

# Współczesna Gospodarka



Contemporary Economy  
Electronic Scientific Journal  
www.wspolczesnagospodarka.pl

Vol. 7 Issue 1 (2016) 81-94  
ISSN 2082-677X

## SPOŁECZNO-EKONOMICZNE DETERMINANTY PRZESTĘPCZOŚCI POWROTNEJ. WYBRANE PROBLEMY STOSOWANIA METOD ILOŚCIOWYCH DO IDENTYFIKACJI ZALEŻNOŚCI

Kinga Kądziołka

### Streszczenie

Artykuł porusza problematykę prognozowania indywidualnych zachowań przestępczych z wykorzystaniem wybranych metod statystycznych i metod *data mining*. Analizowano problem klasyfikacji sprawców do dwóch grup ryzyka recydywy. Porównana została jakość różnych klasyfikatorów, takich jak sieć neuronowa, model logitowy, drzewo klasyfikacyjne i metoda wektorów nośnych. Łączna trafność prognozowania dla wszystkich analizowanych modeli przekroczyła 70%, jednak modele te charakteryzowały się dużym odsetkiem tzw. "fałszów negatywnych".

**Słowa kluczowe:** przestępczość powrotna, ryzyko recydywy, analiza przeżycia, *data mining*, regresja logistyczna

### Wstęp

W artykule poruszona została tematyka prognozowania ryzyka powrotu do przestępczości oraz automatycznej klasyfikacji przestępców do odpowiednich grup ryzyka recydywy z wykorzystaniem metod ilościowych. W wielu krajach (m.in. Kanada, Anglia, Holandia, Szwecja, Dania) podejmowane są próby opracowania narzędzi (zarówno statystycznych jak też ustrukturyzowanych narzędzi klinicznych), które mają na celu prognozowanie ryzyka recydywy<sup>1</sup>. W literaturze wskazuje się, że inspiracją do rozwoju badań w tym zakresie było opracowanie modelu RNR (ang. *Risk – Need - Responsivity*). W różnych krajach stosowane są różne narzędzia. Cechą wspólną tego typu narzędzi jest automatyczna klasyfikacja sprawców do określonej grupy ryzyka na podstawie odpowiedzi udzielonych na pytania dotyczące tzw.

---

<sup>1</sup> Przegląd funkcjonalności wybranych narzędzi prognozowania ryzyka powrotu do przestępczości został dokonany w pracy D. Wójcik, *Stosowanie w postępowaniu karnym narzędzi diagnostyczno - prognostycznych służących oszacowaniu ryzyka powrotności do przestępstwa*, [w:] M. Mozgawa (red.) *Prawo w Działaniu 16*, Wolters Kluwer, Warszawa 2013.

obszarów problemowych, które według wyników badań naukowych są dodatnio skorelowane z powrotem do przestępczości<sup>2</sup>.

Istniejące w Polsce akty prawne nakazują klasyfikację przestępców do grup ryzyka<sup>3</sup>. Jednakże, jak zauważa J. Chojecka (2014), brak jest rzetelnych narzędzi diagnostycznych, które byłyby pomocne w dokonywaniu takiego podziału<sup>4</sup>.

Celem artykułu jest zaprezentowanie możliwości wykorzystania wybranych metod statystycznych i *data mining* do oceny ryzyka powrotu do przestępczości oraz zwrócenie uwagi na pewne problemy pojawiające się przy próbach automatycznej klasyfikacji (przy pomocy metod ilościowych) sprawców do grup ryzyka recydywy.

## 1. Przestępczość powrotna w ujęciu kryminologicznym i kodeksowym

Pojęcie recydywy jest różnie rozumiane w zależności od tego czy rozważa się recydywę kryminologiczną, penitencjarną czy jurydyczną. O recydywie kryminologicznej mówimy, gdy jednostka ponownie popełnia przestępstwo. W tym ujęciu recydywy bez znaczenia pozostaje to, czy za poprzedni czyn była skazana, czy odbyła całą czy część kary oraz ile czasu upłynęło od odbycia kary<sup>5</sup> (w przypadku wcześniejszego skazania). Recydywą w znaczeniu jurydycznym (prawnym) jest ta część zjawiska recydywy, która odpowiada ujętym w ustawie kryteriom. Natomiast za recydywę penitencjarną uznaje się odbywanie co najmniej po raz drugi kary pozbawienia wolności<sup>6</sup>. Analizując rozmiar recydywy penitencjarnej czy jurydycznej należy mieć na uwadze, że istnieje tzw. zatarcie skazania – po pewnym czasie skazanie zostaje usunięte z rejestru karnego i osoba, która popełniła kolejne przestępstwo, po zatarciu poprzedniego wyroku, nie będzie recydywistą. Wykres (rysunek 1) przedstawia liczbę prawomocnie skazanych recydywistów w Polsce w latach 2004 - 2014. Wśród osób, które ponownie trafiają do zakładów karnych, najwięcej jest skazanych po raz drugi (zarówno wśród ogółu osób ponownie skazanych jak i wśród skazanych w warunkach art. 64 kk). Wykres (rysunek 2) przedstawia strukturę populacji ponownie skazanych w zakładach karnych wg ich powrotności do zakładu w 2014 r. Również w poprzednich latach analizowanego okresu w strukturze osób ponownie skazanych największy udział stanowiły osoby odbywające zasadniczą karę pozbawienia wolności po raz drugi. Szczegółowe analizy wskaźników podobieństwa struktur populacji skazanych wg liczby skazań w latach 2004 – 2014 pokazały, że podobieństwo to było bardzo duże lub duże (wartość wskaźnika podobieństwa struktur ponownie skazanych w 2014 r. i kolejnych latach z okresu 2004 – 2013 była większa niż 0,6). Jednocześnie wskaźnik podobieństwa struktur charakteryzował się dużą dynamiką zmian. Przykładowo, dla lat 2012 i 2014 wynosił on 0,682 a dla lat 2013 i 2014 wartość wskaźnika podobieństwa struktur wyniosła aż 0,993. W latach 2009 – 2012 osoby odbywające zasadniczą karę pozbawienia wolności po raz siódmy i więcej stanowiły ok. 10% osób ponownie skazanych. W pozostałym okresie udział tej grupy w strukturze osób ponownie skazanych zmieniał się od 2,28% – 4,3%.

