: JOE CREDIT  
DATE OF BIRTH: 01/01/1970  
ADDRESS: 1234 CREDIT LANE, FINANCE CITY, AA 99999  
EMPLOYER: ANY COMPANY

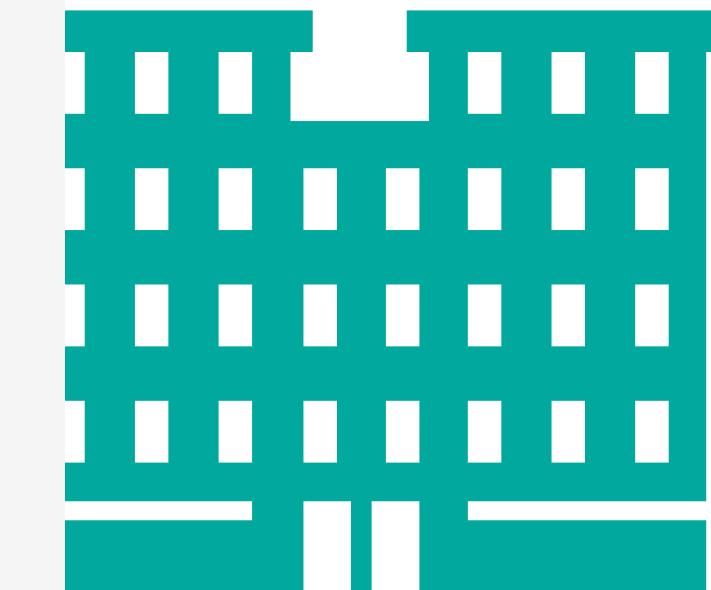
## Summary

	TransUnion	Experian	Equifax
TOTAL ACCOUNTS:	20	21	22
OPEN ACCOUNTS:	7	8	7
CLOSED ACCOUNTS:	13	15	15
DELINQUENT:	0	0	0
DEROGATORY:	0	0	0
BALANCES:	\$1987.00	\$2145.00	\$2145.00
PAYMENTS:	\$321.00	\$433.00	\$433.00
PUBLIC RECORDS:	0	0	0
INQUIRIES (2 YEARS):	1	4	3

## Account History

At-a-glance viewing of your payment history

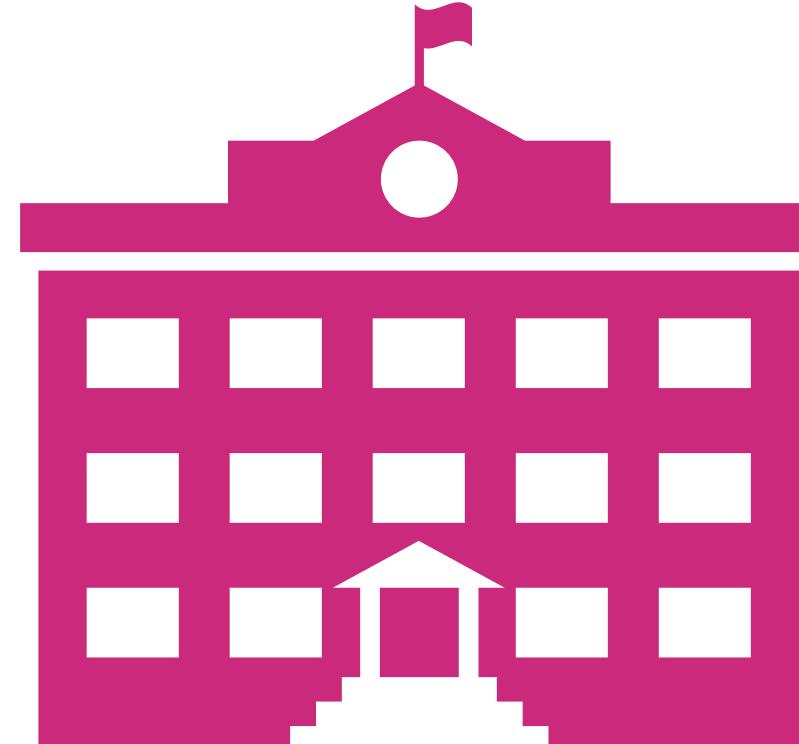
Unknown	OK	30	60	90	120	150	PP	RF	CO
Current		30 days Late	60 days Late	90 days Late	120 days Late	150+ days Late	Payment Plan	Repossession Foreclosure	Collection Chargeoff



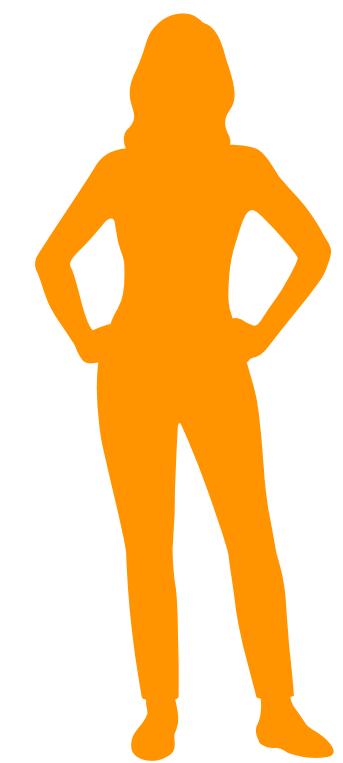
信用機構



信用機構



貸方



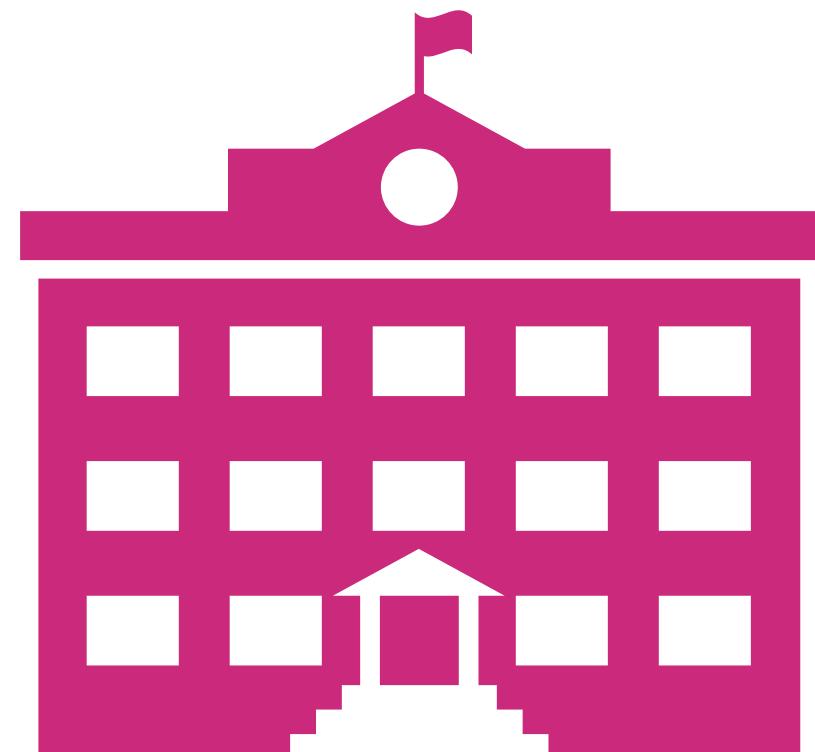
消費者

教育大家  
了解FICO

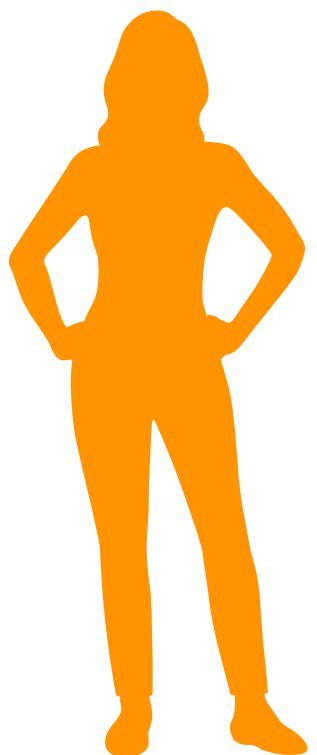
FICO<sup>TM</sup>

信用模型公司





貸方



消費者

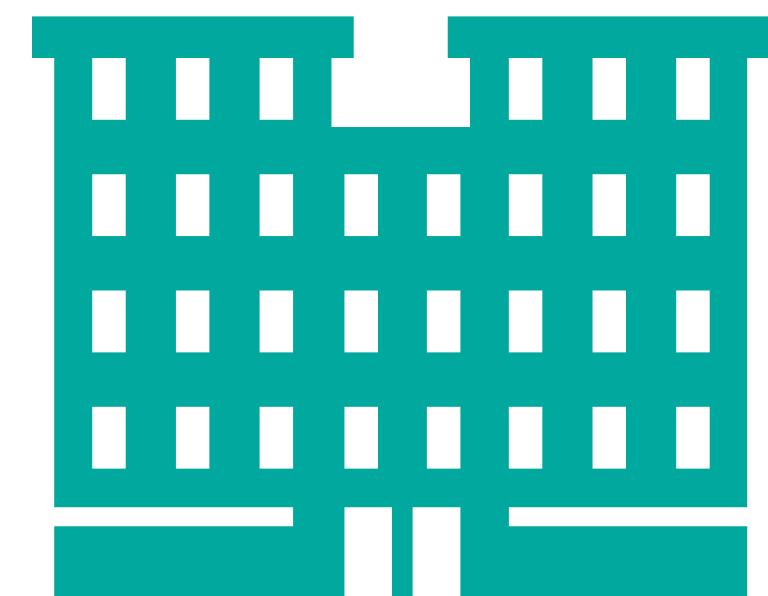
提供付費分數查詢平台  
- myFICO.com



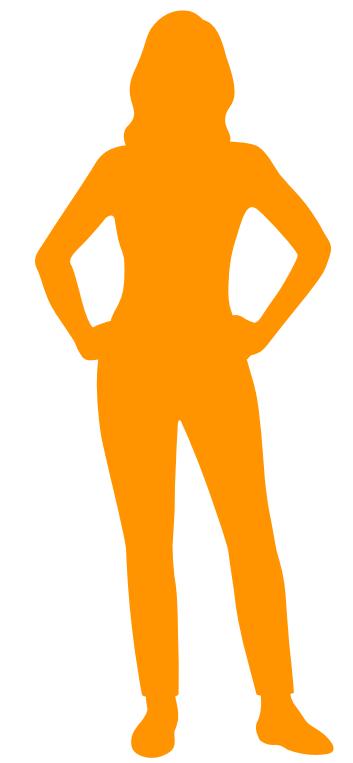
FICO SCORE	INTEREST RATE	MONTHLY PAYMENT
760 - 850	3.75%	\$498
700 - 759	4.0%	\$517
680 - 699	4.25%	\$532
660 - 679	4.4%	\$551
640 - 659	4.82%	\$590
620 - 639	5.38%	\$640

