2019 Fall EE 5183 FinTech - Final Project
Credit Card Default Prediction (Group 2)

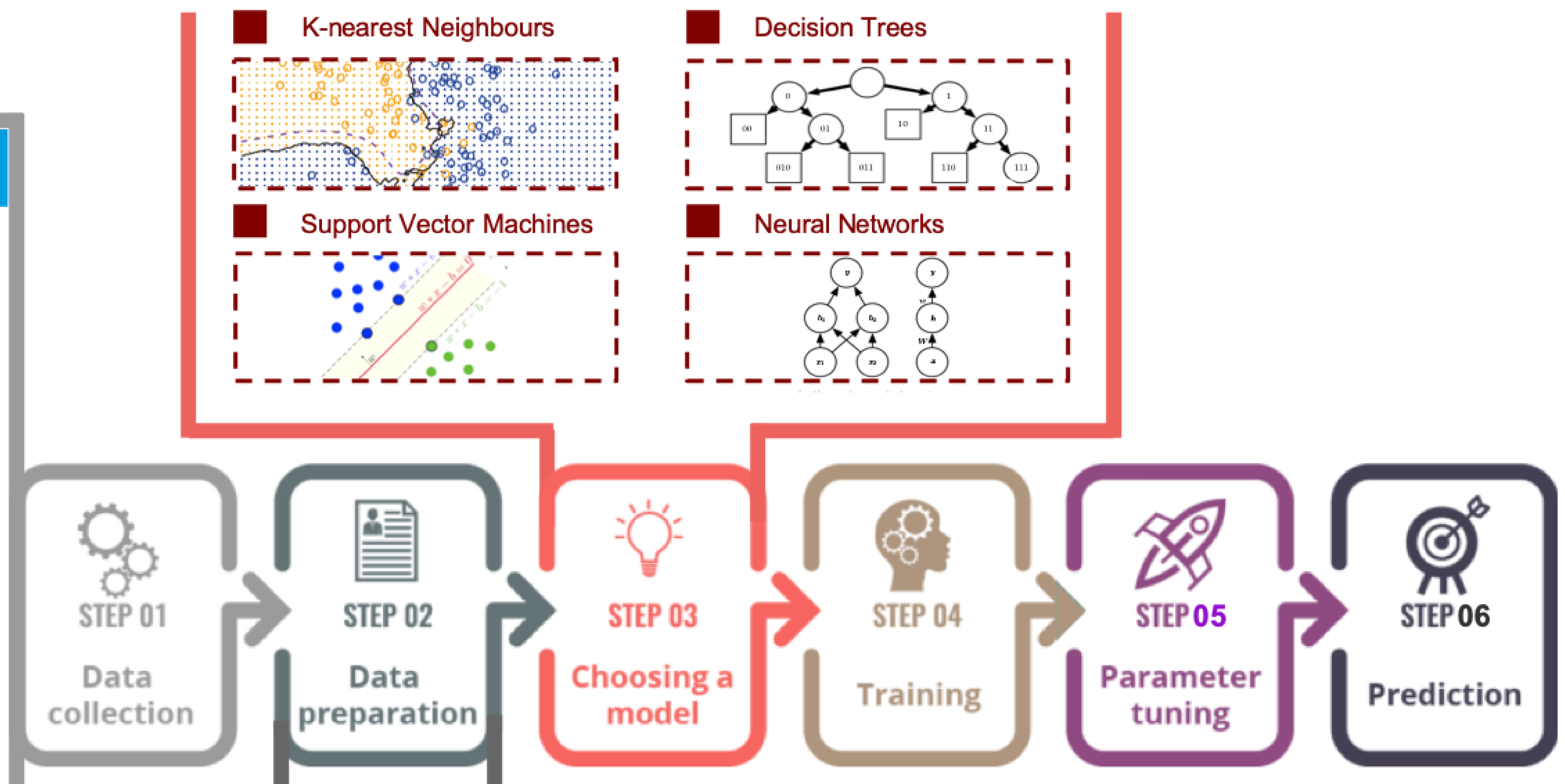
Lewis Mace, David Lin, Guillaume Desermeaux, Dominik Happel, Gregor Lied
2020/01/07, National Taiwan University

Introduction

Credit Card Default happens when clients fail to adhere to the credit card agreement, by not paying the monthly bill. Risk & Credit departments in financial institutions, try to detect the costumers that tend to default in the future, not only to secure money, but even make more money as they could demand higher interest rates for more risky costumers. Meanwhile, the digitalization is not only changing today's society but also companies' business models, in particular of the financial industry. In general, the large variety of digitalized processes and connected devices (Industry 4.0) generates a huge amount of data which can be used to extract valuable insights. In this research project, we want to test the performance and usability of different machine learning and deep learning models to analyze private financial data to predict credit card default, by detecting clients that will not be able to pay their debts next month.

