

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

---

# СКОРИНГОВІ МОДЕЛІ ПОВЕДІНКИ КЛІЄНТІВ-ВЛАСНИКІВ КРЕДИТНИХ КАРТ ДЛЯ ОЦІНКИ ЇХ ПЛАТОСПРОМОЖНОСТІ

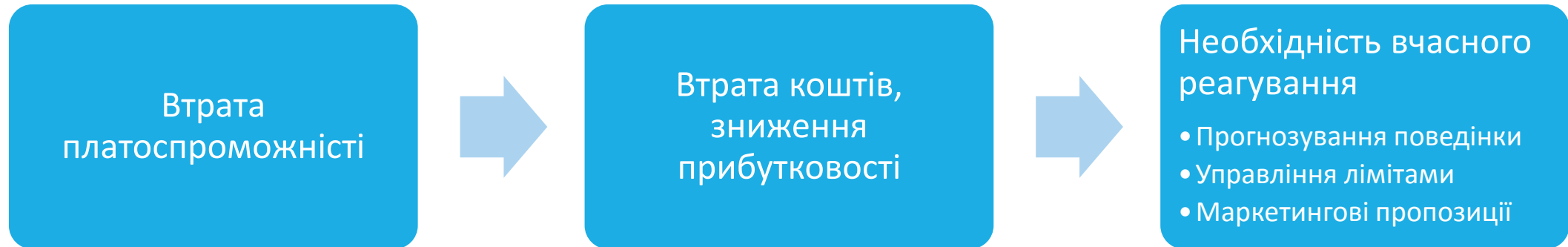
---

ВИКОНАВ: СТУДЕНТ ФОМІН О. В.

НАУКОВИЙ КЕРІВНИК: К.Т.Н КУЗНЕЦОВА Н. В.

# Актуальність теми

---



# Класичний підхід до побудови скорингової моделі

---

**Лінійна регресія** – метод моделювання залежності між скаляром  $y$  та векторною (у загальному випадку) змінною  $X$ . У випадку, якщо змінна  $X$  також є скаляром, регресію називають простою.

$$Y = \beta^T X + \epsilon, \quad Y \in \mathbb{R}^n, \quad X \in \text{Mat}(n \times k)$$

**Логістична регресія** – статистичний регресійний метод, що використовується у випадку коли пояснювана змінна може набувати тільки двох значень (чи, більш загально, скінченну множину значень).

$$p = \mathbb{E}(y|x) = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}} = \sigma(e^{\beta^T x})$$

# Постановка задачі

---

**Мета роботи:** розробка та побудова альтернативних моделей поведінки клієнтів-власників кредитних карт та їх порівняння із існуючими загальноприйнятими.

**Завдання:**

- Розглянути існуючі методи побудови скоринг-моделей;
- Виявити найбільш актуальні та перспективні поведінкові моделі;
- Розробити власні моделі, що б відзначалися динамічністю, ефективністю та предикативністю;
- Порівняти результати та надати практичні рекомендації щодо розробки моделей;

# Кредитні ризики

---

**Кредитний ризик банку** – це вартісне вираження ймовірності відхилення ризикової позиції від очікуваних результатів (настання ризикової події) унаслідок невизначеності дії зовнішніх та внутрішніх щодо банку факторів.

$$LGD = \sum_{i=1}^n PD_i \cdot EAD_i$$

де  $EAD_i$  – ризик дефолту;  $PD_i$  – ймовірність дефолту;  
 $LGD$  – втрати в разі дефолту



# Кредитні ризики. Основні поняття

---

**Платоспроможність** - здатність суб'єкта економічної діяльності погашати свої довготермінові зобов'язання. Характеризується коефіцієнтом заборгованості

**Дефолт (настання неплатоспроможності)** – стан у кредитних відносинах, що настає, коли позичальник не виплачує свої борги або платежі, порушення платіжних зобов'язань позичальника перед кредитором, нездатність проводити своєчасні виплати за борговими зобов'язаннями або виконувати інші умови договору позики.

## Модель дефолту

$$PDD > 90 \wedge OVD \geq 100$$

Що відповідає критерію: прострочено щонайменше 3 платежі (90 днів) та величина простроченої заборгованості складає не менше 100 грн.