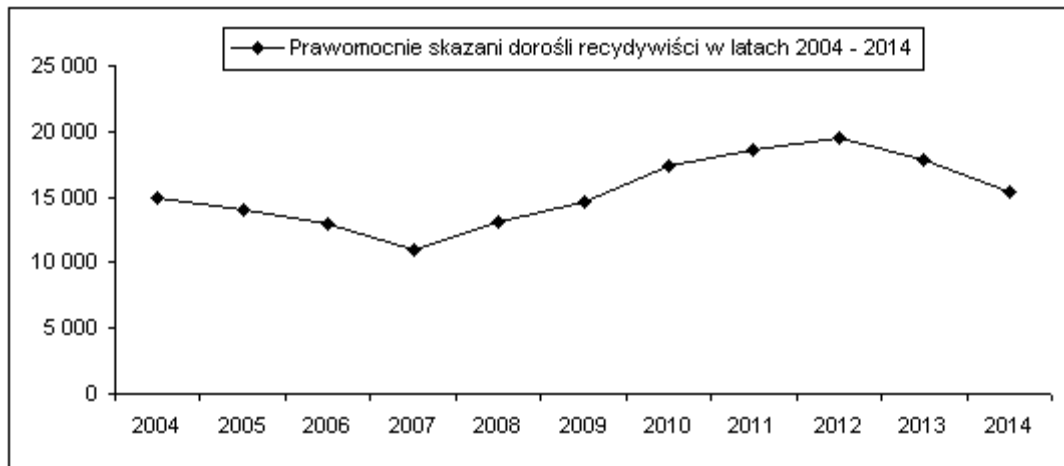
<sup>2</sup> por. B. Stańdo – Kawecka, *Wybrane problemy profesjonalizacji organów probacyjnych i klasyfikacji sprawców oddanych pod dozór do grup ryzyka*, Nowa Kodyfikacja Prawa Karnego. Tom XXXIII, Wrocław 2014, s. 20-22.

<sup>3</sup> Rozporządzenie Ministra Sprawiedliwości z dn. 26 lutego 2013 w sprawie sposobu wykonywania obowiązków i uprawnień przez kuratorów sądowych wprowadziło wymóg szacowania ryzyka powrotu do przestępstwa przez sprawcę oddanego pod dozór kuratora sądowego oraz nałożyło na kuratorów obowiązek zaklasyfikowania dozorowanego do jednej z trzech grup ryzyka.

<sup>4</sup> J. Chojecka, *Model dla wszystkich? Spory wokół koncepcji szacowania ryzyka recydywy*, „Resocjalizacja Polska” 2014, nr 7, s. 97.

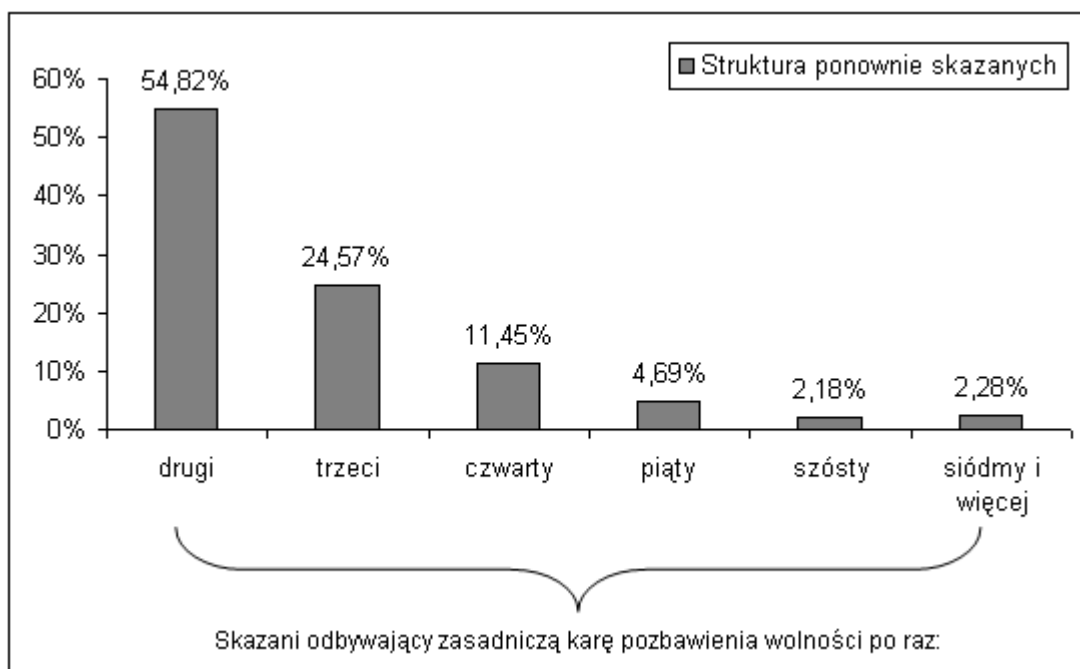
<sup>5</sup> por. R. Szczepanik, *Stawanie się recydywistą. Kariery instytucjonalne osób powracających do przestępczości*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015, s. 17 - 19.

<sup>6</sup> por. J. Zygunt, *Prawne modele zwalczania powrotności do przestępstwa w polskim prawie karnym*, „Czasopismo Prawa Karnego i Nauk Penalnych”, 2008 nr 2, s. 225.



**Rysunek 1.** Prawomocnie skazani dorośli recydywiści w latach 2004 – 2014

Źródło: opracowanie własne na podstawie danych ISWS.



**Rysunek 2.** Ponownie skazani wg powrotności do zakładów karnych w 2014 r.

Źródło: opracowanie własne na podstawie danych Centralnego Zarządu Służby Więziennej.

## 2. Czynniki ryzyka recydywy – przegląd literatury

W kryminologii pozytywistycznej powszechne było przekonanie, że wielokrotnie karani recydywiści muszą być dotknięci jakimiś schorzeniami psychicznymi, gdyż tylko w ten sposób można było wytłumaczyć fakt, że pomimo wielokrotnego karania nie wyciągają z tego

wniosków i nadal popełniają przestępstwa<sup>7</sup>. Ekonomiczne podejście do przestępczości oparte na teorii oczekiwanej użyteczności umożliwiło wyjaśnienie przestępczości powrotnej w kontekście wyborów podyktowanych istniejącymi sposobnościami przestępczymi. I. Ehrlich (1973) zwraca uwagę, że sam fakt ujęcia i ukarania sprawcy nie musi mieć większego wpływu na decyzje o popełnianiu przez niego kolejnych przestępstw<sup>8</sup>. W literaturze wskazywane są różne czynniki skorelowane z ponownym popełnianiem przestępstw. Wśród tych czynników wskazuje się m.in. takie czynniki jak<sup>9</sup>: płeć, wiek, wiek, w którym sprawca popełnił pierwsze przestępstwo, stan cywilny, nadużywanie alkoholu/narkotyków, defekty (obciążenia) dziedziczne, trudności w nauce, kontakty ze środowiskiem przestępczym/nieodpowiednie wykorzystanie czasu wolnego, rodzaj i waga popełnionych czynów przestępczych. Ponadto dokonuje się podziału czynników ryzyka recydywy na tzw. czynniki statyczne i dynamiczne. Czynniki statyczne obejmują takie charakterystyki jak np. płeć, wiek, w którym doszło do pierwszego skazania, rasa. Czynniki te nie ulegają zmianie. Z kolei do czynników dynamicznych zalicza się te charakterystyki, które mogą ulegać zmianie pod wpływem działań korekcyjnych. Są to m.in.: nadużywanie alkoholu/narkotyków, zachowania agresywne, poglądy<sup>10</sup>.

