

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО”
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Тема: Скорингові моделі для оцінки кредитоспроможності
позичальників банку

Виконала :

студентка групи КА-44

Куца Каріна Володимирівна

Науковий керівник:

к.т.н., доц. Кузнецова

Наталія Володимирівна

Київ-2018

Мета дослідження

аналіз та порівняння скорингових моделей на основі дерев рішень та логістичної регресії для оцінки кредитоспроможності позичальників банку

Об'єкт дослідження

кредитні ризики, які виникають в банківській діяльності

Предмет дослідження

скорингові моделі на основі дерев рішень та логістичної регресії для прогнозування неповернення кредиту позичальником банку

Актуальність дослідження

Необхідність побудови найсучасніших аналітичних систем на основі скорингових моделей оцінювання ризиків, які будуть аналізувати кредитоспроможність позичальників банку.

Постановка задачі дослідження

Виконати розгляд методології аналізу даних.

Розглянути існуючі методи скорингових оцінок та провести їх аналіз.

Розробити скорингові моделі для оцінки кредитоспроможності позичальників банку на основі методів дерев рішень та логістичної регресії.

На основі обраної кращої моделі розробити скорингову карту, яка дозволить оцінити кількість скорингових балів за кожну характеристику клієнту і стане зручним інструментом для використання на робочих місцях.

Що таке кредитний скоринг?

Кредитний скоринг — це система оцінки кредитоспроможності позичальника, яка на основі кредитної історії банку оцінює ймовірність дефолту потенційного позичальника, виходячи з його соціально-демографічних характеристик.



Програмне забезпечення для вирішення задач інтелектуального аналізу даних



Методи, які використовуються для побудови прогнозуючих моделей при кредитуванні позичальників банку

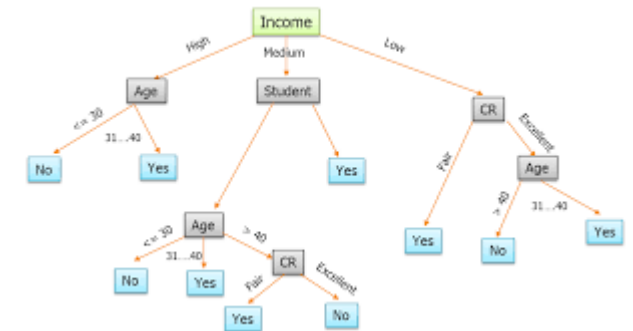
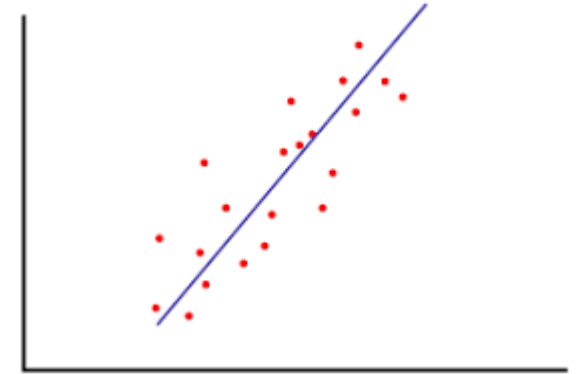
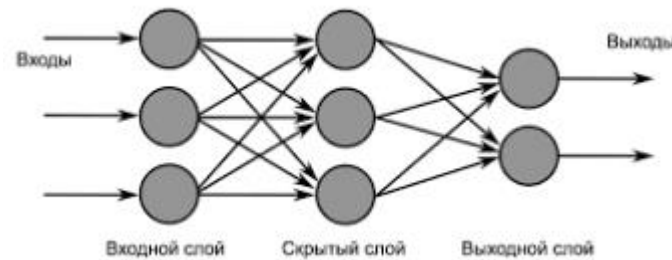
Логістична регресія

