

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
МЕХАНІКО-МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ
КАФЕДРА ТЕОРІЇ ЙМОВІРНОСТЕЙ,
СТАТИСТИКИ ТА АКТУАРНОЇ МАТЕМАТИКИ

КУРСОВА РОБОТА

НА ТЕМУ:

“Аналіз розривів договорів страхування життя за допомогою методів класифікації, за допомогою методів заснованих на регресійних деревах”

Виконала:
студентка 4 курсу
механіко-математичного факультету
групи «Статистика»
Скворцова Єлизавета Олександрівна

Науковий керівник:
доктор фізико-математичних наук
професор кафедри теорії ймовірностей,
статистики та актуарної математики
Голомозий Віталій Вікторович

КИЇВ - 2022

Зміст

| | |
|---|----|
| Вступ | 3 |
| Структура даних..... | 4 |
| Аналіз даних..... | 6 |
| Підготовка даних для моделі | 8 |
| Переробка усіх текстових регресорів на числові..... | 8 |
| Зміна причини розривів..... | 8 |
| Залишаємо тільки «потрібні» регресори | 8 |
| Поділ вибірки на train та test | 9 |
| Побудова моделі прогнозування розриву за допомогою методу класифікації Random Forest Classifier | 9 |
| Перша побудова моделі..... | 9 |
| Друга побудова моделі | 9 |
| Аналіз регресорів. | 10 |
| Третя побудова моделі | 10 |
| Висновок..... | 11 |
| Побудова моделі прогнозування кварталу розриву за допомогою методу класифікації Random Forest Classifier..... | 11 |
| Перша побудова моделі..... | 11 |
| Висновок..... | 12 |
| Порівняння точності прогнозу кварталу розриву з іншими методами класифікації | 12 |
| Висновок..... | 12 |
| Використана література | 13 |
| Додатки | 14 |
| Додаток 1 | 14 |
| Додаток 2 | 15 |
| Додаток 3 | 15 |

Вступ

Випускаючи продукт, страхова компанія прогнозує на багато років вперед прибуток і втрати, щоб зрозуміти наскільки вигідно інвестувати у нові проекти.

Приймаючи людину на страхування життя, страхова компанія аналізує багато ризиків. Окрім розрахунку тарифу на 20 років вперед, андерайтери оцінюють здоров'я людини на даний момент.

І усі ці розрахунки мають за собою базу. Тарифи рахується на таблицях смертності, андерайтери також мають вже таблиці надбавок в залежності від професії людини чи стану здоров'я.

Але якщо розглядати побудову P&L, є два показники, які не отримуються просто з таблиці. Це реальна відсоткова ставка на 20 років вперед та рівні розривів договорів. Зауважимо, що від розривом договору мається на увазі не тільки викуп договору, а і просто припинення надходження платежів, адже для компанії критично зрозуміти скільки грошей вона отримає.

Для прогнозу відсоткової ставки вже є доволі багато моделей, або ніхто і не відміняв прогноз на око. Але ось рівень розривів має дуже суттєвий вплив на прибуток, особливо в перші роки існування договору (статистично найбільший відсоток розривів договору страхування життя припадає на перші 5 років).

Рівень розривів використовується на кожному кроці у лайфовій страховій. Від побудови бюджету на наступний рік, перевірки адекватності резервів, розрахунку Value of new Business (VNB) чи Embedded Value (EV), до звичайної перевірки доброчесності агентів чи самого продукту. Зазвичай тут достатньо використовувати емпіричні значення проаналізовані за декілька минулих років, розбиті тільки на Програму, Канал продажу та генерацію договору, адже це має найбільший вплив. Далі при аналізі даних ми подивимося на скільки є суттєва різниця у цих параметрах.

Але чому коли андерайтери аналізують здоров'я людини чи професію, страхова не аналізує додатково платоспроможність людини, а покладається тільки на агентів у цьому питанні? Маючи взагалі доволі слабкий ринок страхування життя в Україні (5-7% населення, для порівняння у США станом на 2021 рік 52%^(Додаток 2)), страхові компанії нехтують перевіркою клієнтів. Процедури спрощуються, адже треба якнайшвидше і простіше для клієнта укласти якомога більше договорів.

Але проблемою є не тільки видача договорів, а і їх супроводження. Часто, у клієнтів виникають ситуації, при яких він не має більше змогу сплачувати платежі по договору. На такі випадки звичайно є можливість редукування чи розстрочки.

Редукування – призупинення сплати внесків клієнтом, при цьому по договору розраховується нова редукована страхова сума на базі накопиченого резерву. Зазвичай у таких випадках припиняється покриття додаткових ризиків (по яких резерв розраховується на рік, як звичайне ризикове страхування).

Розстрочка – практика деяких страхових, де клієнту надається змога розбити платіж на дві частини. Перша (ризикова, тобто платіж по смерті і додаткових ризиках) сплачується на початку року, а друга (частина платежу по дожиттю) сплачується у кінці року.

Але як я згадувала раніше, редукування ми будемо розглядати як розрив, адже платежі перестали надходити. У такі моменти для клієнта важливо побачити, що компанії не байдуже. Якщо людині нагадати як є важливим страхування життя, показати турботу є більший шанс утримати клієнта (нехай навіть, платіж зменшиться, але це краще ніж нічого). Причому, зауважимо, що часто редукування робиться автоматично,

без дзвінка людини і дзвонити чи писати вже після несплати внеску – пізно, людина просто буде ігнорувати.