W Polsce badaniem przestępczości powrotnej zajmował się m. in. T. Szymanowski (2010), który zwrócił uwagę na pewne zależności między sytuacją społeczno – ekonomiczną sprawców a ponownym popełnianiem przestępstw. Autor zwraca m. in. uwagę na związek między wykształceniem uzyskanym przez badanych skazanych przed odbywaniem kary pozbawienia wolności z ich ponowną karalnością. Przykładowo, wśród osób, które przed uwięzieniem uzyskały wykształcenie podstawowe aż 61,8% dopuściło się kolejnych przestępstw. Ponadto Szymanowski wskazuje, że związek z ponowną karalnością miało również wykonywanie pracy przed skazaniem przez badanych recydywistów. Osoby pracujące przed skazaniem po wykonaniu kary wykazały niższy odsetek recydywy niż osoby niepracujące. Jednocześnie należy mieć na uwadze, że skazanie sprawcy za przestępstwo często powoduje „naznaczenie” takiej osoby i problem w znalezieniu przez nią pracy<sup>11</sup>. Z kolei trudności w znalezieniu pracy i brak środków do życia mogą prowadzić do powrotu na drogę przestępczą.

### 3. Ocena ryzyka recydywy z wykorzystaniem metod ilościowych

W pracach zagranicznych autorów podejmowana była problematyka stosowania wybranych metod statystycznych i data mining do prognozowania ryzyka recydywy i klasyfikacji sprawców do odpowiednich grup ryzyka.

P. van der Heijden i N. Tollenaar (2013) porównali skuteczność prognozowania ryzyka recydywy dla różnych modeli. Analizowali m. in. takie modele jak: model logitowy, LDA, FDA, sieć neuronowa, metoda wektorów nośnych (SVM), drzewo klasyfikacyjne, metoda najbliższych sąsiadów. Autorzy analizowali dane dotyczące dorosłych sprawców skazanych w

<sup>7</sup> por. [J. Błachut i in., *Kryminologia*, InfoTrade, Gdańsk 2001, s. 179].

<sup>8</sup> por. [I. Ehrlich, *Participation in Illegitimate Activities: A Theoretical and Empirical Investigation*, „The Journal of Political Economy”, 81(3), 1973, s. 529].

<sup>9</sup> por. B. Hołyst, *Podstawy i zakres indywidualnej prognozy kryminologicznej*, „Probacja” nr 1/2013; D. Wójcik, *Stosowanie w postępowaniu karnym narzędzi...*; R. Szczepanik, *Stawianie się recydywistą. Kariery instytucjonalne osób powracających do przestępczości*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015; T. Szymanowski, *Recydywa w Polsce: zagadnienia prawa karnego, kryminologii i polityki karnej*, Wolters Kluwer Polska, Warszawa 2010; N. Tinik, D. Hudak, *Examining the Factors Associated with Recidivism*, <https://dspace.rmu.edu/xmlui/handle/11347/25>, 2014, dostęp: 01.12.2015.

<sup>10</sup> por. D. Wójcik, *Stosowanie w postępowaniu karnym narzędzi...*, s. 63.

<sup>11</sup> W Polsce bezrobocie wśród osób, które były skazane lub tymczasowo aresztowane jest dużo wyższe niż bezrobocie wg BAEL, por. K. Kądziołka, *Determinanty przestępczości w Polsce. Aspekt ekonomiczno – społeczny w ujęciu modelowania ekonometrycznego*, niepublikowana rozprawa doktorska, Uniwersytet Ekonomiczny w Katowicach, 2015, s. 79-80.

Holandii. Prognozowane były 3 rodzaje recydywy: recydywa ogólna, recydywa w przypadku przestępstw o charakterze seksualnym i recydywa w przypadku przestępstw przeciwko życiu i zdrowiu<sup>12</sup>. W przypadku prognozowania ryzyka recydywy ogólnej oraz ryzyka recydywy dla przestępstw przeciwko życiu i zdrowiu wykorzystali do estymacji parametrów modeli dane dotyczące 10000 sprawców (losowo wybranych spośród dostępnej bazy zawierającej informacje dotyczące 159298 sprawców). Jako zbiór testowy wykorzystali losowo wybrane dane dotyczące również 10000 sprawców, które nie zostały wykorzystane w procesie estymacji parametrów modeli. W przypadku prognozowania ryzyka recydywy przestępstw seksualnych wykorzystano zbiór danych obejmujący informacje o 1392 sprawcach, przy czym 1/3 elementów tego zbioru posłużyła jako zbiór testowy. Wśród zmiennych objaśniających uwzględnili zarówno czynniki statyczne jak i dynamiczne, przy czym zbiory zmiennych objaśniających były różne dla prognozowania różnych rodzajów recydywy. Autorzy porównywali uzyskane modele dla różnych rodzajów recydywy pod kątem: dopasowania danych teoretycznych do empirycznych, zdolności dyskryminacyjnej modelu oraz tzw. użyteczności klinicznej. Do oceny uzyskanych modeli wykorzystane zostały m. in. wielkość pola pod krzywą ROC, błędy klasyfikacji oraz łączna trafność prognozowania. Model LDA był jednym z najlepszych modeli pod kątem zdolności dyskryminacyjnych w przypadku prognozowania każdego z analizowanych rodzajów recydywy. Jednakże łączna trafność prognozowania analizowanych modeli była porównywalna. Przykładowo, w przypadku prognozowania recydywy ogólnej łączna trafność prognozowania najgorszego modelu (metoda najbliższego sąsiada) wynosiła 67%, natomiast dla najlepszego modelu (MARS) 72,9%.

C. Rudin i in. (2015) badali skuteczność wybranych modeli klasyfikacyjnych (m.in.: drzewa klasyfikacyjne, model logitowy, SVM, las losowy oraz autorski model nazwany SLIM) w zagadnieniu prognozowania czy osoba zwolniona z zakładu karnego dopuści się ponownego przestępstwa w ciągu 3 lat od opuszczenia więzienia. Ponadto autorzy zaprezentowali możliwość zastosowania metod analizy asocjacji do generowania reguł decyzyjnych mogących wspomagać proces identyfikacji grup ryzyka recydywy. Analizowane były następujące rodzaje recydywy: recydywa ogólna, recydywa przestępstw przeciwko życiu i zdrowiu, recydywa przestępstw narkotykowych, recydywa przestępstw seksualnych, recydywa przestępstw przeciwko mieniu, recydywa przestępstw przeciwko porządkowi publicznemu. Analizowane dane dotyczyły 33796 sprawców, którzy opuścili zakłady karne w 1994 r. (dane U.S. Department of Justice, Bureau of Justice Statistics). Wśród zmiennych objaśniających uwzględniono 49 charakterystyk, obejmujących zarówno czynniki dynamiczne jak i statyczne. Oceny zdolności predykcyjnej modeli dokonano na podstawie krzywych ROC (ang. *Receiver Operating Characteristic*<sup>13</sup>) i wartości AUC (ang. *Area Under Curve*), będącej wielkością pola pod krzywą ROC. Nie istniała metoda, która byłaby najlepsza (w sensie przyjętych kryteriów oceny jakości modeli) dla prognozowania ryzyka każdego z analizowanych rodzajów recydywy. Na tle analizowanych modeli słabymi zdolnościami predykcyjnymi charakteryzował się algorytm CART generujący drzewa klasyfikacyjne.