Лінійна регресія

Дерева рішень

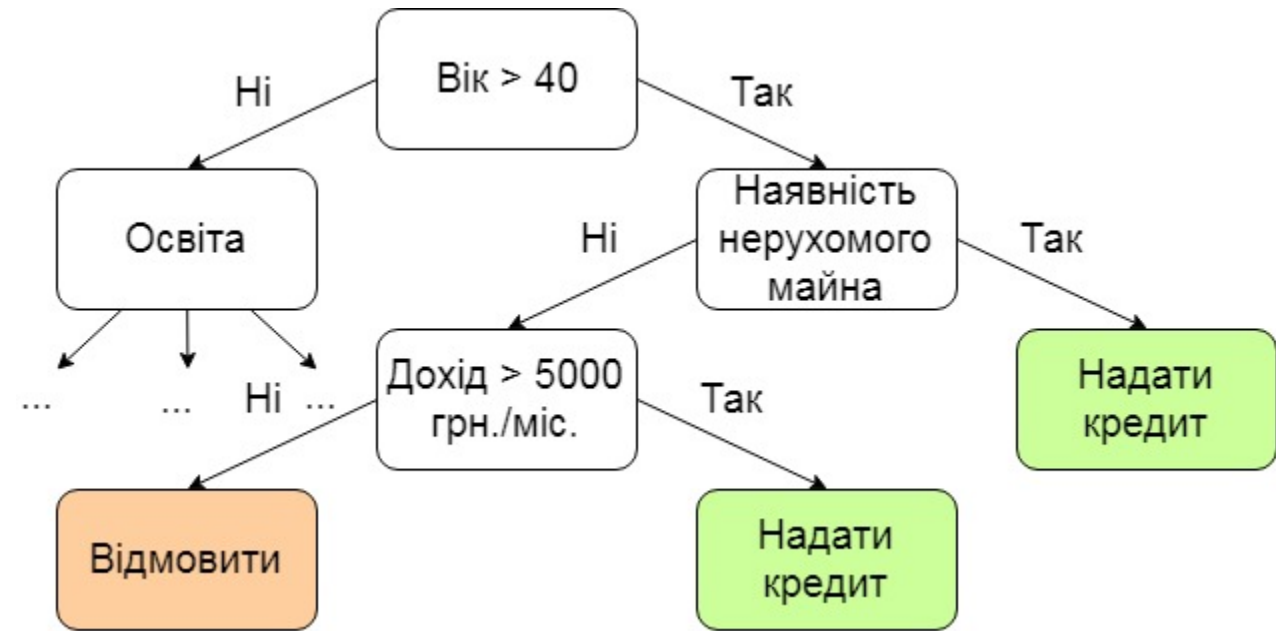
Нейронні мережі

Мережі Байєса



Дерева рішень

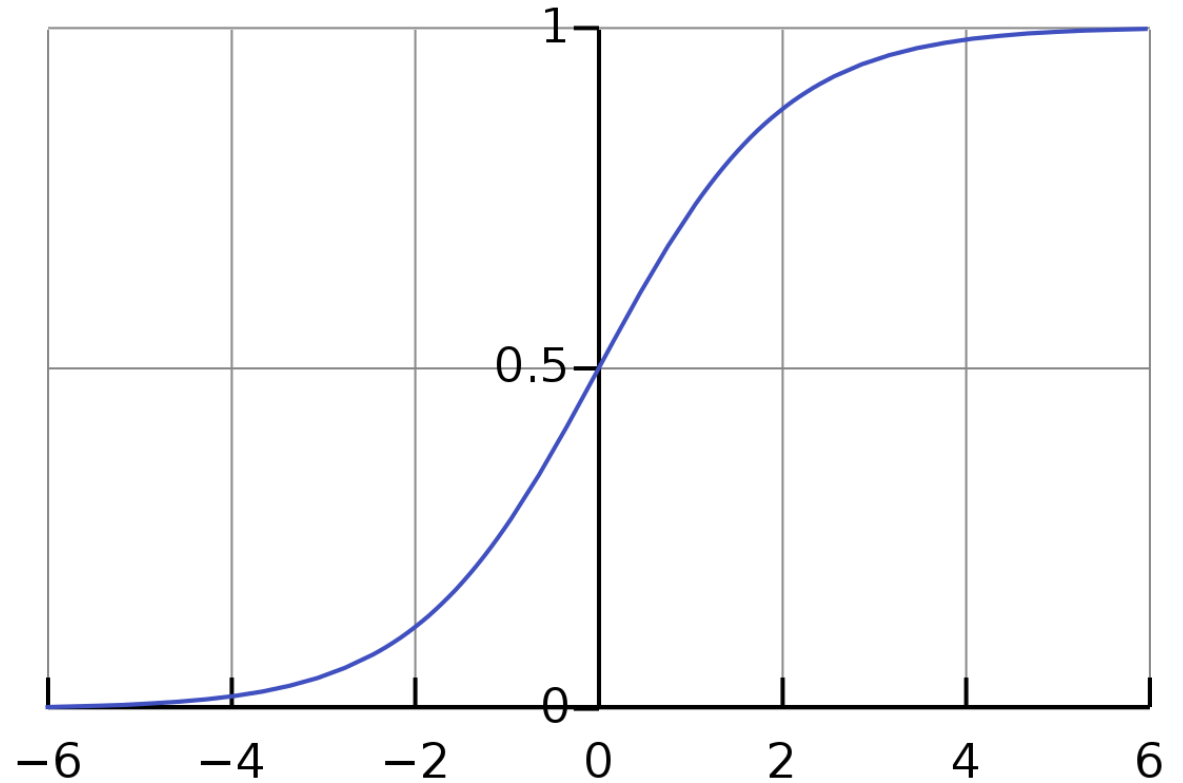
Дерева рішень - це спосіб представлення правил в ієрархічній, послідовній структурі, де кожному об'єкту відповідає єдиний вузол, що дає рішення.



Логістична регресія

Логістична регресія використовується для оцінки ймовірності того, що подія, яка нас цікавить буде мати значення:

$$p = \frac{1}{1 + e^{-y}}$$



Оцінка значущості змінних

IV - це показник, який використовують для оцінки взаємозв'язку між незалежними змінними і бінарною залежною. Він розраховується за наступною формулою:

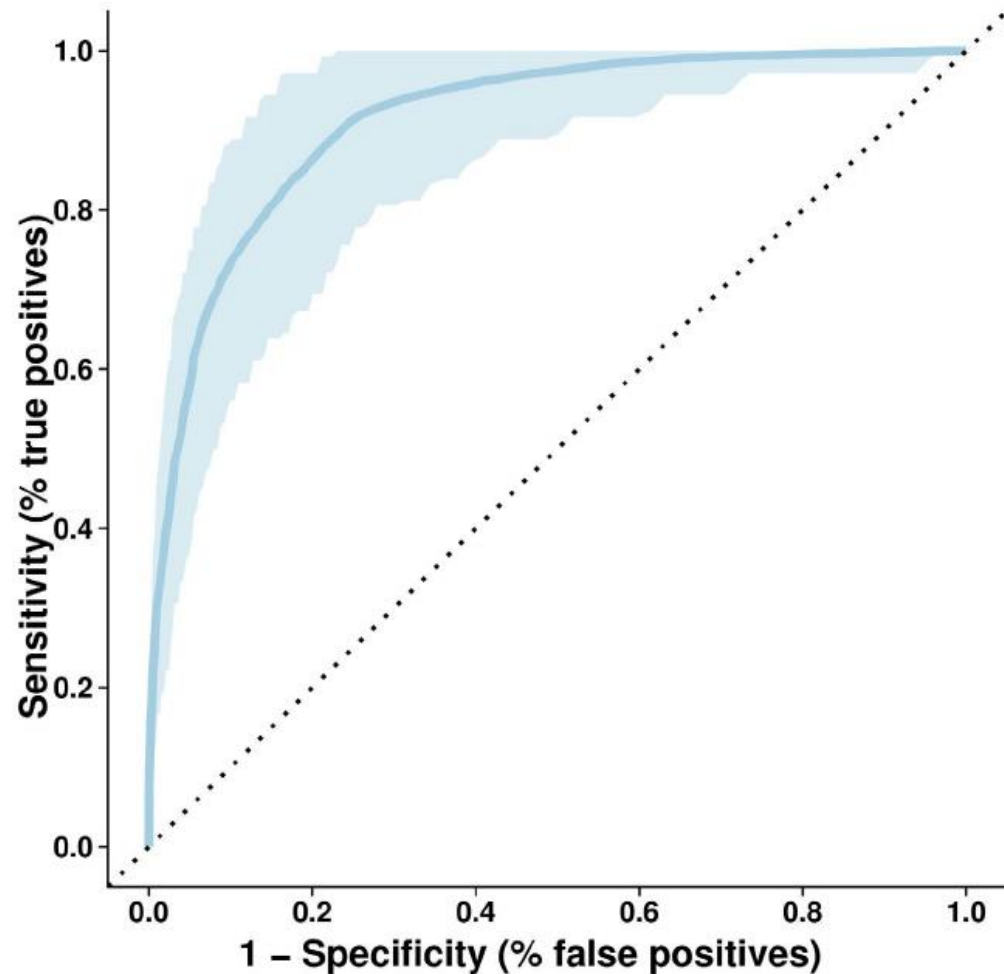
$$IV = \sum_{i=1}^k (G_i - B_i) WOE_i$$

G_i – процент всіх хороших клієнтів i -ої категорії

B_i – процент всіх поганих клієнтів i -ої категорії

Значення IV	Прогнозна здатність
< 0.02	Не має
0.02 – 0.1	Низька
0.1 – 0.3	Середня
0.3 – 0.5	Висока
> 0.5	Відмінна

Оцінки класифікаторів



ROC-крива—показує залежність частки вірно класифікованих позитивних прикладів від частки невірно класифікованих позитивних прикладів.

Оцінки класифікаторів

Accuracy - ймовірність того, що передбачуваний клас буде класифіковано правильно.

