

№1:

Задача 1. Іноземна компанія, яка реалізує свою продукцію в Україні, збирає статистику щодо причин повернення товарів на склад. Розглядають можливі варіанти: товар пошкоджений в процесі транспортування чи товар бракований. За статистикою відомо, що ймовірність повернення товару через пошкодження при транспортуванні становить 0,05, а ймовірність повернення через брак становить 0,3.

1. Обчислити власноруч ймовірність того, що повернений товар був бракований (використати теорему Байєса) та навести обчислення.
2. Побудувати **мережу Байєса** за допомогою програми GeNIe 2.0 та встановити зв'язки між вхідними змінними для задачі логістики.
3. Встановити ймовірність того, що повернений товар був бракований.

Задание 1

A - бракован

B - исправлен

C - помят.

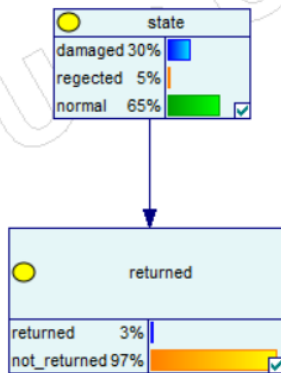
$$P(B) = P(B/A)P(A) + P(B/C)P(C) =$$

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/C)P(C)} =$$

$$= \begin{cases} P(B/A) = 0,3 \\ P(B/C) = 0,05 \end{cases} \begin{cases} P(A) = 0,5 \\ P(C) = 0,5 \end{cases}$$

$$= \frac{0,3 \cdot 0,5}{0,3 \cdot 0,5 + 0,05 \cdot 0,5} = 0,8$$

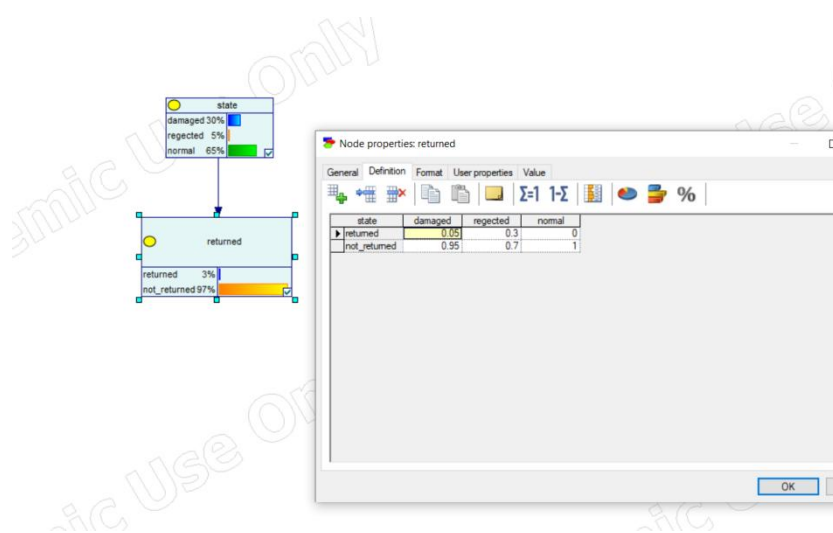
2) Створення мережі Байєса в програмі GeNIe:



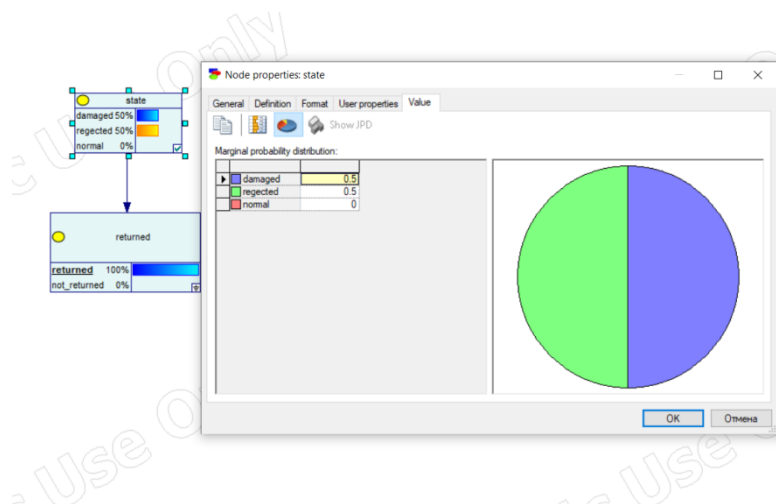
The screenshot shows the GeNIe software interface with the 'Node properties: state' dialog box open. The dialog box has tabs for 'General', 'Definition', 'Format', 'User properties', and 'Value'. The 'Value' tab is selected, showing a table of values for the 'state' node.

	Value
damaged	0.3
rejected	0.05
normal	0.65

The dialog box also has 'OK' and 'Отмена' buttons at the bottom right.



Встановлюємо evidence на returned та перевіряємо умовні вірогідності для таблиці state:



Як бачимо, пораховані вручну результати співпали з отриманими в GeNIe

№2:

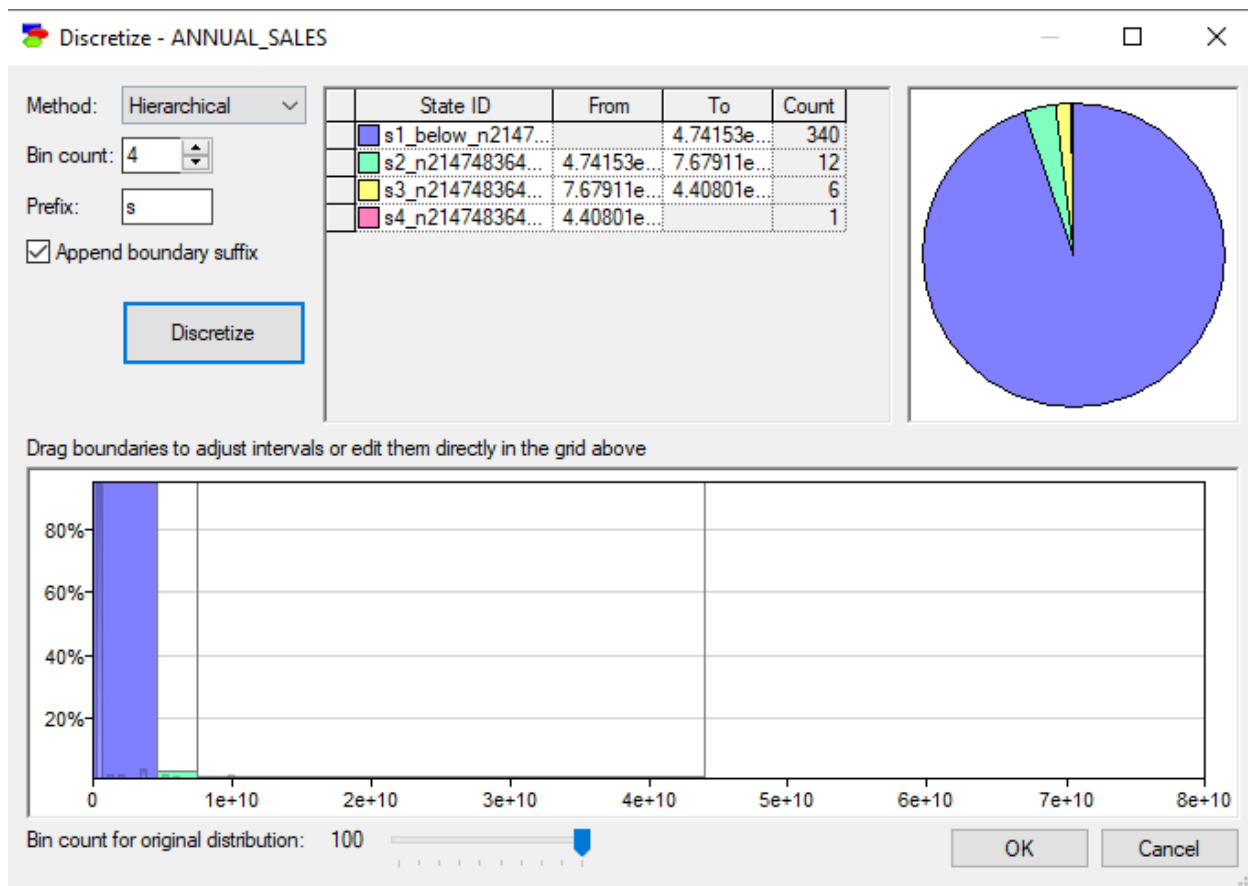
Підготовка даних:

1. Для початку змінюємо символ , на символ .
2. Далі заповнюємо відсутню інформацію нулями
3. Сортуюмо данні та видаляємо значення які сильно відрізняються від інших.

Наприклад:

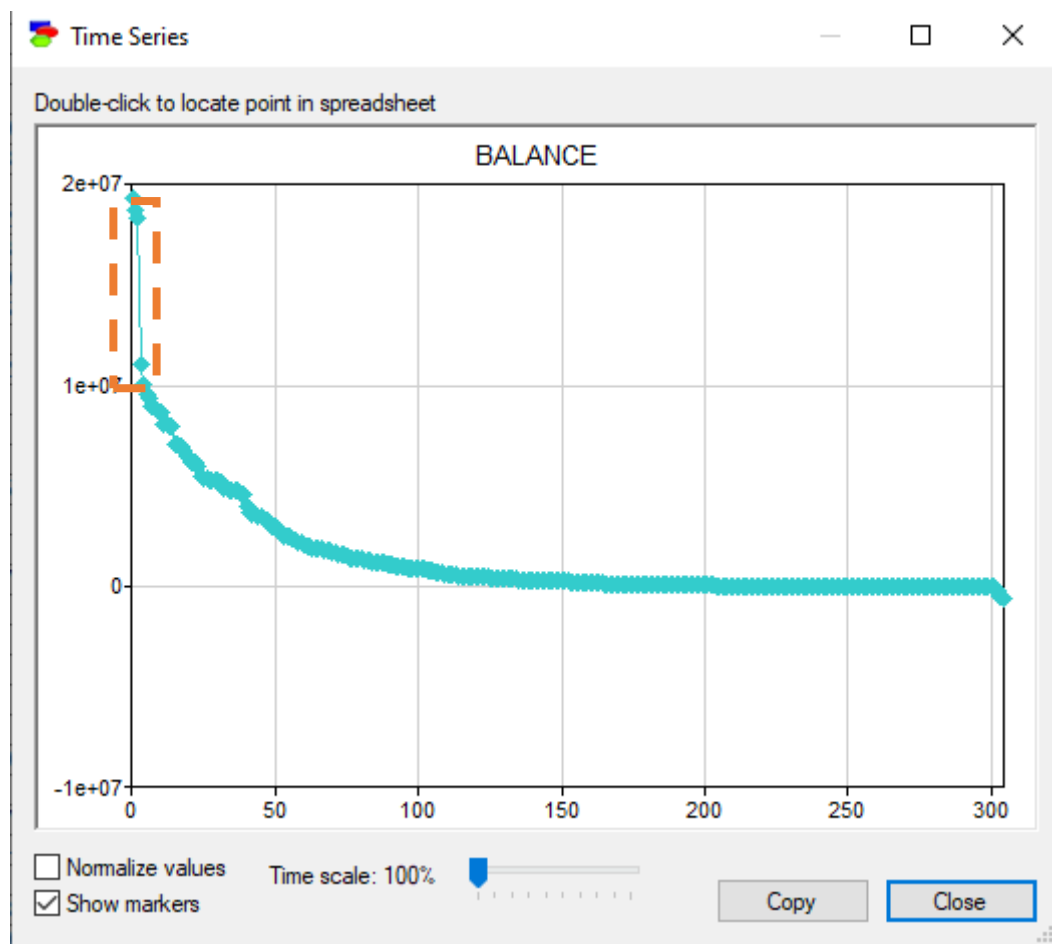
ANNUAL_SALES
78173000000
23960000000
23960000000
9960100000
9960100000
9960100000
9960100000

4. Далі проводимо дискретизацію кожного стовпчика та виводимо гістограму.