T. D. Allen i L. L. Bench (2013) wykorzystali regresję logistyczną do oceny ryzyka recydywy przestępców seksualnych. Analizowane dane dotyczyły 389 mężczyzn, sprawców przestępstw na tle seksualnym, którzy opuścili więzienie stanowe w Utah. W modelu uwzględniono 51 zmiennych objaśniających, m. in.: wiek, rasa, stan cywilny, IQ, nadużywanie

<sup>12</sup> Należy mieć na uwadze, że w różnych krajach obowiązują różne kodeksy karne i pojęcia takie jak „przestępczość” czy „przestępczość przeciwko życiu i zdrowiu” nie są tożsame. Przykładowo jako zabójstwo w niektórych krajach uznawane jest każde umyślne pozbawienie człowieka życia (łącznie z dzieciobójstwem, eutanazją czy pobiciem ze skutkiem śmiertelnym). W innych krajach zbiór czynów ujmowanych w statystyce policyjnej jako zabójstwa nie obejmuje np. eutanazji.

<sup>13</sup> Szczegółowy opis wykorzystania krzywych ROC do oceny jakości klasyfikatorów został dokonany w pracy M. Misztal, *Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań*, „Taksonomia” nr 23, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 328, 2014.

alkoholu/narkotyków. Łączna trafność prognozowania wynosiła 70% , przy czym model w 71,5% poprawnie sklasyfikował recydywistów oraz w 65,3% poprawnie sklasyfikował osoby, które w analizowanym okresie nie popełniły kolejnego przestępstwa na tle seksualnym.

W ocenie modeli prognozowania ryzyka recydywy istotna (oprócz łącznej trafności prognozowania) jest właśnie problematyka tzw. „falszów pozytywnych” i „falszów negatywnych”. Z „falszem pozytywnym” mamy do czynienia, gdy model przewiduje, że recydywa wystąpi, podczas, gdy ona nie występuje. Z kolei z „falszem negatywnym” mamy do czynienia, gdy model przewiduje, że recydywa nie wystąpi, a faktycznie ma ona miejsce. B. Hołyst (2013, s. 22) zwraca uwagę, że „falsze pozytywne” byłyby obciążeniem głównie dla uwięzionych przestępców. Z kolei przewidywanie „falszów negatywnych” powodowałoby obciążenie kosztami błędnych decyzji ofiar przyszłych przestępstw popełnionych przez zwolnionych przestępców. Na problem ten szczególna uwaga zostanie zwrócona w dalszej części pracy.

W Polsce również podejmowane były próby adaptacji metod data mining do analizy problemu przestępczości powrotnej. J. K. Gierowski (2005) wykorzystał metodę k-średnich do badania zależności między wybranymi czynnikami ryzyka recydywy nieletnich sprawców przestępstw. Badaniem objęto grupę 98 chłopców przebywających w zakładach karnych. Uwzględniono czynniki ryzyka ujęte w kwestionariuszu SARVY, obejmujące 4 kategorie danych<sup>14</sup>: czynniki historyczne, czynniki społeczne, czynniki indywidualne oraz czynniki ochronne. Wykorzystanie metody k-średnich umożliwiło zredukowanie całego zbioru danych do średnich poszczególnych grup i identyfikację konfiguracji czynników w ramach poszczególnych skupień.

### 3.1. Charakterystyka analizowanych danych

Do analizy wykorzystano ogólnodostępne dane zebrane i opisane przez R. A. Berk'a i współautorów (1980) dotyczące 432 mężczyzn po odbyciu kary pozbawienia wolności w więzieniu w stanie Maryland<sup>15</sup>. Osoby te były obserwowane przez rok od czasu opuszczenia zakładu karnego. Zbiór danych obejmuje następujące charakterystyki osób:

- arrest* – zmienna binarna przyjmująca wartość 1, jeśli dana osoba została aresztowana za kolejne przestępstwo w okresie 52 tygodni od zwolnienia oraz 0 w przeciwnym przypadku
- week* – liczba tygodni, w czasie których osoba nie została ponownie aresztowana
- fin* – zmienna binarna przyjmująca wartość 1, gdy osoba po opuszczeniu zakładu karnego otrzymywała pomoc finansową oraz 0 w przeciwnym przypadku
- age* – wiek osoby w chwili opuszczenia zakładu karnego
- race* – zmienna binarna przyjmująca wartość 1 w przypadku rasy czarnej i 0 w przeciwnym przypadku
- wexp* – zmienna binarna przyjmująca wartość 1, jeśli osoba przed skazaniem wykonywała pracę zawodową i 0 w przeciwnym przypadku
- mar* – zmienna binarna przyjmująca wartość 1, gdy dana osoba była w związku małżeńskim i 0 w przeciwnym przypadku
- paro* – zmienna binarna przyjmująca wartość 1, jeśli dana osoba została zwolniona warunkowo i 0 w przeciwnym przypadku
- prio* – liczba wcześniejszych skazań danej osoby za przestępstwa
- educ* – liczba lat edukacji

<sup>14</sup> Szczegółowa charakterystyka tych grup czynników dostępna jest w pracy: J. K. Gierowski, *Czynniki ryzyka i opiniowanie przemocy u adolescentów*, „Farmakoterapia w psychiatrii i neurologii”, 2/2005, s. 118.

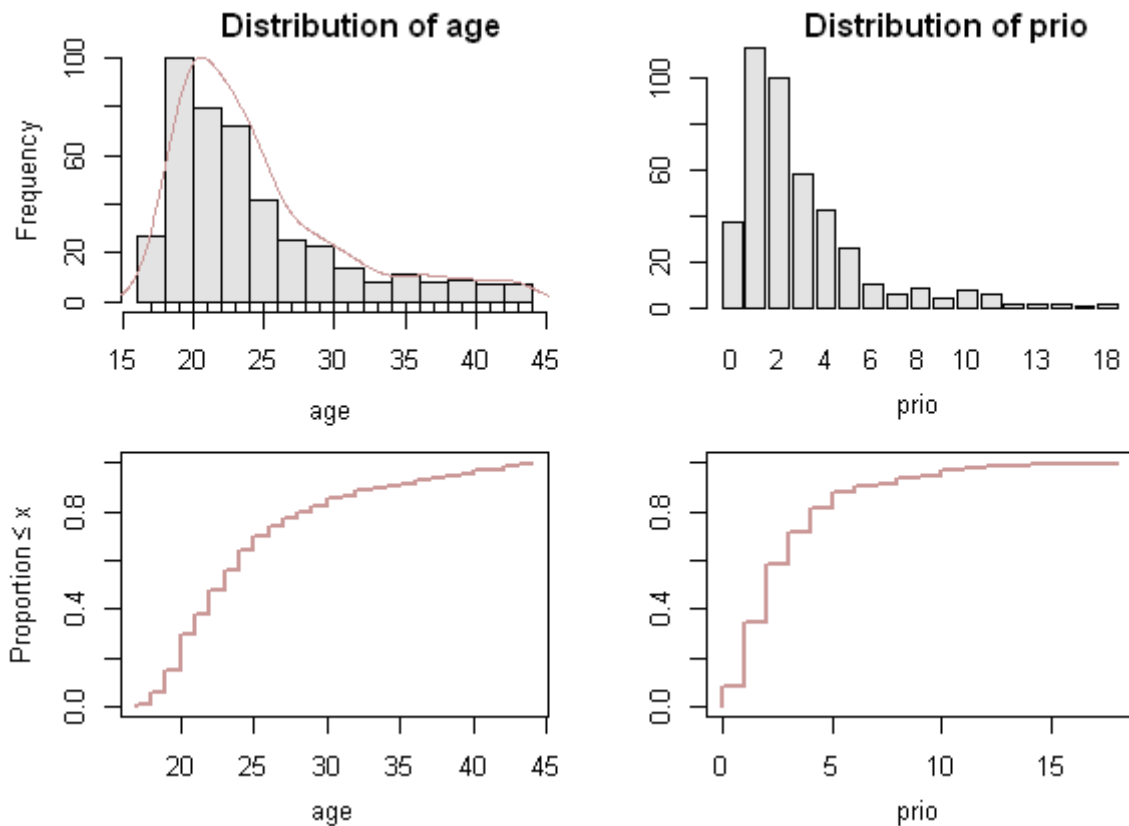
<sup>15</sup> Źródło danych: <http://socserv.mcmaster.ca/jfox/Courses/R/ICPSR/Rossi.txt>, data dostępu 03.10.2015. Prezentowane wyniki uzyskano z wykorzystaniem darmowych programów Gretl i R.

