

Budowa karty scoringowej

Dokument składa się z 7-miu części. Pierwsza to zdefiniowanie dobrego i złego klienta, kolejne dwie dotyczą eksploracji zmiennych – najpierw jakościowych, potem ilościowych. Część czwarta to analiza interakcji między wszystkimi zmiennymi. Kolejna część to budowa modelu logistycznego, a szósta to podsumowanie skuteczności modelu. Ostatnia to budowa karty scoringowej.

Część 1 - Zdefiniowanie dobrego i złego klienta

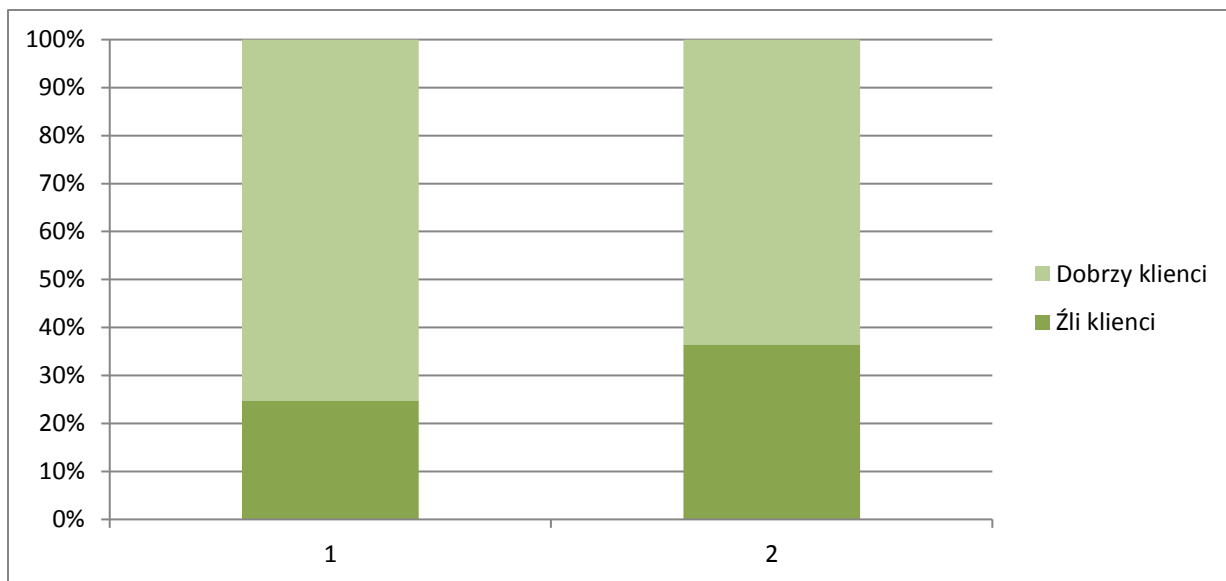
Otrzymane dane składały się z 7122-óch przypadków udzielonych kredytów. Dla kredytów niespłaconych dana jest kolumna Czas - ustalono, że regularny czas spłaty wynosi do 35-ciu dni. Tzn. dla Czasu powyżej 35-ciu kredyt jest uznawany za „zły”. Na początku stworzono więc zmienną binarną *Zaufanie* stosując poniższe reguły:

- pożyczki o stanie „Zakończona” otrzymały *Zaufanie* = 1,
- pożyczki o stanie „Trwająca” i o czasie > 35 otrzymały *Zaufanie* = 0.

Klientów o stanie „Trwająca” i o czasie ≤ 35 nie uwzględniono w analizie, było 16 takich przypadków.

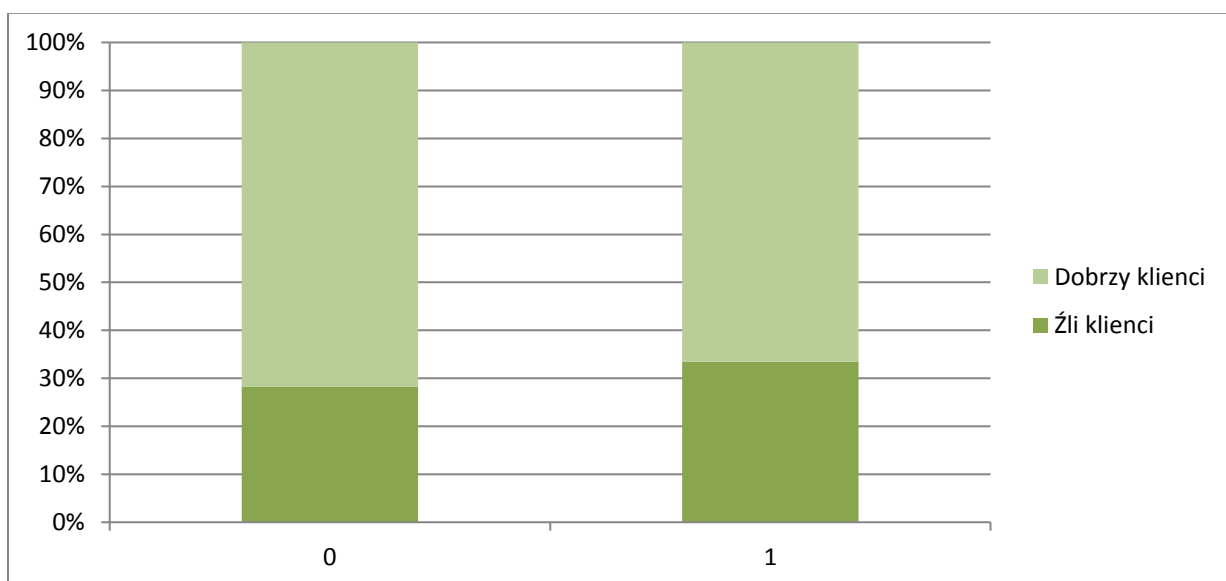
Część 2 – Eksploracja i modelowanie zmiennych jakościowych

Na podstawie nowej zmiennej podzielono dane na dwie zasadnicze grupy. Wartości sześciu zmiennych w podziale na to, jaką część danej odpowiedzi stanowili „dobrzy” i „źli” klienci przedstawione są poniżej. Dzięki temu można zaobserwować zmienność proporcji wartości dla różnych klientów i stworzyć specyfikę złego klienta tak, by w przyszłości przewidywać na podstawie danych, komu nie powinno się udzielać kredytu, ponieważ ma duże ryzyko niewypłacalności. Jednak wnioski z wykresów nie są wystarczająco konkretne, dlatego w dalszej części przedstawiono wyniki testu Chi-kwadrat sprawdzającego, czy te różnice między grupami są istotnie statystycznie. Następnie obliczono współczynniki V Cramera, które pokazują, jak silna jest zależność zmiennych jakościowych od czynnika „dobry/zły” klient.



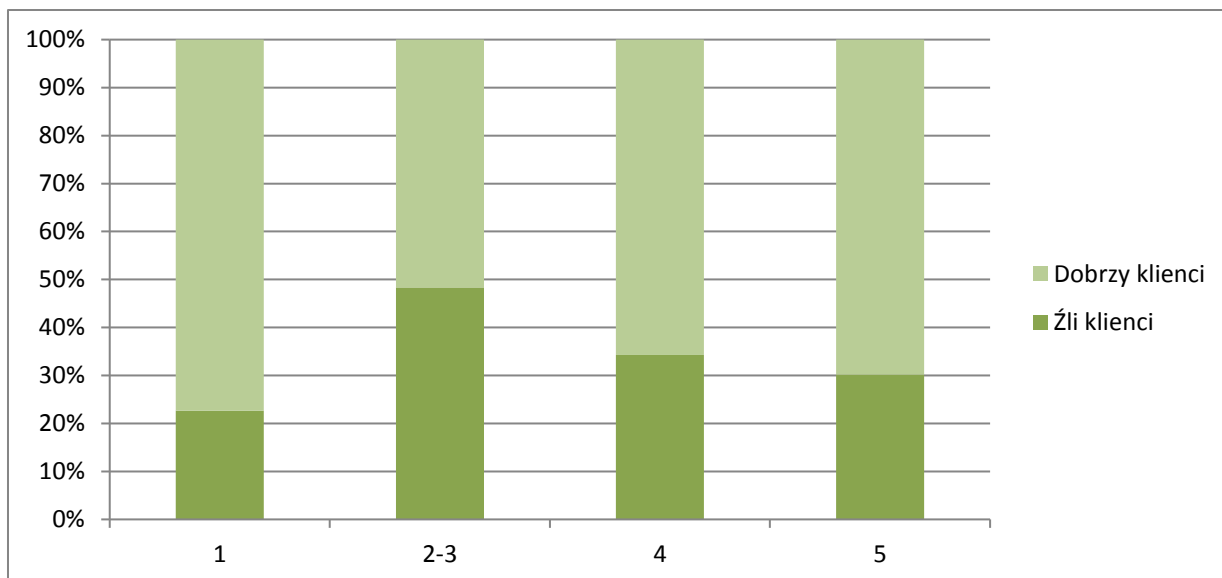
Rysunek 1: Podział pierwszej zmiennej jakościowej

„Źli” klienci stanowili większą część dla wartości równej 2, różnica wyniosła ok. 10%.