Далі розглянемо ці дії на прикладі стовпчика Balance

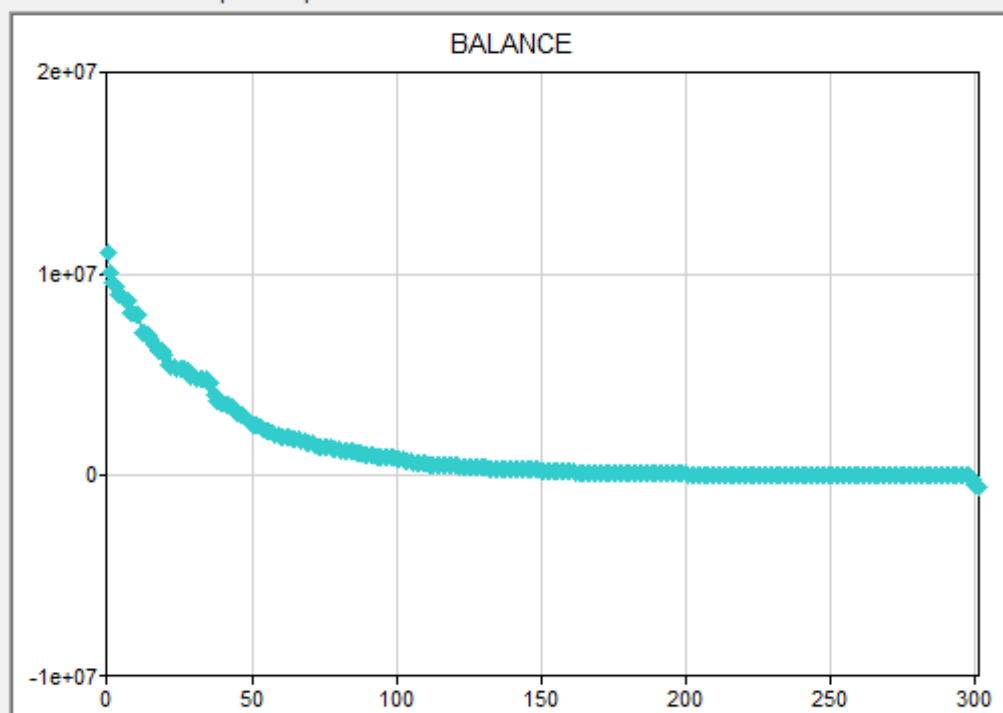
1. Выдаляємо данні які виділяються



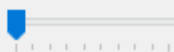
BALANCE
19355637.48
18786994.89
18351656.21
11125835.59
10117581.04
9559780.87
9382447.63
8959866.5
8868985.23000001
8819534.61

2.Проводимо дискретизацію та виводимо результати наших дій.