Тому ми ставимо дві основні задачі у цій роботі:

- 1) Спрогнозувати чи розірветься договір. Тобто зрозуміти наскільки ризикову приймати ту, чи іншу людину на страхування. Таким чином класифікувати договори страхування на ризикові і ні.
- 2) Спробувати спрогнозувати квартал розриву договору.

Обидві ці задачі будуть реалізовані за допомогою методів машинного навчання для класифікації. Перший метод, який буде використовуватися: *Random Forest Classifier*.

Але так як він доволі схилений до перенавчання, також будуть використані такі методи як *K-nearest Neighbors* та *Naive Bayes*, але тільки для порівняння.

Структура даних

Для побудови моделі будуть використовуватися дані договорів лайфвої страхової компанії, з датою укладання починаючи з 2005 року по 2022 рік. Кількість договорів, що аналізується, 318 552. Використовуються різні показники, за якими будемо будувати модель.

Показники, що будуть прогнозуватися

- 1) Причина розриву (0 – не було розриву, 1 – договір закінчено по дожиттю, 2 – договір закінчено через настання страхового випадку, 3 – договір редуковано, 4 – договір розірвано, 5 – розірвано після редукування). Ми будемо загалом аналізувати (<3) та (≥ 3), тобто чи був розрив, чи ні.
- 2) Кількість кварталів до розриву. Для цього параметру і вводилися причини розриву договору.

Подивимося на структуру за більшістю з цих показників.

Далі переглянемо структуру договорів але не за кількістю, а за «важливістю». А саме за частиною внесків по параметрах (програма, канал, тощо).

Компанія має договори у різних 66-х програмах. Структура договорів у розбитті по програмах:

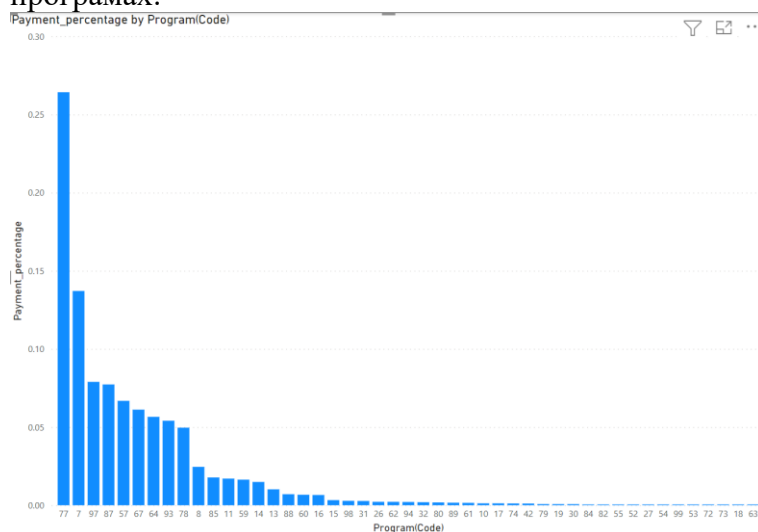


Рисунок 1. Структура портфелю за програмами

Одразу бачимо основні програми: 77, 7, 57, 64, 67, 87, 93, 97, 78.

Також подивимось структуру портфелю по каналах (котрих 13):