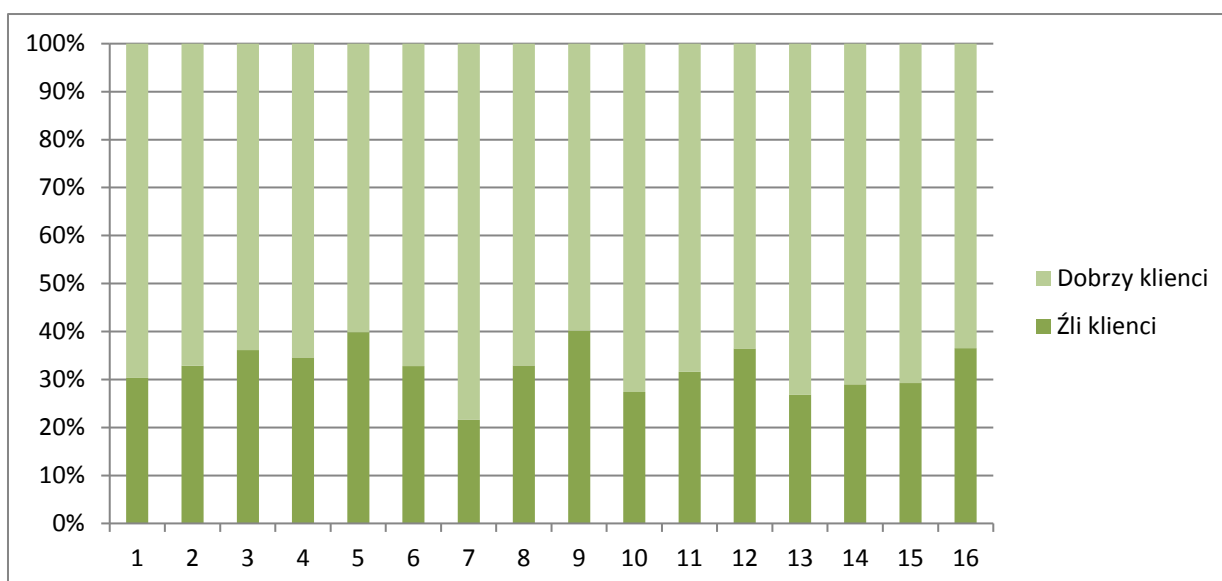
Rysunek 2: Podział drugiej zmiennej jakościowej

Dla tej zmiennej „źli” klienci stanowili większą część dla wartości = 1, jednak różnica jest niewielka – ok. 5%.



Rysunek 3: Podział trzeciej zmiennej jakościowej

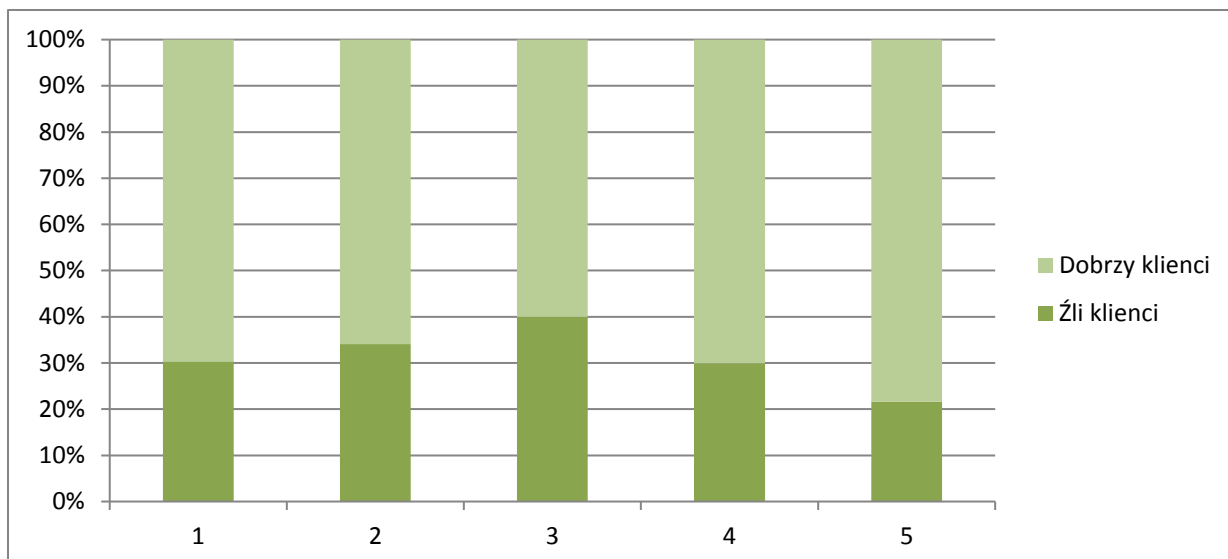
W przypadku tej zmiennej pojawiły się też przypadki o wartości równej 3 – jednak ich liczebności w poszczególnych grupach były mniejsze niż 30, dla poprawności analizy te przypadki zamieniono na wartość 2. Część „złych” klientów jest największa dla wartości = 2-3, natomiast najmniejsza dla wartości = 1.



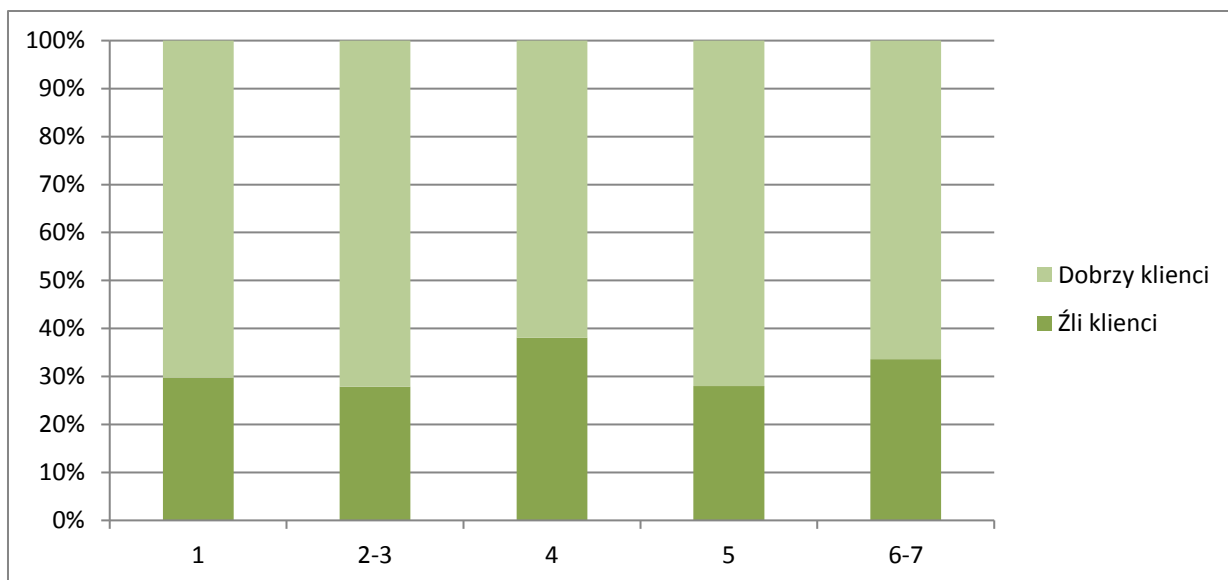
Rysunek 4: Podział czwartej zmiennej jakościowej

Dla tej zmiennej wyróżniły się wartości 5 oraz 9, gdzie część „złych” klientów stanowiła ok. 40%. Natomiast najmniejszą część stanowili dla wartości 7, tylko ok. 22%.

W późniejszej analizie czwarta zmienna jakościowa została podzielona na nowe poziomy:

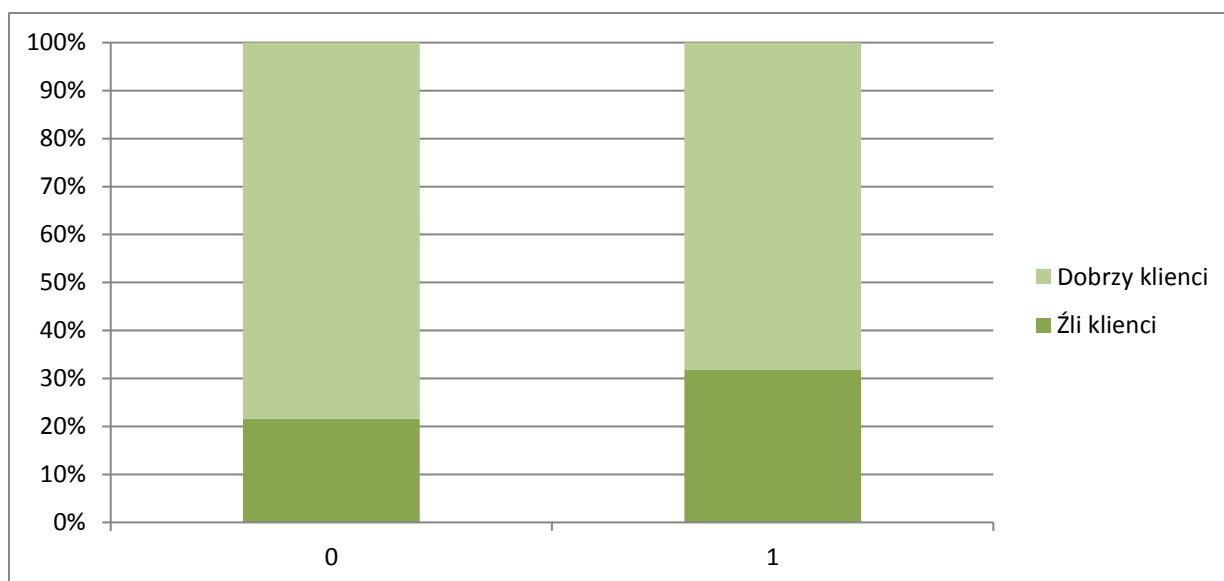


Rysunek 5: Nowy podział czwartej zmiennej jakościowej



Rysunek 6: Podział piątej zmiennej jakościowej

Dla tej zmiennej również należało przypadki z wartością 3 dodać do przypadków o wartości 2, a wartość 7 do 6. Największa część „złych” klientów była dla wartości = 4, najmniejsza dla 2-3 i 5.