Double-click to locate point in spreadsheet

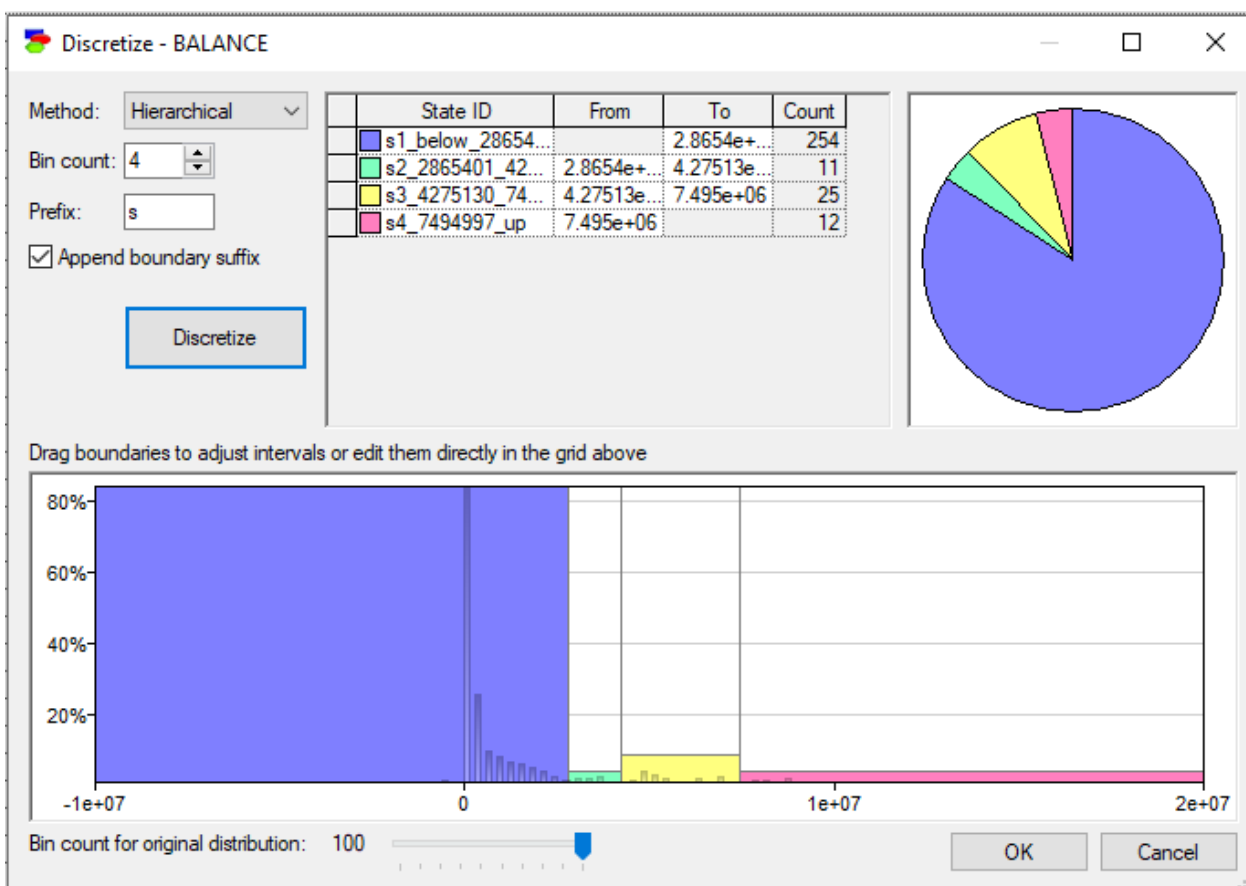
☐ Normalize values

Time scale: 100%

☒ Show markers

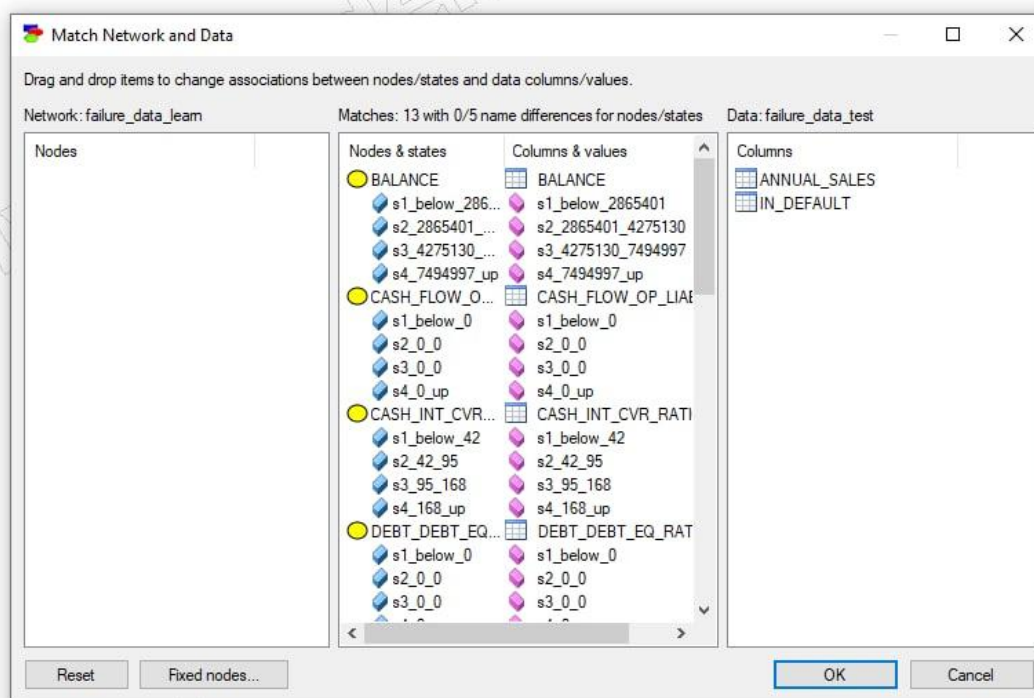
Copy

Close

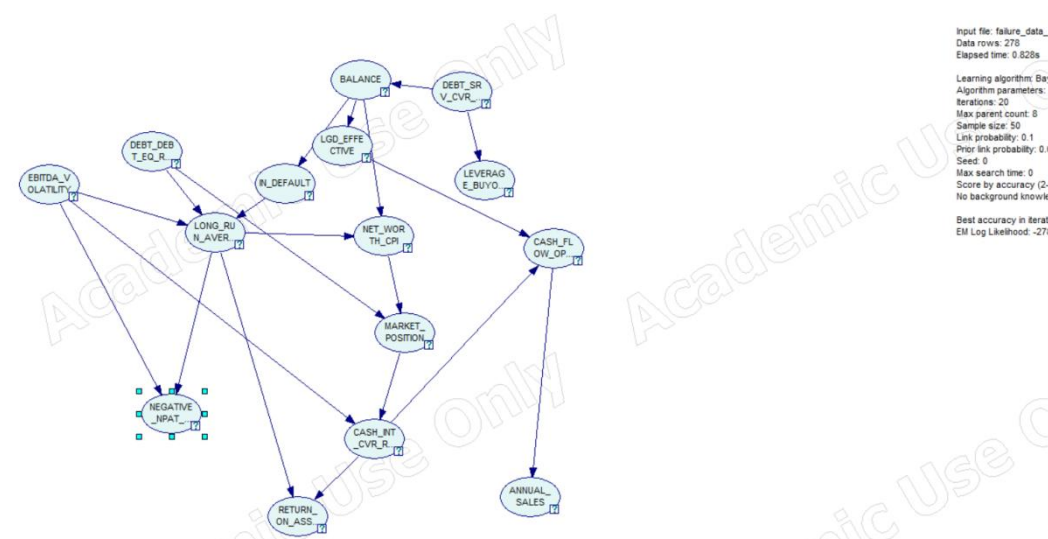


Далі проводимо аналогічні дії стосовно інших стовпчиків.

