Szkoła Główna Handlowa w Warszawie

PODSTAWOWE I ZAAWANSOWANE PROGRAMOWANIE ORAZ STATYSTYKA W SAS

Semestr zimowy 2018/2019

MGR KAMIL STANISZEWSKI

DR KAROL PRZANOWSKI

ANALIZA VINTAGE

Marcin Czarnecki; 67282
Michał Józef Hajdan; 68501
Eryk Marek Mazuś; 68835
Marcin Mandziej; 68373
Yulia Mecherzhak; 72099
Justyna Żak; 64536

SPIS TREŚCI

1. Abstrakt

Raport przygotowany przez zespół projektowy jest formą podsumowania pracy podczas realizacji projektu analizy ryzyka portfela kredytów detalicznych składającego się z kredytów ratalnych i kredytów gotówkowych.

Celem projektu była **ocena stabilności ryzyka portfela kredytowego w czasie**, znalezienie jego maksymalnej asymptotycznej wartości w badanym okresie i przeprowadzenie prognozy na temat na okres następnego roku po ostatnich obserwacjach testowych - oraz **zaproponowanie własnych rekomendacji** nt. strategii kredytowej zleceniodawcy.

Do zbadania stabilności ryzyka portfela kredytowego przeprowadzono analizę udostępnionych danych w programie SAS przy użyciu **statystyki vintage**, która została opisana w rozdziale drugim. W kolejnym rozdziale opisane zostały zbiory danych użyte w prowadzonej analizie i przybliżona została charakterystyka badanych zmiennych, które podczas analizy okazały się najbardziej istotne.

Następnie wszystkie zmienne zostały skategoryzowane za pomocą **algorytmu tree,** który swoje działanie opiera na metodzie drzewa decyzyjnego. Zastosowanie i przebieg tej metody opisano w rozdziale czwartym.

Pierwszym krokiem w tworzeniu modelu służącego do oceny ryzyka portfela kredytowego w czasie, po wcześniejszym oczyszczeniu i przygotowaniu zbiorów do analizy danych, było znalezienie zmiennych, których zależność ze zmienną objaśnianą jest największa. Do wyznaczenia zmiennych najbardziej skorelowanych ze zmienną objaśnianą wykorzystaliśmy analizę wartości **współczynnika V-Cramera**.

Zmienne wyselekcjonowane w poprzednim etapie zostały użyte do stworzenia **modelu regresji logistycznej**, którego celem było obliczenie prawdopodobieństwa i oczekiwanego defaultu poszczególnych segmentów portfeli kredytów gotówkowych i ratalnych. Proces wykonania modelu i zwrócone wyniki zostały szerzej opisane w rozdziale siódmym.

Końcowym krokiem było obliczenie **prognozy vintage_12**, która została wykonana przy użyciu stworzonego wcześniej modelu, co opisano szerzej w rozdziale 8.

2. Analiza vintage

Analiza vintage to metoda używana przez banki oraz podmioty udzielające pożyczek. Jej głównym zastosowaniem jest **monitorowanie stanu portfela kredytów** podczas kolejnych miesięcy spłat rat kredytowych przez kredytobiorców. Metoda może dodatkowo zostać zastosowana do przygotowania raportów dla poszczególnych kombinacji parametrów analizowanego portfela kredytowego i dostępnych obserwacji (np. w zależności od celu, statusu kredytów, liczby dni przeterminowania czy wieku kredytobiorców) oraz do wygenerowania statystyk i wykresów pozwalających na łatwiejsze monitorowanie ryzyka portfela kredytowego w czasie a także interpretację i przygotowanie się do zachodzących zmian. Istnieją dwa rodzaje statystyki vintage: **ilościowa i kwotowa.**

Vintage ilościowy należy interpretować jako wartość procentową, uzyskaną poprzez porównanie liczby kredytów w badanym okresie, dla których określona liczbę ustalonych przedziałów czasowych po uruchomieniu kredytu liczba opóźnionych rat była większa lub równa od tej ustalonej do zbadania. Uzyskana wartość jest następnie porównywana do całkowitej liczby kredytów w tym samym badanym okresie.

Vintage kwotowy należy interpretować jako wartość uzyskaną poprzez porównanie łącznej wartości niespłaconych kredytów do całkowitej łącznej wartości kredytów w badanym okresie i o tych samych pozostałych parametrach.

Co za tym idzie, statystyka vintage może być obliczana dla różnej liczby opóźnionych rat, lub długości okresu od uruchomienia kredytu. W niniejszej analizie zbadano statystyki vintage ilościowe i kwotowe dla **1, 2, i 3 opóźnionych rat w okresach 3, 6, 9 i 12 miesięcy** od początku okresu kredytowania. Pod uwagę wzięliśmy obserwacje z udostępnionych zbiorów danych dla klientów, którzy zaciągnęli **kredyty gotówkowe i ratalne w okresie od stycznia 2006 roku do grudnia 2008 roku.**

3. Zbiór danych - opis

W poniższym rozdziale zostaną omówione zbiory danych na których oparty jest poniższy raport. W projekcie wykorzystane są dwa zbiory danych - Production oraz Transactions. Pierwszy zostanie omówiony zbiór Production.

Zbiór danych Production zawiera listę zmiennych na temat klienta oraz rachunku, a także dane na temat kolejno wnioskowanych kredytów. Składa się on z **119 081 obserwacji oraz 223 zmiennych**. **Kluczem głównym tabeli jest zmienna aid** oznaczająca identyfikator rachunku.

Zmienne zostały podzielone na 4 grupy: **app_, act_, ags_ i agr_**. W grupie **app_** znajdują się dane z formularza aplikacyjnego klienta, dot. jego statusu majątkowego i sytuacji społecznej (np. osiągane dochody, liczba dzieci, wielkość miasta zamieszkanego przez klienta). Grupa zmiennych **act_** zawiera jego dane kredytowe. Zmienne z grupy **ags_** przedstawiają minimum, maksimum oraz średnią liczoną z ostatnich 3,6,9 lub 12 miesięcy dla maksymalnej liczby dni przed lub po 15 dniu danego miesiąca, w którym nastąpiła spłata raty kredytowej, a także minimum, maksimum oraz średnią liczoną z ostatnich 3,6,9 lub 12 miesięcy dla maksymalnej liczby zaległych spłat. W grupie ags_ wyliczane są statystyki agregujące, nawet gdy występują braki danych.

Ostatnia grupa zmiennych **agr_** zawiera minimum, maksimum oraz średnią liczoną z ostatnich 3,6,9 lub 12 miesięcy dla niebrakujących wartości maksymalnej liczby dni przed lub po 15 dniu miesiąca, w którym nastąpiła spłata raty kredytowej oraz minimum, maksimum oraz średnią liczoną z ostatnich 3,6,9 lub 12 miesięcy dla nie brakujących wartości maksymalnej liczby zaległych spłat.

Przy analizie zmiennych z kolejnych dwóch grup należy mieć ponadto na uwadze, iż termin spłaty kolejnej raty kredytu mija **15 dnia każdego miesiąca**. Objaśnienie przykładowych zmiennych z każdej z grup przedstawiono w Tabeli 1.