Project Overview

Feature Name [N = 24]	Discription [Samples = 30000]
ID	User ID
LIMIT_BAL	Amount of the given credit in NTD
SEX	Gender (1=male,2=female)
EDUCATION	Education (1=graduate school,2=University,3=others)
MARRIAGE	Marital status (1=married,2=unmarried)
AGE	Age (year)
PAY_X (X = 1,...6)	Repayment status (from April to September, 2005) (-1=pay duly, 1=one month delay, ..., 9=nine months delay and above)
BILL_AMTX (X = 1,...6)	Amount of bill statement in NTD (from April to September, 2005)
PAY_AMTX (X = 1,...6)	Amount of previous payment in NTD (from April to September, 2005)
Next Month Default	Default payment (1=yes, 0=no)



Data Split
80% Training
10% Validation
10% Test

Data Normalization

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)}$$

Figure 1: Overview of full credit card default prediction project

Results

Model	Type of model	Accuracy (%)	Sensitivity (%)	Specificity (%)
K Nearest Neighbors	*K = 11*	80	31	94
Decision Tree	*Unbalanced Classes*	82.8	39	92
	Balanced Classes	73	60	73
	SMOTE	81	44	91
SVM	*RBF Kernel Unbalanced Classes*	83	34	96
	RBF Kernel Balanced Classes	78	55	85
MLP	*Unbalanced Classes*	81	24	98
	Balanced Classes	73	63	76

Figure 2: Accuracy, Sensitivity and Specificity comparison over all models

References

[1] I-Cheng Yeh and Che-hui Lien. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2):2473–2480, 2009.

Acknowledgements

Special thanks to Prof. Che Lin and the TAs of the Financial Technology class for the interesting lecture and the support.