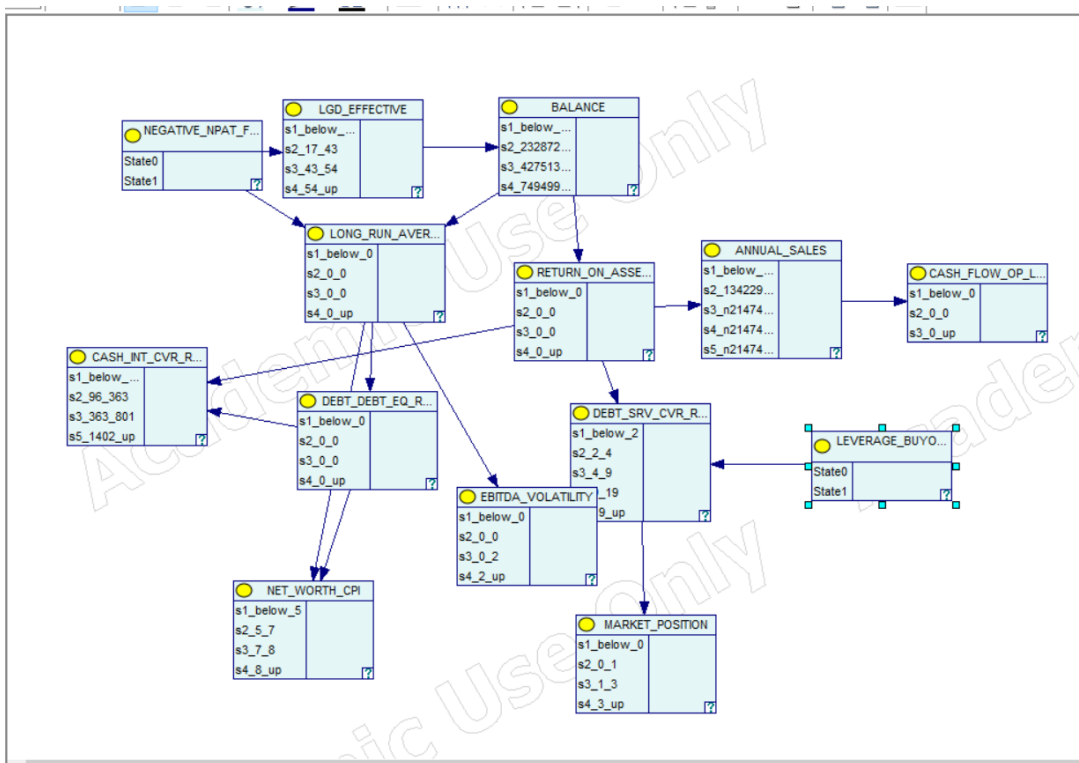
Розділяємо навчальну та тестову вибірку у співвідношенні 85/15 відповідно.



Будуємо структуру навчальної вибірки:



Навчання параметрів мережі за навчальною вибіркою:



7) Валідація тестової вибірки та аналіз:

Accuracy:

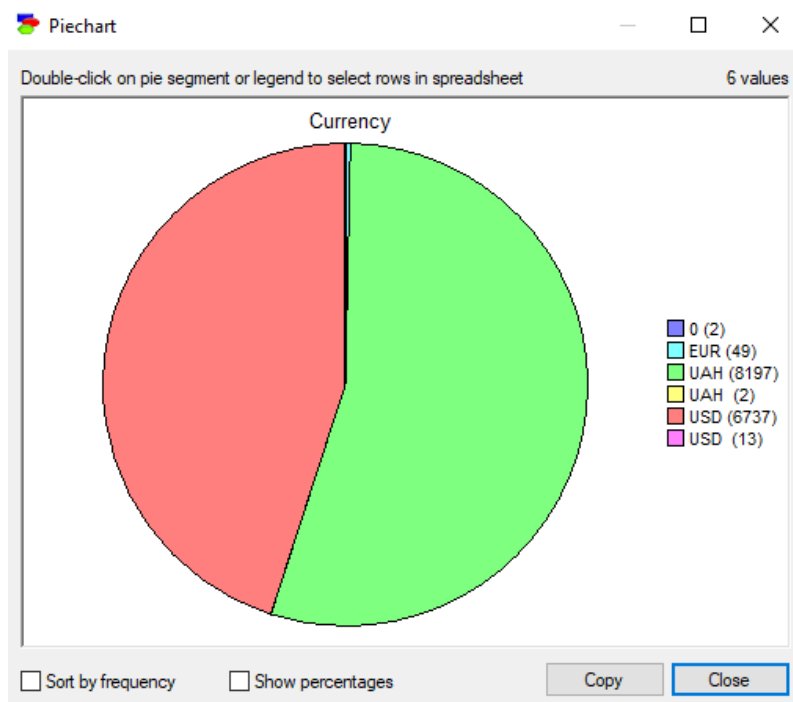
```
LONG_RUN_AVERAGE_PD = 0.850746 (57/67)
s1_below_0 = 0.901961 (46/51)
s2_0_0 = 0.833333 (10/12)
s3_0_0 = nan(ind) (0/0)
s4_0_up = 0.25 (1/4)
```

Мною було протестовано 4 моделі, однак з таким розподіленням була найбільш точною. Вона не гарантує стопроцентний результат, однак, процент похибки в ній найменший.

№3

Попередня обробка даних:

1. Об'єднємо класи з однаковими назвами.