**FICO**  
TM

信用模型公司



信用機構

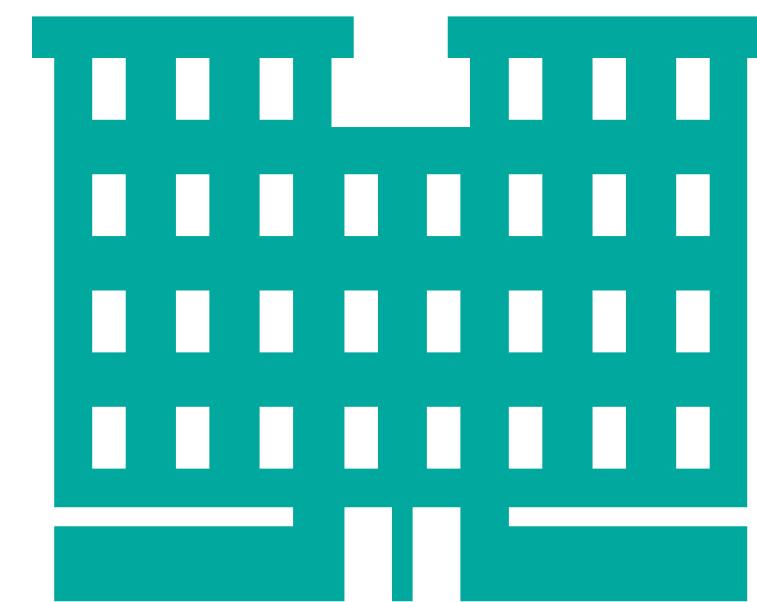


消費者



貸方

四方利用數據和信用分  
串起美國的信用生態



信用機構

FICO<sup>TM</sup>

信用模型公司

# FICO怎麼這麼猛啊？！

- 用信評模型征服世界



Used in over **90%** in US



1956

創辦人們



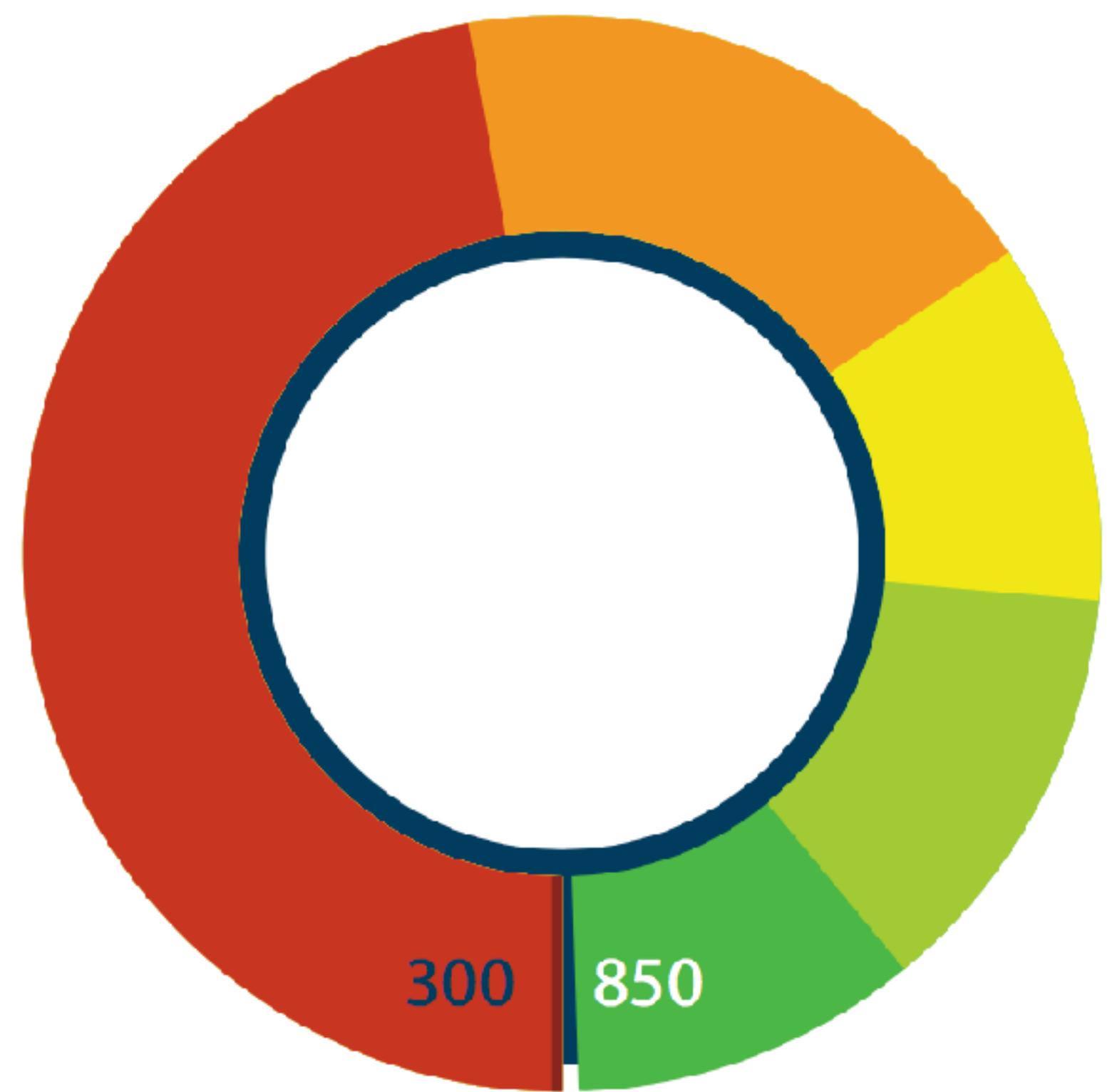
Engineer Bill Fair and mathematician Earl Isaac founded FICO – with an initial investment of \$400 each – on the principle that data, used intelligently, can improve business decisions.



