

Скорингові моделі поведінки гравців для оцінки фінансових показників компанії

Фомін Олександр, КА-61м

Інститут прикладного
системного аналізу

Керівник: к.т.н., доц.
Кузнєцова Наталія

Актуальність теми



Класичний підхід скорингу поведінки

- **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 завдання

- **Мета роботи:** розробка моделей та методів моделювання поведінки клієнтів та їх порівняння із існуючими загальноприйнятими
- **Завдання:**
 - розглянути існуючі методи побудови скоринг-моделей;
 - виявити найбільш актуальні та перспективні підходи до розробки моделей;
 - розробити власні моделі поведінки клієнтів-гравців онлайн-платформи;
 - порівняти отримані результати та зробити висновки;
 - розробити прототип системи прийняття рішення для розв'язку схожих завдань у подальшому.

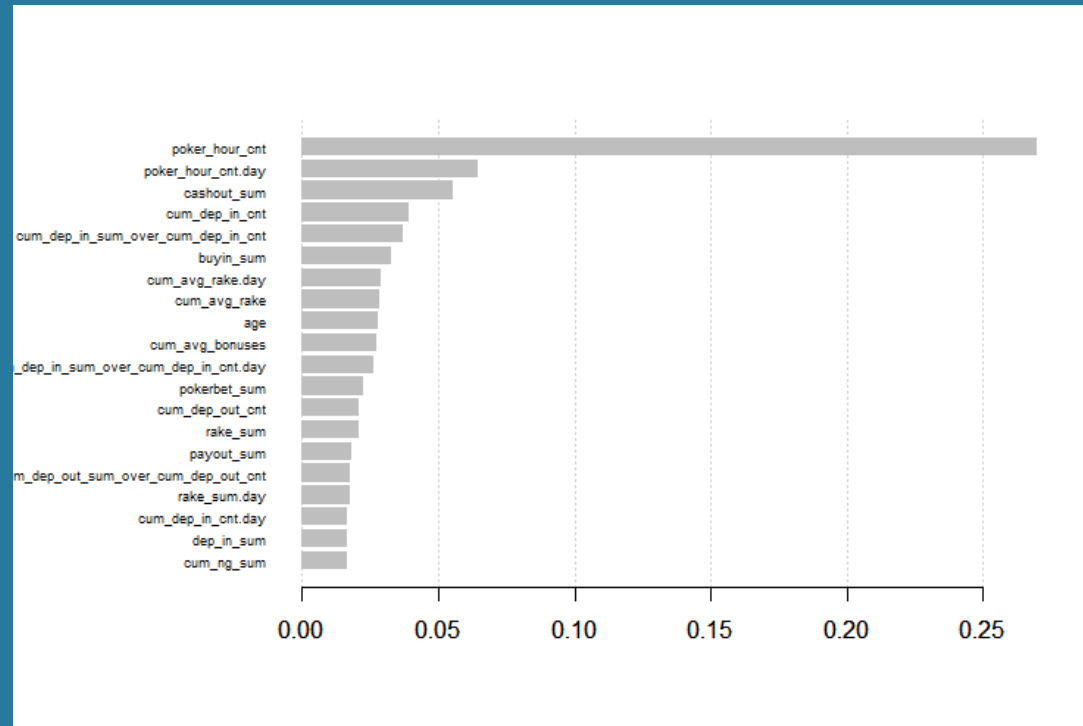


Модель оцінки ймовірності відтоку. Теорія

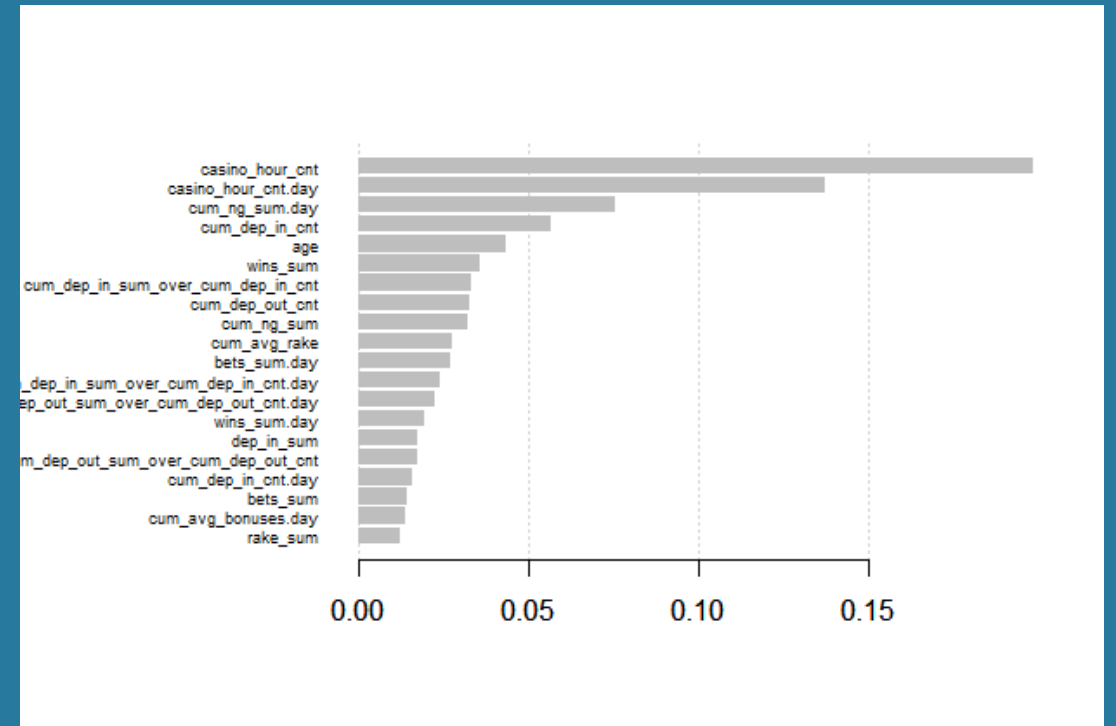
- $y_i^* = \phi(x_i) = \sum_{k=1}^K f_k(x_i)$, $f_k \in F$ – простір дерев прийняття рішень;
- $L(\phi) = \sum_i l(y_i^*, y_i) + \sum_k \Omega(f_k)$, $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$;
- $L^{(t)} \approx \sum_{i=1}^n \left[l(y_i, y_i^{*(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$;
- $L_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma$.