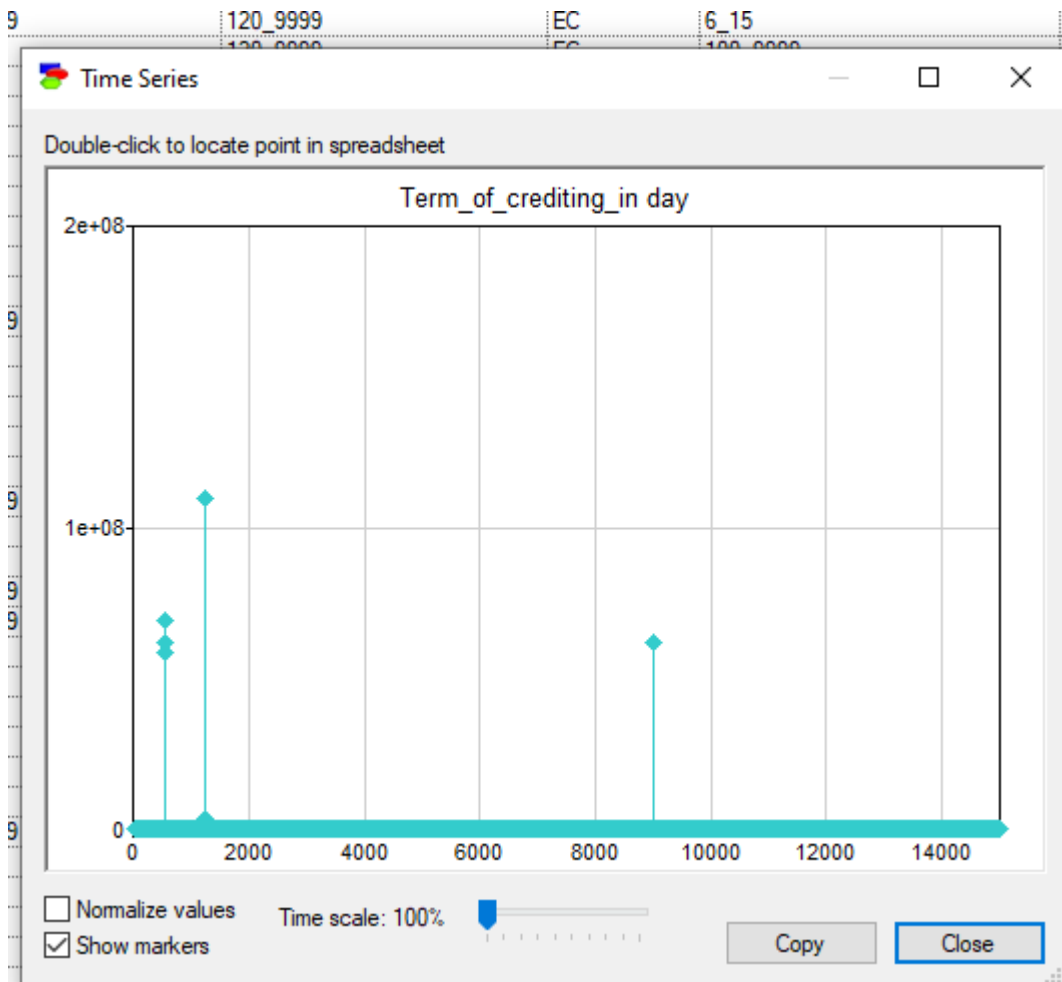


The Merge States dialog box is used to select states to merge. It features a table with 'State...' and 'Occurrences' columns. The table lists the following states and their occurrences: 0 (2), EUR (49), UAH (8197), UAH (2), USD (6737), and USD (13). The 'UAH 2' row is selected. Below the table, there is a text field for 'Name of new state:' containing the text 'UAH'. At the bottom, there are 'OK' and 'Cancel' buttons.

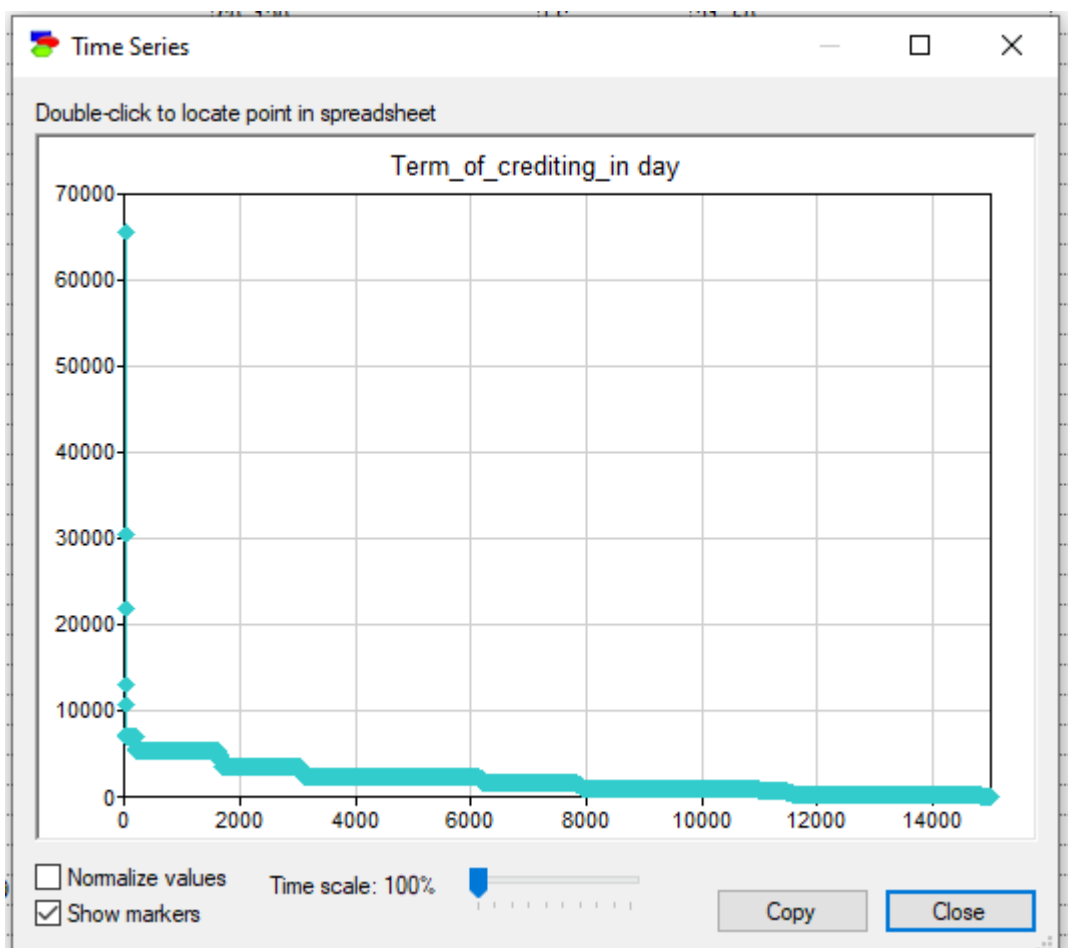
2. Видаляємо записи з нулями.