Rysunek 7: Podział szóstej zmiennej jakościowej

Większa część „złych” klientów miała wartość 1, różnica między drugą wartością wyniosła ok. 10%.

By przetestować, czy różnice między grupami są istotnie statystycznie, zastosowano test niezależności Chi-kwadrat. Do sprawdzenia jaka jest siła tych zależności policzono współczynnik V Cramera, który przyjmuje wartości od 0 do 1. Bardzo słabą zależność wykazał dla zmiennych drugiej, piątej i szóstej ($< 0,1$), trochę większą dla pierwszej, trzeciej i czwartej. Ponieważ analiza wykazała, że wszystkie zmienne są istotnie zależne od zmiennej Zaufanie, zostały one użyte do dalszego modelowania. Poniżej tabela z dokładnymi wynikami uporządkowana w kolejności od najmniejszego współczynnika Cramera:

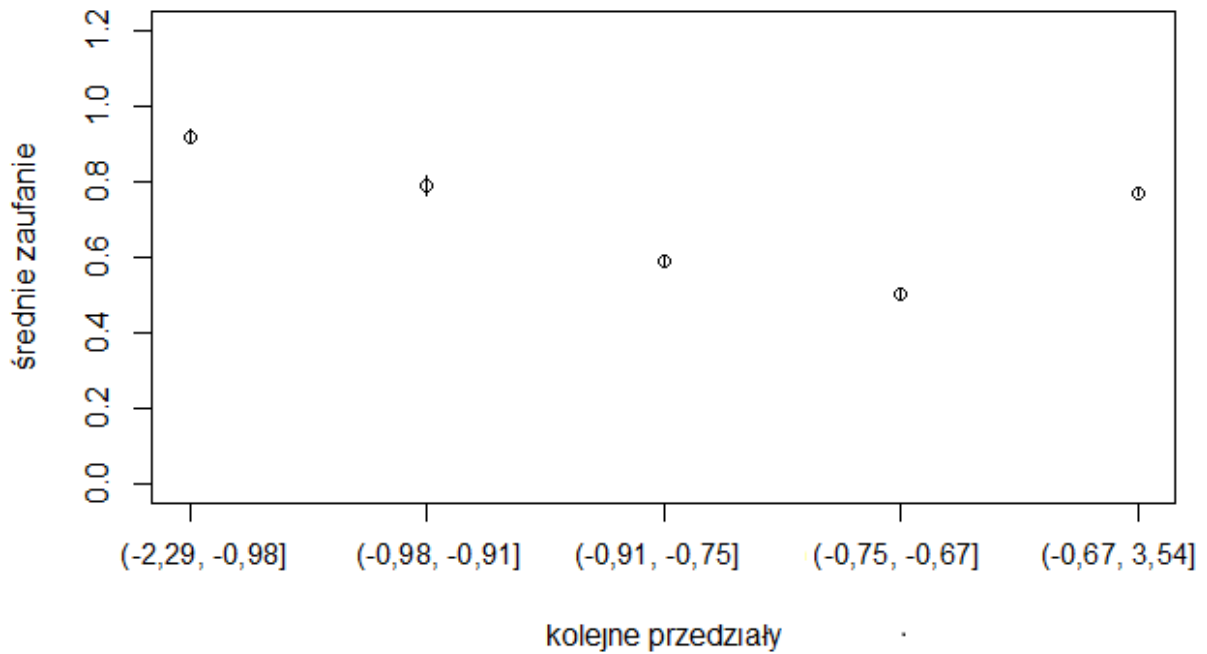
Zmienna	Statystyka χ^2	Istotność T	wsp. V Cramera
Jakościowa6	21,12	0,00	0,05
Jakościowa2	22,22	0,00	0,06
Jakościowa5	33,70	0,00	0,07
Jakościowa4	91,29	0,00	0,11
Jakościowa1	111,26	0,00	0,13
Jakościowa3	232,23	0,00	0,18

Tabela 1: Wyniki testu Chi-kwadrat i współczynniki V Cramera

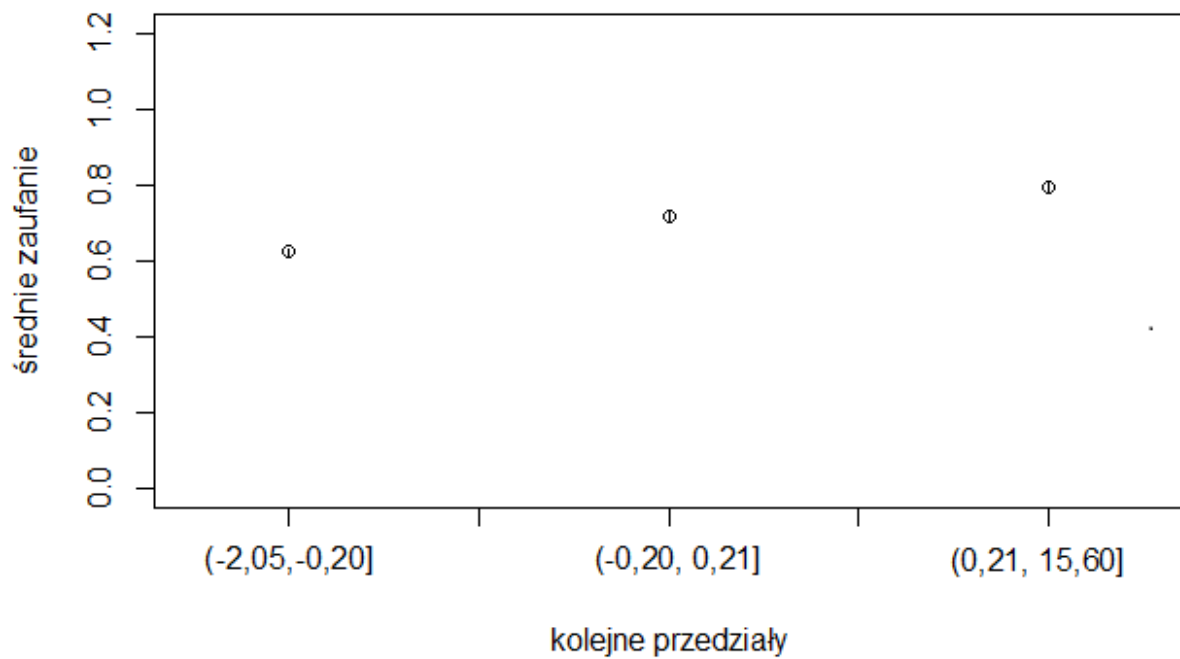
Część 3 – Eksploracja i modelowanie zmiennych ilościowych

Oprócz przeanalizowanych zmiennych jakościowych, dane składały się również z sześciu zmiennych ilościowych. Zostały one podzielone na przedziały przez drzewa decyzyjne. Poniżej przedstawiono średnie zaufanie i błędy standardowe dla kolejnych przedziałów

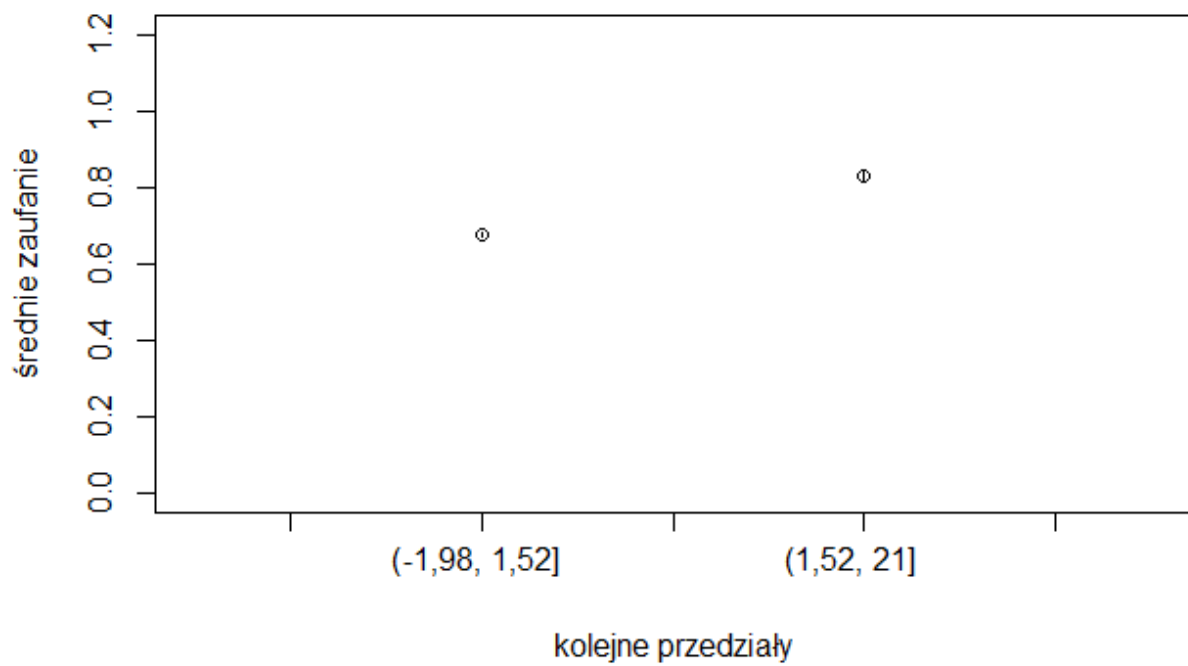
przekształconej zmiennej ilościowej. Metoda ta nie rozdzieliła drugiej zmiennej, tzn. że nie ma jej grup istotnie różnych pod względem zaufania.