業務範圍

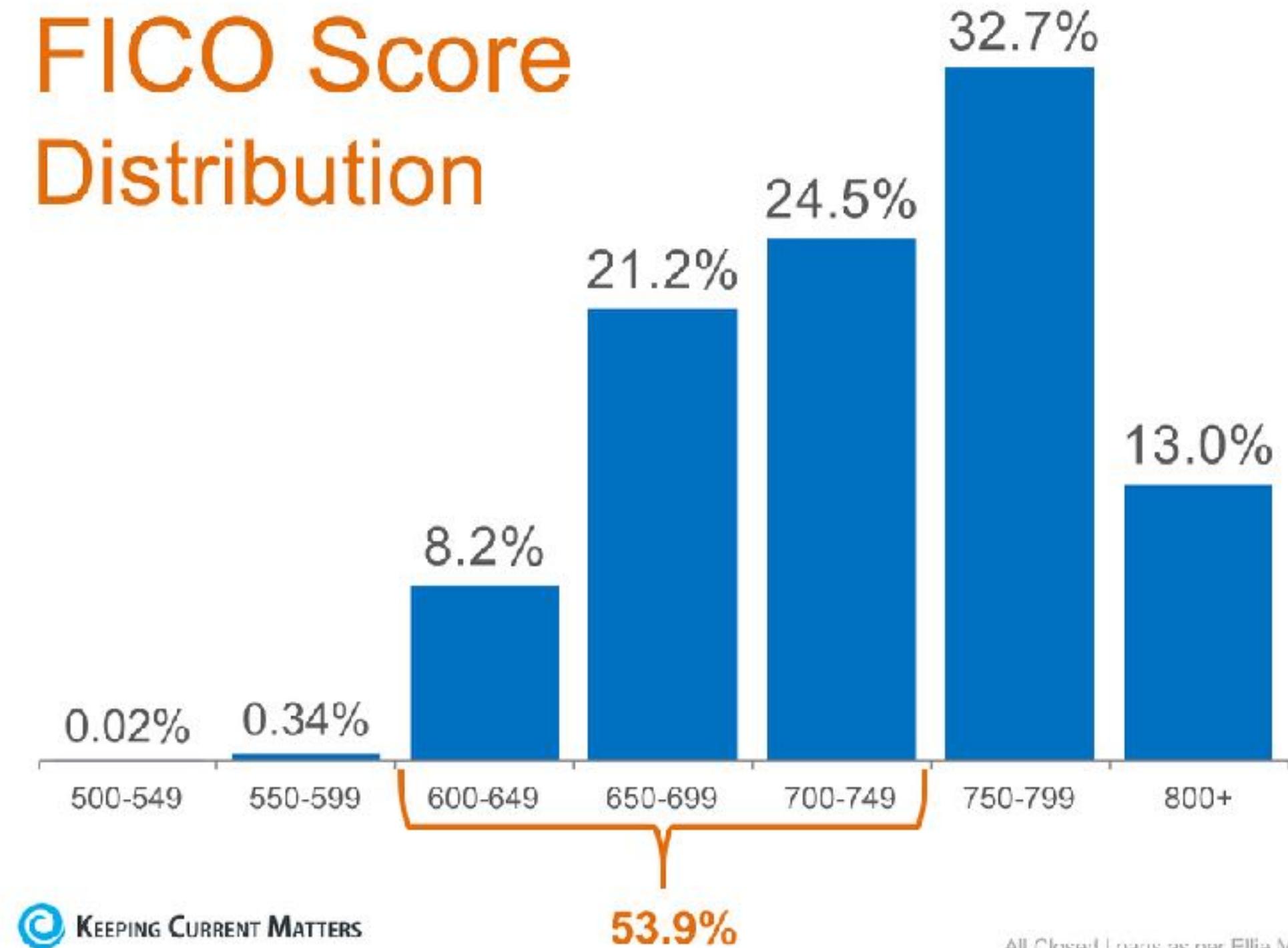


分數 系統



Poor	< 580
Fair	580 - 669
Good	670 - 739
Very Good	740 - 799
Exceptional	800+

## FICO Score Distribution



All Closed Loans as per Ellie Mae

2016 US

區間與意義



分數成分

35% 付款與違約歷史

繳款違約紀錄

各類帳戶繳費狀況

延遲繳款繳款紀錄

重點：你有沒有準時繳款啊

30% 債務負擔

各帳戶欠多少錢

信用額度用多少了

分期還有多少要繳

重點：你還欠多少錢啊

15% 信用歷史長度

最久的帳戶開多久了

特定種類帳戶年齡

平均帳戶年齡

重點：你是金融老江湖嗎

10% 新信用申請

最近有沒有開新帳戶

短期內有沒有太多信用相關行為

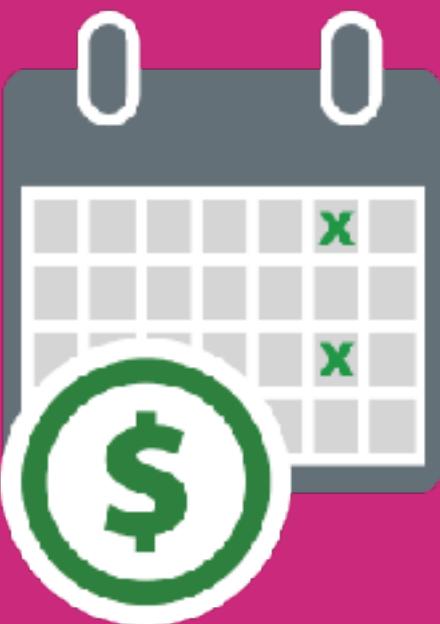
重點：你近一年有沒有一直借錢

# 10% 信用種類

持有帳戶類型數



信用卡



分期繳款



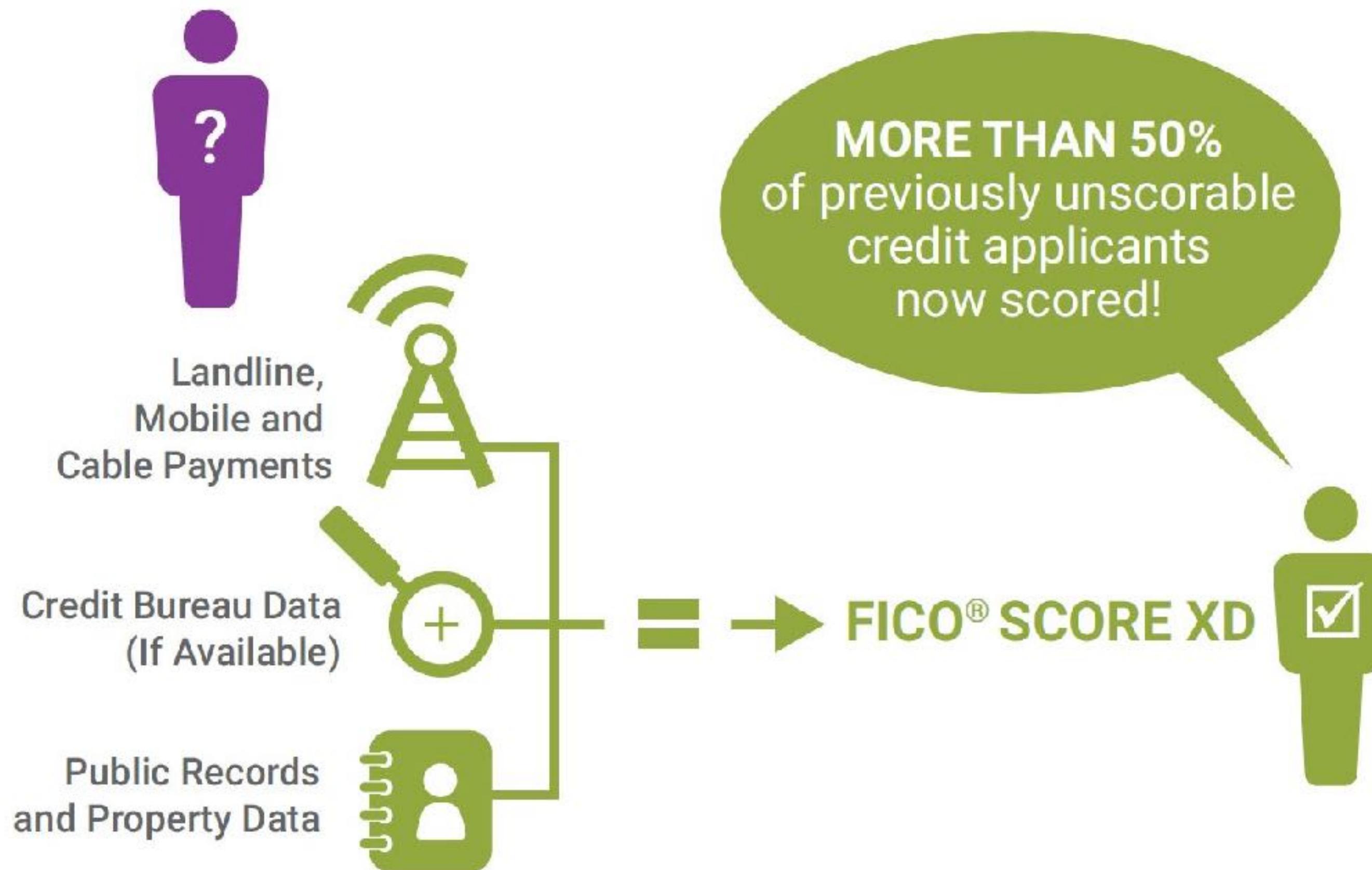
個人信用貸款



房貸抵押

重點：帳戶類型是不是太多

# FICO Score XD



新資料源、新方法

# Fintech怎麼做信用風險

# 案例一：結合生活的大數據評分 強國人最愛



芝麻信用  
ZHIMA CREDIT

芝麻分

信用生活

異業合作

企業信用

芝麻分

範圍

350

不好

# 芝麻分是什么？

芝麻分是一个代表您信用程度的数字  
参考国际通用的信用分体系



分数越大，代表信用程度越好



950

超好

# 阿里集團



芝麻分  
數據源



## 生活合作



## 政府合作



## 金融機構合作

# 異業合作

# 芝麻分 成分



15%

# 身份特徵

實名認證

身份訊息



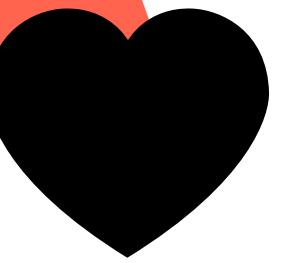
25%

# 行為偏好

支付寶消費

消費偏好

繳費層次



5%

## 人脈關係

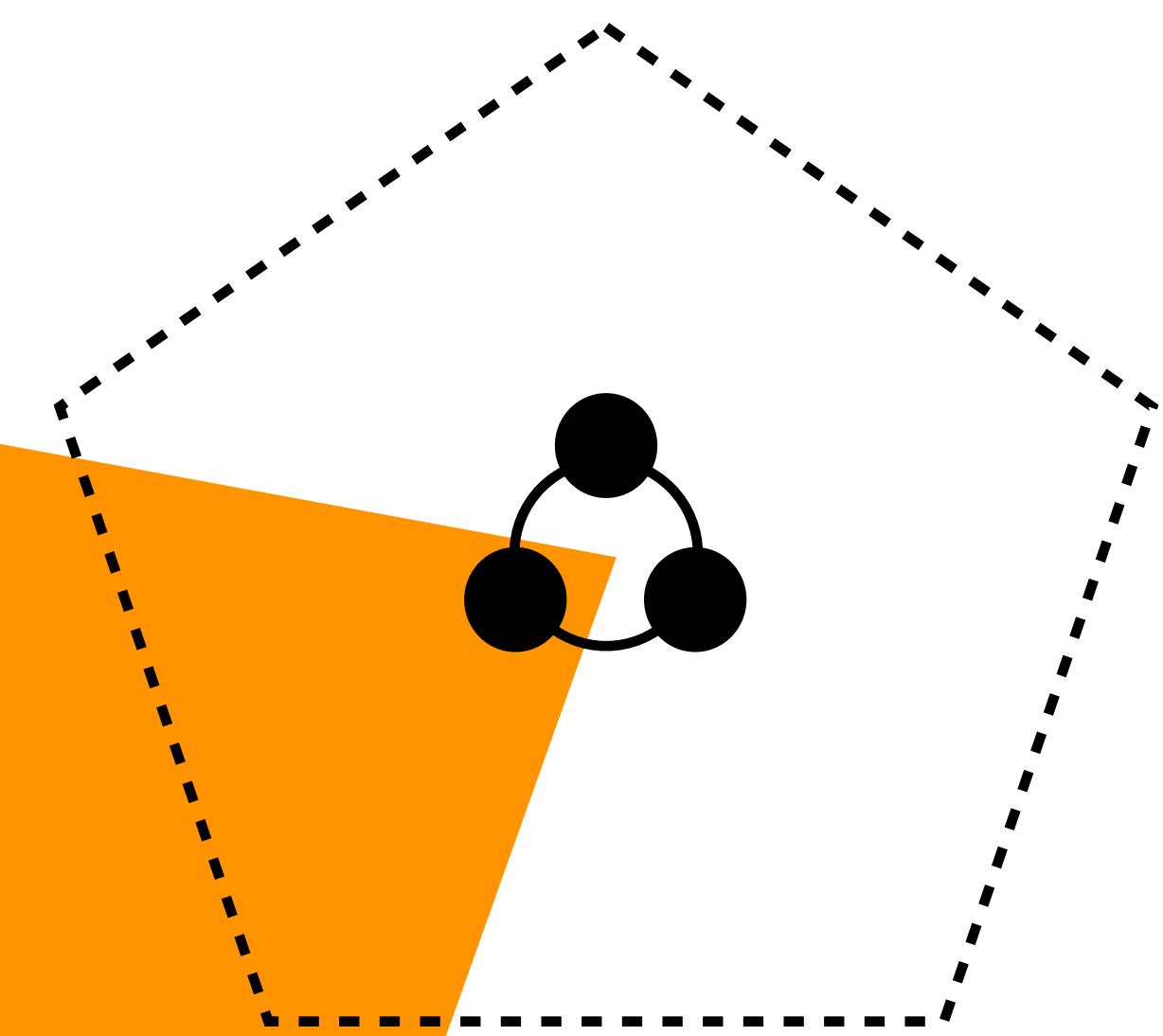
支付寶通訊錄

轉帳

關係圈

社交影響力

朋友圈信用水平



# 信用歷史

35%

信用卡還款歷史

水電繳費

罰單

微貸還款

徵信系統

支付寶帳戶歷史



20%

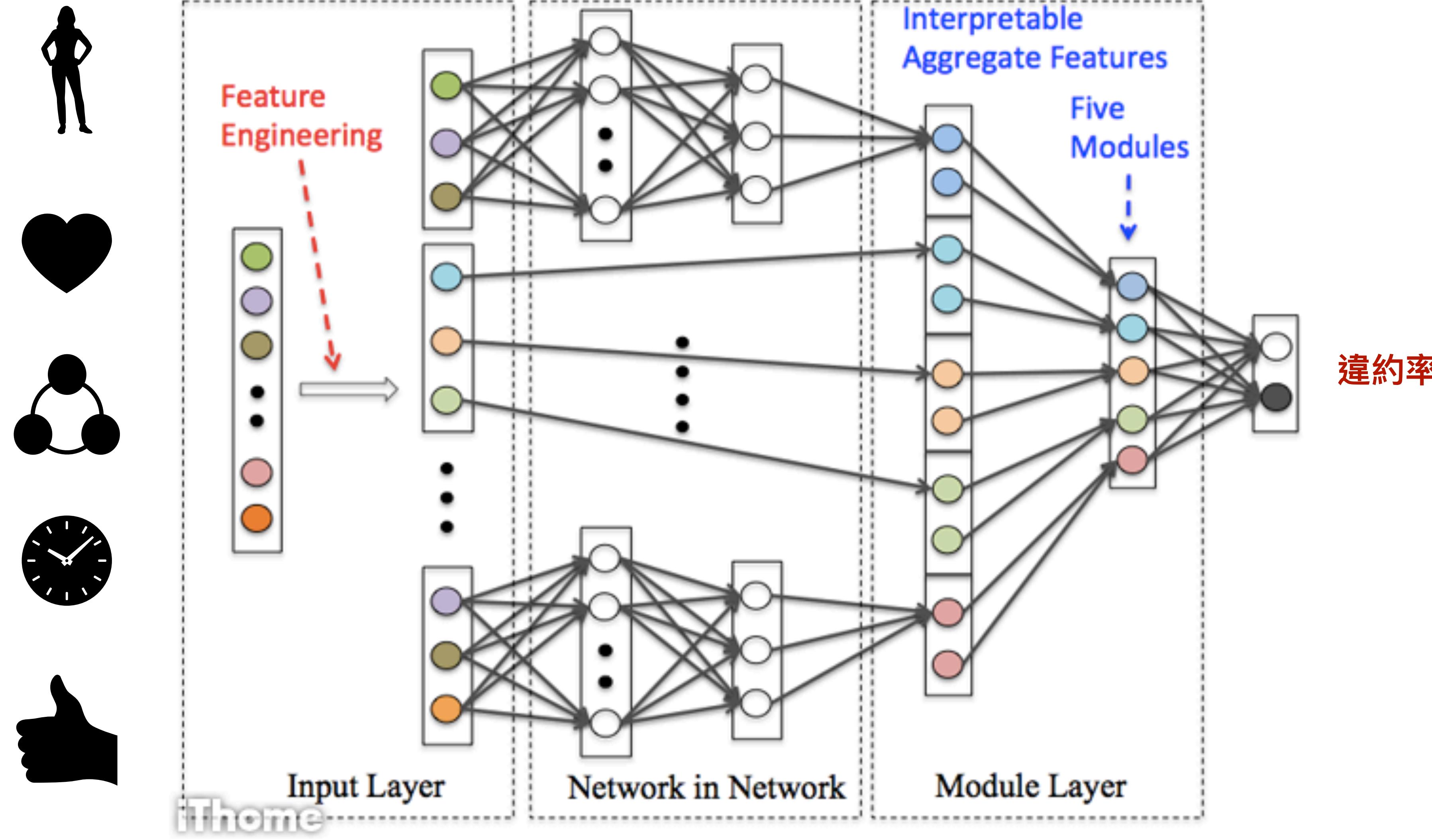
# 履約能力

支付寶餘額

餘額寶餘額

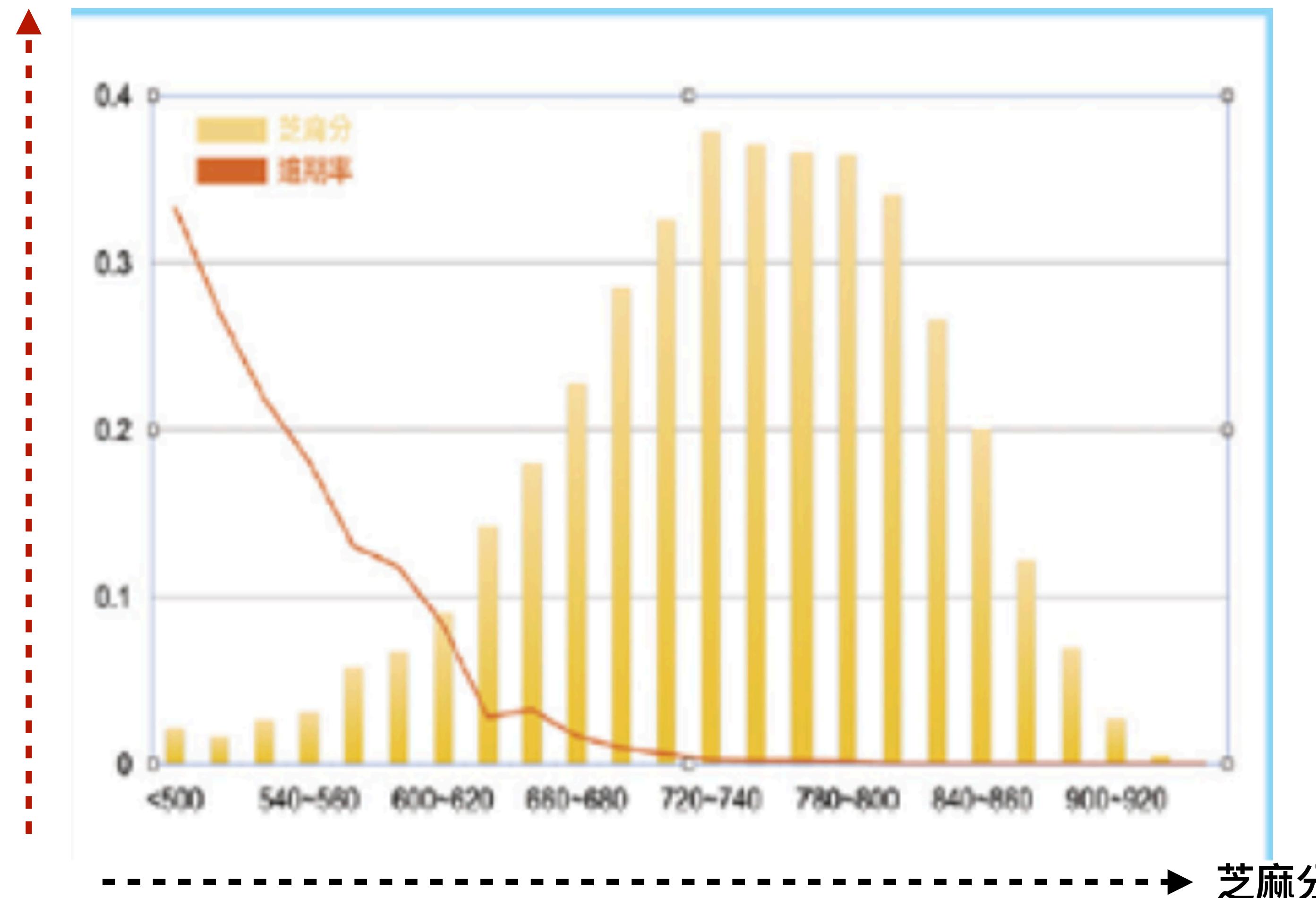
有車有房？





深度學習模型

違約率



分數成效

# 信用生活

## 免押出行



信用出行，免押金骑行  
芝麻信用让绿色出行如此简单

免押金骑行是芝麻信用联合公共自行车服务商为信用良好（芝

点击网站右上角下载手机版体验

点击网站右上角下载手机版体验



信用出行，免押金租车  
芝麻信用让轻松出行变为可能

免押金租车是芝麻信用联合多家汽车租赁公司（神州租车、一嗨租车等）为信用良好（芝麻分650分及以上）的用户在提供更为便捷的租车方式，无需排队刷卡，在线预定即可免押金租赁车辆，仅需要2-3分钟，即可轻松踏上旅程。

## 免押住宿



信用住宿，免押月付  
芝麻信用开创租房新模式



信用租房，芝麻信用联合多家租房平台（相寓、优客逸家等）合作，希望能为漂泊在外的年轻人带来全新租房体验。根据芝麻分的不同，租房平台会提供不同的租房方案，租房免押，房租月付，为年轻人减轻租房的押金压力。



信用相伴，轻松出游  
芝麻信用让旅行不再繁琐



芝麻信用联合多家酒店、民宿平台（飞猪旅行、小猪短租等）为信用良好（芝麻分600分及以上）的用户，推出先住后付的免押金服务，可以享受免押金、免排队、免查房的服务。高效快捷的入住方式让用户入住时间从原来3分钟下降到45秒，退房时间从2分钟下降到18秒，再疲劳的旅程都像“回家”一样简单舒心。