Модель оцінки ймовірності відтоку

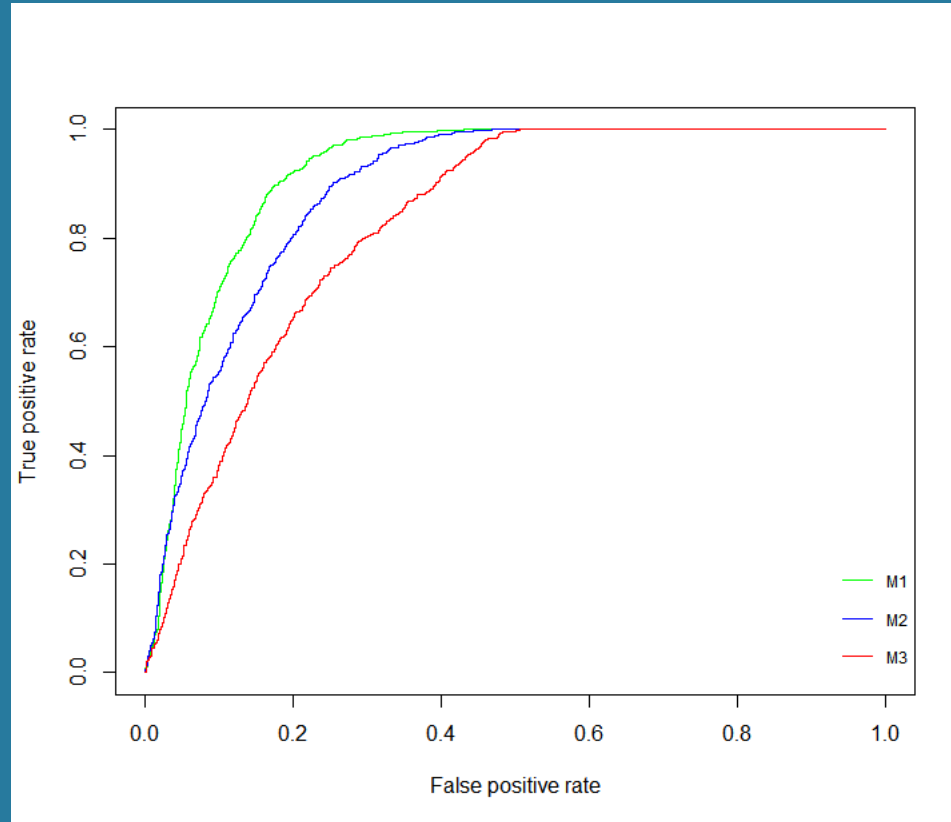
Покер



Казино