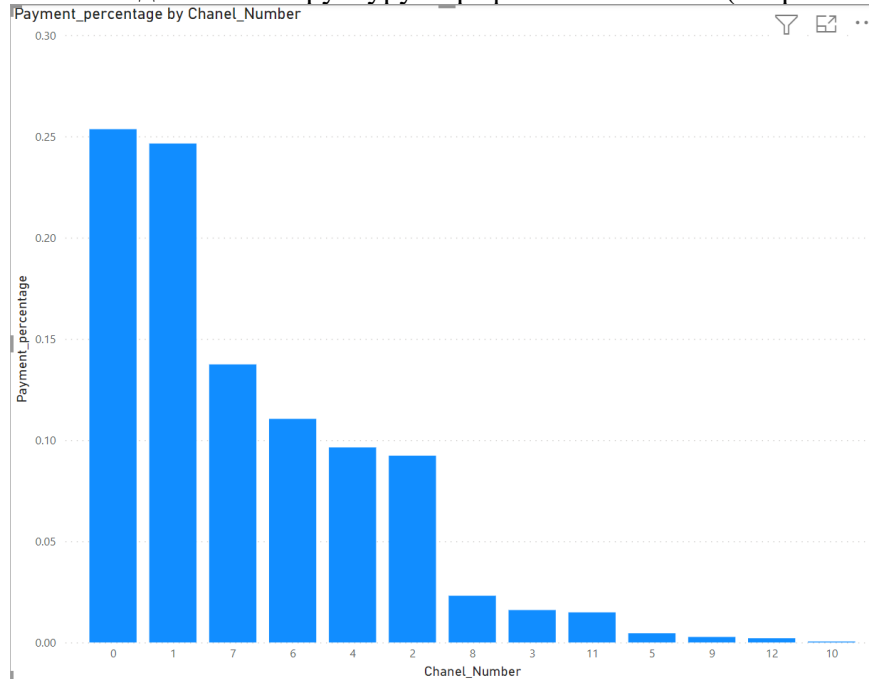
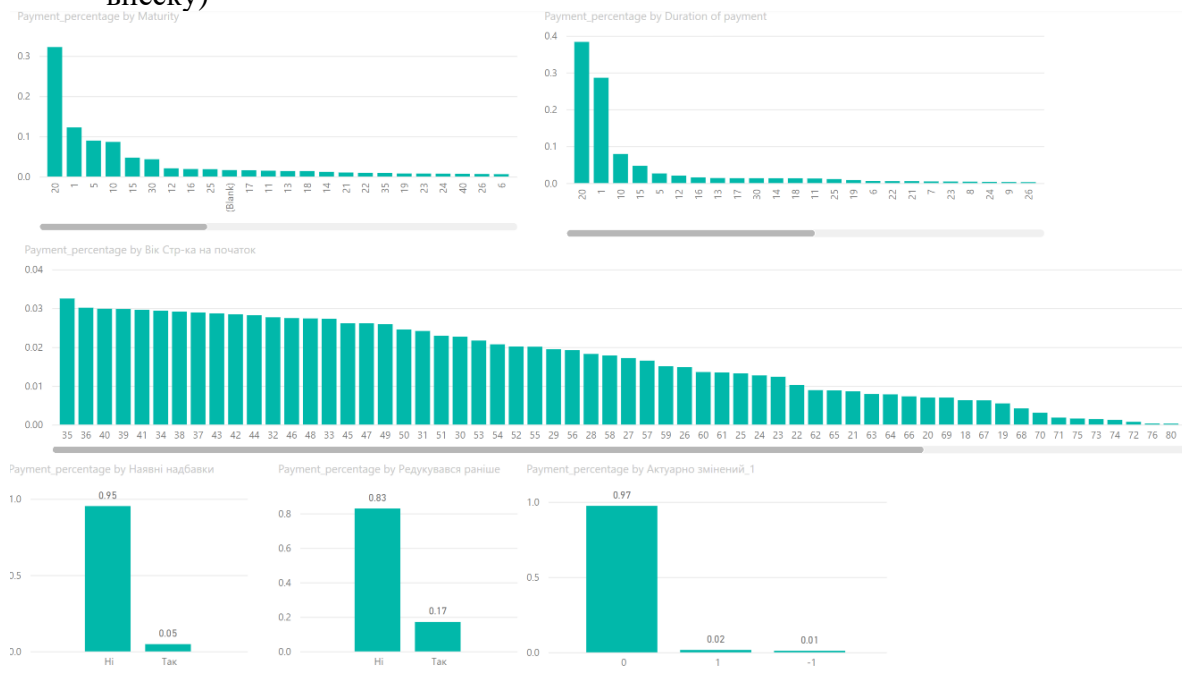


Рисунок 2. Гістограма структури портфелю по каналах

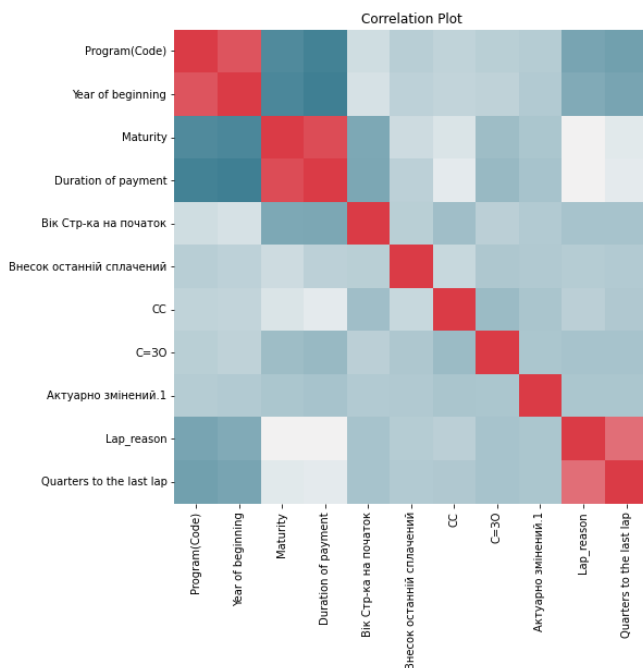
Бачимо, що домінує 5-й канал, і що також важливими є: (0, 1, 2, 4, 6, 7).

Також переглянемо структуру портфелю у розрізі:

- 1) Строк дії договору
- 2) Строку сплати внесків
- 3) Віку страхувальника на момент укладання
- 4) Чи наявні надбавки по договору
- 5) Чи редукувався договір до цього (редукування на даний момент не враховується)
- 6) Чи були актуарні зміни (1 – збільшення внеску, 0 – не було змін, -1 – зменшення внеску)



Також переглянемо лінійну кореляцію Пірсона у основних показниках.



Бачимо, що сильна кореляція має місце у таких показниках:

1) Рік укладання і програма. (Очевидно, продукти змінюються і продуються нові з часом). Тож можемо прибрати з аналізу рік укладання як регресор. (або розглянемо як у пункті 2))

2) Програма і строки договору та строк сплати платежів. Тут також очевидна кореляція, адже різні програми зосередженні на різні строки (Страхування життя класичне та пенсійне, наприклад). Для того, щоб прибрати цей момент, подивимося як буде себе показувати моделі, якщо розділити їх для різних програм.

