

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
ФАКУЛЬТЕТ ІНФОРМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ
Кафедра автоматизованих систем обробки інформації та управління

ЗВІТ

до лабораторної роботи №4
з дисципліни «Інтелектуальний аналіз даних»

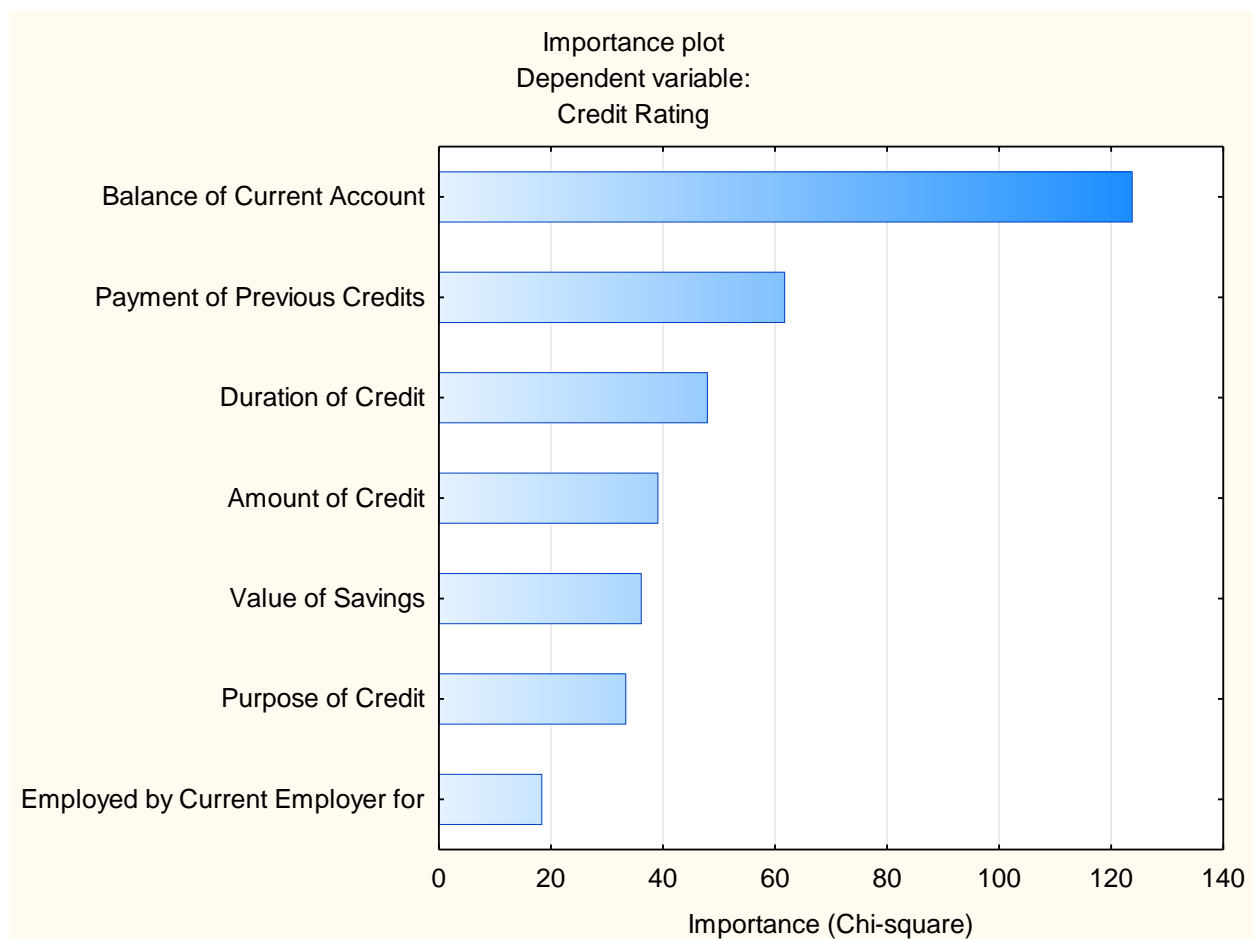
Виконав:
студент гр. ІС-32
Капорін Р. М.