## 案例二：致力運用機器學習 改變美國信評生態



# 創辦人



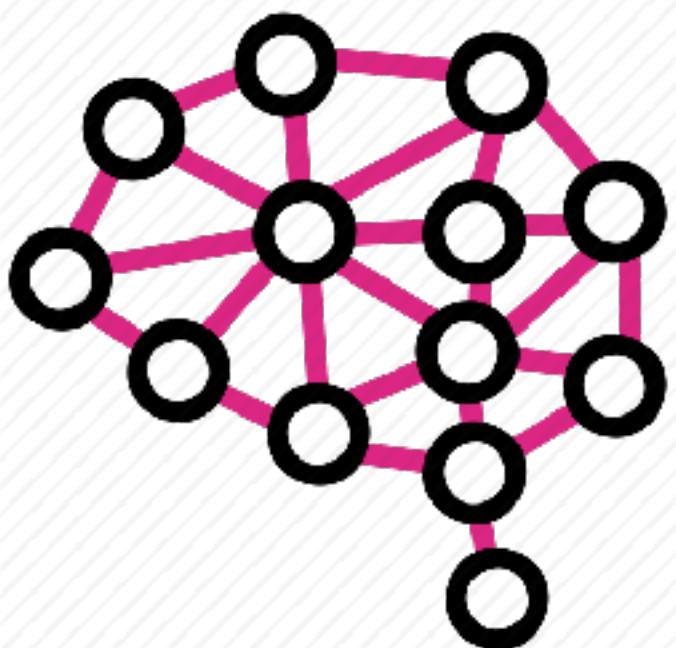
**Google**

**Douglas Merrill**  
Founder & CEO



**CapitalOne®**

**Shawn Budde**  
Former President



# 大數據徵信

7萬變數 + 10個機器學習模型

致力於給大眾創造更透明公正的信用評分

- 15%美國人沒有傳統信評

向代理商購買或交換  
銀行、法律、搬家數據

結構化+非結構化

網路數據

IP、瀏覽器版本、螢幕解析度

典當紀錄

借款人房租繳款

用戶激勵問卷  
水電帳單、手機帳單

社交網路訊息

用戶申請訊息  
閱讀申請說明書?  
填表習慣



月消費、壞賬、所用通信網絡其中有缺失  
→ 故意隱匿

填表喜歡全部用大寫字母  
→ 違約率更**高**

有趣分析

淨收入+地理因素  
→ 對還款產生更大的**預測能力**

月收入沒有經過驗證  
自填月收入7500美元→違約率是**最低**  
填寫7500美元以上 → 數字越大違約率就**更高**

月收入經過驗證的情況下  
收入越高→違約率越**低**

身份驗證詐欺

衡量還款能力

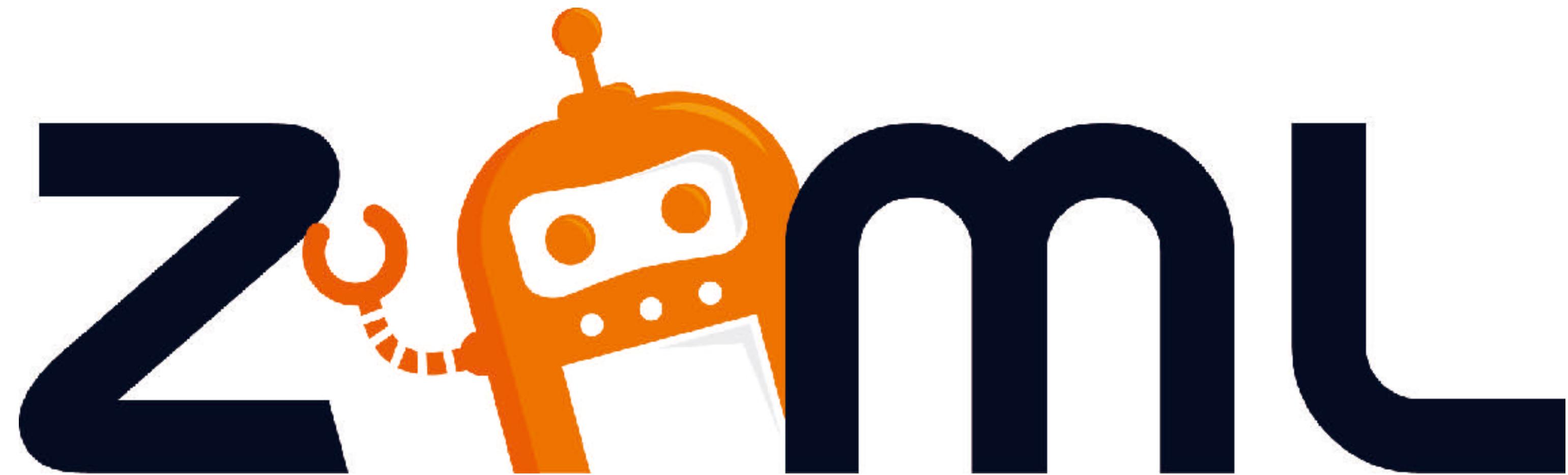
預測提前還款機率

模型

判斷還款意願

每半年一個新版本

Ensemble Model



### Data Assimilation

Rapidly discover, acquire, and onboard data sources at a massive scale.



### Modeling Tools

Train, ensemble, and productionalize machine learning models in one streamlined workflow.



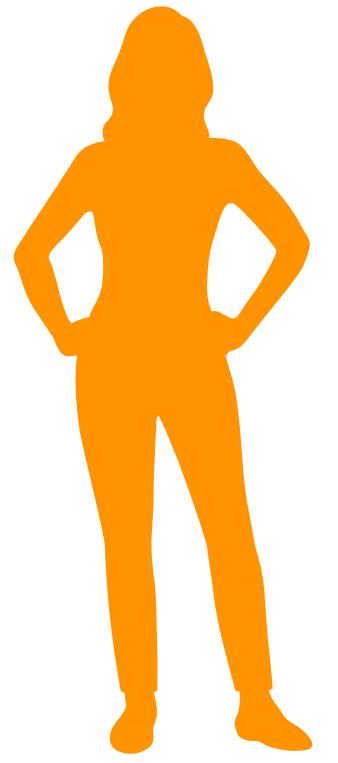
### Modeling Explainability

Unpack the "black box" of machine learning models to clearly communicate economic value and support compliance.

# 案例三：金融商業模式創新 查信用分最快

**credit karma<sup>TM</sup>**  
*Your Credit Scores Should Be Free. And Now They Are.*





消費者

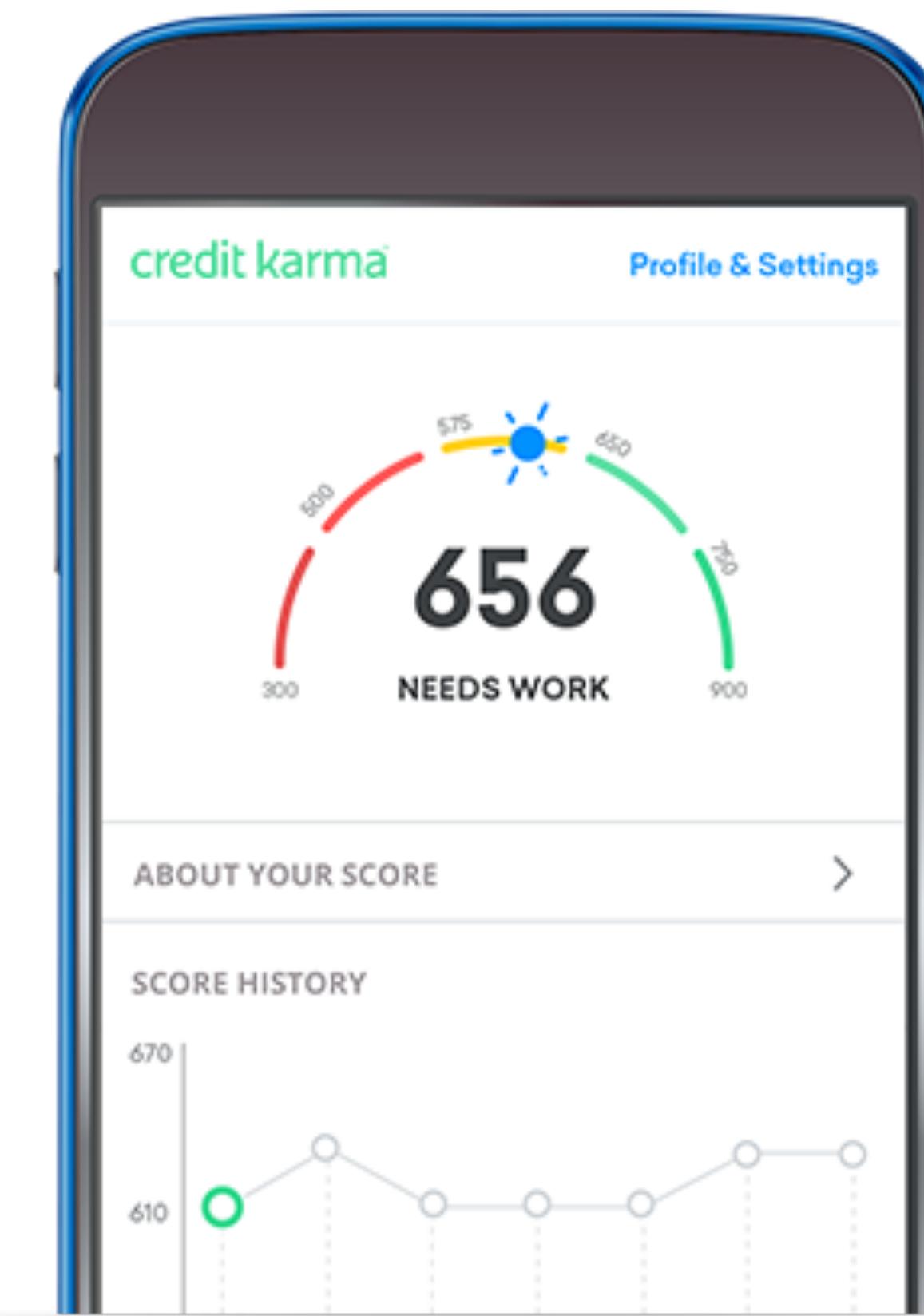
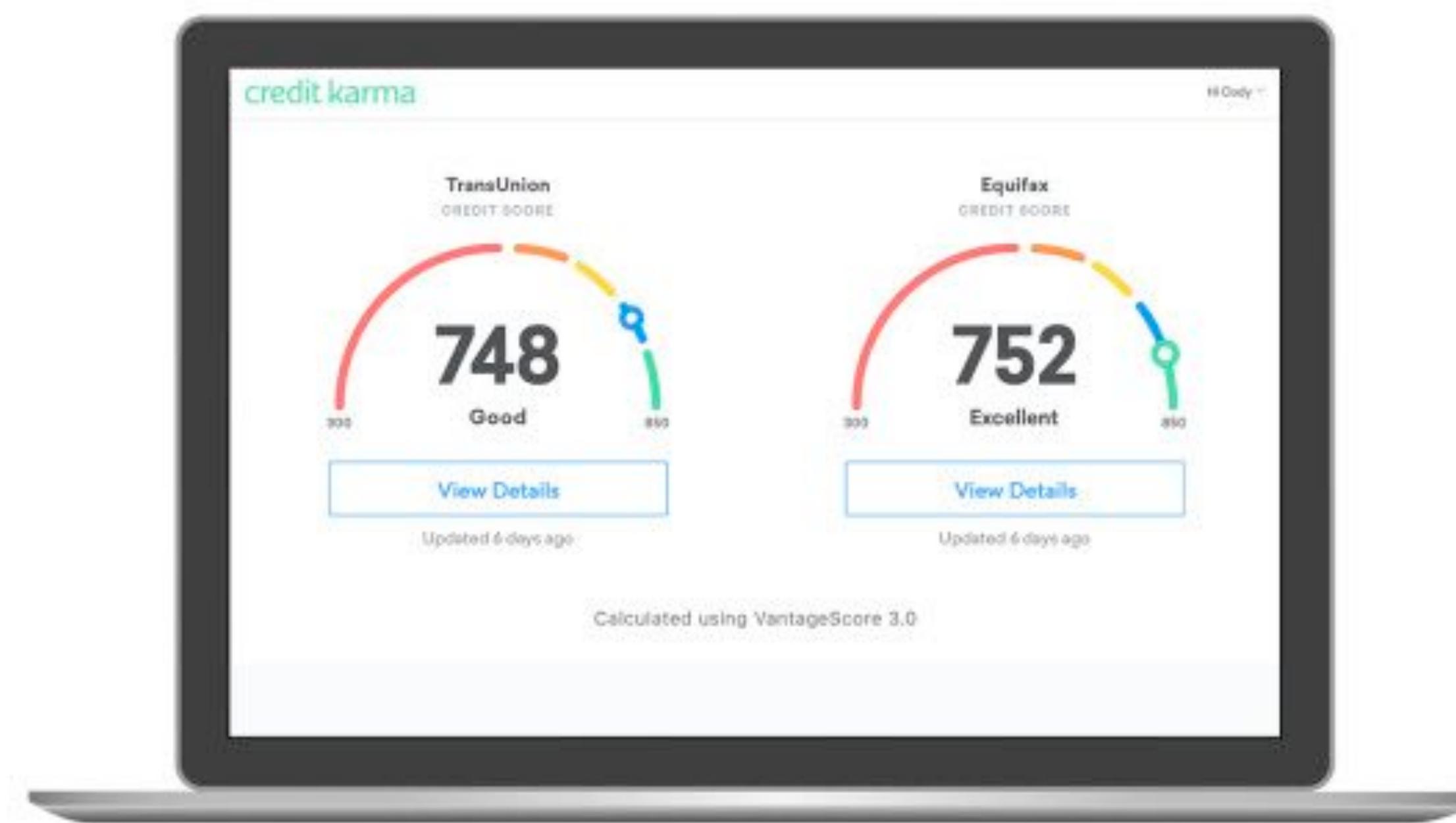
多方信用評等報告  
信用警示通知  
金融方案推薦



←... EQUIFAX®  
←... Experian®  
←... TransUnion®



# 多方信用分數報告



# 案例四：利用社群網路分析 反詐欺最厲害



[www.yirendai.com](http://www.yirendai.com)