3) Причина розриву і квартал розриву. Про це я вже робила зауваження вище.

4) Також цікавим моментом є кореляція між Віком страхувальника та строком укладання договору, але вона не є дуже сильною.

Тепер, що дивно, лінійна кореляція відсутня для причини розриву та строку дії. Проаналізуємо цей момент окремо, використовуючи χ^2 тест, та при цьому замінимо значення від 0 до 5 на бінарну змінну, що приймає значення 0 та 1, де 0 – не було розриву, а 1 – був розрив.

У результаті, $p - value \approx 0$, що означає, що нульова гіпотеза про незалежність не правильна. Тож залишаємо даний параметр.

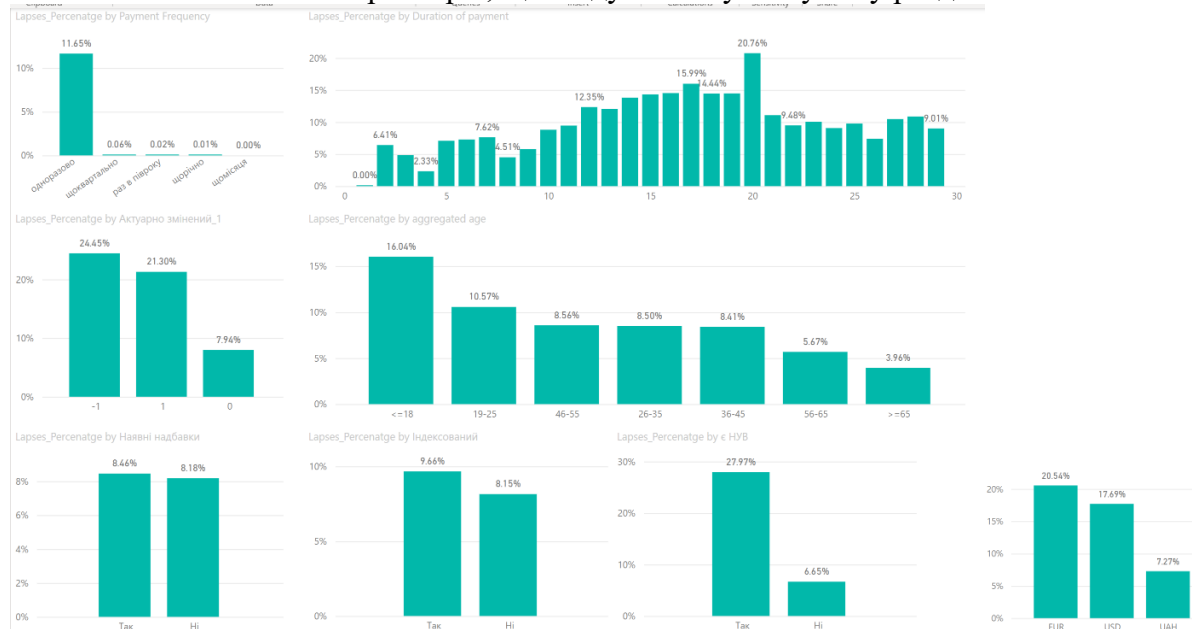
У Додатку 1 можна переглянути усі регресори, що були завантажені і згадаються впродовж роботи. Про частину з них вже було зазначено і вони не будуть використовуватися у моделі.

Аналіз даних

Проаналізуємо емпірично наші дані щодо розривності, в розрізі головних програм та каналів. З однієї сторони подивимося кількісно в яких програмах найбільше мають місце розриви. З іншої розглянемо який відсоток договорів по програмі розривається.



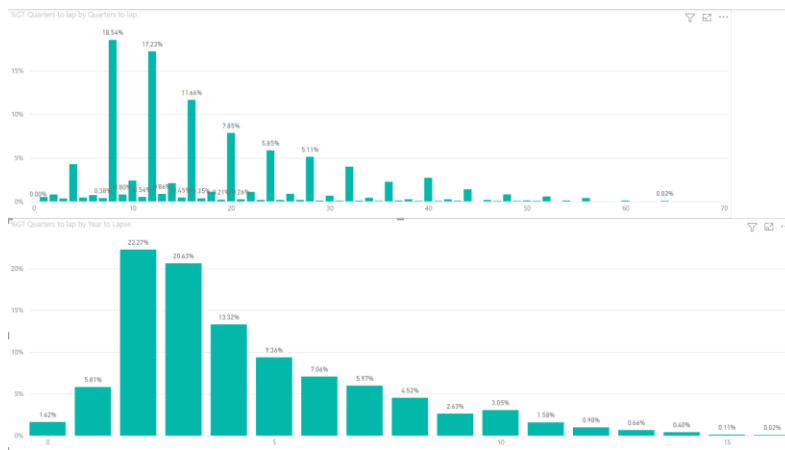
Тепер також переглянемо відсоток розривів по відношенню до договорів, що відповідають певному з параметрів, щоб побачити саме тенденцію по параметру. Розглянемо більшість з параметрів, що згадувалися у минулому розділі.