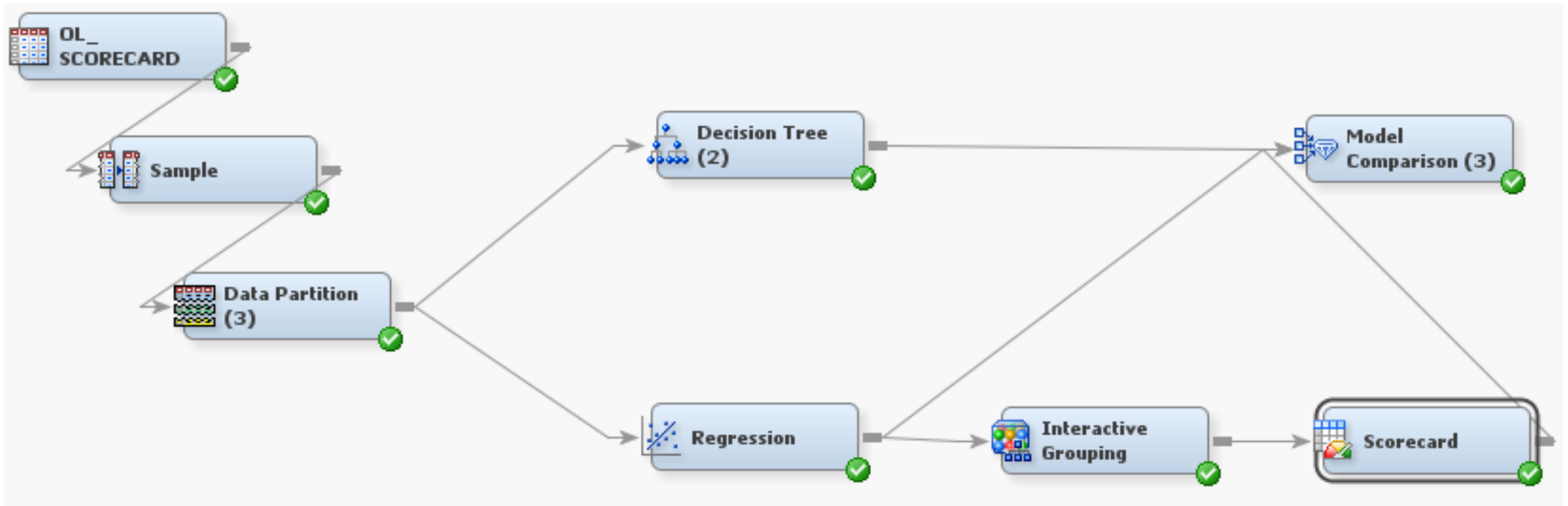
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Змінна	Опис
ID	ID позичальника
Work_experience	Стаж роботи у місяцях
Childrens	Кількість дітей в сім'ї
Num_workers	Кількість співробітників у компанії, в котрій працює позичальник
Education	Рівень освіти
Living_Region	Регіон, в якому проживає позичальник
Number_of_add_phones	Кількість вказаних номерів телефону
Gender	Стать
income1	Дохід
Family_status	Статус одруження
Time_Activation	Дата видачі кредиту
Time_Deactivation	Дата виплати кредиту
Bad	Статус дефолт/не дефолт