Tabela 1. Przykładowe zmienne z bazy danych

Grupa zmiennych	Objaśnienie grupy zmiennych	Nazwa zmiennej	Objaśnienie zmiennej
*app_*	dane z formularza aplikacyjnego	*app_char_marital_status*	status cywilny klienta
*act_*	dane kredytowe klienta	*act_ccss_min_pninst*	minimalna liczba rat kredytu gotówkowego płaconych przez klienta
*ags_*	zmienne behawioralne liczone niezależnie od pojawienia się klienta w bazach banku	*ags6_MeanCMaxC_Due*	średnia kalkulowana z ostatnich 6 miesięcy dla maksymalnej liczby zaległych rat kredytu gotówkowego w danym miesiącu
*agr_*	zmienne behawioralne liczone jeśli klient był wcześniej znany dla banku	*agr9_Max_CMaxC_Days*	maksimum kalkulowane z ostatnich 9 miesięcy dla niebrakujących wartości maksimum liczonych dla dni spłaty kredytu gotówkowego

Zmienne nieprzypisane do żadnej z grup zostały wyjaśnione poniżej:

- *cid* – identyfikator klienta;
- *aid* – identyfikator rachunku;
- *product* – rodzaj kredytu (*css* – kredyt gotówkowy, *ins* – kredyt ratalny);
- *period* – data raportowa.

Tabela danych transakcyjnych zawiera miesięczne informacje o spłatach kolejnych rat kredytowych. Składa się ona z **1 718 727 obserwacji oraz 9 zmiennych**. Klucz główny tabeli jest taki sam jak w przypadku wcześniej omawianej tabeli danych produkcji (*aid* – identyfikator rachunku). Ze względu na niewielką liczbę zmiennych każda z nich została opisana poniżej:

- *cid* – identyfikator klienta;
- *aid* – identyfikator rachunku;
- *due_installments* – zaległe raty kredytowe;
- *fin_period* – data udzielenia kredytu;
- *leftn_installments* – liczba rat kredytowych pozostałych do spłaty;
- *paid_installments* – ilość zapłaconych rat kredytowych;
- *period* – data raportowa;
- *product* – rodzaj kredytu (*css* – gotówkowy, *ins* – ratalny);
- *status* – status kredytu.

W celu lepszego zrozumienia zmiennych objaśniających należy dodać, iż status kredytu przyjmuje 3 następujące wartości:

A – active (kredyt spłacany w terminie);

B – bad (kredyt niespłacany przez 7 mies. – klient staje się niewypłacalny):

C – closed (nastąpiła spłata wszystkich opóźnień).

Zmienna *product* przyjmuje 2 wartości, *css* i *ins*. Kredyt gotówkowy (pożyczka pieniężna) – jest przyznawany w gotówce na dowolny cel. Podstawą do udzielenia kredytu ratalnego jest zdolność kredytowa kredytobiorcy obliczana na podstawie jego zarobków, sytuacji rodzinnej i stałych zobowiązań. Pod pojęciem kredyt gotówkowy rozumie się kredyt związany z kupnem towaru w sklepie, gdzie jednocześnie załatwiane są wszystkie formalności, związane z zakupem na raty. Kredyt gotówkowy też mieści się w pojęciu kredytu ratalnego, bowiem spłata takiego zobowiązania następuje zwykle przez spłatę miesięcznych rat.

4. Kategoryzacja zmiennych

Kategoryzacja polega na tworzeniu kategorii, które w wystarczającym stopniu **różnicowałyby obserwacje** między sobą. Część danych wprowadza redundancję informacji, opisując zbliżone właściwości obiektów - co może przysłonić rzeczywiste współzależności, między zmiennymi objaśniającymi a zmienną objaśnianą w modelu. Kategoryzacja może go przed tym zjawiskiem uchronić.

Poza tym grupowanie zmiennych ułatwia analizę. Kategoryzacji używa się by w sposób bardziej reprezentatywny i pozbawiony wpływu absolutnych wartości przedstawić wybrane zmienne. Dzięki temu można sprawdzić, jak na wybrane zjawisko wpływa fakt znalezienia się w danej grupie.

Drzewo decyzyjne jest jednym z wielu sposobów kategoryzacji zmiennych. W działalności badawczej, a w szczególności w teorii decyzji, drzewa decyzyjne są narzędziami wspomagającymi podejmowanie decyzji. Istnieje wiele parametrów, które mogą wpłynąć na jakość procesu i finalny wynik, a najważniejsze cechy spośród nich to:

- **ustalenie ilości podziałów** – zabieg ten ma uchronić przed budowaniem zmiennych o zbyt dużej liczbie kategorii i zbyt wysokim dopasowaniem modelu do danych uczących. Stąd maks. ilość podziałów w tym modelu ustalono na 3 ;
- **kryterium podziału** - zmienne zostały rozdzielone do 4 różnych grup, zgodnie z ich przedrostkami (app_, act_, ags_ i agr_). Wybór podziału jest zdeterminowany cechami, które ma każda grupa zmiennych (zob: *Tabela 1. Przykładowe zmienne z bazy danych);*
- **głębokość drzewa** - drzewo decyzyjne ma dwa poziomy głębokości, dzięki czemu minimalizuje się ryzyko ,,przeuczenia" modelu i uzyskania zbyt dużej liczby kategorii.

5. Zależność zmiennych

Poniższy rozdział poświęcony jest wybraniu najbardziej skorelowanych zmiennych, które w dalszej części tego raportu użyte są do modelu predykcyjnego. Skorelowanie zmiennych ze zmienną objaśnianą Vin3_12 obliczono za pomocą **współczynnika V-Cramera**. Rozdział składa się z przybliżenia teorii V-Cramera oraz przedstawienia oszacowań współczynnika dla zmiennych w zbiorze Production.

Współczynnik V-Cramera jest stosowany jako miara **trafności dopasowania.** Jest on obliczany za pomocą następującego wzoru:

$$V = \sqrt{\frac{\chi^2}{n(m-1)}}$$

gdzie:

V - współczynnik V-Cramera

χ2 - wynik testu chi kwadrat

n - liczba obserwacji

m = mniejsza z liczb (k i l) określających liczbę wierszy i kolumn

Wartość współczynnika V mieści się w **przedziale <0; 1>**. Im wartość ta jest bliższa 0, tym siła związku pomiędzy badanymi cechami jest mniejsza, a im bliższa 1 - tym siła badanego związku jest większa. Współczynnik V-Cramera uznaje się za istotny statystycznie, jeśli wartość p (*p-value*) wyznaczona na podstawie **statystyki testu chi kwadrat** i rozkładu chi-kwadrat (wyznaczonego dla tabeli) jest równa bądź mniejsza niż poziom istotności α.

Po obliczeniu współczynnika V-Cramera dla każdej ze zmiennych postanowiono wybrać po 5 zmiennych z każdej z grup (**app, agr, act i ags**) co dało ostatecznie 20 zmiennych. Uzyskane wartości przedstawione zostały w Tabeli 3.

