

Міністерство освіти і науки України
Національний технічний університет України «Київський Політехнічний інститут»
Навчально-науковий комплекс «Інститут прикладного системного аналізу»
Кафедра математичних методів системного аналізу

Система для оцінювання і менеджменту банківських кредитних ризиків

Виконав: студент групи КА-23 Роговий А.В.

Науковий керівник: проф. Бідюк П.І.

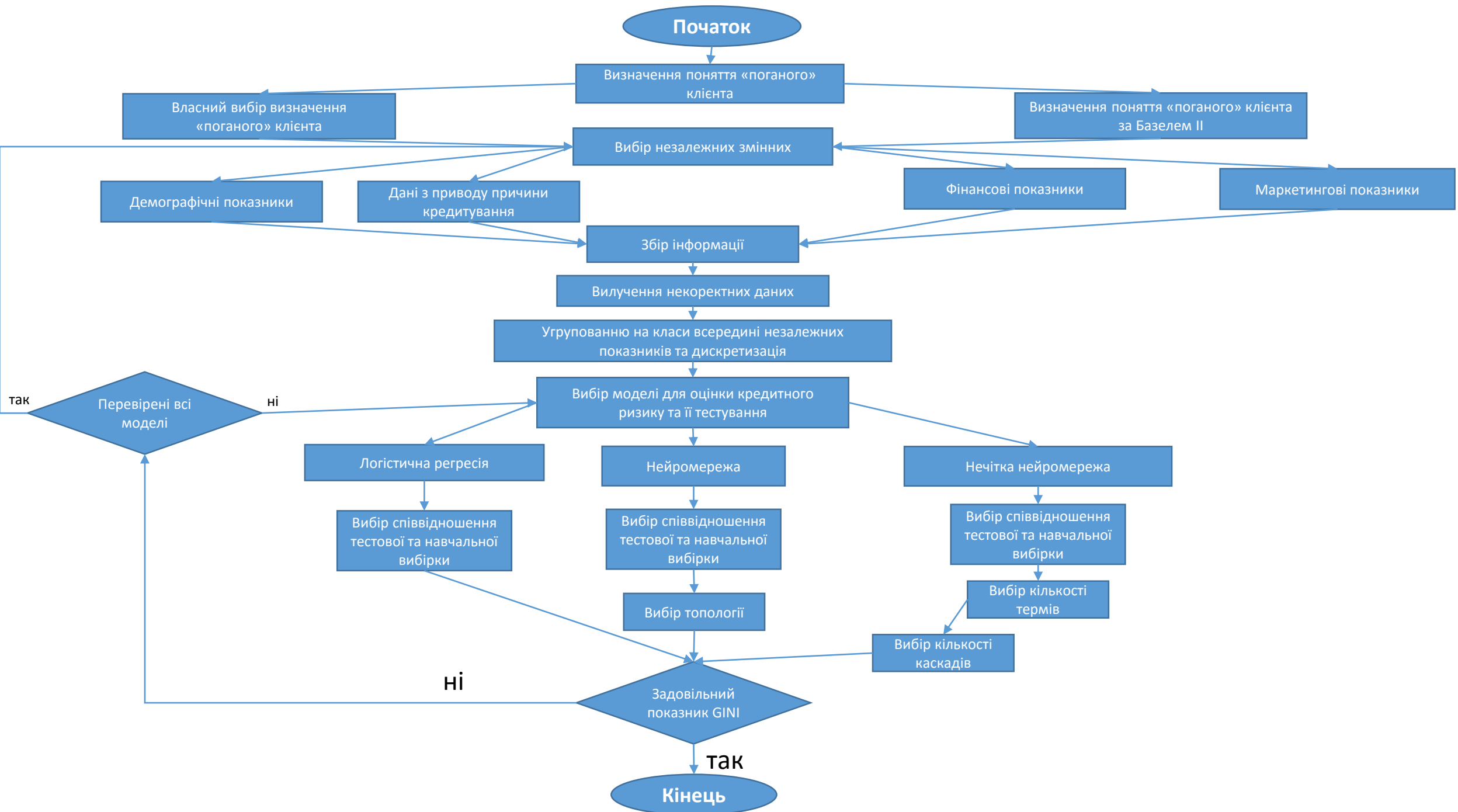
Київ 2016

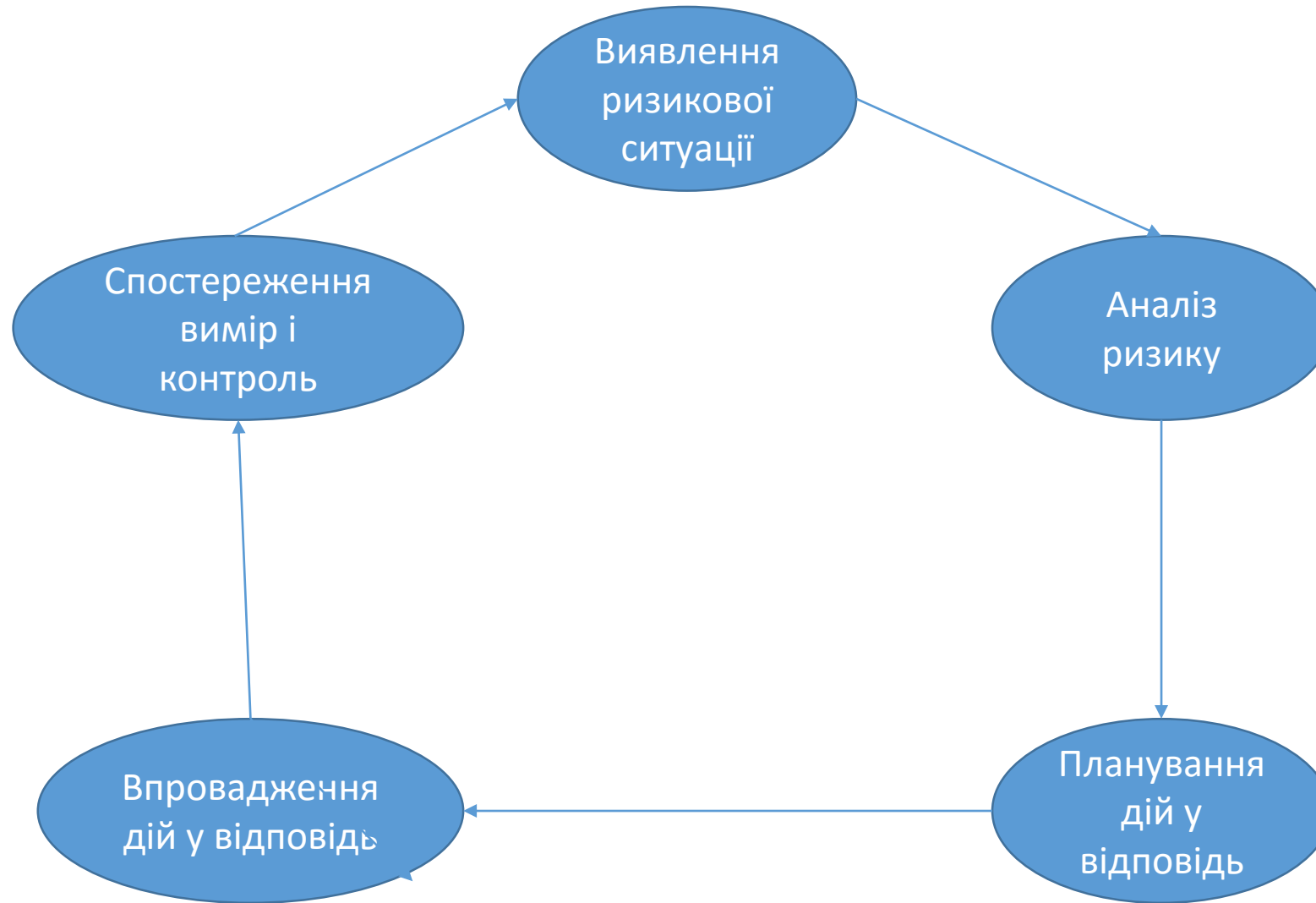
Актуальність теми, методи і мета дослідження

- **Актуальність теми** – у банківській сфері при управлінні кредитними ризиками одним з ключових завдань є оцінка кредитоспроможності позичальників. Незважаючи на великий світовий досвід кредитування осіб, побудова систем для оцінювання і менеджменту кредитних ризиків не має чітко алгоритмізованого підходу. Тому дана предметна область потребує окрім необхідного багажу знань, також наявність нових ідей та методів
- **Мета роботи** – розробка системи для оцінювання і менеджменту банківських кредитних ризиків і для оцінювання кредитного ризику за допомогою створеного програмного продукту.
- **Методи дослідження** – логістична регресія, нейронна мережа зі зворотнім поширенням похибки та нечітка нео-фаззі каскадна нейронна мережа.
