



МОДЕЛІ І МЕТОДИ ОЦІНЮВАННЯ АКТУАРНИХ РИЗИКІВ У БАНКІВСЬКІЙ СФЕРІ

Керівник: д.т.н., професор
каф.математичних методів
системного аналізу
Бідюк Петро Іванович
Виконала: студентка групи
КА-44, Поркалова Катерина.

АКТУАЛЬНІСТЬ

- Усвідомлення й прийняття ризиків – основа банківської справи. Успіх банку, як і будь-якої фінансової установи, залежить від того наскільки правильно й своєчасно прийняті ними ризики розумні, компетентні й контрольовані.
- Ризик – це ймовірність, або загроза, втрати частини своїх ресурсів, ймовірність появи додаткових витрат, або ж навпаки – недоотримання доходу підприємцем.
- Актуарний ризик – це ризик, який покривається страховою компанією чи банком в обмін на сплату премії, як правило, у грошових одиницях.

- Мета дослідження: побудова регресійної моделі для банківської системи України з метою визначення актуарних ризиків.
- Об'єкт дослідження: статистичні та розрахункові дані окремих фінансово економічних показників діяльності банку
- Предмет дослідження: методи регресійного аналізу, засоби програмного середовища SAS Enterprise Miner/Guide для вирішення цих завдань.

ПОСТАНОВКА ЗАДАЧІ

- Виконати аналіз предметної області, обрати математичні методи для визначення оптимальної кількості вхідних параметрів регресійної моделі.
- Побудувати математичні моделі лінійної і логістичної регресії, дерева рішень за допомогою програмного середовища SAS Enterprise Miner.
- Провести аналіз й порівняти отримані моделі, обрати кращі для подальшого використання моделі.
- Побудувати обрані моделі на іншому розподілі даних.
- Проаналізувати й порівняти отримані результати.
- Запропонувати кращі математичні моделі для прогнозування актуарних ризиків.

МАТЕМАТИЧНИХ МОДЕЛІ ДЛЯ ОПИСУ І ОЦІНКИ АКТУАРНИХ РИЗИКІВ

- BIA (Basic Indicator Approach – Метод базових показників) $\hat{Q}_{99\%} = \overline{GI} \cdot \alpha$;
- LDA (Loss Distribution Approach – Метод розподілу збитків) $x = \sum_{i=1}^{n(t)} L_i$;
- IMA (Internal Measurements Approach – Метод внутрішніх вимірів)

$$\hat{Q}_{99\%} = EI \cdot PE \cdot LGE = \gamma \cdot EL$$

- Регресійні моделі $x = A \cdot F + b + \varepsilon$

КРИТЕРІЇ АДЕКВАТНОСТІ МОДЕЛІ

- Коефіцієнт множинної детермінації R^2 :
$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST},$$
- Уточнений коефіцієнт детермінації:
$$\text{adjusted } R^2 = 1 - (1 - R^2) \cdot \frac{T - 1}{T - k - 1}$$
- Стандартна похибка регресії:
$$\text{CKП}(y, \hat{y}) = \text{RMSE}(y, \hat{y}) = \sqrt{\text{MSE}(y, \hat{y})} = \sqrt{E((y - \hat{y})^2)} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$
- Сума квадратів похибок моделі:
$$SSE = \sum_{t=1}^N (y - \hat{y})^2$$
- Критерій Акайке:
$$AIC = -\frac{2 \cdot l}{T} + \frac{2 \cdot k}{T}$$