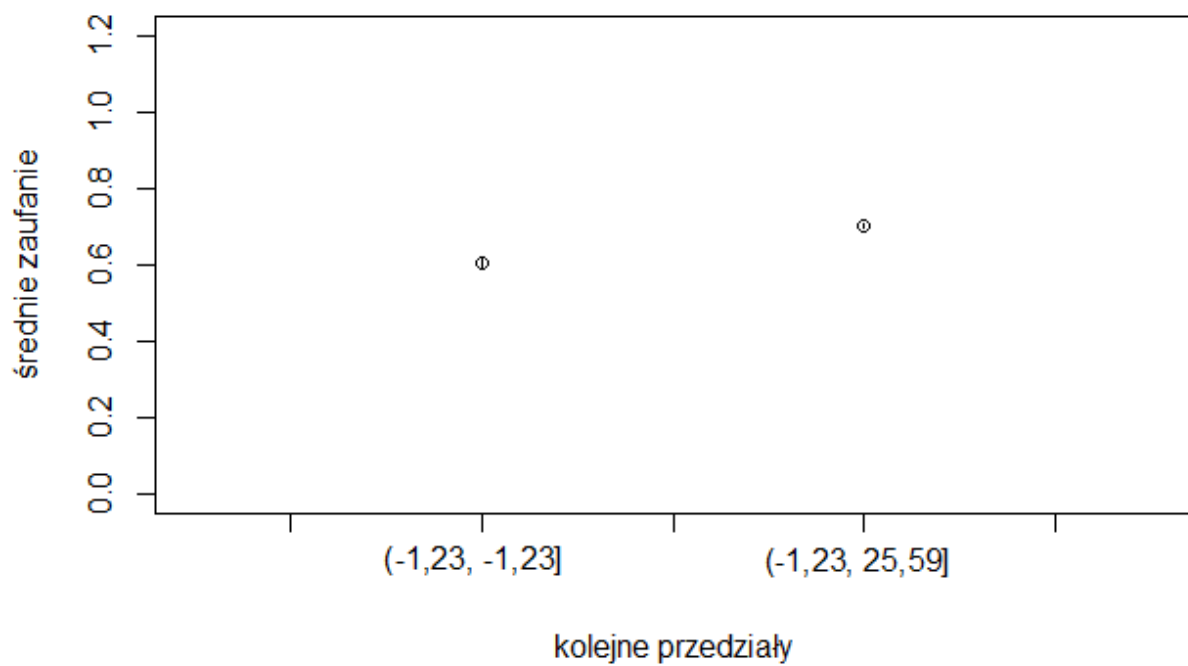
Rysunek 8: Wykres średniej i błędów standardowych dla pierwszej zmiennej ilościowej



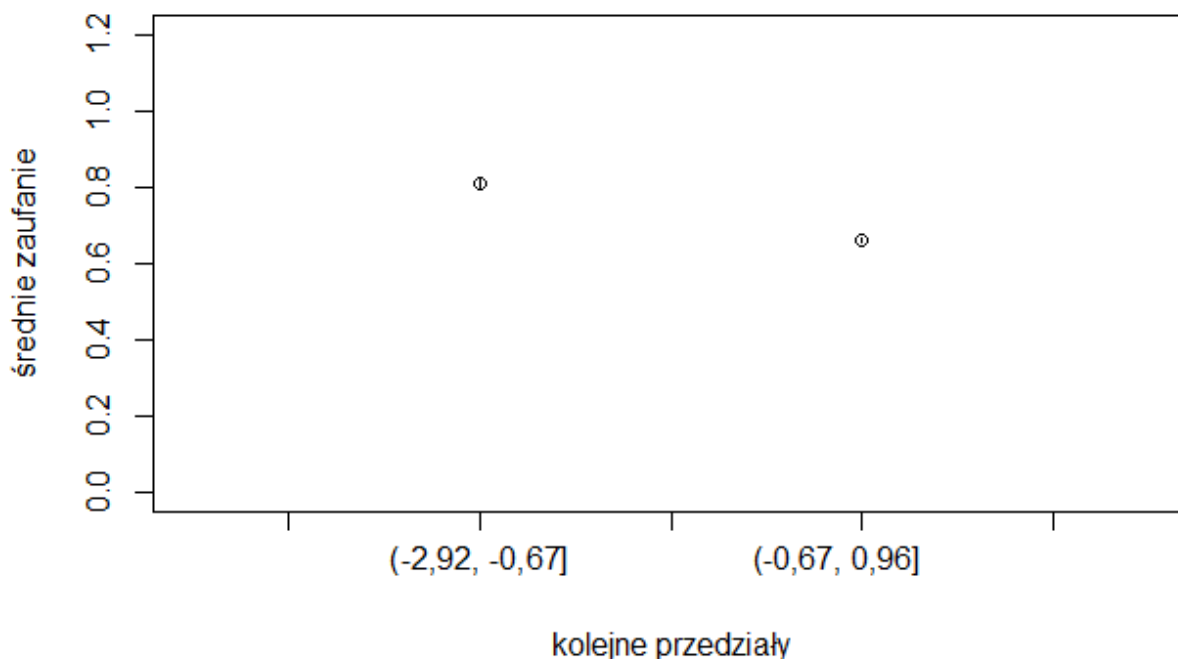
Rysunek 9: Wykres średniej i błędów standardowych dla trzeciej zmiennej ilościowej



Rysunek 10: Wykres średniej i błędów standardowych dla czwartej zmiennej ilościowej



Rysunek 11: Wykres średniej i błędów standardowych dla piątej zmiennej ilościowej



Rysunek 12: Wykres średniej i błędów standardowych dla szóstej zmiennej ilościowej

Policzono współczynnik η^2 , który określa, w jakim stopniu zmienność zaufania jest wyjaśniona przez zmienność danej zmiennej oraz przeprowadzono analizę wariancji.

Zmienna	Df efektu	Df błędu	Suma kwadratów	F	Istotność	η^2
ilościowa1	4	7105	102,78	132,73	0,00	5,75%
ilościowa6	1	7105	20,31	104,90	0,00	1,33%
ilościowa3	2	7105	23,21	59,95	0,00	1,02%
ilościowa5	1	7105	2,75	14,21	0,00	0,19%
ilościowa4	1	7105	2,15	11,13	0,00	0,08%

Tabela 2: Wyniki analizy wariancji i współczynnik η^2

Zaufanie jest najlepiej wyjaśnione przez zmienną ilościową1 – w 5,75%. Jest to mała wartość, jednak pokazuje zwiększoną zmienność zaufania na różnych poziomach tej zmiennej. Wszystkie zmienne są istotne, posortowane zostały wg współczynnika η^2 .

Część 4 – Analiza interakcji między zmiennymi ilościowymi, a jakościowymi

Przeprowadzono analizę wariancji dla pięciu zmiennych ilościowych i sześciu jakościowych wraz z ich interakcjami pierwszego rzędu. Dla modelu, z którego wcześniej wyłączono nieistotne zmienne metodą eliminacji wstecznej wyniki analizy wariancji przedstawiają się następująco:

Zmienna	Df efektu	Df błędu	Suma kwadratów	F	Istotność	η^2
ilościowa1	4	7105	102,78	152,6	0	2,32%
jakościowa4:jakościowa5	59	7105	33,46	3,37	0	2,20%
jakościowa3:jakościowa4	39	7105	23,49	3,58	0	1,59%
ilościowa1:jakościowa4	59	7105	23,92	2,41	0	1,39%
ilościowa3:jakościowa4	30	7105	15,6	3,09	0	1,25%
ilościowa6	1	7105	20,31	120,6	0	1,10%
jakościowa3	3	7105	18,88	37,38	0	0,90%
ilościowa1:jakościowa3	11	7105	12,64	6,83	0	0,76%
jakościowa2:jakościowa4	15	7105	11,58	4,59	0	0,73%
ilościowa6:jakościowa4	15	7105	8,61	3,41	0	0,66%
ilościowa3	2	7105	23,21	68,93	0	0,62%
ilościowa1:jakościowa5	16	7105	9,81	3,64	0	0,54%
jakościowa1	1	7105	12,72	75,57	0	0,51%
jakościowa4	15	7105	4,89	1,94	0,02	0,44%
ilościowa4:jakościowa4	15	7105	4,31	1,71	0,04	0,44%
jakościowa4:jakościowa6	14	7105	5,42	2,3	0	0,41%
ilościowa1:jakościowa1	4	7105	11,56	17,16	0	0,32%
jakościowa3:jakościowa5	11	7105	3,94	2,13	0,02	0,31%
ilościowa3:jakościowa5	8	7105	5,43	4,03	0	0,30%
ilościowa3:jakościowa3	6	7105	3,13	3,09	0,01	0,29%
ilościowa1:ilościowa6	4	7105	2,52	3,74	0	0,24%
ilościowa3:ilościowa5	2	7105	4,19	12,45	0	0,21%
ilościowa1:ilościowa4	4	7105	2,74	4,07	0	0,20%
ilościowa6:jakościowa5	4	7105	1,94	2,89	0,02	0,20%
jakościowa6	1	7105	1,68	9,99	0	0,18%
jakościowa1:jakościowa3	3	7105	1,4	2,77	0,04	0,16%
jakościowa2:jakościowa5	4	7105	2,47	3,66	0,01	0,13%
jakościowa2	1	7105	0,74	4,4	0,04	0,13%
jakościowa5:jakościowa6	4	7105	1,83	2,71	0,03	0,12%
jakościowa5	4	7105	2,68	3,97	0	0,11%
jakościowa3:jakościowa6	2	7105	2,63	7,8	0	0,11%
ilościowa1:ilościowa5	4	7105	0,84	1,25	0,29	0,10%
ilościowa5:jakościowa3	3	7105	1,82	3,6	0,01	0,10%
ilościowa5	1	7105	2,75	16,33	0	0,09%
ilościowa3:jakościowa2	2	7105	1,53	4,55	0,01	0,08%
ilościowa3:jakościowa6	2	7105	1,27	3,78	0,02	0,07%