Бачимо деякі тенденції:

- 1) Договори з одноразовими платежами розриваються частіше
- 2) Більш схильні до розривів договори, де платежі сплачуються 13-21 рік.
- 3) Договори, по яким були актуарні зміни, більш схильні до розривів. Незважаючи на те, чи були актуарні зміни на збільшення внеску, чи навпаки.
- 4) Наявність надбавок не впливає на розривність.
- 5) А ось наявність НУВ (Авансових платежів) показує на можливість розриву.
- 6) Також частково бачимо, відсутність сильного впливу віку. (Великий відсоток для віку зумовлений малою кількістю таких договорів, адже вони виникли при зміні Страхувальника з батьків на дитину).

Також зосередимо увагу на кварталі та року розриву договору.



Квартал нас цікавить для прогнозу, але аналізувати його важко. Бачимо доволі яскравий розподіл року розриву, де найбільше розривів припадає на 3-й рік (це зумовлено тим, що страхова не виплачує викупну суму до 3-го року), і потім йде на спад.

Підготовка даних для моделі

Перед тим як будувати саму модель прогнозу нам потрібно побудувати дані так, щоб вони були придатні для обробки методами машинного навчання.

Переробка усіх текстових регресорів на числові.

1. Зміна усіх каналів на числові, що ми вже робили в розділі «Структура даних». Таким чином ми маємо замінені канали на числа від 0 до 12.
2. Зміна періодичності:
одноразово: 0,
щорічно: 1,
раз в півроку: 2,
щоквартально: 4,
щомісяця: 12
3. Змінюємо на 0 та 1 такі регресори: “Чи було редукування”, “Стать страхувальника”, “Наявність авансових платежів”, “Чи є додатково дитячий договір” (цей регресор буде використовуватися при окремому аналізі по програмах), “Чи договір додатково індексувався”, “Спосіб оплати”.
4. Також валюта була замінена на значення 0-7 з причин конфіденційності.

Зміна причини розривів

Для аналізу чи розірветься договір, прибираємо причину розриву. Таким чином причини 0-2 на 0 (не було/дожиття/смерть) та 3-5 на 1 (редукування/викуп/викуп після редукування).

Залишаємо тільки «потрібні» регресори

При попередньому аналізі ми виявили, що деякі регресори не потрібно використовувати або чи ми будемо використовувати їх пізніше при поділі на програми (див. Додаток 1). Таким чином маємо такі:

Код програми, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався договір до цього, Останній сплачений внесок, Страхова Сума, Валюта, Чи Страхувальник = Застрахований особі, Чи є авансові платежі, Спосіб оплати, Чи додатково індексований договір, Чи були актуарні зміни.

Поділ вибірки на train та test

Для того, щоб передати нашу вибірку методу Random Forest Classifier, потрібно поділити її на тренувальну та тестувальну. Загальна вибірка має розмір 318 552.

Так як даний метод дуже схилений до перенавчання, то ми почнемо з 70% на тренування. Далі будемо або змінювати параметри моделі, або розмір.

Побудова моделі прогнозування розриву за допомогою методу класифікації Random Forest Classifier

Дані вже підготовлені для побудови моделі. Для попереднього аналізу, щоб приблизно розуміти наскільки модель приблизно працює та чи вона неперенавчена будемо використовувати False Acceptance Rate (FAR) та False Rejection Rate (FRR) для тренуваних даних та тестувальних.

Додамо, що для нас кращим є результат, якщо FAR має достатній рівень, але FRR при цьому повинен бути близьким до нуля. Адже якщо модель припускає, що договір розірветься, то ми про це будемо знати і зможемо прийняти міри. Але якщо ми навпаки не будемо знати, що наче як договір не розірветься, то ми його втратимо.

Також для порівняння моделей будемо використовувати функцію score методу Random Forest Classifier, яка вказує на середню точність моделі.

Зауважимо до того ж, що поки не додаємо параметри моделі, це буде проведено на більш глибокому аналізі моделей.

Перша побудова моделі

Відсоток даних, на яких тренується: 70%.

Регресори:

Код програми, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався договір до цього, Останній сплачений внесок, Страхова Сума, Валюта, Чи Страхувальник = Застрахованій особі, Чи є авансові платежі, Спосіб оплати, Чи додатково індексований договір, Чи були актуарні зміни.

| | Train Data | Test Data |
|----------|------------|-----------|
| FAR | 1.52% | 14.0% |
| FRR | 0.03% | 0.4% |
| Accuracy | 99.85% | 98.55% |

Бачимо, що при тесті модель показала себе гірше, а отже можна припустити, що присутнє маленьке перенавчання. Тому одразу проведемо ще одну модель, але зменшимо відсоток даних, що йдуть на навчання.

Друга побудова моделі

Провівши одразу декілька тренувань, помітили, що FAR та FRR зберігаються приблизно на однаковому рівні. Тому я вирішила взяти 50% і далі покращувати модель іншими способами. Адже видно, що через великий об'єм даних, моделі достатньо 50% для тренування.

Відсоток даних, на яких тренується: 50%.