КРИТЕРІЇ АДЕКВАТНОСТІ МОДЕЛІ

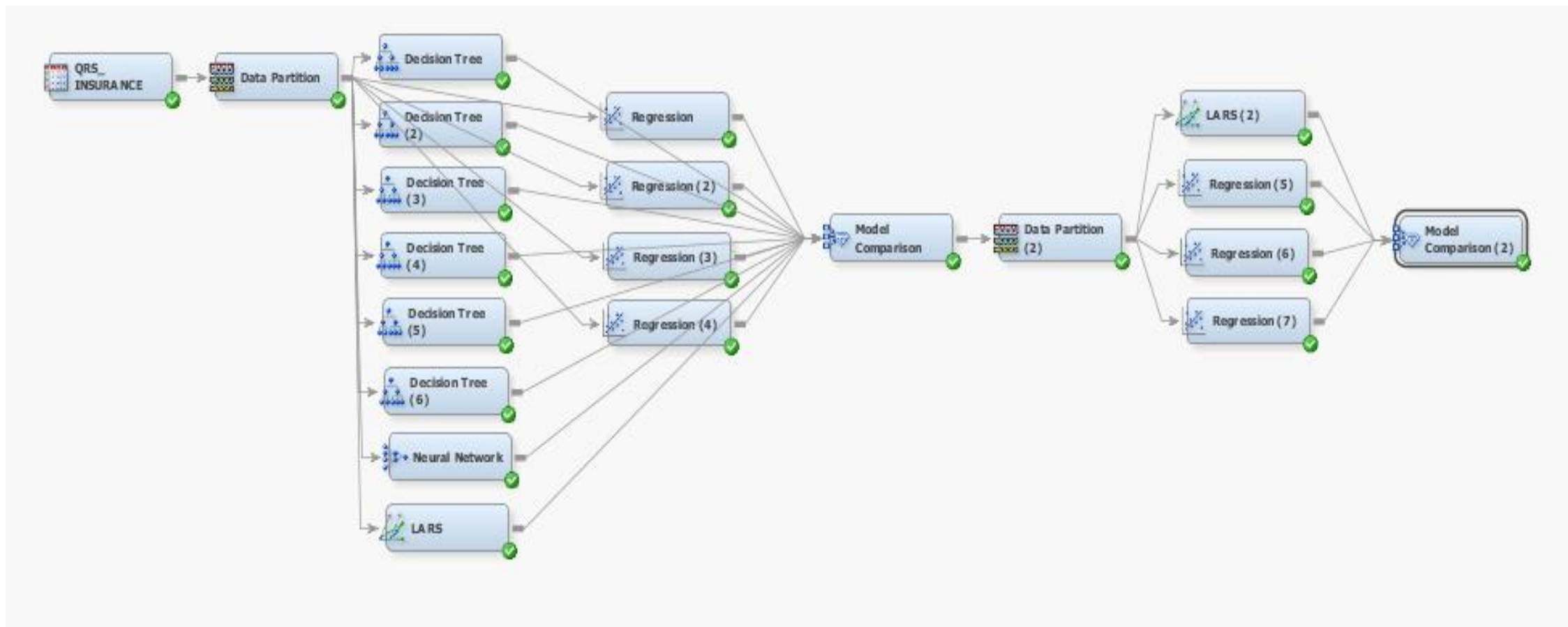
- Критерій Шварца : $SC = -\frac{2 \cdot l}{T} + \frac{k \cdot \log(T)}{T}$
- Байєсівський інформаційний критерій : $BIC = -2 \cdot \ln(l) + k \cdot \ln(T)$
- Статистика Стюдента або t-статистика : $t - \text{Statistic} = \frac{\text{Coefficient}}{\text{Std.Error}}$
- Критерій Колмогорова-Смірнова : $D = \sup_x |F_n(x) - F(x)|$
- Коефіцієнт Дурбина-Уотсона : $DW = \frac{\sum_{t=2}^N (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^N \varepsilon_t^2}$