Порівняння із логістичною регресією

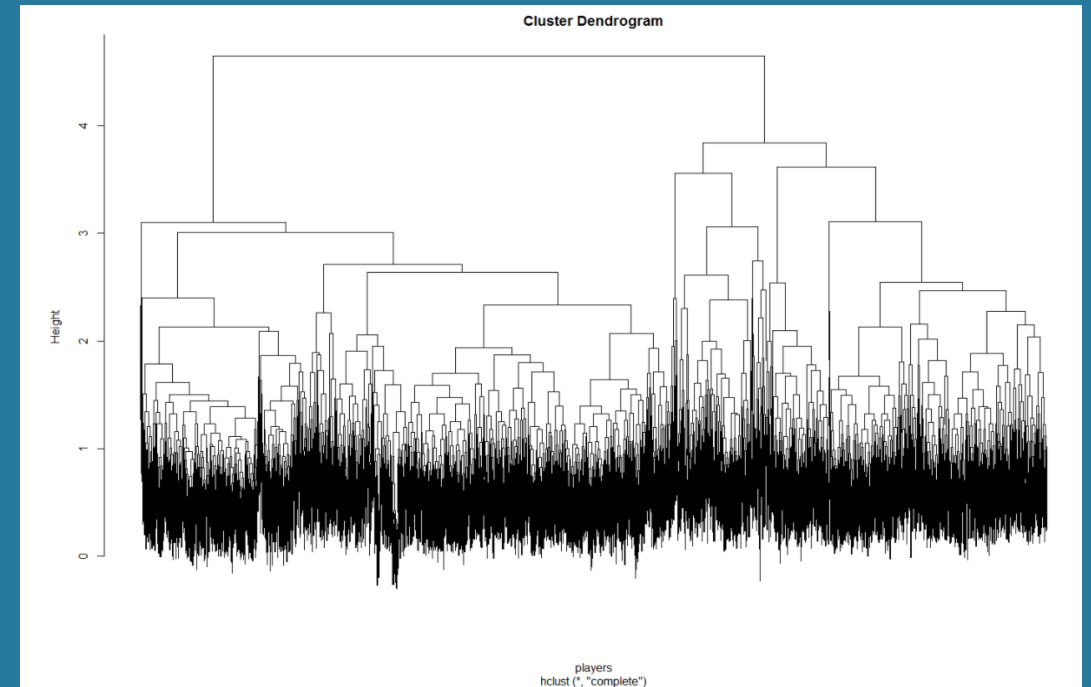
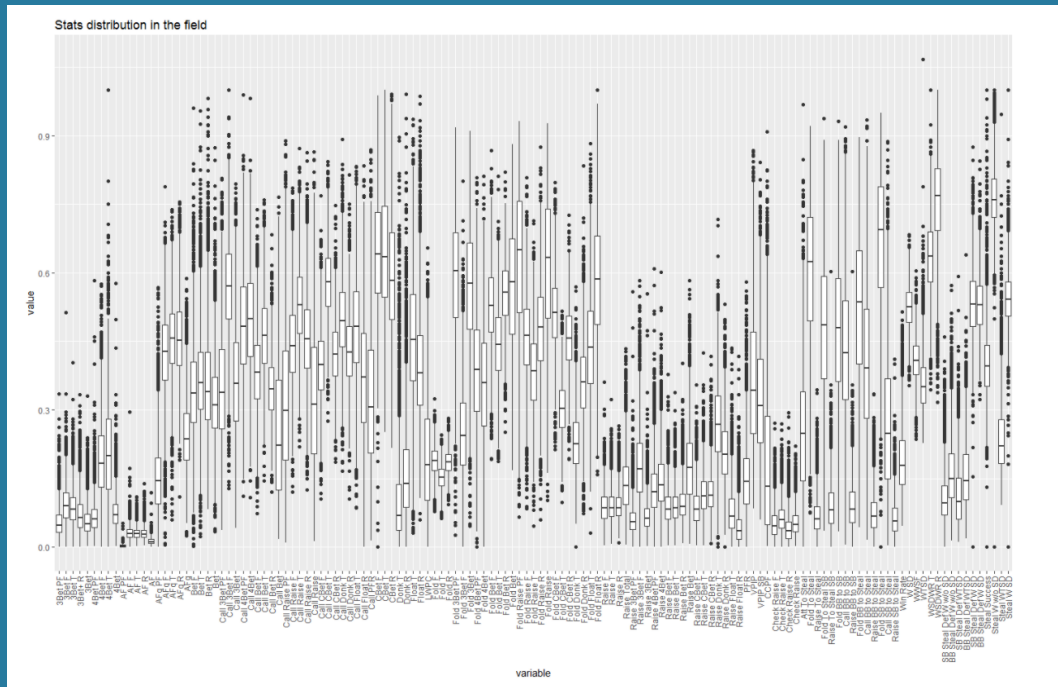