Tabela 3. Dwadzieścia najbardziej skorelowanych zmiennych ze zmienną Vin3_12; oszacowania współczynnika V-Cramera

Zmienna	V-Cr	Zmienna	V-Cr	Zmienna	V-Cr	Zmienna	V-Cr
act_CCss_Acp5y	0.21	*agr12_Mean_CMaxA_Days*	0.19	*ags12_Mean_CMaxC_Days*	0.18	*app_char_branch*	0.22
act_call_n_loan	0.21	*agr9_Mean_CMaxA_Days*	0.19	*ags9_Mean_CMaxC_Days*	0.17	*app_loan_amount*	0.18
act_cins_min_seniority	0.21	*agr12_Min_CMaxA_Days*	0.19	*ags12_Min_CMaxC_Days*	0.17	*app_installment*	0.15
act_ccss_n_loans_act	0.20	*agr9_Min_CMaxA_Days*	0.19	*ags12_Mean_CMaxA_Days*	0.17	*app_n_installments*	0.14
act_CAll_Acp5y	0.19	*agr6_Mean_CMaxA_Days*	0.19	*ags9_Min_CMaxC_Days*	0.17	*app_income*	0.08

Z Tabeli 3. wynika, że wartość współczynnika dla żadnej ze zmiennych jest nie większa niż 0.21 co oznacza że zmienne ze zbioru Production są **słabo skorelowane ze zmienną objaśnianą.**

6. Analiza wybranych zmiennych

Po wyliczeniu współczynnika V Cramera wybrano po 5 najlepszych zmiennych z wcześniej opisanych grup zmiennych app, act, ags i agr.

W tym rozdziale wykorzystano dwa współczynniki - **korelacji Pearsona i korelacji rang Spearmana** - celem analizy wybranych zmiennych.

Użycie współczynników korelacji Pearsona i korelacji rang Spearmana ma na celu sprawdzenie, czy pomiędzy dwiema zmiennymi istnieje związek (współzależność). Korelacja Pearsona należy do grupy testów parametrycznych, zaś korelacja Spearmana jest jej nieparametrycznym odpowiednikiem. Wartość tych współczynników mieści się w **przedziale <-1,1>**. Wartości skrajne oznaczają idealną korelację między zmiennymi (przy czym -1 - korelacja odwrotna). Wynik równy 0 oznacza brak korelacji badanych zmiennych.

W poniższej tabeli zostały przedstawione wartości opisanych wyżej współczynników przykładowych zmiennych.

Tabela 4. Wartości współczynników przykładowych zmiennych

	Zmienne			
	app_loan_amount	*act_call_n_loan*	*act_cins_min_seniority*	*act_CCss_Acp5y*
Pearson	**0,1683**	**0,2101**	**-0,2049**	**-0,2113**
Spearman	**0,1745**	**0,2101**	**-0,2069**	**-0,2113**

Na podstawie powyższej tabeli można stwierdzić, że im wyższa wartość kredytu (*app_loan_amount*) oraz im więcej kredytów zaciągnął klient (*act_call_n_loan*), tym większe prawdopodobieństwo defaultu, ponieważ korelacja jest dodatnia dla podanych współczynników korelacji. Dodatkowo występuje ujemna współzależność dla zmiennych *act_cins_min_seniority* (minimalna długość trwania kredytu ratalnego) oraz *act_CCSS_Acp5y* (ilość zaakceptowanych wniosków w ostatnich 5 latach).

7. Regresja logistyczna i jej wyniki

W tym rozdziale zostanie przedstawione modelowanie danych przy wykorzystaniu regresji logistycznej z wykorzystaniem 20 najbardziej skorelowanych zmiennych. W ten sposób można poznać zależności między tymi zmiennymi a kształtowaniem się prawdopodobieństwa *defaultu* w odniesieniu do vintage 3. Skonstruowany model regresji logistycznej zostanie również wykorzystany w rozdziale 8. do predykcji „zachowania się" vintage 3_12 (na 12 miesięcy wstecz) dla roku 2008.

Regresja logistyczna

Do regresji wykorzystaliśmy **model regresji logistyczne**j ze względu na fakt, że zmienna zależna, czyli objaśniana, jest dychotomiczna - co oznacza, że **przyjmuje wartość albo 0 albo 1**. Wartość 1 dla zmiennej vin3 w odniesieniu do naszego modelu oznacza, że wystąpiły **warunki defaultu**, tzn w ciągu określonej liczby miesięcy od początku udzielenia pożyczki nie zostały zapłacone 3 raty. Wartość 0 ma przeciwne znaczenie tzn. nie występują warunki defaultu.

Zmienne niezależne w analizie regresji logistycznej mogą przyjmować charakter nominalny, porządkowy, przedziałowy lub ilorazowy. W przypadku **zmiennych nominalnych oraz porządkowych** następuje ich przekodowanie w liczbę zmiennych zero-jedynkowych taką samą lub o 1 mniejszą niż liczba kategorii w jej definicji (tak jest w przypadku zmiennej *app_char_branch*) W regresji logistycznej, zamiast określać prawdopodobieństwo „klasycznie", za pomocą stosunku liczby sukcesów do liczby wszystkich prób, oblicza się **iloraz szans**, czyli stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki czyli **pi/(1-pi)** (gdzie pi to prawdopodobieństwo sukcesu - otrzymania wartości 1). Logit z kolei to **logarytm ilorazu szans czyli ln(pi/(1-pi))** i jest to zmienna objaśniana.

Jeżeli chodzi o interpretację wyników regresji logistycznej, to wzrost wartości jednej ze zmiennych objaśniających o jednostkę przy założeniu c.p. to iloraz szans, o jaką średnio zmienia się zmienna objaśniana i wynosi **exp** (oszacowana wartość parametru stojącego przy tej zmiennej). W proponowanym modelu przykładowo: exl(0.1644)=1,18 dla zmiennej act_CAll_Acp5y czyli wzrost wartości act_CAll_Acp5y o 1 zwiększa iloraz szans (prawdopodobieństwo, że vin3 osiągnie wartość 1, czyli default) o 18%.

Wyniki modelu regresji logistycznej

Do stworzenia modelu regresji logistycznej wykorzystano **20 najbardziej skorelowanych** (a więc tych z największą wartością współczynnika V-Cramera) zmiennych ze zmienną objaśnianą Vin3. Do wytrenowania modelu posłużono się obserwacjami ze **zbiorów Production i Transaction z okresu do początku roku 2008.** W celu wybrania postaci funkcyjnej modelu posłużono się **metodą krokową wsteczną** (*selection = stepwise*). Powyższa metodyka została zastosowana trzykrotnie: do budowy modelu przewidującego Vin3 dla kredytów gotówkowego, dla kredytu ratalnego oraz obu kredytów naraz.

Szczegółowe wyniki wyestymowane dla modelu regresji dla obu rodzajów kredytów przedstawia poniższa tabela. Zostały w niej przedstawione zmienne, wartości oszacowanych parametrów oraz wnioski dotyczące kierunku i siły zależności.