- **Отримані результати** – створена система для оцінювання та менеджменту кредитного ризику, також створено програмний продукт, який дає можливість користувачу отримати кредитний прогноз за допомогою декількох моделей, а також порівняти їхні результати. Був проведений порівняльний аналіз результатів роботи декількох моделей.

Постановка задачі

- 1) Розробити систему для аналізу та менеджменту кредитних ризиків
- 2) Обробка первинних даних
- 3) Вибрати методи моделювання ризику неповернення кредиту фізичними особами.
- 4) Розробити програмний продукт для аналізу кредитоспроможності фізичних осіб і застосувати його до фактичних даних.
- 5) Виконати аналіз отриманих результатів





Методи аналізу якості моделі

Основні критерії оцінки:

- 1) загальна точність моделі;
- 2) помилки I-го та II-го роду;
- 3) ROC -крива та індекси AUC та GINI.

Загальна точність моделі (CA – common accuracy) визначається як:

$$CA = \frac{CorrectForecast}{N},$$

де *CorrectForecast* – кількість вірно спрогнозованих випадків, а *N* – загальна кількість випадків. Загальна точність моделі є дещо суб'єктивною оцінкою, оскільки вона залежить від долі дефолтів у моделі, а також від порогу відсікання. Для різних значень порогу точність моделі також буде приймати різні значення.