# Аналіз виживання. Основні поняття

---

**Аналіз виживання** використовується для дослідження моменту загибелі представників деякої популяції.

Час, що проходить до настання цього моменту, називається часом виживання ( $T$ ).

**Момент смерті:**

$T$  – випадкова величина

**Функція виживання:**

$$S(t) = \mathbb{P}(T > t)$$

**Функція ризику:**

$$h(t) = \lim_{\delta \rightarrow 0} \left( \frac{\mathbb{P}(t \leq T < t + \delta | T \geq t)}{\delta} \right)$$

**Основні співвідношення:**

$$S(t) = e^{-\int_0^t h(u) du}, \quad h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)}$$

# Модель Кокса

---

Моделюється функція ризику:

$$h(t, x(t), \beta) = h_0(t)e^{\beta^T x(t)}$$

де  $x$  – вектор незалежних змінних;  $\beta$  – вектор коефіцієнтів, що оцінюються;

Припускається, що відношення функцій ризику постійне:

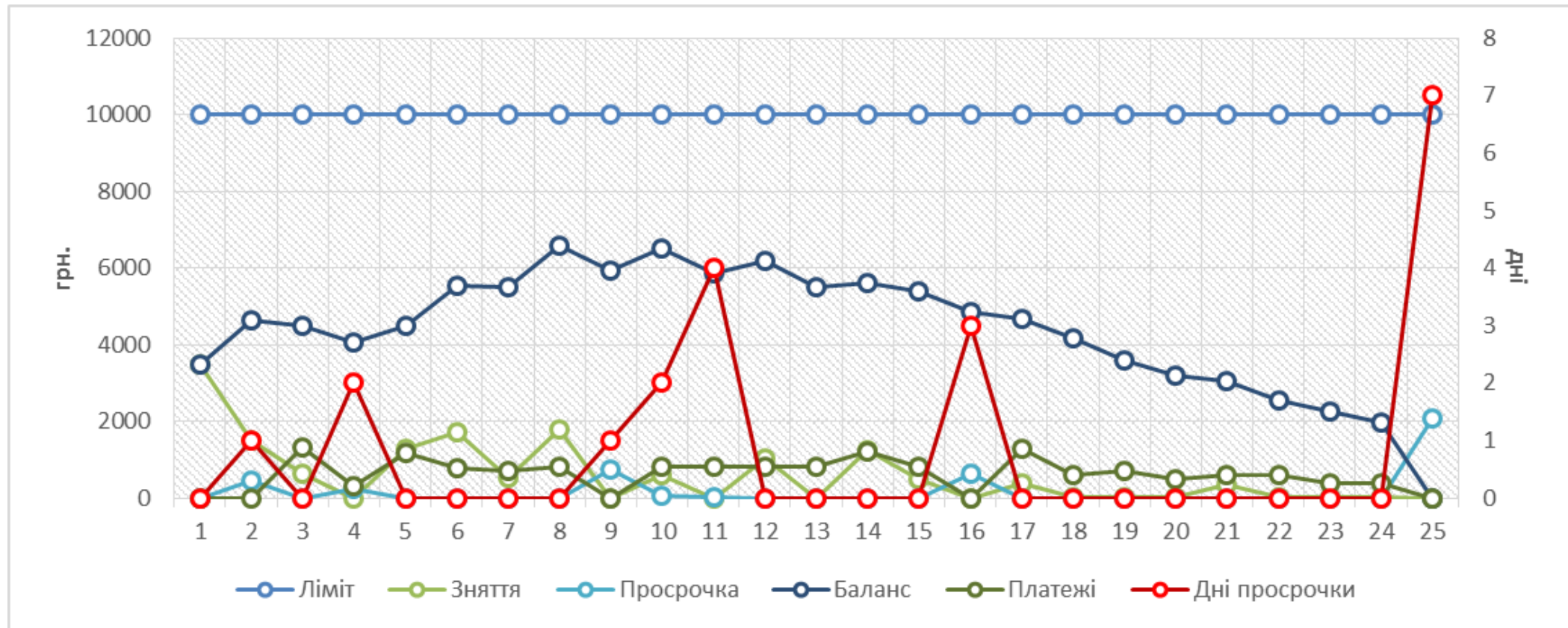
$$\frac{h(t, x_1, \beta)}{h(t, x_2, \beta)} = e^{\beta(x_1 - x_2)} = \text{const}$$

Для врахування змінних, що залежать від часу застосовується її **узагальнена** варіація

$$\frac{h(t, x_1(t), \beta)}{h(t, x_2(t), \beta)} = e^{\beta(x_1(t) - x_2(t))} \neq \text{const}$$



# Характер наявних даних



# Дані для логістичної регресії

---

Для дослідження використовувалися спосетережння за 4 тис. такими кредитними картами, виданими в 2013-15 рр. Загальна кількість спосетережнь – 55 тис. За весь період було зафіксовано 714 випадків дефолту.