*emp1* – zmienna binarna przyjmująca wartość 1, jeśli w pierwszym tygodniu po opuszczeniu więzienia dana osoba wykonywała pracę zawodową i 0 w przeciwnym przypadku

⋮

*emp52* – zmienna binarna przyjmująca wartość 1, jeśli w 52 tygodniu po opuszczeniu więzienia dana osoba wykonywała pracę zawodową i 0 w przeciwnym przypadku

Z uwagi na dużą liczbę brakujących wartości pominięto w dalszych analizach zmienne *emp1* – *emp52*. Rysunek 3 przedstawia histogramy oraz wykresy wartości skumulowanych dla zmiennych *age* i *prio*. Można zauważyć, że większość osób w momencie zwolnienia była w wieku do 25 lat. Ponadto, większość z nich nie była wcześniej karana więcej niż trzy razy.



**Rysunek 3.** Histogramy i wartości skumulowane dla wybranych zmiennych

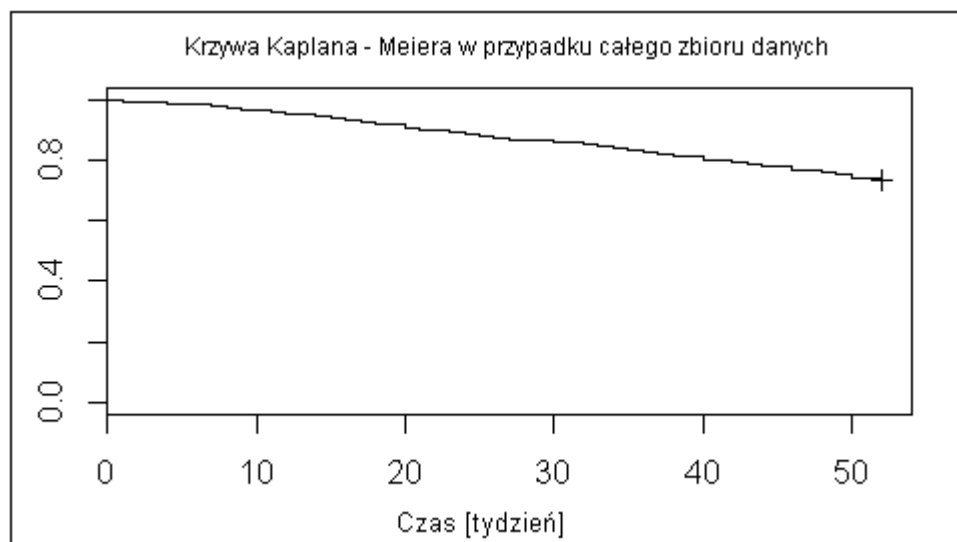
Źródło: opracowanie własne.

Do graficznej prezentacji zależności między prawdopodobieństwem recydywy a czasem od zwolnienia z zakładu karnego wykorzystano metodę analizy przeżycia. Analiza przeżycia jest zbiorem procedur statystycznych w analizie danych, dla których badaną zmienną jest czas, w którym pojawia się dane zdarzenie<sup>16</sup>. Z kolei czas przeżycia to okres między stanem początkowym a momentem wystąpienia zdarzenia. Czas przeżycia oznaczany będzie przez  $T$ . Podstawą analizy przeżycia jest tzw. funkcja przeżycia (ozn.  $S(t)$ ), która określa jakie jest prawdopodobieństwo przeżycia (tutaj rozumiane jako prawdopodobieństwo nie zostania

<sup>16</sup> por. A. Stanisławski, *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*. Tom 3. Analizy wielowymiarowe, Kraków 2007, s. 355.

aresztowanym za kolejne przestępstwo) dłużej niż do czasu  $t$ , tzn.  $S(t)=Pr(T>t)$ . Do estymacji krzywej przeżycia wykorzystana została metoda Kaplana – Meiera opisana szczegółowo m. in. w pracy R. Korzonka (2009). Wykres (rysunek 4) przedstawia funkcję przeżycia w przypadku całego analizowanego zbioru danych. W ciągu 52 tygodni od opuszczenia więzienia ponownie aresztowanych zostało 114 mężczyzn, co stanowiło 25,38% ogółu badanych osób. Następnie analizowany zbiór danych podzielono na podgrupy, według wybranych cech, celem zobrazowania różnic między krzywymi przeżycia w ramach poszczególnych grup. Formalnej oceny istotności różnic między uzyskanymi funkcjami dokonano za pomocą testu log – rank, którego wynik potwierdził przypuszczenia o różnicy między krzywymi w ramach poszczególnych podgrup. Osoby, które przed skazaniem wykonywały pracę wykazywały mniejszy odsetek recydywy niż osoby nie pracujące przed skazaniem. W grupie osób, które przed skazaniem wykonywały pracę 21,05% mężczyzn zostało ponownie aresztowanych za przestępstwo. W grupie osób, które nie pracowały przed skazaniem odsetek ten wynosił 33,51%. Podobnie, osoby będące w związku małżeńskim wykazywały mniejszy odsetek recydywy niż osoby nie będące w związku (rysunek 5). Z kolei dokonując podziału wg zmiennej *race* okazało się, że różnice między uzyskanymi krzywymi przeżycia w ramach wyróżnionych grup nie były istotne statystycznie.