Опис набору даних

Для аналізу були обрані дані одного з банків України. Вибірка містить 30440 спостережень по видачі кредитів фізичним особам та 13 характеристик, одна з яких - відповідь на питання повернув позичальник кредит чи ні.

Технологічний процес



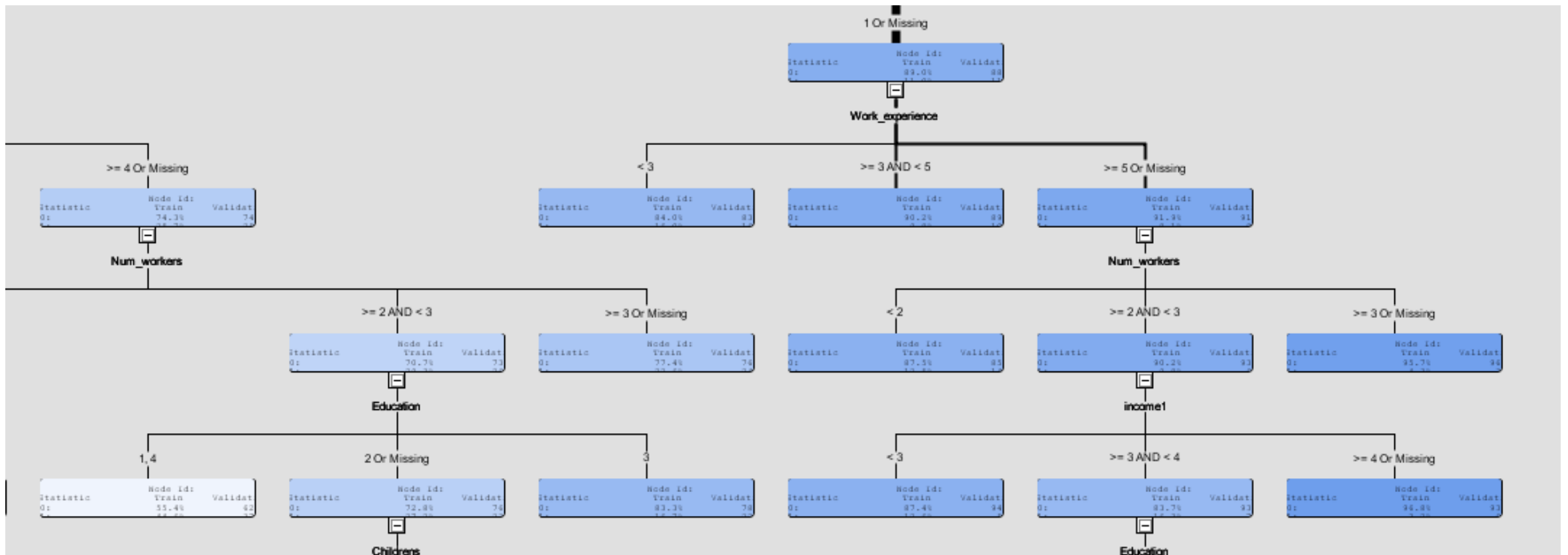
Параметри моделі логістичної регресії

Entered	Score Chi-Square
Family_status	780,85
Work_experience	156,47
Num_workers	130,08
Number_of_add_phones	75,25
Education	73,20
Childrens	53,67
income1	37,82
Living_Region	31,30

