

Семинарски рад из Статистичког софтвера 1

Истраживање о одливу клијената у банкарском сектору

Стефан Малбашић

3. јануар, 2024.

Одлив клијената у банкарском сектору (енг. Bank Customer Churn) дешава се када клијенти престану да користе производе и услуге које банка нуди на одређено време, а затим прекину свој однос с банком. Због тога је задржавање клијената од суштинског значаја на данашњем изузетно конкурентном банкарском тржишту. Осим тога, стабилна база клијената помаже у привлачењу нових корисника јачањем поверења и препорука постојећих корисника.

Истраживање се спроводи како би се идентификовали кључни фактори који утичу на одлуку клијената да напусте банку. Разумевање понашања клијената омогућава банкама да унапреде програме лојалности и задрже већи број корисника, што је значајно јефтиније од аквизиције нових клијената. Ова анализа је посебно важна за креирање циљаних маркетиншких стратегија, побољшање корисничког искуства и повећање укупне профитабилности. Најновији доступни подаци на основу којих је потребно извршити анализу налазе се на линку:

[Преглед показатеља о одливу клијената \(Customer-Churn-Records.csv\)](#)

У Табели 1. налази се опис варијабли приказаних у приложеној бази података.

| Назив варијабле | Опис |
|-------------------|--|
| RowNumber | Редни број записа |
| CustomerId | Идентификациони број клијента |
| Surname | Презиме клијента |
| CreditScore | Кредитни скор клијента |
| Geography | Локација клијента |
| Gender | Пол клијента |
| Age | Старост клијента |
| Tenure | Број година колико је клијент корисник услуга банке |
| Balance | Стање на рачуну |
| NumOfProducts | Број производа које клијент користи у банци |
| HasCrCard | Да ли клијент има кредитну картицу (1 - да, 0 - не) |
| IsActiveMember | Да ли је клијент активан корисник (1 - да, 0 - не) |
| EstimatedSalary | Процењена плата клијента |
| Exited | Да ли је клијент напустио банку (1 - да, 0 - не) |
| Complain | Да ли је клијент имао жалбу (1 - да, 0 - не) |
| SatisfactionScore | Оцена задовољства клијента у вези са решавањем жалби |
| CardType | Тип картице коју клијент користи |
| PointsEarned | Број поена које је клијент зарадио коришћењем кредитне картице |

Табела 1: Преглед показатеља о одливу клијената

АНАЛИЗА ПОДАТАКА - ЗАДАЦИ

ЗАДАТАК 1 [2п]. Одредити основне дескриптивне статистике (средња вредност, стандардна девијација, минимум и максимум) за показатеље CreditScore, Age, Tenure, Balance и EstimatedSalary.

ЗАДАТАК 2 [1п]. Приказати на једном графичком прозору хистограме густина за све нумеричке показатеље о одливу клијената из банке.

ЗАДАТАК 3 [1п]. Категоричке показатеље дефинисати као варијабле факторског типа.

ЗАДАТАК 4 [2п]. Формирати листу идентификационих бројева клијената тако да први елемент листе садржи 10 клијената са највишим кредитним скором, други елемент садржи 5% клијената који су најдуже корисници услуга банке, а трећи елемент садржи оне клијенте чији је баланс једнак 0\$.

ЗАДАТАК 5 [2п]. Имплементирати функцију `GeoBalance()` која рачуна просечно стање на рачуну клијената за сваку од земаља.

ЗАДАТАК 6 [3п]. Приказати на тракастом дијаграму број клијената који су напустили банку и оних који нису у зависности од пола клијената, од тога из које државе су клијенти, броја производа које клијенти користе у банци као и тога да ли су активни или не. Интерпретирати добијене резултате.

ЗАДАТАК 7 [3п]. Приказати кутијасте дијаграм стања на рачуну (`Balance`) и процењене плате (`EstimatedSalary`) клијената који су напустили банку и оних који нису. Да ли стање на рачуну и плата утичу на одлазак клијената из банке? Одговор образложити.

ЗАДАТАК 8 [3п]. Приказати график густине (енг. `density plot`) показатеља `CreditScore`, `Age`, `Balance`, `EstimatedSalary` и `PointsEarned` за клијенте који су напустили банку и оне који нису. Шта закључујете на основу ових графика?

ЗАДАТАК 9 [3п]. Анализирајте да ли постоји веза између оцене задовољства (`SatisfactionScore`) и одласка клијената, као и како поседовање кредитне картице (`HasCrCard`) утиче на одлазак клијената.

ЗАДАТАК 10 [2п]. Истражите да ли постоји значајна разлика у просечној вредности оцене задовољства (`SatisfactionScore`) између клијената који су остали и оних који су напустили банку користећи `t-test`.

ЗАДАТАК 11 [5п]. Клијенти са високим кредитним скором су они који имају скор већи од 840. Издвојити подскуп клијената са високим кредитним скором и приказати график густине скор за клијенте који су напустили банку и оне који нису. Упоредити резултате добијене са овом подскупу са резултатима на читавом скупу података. Колики је проценат клијената који напусте банку у оба посматрана скупа? Упоредити симетрију и спљоштеност расподеле на оба скупа. Образложити закључке.