Період «дозрівання» – 7 місяців. Для формування параметрів проводилась агрегація, внаслідок якої було отримано 3000 спостережень за кредитними картами. До списку параметрів увійшли як характеристики поведінки, такі і аплікаційні дані. Серед інших:

**Поведінкові параметри:** залишок за тілом кредиту, залишок за комісією/відсотками, прострочена заборгованість за тілом, прострочена заборгованість за комісією/відсотками, кількість днів прострочки за тілом,...

**Аплікаційні дані:** тип клієнта, вік, рік видачі, тривалість угоди, ліміт на початку, сума при подачі заявки, признак рівності ліміту та суми, адреса прописки, адреса місця постійного проживання, кількість дітей, ...

**Агреговані дані:** максимальна кількість прострочених місяців, середнє значення щомісячного зняття, максимальне значення відношення простроченої заборгованості до встановленого ліміту,...

# Відібрані параметри та модель на основі логістичної регресії

$$\ln \frac{p}{1-p} = \sum_{i=1}^{20} \beta_i x_i$$

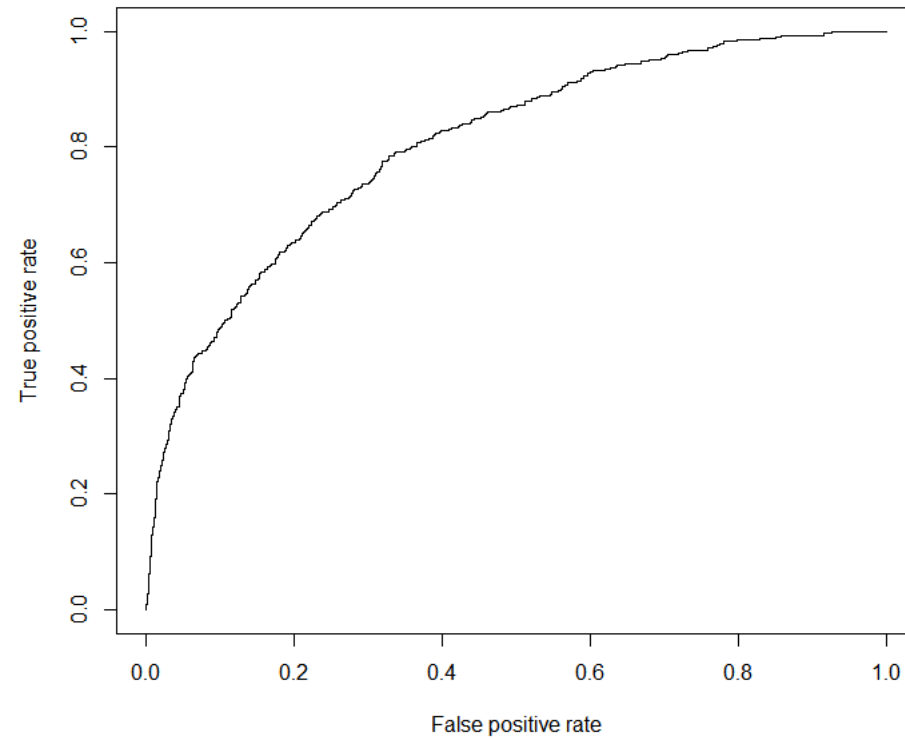
	Estimate	Std. Error	z value	Pr(> z )							
(Intercept)	-0.863284	0.613071	-1.408	0.159093							
age	0.004807	0.005337	0.901	0.367745							
is_satisfiedt	-0.411450	0.119419	-3.445	0.000570	***						
liv_is_regt	-0.018226	0.129879	-0.140	0.888402							
childcnt	0.030690	0.078934	0.389	0.697421							
dependantcnt	0.011642	0.124057	0.094	0.925233							
log(1 + income)	0.029583	0.022285	1.327	0.184346							
has_u_scoret	0.879906	0.580642	1.515	0.129671							
with_bank_mon	-0.018143	0.009391	-1.932	0.053368	.						
reg_mon	-0.004135	0.003742	-1.105	0.269086							
clienttypeClient	0.479826	0.319386	1.502	0.133009							
clienttypeEmployee	0.275121	0.411532	0.669	0.503796							
clienttypeNormal	0.690322	0.242271	2.849	0.004380	**						
clienttypeSalary	0.050843	0.246264	0.206	0.836434							
log(1 + max_outbody)	0.142916	0.140520	1.017	0.309129							
log(1 + max_ovdbody)	0.380554	0.186531	2.040	0.041334	*						
log(1 + max_ovd)	-0.023716	0.183434	-0.129	0.897131							
log(1 + max_limit)	-0.333863	0.087932	-3.797	0.000147	***						
log(1 + avg_montake)	0.367150	0.197322	1.861	0.062792	.						
log(1 + avg_monpay)	-0.556334	0.065413	-8.505	< 2e-16	***						
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