Київ – 2016

ЗАВДАННЯ

В процессе предоставления кредита банки заинтересованы в изучении платежеспособности будущего потребителя кредита. Цель этого изучения - моделирование или предсказание вероятности, с которой претендент на кредит может быть отнесен к привлекательным или непривлекательным клиентам.

Методы, описанные в этом примере, демонстрируют построение кредитно-скоринговой модели с помощью *STATISTICA Data Miner* - системы для проведения интеллектуального анализа данных, позволяющей идентифицировать входы или предикторы, которые отделяют рискованных клиентов от всех остальных. Предиктивные методы, успешно применяемые на тестовых данных, и в дальнейшем могут быть использованы для предсказания новых рискованных клиентов.



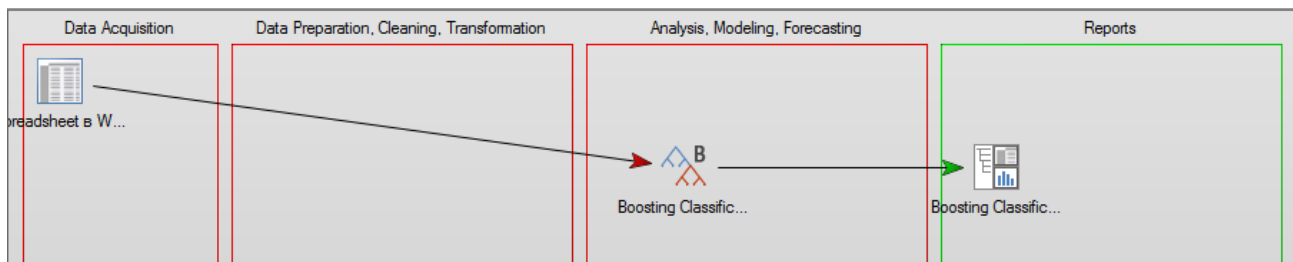


Рис. 1 - Графік значимості предикторів для залежної змінної «кредитоздатність»

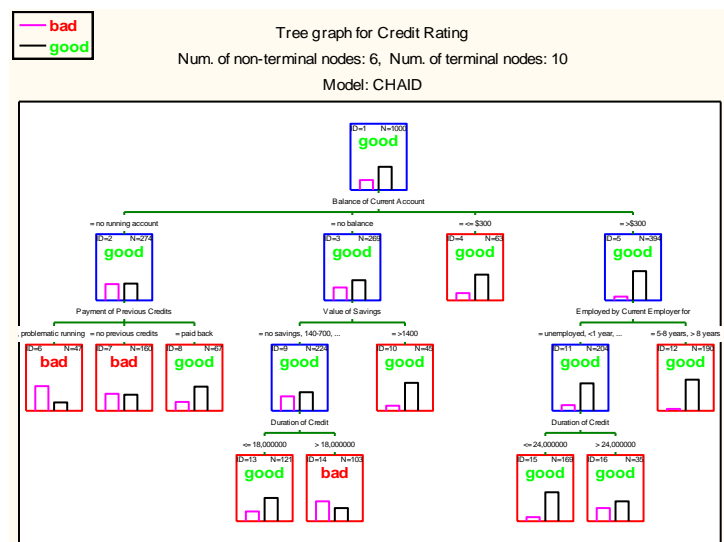


Рис. 2 - Приклад структури робочої області DataMiner

Classification matrix (CreditScoring.sta in Workbook1. Response: Credit Rating Model: CHAID				
	Observed	Predicted bad	Predicted good	Row Total
Number	bad	179	121	300
Column Percentage		57.74%	17.54%	
Row Percentage		59.67%	40.33%	
Total Percentage		17.90%	12.10%	30.00%
Number	good	131	569	700
Column Percentage		42.26%	82.46%	
Row Percentage		18.71%	81.29%	
Total Percentage		13.10%	56.90%	70.00%
Count	All Group:	310	690	1000
Total Percent		31.00%	69.00%	

