Математичні моделі інтелектуального аналізу даних для виявлення шахрайства з кредитними картками

Дипломна робота
студентки ННК “ІПСА” групи КА-44
Штогрін Світлани Романівни

Науковий керівник
к. т. н., доцент Кузнєцова Н. В.

Об'єкт дослідження:

Методи виявлення шахрайства з кредитними картками.

Предмет дослідження:

Моделі інтелектуального аналізу даних для виявлення шахрайства з кредитними картками.

Мета дослідження:

Побудувати та порівняти моделі виявлення шахрайства з кредитними картками.

Актуальність роботи

Згідно з інформацією НБУ у 2017 році зафіксовано 77,6 тис. випадків шахрайських дій з платіжними картками. За розрахунками Української міжбанківської асоціації членів платіжних систем ЕМА, у 2017-му році внаслідок вішингу (телефонне шахрайство з виманюванням реквізитів банківських карток і переказом коштів на карту злодіїв) з рахунків українців було вкрадено 509,72 млн. грн, а внаслідок фішингу (виманювання конфіденційних даних – паролів, номерів банкінських карток, PIN-кодів тощо) – 63,68 млн. грн. Загалом – 669,63 млн. грн. Для порівняння, у 2016 році шахраї викрали з рахунків українців 339,13 млн. грн, а у 2015 – 84,36 млн. грн.

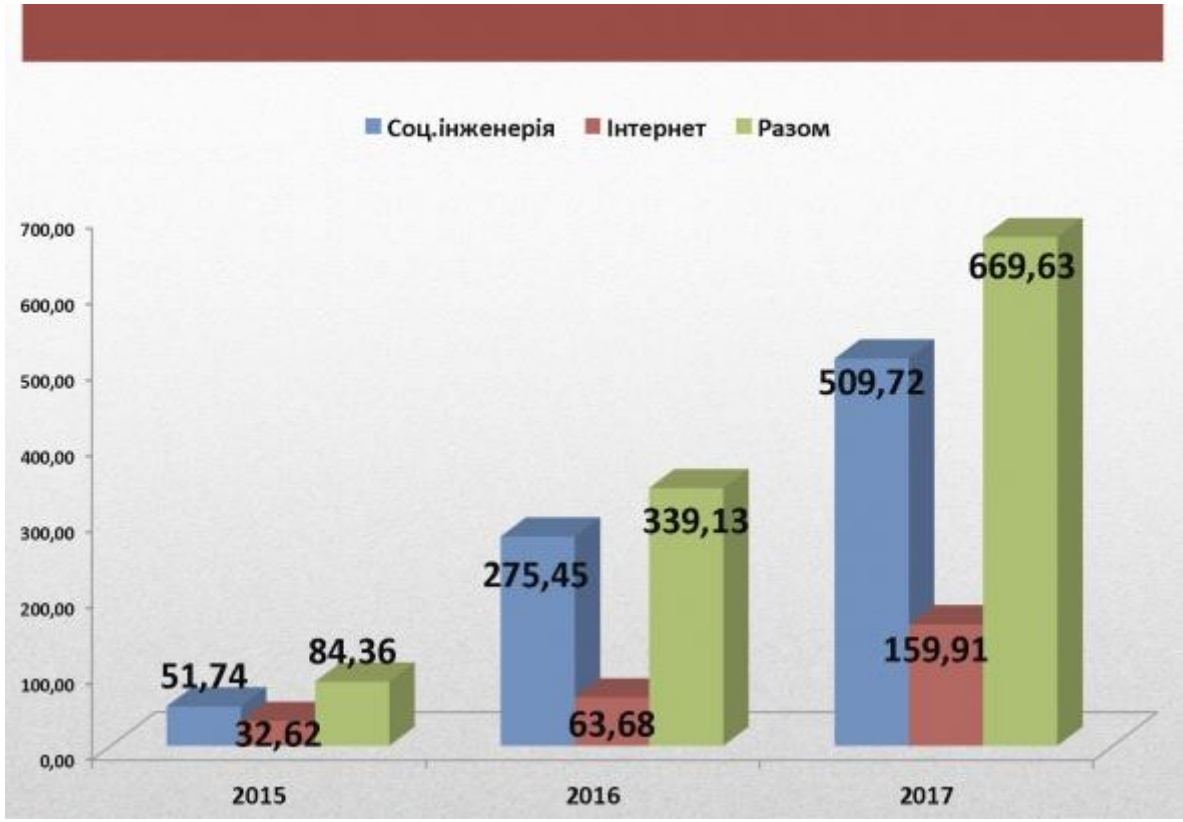


Рисунок 1 - Обсяги шахрайства з банківськими картками в Україні за інформацією ЄМА. Сукупний дохід шахраїв, млн. грн