Gender	Age	#credit_sum	Currency	Term_of_crediting_in_day	Interest_rate_%	Duration_of_stay_in_a_city	Marital_status	Children	Job_position	Tenure_with_current_employer	Term_of_existence_of_enterprise	Company_type	Number_of_employees_in_company	Income_customer	Costs_customer	goal_credit	Result
Female	44	2619	0	360	28	120_9999	MARRIED	c2	AS	120_9999	120_9999	OT	100_9999	1138	155	consumer_credit	good
Female	40	96000	0	5400	13	6_12	MARRIED	c1	AS	24_60	24_60	WB	0_5	8397	0	flat	good
Female	45	15000	EUR	1080	16	120_9999	SINGLE	c0	TM	60_120	60_120	OT	6_15	5296	430	overhaul	good
Male	44	2500	EUR	1095	17	120_9999	MARRIED	c1	TM	0_3	0_12	LI	6_15	3000	192	overhaul	good
Male	30	10980	EUR	1080	13	120_9999	MARRIED	c1	TM	12_18	12_24	WB	6_15	7238	800	overhaul	good
Male	60	3200	EUR	1800	12	120_9999	MARRIED	c1	TM	24_60	24_60	CA	31_50	2960	810	consumer_credit	good
Female	30	8000	EUR	2520	14	120_9999	MARRIED	c3	SP	120_9999	12_24	WB	31_50	2634	764	overhaul	good

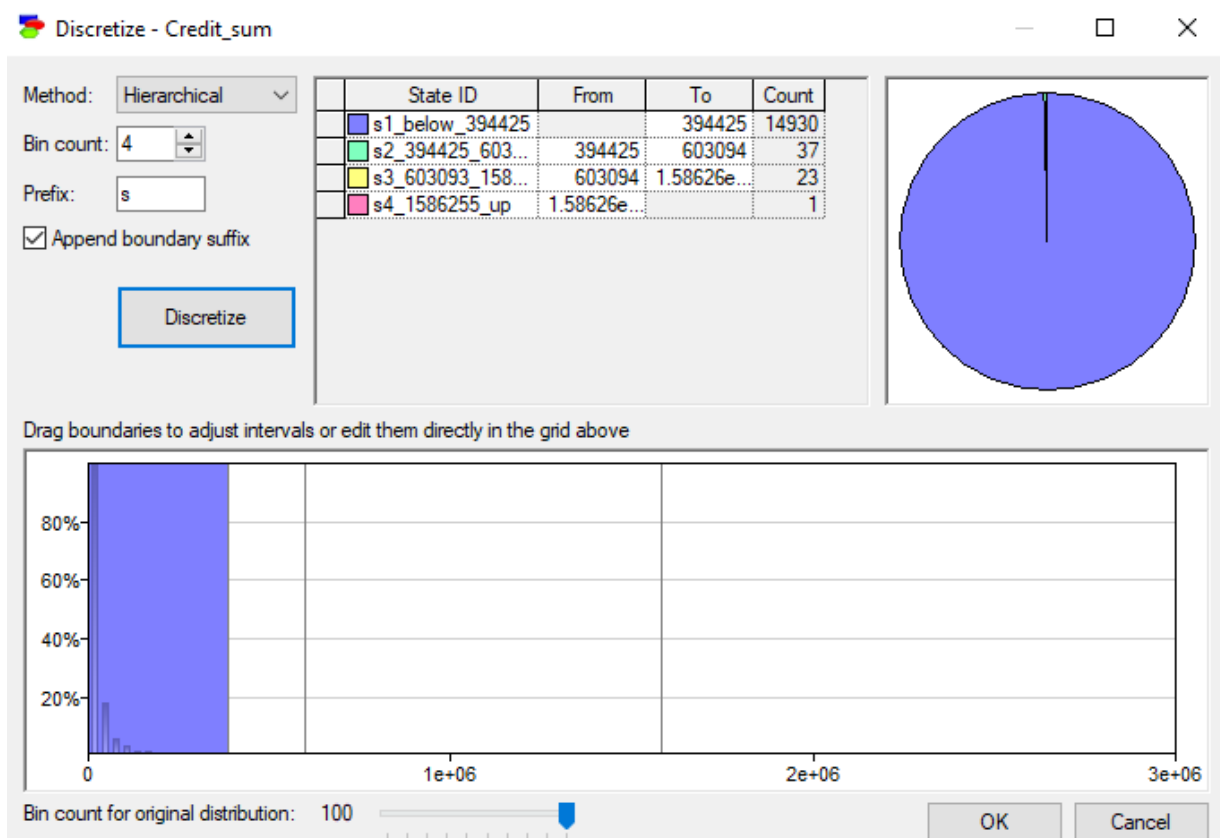
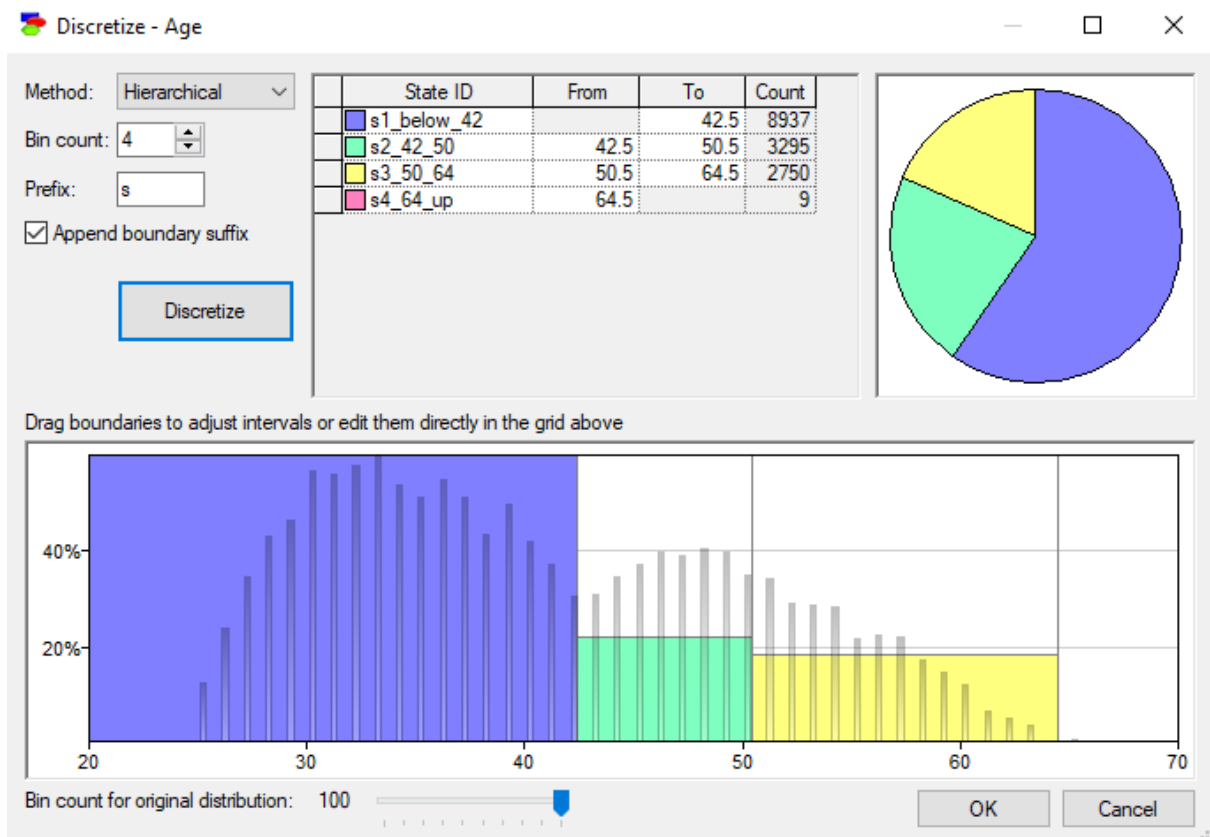
3. Знов видаляємо данні які сильно виділяються від середніх.