Помилки I-го та II-го роду

	Прогноз моделі: Повернення кредиту (0)	Прогноз моделі: Дефолт (1)
Фактично: Повернення кредиту (0)	Вірно класифіковані (TP)	Помилки II-го роду (FN)
Фактично: Дефолт (1)	Помилки I-го роду (FP)	Вірно класифіковані (TN)

Відносні показники у процентах:

➤ частка істинно позитивних прикладів (True Positives Rate):

$$TPR = \frac{TP}{TP + FN}$$

➤ частка хибно позитивних прикладів (False Positives Rate):

$$FPR = \frac{FP}{TN + FP}$$

Чутливість моделі – це частка істинно позитивних випадків, тобто

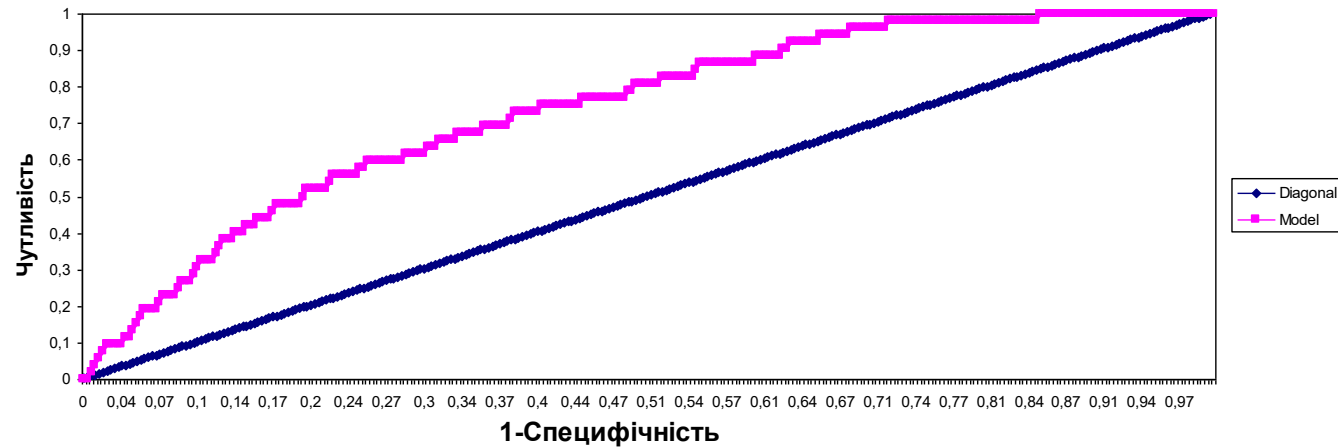
$$Se = TPR = \frac{TP}{TP + FN}$$

Специфічність моделі – це частка істинно негативних випадків, які були вірно класифіковані моделлю:

$$Sp = \frac{TN}{TN + FP}$$

$$Sp = \frac{TN + FP - FP}{TN + FP} = 1 - \frac{FP}{TN + FP} = 1 - FPR$$

ROC - крива



ROC-крива для задачі кредитування

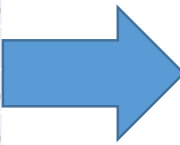
Чисельний показник площі під кривою AUC (Area Under Curve) і обчислюється, наприклад, за методом трапецій:

$$AUC = \int f(x)dx = \sum_i \left[\frac{Y_{i+1} + Y_i}{2} \right] \cdot (X_{i+1} - X_i)$$
$$GINI = 2 \cdot AUC - 1$$

Об'єкт дослідження

<u>Ім'я змінних</u>	<u>Пояснення</u>	<u>Тип змінних</u>
<i>Age years</i>	Вік позичальника в роках	integer
<i>Sex</i>	Стать особи	female/male
<i>Education</i>	Освіта особи	string
<i>Marital status</i>	Сімейний статус	string
<i>Credit</i>	Сума взятого позичальником кредиту	real
<i>Region</i>	Регіон проживання	integer
<i>PrevEmployerSector</i>	Сектор зайнятості особи	string
<i>OccupationStatus</i>	Професія	string
<i>WorkExperienceMonth</i>	Кількість місяців стажу .	integer
Good/Bad	Статус повернення кредиту	good/bad

Спеціальність	Загальна сума	Кількість дефолтів
автоелектрик	1	0
Авто-механик	4	0
автомийник	3	0
Автомойщик	1	0
автослесарь	5	3
авто-слесарь	2	0
Автослюсар	1	0
агент комерційний	1	0
агент по нерухомості	1	0
адміністратор торгового зала	5	1
Адміністратор	11	6
....
Юрист	1	1

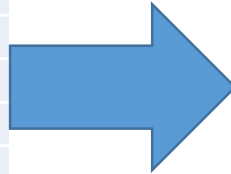


Спеціальність	Загальна сума	Кількість дефолтів	Bad rate	Weight of Evidence
автоелектрик	1	0	0	0
Авто-механик	4	0	0	0
автомийник	3	0	0	0
Автомойщик	1	0	0	0
автослесарь	5	3	0.6	0.405465
авто-слесарь	2	0	0	0
Автослюсар	1	0	0	0
агент комерційний	1	0	0	0
агент по нерухомості	1	0	0	0
адміністратор торгового зала	5	1	0.2	-1.38629
Адміністратор	11	6	0.54545	0.182322
....
Юрист	1	1	1	0



Спеціальність	Загальна сума	Кількість дефолтів	Bad rate	Weight of Evidence
Employee	1389	1313	0.945284	2.849337
Government officer	61	60	0.983607	4.094345
Military	9	9	1	0
Owner/coowner	4	4	1	0
Pensioner	137	131	0.956204	3.083438