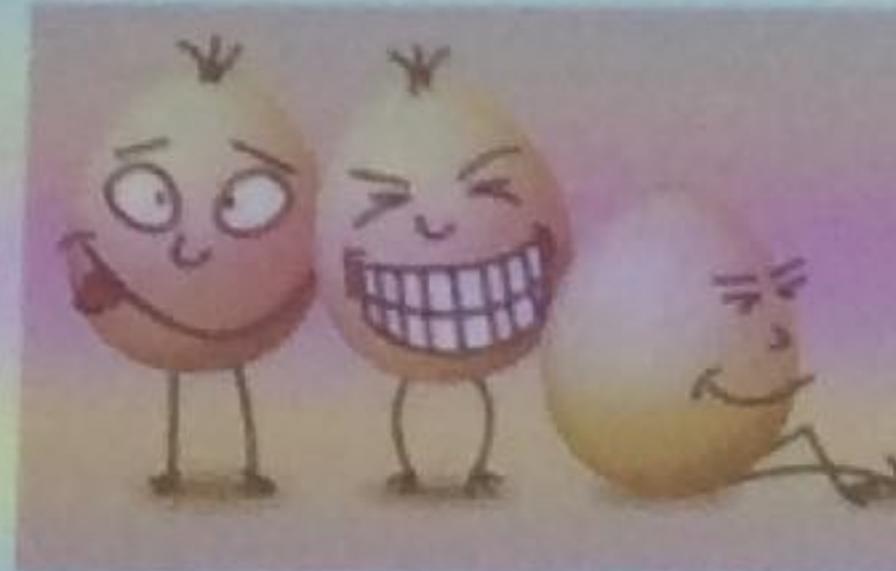
- 雖然講到爛，但真的很強

# 为什么找团伙

宜信  
CreditEase

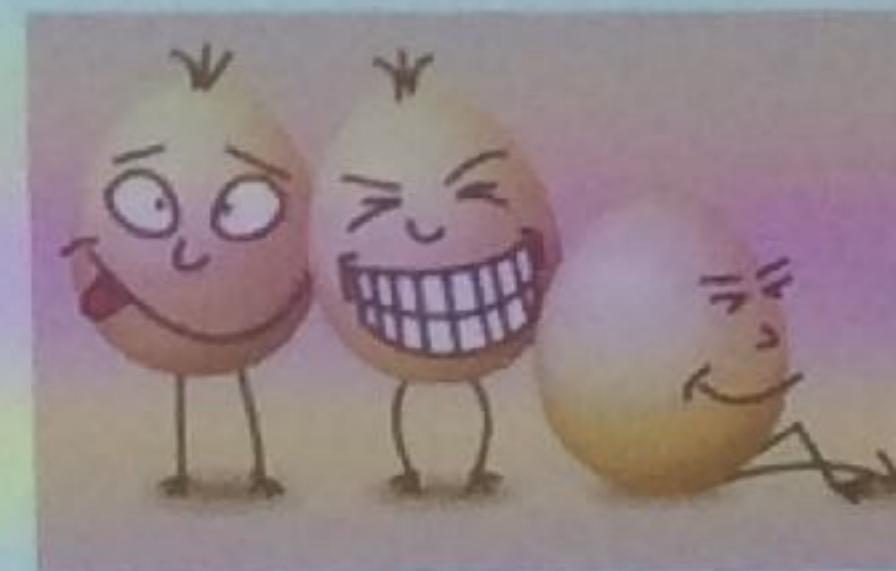
宜人贷  
宜信旗下P2P网贷平台

## 用户预警



- 新进件用户和某一个团伙联系密切

## 团伙监控

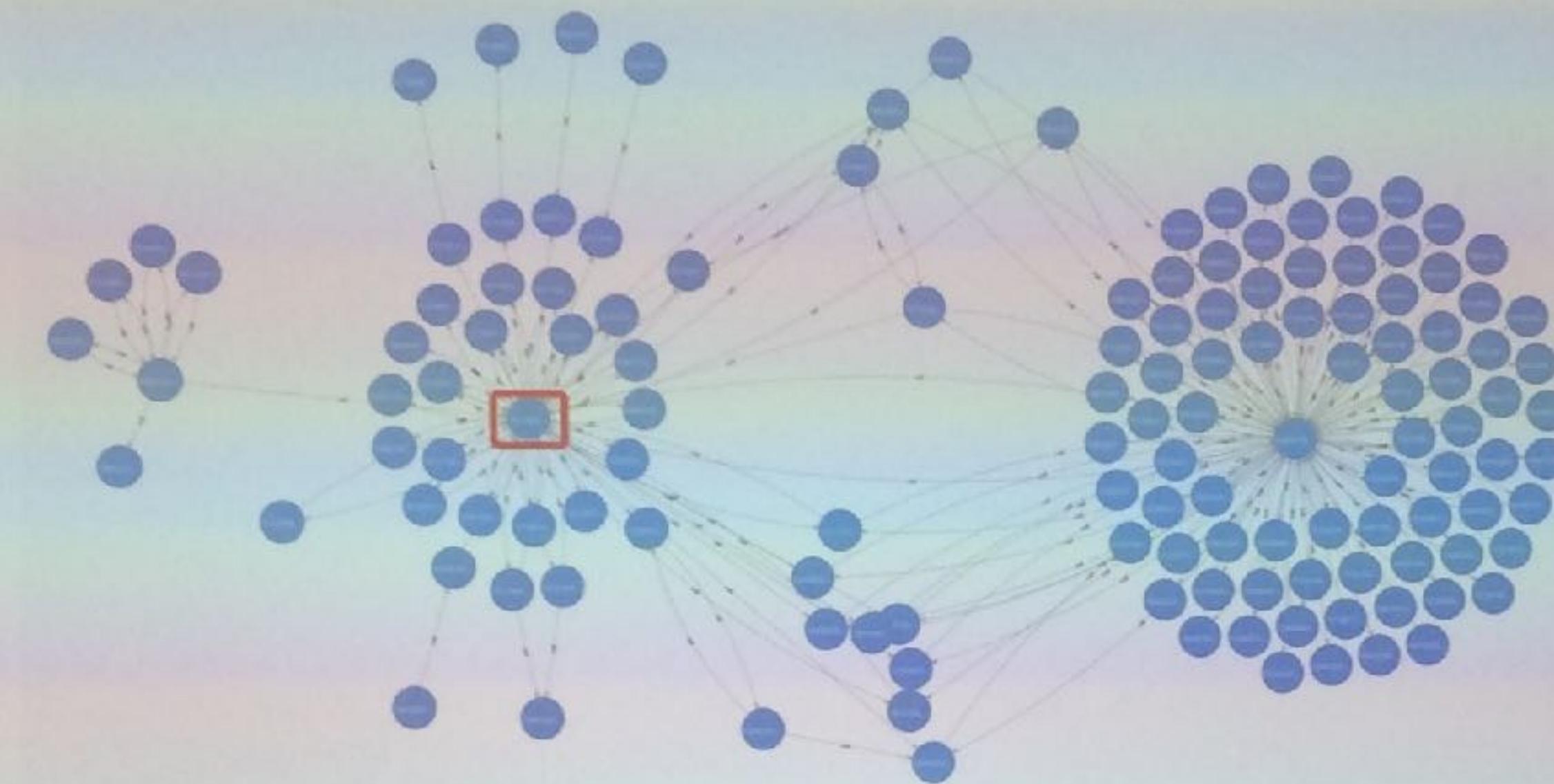


- 团伙在变化，每一段时间都有新的用户加入
- 大量进件用户联系同一个团伙
- 相似用户联系同一个团伙

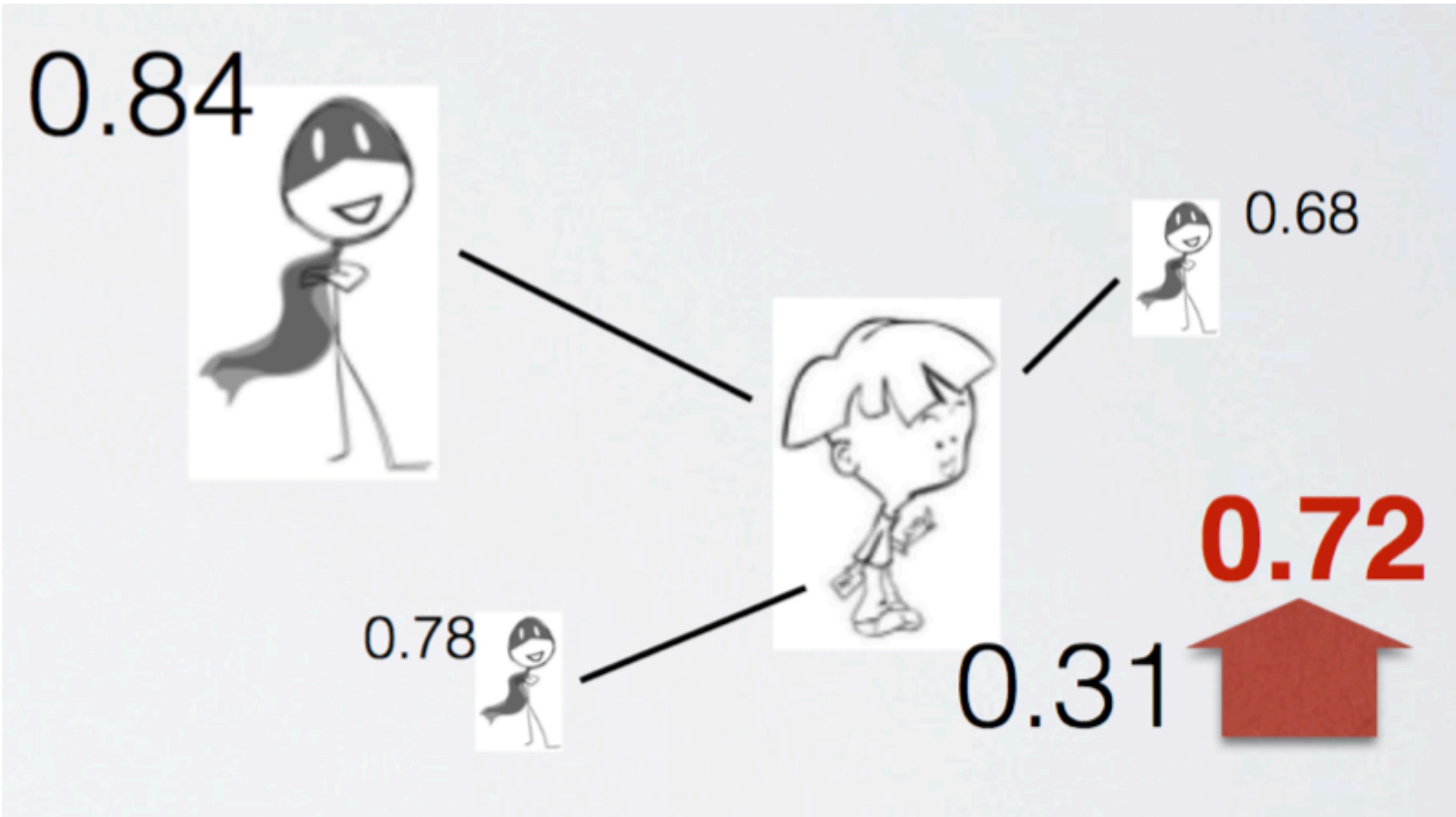
# 在Neo4j中实现

```
match (p:phone)-[r1:call]-(mp:phone)-[r2:call]-(lp:phone)
where p.phone_num = '187 **** 91'
and size(mp.phone_num) = 11
and mp.phone_num =~ "1.**"
and size(lp.phone_num) = 11
and lp.phone_num =~ "1.**"
return r1, r2
```

# 在Neo4j中实现



# 用PageRank穩定信用分



## 1、Active Learning

团伙骨干标签可以预测团，同时团伙中的未知骨干又被算法选择性送到人工调查。人工标记后标签返回到团中再次训练。我们投入了专业的反欺诈调查组，Active Learning在正确的人工标注下，效果值得期待。

人工调查需要及时，很多线索／欺诈证据耽搁了就不能有效收集了。我们还需不断监测调查的实时性。我们可以根据调查结果，设计自动化收集新的更有鉴别能力的数据源和特征，提升欺诈识别率。

## 2、Deep Learning

用户进件提交的信用数据，可以通过知识图谱结合深度学习，挖掘信用数据在时间序列上的新特性。

# kaggle<sup>™</sup> 上的風險分析題目

# **Default of Credit Card Clients Dataset**

Default Payments of Credit Card Clients in Taiwan from 2005

# 目標

## 顧客變數

### 基本資料

- ID
- LIMIT\_BAL
- SEX
- EDUCATION
- MARRIAGE
- AGE

### 付款歷史

- PAY\_N (N = 1~ 6): Repayment status
- BILL\_AMT\_N (N = 1~ 6): Amount of bill statement
- PAY\_AMT\_N (N = 1~ 6): Amount of previous payment

Predict

下個月是否違約

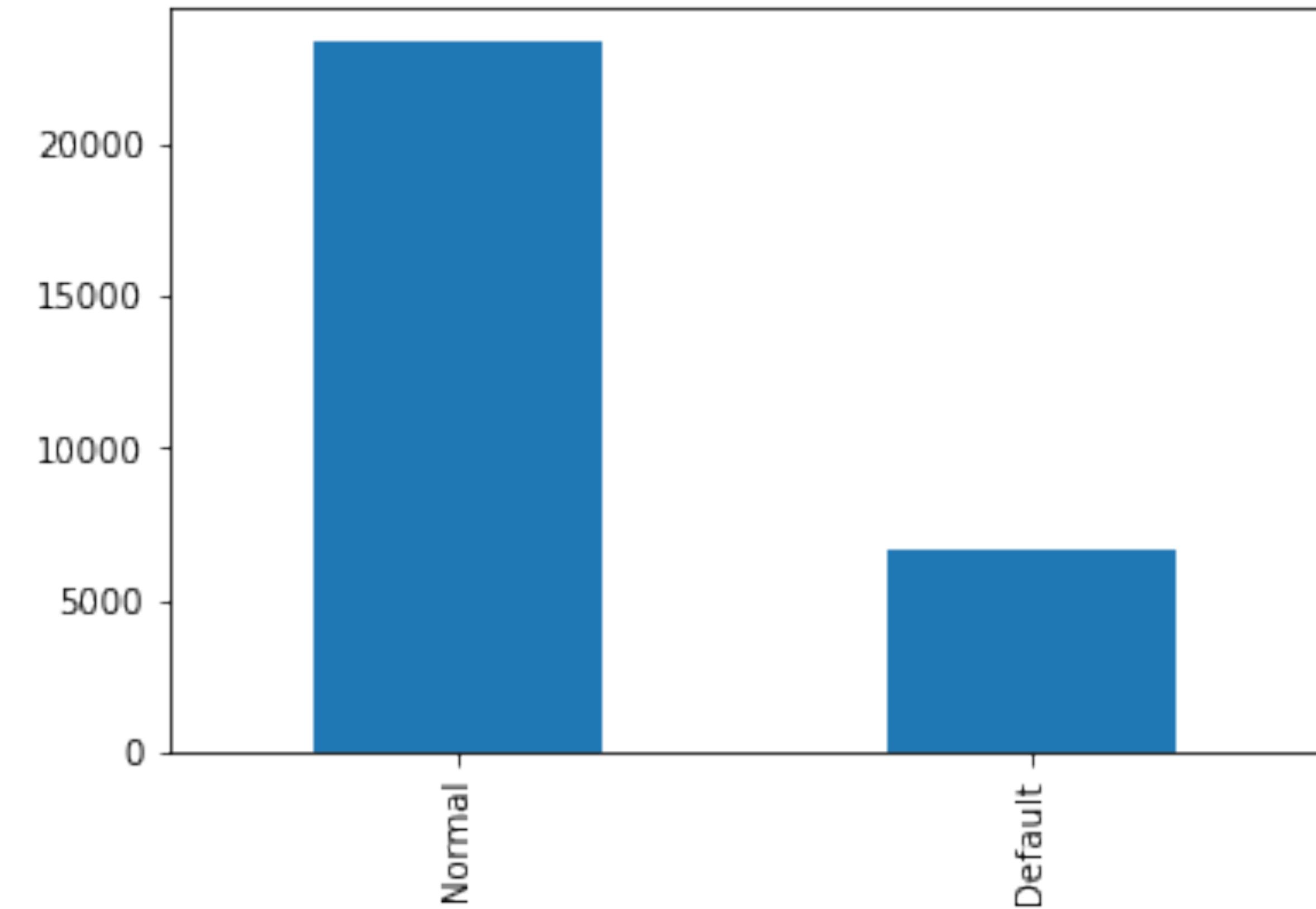
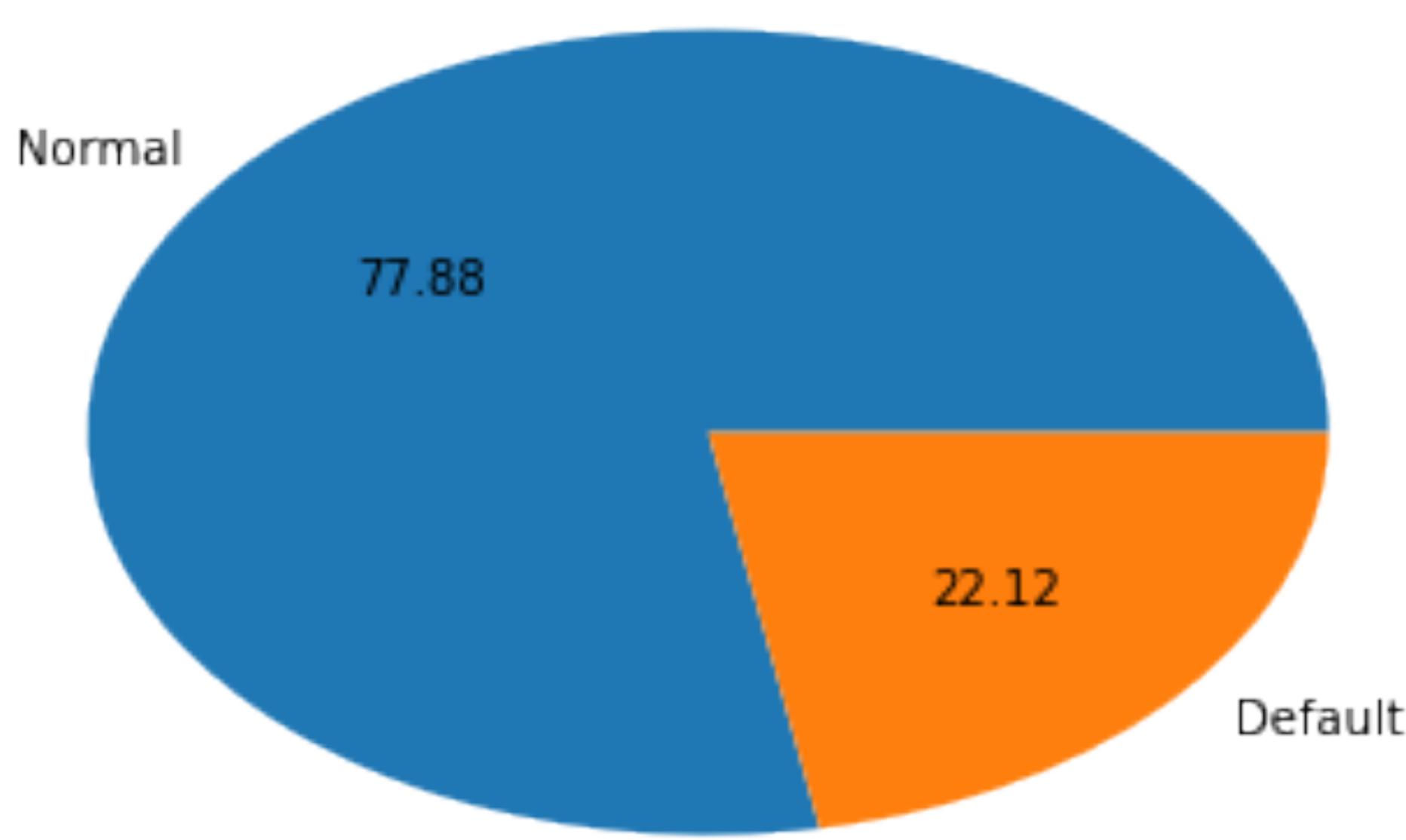
default.payment.next.month: 1=yes, 0=no