Tabela 5. Oszacowania parametrów modelu regresji logistycznej

Parametr	DF	Ocena	Błąd standardowy	Chi-kwadrat Walda	Pr. > chi-kw.
Intercept	1	-6.4000	0.1868	1173.8555	<.0001
act_CCss_Acp5y	1	-0.4218	0.00933	2045.7361	<.0001
act_call_n_loan	1	-0.0239	0.00882	7.3487	0.0067
act_cins_min_seniori	1	0.00225	0.000161	197.0534	<.0001
act_ccss_n_loans_act	1	0.2305	0.00983	549.8002	<.0001
act_CAll_Acp5y	1	0.1644	0.00891	340.3674	<.0001
agr9_Mean_CMaxA_Days	1	0.1902	0.0165	133.2498	<.0001
agr12_Min_CMaxA_Days	1	-0.0187	0.00590	10.0785	0.0015
*agr9_Min_CMaxA_*	1	-0.0468	0.00631	54.9597	<.0001

Days					
ags12_Mean_CMaxC_Day	1	0.2203	0.0209	111.0700	<.0001
ags9_Mean_CMaxC_Days	1	-0.1328	0.0194	47.0133	<.0001
ags12_Min_CMaxC_Days	1	0.0217	0.00476	20.8666	<.0001
app_char_branch_Computers	1	-0.8615	0.0689	156.2683	<.0001
app_char_branch_DiY	1	0.2037	0.0461	19.4829	<.0001
app_char_branch_Empty	1	0.9855	0.0252	1525.7545	<.0001
app_char_branch_Fenitures	1	-0.1938	0.0401	23.3762	<.0001
app_loan_amount	1	-0.00008	0.000021	14.4334	0.0001
app_installment	1	0.00418	0.000445	88.1383	<.0001
app_n_installments	1	0.0267	0.00560	22.6934	<.0001
app_income	1	-0.00030	8.387E-6	1237.9797	<.0001

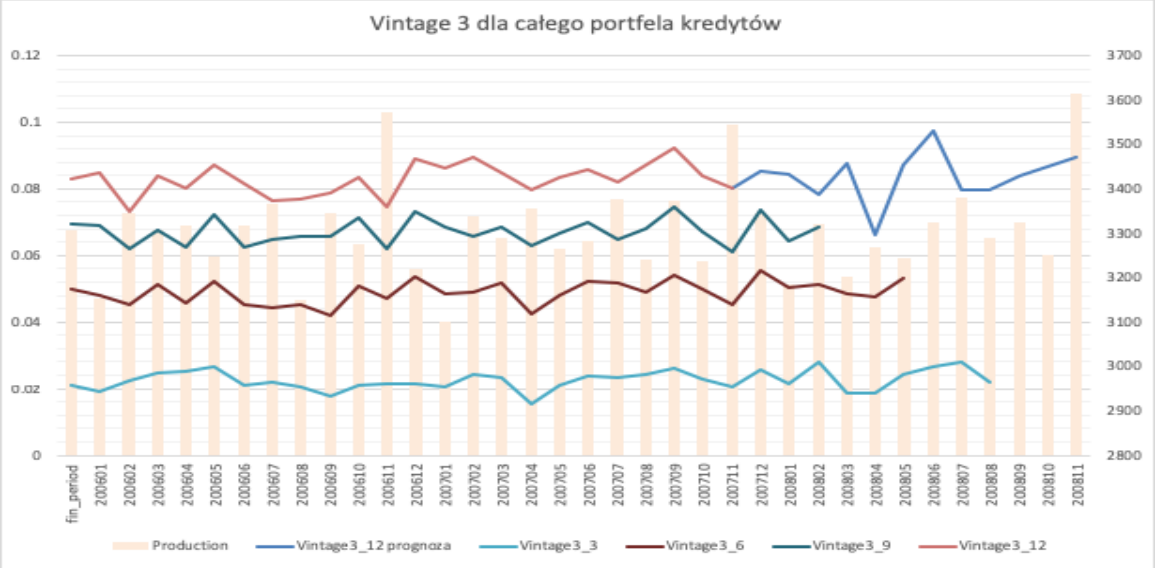
8. Prognoza vintage

Do przeprowadzenia prognozy wykorzystany został model regresji logistycznej, który omówiono szerzej w rozdziale 7. Model został wykorzystany do prognozowania kształtowania się przyszłej wartości vintage 3 w okresie całego roku 2008. Do wykonania prognozy wykorzystano obserwacje historyczne, dla których długość trwania okresu kredytowania wynosiła **co najmniej 12 miesięcy.** Prognoza Vintage została wykonana oddzielnie dla dwóch segmentów portfela kredytowego - kredytów gotówkowych i ratalnych oraz dla całego portfela kredytowego. Wykorzystana metoda miała na celu sprawdzenie, czy istnieją istotne różnice pomiędzy kredytami gotówkowymi a kredytami ratalnymi odnośnie trendów kształtowania się ryzyka w czasie oraz jak poszczególne segmenty wpływają na wyniki dla całości portfela kredytowego.

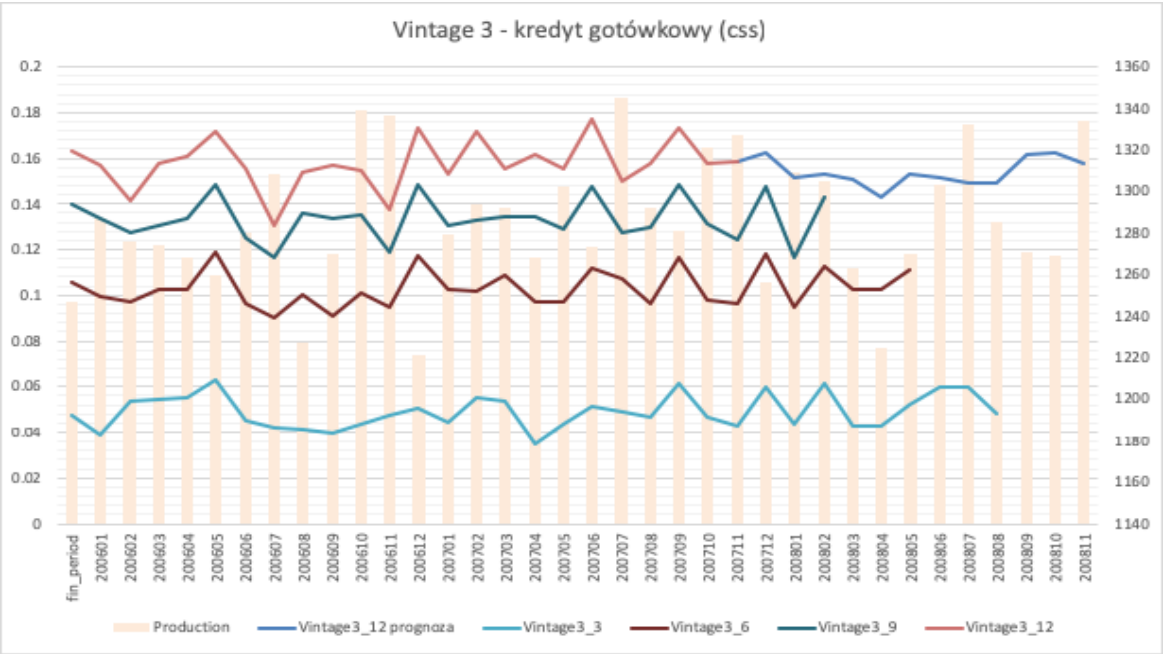
W modelu regresji logistycznej do prognozowania trzeba zaklasyfikować daną obserwację do kategorii 1 lub 0 na podstawie progu odcięcia tzn. vin3_12=1, jeśli $\pi>x$, gdzie x to próg odcięcia, a π - oszacowana wartość prawdopodobieństwa sukcesu π i-tej obserwacji. W modelu zastosowano następujące **progi odcięcia**:

- VIN i CSS - 0.22
- CSS - 0.2
- VIN - 0.35

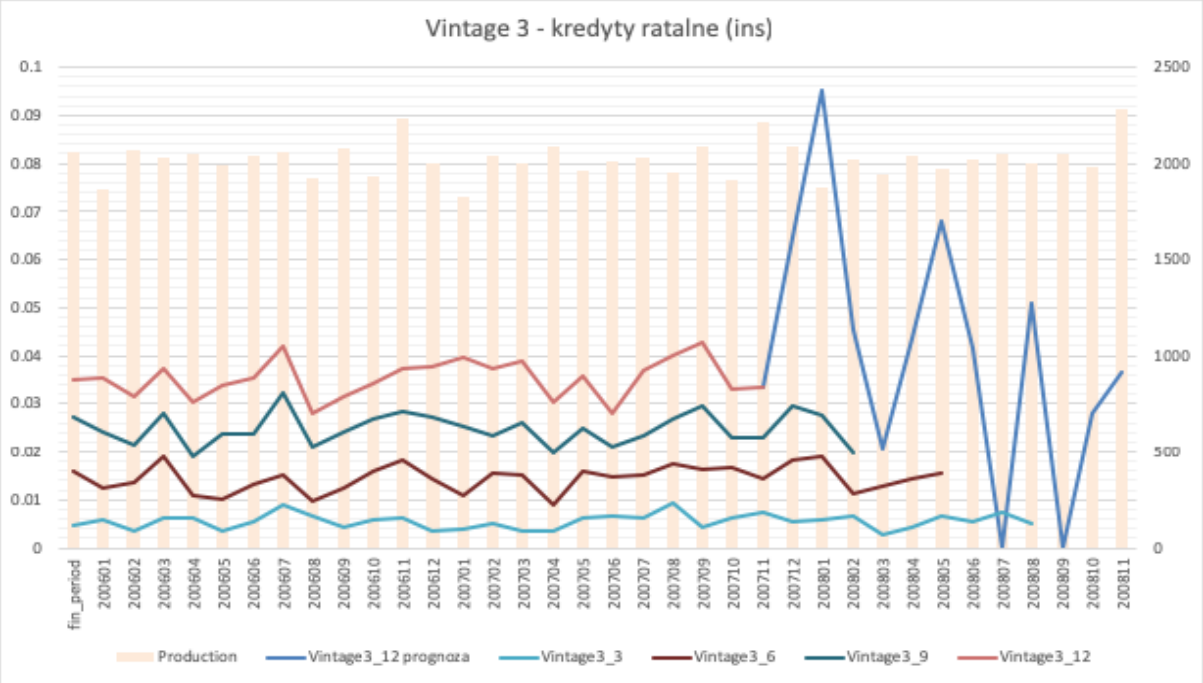
Wyniki prognozy przedstawione zostały na poniższych 3 wykresach. Wynika z nich, że **portfel kredytowy jest narażony na wzrost ryzyka w czasie**. Poziom ryzyka dla całego portfela kredytowego w szczytowym momencie roku 2008 może zwiększyć się w porównaniu do lat 2006 i 2007. Według prognozy **wzrośnie również amplituda wahań**. Ze względu na znaczne różnice w liczbie i wolumenie niespłacanych kredytów w poszczególnych miesiącach, większe niż miało to miejsce w poprzednich latach, **ryzyko całego portfela powinno istotnie wzrosnąć.**



Według prognozy w 2008 roku poziom ryzyka segmentu kredytów gotówkowych powinien utrzymywać się na **poziomie podobnym** do tego, który miał miejsce w latach 2006 - 2007. Nieznacznie zmniejszyć powinna się jednak amplituda wahań, co może świadczyć o tym, że **ryzyko dla tego segmentu portfela ustabilizuje się** i zarządzanie nim będzie prostsze.



W przypadku segmentu kredytów ratalnych w roku 2008 ryzyko powinno cechować się **znacznym wzrostem amplitudy wahań w porównaniu do lat 2006 - 2007.** Z tego powodu należy spodziewać się **wzrostu poziomu ryzyka dla tej części portfela kredytowego.**



9. Wnioski końcowe i rekomendacje

Podsumowując, po przeprowadzeniu analizy można stwierdzić, że segment kredytów gotówkowych jest bardziej ryzykowny, niż segment kredytów ratalnych. Wniosek może wydawać się intuicyjny, ponieważ wymagania, które należy spełnić do otrzymania kredytu gotówkowego są zazwyczaj mniej restrykcyjne w porównaniu do kredytów ratalnych. Dodatkowo, kredyty gotówkowe są często gorzej zabezpieczone, niż kredyty ratalne i nie są udzielane na konkretny cel, przez co dłużnik odczuwa mniejszą motywację do spłaty takiej pożyczki. Dlatego podmioty udzielające kredytów gotówkowych są przygotowane na większy poziom ryzyka i rekompensują wyższą oczekiwaną wartość defaultu odpowiednio wyższą marżą na tego rodzaju produktach.

Całościowo portfel kredytowy nie może zostać uznany za stabilny, oba segmenty cechują się zbyt wysokim poziomem ryzyka w czasie, choć nieco z innych powodów.

Portfel kredytów ratalnych cechuje się częstymi zmianami poziomów niespłacanych kredytów. Zmiany te są stosunkowo nieregularne, przez co trudne do przewidzenia. Ich amplituda, mimo że znacznie niższa niż w przypadku kredytów ratalnych w opinii zespołu projektowego jest w dalszym ciągu zbyt wysoka, by ryzyko tego segmentu w czasie można uznać za dopuszczalne.

Segment kredytów gotówkowych cechuje się mniejszą częstotliwością wahań poziomu niespłacanych pożyczek. Amplituda tych wahań jest jednak istotnie wyższa. Ten czynnik według zespołu projektowego stanowi główny powód, dla którego ta część portfela kredytowego nie może zostać uznana za bezpieczną.

Różnice w poziomie i fluktuacjach ryzyka dla tych dwóch grup produktów potwierdzają **zasadność prowadzenia oddzielnych analiz dla tych grup**. Zespół projektowy po zastosowaniu niżej przedstawionych rekomendacji zaleca przeprowadzenie dalszych analiz dla całości portfela, jak i obu jego segmentów w dłuższym horyzoncie czasowym, po zwiększeniu ilości obserwacji w najbliższym okresie. Takie rozwiązanie pozwoli na ocenę skuteczności zastosowanych rozwiązań i wyciągnięcie kolejnych wniosków, które pozwolą na dostosowanie polityki klienta do realizacji założonego celu, czyli ustabilizowania i ograniczenia ryzyka obu części portfela kredytowego.

Rekomendacje:

1) czasowe **przekształcenie polityki kredytowej na bardziej konserwatywną** do czasu ustabilizowania się poziomu ryzyka całego portfela kredytowego na zadowalającym poziomie, w szczególności:

a) **zwiększenie poziomu rezerw**, który będzie służył jako bufor bezpieczeństwa w razie zrealizowania się scenariusza wzrostu poziomu ryzyka;

b) **ograniczenie akcji kredytowej dla kredytów gotówkowych** i z**większenie procentowej wartości wolumenu i liczby kredytów ratalnych** w łącznym portfelu. Jednakże, po przeanalizowaniu wyników prognozy nie zaleca się wykonania takiego zabiegu poprzez zwiększanie akcji kredytowej dla portfela kredytów ratalnych. Zamiast tego sugeruje się stopniowe ograniczanie liczby i wolumenu kredytów ratalnych.