Сфера діяльності	Загальна сума	Кількість дефолтів
агрегатный завод	1	0
Адміністратор	5	0
Ательє	1	0
Банки та фінансові установи	4	1
безпека	7	1
бильярдный клуб	1	0
благотворительность	1	0
Будівництво	44	15
Бульно-кондитерський комбінат	1	0
бухгалерия	9	3
бухгалтерия	2	0
Вагонобудівництво	4	2
....
Излтельство	1	1



Сфера діяльності	Загальна сума	Кількість дефолтів	Bad rate	Weight of Evidence
агрегатный завод	1	0	0	0
Адміністратор	5	0	0	0
Ательє	1	0	0	0
Банки та фінансові установи	4	1	0.25	-1.09861
безпека	7	1	0.1428	-1.79176
бильярдный клуб	1	0	0	0
благотворительность	1	0	0	0
Будівництво	44	15	0.3409	-0.65925
Бульно-кондитерський комбінат	1	0	0	0
бухгалерия	9	3	0.3333	-0.69315
бухгалтерия	2	0	0	0
Вагонобудівництво	4	2	0.5	0
....
Излтельство	1	1	1	0



Сфера діяльності	Загальна сума	Кількість дефолтів	Bad rate	Weight of Evidence
Agriculture	198	193	0.974747	3.653252
Building	201	190	0.945274	2.849129
Education	415	403	0.971084	3.51403
Finance	74	72	0.972973	3.583519
Manufacturing	181	173	0.955801	3.07385
Medicine	333	320	0.960961	3.203372
Mining	385	374	0.971429	3.526361
none	2521	2391	0.948433	2.911933
Other	2218	2095	0.944545	2.835124
Services	600	566	0.943333	2.812234
Trade	442	419	0.947964	2.902377