	M1	M2	M3
Покер	86%	71%	75%
Казино	82%	70%	72%

Модель виживання. Теорія

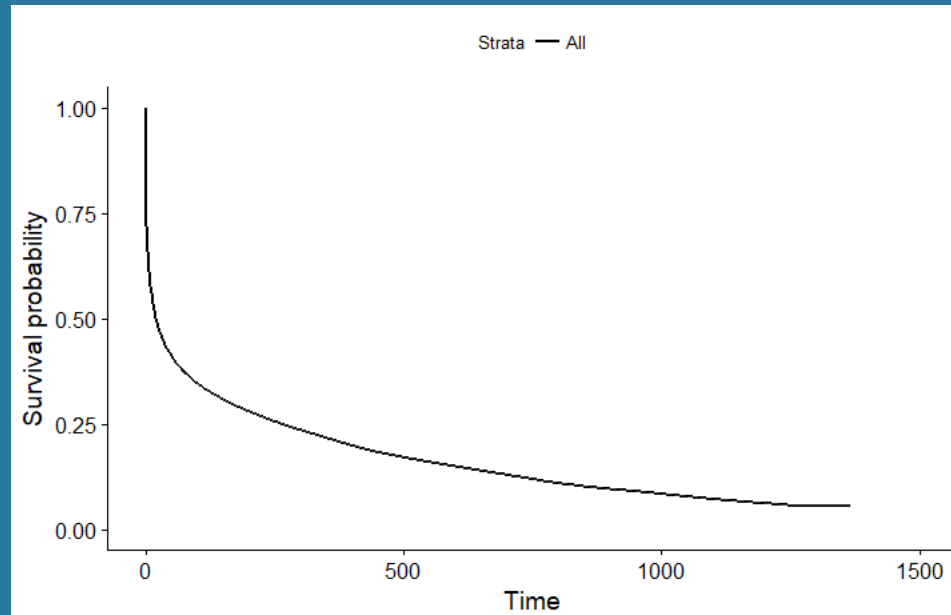
- $h(t, x(t), \beta) = h_0(t)e^{\beta^T x(t)} ;$
- $\mathbb{P}(T \geq t) = \frac{h(t, x_i(t), \beta)}{\sum_{j \in R(t_i)} h(t_i, x_j(t), \beta)} = \frac{e^{\beta^T x_i(t)}}{\sum_{j \in R(t)} e^{\beta^T x_j(t)}} , R(t) = \{j | t_j \geq t\}.$