Типи шахрайства з кредитними картками

Незаконне використання кредитної картки або її інформації без відома власника називається *шахрайством з кредитною карткою.*

Серед найбільш розповсюджених типів шахрайства можна виділити наступні:

Фишинг (англ. Phishing)

- Різновидами фішингу є: 1) фішингові сайти; 2) фішингові електронні листи; 3) фішингові смс-повідомлення.

Вішинг (англ. Vishing)

- Шахрайство з використанням телефонної комунікації: 1) виманювання карткових реквізитів; 2) стимулювання власника до зняття лімітів по карті, відключення перевірки CVV2 / CVC2-коду і т.д.

Кардинг (англ. carding)

- різновид шахрайства, при якому проводиться операція з використанням банківської картки або її реквізитів, яка не ініційована або не підтверджена її власником.

Скіммінг (англ. Skimming, Card Skimming)

- частковий випадок кардингу; незаконне копіювання інформації з магнітної смуги дійсної картки законного власника.

Шиммінг (англ. Shimming, Card Shimming)

- різновид скімінгу; встановлення шахраями всередину банкомату пристрою для копіювання даних з чіпа платіжних карт законних власників.

Методи виявлення шахрайства з кредитними картками

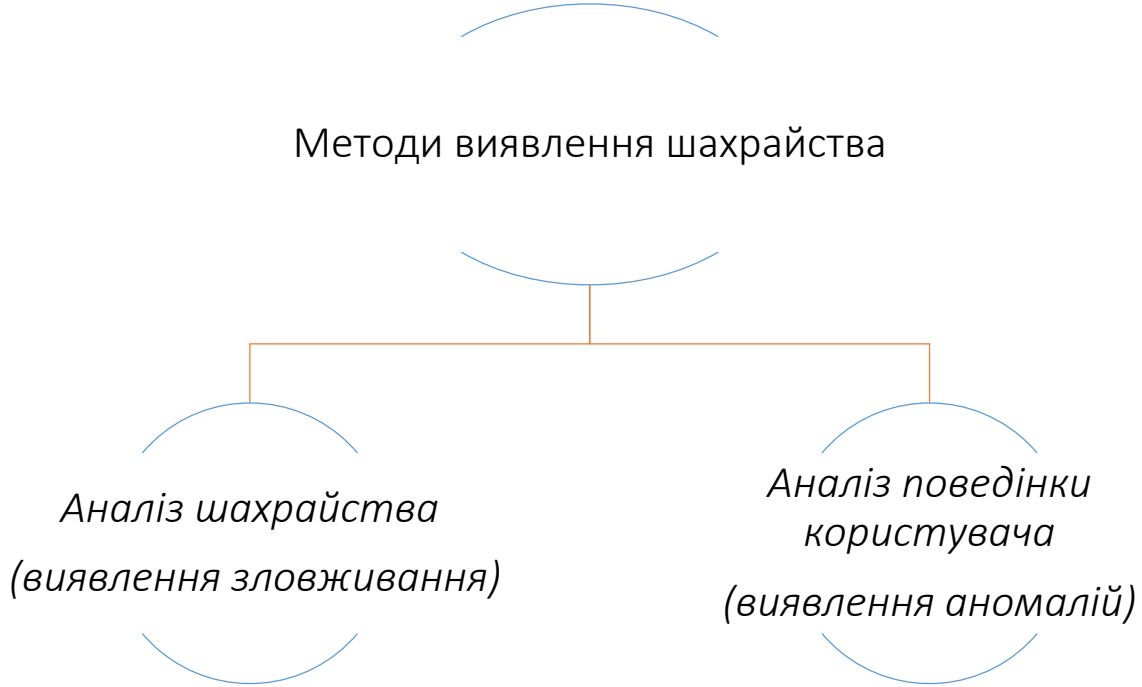


Рисунок 2 - Методи виявлення шахрайства з кредитними картками

Інструментарій

Потреба в якісній аналітичній інформації зростає у всіх галузях, через що фахівцям з інтелектуального аналізу даних і бізнес-аналітикам постійно доводиться створювати нові і більш ефективні моделі в більш стислі терміни.