ВИБІРКА

Variable Name	Role	Measurement Level	Label	Drop
CustID	Input	Interval	Customer ID Number	No
Premium_inc	Input	Interval	Premium increase in past yr	No
TARGET	Target	Interval	TARGET	No
bin_CHLDBL	Input	Interval	FLAG: Children UNK	No
bin_CHLDN	Input	Interval	FLAG: No Children in HH	No
bin_CHLDY	Input	Interval	FLAG: Children in HH	No
bin_LORES	Input	Interval	FLAG: LORES > 1 yr	No
bin_MARBL	Input	Interval	FLAG: Blank Marital Status	No
bin_MARI	Input	Interval	FLAG: Inferred Marital Status	No
bin_MARIM	Input	Interval	FLAG: Inferred Married	No
bin_MARIS	Input	Interval	FLAG: Inferred Single	No
bin_MARK	Input	Interval	FLAG: Known Marital Status	No
bin_MARKM	Input	Interval	FLAG: Known Married	No
bin_MARKS	Input	Interval	FLAG: Known Single	No
bin_REDPD	Input	Interval	FLAG: FLAG: New Policies Purchased with cash value of old policies	No
bin_eduBL	Input	Interval	FLAG: EDU blank	No
bin_eduCO	Input	Interval	FLAG: EDU College	No
bin_eduGS	Input	Interval	FLAG: EDU Grad Sch	No
bin_eduHS	Input	Interval	FLAG: EDU High Sch	No
bin_eduVT	Input	Interval	FLAG: EDU Vo/Tech	No
int_ETI	Input	Interval	INT: Extended Term Count	No
int_HHVPI	Input	Interval	INT: Potential HH VPI	No
int_INCOME	Input	Interval	INT: Income Group	No
int_LORES	Input	Interval	INT: Length of Residence	No
int_NTWRTH	Input	Interval	INT: Net Worth Indicator	No
int_POINTS	Input	Interval	INT: min(1.2, tot_mode_code_points)	No
int_POLCT	Input	Interval	INT: Policy Count	No
int_REDPD	Input	Interval	INT: # of new policies paid for with cash value of existing policies, max 3	No
int_STAR	Input	Interval	INT:HH_VPI_STAR_RATING	No
int_age	Input	Interval	INT: Age	No
int_nchld	Input	Interval	INT: Number of Children	No
nom_occup	Input	Nominal	NOM: Occupation	No
ord_STAR	Input	Interval	ORD:HH_VPI_STAR_RATING	No

Набір даних містить 74372 записів і 27 показників по клієнтам.

МОДЕЛЬ



АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ

Вибір кращої моделі для обраної задачі був проведений у два етапи. На першому етапі було побудовано 12 моделей, з яких 6 – дерева рішень, 3 логістичних регресії, лінійна регресія та LARS. Після отриманні результати порівнювались за двома критеріями: за загальною точністю та за індексом Gini.

Тип моделі	Навчальна вибірка (Trainе)		Тестова вибірка (Test)		Валідаційна вибірка (Validation)	
	Загальна точність (CA)	Індекс Gini	Загальна точність (CA)	Індекс Gini	Загальна точність (CA)	Індекс Gini
Дерево рішень	0	1	0	1	0	1
LARS	0,268004	0,693	0,266832	0,687	0,268004	0,691
Логістична регресія	0,438703	0,246	0,438469	0,234	0,438703	0,228
Лінійна регресія	0,438703	- 0,246	0,438469	- 0,234	0,438703	- 0,228
Нейрона мережа	0,438703	0,246	0,439469	0,234	0,438703	0,228

АНАЛІЗ РЕЗУЛЬТАТІВ РОБОТИ

За результатами першого етапу були обрані дві кращі моделі – логістична регресія та LARS. Порівняння цих двох моделей виконувалось на основі статистичних результатів обох моделей на переросподіленій вибірці.