Рис. 3 - Дерево рішень CHAID для кредитоздатності

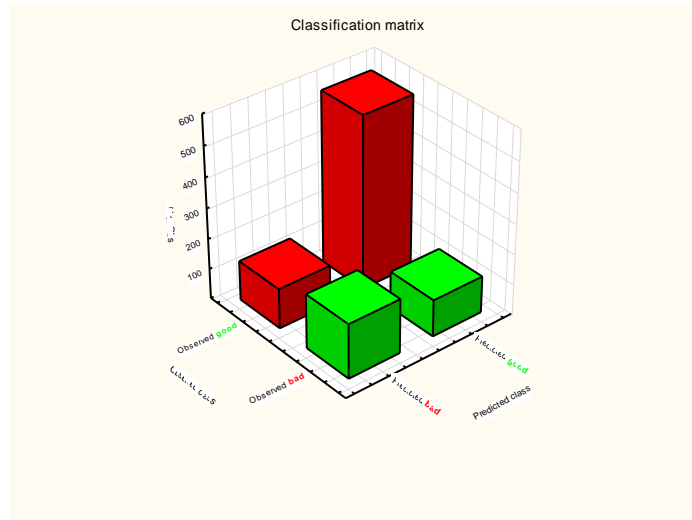


Рис. 4 - Матриця класифікації

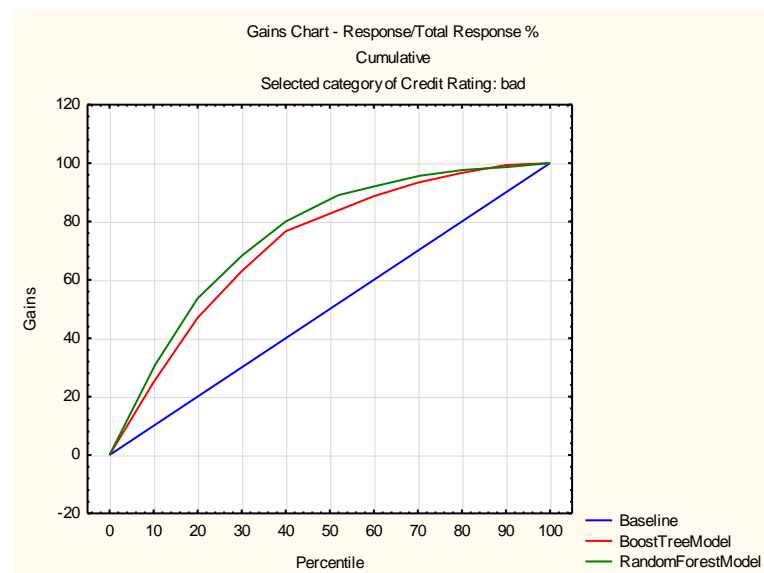


Рис. 4 - Візуалізація матриці класифікації

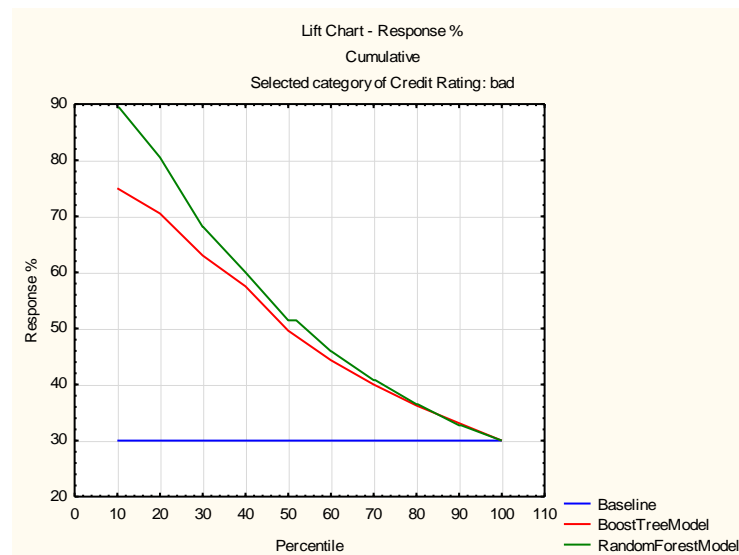


Рис. 5 - Карта виграшів для кредитоздатності = погано

Summary of Deployment (Error rates) (CreditScoring.sta in Workbook1.stw)						
	BoostTreeModel	RandomForest Model				
Error rate	0,270000	0,255000				