SAS Enterprise Miner - потужне і повнофункціональне рішення для прогнозуючої аналітики і інтелектуального аналізу даних, що дозволяє створювати високоефективні прогнозуючі та описові моделі на основі величезної кількості інформації.

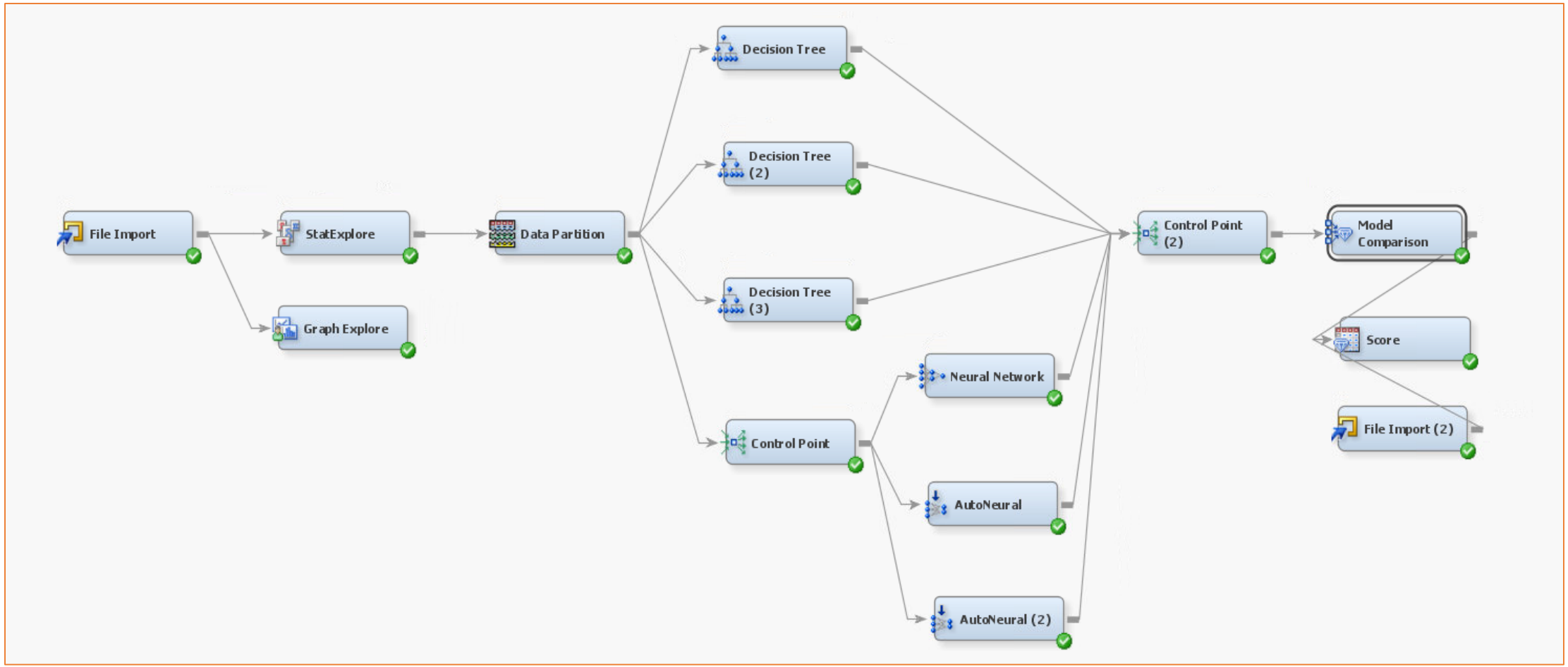
Опис набору даних

Набір даних містить транзакції, здійснені кредитними картками у вересні 2013 року власниками карток Європи. Цей набір даних представляє транзакції, які відбулися за два дні та містить 492 шахрайства з 284 807 операцій.