Exploratory Data Analysis

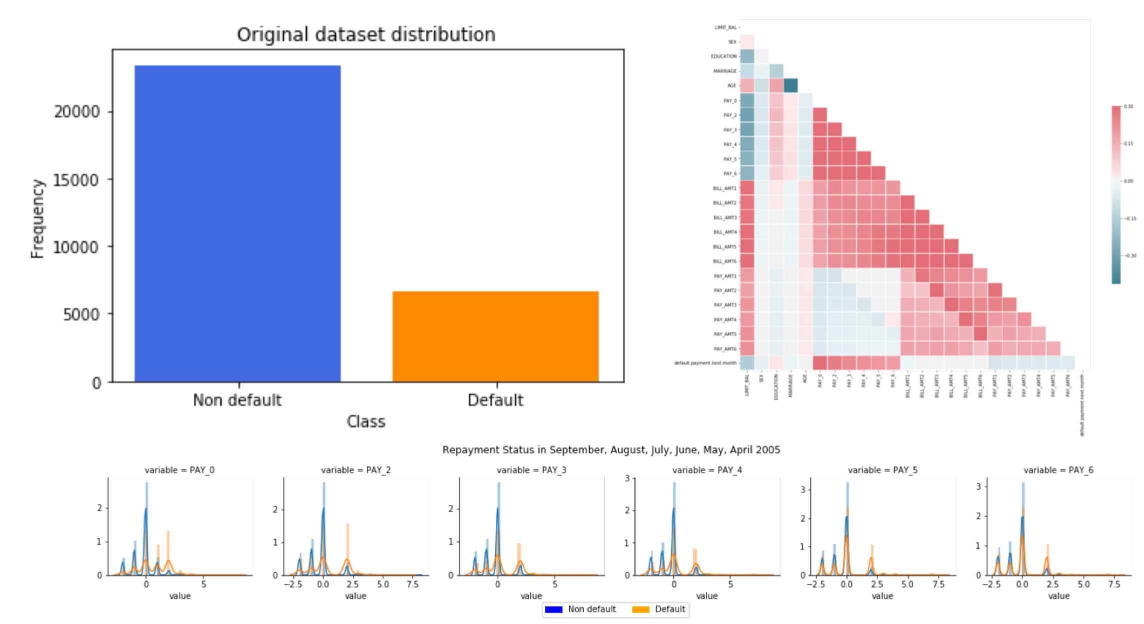


Figure 3: Data Analysis plots

K-nearest Neighbours

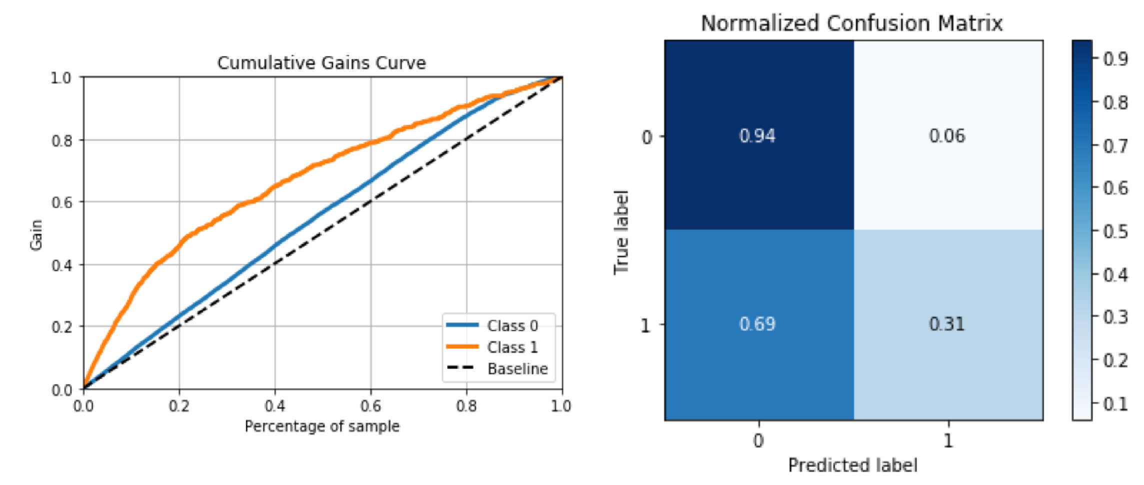


Figure 4: KNN: Gain Curve and Confusion Matrix

Decision Trees

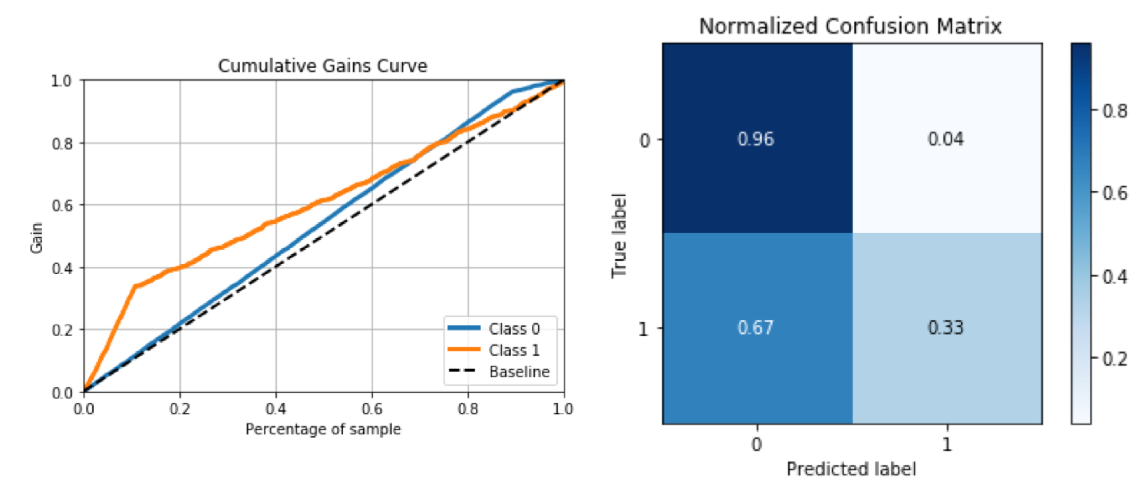