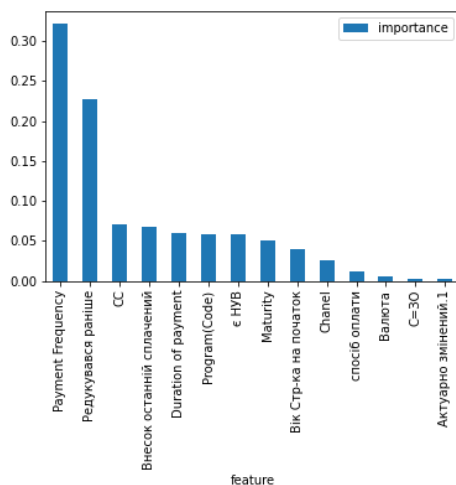
Регресори:

Код програми, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався договір до цього, Останній сплачений внесок, Страхова Сума, Валюта, Чи Страхувальник = Застрахованій особі, Чи є авансові платежі, Спосіб оплати, Чи додатково індексований договір, Чи були актуарні зміни.

| | Train Data | Test Data |
|----------|------------|-----------|
| FAR | 1.51% | 14.27% |
| FRR | 0.02% | 0.4% |
| Accuracy | 99.86% | 98.52% |

Бачимо, що результати прийнятні. Подивимось далі на вплив регресорів, наскільки вони важливі і потрібні.

Аналіз регресорів.



Одразу бачимо те, що вже частково було видно при попередньому аналізі:

1. Спосіб оплати – не впливає на нашу модель.
2. Страхувальник = Застрахована Особа – немає впливу.
3. Валюта – трохи дивно, що показує, що немає впливу, адже було видно більший відсоток для валютних договорів. Але це можливо, адже відсоток валютних договорів набагато менший.
4. Наявність попередніх актуарних змін – не впливає, що дивно також, адже ми вже помічали схильність розривати договір після.

Приберемо з моделі найменші регресори і спробуємо подивитися ще раз.

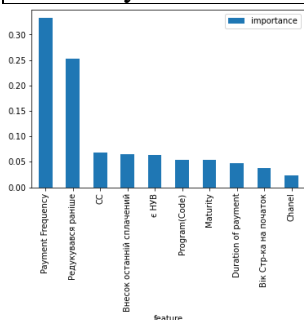
Третя побудова моделі

Відсоток даних, на яких тренується: 50%.

Регресори:

Код програми, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався договір до цього, Останній сплачений внесок, Страхова Сума, Чи є авансові платежі, Чи додатково індексований договір.

| | Train Data | Test Data |
|----------|------------|-----------|
| FAR | 1.43% | 14.4% |
| FRR | 0.03% | 0.39% |
| Accuracy | 99.86% | 98.52% |



Покращення моделі не видно, але це зумовлено тим, що вона і так є доволі добре побудованою.

Подивимось ще на оцінку точності:

Середня точність – 98.5%

Дисперсія точності – 0.001

Висновок

Показники точності дуже високі на рівні 98-99%.

Водночас показники FAR/FRR доволі низькі.

І хоча спочатку були припущення розбиття на різні програми, видно, що це не потрібно. Це все дає нам можливість зробити висновок, що модель побудована точно. І тепер за допомогою функції *predict_proba* модель дає можливість оцінювати ймовірність розірвання того чи іншого договору.

Побудова моделі прогнозування кварталу розриву за допомогою методу класифікації Random Forest Classifier

Модель прогнозу розриву показала себе дуже добре, але окрім знання чи розірветься договір, нас цікавить питання *Коли?*. Тому тепер спробуємо реалізувати саме прогнозування кварталу розриву договору. Регресори залишаємо ті самі, що і для минулої моделі.

Зауважимо, що дана модель є більш складною, адже прогноз вже йде не на 0 та 1, а на доволі розкинуті цілі значення від 1 до 64 (найбільше представлене значення).

На жаль, для аналізу тут не вийде використовувати показники *FAR/FRR*, адже вони себе краще показують на бінарних даних.

Тому в даному випадку будемо користуватися знайомою функцією *score*, яка дає змогу оцінити коректність прогнозу, порівнюючи прогноз та реальні дані.

Перша побудова моделі

Відсоток даних, на яких тренується: 50%.

Регресори:

Код програми, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався договір до цього, Останній сплачений внесок, Страхова Сума, Чи є авансові платежі.

Параметри моделі: базові.

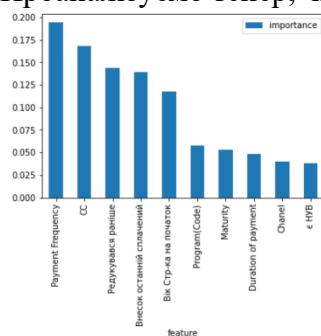
| | Train Data | Test Data |
|-------|------------|-----------|
| Score | 99.5% | 94.46% |

K-Fold Cross Validation допомагає нам переглянути наскільки нам повезло, чи такі параметри майже завжди притаманні даним моделі.

Середня точність – 94.42%

Дисперсія точності – 0.0008

Проаналізуємо тепер, чи потрібні нам усі регресори:



Бачимо, що важливість усіх регресорів не нульова. Останні 5 мають важливість менше 5%, але більш менш рівномірною між ними, тому залишимо їх для нашої моделі.