ilościowa3:ilościowa4	2	7105	0,72	2,12	0,12	0,05%
ilościowa6:jakościowa2	1	7105	0,61	3,65	0,06	0,03%
ilościowa5:jakościowa2	1	7105	0	0,02	0,89	0,03%
ilościowa4	1	7105	2,15	12,79	0	0,01%

Tabela 3: Wyniki analizy wariancji i współczynnik η^2

Dane posortowano malejąco według współczynnika η^2 . Na czerwono zaznaczono zmienne nieistotne statystycznie. Sprawdzone współliniowość wszystkich istotnych czynników wskaźnikiem VIF. Jego wartości okazały się bardzo duże, dlatego zdecydowano zastosować model bez interakcji.

Część 5 – Budowa modelu regresji logistycznej

Zaufanie jest zmienną binarną, dlatego wybrano model logistyczny z funkcją wiążącą zwaną logitem, tzn.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right),$$

gdzie p to prawdopodobieństwo sukcesu (w tym przypadku zaufanie = 1).

Wszystkie zmienne jakościowe przekształcono na zmienne binarne, pierwszy poziom nazywa się bazowym, natomiast kolejne się do niego odnoszą.

Ponieważ wiele czynników dla zmiennej jakościowej4 było nieistotnych, pogrupowano je wg podobnego wpływu na zaufanie. Wyniki modelu z tak przekształconą zmienną są następujące:

Zmienna	β	95% p. ufn. β	SE	Z	Istotność	exp(β)
(Intercept)	3,61	(3; 4,25)	0,32	11,34	0,00	36,94
ilościowa3 – poziom 3	0,51	(0,35; 0,68)	0,08	6,08	0,00	1,67
jakościowa5 – poziom 4	0,4	(0,14; 0,67)	0,14	2,96	0,00	1,5
jakościowa4 – poziom 5	0,36	(0,12; 0,59)	0,12	2,96	0,00	1,43
ilościowa4 – poziom 2	0,28	(0,02; 0,55)	0,13	2,06	0,04	1,32
ilościowa3 – poziom 2	0,25	(0,12; 0,38)	0,07	3,81	0,00	1,29
jakościowa4 – poziom 4	0,2	(0; 0,41)	0,11	1,92	0,05	1,23
ilościowa5 – poziom 2	0,19	(0,04; 0,34)	0,08	2,43	0,02	1,21
jakościowa5 – poziom 2	0,1	(-0,08; 0,28)	0,09	1,11	0,27	1,11
jakościowa5 – poziom 5	0,09	(-0,09; 0,26)	0,09	0,95	0,35	1,09

jakościowa4 – poziom 2	0,02	(-0,18; 0,21)	0,1	0,15	0,88	1,02
------------------------	------	---------------	-----	------	------	------

jakościowa5 – poziom 3	-0,1	(-0,35; 0,15)	0,13	-0,78	0,43	0,91
jakościowa3 – poziom 4	-0,14	(-0,59; 0,32)	0,23	-0,61	0,54	0,87
jakościowa4 – poziom 3	-0,21	(-0,48; 0,05)	0,13	-1,58	0,11	0,81
jakościowa2 – poziom 2	-0,22	(-0,36; -0,08)	0,07	-3,11	0,00	0,8
jakościowa1 – poziom 2	-0,34	(-0,46; -0,23)	0,06	-5,96	0,00	0,71
jakościowa3 – poziom 3	-0,37	(-0,5; -0,25)	0,06	-5,84	0,00	0,69
jakościowa6 – poziom 2	-0,39	(-0,63; -0,15)	0,12	-3,12	0,00	0,68
jakościowa3 – poziom 2	-0,81	(-0,99; -0,63)	0,09	-8,87	0,00	0,44
ilościowa6 – poziom 2	-0,88	(-1,04; -0,73)	0,08	-10,86	0,00	0,41
ilościowa1 – poziom 2	-1,08	(-1,65; -0,54)	0,28	-3,8	0,00	0,34
ilościowa1 – poziom 5	-1,3	(-1,8; -0,85)	0,24	-5,37	0,00	0,27
ilościowa1 – poziom 3	-2,1	(-2,61; -1,64)	0,25	-8,55	0,00	0,12
ilościowa1 – poziom 4	-2,2	(-2,71; -1,74)	0,25	-8,94	0,00	0,11

Tabela 4: Wyniki regresji logistycznej dla wszystkich zmiennych

Zależności między zmiennymi bazową i pozostałymi odczytuje się z kolumny $\exp(\beta)$ w następujący sposób: gdy zmienna ilościowa1 = 2, to szansa na zaufanie = 1 w tej grupie wynosi $0,34 * \text{szansa na zaufanie} = 1$ w grupie dla zmienna ilościowa1 = 1, analogicznie dla pozostałych poziomów. Dla wartości > 1 jest to czynnik wzmacniający, a dla < 1 – czynnik ograniczający względem zmiennej bazowej. Tabela posortowana jest malejąco względem $\exp(\beta)$, na początku są więc zmienne wzmacniające, następnie neutralne, a na końcu ograniczające. Trzy czynniki mające największy wpływ to ilościowa3 – poziom 3, ilościowa1 – poziom 3 i ilościowa1 – poziom 4. W teście ilorazu wiarygodności, który porównuje wybrany model do modelu tylko z wyrazem wolnym otrzymana statystyka miała wartość 950, oznacza, że nasz model był istotny statystycznie.

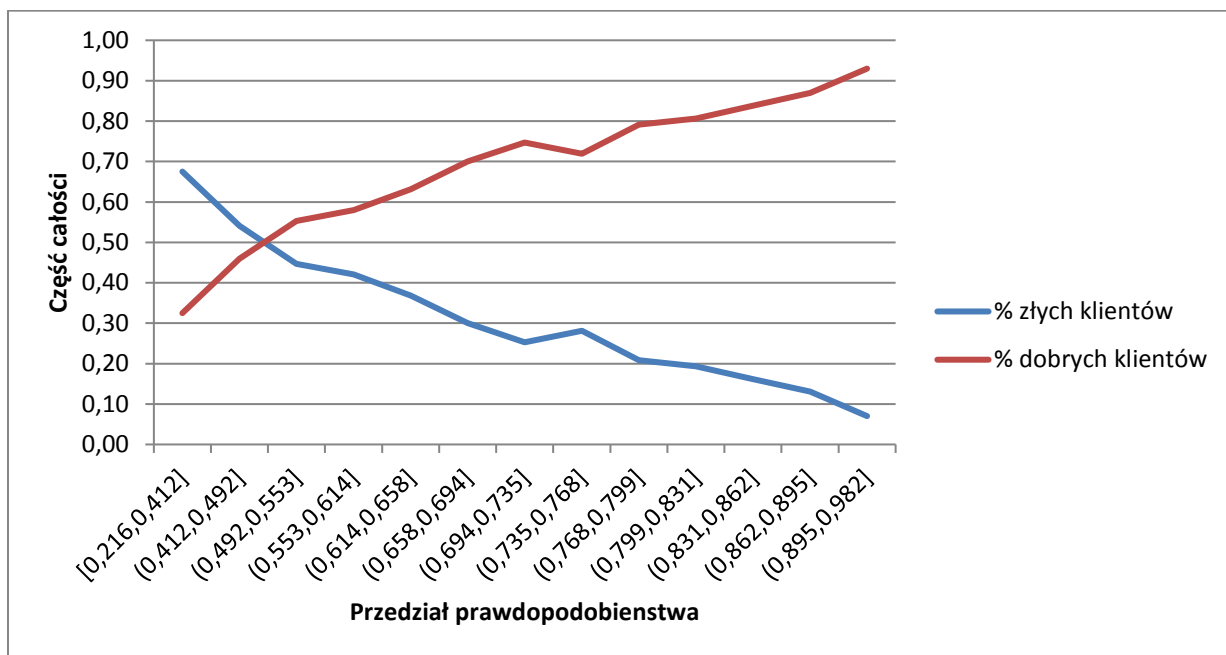
Przeprowadzono test Hosmera-Lemeshowa, który sprawdza, jak dobrze dopasowany jest model. Po kilku próbach wartości prawdopodobieństw dopasowanych przez model zostały podzielone na 13 przedziałów, ich zestawienie wraz z rzeczywistymi wartościami poniżej:

przedział prawdopodobieństwa	rzeczywiste		przewidywane	
	zaufanie = 0	zaufanie = 1	zaufanie = 0	zaufanie = 1
[0,216,0,412]	374	180	360	194
(0,412,0,492]	292	248	296	244
(0,492,0,553]	244	302	258	288