Змінна Class приймає значення 1 у випадку шахрайства і 0 в іншому випадку.

Time	Кількість секунд, що пройшли між першою та кожною транзакцією	Числовий
V1	Атрибут 1	Числовий
V2	Атрибут 2	Числовий
…	…	…
V28	Атрибут 28	Числовий
Amount	Сума грошей для цієї операції	
Class	Шахрайство/не шахрайство	Булевий

Технологічний процес



Штучні нейронні мережі

Дерева рішень

У роботі побудовано моделі дерев рішень та штучних нейронних мереж.

Оцінка якості побудованих моделей

Критерії оцінки:

- Помилки I та II роду
- ROC-крива та індекс GINI

	Прогноз моделі: Не шахрайство (0)		Прогноз моделі: Шахрайство (1)	
Фактично: Не шахрайство (0)	Вірно класифіковані (TN)		Помилки II роду (FP)	
	Дерево рішень Train: 170581 Validate: 56860	Нейронна мережа Train: 170542 Validate: 56846	Дерево рішень Train: 7 Validate: 3	Нейронна мережа Train: 46 Validate: 17
Фактично: Шахрайство (1)	Помилки I роду (FN)		Вірно класифіковані (TP)	
	Дерево рішень Train: 72 Validate: 20	Нейронна мережа Train: 61 Validate: 16	Дерево рішень Train: 223 Validate: 77	Нейронна мережа Train: 234 Validate: 81

Оцінка якості побудованих моделей

Критерії оцінки:

- Помилки I та II роду
- ROC-крива та індекс GINI

Кількісну інтерпретацію ROC-кривої дає показник **AUC** (англ. area under ROC curve, площа під ROC-кривою)

Чим ближче до 1 показник AUC, тим кращу прогностичну силу має модель.

GINI = 2*AUC - 1

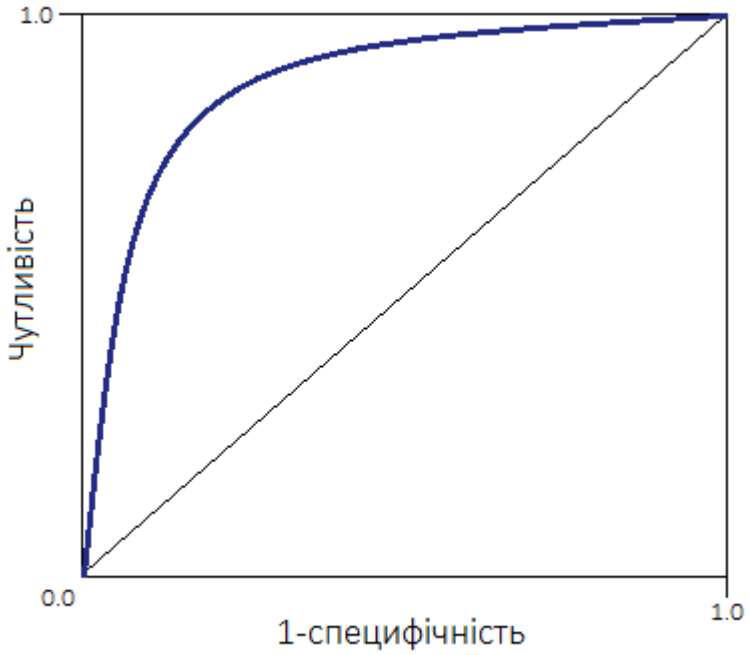


Рисунок 3 – ROC-крива

Оцінка якості побудованих моделей

Критерії оцінки:

- Помилки I та II роду
- ROC-крива та індекс GINI

Специфічність моделі:	$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = \frac{\mathrm{TN}+\mathrm{FP}-\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}} = 1 - \frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}} = 1 - \mathrm{FPR}$
Чутливість моделі:	$\mathrm{Se} = \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
FPR	*частина вірно позитивних прикладів (True Positives Rate)*
TPR	*частина помилково позитивних прикладів (False Positives Rate)*