2) **zwiększenie wymagań** koniecznych do otrzymania pożyczki dla potencjalnych nowych klientów, w szczególności dla kredytów gotówkowych w celu obniżenia poziomu ryzyka niewypłacalności tego segmentu, co przełoży się na zmniejszenie ryzyka w czasie dla całego portfela kredytowego,

3) **zaoferowanie klientom, którzy spłacili już swoje zobowiązania w wyznaczonym terminie, lub wcześniej bardziej preferencyjnych warunków** kredytowania podczas zaciągania następnych kredytów. Takie rozwiązanie powinno przynieść korzyści w postaci zwiększenia retencji wiarygodnych klientów, którzy cechują się niskim poziomem ryzyka, co przełoży się na obniżenie poziomu ryzyka całego portfela kredytowego,

4) **dostosowanie procesu i narzędzi służących do obliczania zdolności kredytowej** potencjalnych kredytobiorców poprzez zwrócenie większej uwagi na zmienne najbardziej skorelowane ze zmienną objaśnianą przedstawione rozdziale 5. Zaleca się nadanie większej wagi testowym scoringowym, które biorą pod uwagę następujące kryteria:

a) źródło, z którego kredytobiorca dowiedział się o pożyczce - statystycznie kredytobiorcy, którzy dowiedzieli się o możliwości otrzymania pożyczki ze źródeł takich jak reklamy radiowe, czy telewizyjne są bardziej ryzykownymi klientami niż osoby, które prowadziły poszukiwania samodzielnie, lub dowiedziały się o takiej możliwości bezpośrednio w oddziale;

b) liczba zaakceptowanych wniosków o kredyt w ciągu ostatnich 5 lat (w szczególności kredyt gotówkowy) - im więcej zaakceptowanych wniosków o kredyty w swojej historii posiada potencjalny klient, tym większe prawdopodobieństwo, że będzie spłacał swoje zobowiązania w terminie;

c) łączna liczba zobowiązań klienta - im większa liczba wszystkich zobowiązań klienta, tym większe prawdopodobieństwo, że klient nie będzie spłacał swoich zobowiązań w terminie;

d) wysokość sumy kredytowej - im wyższa wysokość sumy kredytowej, tym większe prawdopodobieństwo, że potencjalny klient nie będzie spłacał swoich zobowiązań w terminie.