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2
0	1	200000.0	2	2	1	24	2
1	2	120000.0	2	2	2	26	-1
2	3	90000.0	2	2	2	34	0
3	4	50000.0	2	2	1	37	0
4	5	50000.0	1	2	1	57	-1

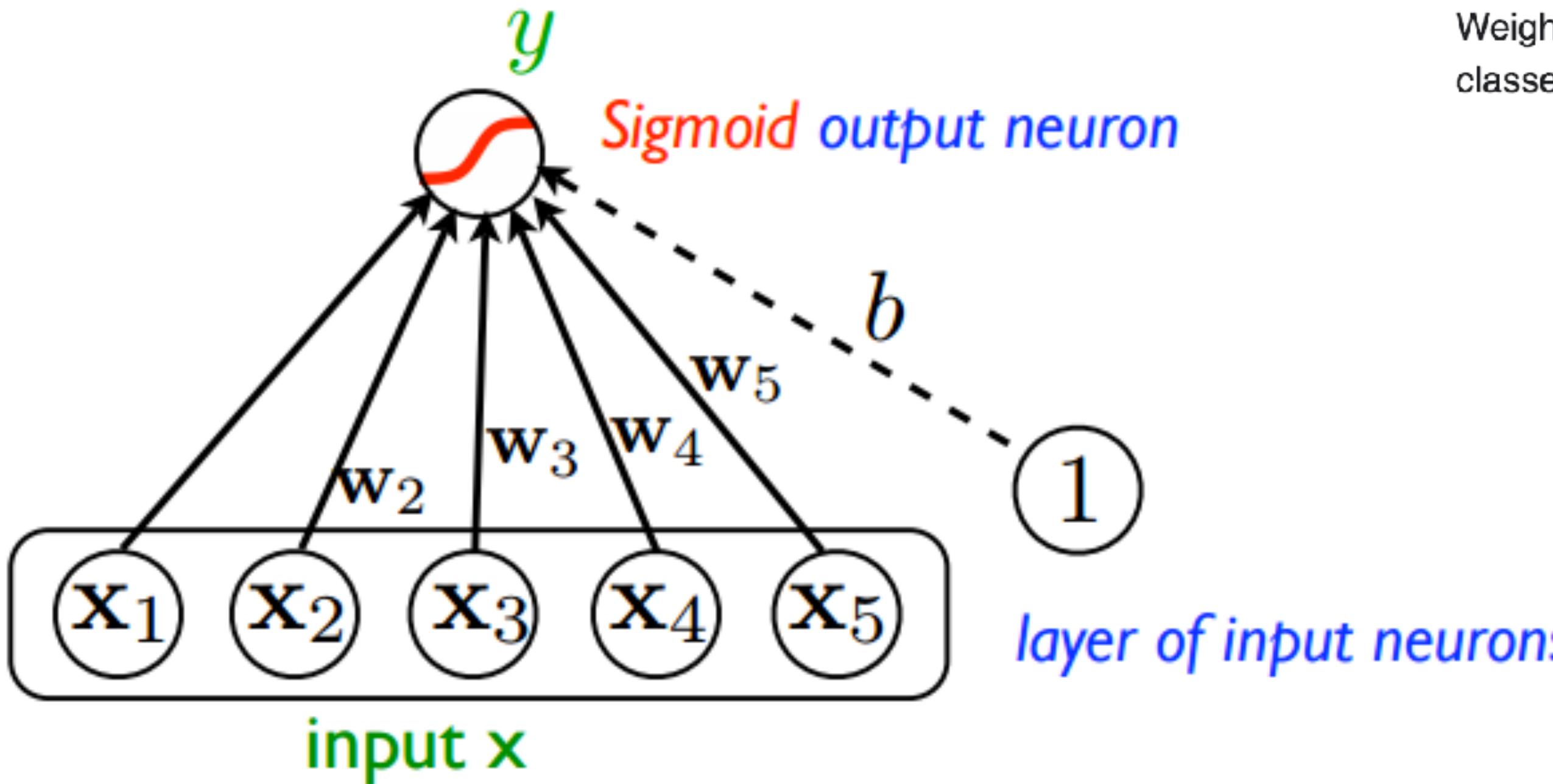
BILL_AMT6	PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6	default.payment.next.month
0.0	0.0	689.0	0.0	0.0	0.0	0.0	1
3261.0	0.0	1000.0	1000.0	1000.0	0.0	2000.0	1
15549.0	1518.0	1500.0	1000.0	1000.0	1000.0	5000.0	0
29547.0	2000.0	2019.0	1200.0	1100.0	1069.0	1000.0	0
19131.0	2000.0	36681.0	10000.0	9000.0	689.0	679.0	0

# Normal vs Default

Normal vs Default



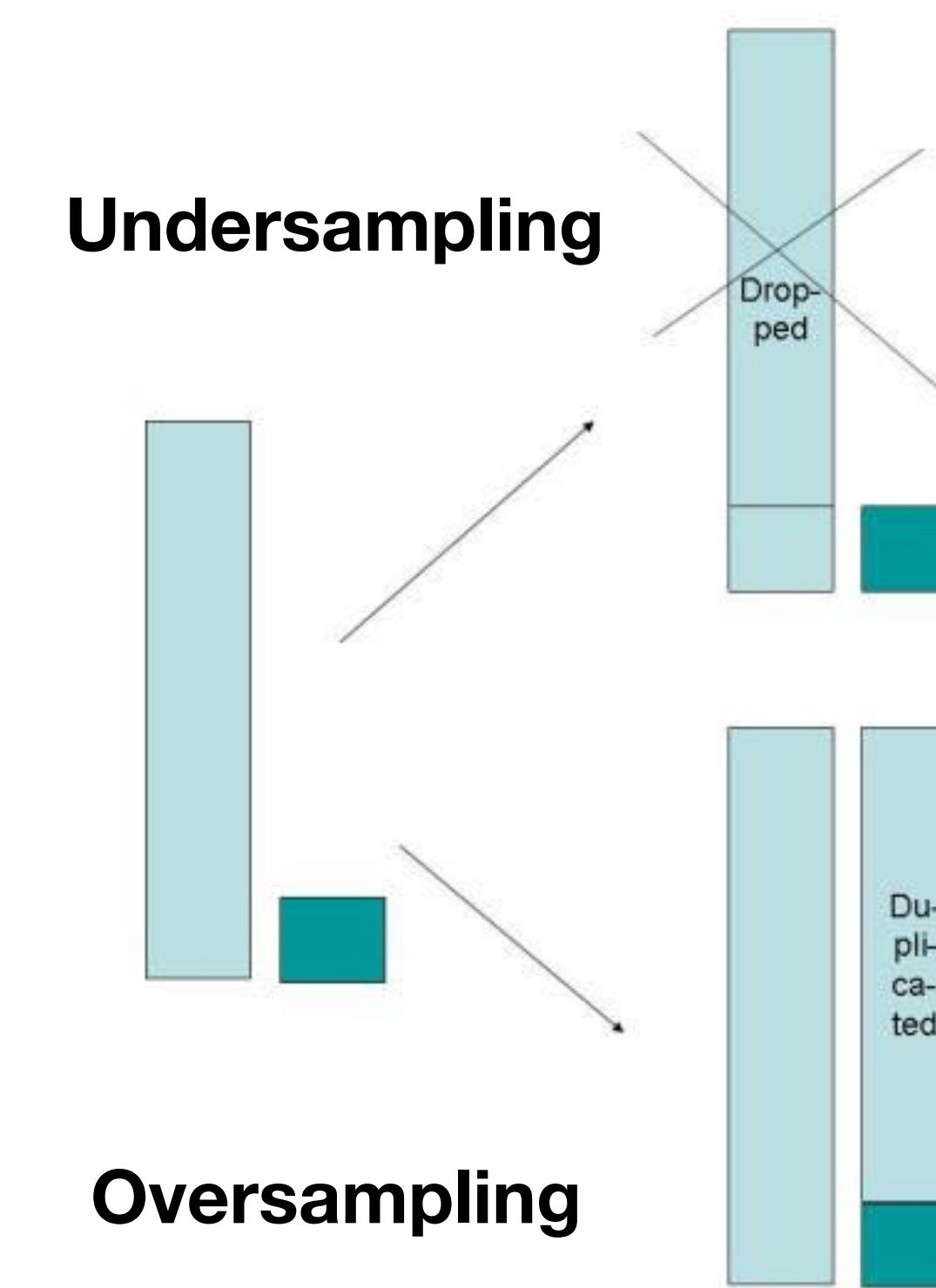
# Imbalanced data classification



## Logistic Regression

`class_weight` : dict or 'balanced', default: None

Weights associated with classes in the form `{class_label: weight}`. If not given, all classes are supposed to have weight one.



# Credit Card Fraud Detection

Anonymized credit card transactions labeled as fraudulent or genuine

# 目標

Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...

V25	V26	V27	V28	Amount	Class
0.128539	-0.189115	0.133558	-0.021053	149.62	0
0.167170	0.125895	-0.008983	0.014724	2.69	0
-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
0.647376	-0.221929	0.062723	0.061458	123.50	0
-0.206010	0.502292	0.219422	0.215153	69.99	0