Figure 5: DT: Gain Curve and Confusion Matrix

Support Vector Machine

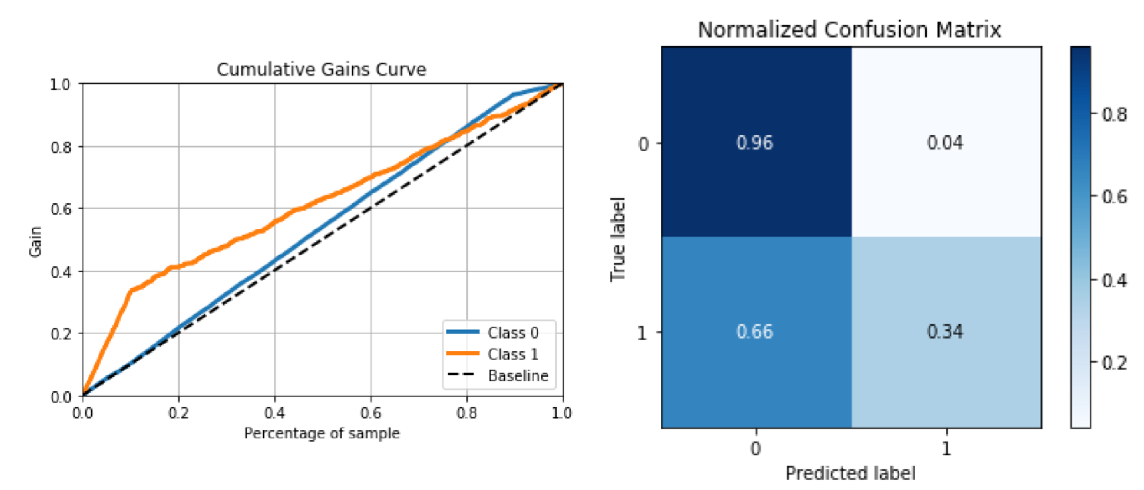


Figure 6: SVM: Gain Curve and Confusion Matrix

Neural Network

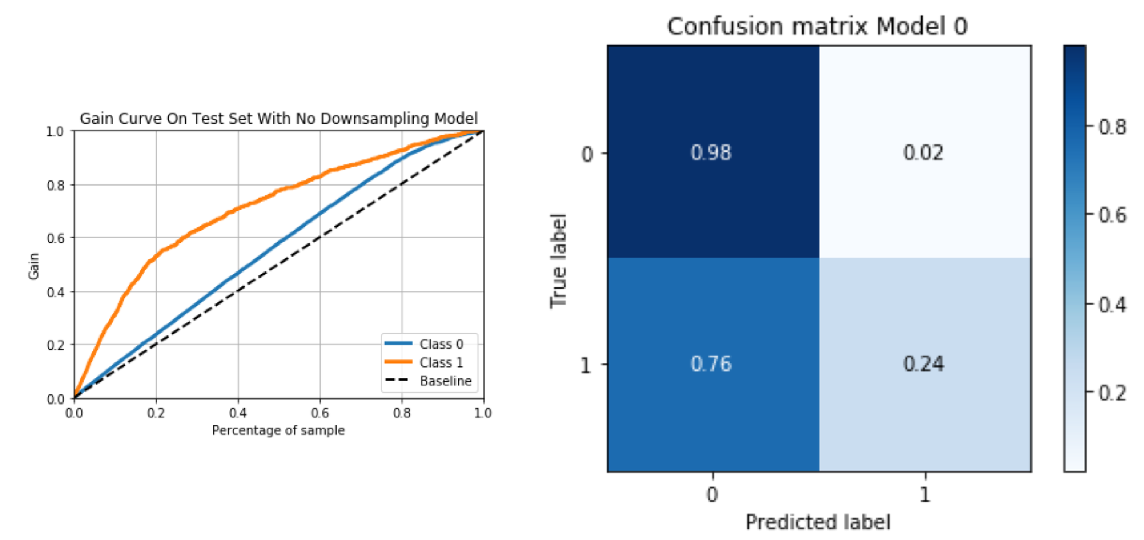


Figure 7: NN: Gain Curve and Confusion Matrix