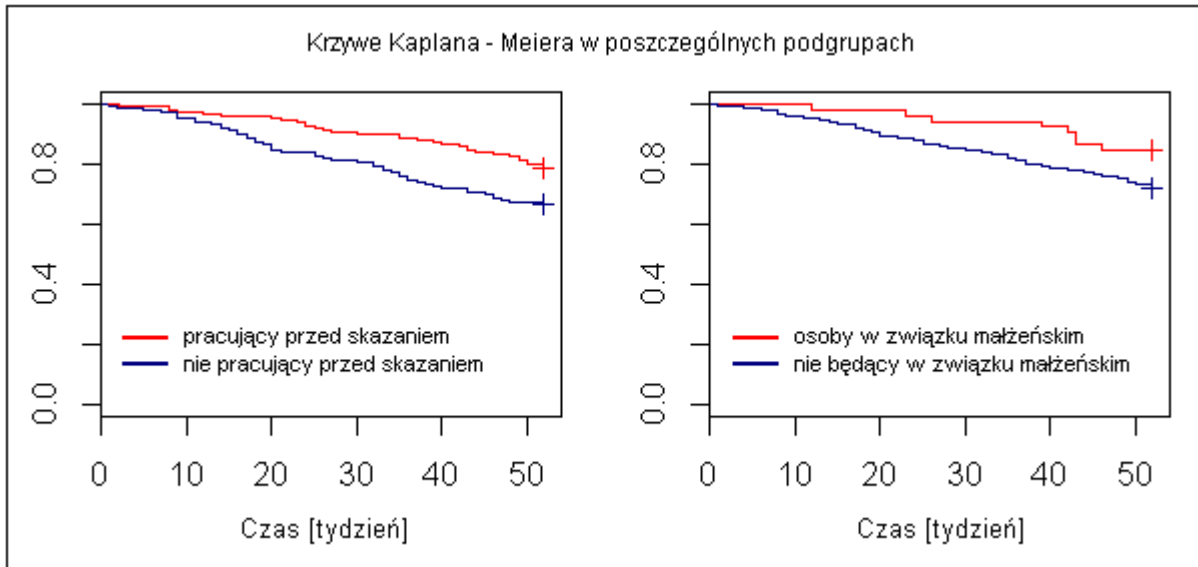
Analiza przeżycia obejmuje również m. in. wykorzystanie modeli hazardu. Podejście to w badaniach przestępczości powrotnej wykorzystali m. in.: Bierens H. J., Carvalho R. J. (2002), K. Masyn i B. Muthen (2005), B. Monnery (2013).



**Rysunek 4.** Krzywa Kaplana – Meiera dla całego zbioru danych

Źródło: opracowanie własne.





Rysunek 5. Krzywe przeżycia w poszczególnych podgrupach

Źródło: opracowanie własne.

### 3.2. Problemy identyfikacji grup ryzyka recydywy

Na podstawie analizowanego zbioru danych oszacowano parametry modelu logitowego, w którym zmienną objaśnianą była zmienna zero-jedynkowa, przyjmująca wartość 1, gdy dana osoba została ponownie aresztowana za przestępstwo w ciągu 52 tygodni od opuszczenia zakładu karnego oraz 0 w przeciwnym przypadku. Zmiennymi objaśniającymi były: *fin*, *age*, *race*, *wexp*, *mar*, *paro*, *prio*, *educ*. Tablica 1 przedstawia oszacowania parametrów modelu.

Tablica 1. Oszacowania parametrów modelu logitowego

	<i>Współczynnik</i>	<i>Błąd stand.</i>	<i>z</i>	<i>wartość p</i>	
<i>const</i>	1,21814	0,857686	1,4203	0,15553	
<i>fin</i>	-0,415285	0,22947	-1,8098	0,07033	*
<i>age</i>	-0,0684669	0,0244432	-2,8011	0,00509	***
<i>race</i>	0,407562	0,366456	1,1122	0,26606	
<i>wexp</i>	-0,0946897	0,252806	-0,3746	0,70799	
<i>mar</i>	-0,478402	0,42433	-1,1274	0,25956	
<i>paro</i>	-0,0959377	0,236777	-0,4052	0,68534	
<i>prio</i>	0,0938566	0,0382573	2,4533	0,01416	**
<i>educ</i>	-0,267199	0,157294	-1,6987	0,08937	*

Źródło: opracowanie własne.

Uzyskane znaki współczynników przy poszczególnych zmiennych objaśniających są zgodne z oczekiwaniami. Skłonność do popełniania kolejnych przestępstw okazała się większa w przypadku osób, które w przeszłości były karane. Istotnymi czynnikami okazały się również wiek, liczba lat edukacji oraz otrzymywane wsparcie finansowe po wyjściu z więzienia. Osoby, które otrzymywały wsparcie finansowe po opuszczeniu zakładu karnego były mniej skłonne do popełnienia kolejnych przestępstw. Czynnikiem zwiększającym ryzyko recydywy był młody

wiek sprawców. Prawdopodobieństwo powrotu do przestępczości malało wraz z wiekiem. Malało ono również ze wzrostem poziomu wykształcenia sprawców.

Do oceny trafności prognozowania modelu wykorzystano odpowiednie wskaźniki uzyskane na podstawie tzw. tablicy trafień (tablica 2).

**Tablica 2.** Tablica trafień

Faktyczne	Przewidywane		Razem
	Y*=0	Y*=1	
Y=0	$n_{00}$	$n_{01}$	$n_{0.}$
Y=1	$n_{10}$	$n_{11}$	$n_{1.}$
Razem	$n_{.0}$	$n_{.1}$	$n$

Źródło: T. Kufel, *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2011, s. 149.

Procentową trafność prognozowania możemy oceniać za pomocą następujących wskaźników<sup>17</sup>:

$$\text{łączna trafność prognozowania: } \frac{n_{00} + n_{11}}{n}$$

$$\text{trafność prognozowania dla Y=1: } \frac{n_{11}}{n_{1.}}$$

$$\text{trafność prognozowania dla Y=0: } \frac{n_{00}}{n_{0.}}$$

Dla analizowanego modelu łączna trafność prognozowania wyniosła 73,6%, trafność prognozowania dla Y=0 wyniosła 97,17%, jednakże trafność prognozowania dla Y=1 wyniosła tylko 7,89%. Uzyskany model charakteryzuje się dużym odsetkiem „fałszów negatywnych”. Trafność prognozowania dla Y=1 jest bardzo niska. Dla większości przypadków, w których osoby dokonały kolejnych przestępstw, model przewidywał, że recydywa nie wystąpi. Dla analizowanego zbioru danych porównano również trafność prognoz uzyskanych z wykorzystaniem innych modeli klasyfikacyjnych takich jak: drzewo klasyfikacyjne, metoda wektorów nośnych i sieć neuronowa<sup>18</sup>. W przypadku wszystkich rozważanych modeli problemem była niska trafność prognoz dla Y=1 (tablica 3).

**Tablica 3.** Trafność prognozowania modeli

Model	Łączna trafność prognozowania	Trafność prognozowania dla Y=1	Trafność prognozowania dla Y=0
drzewo klasyfikacyjne	75,69%	11,40%	98,74%
metoda wektorów nośnych	74,77%	4,39%	100%
sieć neuronowa	75,23%	12,28%	97,80%
model logitowy	73,60%	7,89%	97,17%

Źródło: opracowanie własne.

Dokonano również porównania zdolności predykcyjnej rozważanych klasyfikatorów z wykorzystaniem krzywych ROC i wartości AUC (tablica 4, rysunek 6). W przypadku każdego z klasyfikatorów  $AUC > 0,5$ . Ponadto, dla sieci neuronowej i metody wektorów nośnych  $AUC > 0,7$ . W literaturze przyjmuje się, że jeśli  $AUC > 0,7$  to klasyfikacja jest akceptowalna.

<sup>17</sup> Por. T. Kufel, *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2011, s. 149 – 150.