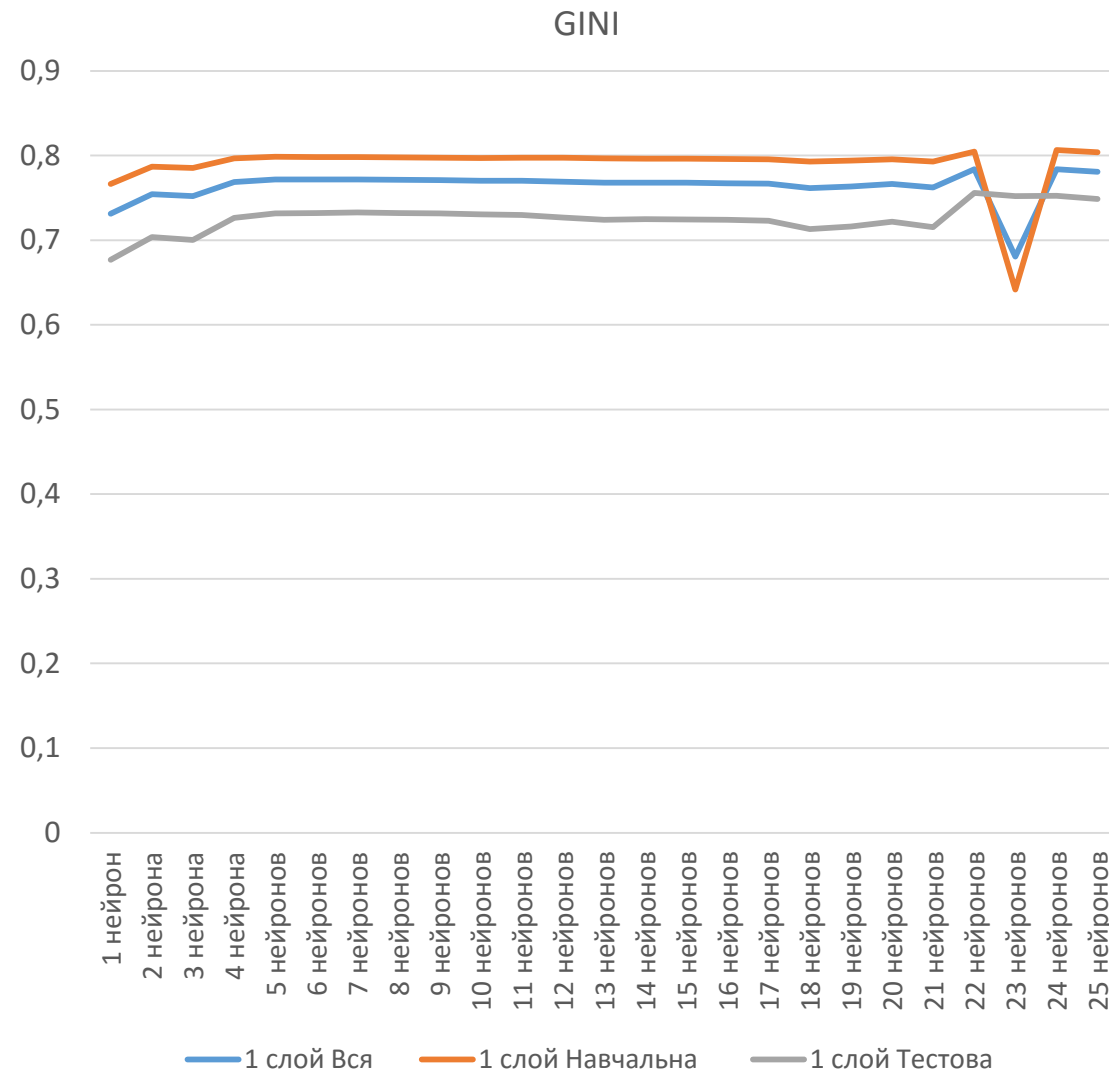
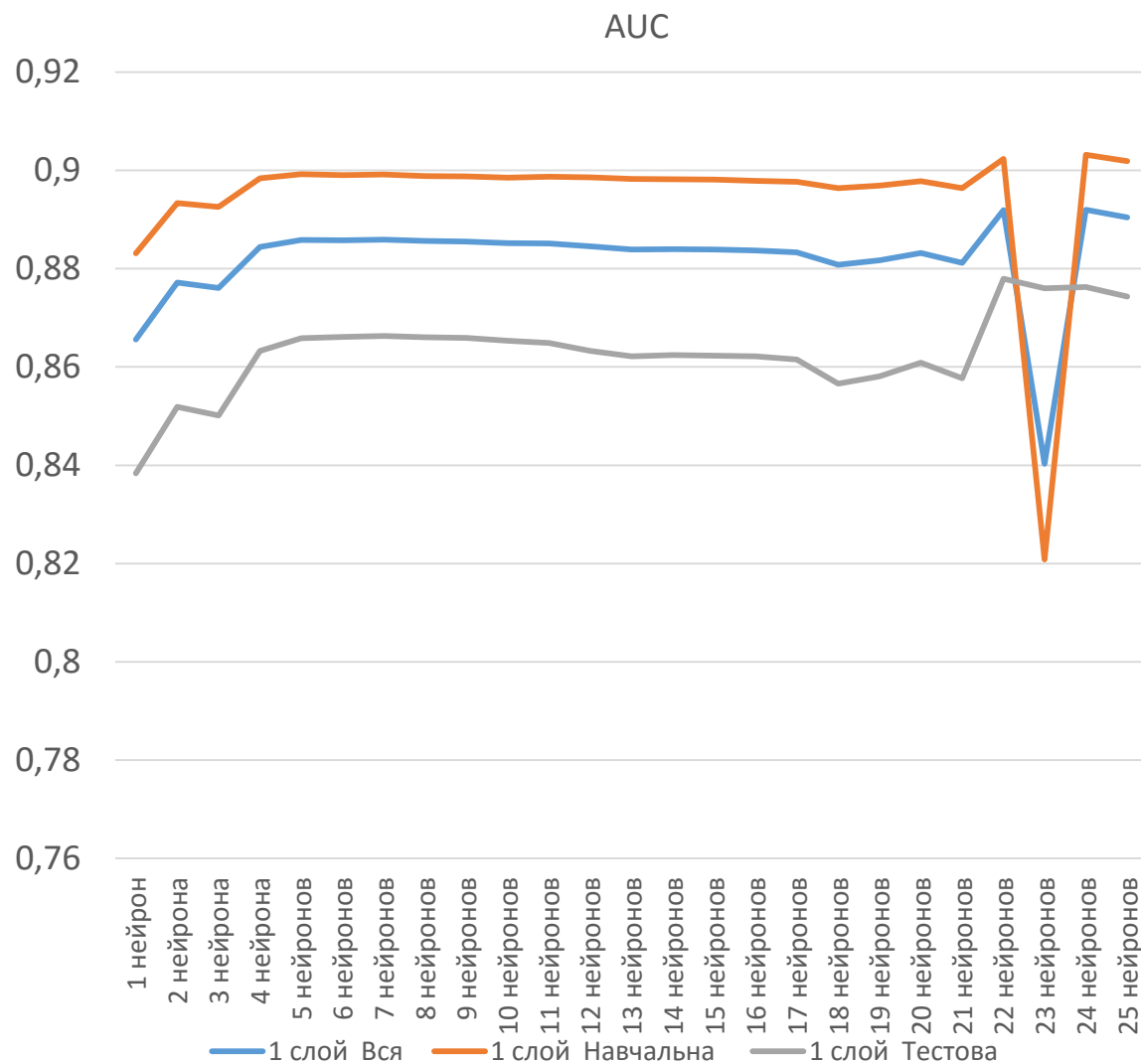
Логістична регресія

$$Y = - 0.0308 + 0.0231 \times \text{Age years} - 0.6233 \times \text{Sex} + 0.6047 \times \text{Education} + 0.0616 \times \text{Marital status} - 0.00002 \times \text{Credit} - 0.2673 \times \text{Region} + 0.09256 \times \text{PrevEmployerSector} - 0.103 \times \text{OccupationStatus} + 0.0064 \times \text{WorkExperienceMonth}$$

Логістична регресія

	Вся вибірка	Навчальна вибірка	Тестова вибірка
AUC	0,8771	0,8875	0,8456
GINI	0,7542	0,775	0,6912
ROC-крива			