# Показники побудованої моделі

---

$AUC = 0,804$

$Gini = 60,8\%$



# Вхідні дані для моделей Кокса (РН)

---

Застосовувалася та сама вибірка без агрегації у вигляді 55 тис. записів.

- ❑ середній вік КК – 9,113 місяців; максимальний – 31 місяць;
- ❑ максимальна заборгованість за тілом – 100000 грн.; середнє значення такої заборгованості – 4105,5 грн.;
- ❑ середнє значення простроченої заборгованості – 171 грн.; максимальне – 51474 грн.;
- ❑ найбільша величина ліміту – 250000 грн.; середнє значення – 6884 грн.;
- ❑ в середньому щомісяця клієнти користувалися лімітом у розмірі 942,9 грн.; при цьому з їхнього боку в середньому надходило менше – 740,2 грн., що свідчить про тенденцію до прострочування клієнтів у даній вибірці;
- ❑ середній вік власників КК – 38,67 років; медіана – 37 років; максимальний – 66 роки; мінімальний – 20 років;
- ❑ середній дохід – 3827 грн.; максимальний – 120000 грн.; мінімальний – 0 грн.

# Відібрані параметри та модель М0 на основі РН

$$\ln h(t, x(t), \beta) = \sum_{i=1}^{21} \beta_i x_i$$

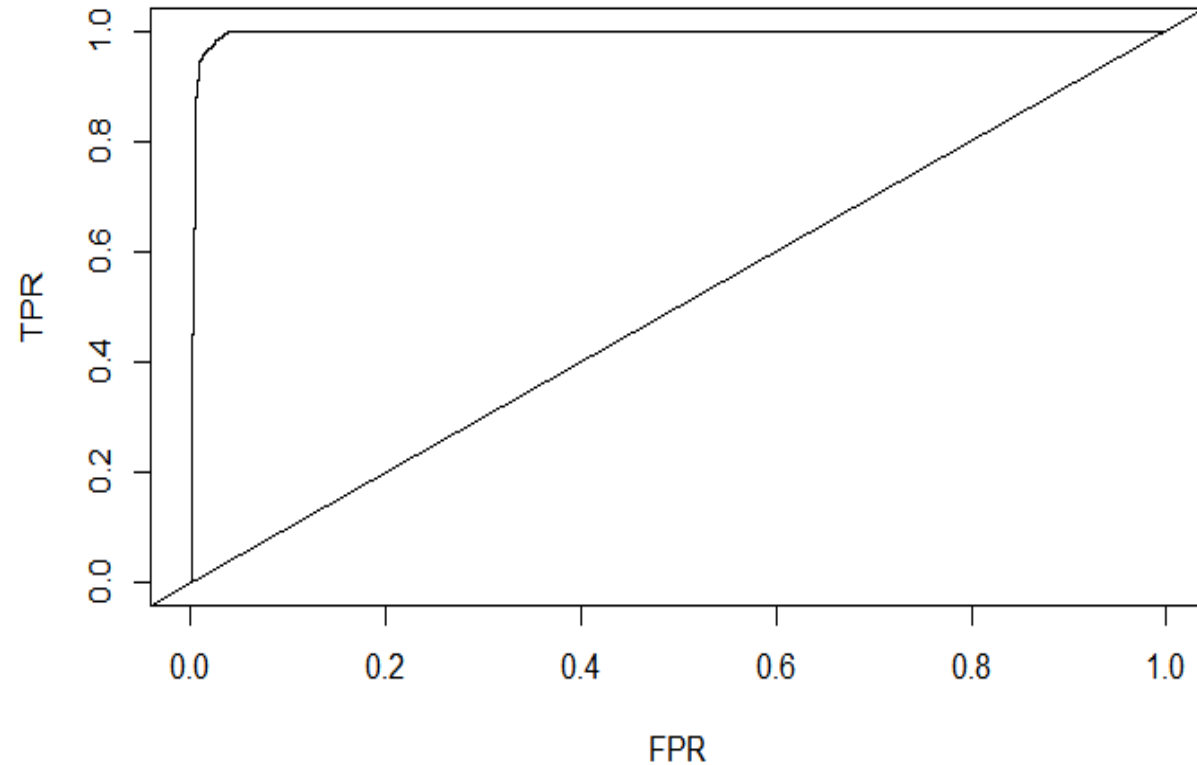
	coef	exp(coef)	se(coef)	z	p
log(1 + outbody)	0.14317	1.15392	0.03477	4.12	3.8e-05
log(1 + ovd)	0.74007	2.09607	0.04711	15.71	< 2e-16
