Group 4: Jimmy Nguyen, Kyle Jow, Kyle Pickle, Jacob Tuttle, Kris Wong

Credit Score

Classifier

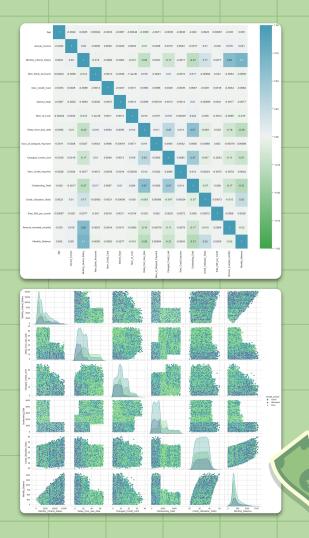
Background

Standard

000

GO

- Credit scoring necessary to mitigate lending risk
- Banks use it to predict whether a loan will be paid back in full or not
- One of the oldest implementations of ML (1950s)
- 5 C's approach ⇒ Decision Trees ⇒ Log. Reg. ⇒ ??
- Performance vs Transparency
- We use credit Score data from Kaggle
- Features based on FICO credit standards
- Used to classify credit score into:
 - Good
 - Standard
 - Poor



Exploratory Analysis

- Fairly imbalanced dataset
 - Not enough 'Good' classifications
- Outliers and unnecessary features
- Decided to use different models to test the dataset:
 - SVM Linear Kernel
 - SVM RBF Kernel
 - Multinomial Logistic Regression
- Kaggle dataset had 27 features
 - Cut down these features to 5
- 80:20 split

Results

Logistic Regression

SVM Linear Kernel

SVM RBF Kernel

			precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
		Good	0,50	0.14	0.22	2801	Good	0.00	0.00	0.00	2801	Good	0.59	0.06	0.10	2801
		Poor	0.58	0.39	0.47	4560	Poor	0.59	0.44	0.50	4560	Poor	0.66	0.57	0.61	4560
	S	tandard	0.57	0.81	0.67	8213	Standard	0.56	0.84	0.68	8213	Standard	0.60	0.83	0.70	8213
	а	ccuracy			0.57	15574	accuracy			0.57	15574	accuracy			0.62	15574
		cro avg	0.55		0.45	15574	macro avg	0.39	0.43	0.39	15574	macro avg	0.62	0.49	0.47	15574
we	igh	ted avg	0.50	0.57	0.53	15574	weighted avg	0.47	0.57	0.50	15574	weighted avg	0.62	0.62	0.57	15574
		Good	403	48	2350	- 6000	Good	0	52	2749	- 6000	Good	157	48	2596	- 6000
		Poor	68	1784	2708	- 4000	Poor True label	Ō	1986	2574	- 4000 - 3000	Poor Poor	23	2601	1936	- 4000 - 3000
		Standard	343	1228	6642	- 2000	Standard	0	1311	6902	- 2000 - 1000	Standard	84	1273	6856	- 2000
5			Good	Poor Predicted label	Standard			Good	Poor Predicted label	Standard			Good P	Poor Predicted label	Standard	

Results - Cont'd

SVM Linear Kernel

Logistic Regression

- Outliers

precision

Good

Poor

Standard

accuracy

macro avo

weighted avg

0.52

0.62

0.61

0.58

0.59

recall f1-score

0.37

0.49

0.69

0.60

0.52

0.58

0.28

0.40

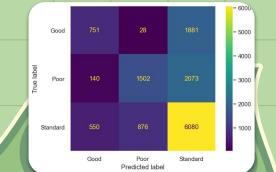
0.81

0.50

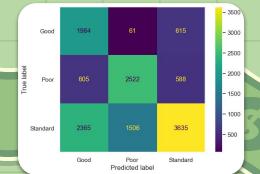
0.60

SVM RBF Kernel

+ Oversampling - Outliers + Oversampling - Outliers recall f1-score support precision precision recall f1-score support support 2660 Good 0.40 0.75 0.52 2660 0.40 Good 0.75 0.52 2660 Poor 0.62 0.68 0.65 3715 0.62 0.68 0.65 Poor 0.75 0.48 0.59 7506 Standard 0.49 0.59 7506 Standard 0.75 0.59 13881 13881 accuracy 0.59 13881 accuracy 13881 macro avg 0.59 0.64 0.59 13881 0.59 0.64 0.59 13881 macro avg 13881 weighted avg 0.65 0.59 0.59 13881 weighted avg 0.65 0.59 0.59 13881 - 3500 - 3500







Conclusion & Analysis

- SVM with RBF kernel yielded the highest accuracy
 - Highest recall for standard credit scores
 - Highest Precision for poor credit scores
- Very low performance in all models when predicting high credit scores
 - Possibly increased by sampling an equal amount of each credit score
- Removed outliers in logistic model:
 - Slightly better at predicting scores, especially 'Good'
- Oversampling and removing outliers in SVM models:
 - Greatly improved predictions of edge 'Poor' and 'Good' credit scores
 - ...at the cost of 'Standard'

Demo

https://ecs171-project.streamlit.app/

32

I manual harmony with		
· · · · · · · · · · · · · · · · · · ·	1.45	
from the second se		
the second se		
1		
the state of the s		
	100	
100000000000000000000000000000000000000		
a la companya da ser		
Construme by		
1		
	17	
 Specify Weight and Constitution of Manufacture 1 American Soft Print Act Constitution Act Soft 		
This is an Unit Francisco Deal (Sam Science)		
		C Respond

4.4.4.4.8.8