Класифікація покерних гравців

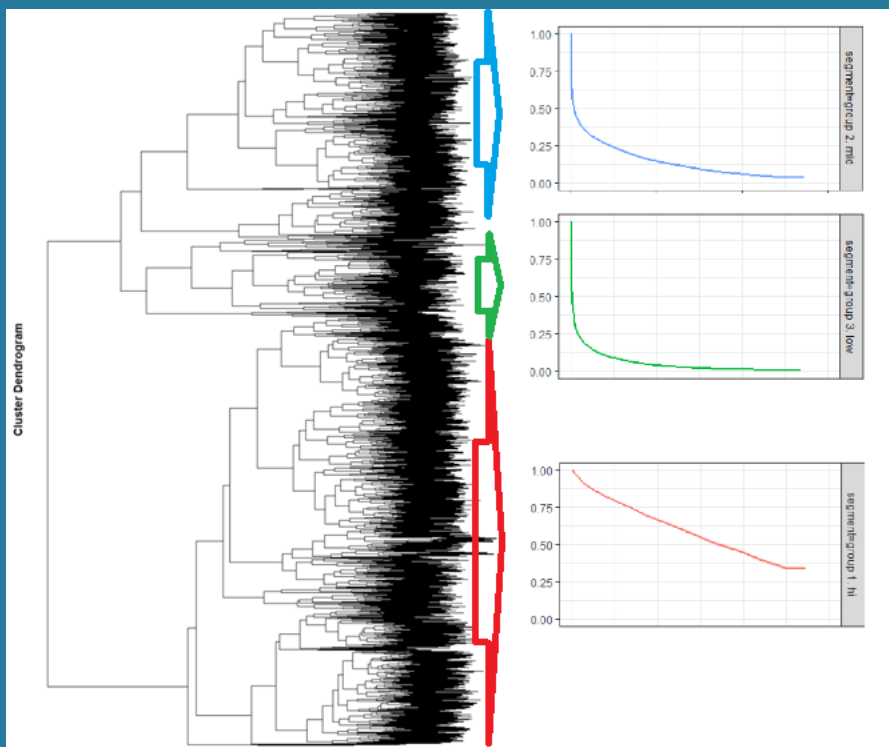


Модель виживання для всієї популяції

- Медіанний час життя – 7 днів:



Модель виживання



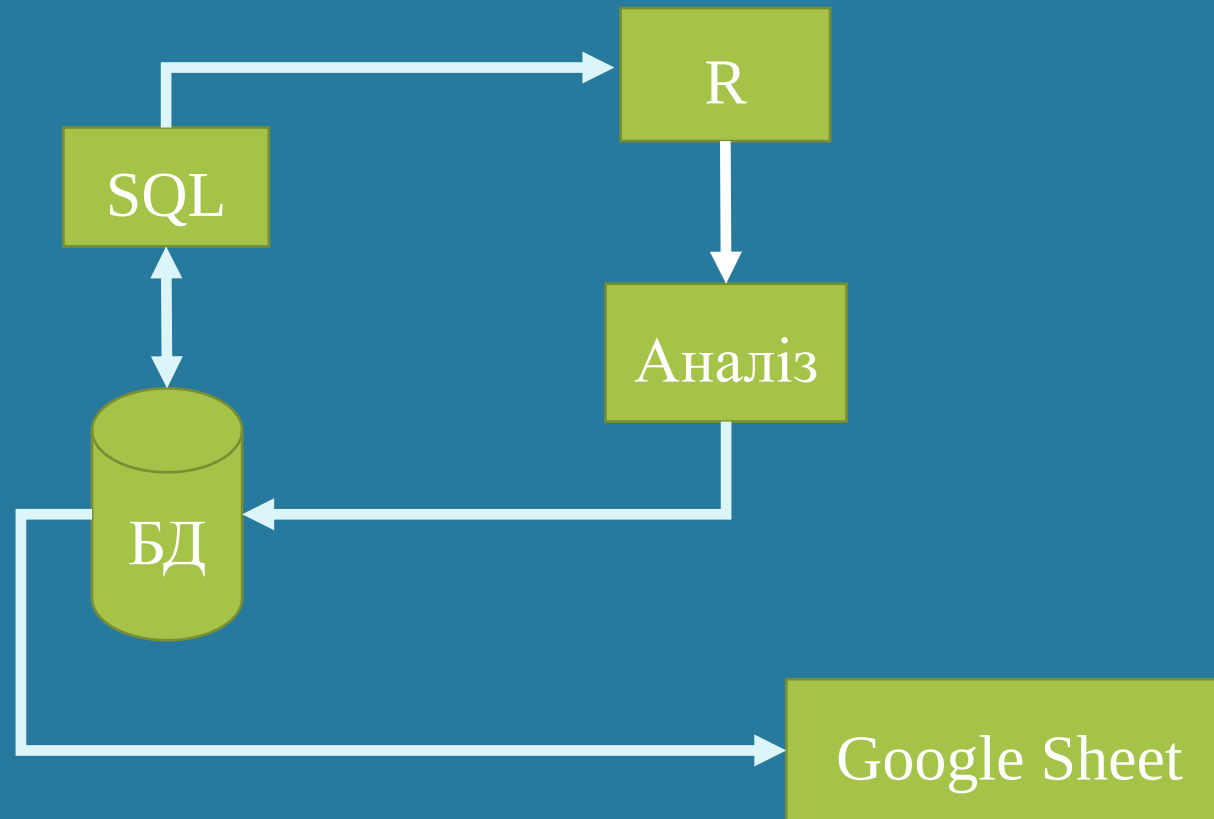
- Медіанний час життя – 3 дні;
- Медіанний час життя – 7 днів;
- Медіанний час життя – 100 днів;

Оцінка можливих втрат доходів

$$E = \sum_{i=1}^n PD_i \cdot (T_i^{(m)} - T_i) \cdot R_{avg_i}$$

- PD_i – ймовірність відпаду;
- $T_i^{(m)}$ – медіанний час життя клінта;
- T_i – це час життя на момент розрахунку;
- R_{avg} – середньоденний дохід від клієнта

Прототип СППР для автоматизації процесу



Висновки

- Побудовано модель на основі бустингового ансамблю дерев, що значно краще оцінює ймовірність відвалу порівняно із логістичною регресією;
- Здійснено класифікацію клієнтів;
- Проведно моделювання часу життя клієнта на основі проведеної класифікації, що значно точніше описує гравців;
- Результати проведеного аналізу синтезовано в показник оцінки можливих втрат доходу, як один із основних фінансових показників;
- Імплементовано прототип СППР на основі вище описаних експериментів.

Перспективи для подальших досліджень

- Проведення аналізу інших ансамблів дерев, таких як random forest та моделей на основі нейронних мереж;
- Дослідження можливості кластеризації гравців казино;
- Розвиток прототипу СППР.

Дякую за увагу