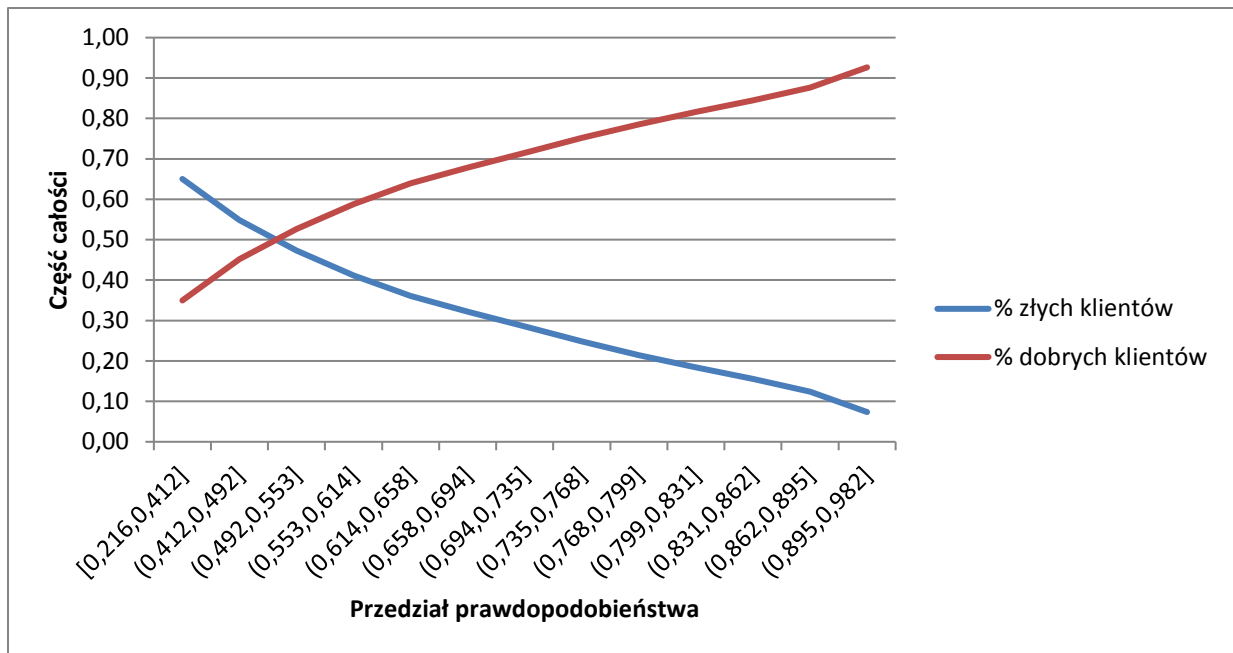
(0,553,0,614]	232	320	227	325
(0,614,0,658]	202	346	198	350
(0,658,0,694]	162	378	174	366
(0,694,0,735]	138	408	156	390
(0,735,0,768]	154	394	136	412
(0,768,0,799]	114	432	117	429
(0,799,0,831]	106	442	101	447
(0,831,0,862]	88	456	84	460
(0,862,0,895]	72	480	68	484
(0,895,0,982]	38	504	40	502

Tabela 5: Przewidywane i rzeczywiste wartości w teście H-L

Wyraźnie widać, że dla kolejnych przedziałów prawdopodobieństwa wzrastała liczba dobrych klientów i malała liczba złych. Na poniższych wykresach przedstawiono, jaką część stanowili dobrzy i źli klienci w danym przedziale.



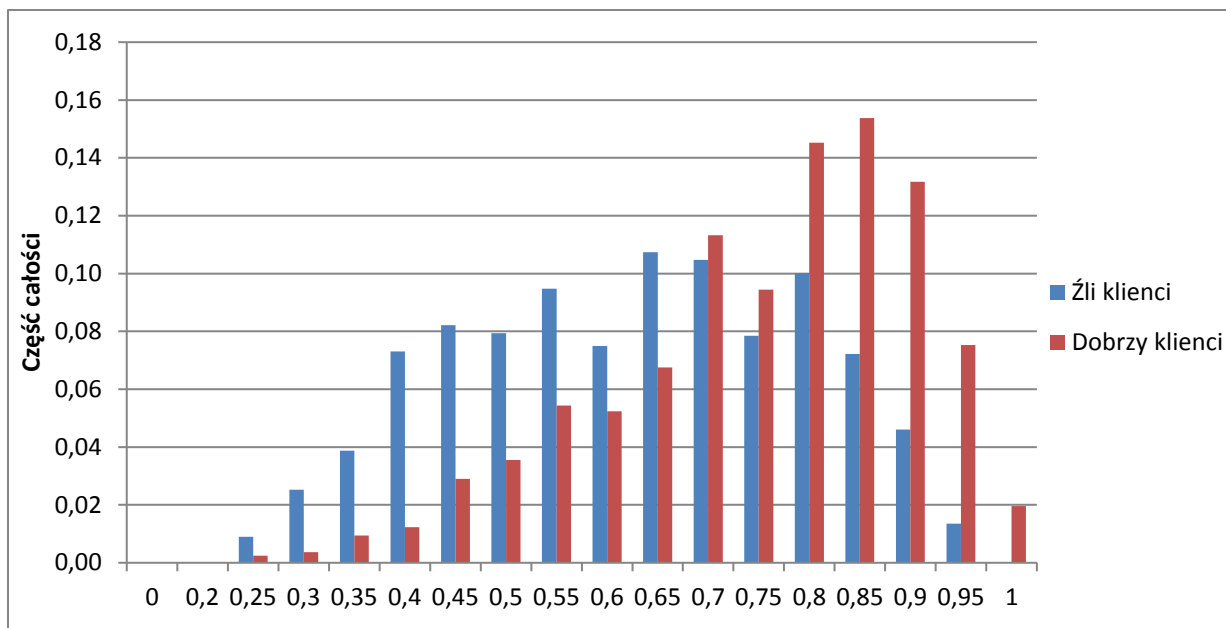
Rysunek 13: Wykres przedstawiający rzeczywisty % dobrych i złych klientów



Rysunek 14: Wykres przedstawiający przewidywany % dobrych i złych klientów

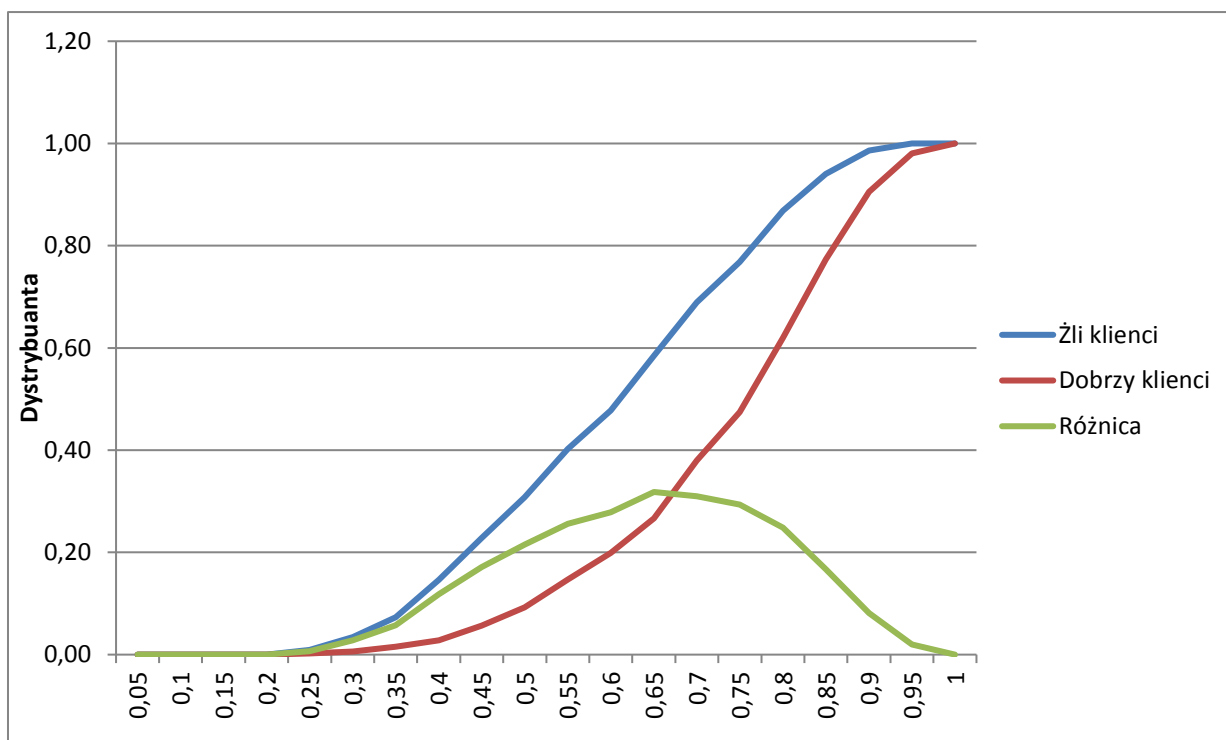
Na wykresach widać duże podobieństwo w przewidywanych, a rzeczywistych wartościach. Istotność testu H-L wyniosła 0,39, więc nie było podstaw do odrzucenia hipotezy, że model był dobrze dopasowany. Jego poprawność na całym modelu wyniosła 72%, dla różnych wylosowanych podzbiorów treningowych i testowych, za każdym razem ta wartość wahała się między 71%, a 73%. Współczynnik R^2 Nagelkerke'a wyniósł 0,18, co oznacza, że model objaśnia 18% zmienności zaufania. Pozostała zmienność to błąd oraz inne czynniki, których nie było w modelu.