10. Wykorzystane kody

```
%let dir_projekt=C:\Users\Eryk\Desktop\dane\;
%let dir=C:\Users\Eryk\Desktop\dane\tree\;

libname wej "&dir_projekt" compress=yes;
libname wyj "&dir" compress=yes;

/*makro liczące zmienne vintage (zero-jedynkowe)
	zmienna vitage zależna jest od:
			m_prod - miesiąc w którym wszystkie kredyty uruchomiono
			m - liczba miesięcy po uruchomieniu
			due - min. liczba opóźnionych rat ustalana przez nas na poziomie 1,2,3
	ostatecznie otrzymujemy 3 zmienne vin: vin1, vin2, vin3*/

%macro DataPreparation();

data wyj.DaneDoVintage;
	set wej.Transactions;
	seniority=intck('month',input(fin_period,yymmn6.),input(period,yymmn6.));
	/*intck - zwraca liczbę interwałów pomiędzy danymi okresami czasowymi*/
	/*liczenie ile miesięcy jest pmiędzy period a fin period*/
	/* fin_period - data udzielenia kredytu , period - data raportowa*/
	vin3=(due_installments>=3);
	vin2=(due_installments>=2);
	vin1=(due_installments>=1); *klient nie spłacił conajmniej jednej raty;
	output;

	if status in ('B','C') and period<='200812' then do;
	/*B - bankrupcy, C - closed*, A - available*/
		n_steps=intck('month',input(period,yymmn6.),input('200812',yymmn6.));
		do i=1 to n_steps;
			period=put(intnx('month',input(period,yymmn6.),1,'end'),yymmn6.);
			/*intnx - zwiększa czas o podaną wartość */
			seniority=intck('month',input(fin_period,yymmn6.),input(period,yymmn6.));
			output;
		end;
	end;
	keep vin3 vin2 vin1 seniority aid fin_period period;
run;

%do i=1 %to 3;
proc means data=wyj.DaneDoVintage noprint nway;
	class fin_period seniority;
	var vin&i;
```

```
	output out=wyj.vintagr_vin&i(drop=_freq_ _type_) n()=production mean()=vin&i;
	format vin&i percentn10.;
run;

proc means data=wyj.DaneDoVintage noprint nway;
	class fin_period;
	var vin&i;
	output out=wyj.production(drop=_freq_ _type_) n()=production;
	where seniority=0;
run;

proc transpose data=wyj.vintagr_vin&i out=wyj.vintage&i prefix=months_after_;
	by fin_period;
	var vin&i;
	id seniority;
run;

data wyj.vintage&i;
	set wyj.vintage&i;
	drop _NAME_;
run;

proc sql noprint;
	create table wyj.Seniority_vin&i as
	select aid,seniority, vin&i
	from wyj.DaneDoVintage
	where seniority in (3,6,9,12)
	order by aid, seniority;
run;

proc transpose data=wyj.Seniority_vin&i out=wyj.Seniority_trans_vin&i prefix=vin&i._;
	by aid;
	var vin&i;
	id seniority;
run;

proc sql noprint;
	create table wyj.production_vin&i as
	select a.*, b.*
	from wej.production a
	left join wyj.Seniority_trans_vin&i b
	on a.aid=b.aid;
quit;

/*eksport vintage do xlsx*/
%let arkusz = vintage&i;
proc export data = wyj.vintage&i dbms=xlsx outfile="C:\Users\Eryk\Desktop\SAS Projekt\" replace;
sheet=&arkusz;
run;

%end;
```

```
%mend;

%DataPreparation();

/*Kategoryzacja zmiennych
Usunięcie/przekodowanie zmiennych tekstowych:
app_char_branch
app_char_gender
app_char_job_code
app_char_marital_status
app_char_city
app_char_home_status
app_char_cars
*/
data wyj.vin;
set wyj.Production_vin3;
run;


%let zb=wyj.vin;
%put &zb;
%let tar=vin3_12;
%put &tar;

proc contents data= wej.Production noprint out=varlist (keep = name);
run;

/*Wykluczenie zmiennych, które nie mają być kategoryzowane*/
data varlist;
set varlist;
where name not in ('cid', 'aid', 'product', 'period', 'app_char_branch', 'app_char_cars', 'app_char_city',
'app_char_gender', 'app_char_home_status', 'app_char_job_code', 'app_char_marital_status');
run;

/*Utworzenie listy zmiennych*/                /*slect into - zapis do zmienne_int_ord*/
proc sql noprint;
select name
into :zmienne_int_ord separated by ' '
from varlist;
quit;


/*Kategoryzacja zmiennych uwzględniająca algorytm tree*/

data wyj.vin;
set wyj.Production_vin3;
run;
%let zb=wyj.Production_vin3;
%put &zb;
%let tar=vin3_12;
%put &tar;
```

```
%let il_zm = &sqlobs;

%put ***&il_zm***&zmienne_int_ord;

/*maksymalna liczba podziałów minus 1*/
%let max_il_podz=2;
/*minimalna liczba obs w liściu*/
%let min_percent=3;

/*Odwołanie się do makra tree.sas*/
%include "&dir.tree.sas" / source2;

/*Zdefiniowanie warunków podziału*/

data wyj.warunek;
set wyj.podzialy_int_niem;
low = scan(war,1,'<');
if substr(low,1,1) = 'n' then low = '1=1and';
/*substr - wycina kawałek tekstu*/
if substr(low,1,3) ne '1=1' then low = catt(lowcase(zmienna),'>',low,'and');
/*ne - not exist*/
high = scan(war,2,'=');
if high='' then high ='1=1';
else high = catt(lowcase(zmienna),'=<',high);
if substr(war,1,1) = 'n' then miss = cats('not missing(',lowcase(zmienna),')and');
warunek = catt(miss,low,high);
warunek = transtrn(warunek,'and',' and ');
warunek = transtrn(warunek,'=<',' =< ');
warunek = transtrn(warunek,'>',' > ');
klucz=catt(lowcase(zmienna),'#',grp,'#',warunek);
keep zmienna grp warunek klucz;
run;

proc sql noprint;
select klucz into :warunki separated by '^'
from wyj.warunek;
quit;

%put &warunki;

/*Przyporządkowanie obserwacji do grup zmiennych*/
%macro groups;
data wyj.vin_groups;
set wyj.vin nobs=n_obs;
%do i = 1 %to &sqlobs;
%let warunek = %scan(%scan(&warunki,&i,'^'),3,'#');
%let zmienna = %scan(%scan(&warunki,&i,'^'),1,'#');
%let grupa = %scan(%scan(&warunki,&i,'^'),2,'#');
if &warunek then &zmienna=&grupa;
%end;
```

```
output;
run;
%mend;
%groups;


/*Badanie zaleĹĽnoĹ›ci pomiÄ™dzy zmiennymi a zmiennÄ… vin3_12 poprzez wspĂłĹ‚czynnik V-Cramera*/
%let groups = act#ags#agr#app#; *4 nazwy w jednej makrozmiennej, do wybrania;
%macro vcram;
%do i = 1 %to 4;
%let j = %scan(&groups,&i,'#');

data wyj.vin_groups_&j;
set wyj.vin_groups;
keep vin3_12 &j.:; *wszystkie zmienne z prefixem act;
run;

proc contents data=wyj.vin_groups_&j out=varlist_&j noprint;
run;

proc sql noprint;
select name into:zm_&j separated by '#'
from varlist_&j;
quit;

%let liczba=&sqlobs;

data vcram_&j;
length vcram _cramv_ 8 zmienna $30; *8 znaków każda zmienne;
run;

/* Obliczanie wspĂłĹ‚czynnika chi2 i V-Cramera */

%do k = 1 %to (&liczba-1);
%let zm=%scan(&&zm_&j,&k,'#');
proc freq data = wyj.vin_groups_&j noprint;
tables vin3_12*&zm/chisq; *chisq - dodaje nam wartości do vcramera;
output out = vcram_&zm cramv;
run;

data vcram_&zm;
set vcram_&zm;
zmienna = "&zm";
vcram = abs(_cramv_); *wartośc bezwzględna ze współczynnika vcramera;
run;

data vcram_&j;
set vcram_&j vcram_&zm;
run;
```

```
%end;

/* Sortowanie wedĹ‚ug malejÄ…cej wartoĹ›ci wspoĹ‚czynnika V-Cramera */

proc sort data = vcram_&j out = wyj.vcram_&j;
by descending vcram;
run;

/*WybÄłr 5 najbardziej skorelowanych zmiennych z kaďż˝dej grupy*/
data wyj.vcram_&j;
set wyj.vcram_&j (obs = 5);
run;

%end;
%mend;
%vcram;

*;

%let b = 'ins#css#12#';
%let c = '3#6#9#12#';

%macro tabulate;
%do i = 1 %to 4;
%let typ = %scan(&groups,&i,'#');

proc sql;
select zmienna into :zmienne separated by ' '
from wyj.vcram_&typ;
quit;

data tabulate_&typ;
run;

%do j = 1 %to 3;
/*%do k = 1 %to 3;*/
/*%let kred = %scan(&b,&k,'#');*/
%do l = 1 %to 4;
%let sen = %scan(&c,&l,'#');

*sen - seniority;

proc tabulate data = wyj.Production_vin&j out = tabulate_&typ._vin&j._sen&sen;
class vin&j._&sen.;
var &zmienne;
table (&zmienne), vin&j._&sen.*mean;
run;

data tabulate_&typ._vin&j._sen&sen;
length zmienna $15;
set tabulate_&typ._vin&j._sen&sen;
```

```
zmienna = "vin&j._&sen.";

due = &j;
sen = &sen;
run;

data tabulate_&typ;
set tabulate_&typ tabulate_&typ._vin&j._sen&sen;
run;

/*%end;*/
%end;
%end;