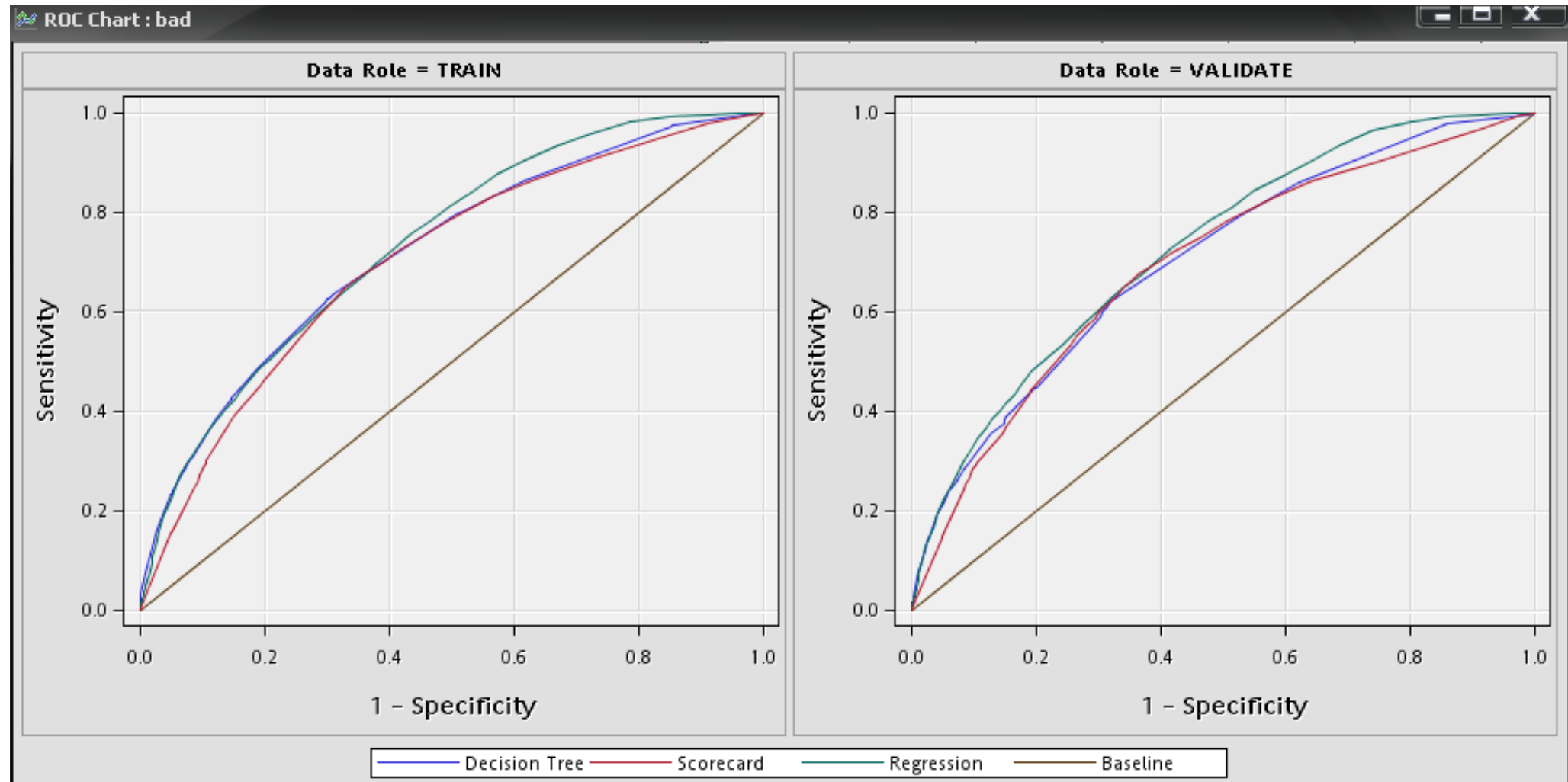
Значення важливості змінних в топології дерева рішень