Рис. 6 - Загальний результат помилок

СПОЖИВЧИЙ КРЕДИТНИЙ СКОРИНГ

1. Побудова моделі

Відкриємо CreditScoring2.sta. Застосуємо аналіз нейронних мереж. Основною ціллю буде тип ризику, несуттєвими даними – вік клієнта, а важливими даними – Number of Loans (к-сть займів), How Paid (скільки заплатили), Monthly Income (Місячний прибуток), Income Range (діапазон прибутків), Mortgage (іпотека).

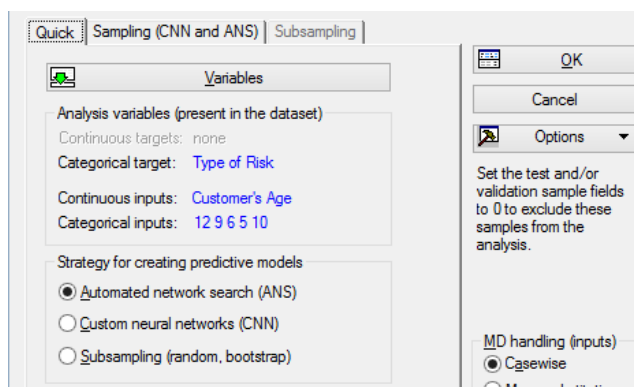


Рис. 1.1 – Налаштування інструменту «Побудова моделі»

2. Налаштування параметрів

Задамо значення мінімуму і максимуму нейронів: 2 і 13 відповідно. Вкажемо кількість мереж для навчання – 100. А для збереження – 5.

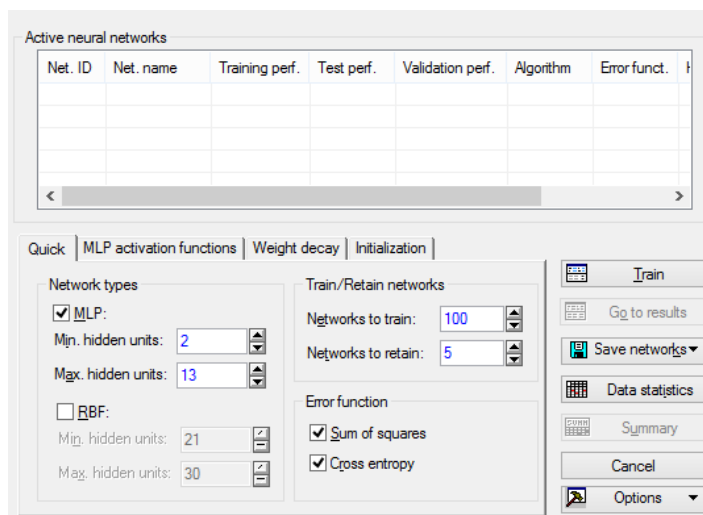


Рис. 2.1 – Завдання параметрів

Отримали результати: точність моделі – 65%.