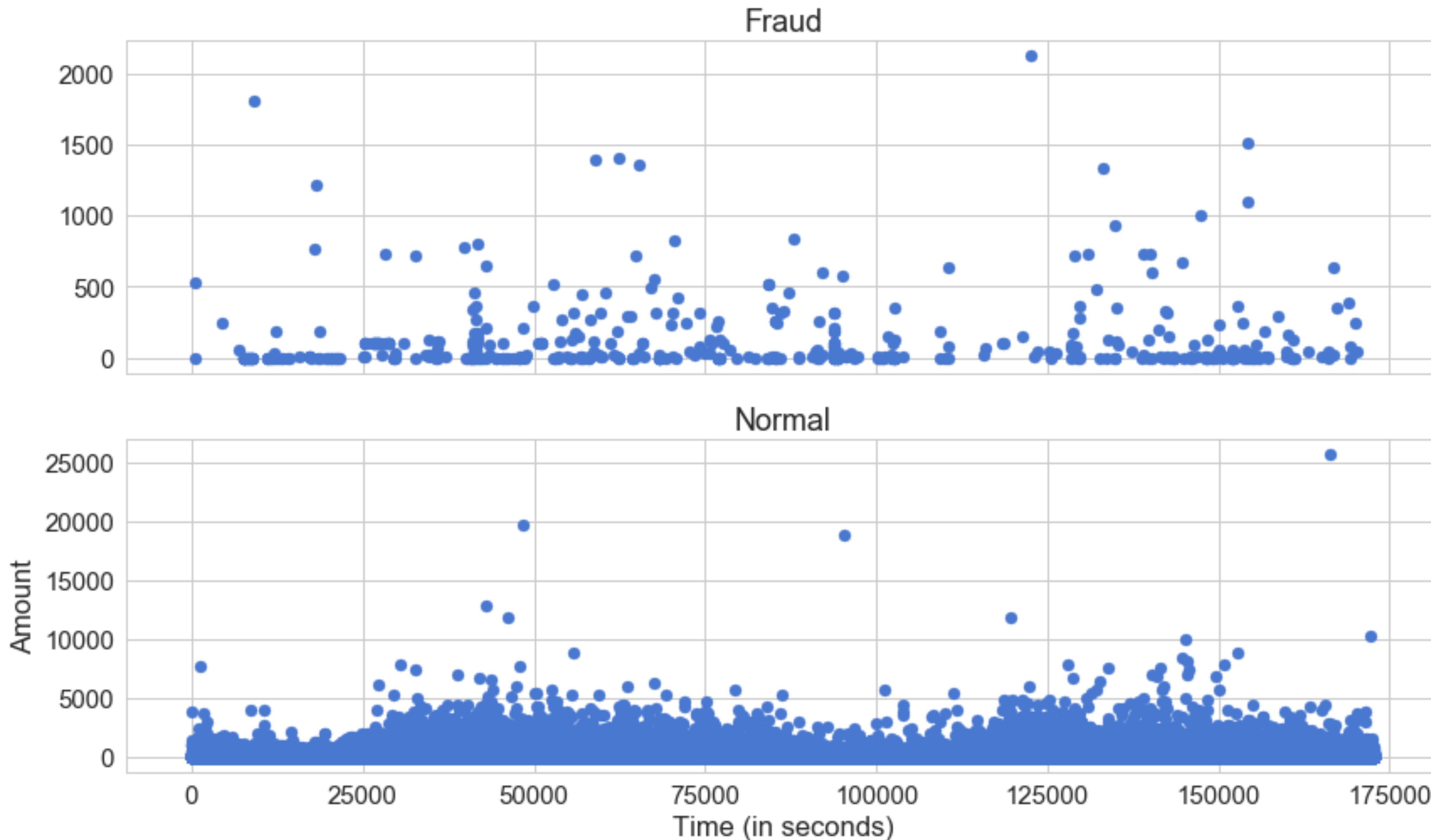
交易變數 Predict -----> 是否詐欺

# Super imbalanced!

**492** frauds  
—  
**284,807** transactions = **0.172%**

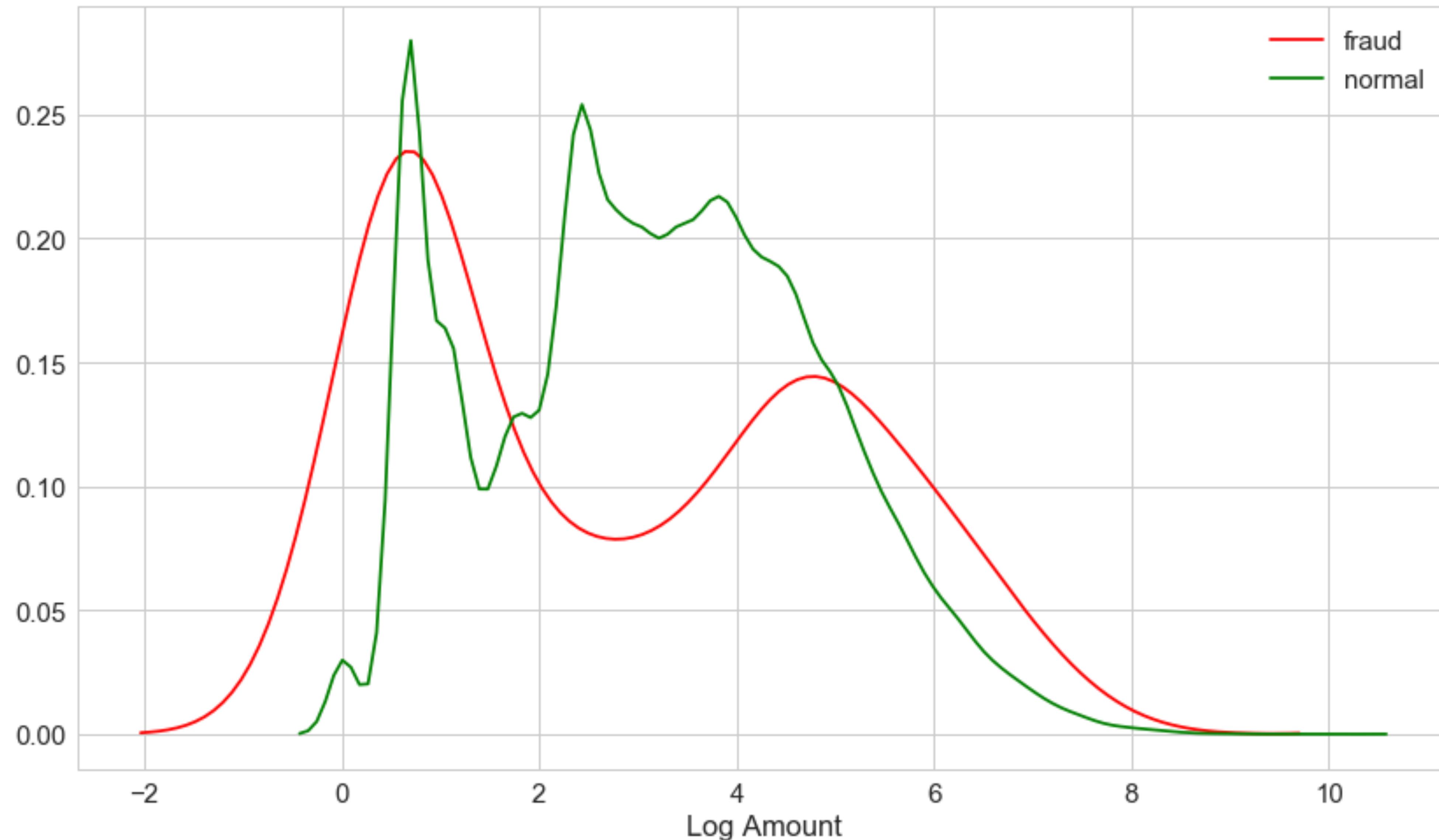
# Time

Time of transaction vs Amount by class



# Amount

Log amount per transaction by class

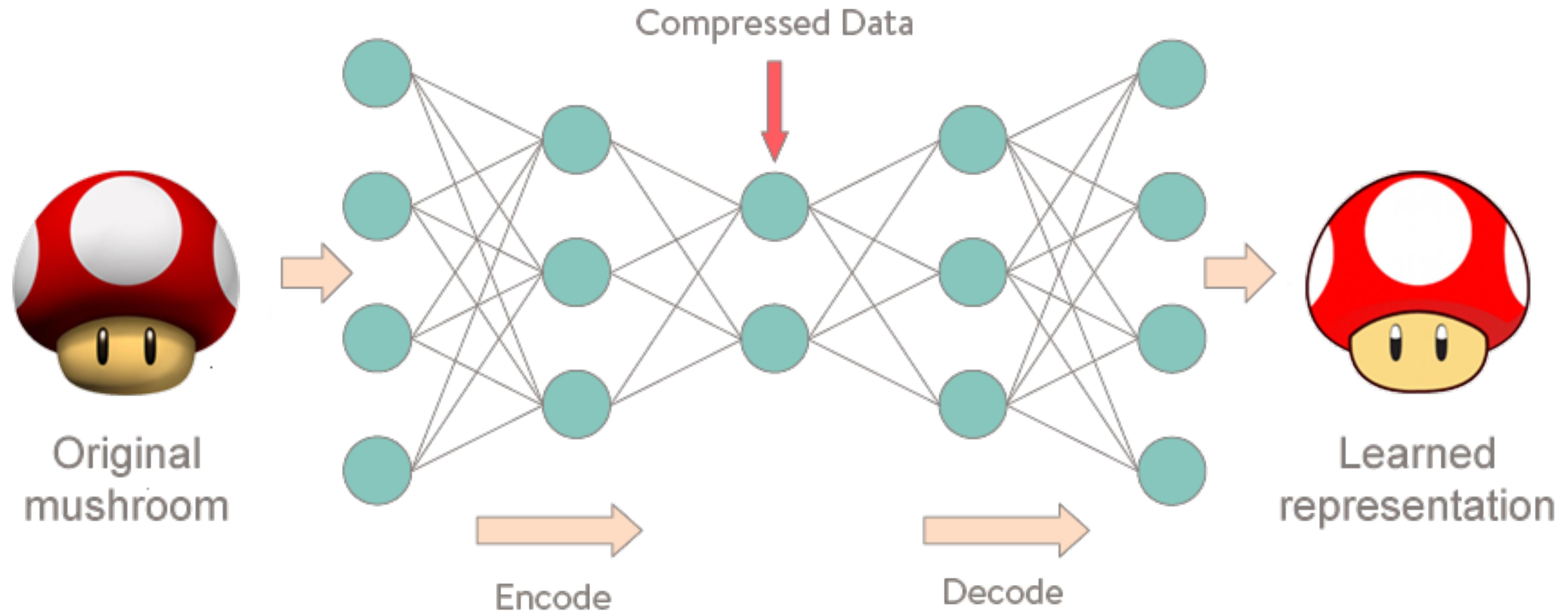


# Variable Selection

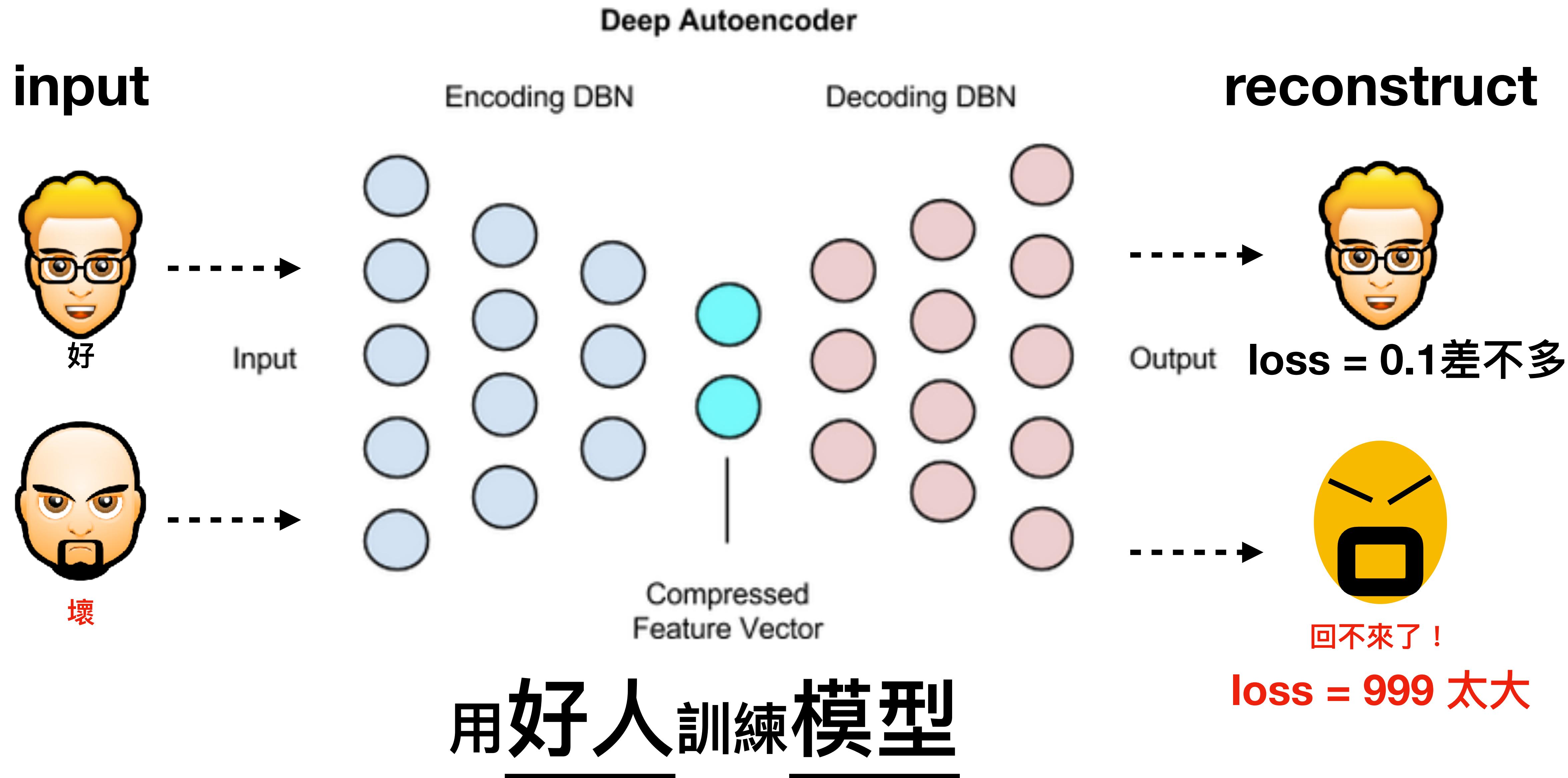
from PCA

V1 ~ V28 + Amount

# Model: AutoEncoder



# Anomaly Detection using AutoEncoder



# AutoEncoder in Keras

```
input_dim = x_train.shape[1]
encoding_dim = 14

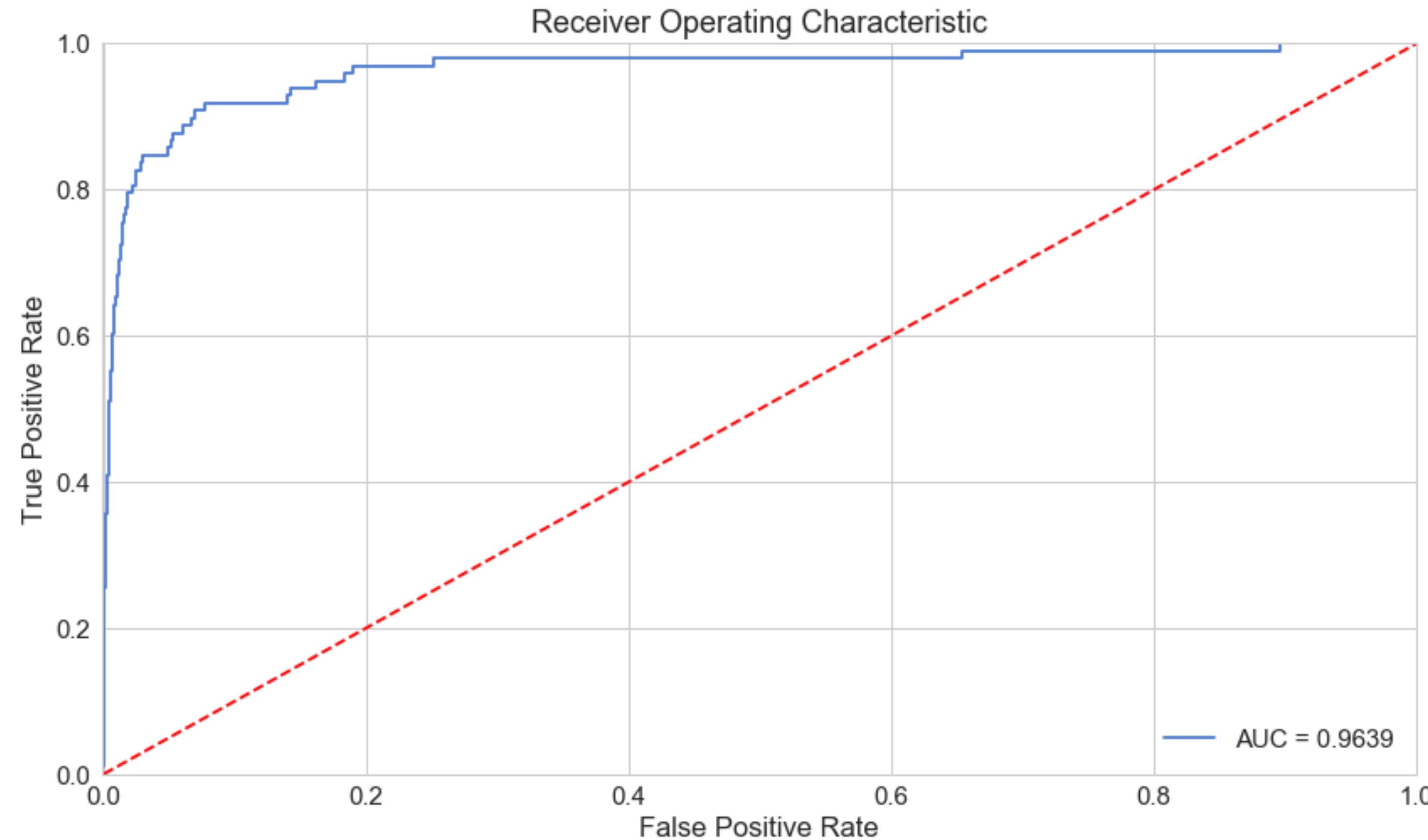
input_layer = Input(shape=(input_dim,))

encoder = Dense(encoding_dim, activation="tanh",
                 activity_regularizer=regularizers.l1(10e-5))(input_layer)