NAME	IMPORTANCE
Family_status	1.0000
Work_experience	0.4696
Num_workers	0.4305
Number_of_add_phones	0.3182
Education	0.2560
Living_Region	0.2320
income1	0.2242
Childrens	0.1813
Gender	0.0473

Дерево рішень



ROC-криві моделей на навчальній і валідаційній вибірках



Порівняння результатів моделювання

На тренувальній

На тестовій

	Logistic Regression	Decision Tree	Logistic Regression	Decision Tree
Accuracy	0,813	0,817	0,811	0,817
AUC-ROC	0,741	0,701	0,725	0,672

Скорингова карта

Перед тим як приступити до впровадження розробленої скорингової карти, необхідно на зовнішніх даних перевірити точність розрахунку балів і адекватність карти. Задача полягає у тому, щоб підтвердити, що розроблена скорингова карта придатна для поточної популяції заявників.

Childrens	1, _MISSING_	27
	2	35
	3	39
	4	34
Family_status	2	13
	1, _MISSING_, _UNKNOWN_	51
Living_Region	2	32
	3	29
	4	29
	1, _MISSING_, _UNKNOWN_	36
Work_experience	2	26
	3	34
	4	38
	1	19
	5, _MISSING_	41

Висновки

Розглянуто існуючі методи скорингових оцінок та проведено їх аналіз.

Побудовано скорингові моделі оцінки кредитоспроможності позичальників банку на основі методів дерев рішень та логістичної регресії.

Розроблено скорингову карту на основі кращої моделі як ефективний інструмент для роботи кредитних спеціалістів.

Виконано порівняльний аналіз вищезазначених методів прогнозування на основі статистичних критеріїв.

Подальшими напрямками удосконалення роботи можуть бути питання, що стосуються:

Реалізації методів машинного навчання, наприклад, логістичної регресії, нейронної мережі або інших;

Розробка автоматизованого інструментарію для розрахунку скорингового балу та прийняття рішення щодо видачі кредиту з можливістю варіації порогового балу відсікання.

Статті

1. Кузнєцова Н.В., Куца К.В., Штогрін С.Р. Застосування методології аналізу виживання для дослідження споживчих ризиків. *Системні Науки Та Кібернетика*, Науковий електронний збірник НТУУ «КПІ». 2017. № 6. С. 126–136.

Участь у конференціях

- 1) Куца К.В., Штогрін С.Р. к.т.н., каф. ММСА Терентьев О. М. Інформаційна технологія прогнозування ціни золота. *Сучасні інформаційні технології 2017: Матеріали VIII міжнародної наукової конференції студентів та молодих вчених "Сучасні інформаційні технології 2017»*. Одеса: Одеський національний політехнічний університет, 2017. 81-82 с.
- 2) Куца К.В. Анализ кредитоспособности малого и среднего бизнеса в Украине с использованием нечеткой логики. *Системный анализ и информационные технологии: материалы 19-й Международной научно-технической конференции SAIT 2017*. Киев, 2017 г. 22 – 25 мая. УНК “ИПСА” НТУУ “КПИ им. Игоря Сикорского”. Киев: УНК “ИПСА” НТУУ “КПИ им. Игоря Сикорского”, 2017. 122 с.

Участь у конкурсах наукових робіт

Куца К.В. Оцінювання кредитоспроможності осіб за допомогою інтелектуального аналізу даних. Київ, 2017.

Дякую за увагу!