Term_of_crediting_in day	lr
109685055	
69535055	
62236880	
62235055	
62235055	
58585420	
3871920	
65700	
65700	



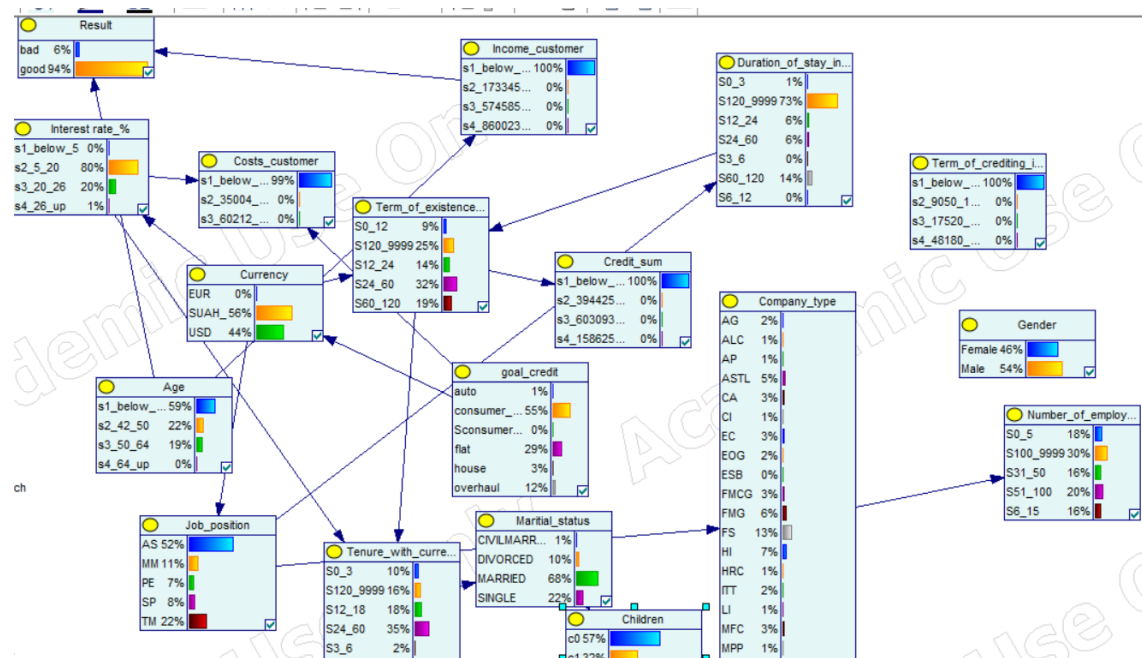
Далі проводимо дискретизацію та виводимо результати аналогічно другому завданню.



Аналогічно робимо для інших стовпчиків.

Розділюємо на тренувальну та тестову вибірки у співвідношенні 90/10

Та проводимо навчання



Результати валідації та аналіз:

Accuracy:

```
Result = 0.939301 (7923/8435)  
bad = 0.00197239 (1/507)  
good = 0.999243 (7922/7928)
```

Для покращення моделі треба збільшити данні с поганою кредитною історією, бо інакше мережа предикетить майже завжди результат близький до 1.