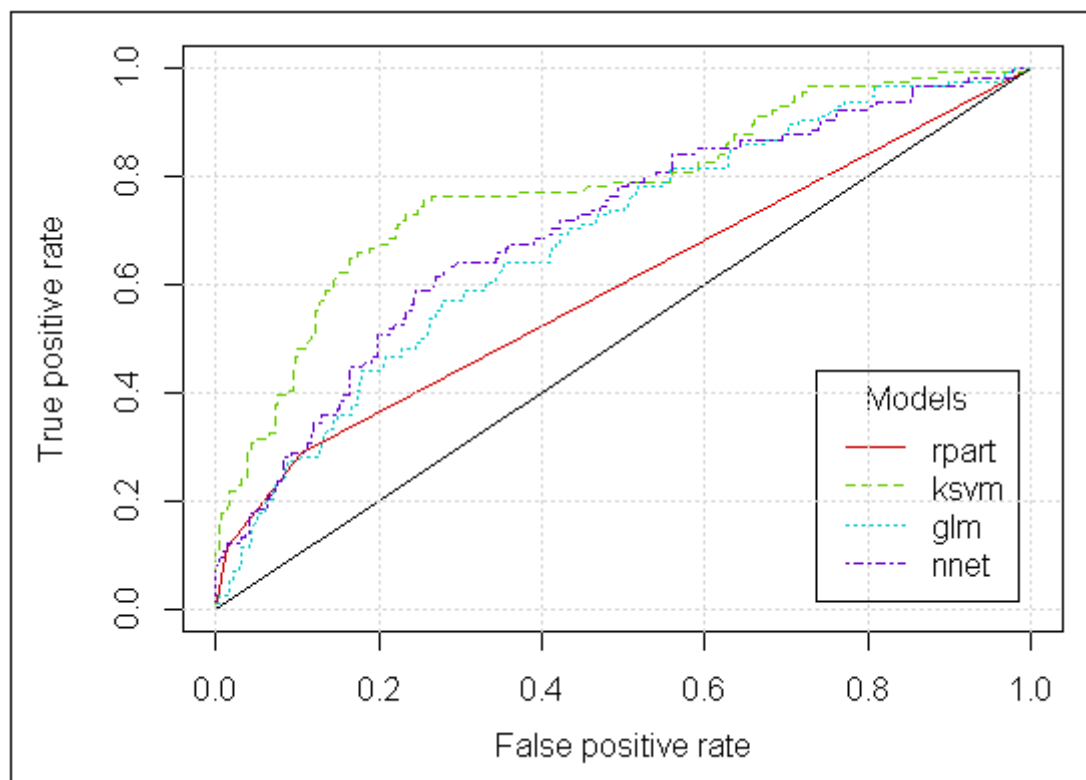
<sup>18</sup> W tym celu wykorzystano pakiet *rattle* programu R.

Krzywe ROC są pomocnym narzędziem w ocenie zdolności predykcyjnej modeli, jednakże do uzyskiwanych wyników nie należy podchodzić bezkrytycznie. Trudno bowiem w analizowanym przypadku mówić o zaakceptowaniu któregośkolwiek z rozważanych klasyfikatorów wobec bardzo niskich trafności prognoz dla  $Y=1$ .

**Tablica 4.** Wartość AUC dla analizowanych modeli

Model	AUC
drzewo klasyfikacyjne (rpart)	0,5969
metoda wektorów nośnych (ksvm)	0,7774
sieć neuronowa (nnet)	0,7037
model logitowy (glm)	0,6835

Źródło: opracowanie własne.



**Rysunek 6.** Krzywe ROC dla analizowanych klasyfikatorów

Źródło: opracowanie własne.

W przypadku stosowania metod *data mining* do problemów klasyfikacji typowym podejściem jest estymacja parametrów modelu na zbiorze treningowym i ocena jakości modelu na danych zbioru testowego, które nie zostały wykorzystane w trakcie estymacji parametrów modelu. Tutaj do oceny jakości modeli i estymacji ich parametrów wykorzystano cały zbiór. Zrezygnowano z dalszych ocen (w szczególności z analizy rezultatów uzyskanych w wyniku podziału danych na zbiór treningowy i testowy) rozważanych klasyfikatorów, które w tym przypadku okazały się z praktycznego punktu widzenia bezwartościowe z uwagi na bardzo duży odsetek przewidywanych „fałszów negatywnych”.

## Podsumowanie

Dokonując oceny jakości klasyfikatorów wykorzystywanych do prób identyfikacji grup ryzyka recydywy istotna jest nie tyle ocena łącznej trafności prognoz<sup>19</sup>, co odsetek „fałszów pozytywnych” i „fałszów negatywnych”. W analizowanym przykładzie łączna trafność prognozowania wybranych klasyfikatorów wynosiła od 73,6% do 75,69% a wartości AUC dla sieci neuronowej i SVM sugerowały, że klasyfikacja jest akceptowalna. Jednakże odsetek prognozowanych „fałszów negatywnych” był dla wszystkich modeli bardzo duży, co uniemożliwiałoby w praktyce wnioskowanie na podstawie takich modeli. Wśród zmiennych objaśniających wykorzystanych w analizowanych modelach znalazły się czynniki statyczne. Stosowanie jako predyktorów jedynie czynników statycznych nie pozwoli w pełni oszacować ryzyka przestępczości powrotnej. Z drugiej strony istnieją problemy z pomiarem czynników dynamicznych, czego przykładem był duży odsetek brakujących wartości dla zmiennych *emp1* – *emp52* określających czy w kolejnych tygodniach po zwolnieniu z zakładu karnego osoby wykonywały pracę. Identyfikacja indywidualnych czynników ryzyka powrotu do przestępczości polega na „ustaleniu u sprawcy obszarów problemowych, które w świetle badań empirycznych są skorelowane z przestępczością<sup>20</sup>”. Należy mieć jednak na uwadze, że brak jest w literaturze jednoznacznych wskazań, które czynniki ryzyka recydywy charakteryzują się najlepszą zdolnością predykcyjną. Ponadto metoda dająca dobre rezultaty dla danych z jednego kraju w innym mogłaby nie być najlepsza (w sensie przyjętych kryteriów oceny jakości klasyfikatorów), gdyż nie opracowano jak dotąd klasyfikatora, który byłby najlepszy dla każdego zbioru danych. Przykładowo, dla danych, które wykorzystali C. Rudin i in. (2015) drzewa klasyfikacyjne charakteryzowały się gorszą jakością klasyfikacji (w sensie wartości AUC) niż regresja logistyczna. Z kolei Stalans L. i in. (2004), na podstawie uzyskanych rezultatów wskazali na lepszą jakość prognozowania recydywy przestępstw przeciwko życiu i zdrowiu w przypadku wykorzystania drzew klasyfikacyjnych niż modeli logitowych. Sytuacja ta może powodować ewentualne problemy adaptacji w danym kraju metod statystycznych stosowanych w innym kraju.

Narzędzia diagnostyczne umożliwiające prognozowanie ryzyka przestępczości powrotnej mogłyby wspomóc proces podejmowania decyzji np. przez kuratorów sądowych, jednakże nie zastąpią one diagnozy klinicznej. Ponadto narzędzi takich nie można sprowadzać jedynie do ilościowych metod zarządzania ryzykiem recydywy, gdyż istotnym elementem jest także m. in. dobór programów resocjalizacyjnych do odpowiednich grup ryzyka<sup>21</sup>.