proc export data = tabulate_&typ dbms=xlsx outfile  = "C:\Users\Eryk\Desktop\dane\testy.xlsx" replace;
sheet = tabulate_&typ;
run;

%end;
%mend;
%tabulate;

/*Analiza vintage zbiorczo i w podziale na grupy produktów - export excel*/
%macro vintage_zbiorczy;
%do i=1 %to 3;
%let arkusz = vintage&i;
proc export data=wyj.vintage&i dbms=xlsx outfile="C:\Users\Agata\Desktop\projekt zaliczeniowy
sas\zbiorczy.xlsx" replace;
sheet=&arkusz;
run;
%end;
%mend;
%vintage_zbiorczy;

/* PROGNOZA dla Vintage3_12 */

/* Stworzenie oddzielnego zbioru bazowego do modelu regresji logistycznej */
data wyj.dane_do_regresji;
set wej.Transactions;
seniority=intck('month',input(fin_period,yymmn6.),input(period,yymmn6.)); *liczba okresów między
datami - miesięcy;
vin3=(due_installments>=3);
output;
if status in ('B','C') and period<='200812' then do;
        n_steps=intck('month',input(period,yymmn6.),input('200812',yymmn6.)); *różnica między datą
raportową a datą prognozy;
        do i=1 to n_steps;
        period=put(intnx('month',input(period,yymmn6.),1,'end'),yymmn6.); *zwiększa datę o keden,
'end' że do końca okresu;
        seniority=intck('month',input(fin_period,yymmn6.),input(period,yymmn6.));
        output;
```

```
        end;
end;
keep vin3 seniority aid fin_period;
run;
*;

/* Sortowanie zbioru z danymi do regresji oraz zbioru zawierďż˝cego chorakterystyki klientďż˝w */
proc sort data=wyj.dane_do_regresji;
by aid;
run;

proc sort data=wej.Production(keep= aid act_CCss_Acp5y act_call_n_loan act_cins_min_seniority
act_ccss_n_loans_act act_CAll_Acp5y agr12_Mean_CMaxA_Days agr9_Mean_CMaxA_Days
agr12_Min_CMaxA_Days agr9_Min_CMaxA_Days agr6_Mean_CMaxA_Days ags12_Mean_CMaxC_Days
ags9_Mean_CMaxC_Days ags12_Min_CMaxC_Days ags12_Mean_CMaxA_Days ags9_Min_CMaxC_Days
app_char_branch app_loan_amount app_installment app_n_installments app_income) out=prod_regresja;
by aid;
run;

/* laczenie zbiorďż˝w */
data wyj.vin_regresja;
merge wyj.dane_do_regresji(in=z) prod_regresja; /*lacza dwa zbiory*/
by aid;
if z; /*to oznacza de facto left join*/
run;

/* Dodanie do bazy danych zmiennych okreďż˝lajďż˝cych rodzaj produktu oraz numer miesiďż˝ca */
data wyj.vin_reg;
set wyj.vin_regresja;
nazwa_produktu = substr(aid, 1, 3); /*tworza nazwe produktu wycinajac ja z aid*/
if substr(fin_period, 5,1) = 0 then miesiac = substr(fin_period,6,1); /*wyciagaja tylko miesiac*/
        else miesiac = substr(fin_period,5,2);
run;

/*
data wyj.vin_reg2;
set wyj.vin_reg;
if fin_period<='200712' then vin3 = vin3;
        else vin3 = .;
run;
*/

data wyj.kamil_trenuje_all;
set wyj.vin_reg;
if fin_period<='200712';
run;

data wyj.kamil_testuje_all;
set wyj.vin_reg;
if fin_period>'200712';
run;
```

```
%macro product(name);

data wyj.trenuje_&name;
	set wyj.kamil_trenuje_all;
	where nazwa_produktu = "&name";
run;

data wyj.testuje_&name;
	set wyj.kamil_testuje_all;
	where nazwa_produktu = "&name";
run;

%mend;
%product (css);
%product (ins);

/* Procedura regresji logistyzcnej z metoda selekcji krokowej */
proc logistic data=wyj.trenuje_css;
class app_char_branch;
model vin3 (Event="1") = act_CCss_Acp5y act_call_n_loan act_cins_min_seniority act_ccss_n_loans_act
act_CAll_Acp5y agr12_Mean_CMaxA_Days
agr9_Mean_CMaxA_Days agr12_Min_CMaxA_Days agr9_Min_CMaxA_Days agr6_Mean_CMaxA_Days
ags12_Mean_CMaxC_Days ags9_Mean_CMaxC_Days
ags12_Min_CMaxC_Days ags12_Mean_CMaxA_Days ags9_Min_CMaxC_Days app_char_branch
app_loan_amount app_installment app_n_installments
app_income /selection=stepwise;
score data=wyj.testuje_css out=wyj.prediction_css;
run;

proc logistic data=wyj.trenuje_ins;
class app_char_branch;
model vin3 (Event="1") = act_CCss_Acp5y act_call_n_loan act_cins_min_seniority act_ccss_n_loans_act
act_CAll_Acp5y agr12_Mean_CMaxA_Days
agr9_Mean_CMaxA_Days agr12_Min_CMaxA_Days agr9_Min_CMaxA_Days agr6_Mean_CMaxA_Days
ags12_Mean_CMaxC_Days ags9_Mean_CMaxC_Days
ags12_Min_CMaxC_Days ags12_Mean_CMaxA_Days ags9_Min_CMaxC_Days app_char_branch
app_loan_amount app_installment app_n_installments
app_income /selection=stepwise;
score data=wyj.testuje_ins out=wyj.prediction_ins;
run;

*act_CCss_Acp5y act_call_n_loan act_cins_min_seniority act_ccss_n_loans_act act_CAll_Acp5y
agr12_Mean_CMaxA_Days agr9_Mean_CMaxA_Days agr12_Min_CMaxA_Days agr9_Min_CMaxA_Days
agr6_Mean_CMaxA_Days ags12_Mean_CMaxC_Days ags9_Mean_CMaxC_Days ags12_Min_CMaxC_Days
ags12_Mean_CMaxA_Days ags9_Min_CMaxC_Days app_char_branch app_loan_amount app_installment
app_n_installments app_income
;
/*
cutoff values:
css - 0.155
```

```
ins - 0.051
*/

data wyj.prediction_css_labeled;
set wyj.prediction_css;
if P_1 ^= .;
if P_1 >= 0.2 then LABEL = 1;
else LABEL = 0;
run;

data wyj.prediction_ins_labeled;
set wyj.prediction_ins;
if P_1 ^= .;
if P_1 >= 0.35 then LABEL = 1;
else LABEL = 0;
run;

*predykcja vintage 3 w 2008;
*css;
proc means data=wyj.prediction_css_labeled noprint nway; *noprint - bez raportu; *nway - ;
        class fin_period seniority;
        var LABEL;
        output out=wyj.prediction_css_vintage (drop=_freq_ _type_) n()=production
mean()=vintage3_p_css;
        format LABEL nlpct12.2;
run;

proc transpose data=wyj.prediction_css_vintage out=wyj.prediction_css_vintage2 prefix=months_after_;
        by fin_period;
        var vintage3_p_css;
        id seniority;
run;

*ins;
proc means data=wyj.prediction_ins_labeled noprint nway; *noprint - bez raportu; *nway - ;
        class fin_period seniority;
        var LABEL;
        output out=wyj.prediction_ins_vintage (drop=_freq_ _type_) n()=production
mean()=vintage3_p_ins;
        format LABEL nlpct12.2;
run;

proc transpose data=wyj.prediction_ins_vintage out=wyj.prediction_ins_vintage2 prefix=months_after_;
        by fin_period;
        var vintage3_p_ins;
        id seniority;
run;

*all = ins + css;
*wyj.kamil_trenuje_all;
*wyj.kamil_testuje_all;
```

```
proc logistic data=wyj.kamil_trenuje_all;
class app_char_branch;
model vin3 (Event="1") = act_CCss_Acp5y act_call_n_loan act_cins_min_seniority act_ccss_n_loans_act
act_CAll_Acp5y agr12_Mean_CMaxA_Days
agr9_Mean_CMaxA_Days agr12_Min_CMaxA_Days agr9_Min_CMaxA_Days agr6_Mean_CMaxA_Days
ags12_Mean_CMaxC_Days ags9_Mean_CMaxC_Days
ags12_Min_CMaxC_Days ags12_Mean_CMaxA_Days ags9_Min_CMaxC_Days app_char_branch
app_loan_amount app_installment app_n_installments
app_income /selection=stepwise;
score data=wyj.kamil_testuje_all out=wyj.prediction_all;
run;
*cutoff = 0.22;

data wyj.prediction_all_labeled;
set wyj.prediction_all;
if P_1 ^= .;
if P_1 >= 0.22 then LABEL = 1;
else LABEL = 0;
run;

proc means data=wyj.prediction_all_labeled noprint nway; *noprint - bez raportu; *nway - ;
        class fin_period seniority;
        var LABEL;
        output out=wyj.prediction_all_vintage (drop=_freq_ _type_) n()=production
mean()=vintage3_p_all;
        format LABEL nlpct12.2;
run;

proc transpose data=wyj.prediction_all_vintage out=wyj.prediction_all_vintage2 prefix=months_after_;
        by fin_period;
        var vintage3_p_all;
        id seniority;
run;

proc export
 data=wyj.prediction_all_vintage2
 dbms=xlsx
 outfile="C:\Users\Eryk\Desktop\dane\export logistic\wykresy\all_pred_vintage3.xlsx"
 replace;
 sheet="Predykcja";
run;

* tablea production do wyresów;
proc freq data=wej.Production;
tables product * period /nocum nopercent norow nocol;
run;
```