pdd_new	1.49379	4.45395	0.04751	31.44	< 2e-16
log(1 + limit)	-0.21662	0.80524	0.01981	-10.93	< 2e-16
log(1 + montake)	-0.43572	0.64680	0.13098	-3.33	0.00088
log(1 + abs(monpay))	-0.03364	0.96692	0.02740	-1.23	0.21953
clienttypeClient	-0.05674	0.94484	0.23544	-0.24	0.80957
clienttypeEmployee	-0.19553	0.82240	0.40485	-0.48	0.62912
clienttypeInsider	0.65419	1.92359	0.64436	1.02	0.30998
clienttypeNormal	-0.24745	0.78079	0.18826	-1.31	0.18872
clienttypeOldSalary	0.04742	1.04856	0.24114	0.20	0.84410
clienttypeSalary	0.23902	1.27000	0.19967	1.20	0.23129
age	-0.01370	0.98639	0.00395	-3.47	0.00052
is_satisfiedt	-0.44390	0.64153	0.08898	-4.99	6.1e-07
liv_is_regt	-0.02701	0.97335	0.09588	-0.28	0.77814
childcnt	0.12223	1.13001	0.05829	2.10	0.03600
dependantcnt	-0.23561	0.79009	0.09375	-2.51	0.01196
log(1 + income)	-0.13613	0.87273	0.01892	-7.19	6.3e-13
has_u_scoret	0.38762	1.47347	0.49089	0.79	0.42975
with_bank_mon	-0.01518	0.98493	0.00874	-1.74	0.08254
reg_mon	0.00275	1.00275	0.00307	0.90	0.36960

# Показники побудованої моделі M0

---

$AUC = 0,95$

$Gini = 90\%$



# Порівняння моделей М1-М3

$$\ln h_k(t, x(t - k), \beta) = \sum_{i=1}^5 \beta_i x_i(t - k) + \sum_{i=6}^{21} \beta_i x_i$$