Висновок

Отже, дана модель є доволі доброю з високим рівнем точності. Усі регресори використовуються при побудові моделі, прибирати чи змінювати немає сенсу. Побудова моделі з іншими параметрами також не дасть великого покращення моделі. Тому приймаємо дану модель для прогнозування кварталу розриву договору.

Порівняння точності прогнозу кварталу розриву з іншими методами класифікації

Цікаво перевірити чи має наша побудована модель за допомогою Random Forest Classifier найвищу точність. Тому проведемо порівняння оцінок точності нашої моделі та таких методів класифікації:

Logistic Regression, K-Neighbors Classifier, Gaussian Naive Bayes.

| | Score for Train Data | Score for Test Data |
|--------------------------|----------------------|---------------------|
| Random Forest Classifier | 99.5% | 94.46% |
| Logistic Regression | 92.29% | 92.29% |
| K-Neighbors Classifier | 93.35% | 91.67% |
| Gaussian Naive Bayes | 73.69% | 73.54% |

Як бачимо, вибір *RFC* був правильний для даної роботи, адже точність для інших методів набагато нижча.

Висновок

Однією з головних проблем у довгостроковому страхуванні є редукування договорів чи повноцінне їх розірвання. Маючи розуміння які договори знаходяться під ризиком розриву, компанії матимуть змогу запровадити превентивні заходи спрямовані на зниження потенціального негативного ефекту. Тому метою цієї роботи було саме прогнозування ймовірності розірвання договорів та конкретного кварталу розриву.

Перед побудовою моделі прогнозування, головною задачею є підготовка даних. Вдалось зібрати вибірку достатнього великого обсягу з усіма договорами, що діяли з 2005 року. Провівши ретельну підготовку з попереднім аналізом, було побудовані добре уявлення о структурі даних та розуміння портфелю. Це допомогло виявити основні регресори, які впливають на розриви договорів. Також була проведена підготовка даних для моделювання. Під час якої дані було замінено на числові та розбито на тренувальну та тестувальну вибірки.

Таким чином за допомогою методу класифікації Random Forest Classifier було побудовано прогноз ймовірності розірвання договору страхування життя та коли він розірветься. Через великий об'єм якісних даних модель одразу показала високу точність (Середня точність – 98.5%) прогнозування та низькі рівні False Acceptance Rate (1.4% для тренувальної вибірки та 14.4% для тестувальної) та False Rejection Rate (0.03% для тренувальної вибірки та 0.39% для тестувальної).

При додатковому аналізі моделі прогнозу розривів було прибрано регресори з малою важливістю і виявлено основні регресори, що мають вплив на розриви договорів. Такими виявилися: Програма, Канал, Термін дії договору, Термін сплати платежів, Частота сплати внесків, Вік страхувальника на початок дії договору, Чи редукувався

договір до цього, Останній сплачений внесок, Страхова Сума, Наявність авансових платежів, Наявність індексації у договорі.

Модель прогнозування кварталу розриву також показала себе дуже точною (Середня точність – 94.42%). У даному випадку використовувалися усі регресори, що й для минулої моделі. Також було проведено порівняння точності даної моделі, побудованої за допомогою Random Forest Classifier, та середніми точностями щоб дали інші моделі такі як: Logistic Regression, K-Neighbors Classifier, Gaussian Naive Bayes. Де Random Forest Classifier мала найвищу точність (94.42% проти 73.54 – 92.29%).

Незважаючи на високі показники ефективності моделі на вибірці треба проаналізувати такі дані у реальному житті, як мінімум побачити наскільки імплементація прогноз збільшить прибуток компанії.

Але важливо зазначити, що найбільшим вразливим місцем цього прогнозу є викиди під час криз та війн, а також зміна структури взаємних ефектів між змінними. Сьогоднішній досвід компаній показує, що у такі часи багато клієнтів розривають договори не через малі проблеми, а через абсолютну недостачу грошей. Тому дана модель не буде давати велику точність при прогнозі розривів у період війни в Україні.

Тож дослідження таких структурних зламів є не тільки цікавою темою для подальших досліджень і побудов моделей, а й важливим чинником зменшення загального ризику портфелю полісів компанії, що в даних умовах може бути гранню між виживанням та банкрутством.

Використана література

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Second Edition, 2001.
2. Vrushali Y Kulkarni, Pradeep K Sinha, Effective Learning and Classification using Random Forest Algorithm, International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May 2014.
3. Erwan Scornet, Random forests and kernel methods, Sorbonne Universites, UPMC Univ Paris 06, F-75005, Paris, France
4. Segal, M., Machine learning benchmarks and random forest regression, Technical report, eScholarship Repository, University of California, 2004.
5. Leo Breiman. "Random Forests." Machine Learning, 2001.