Dopasowane prawdopodobieństwa podzielono według zaufania. Rozkłady tak utworzonych zmiennych przedstawiają się następująco:



Rysunek 15: Wykres przedstawiający rozkłady dopasowanych prawdopodobieństw w podziale na zaufanie

W teście Kolmogorowa Smirnowa statystyka miała wartość 0,33, rozkłady były istotnie różne i wyraźnie od siebie oddalone. Na poniższym wykresie przedstawione są dystrybuanty obu rozkładów oraz różnica między nimi.



Rysunek 16: Wykres przedstawiający dystrybuanty rozkładów dopasowanych prawdopodobieństw

Dystrybuanta złych klientów zaczyna znacznie rosnąć w okolicach prawdopodobieństwa sukcesu wynoszącego 0,35, natomiast dobrych dopiero ok. wartości 0,5, przy czym najszybszy wzrost ma od 0,75 wzwyż, co oznacza, że wśród klientów z zaufaniem = 1 dla znacznej większości model dopasował prawdopodobieństwa sukcesu większe od ¾ - intuicyjnie model wydaje się więc prawidłowy. Do lepszej jego oceny przeprowadzono analizę krzywej ROC, która opiera się na wskaźnikach zwanych czułość i specyficzność. Tabela przedstawiająca porównanie rzeczywistego wyniku do przewidywanego:

		Rzeczywistość	
		zaufanie = 0	zaufanie = 1
Przewidywanie	zaufanie = 0	TN	FN
	zaufanie = 1	FP	TP

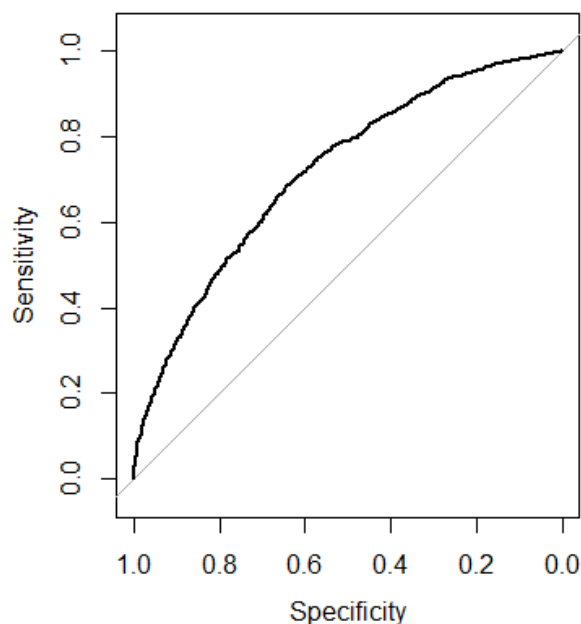
Tabela 6: Podział przypadków na TN, FN, FP i TP

Na jej podstawie:

$$\text{czułość} = \frac{TP}{TP+FN}$$

$$\text{specyficzność} = \frac{TN}{TN+FP}$$

Krzywa ROC to wykres tych wskaźników dla kolejnych punktów odcięć dobrych klientów od złych. Przedstawia się on następująco:



Rysunek 17: Wykres przedstawiający krzywą ROC

Wraz ze zwiększeniem punktu odcięcia maleje specyficzność, czyli część klientów poprawnie sklasyfikowanych jako złych, a rośnie czułość, czyli część klientów poprawnie sklasyfikowanych jako dobrych. Sumarycznie największą wartość dają dla odcięcia równego 0,67 – ten punkt powinien być wybrany wg kryterium Youdena. Jednak zależnie od tego, które przypadki – FP czy FN są w tym wypadku istotniejsze, można wybrać inny punkt graniczny. Pole pod krzywą wynosi 0,72, jest to prawdopodobieństwo, że dla losowo wybranych osób z obu grup, dobry klient będzie miał większe prawdopodobieństwo spłaty od złego. Model ma zadowalającą zdolność rozdzielenia klientów dobrych od złych.

Część 6 – Podsumowanie skuteczności modelu

Zastosowane testy potwierdziły, że dopasowany model jest istotny statystycznie i dobrze dopasowuje prawdopodobieństwa sukcesu. Na całym zbiorze danych przy odcięciu 0,5 poprawnie sklasyfikował 72% klientów. Czynniki objaśniły 18% wariancji zmiennej zaufanie. Największy wpływ zwiększający prawdopodobieństwo sukcesu wraz z ich współczynnikiem β miały:

Zmienna	β
ilościowa3 – poziom 3	0,51
jakościowa5 – poziom 4	0,40
jakościowa4 – poziom 5	0,36

Tabela 7: Najbardziej znaczące zmienne wzmacniające

natomiast zmniejszający:

Zmienna	β
jakościowa3 – poziom 2	-0,81
ilościowa6 – poziom 2	-0,88
ilościowa1 – poziom 2	-1,08
ilościowa1 – poziom 5	-1,30
ilościowa1 – poziom 3	-2,10
ilościowa1 – poziom 4	-2,20

Tabela 8: Najbardziej znaczące zmienne ograniczające

Istotność w teście Hosmera Lemeshowa wyniosła 0,39, co potwierdziło przyjęcie hipotezy o dobrym dopasowaniu modelu do danych.

Wyraźnie widać, że wraz ze wzrostem dopasowanego prawdopodobieństwa sukcesu wzrastała również część dobrych klientów. Rozkłady tych wartości w podziale na grupy w teście Kolmogorowa Smirnowa miały wartość statystyki 0,33, co potwierdziło, że są one istotnie różne i wyraźnie od siebie oddalone.

Pole pod krzywą ROC wyniosło 0,72, co oznacza, że model w zadowalający sposób rozdzielił obie grupy klientów.

Część 7 – Budowa karty scoringowej

Aby scoring był łatwy do zinterpretowania, chcemy, by jego wartość dla każdego klienta była w przedziale 0 – 100. W tym celu przeprowadzono następujące przekształcenia:

$$S_j = (\sum a_{ji} \beta_i - \min(\sum a_{ji} \beta_i)) * 100 / \max(\sum a_{ji} \beta_i),$$

gdzie S_j to wartość scoringu dla j -tego klienta, a_{ji} to wartość z macierzy zero-jedynkowej określającej poziomy zmiennych dla danego klienta – macierzy A .

Mając nową zmienną S możemy policzyć, jaki musiałby być wektor β' , taki, by S otrzymać jako wynik mnożenia tablicy A przez β' , bez dodatkowych przekształceń. Konieczne jest do tego znalezienie 24-ch liniowo niezależnych wektorów (bo tyle jest poziomów zmiennych). Z takiego układu równań wyznaczony jest wektor współczynników β' .

Zmienna	β	β'
(Intercept)		93,09
	3,61	

ilościowa3 - poziom 3	0,51	9,76
jakościowa5 - poziom 4	0,40	7,69
jakościowa4 - poziom 5	0,36	6,79
ilościowa4 - poziom 2	0,28	5,28
ilościowa3 - poziom 2	0,25	4,82
jakościowa4 - poziom 4	0,20	3,89
ilościowa5 - poziom 2	0,19	3,59
jakościowa5 - poziom 2	0,10	1,93
jakościowa5 - poziom 5	0,09	1,63
jakościowa4 - poziom 2	0,02	0,29
jakościowa5 - poziom 3	-0,10	-1,88
jakościowa3 - poziom 4	-0,14	-2,69
jakościowa4 - poziom 3	-0,21	-4,04
jakościowa2 - poziom 2	-0,22	-4,18
jakościowa1 - poziom 2	-0,34	-6,54
jakościowa3 - poziom 3	-0,37	-7,04
jakościowa6 - poziom 2	-0,39	-7,32
jakościowa3 - poziom 2	-0,81	-15,40
ilościowa6 - poziom 2	-0,88	-16,77
ilościowa1 - poziom 2	-1,08	-20,47
ilościowa1 - poziom 5	-1,30	-24,61
ilościowa1 - poziom 3	-2,10	-39,82
ilościowa1 - poziom 4	-2,20	-41,77

Tabela 9: Karta scoringowa posortowana malejąco wg β'