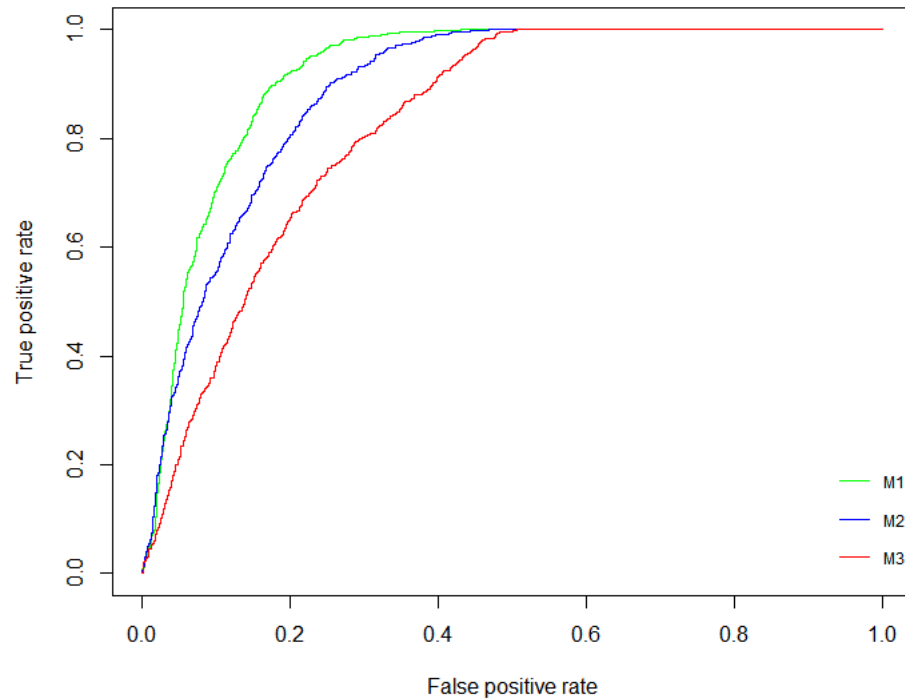
Динамічні

Статичні

Тип змінних	Скорочена назва змінної	Оцінені коефіцієнти моделі Mi		
		M1	M2	M3
Динамічні	log(1 + outbody(t-i))	-0.1984	-0.1679	-0.0500
	log(1 + ovd(t-i))	1.0041	0.9217	0.6563
	log(1 + limit(t-i))	-0.1422	-0.1458	-0.2052
	log(1 + montake(t-i))	-0.8695	-0.5845	-0.0513
	log(1 + abs(monpay(t-i)))	-0.2420	-0.1002	-0.0302
Статичні	clienttypeClient	0.2738	0.0777	0.3777
	clienttypeEmployee	0.3499	0.0680	-0.4239
	clienttypeInsider	0.9472	0.8103	0.6047
	clienttypeNormal	0.2748	0.1149	0.3139
	clienttypeOldSalary	0.6650	0.4982	0.5961
	clienttypeSalary	0.6318	0.1192	-0.0524
	age	-0.0021	0.0018	0.0053
	is_satisfiedt	-0.2529	-0.3165	-0.2657
	liv_is_regt	0.1521	0.1887	0.1976
	childcnt	0.0275	-0.0015	0.0083
	dependantcnt	-0.0610	-0.0747	-0.0593
	log(1 + income)	-0.0862	-0.0363	0.0012
	has_u_scoret	0.1002	0.1747	0.4204
	with_bank_mon	-0.0142	-0.0146	-0.0087
	reg_mon	-0.0040	-0.0077	-0.0128