Оцінка якості побудованих моделей

Критерії оцінки:

- Помилки I та II роду
- ROC-крива та індекс GINI

	ROC index			GINI		
	Train	Validate	Test	Train	Validate	Test
Дерево рішень	0.905	0.928	0.895	0.81	0.855	0.79
Нейронна мережа	0.986	0.975	0.98	0.972	0.95	0.96

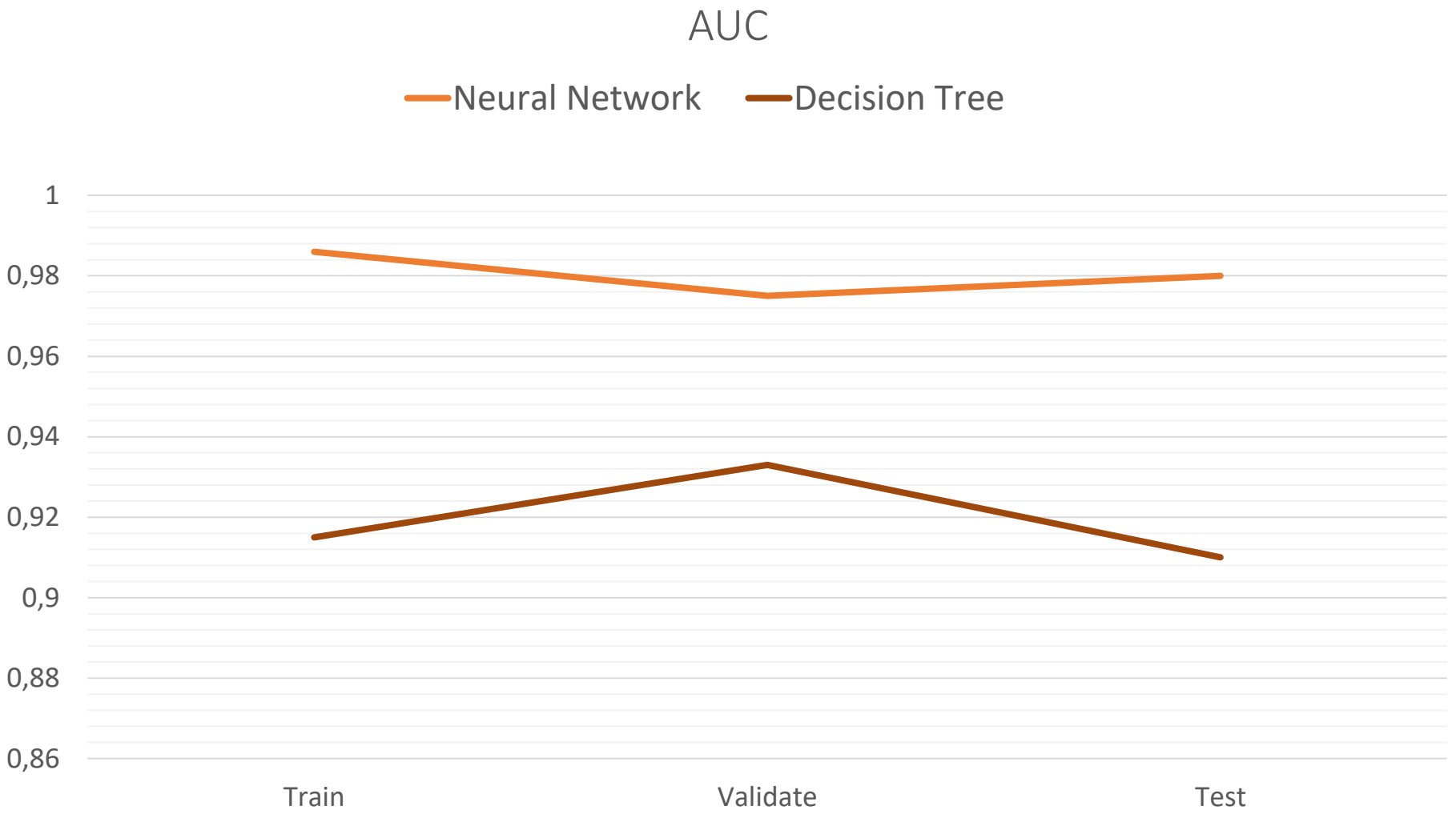


Рисунок 3 – Значення показника AUC для навчальної, тестової та перевірочної вибірок відповідних моделей