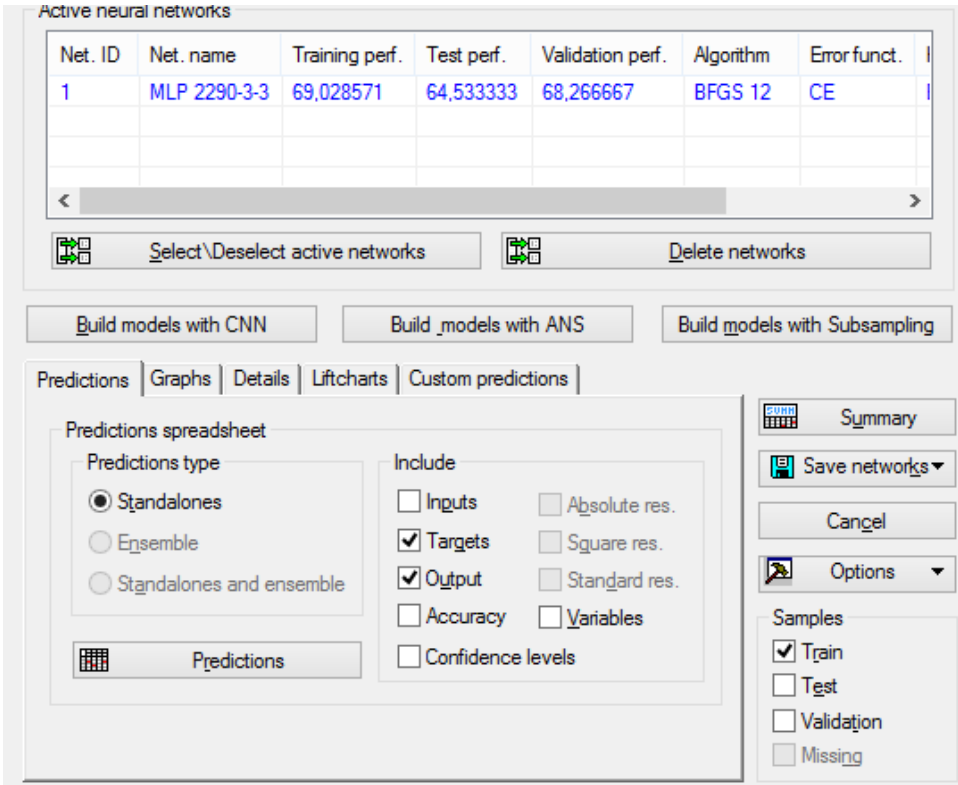


Рис. 2.2 – Отримання результатів з похибками

Отримали наступні прогнози.

Case name	Type of Risk Target	Type of Risk - Output 1. MLP 2290-3-3
105	Good Profit	Good Profit
106	Good Profit	Good Profit
107	Bad Loss	Good Profit
108	Good Profit	Good Profit
109	Bad Profit	Bad Loss
110	Bad Profit	Good Profit
111	Bad Loss	Good Profit
112	Good Profit	Good Profit
115	Bad Loss	Good Profit
117	Bad Loss	Good Profit
119	Good Profit	Good Profit
120	Bad Loss	Good Profit
121	Bad Profit	Bad Loss
122	Bad Loss	Good Profit
125	Good Profit	Good Profit
127	Bad Profit	Bad Loss
128	Bad Profit	Good Profit
129	Good Profit	Good Profit
130	Good Profit	Good Profit
131	Good Profit	Good Profit
132	Bad Profit	Bad Loss
133	Good Profit	Good Profit
135	Bad Profit	Bad Loss
136	Good Profit	Good Profit
137	Bad Profit	Good Profit

Рис. 2.3 – Прогноз

3. Виконання аналогічних налаштувань моделі для CreditRisk.sta

В даному випадку цільовим є Credit Standing. Важливими є статус активів, кредитна історія, кількість місяців вкладу активів, збережені активи.