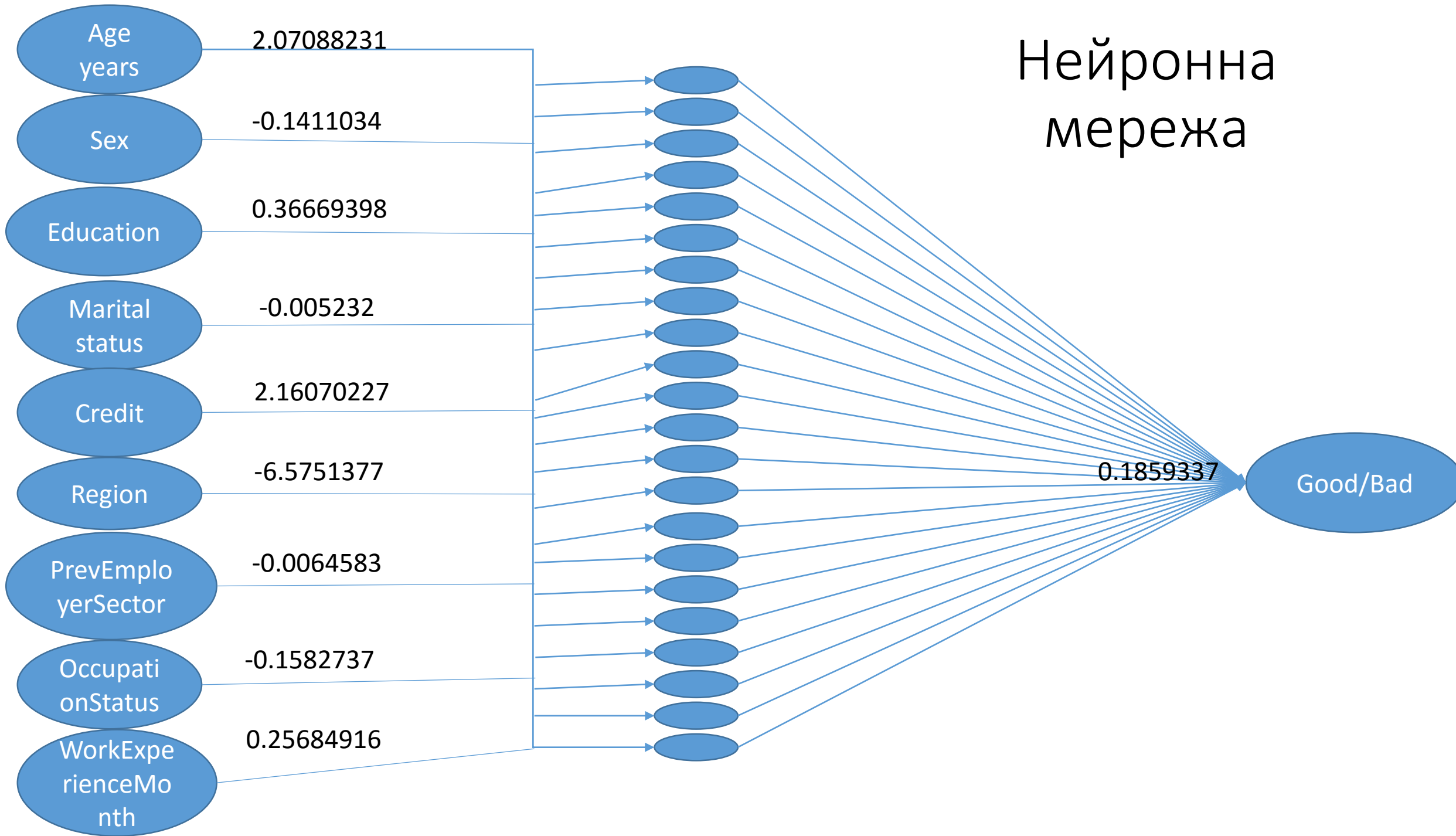
Нейронна мережа (вибір топології)