Додатки

Додаток 1

Регресори та їх можливі значення.

| Назва регресора | Можливі значення | Коментарі |
|--|---|--|
| Номер програми | Числове значення (усього 66) | |
| Канал | Змінено на числове значення від 0 до 12 | |
| Рік початку дії договору | 2005:2022 | Буде використовуватися тільки при поділі на програму |
| Тривалість платежів | Числове | |
| Тривалість дії договору | Числове | |
| Частота сплати внесків | Одноразово/щорічно/щоквартально/раз у піврок/щомісячно | |
| Вік стр-ка на початок дії договору | Числове | |
| Чи редукувався договір до | Так/Ні – змінено на 1/0 | |
| Стать страхувальника | M/F – змінено на 1/0 | |
| Чи наявні надбавки по договору | Так/Ні – змінено на 1/0 | |
| Розмір внеску останнього сплаченого | Числове | |
| Страхова сума | Числове | |
| Валюта | Змінено на числові (0-7) з причин конфіденційності | |
| Чи у договорі = Страховальник Застрахована Особа | 0/1 | |
| Чи є Авансові платежі | Так/Ні – змінено на 1/0 | Клієнт може сплатити по договору частково наперед та отримувати таким чином додаткові бонуси по платежам |
| Спосіб оплати | Каса/Сайт – змінено на 1/0 | |
| Чи додатково індексувався договір до цього | Так/Ні – змінено на 1/0 | По договорам клієнтам раз в рік пропонують збільшити внески по договору |
| Чи є у даного страхувальника дитячий договір | Так/Ні – змінено на 1/0 | Таким чином бачимо: 1) у людини є дитина; 2) дитина застрахована. Буде аналізуватися тільки при розбитті на програми і для програм для дорослих. |
| Чи були актуарні зміни | 0 – не було, 1 – були зі збільшенням внеску, -1 – були зі зменшенням внеску | |

Додаток 2

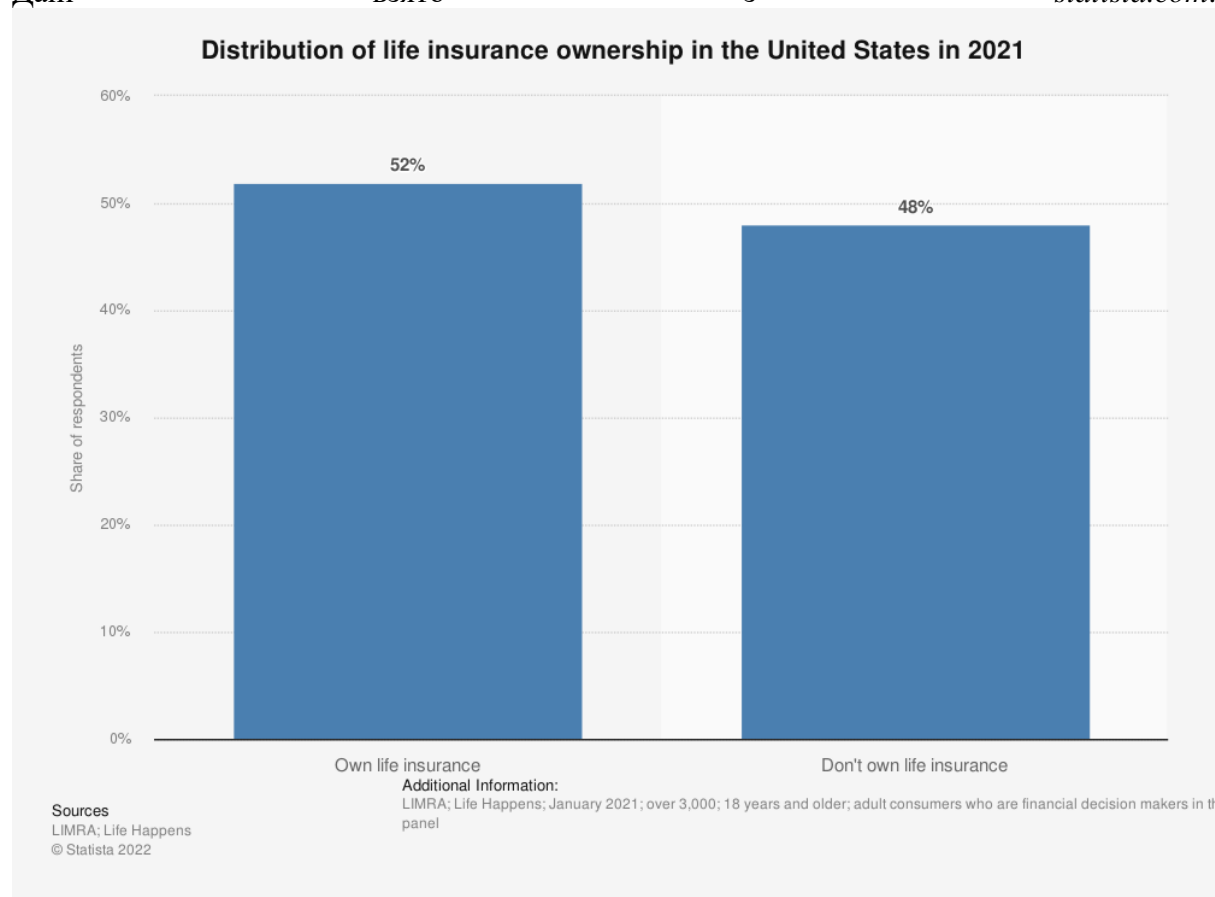
Розподіл населення, що мають поліси страхування життя у Сполучених Штатах Америки станом на 2021 рік.

Дані

взято

3

statista.com.



Додаток 3

Код програми

<https://github.com/ElizabethSkvortsova/Diploma/>