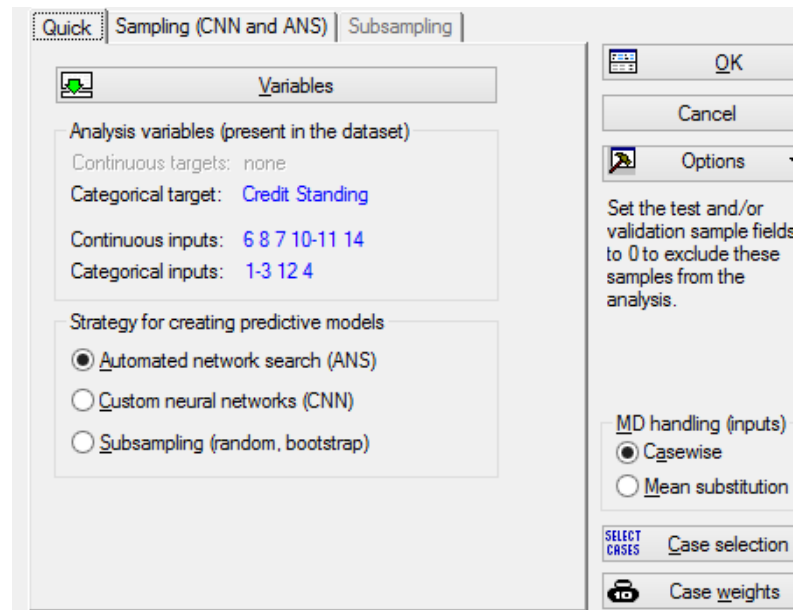


Рис. 3.1 – Завдання параметрів

Отже, точність моделі – 70%.

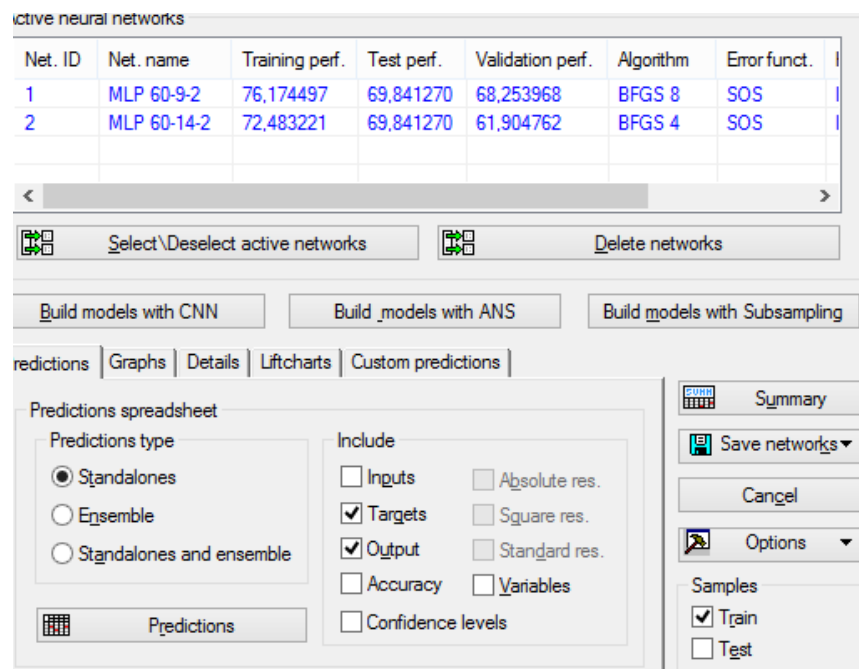


Рис. 3.2 – Отримання результатів з похибками

Case name	Predictions spreadsheet for Credit Standing (CreditRisk.sta)		
	Samples: Train		
	Credit Standing Target	Credit Standing - Output 1. MLP 60-9-2	Credit Standing - Output 2. MLP 60-14-2
2	Bad	Bad	Bad
3	Bad	Bad	Good
4	Bad	Good	Good
5	Good	Bad	Bad
6	Good	Good	Good
8	Good	Good	Good
10	Bad	Good	Good
11	Bad	Bad	Good
12	Good	Good	Good
13	Good	Bad	Bad
14	Bad	Bad	Bad
15	Good	Good	Good
16	Good	Good	Bad
17	Bad	Bad	Bad
18	Good	Bad	Bad
19	Good	Good	Good
22	Bad	Bad	Bad
24	Bad	Good	Good
25	Good	Good	Good
27	Bad	Bad	Good
28	Good	Good	Good
29	Good	Bad	Bad
31	Bad	Bad	Bad
32	Good	Good	Good
34	Good	Good	Good

Рис. 3.3 – Прогноз

ВИСНОВОК

Під час виконання даної лабораторної роботи я навчилася використовувати інструменти Data Mining та застосувала набуті навички на побудові споживчої кредитної оцінки ризиків (скорингу).