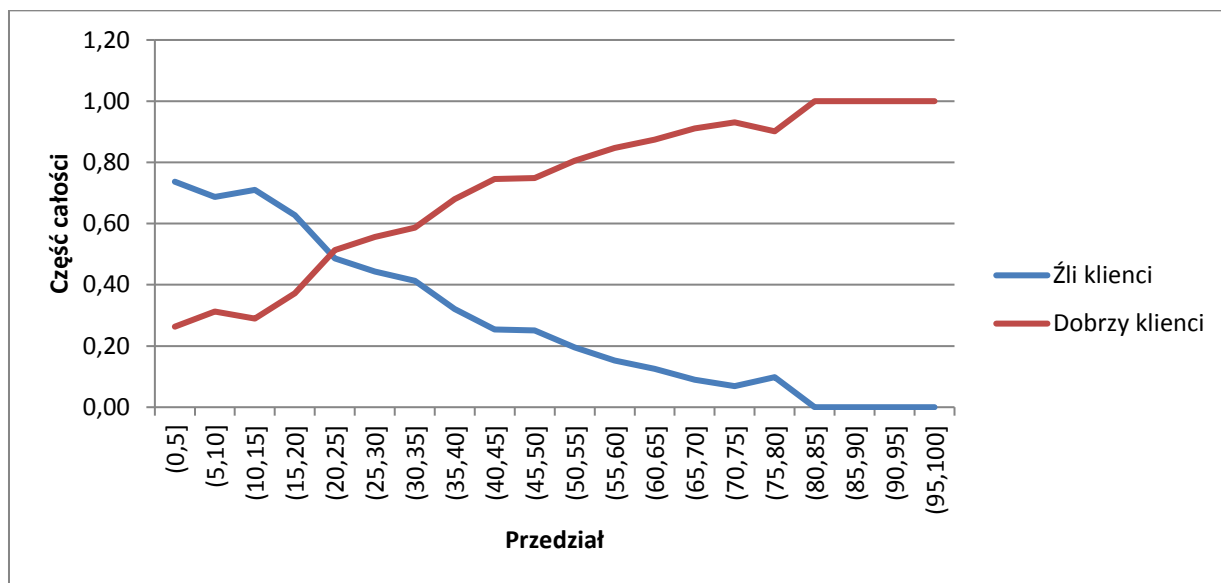
Tabela scoringu dla 20 przedziałów w podziale na rzeczywistą zmienną zaufanie:

przedział	zaufanie = 0	zaufanie = 1	część klientów z zaufaniem = 0	część klientów z zaufaniem = 1
(0,5]	28	10	0,74	0,26
(5,10]	66	30	0,69	0,31
(10,15]	162	66	0,71	0,29
(15,20]	236	140	0,63	0,37
(20,25]	220	232	0,49	0,51
(25,30]	252	316	0,44	0,56
(30,35]	256	364	0,41	0,59
(35,40]	286	606	0,32	0,68
(40,45]	184	540	0,25	0,75
(45,50]	198	590	0,25	0,75
(50,55]	142	586	0,20	0,80
(55,60]	94	522	0,15	0,85
(60,65]	52	362	0,13	0,87
(65,70]	22	224	0,09	0,91

(70,75]	8	108	0,07	0,93
(75,80]	10	92	0,10	0,90
(80,85]	0	56	0,00	1,00
(85,90]	0	28	0,00	1,00
(90,95]	0	6	0,00	1,00
(95,100]	0	10	0,00	1,00

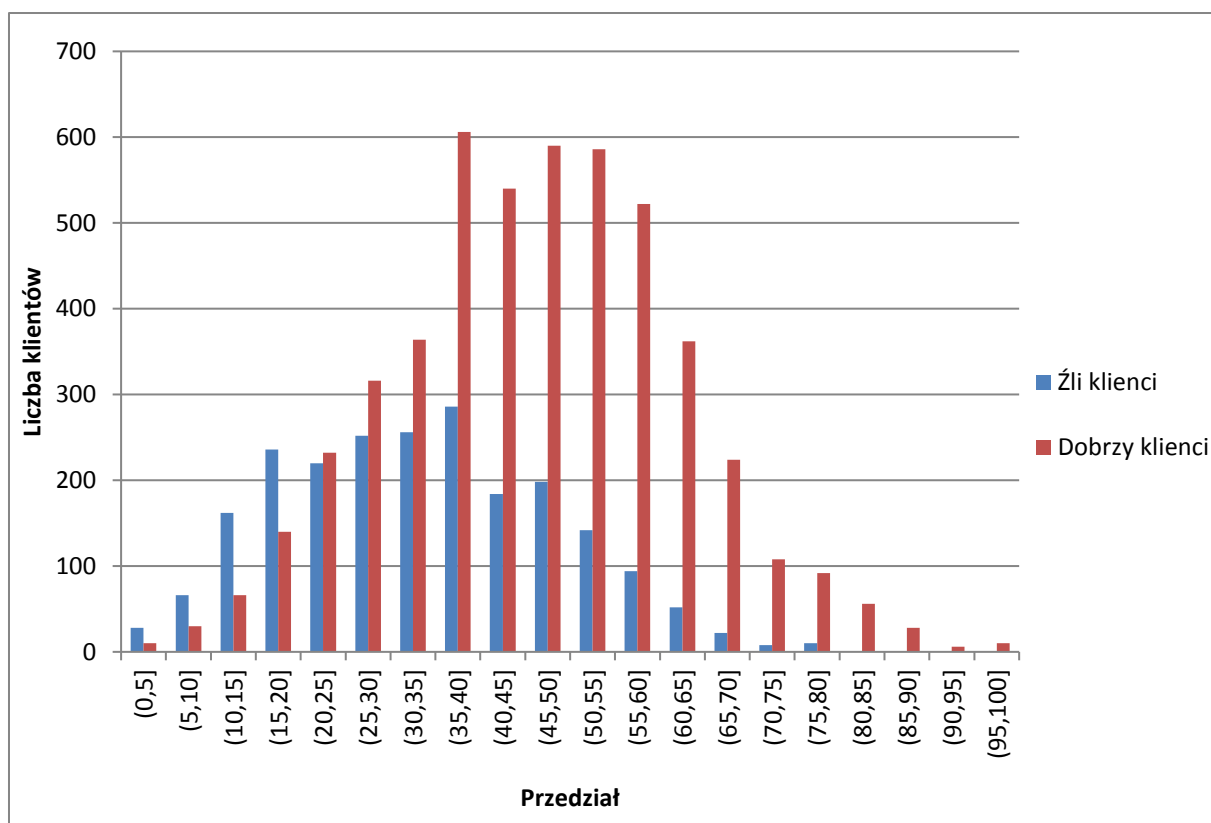
Tabela 10: Liczebność klientów w poszczególnych przedziałach scoringu

Na wykresie część dobrych i złych klientów przedstawia się to następująco:



Rysunek 18: Wykres przedstawiający część dobrych i złych klientów w danych przedziałach scoringu

Natomiast liczbowo wygląda to tak:



Rysunek 19: Wykres przedstawiający liczebność klientów w danych przedziałach scoringu

Dobrzy klienci przeważali już w przedziale (20, 25], jednak poziom odcięcia wybieramy zależnie od poziomu ryzyka, jakie możemy podjąć. Przypuśćmy, że dopuszczamy ryzyko 10% i przyjmujemy punkt odcięcia równy 80.

Część 8 – Klasyfikacja klienta złego i dobrego.

Poniżej przedstawiono przykład zastosowania karty scoringowej dla dwóch klientów z różnych grup. Próg odcięcia dla klienta dobrego wynosił 0,80 w punktacji β' . Poniżej wartości 0,80 traktowano klienta jako ryzykownego pod względem spłacalności.

Zmienna	Zły klient	Dobry klient	β'	Zły klient * β'	Dobry klient * β'
(Intercept)	1	1	93,09	93,09	93,09
ilościowa1 - poziom 2	0	0	-20,47	0,00	0,00
ilościowa1 - poziom 3	1	0	-39,82	-39,82	0,00
ilościowa1 - poziom 4	0	0	-41,77	0,00	0,00
ilościowa1 - poziom 5	0	0	-24,61	0,00	0,00
ilościowa3 - poziom 2	0	1	4,82	0,00	4,82
ilościowa3 - poziom 3	0	0	9,76	0,00	0,00

ilościowa4 - poziom 2	0	0	5,28	0,00	0,00
ilościowa5 - poziom 2	1	1	3,59	3,59	3,59
ilościowa6 - poziom 2	1	0	-16,77	-16,77	0,00
jakościowa1 - poziom 2	1	0	-6,54	-6,54	0,00
jakościowa2 - poziom 2	0	0	-4,18	0,00	0,00
jakościowa3 - poziom 2	1	0	-15,40	-15,40	0,00
jakościowa3 - poziom 3	0	1	-7,04	0,00	-7,04
jakościowa3 - poziom 4	0	0	-2,69	0,00	0,00
jakościowa5 - poziom 2	0	0	1,93	0,00	0,00
jakościowa5 - poziom 3	0	0	-1,88	0,00	0,00
jakościowa5 - poziom 4	0	0	7,69	0,00	0,00
jakościowa5 - poziom 5	0	1	1,63	0,00	1,63
jakościowa6 - poziom 2	1	1	-7,32	-7,32	-7,32
jakościowa4 - poziom 2	0	0	0,29	0,00	0,00
jakościowa4 - poziom 3	0	0	-4,04	0,00	0,00
jakościowa4 - poziom 4	1	0	3,89	3,89	0,00
jakościowa4 - poziom 5	0	1	6,79	0,00	6,79
			Suma punktacji scoringowej	14,71	95,55

Tabela 11: Zastosowanie karty scoringowej dla dwóch klientów

Zły klient ma w scoringu wartość 14,71, jest dużo poniżej przyjętego ograniczenia, wg modelu dopasowane prawdopodobieństwo spłaty to 0,37. Natomiast dobry klient ma wartość 95,55, natomiast prawdopodobieństwo równe 0,98. Czytelniej ich poziomy wszystkich zmiennych w tabeli poniżej:

Zmienna	Poziom złego klienta	Poziom dobrego klienta
ilościowa1	3	1
ilościowa3	1	2
ilościowa4	1	1
ilościowa5	2	2
ilościowa6	2	1
jakościowa1	2	1
jakościowa2	1	1
jakościowa3	2	3
jakościowa4	4	5
jakościowa5	1	5
jakościowa6	2	2

Tabela 12: Poziomy wartości przyjmowanych zmiennych dla dwóch klientów