ЗАДАТАК 12 [5п]. Филтрирати клијенте у групе по стању на рачуну тако да су подеоне тачке следећи биланси: 0\$, 50000\$, 100000\$, 150000\$, 200000\$, 250000\$. Размотрити клијенте чији је баланс једнак 0\$ и оне чији је баланс већи од 250000\$ (укупно 7 група). Извршити анализу одласка клијената из банке по датим групама. Шта закључујете? Да ли клијенти чији је баланс једнак 0\$ значајно утичу на стопу одласка клијената из банке или не?

ЗАДАТАК 13 [5п]. Предложити поделу клијената у старосне групе у зависности од година (`Age`), а затим извршити анализу одласка клијената из банке по датим групама.

ЗАДАТАК 124 [4п]. Испитати графички и аналитички да ли стање на рачуну (`Balance`) клијената чији је баланс већи од 50000\$ прати неку нормалну $\mathcal{N}(m, \sigma^2)$ расподелу. Ако да, одредити 95% интервал поверења за непознате параметре m и σ^2 .

ЗАДАТАК 15 [4п]. За функцију расподеле случајне величине W важи да је

$$F_W(w; \alpha, m, s) = \exp\left(-\left(\frac{w-m}{s}\right)^{-\alpha}\right), \quad w \geq m.$$

Методом инверзне функције генерисати 5000 вредности случајне величине W где је скуп параметара функције расподеле дат са $(\alpha, m, s) = (0.02, 34.27, 8.38)$, а затим на случајан начин издвојити 5000 вредности показатеља `Age` и испитати независност ова два обележја.

ЗАДАТАК 16 [2п]. Издвојити базу `Clients.Spain` која садржи све клијенте из Шпаније. Из нове базе издвојити све клијенте млађе од 20 или старије од 70 година.

ЗАДАТАК 17 [3п]. Сортирати клијенте у опадајућем поретку по стању на рачуну. Издвојити само активне чланове (`IsActiveMember`), па их сортирати по презимену (`Surname`), па у опадајућем поретку по просечном број бодова (`PointsEarned`) по години чланства (`Tenure`).

ЗАДАТАК 18 [5п]. Монте Карло симулацијама проценити очекиване губитке услед одлива клијената из банке. За процену губитка по клијенту користити неку линеарну комбинацију стања на рачуну (`Balance`) и процењене плате (`EstimatedSalary`). Приказати густину очекиваних финансијских губитака хистограмом.

ЗАДАТАК 19 [5п]. Симулирати какав ефекат имају различити сценарији на стопу одлива клијената и упоредити са реализованим вредностима.

- сценарио 1: клијенти чији је кредитни скор мањи од 500 напуштају банку са вероватноћом 0.85;
- сценарио 2: клијенти старији од 60 година напуштају банку са вероватноћом 0.75;
- сценарио 3: клијенти чије је стање на рачуну веће од 100000\$ и активни су корисници напуштају банку са вероватноћом 0.98.

Предложити још један сценарио који је у складу са резултатима истраживања и испитати његове ефекте на стопу одлива клијената.

ЗАДАТАК 20 [5п]. Користећи Монте Карло симулације процените број клијената који ће напустити банку у наредних годину дана. Претпоставите да је вероватноћа одлива клијената логистичка функција показатеља као што су Age, CreditScore, Balance, и IsActiveMember. Одредити 95% интервал поверења за број клијената који остају у банци и број оних који напуштају.

СЕМИНАРСКИ РАД - НАПОМЕНЕ

НАПОМЕНА 1. Рок за израду семинарског рада је крај испитних рокова. У зависности од испитног рока у којем желите да браните семинарски рад, накнадно ће бити договорен термин за одбрану са колегама који пошаљу радове. Радове слати на мејл stefan.malbasic@matf.bg.ac.rs са назнаком "С1 - семинарски рад".

НАПОМЕНА 2. Рад је могуће послати у два формата:

- RMarkdown документ, у коме би били интегрисани код и детаљна анализа заједно (препоручени формат), детаљније на линку: rmarkdown.rstudio.com;
- LaTeX документ и сл, у коме би било описано шта је урађено са детаљним закључцима, а искоментарисан код као додаток у обичном .r фајлу.

НАПОМЕНА 3. За визуализацију података је потребно користити пакет ggplot2 јер семинарски рад осликава један мањи истраживачки пројекат који треба да буде прегледан и лако интерпретабилан.

НАПОМЕНА 4. У оквиру дефинисаних задатака дозвољено је (и пожељно) одрадити детаљнију анализу траженог сегмента а у циљу што прецизнијих и интерпретабилнијих резултата.

НАПОМЕНА 5. Семинарски рад се ради самостално. Дозвољено је коришћење материјала за курс, као и додатне литературе која је на истим истакнута. На одбрани је могуће добити и питање из градива које је излагано у току курса.

НАПОМЕНА 6. Рад треба да садржи закључак са сажетом и прегледном интерпретацијом добијених резултата.