encoder = Dense(int(encoding_dim/2), activation="relu")(encoder)

decoder = Dense(encoding_dim, activation="tanh")(encoder)
decoder = Dense(input_dim, activation="relu")(decoder)

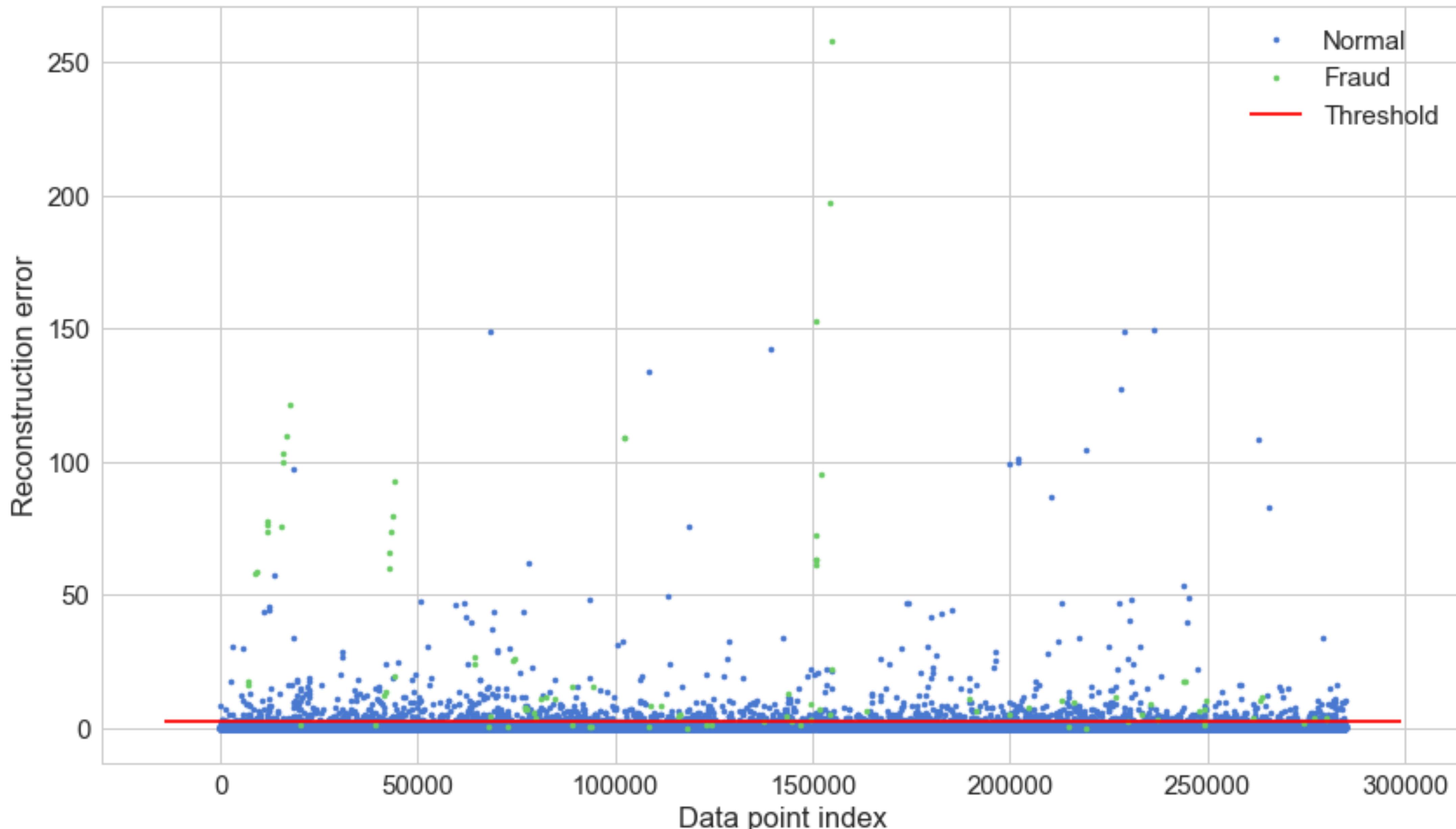
autoencoder = Model(inputs=input_layer, outputs=decoder)
```

# ROC Curve

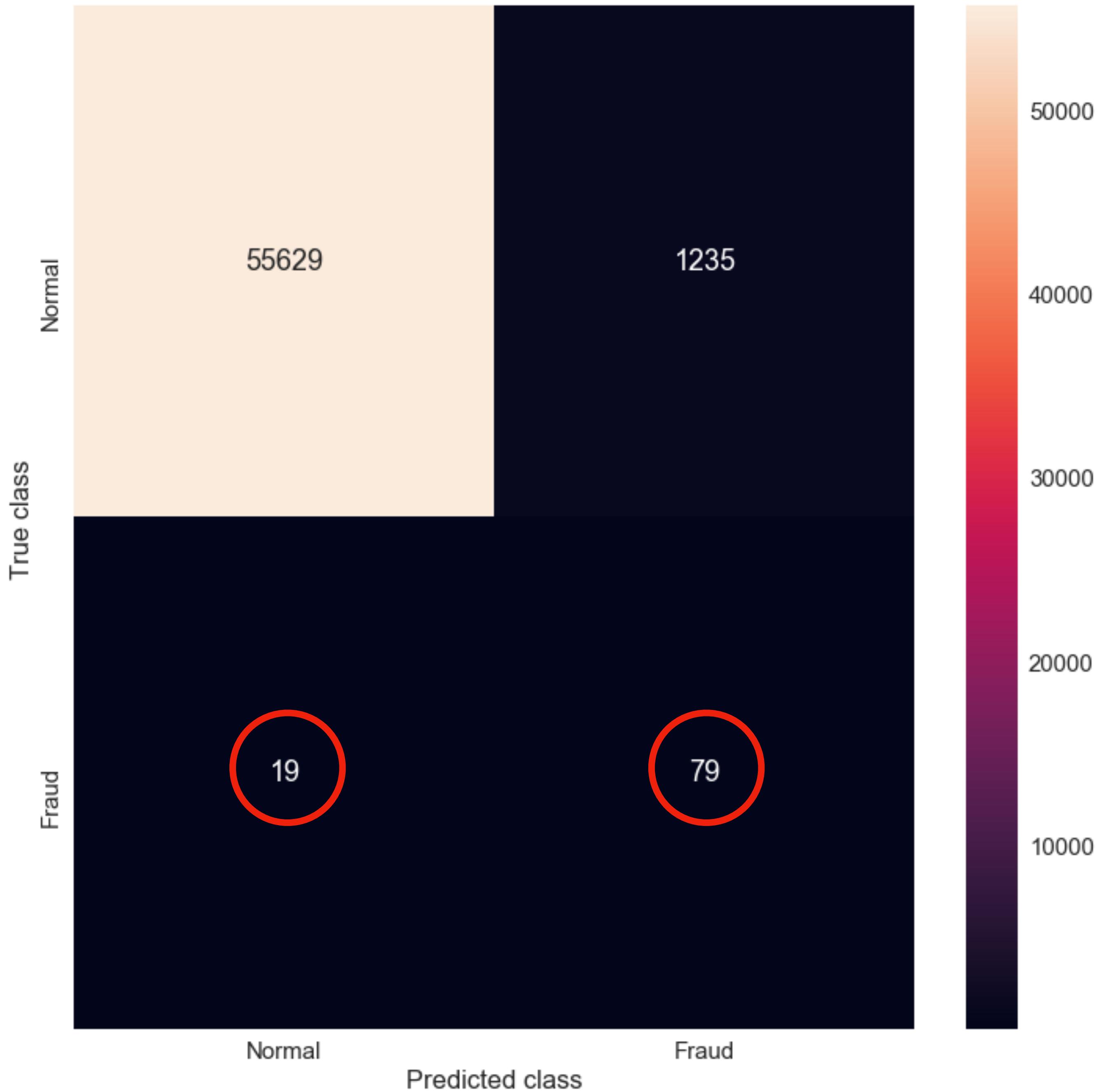


**loss threshold = 2.9**

Reconstruction error for different classes



Confusion Matrix



# 風險偵測大數據架構

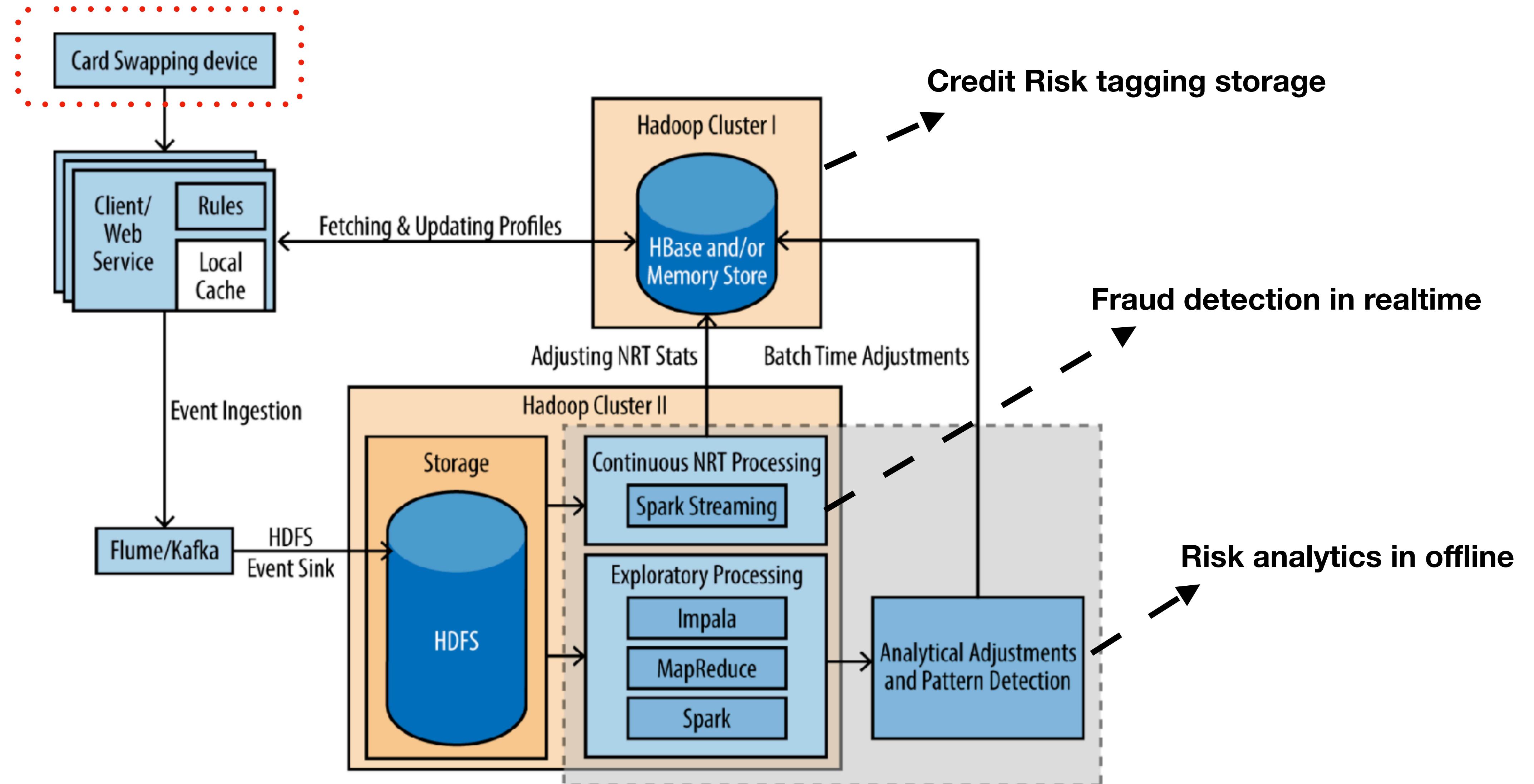


Figure 9-1. High-level view of fraud detection system

總結



信用評等開宗始祖 統計模型準到High



生活數據信用評分 強國人們的最愛



致力運用機器學習 改變信評的生態

credit karma™

金融商業模式創新 查信用分數最快



利用社群網路分析 反詐欺超級厲害



金融科技改變台灣 相



**讓我們朝風險大數據邁進吧！**