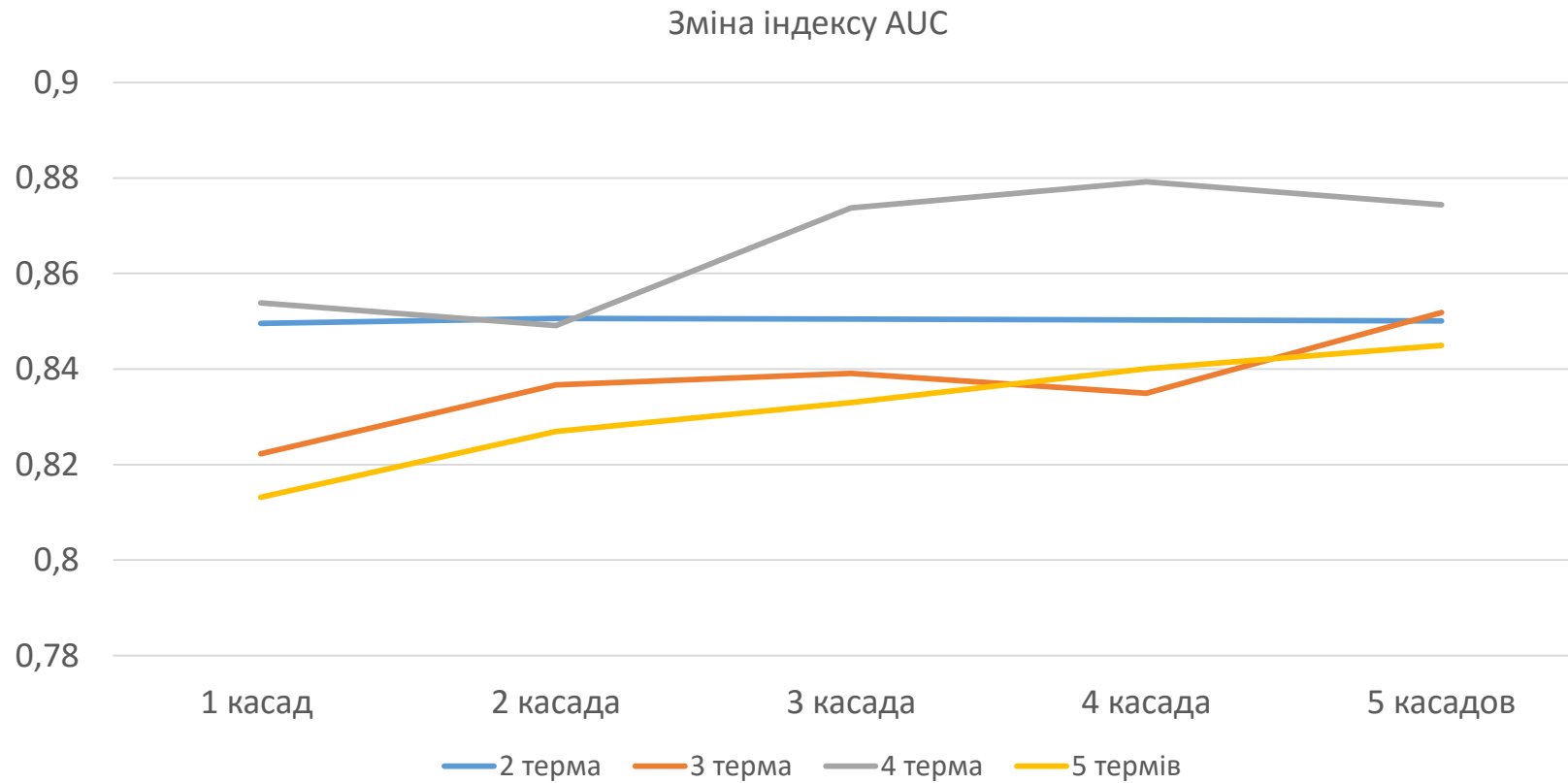
Нейронна мережа

Таким чином була обрана топологія з 1-м прихованим шаром з 22 нейронами

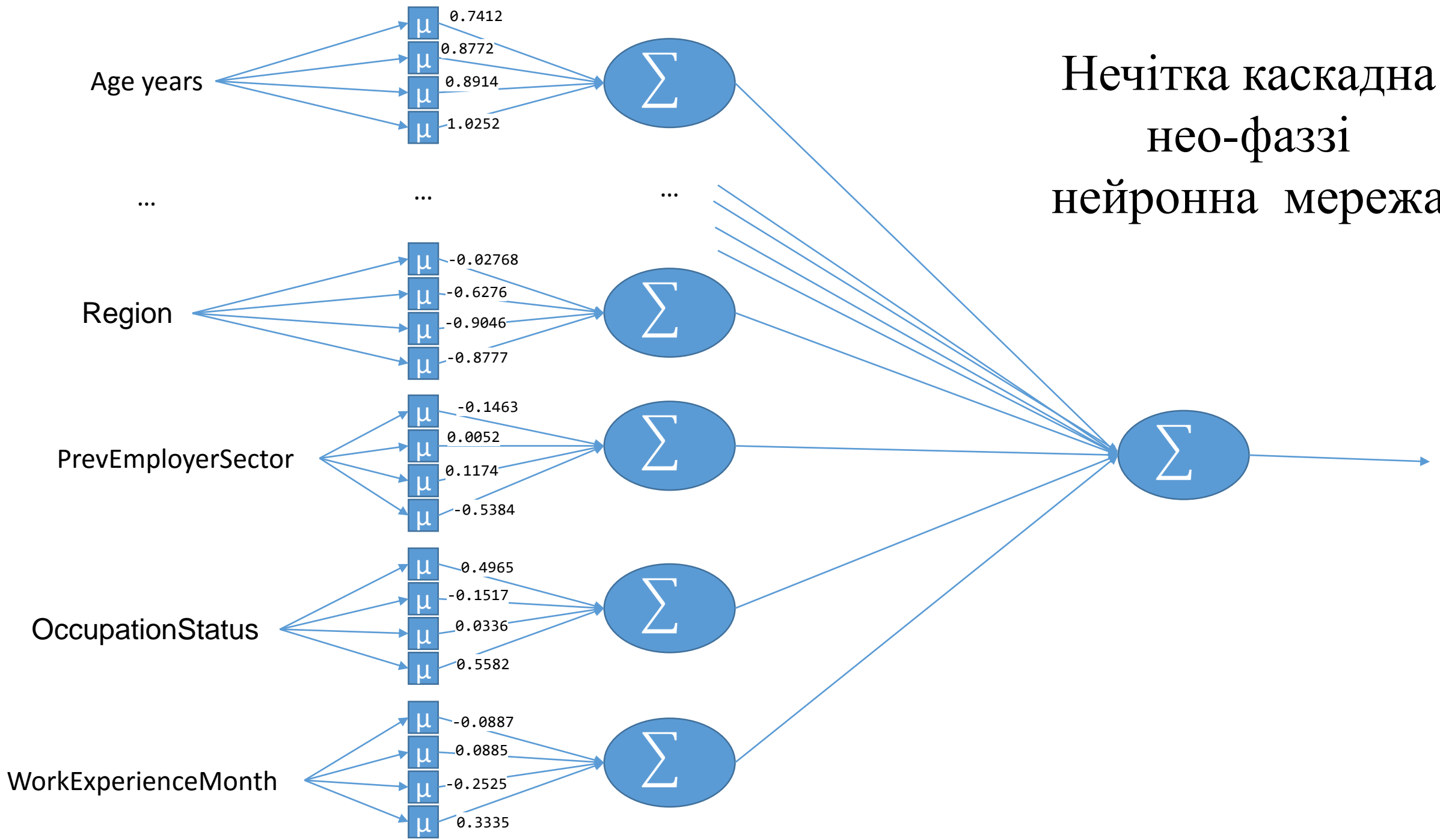
	Вся вибірка	Навчальна вибірка	Тестова вибірка
AUC	0,8919	0,9023	0,8779
GINI	0,7839	0,8046	0,7559
ROC-крива			