Тип моделі	Навчальна вибірка (Train)		Тестова вибірка (Test)		Валідаційна вибірка (Validation)	
	Сума квадратів в похибки	Індекс Gini	Сума квадратів в похибки	Індекс Gini	Сума квадратів в похибки	Індекс Gini
LARS	0,003098	1	0,00243	1	0,00243	1
Логістична регресія	0,000125	1	6,574E-5	1	0,0001782	1

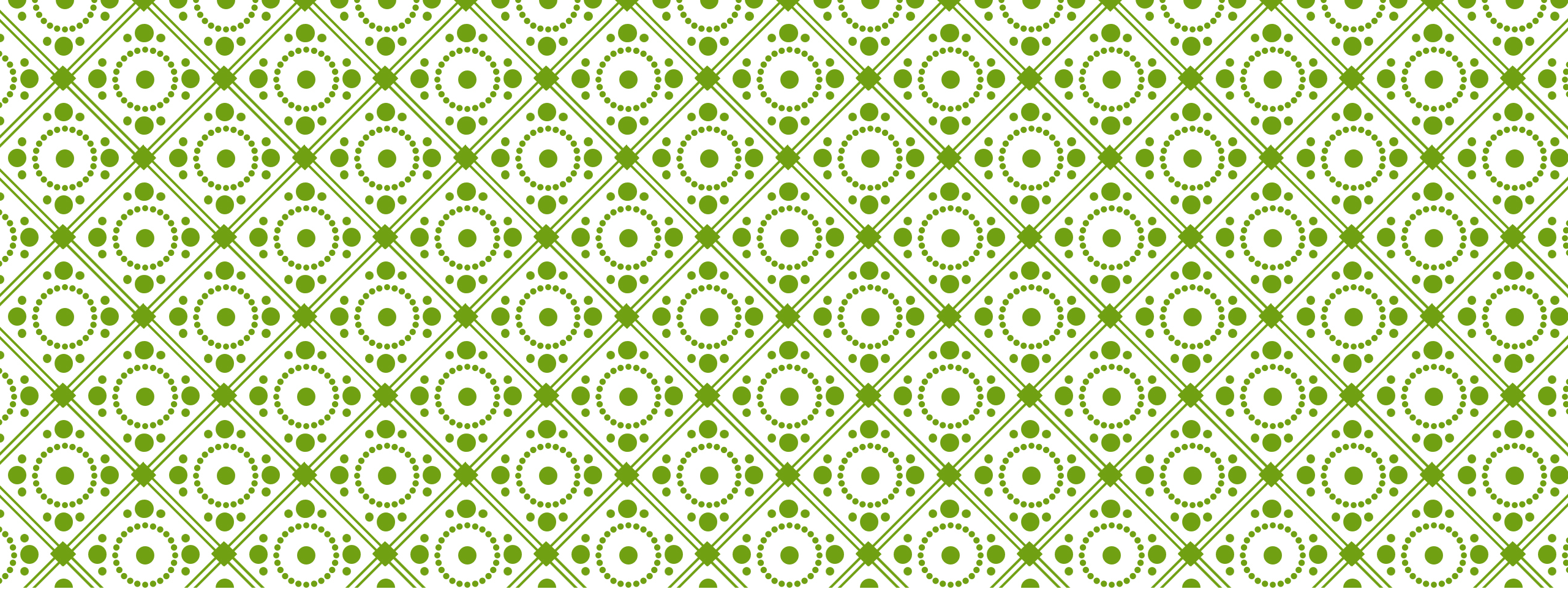
Кращі моделі порівнювалися за індексом Gini та за сумою квадратів похибки. Обидві моделі індекс Gini показали рівним 1, а сума квадратів похибки у моделі LARS була значно більша тому кращою було обрано логістичну регресію.

ВИСНОВОК

- Проведено аналіз і опрацювання джерел інформації з актуарних ризиків в фінансовій сфері.
- Побудовані моделі лінійної та логістичної регресії за допомогою засобів програмного середовища SAS.
- Проведений аналіз та порівняння отриманих моделей та обрано найзмістовніші моделі для подальшого використання.

ПОДАЛЬШЕ ДОСЛІДЖЕННЯ

Подальшим дослідженням для цієї роботи могли б стати подальша адаптація моделі для досягнення більшої точності, доцільно створити спеціалізовану СППР з використанням альтернативних моделей – мережі Байєса, пробіт-регресія, модель бінарного вибору тощо і застосувати сучасні методи оцінювання параметрів



ДЯКУЮ ЗА УВАГУ!