## Literatura

1. Allen T. D., Bench L. L., *Assessing Sex Offender Recidivism Using Multiple Measures: A Longitudinal Analysis*, „The Prison Journal” 93(4), 2013

<sup>19</sup> Przy ograniczeniu oceny dokładności klasyfikacji do łącznej trafności prognozowania oba rodzaje błędów traktowane są tak samo. Ponadto nie uwzględnia się wtedy niezrównoważenia klas. W analizowanym przykładzie problemem był wysoki odsetek „fałszów negatywnych”. Istnieją również prace, przytaczane przez B. Hołysta (2013), w których autorzy podejmując próby klasyfikacji sprawców do grup ryzyka recydywy, uzyskiwali wysoki odsetek „fałszów pozytywnych”. W takim przypadku również pojawiają się wątpliwości, co do możliwości wykorzystywania danych narzędzi w procesie podejmowania decyzji dotyczących sprawców. W takiej sytuacji problemem nie jest obciążenie ofiar kosztami przestępstw popełnionych przez przyszłych recydywistów, którzy wskutek błędnej decyzji opuścili zakład karny, ale za to pojawia się pytanie: „czy zapobieganie przyszłej szkodzi jest dostatecznym powodem, by zamknąć w więzieniu przestępcę, za to, czego jeszcze nie popełnił?”, por. B. Hołyst, *Podstawy i zakres...*, s. 23.

<sup>20</sup> cyt. B. Stańdo – Kawecka, *Wybrane problemy profesjonalizacji...*, s. 20.

<sup>21</sup> Na problem ten zwraca uwagę m. in. J. Chojecka, *Model dla wszystkich? Spory wokół koncepcji szacowania ryzyka recydywy*, „Resocjalizacja Polska” 2014, nr 7, s. 96.

2. Berk, R. A. i in., *Money, work, and crime: Some experimental results*, Academic Press, New York, 1980
3. Bierens H. J., Carvalho R. J., *A Competing Risk Analysis of Recidivism*, [www.researchgate.net/publication/255581244\\_A\\_Competing\\_Risk\\_Analysis\\_of\\_Recidivism](http://www.researchgate.net/publication/255581244_A_Competing_Risk_Analysis_of_Recidivism), 2002, dostęp: 01.12.2015
4. Błachut J., Gaberle A., Krajewski K., *Kryminologia*, InfoTrade, Gdańsk 2001
5. Chojecka J., *Model dla wszystkich? Spory wokół koncepcji szacowania ryzyka recydywy*, „Resocjalizacja Polska” 2014, nr 7
6. Ehrlich I., *Participation in Illegitimate Activities: A Theoretical and Empirical Investigation*, „The Journal of Political Economy”, 81(3), 1973
7. Gierowski J. K., *Czynniki ryzyka i opiniowanie przemocy u adolescentów*, „Farmakoterapia w Psychiatrii i Neurologii”, 2005 nr 2
8. Hołyst B., *Podstawy i zakres indywidualnej prognozy kryminologicznej*, „Probacja” 2013, nr 1
9. Hudak D., Tinik N., *Examining the Factors Associated with Recidivism*, <https://dspace.rmu.edu/xmlui/handle/11347/25>, 2014, dostęp: 01.12.2015
10. Kądziołka K., *Determinanty przestępczości w Polsce. Aspekt ekonomiczno – społeczny w ujęciu modelowania ekonometrycznego*, niepublikowana rozprawa doktorska, Uniwersytet Ekonomiczny w Katowicach, 2015
11. Korzonek R., *Estymacja funkcji przeżycia – estymator Kaplana – Meiera*, [w:] E. Dziwok i P. Chrzan (red.) *Metody matematyczne, ekonometryczne i komputerowe w finansach i ubezpieczeniach 2008*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2009
12. Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2011
13. Masyn K., Muthen B., *Discrete – Time Survival Mixture Analysis*, „Journal of Educational and Behavioral Statistics”, 30(1), 2005
14. Misztal M., *Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań*, „Taksonomia” nr 23, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 328, 2014
15. Monery B., *The determinants of recidivism among ex-prisoners: a survival analysis on French data*, <https://halshs.archives-ouvertes.fr/halshs-00822847>, 2013, dostęp: 03.10.2015
16. Rudin C. i in., *Interpretable Classification Models for Recidivism Prediction*, [web.mit.edu/rudin/www/ZengUsRu15.pdf](http://web.mit.edu/rudin/www/ZengUsRu15.pdf), 2015, dostęp: 01.12.2015
17. Stalans, L., i in., *Identifying three types of violent offenders and predicting violent recidivism while on probation: A classification tree analysis*, „Law and Human Behavior”, 28(3), 2004
18. Stanisław A., *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe*, Kraków 2007
19. Stańdo – Kawecka B., *Wybrane problemy profesjonalizacji organów probacyjnych i klasyfikacji sprawców oddanych pod dozór do grup ryzyka*, [w:] Nowa Kodyfikacja Prawa Karnego. Tom XXXIII, Wrocław 2014
20. Szczepanik R., *Stawanie się recydywistą. Kariery instytucjonalne osób powracających do przestępczości*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015
21. Szymanowski T., *Recydywa w Polsce: zagadnienia prawa karnego, kryminologii i polityki karnej*, Wolters Kluwer Polska, Warszawa 2010

22. Tollenaar N., van der Heijden P., *Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models*, „Journal of the Royal Statistical Society: Series A (Statistics in Society)”, 176(2), 2013
23. Wójcik D., *Stosowanie w postępowaniu karnym narzędzi diagnostyczno - prognostycznych służących oszacowaniu ryzyka powrotności do przestępstwa*, [w:] M. Mozgawa (red.) *Prawo w Działaniu* 16, Wolters Kluwer, Warszawa 2013
24. Zygmunt J., *Prawne modele zwalczania powrotności do przestępstwa w polskim prawie karnym*, „Czasopismo Prawa Karnego i Nauk Penalnych”, 2008, nr 2
25. Strona internetowa Centralnego Zarządu Służby Więziennej, <http://sw.gov.pl/pl/o-sluzbie-wieziennej/statystyka/statystyka-roczna/>, dostęp: 03.10.2015
26. Informator Statystyczny Wymiaru Sprawiedliwości (ISWS), <http://isws.ms.gov.pl/pl/baza-statystyczna/opracowania-wieloletnie/>, dostęp: 03.10.2015

## **SOCIO-ECONOMIC DETERMINANTS OF RECIDIVISM. SOME PROBLEMS OF IDENTIFICATION RELATIONSHIPS USING QUANTITATIVE METHODS**

### **Summary**

The aim of the author was to discuss an application of data mining and statistical methods to recidivism prediction. There was analysed a binary classification problem where the goal was to predict if a prisoner will be arrested for a certain type of crime within one year of being released from prison. There were compared different models such as neural network, classification tree, logistic regression and SVM. General accuracy of all the models exceeded 70% correctly classified instances, but all of the analysed classifiers were characterized by high “false negatives” ratio and so they would be useless in practice.

**Keywords:** risk of recidivism, survival analysis, logistic regression, data mining

Kinga Kądziołka  
Prokuratura Okręgowa w Katowicach  
e-mail: kinga\_kadziolka@onet.pl