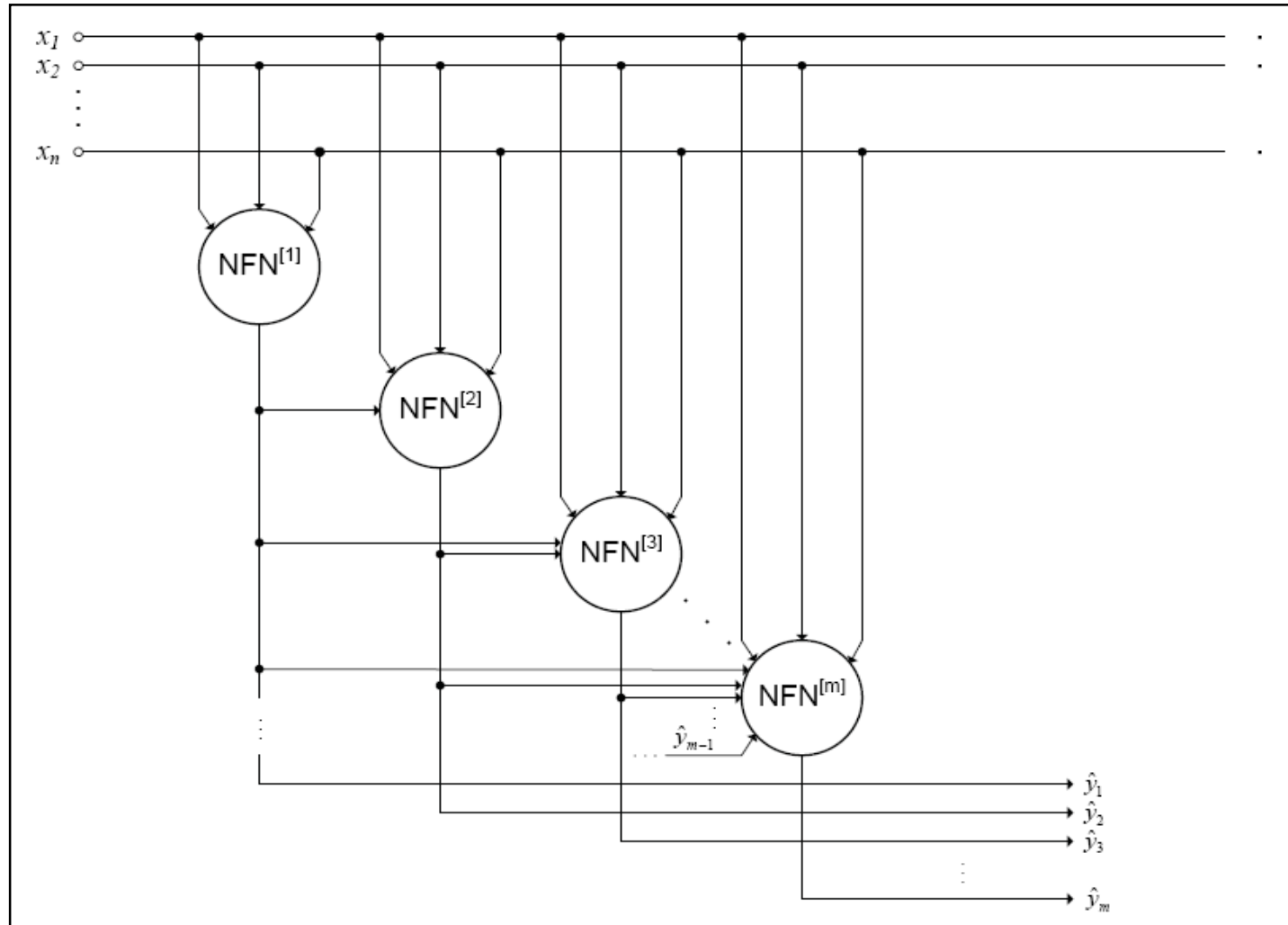
Нечітка каскадна нео-фаззі нейронна мережа (вибір кількості термів)



Нечітка каскадна нео-фаззі нейронна мережа



Нечітка каскадна нео-фаззі нейронна мережа



Нечітка каскадна нео-фаззі нейронна мережа

Таким чином була обрана топологія з 1-м прихованим шаром з 22 нейронами

	Вся вибірка	Навчальна вибірка	Тестова вибірка
AUC	0,938	0,944	0,931
GINI	0.876	0.887	0.862
ROC-крива			

Порівняння результатів

Назва методу	Індекс GINI	Значення AUC
Логістична регресія	0.691	0.846
Нейромережа	0.805	0.902
Нечітка нейромережа	0.862	0.931

Дякую за увагу!