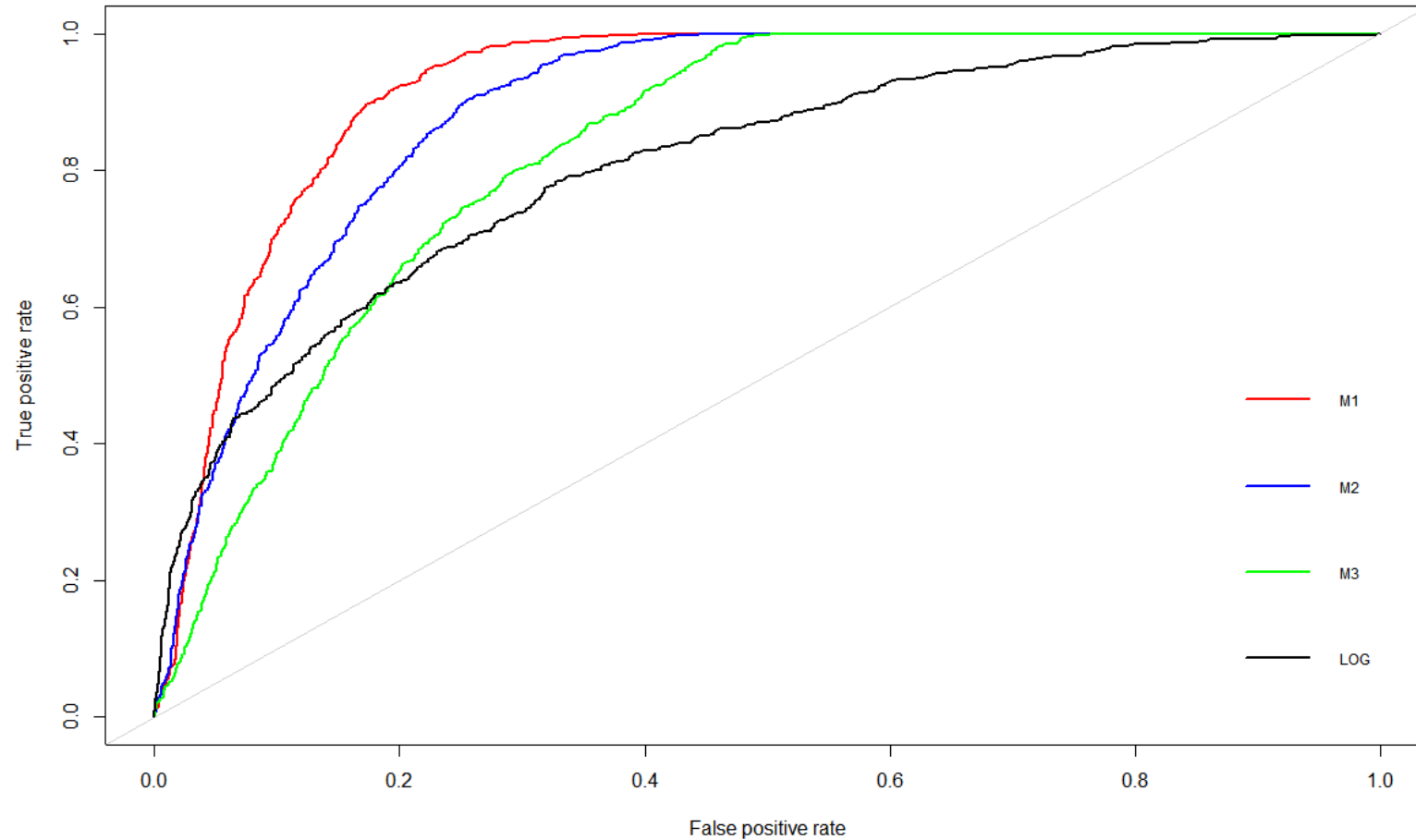


# Вибір кращої моделі (M1-M3)



Mi	M1	M2	M3
AUC	0.92	0.89	0.829
GINI	83.78%	77.55%	65.78%
AIC	7495.905	7819.972	8946.005
WT	1410.135	1384.005	1308.561

# Порівняння моделей РН та логістичної регресії



# Публікація результатів

---

На основі проведеного дослідження написана стаття. Отримані результати будуть опубліковані в Університетський науковий електронний збірник «Системні науки та кібернетика» №5, 2016 (сторінки 56 – 68).

Збірник створено для публікації студентських наукових статей при проведенні науководослідної роботи у магістратурі.

З ними можна ознайомитися за посиланням: [www.mmsa.kpi.ua/ssc](http://www.mmsa.kpi.ua/ssc)

# Висновки

---

У даній роботі проведено огляд різних підходів до оцінювання платоспроможності клієнтів-власників кредитних карт.

У процесі проведення дослідження було:

- ❑ визначено, що поведінковий скоринг – найбільш зручний та універсальний спосіб порівняння клієнтів між собою.
- ❑ виявлено, що стандартні підходи до вирішення цієї задачі не забезпечують необхідного рівня ефективності.
- ❑ розроблено та побудовано альтернативну модель, що виявилась значно кращою порівняно із загальноприйнятими.
- ❑ виявлено переваги подібного підходу, що полягає перш за все в можливості прогнозування поведінки клієнта.
- ❑ наведено практичні рекомендації до побудови моделей (у вигляді набору команд мови R).

# Перспективи для подальших досліджень

---

До подальших перспектив можна віднести:

- застосування пропорційних моделей Кокса до аплікаційного скорингу та скорингу шахрайства.
- комбінування оцінок різних типів моделей для покращення предикативності, зокрема з моделями на основі нейронних мереж та SVM.
- аналіз предикативних можливостей побудованих моделей
- аналіз залежності предикативності від додавання нових параметрів або агрегуванні інших, що в теорії забезпечить використання переваг як статичності логістичної регресії, так і динамічності моделей Кокса.

Дякую за увагу!