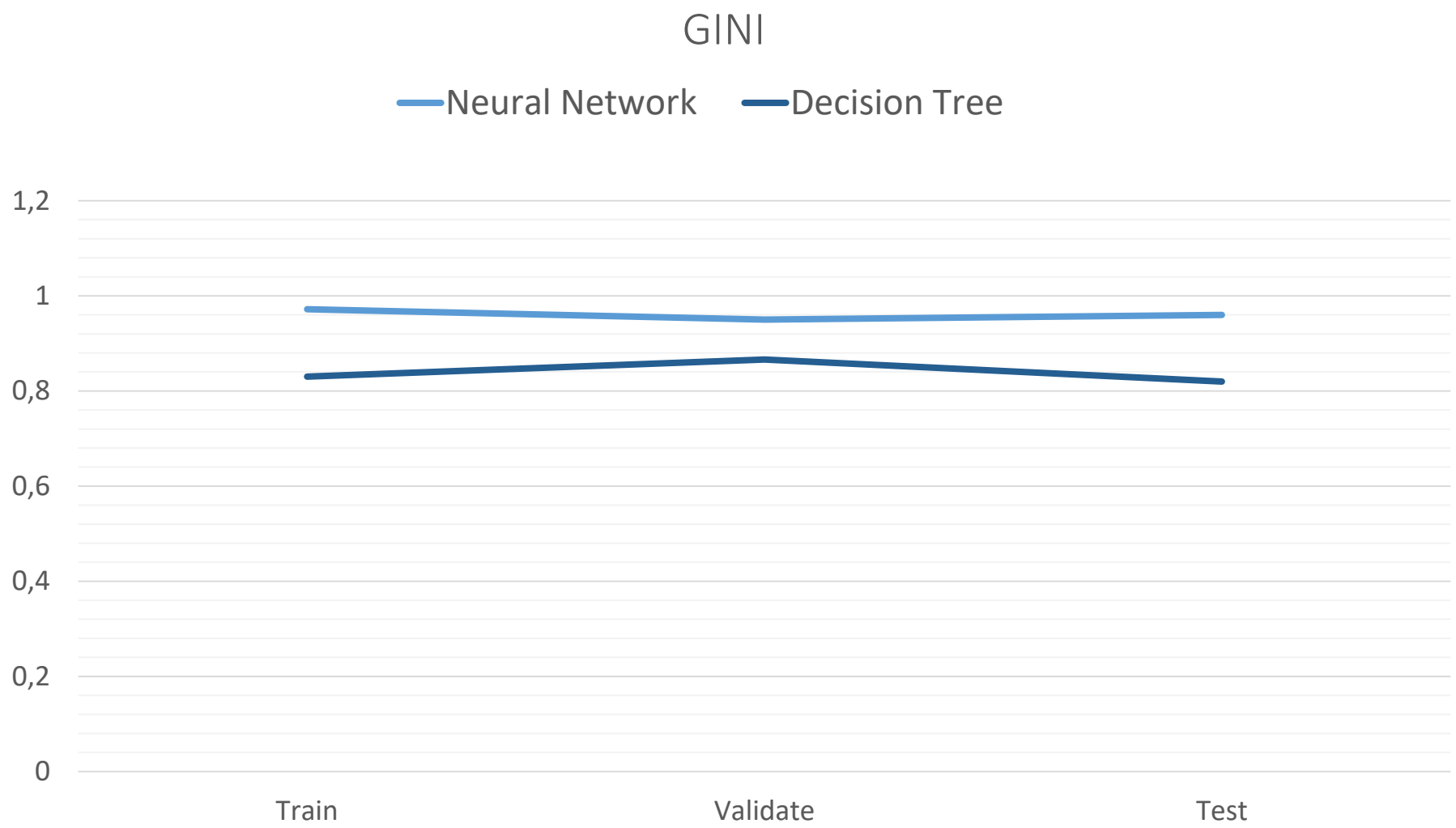
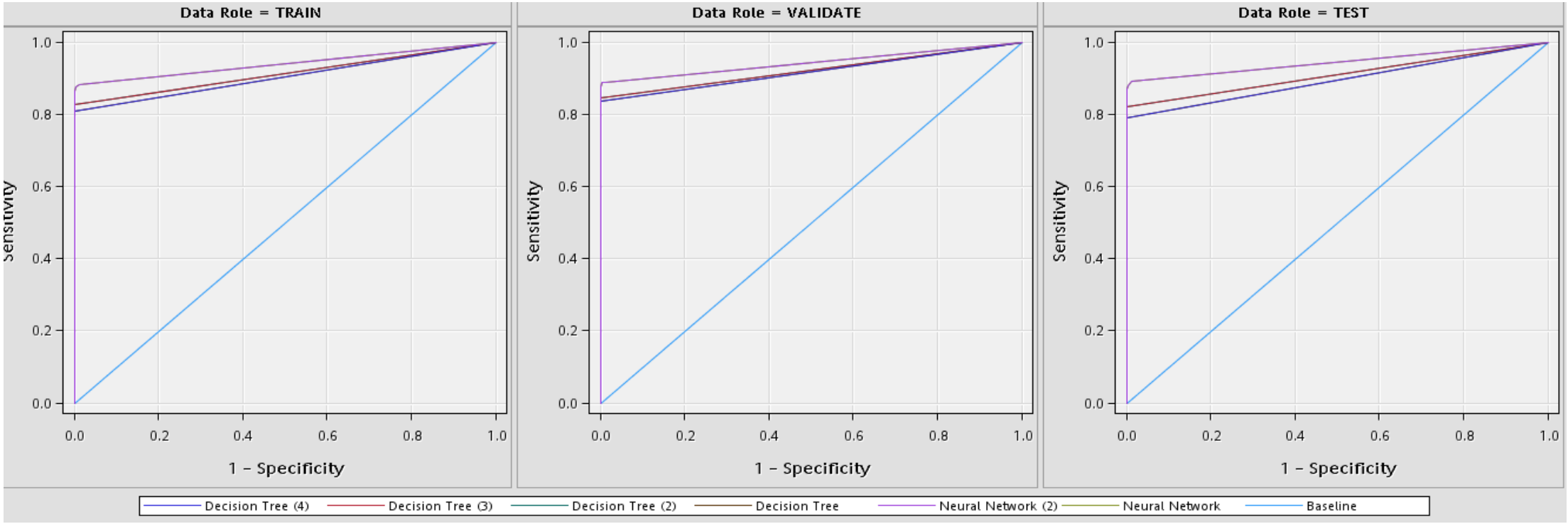


Рисунок 4 – Значення індексу GINI для навчальної, тестової та перевірочної вибірок відповідних моделей

Вибір кращої моделі



На основі ROC-індексу кращою моделлю виявилася модель нейронної мережі.

Висновки

У роботі розглянуто:

- основні типи шахрайства з кредитними картками;
- основні методи виявлення шахрайства з кредитними картками.

Побудовано моделі дерев рішень і штучних нейронних мереж для виявлення шахрайства з кредитними картками та проведено їх аналіз.

Модель штучної нейронної мережі показала кращу якість класифікації, аніж модель дерева рішень.

Рекомендації щодо подальшої роботи

Розробка і аналіз моделей виявлення шахрайства з кредитними картками у реальному часі.

Публікації

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.126-135. URL: http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_6_2017.pdf (дата звернення: 27.05.2018).
2. Куца К.В., Штогрін С.Р., к.т.н. каф. ММСА Терентьєв О. М. Інформаційна технологія прогнозування ціни золота. *Сучасні інформаційні технології 2017:* матеріали VIII міжнародної наукової конференції студентів та молодих вчених "Сучасні інформаційні технології 2017». О.: Одеський національний політехнічний університет, 2017. С. 81-82.

Дякую за увагу!