

### **Barbara Pawelek**

Uniwersytet Ekonomiczny w Krakowie  
e-mail: barbara.pawelek@uek.krakow.pl

### **Dorota Grochowina**

Szkoła Główna Handlowa w Warszawie  
e-mail: Dorota.Grochowina@doktorant.sgh.waw.pl

---

## **PODEJŚCIE WIELOMODELOWE W PROGNOZOWANIU ZAGROŻENIA PRZEDSIĘBIORSTW UPADŁOŚCIĄ W POLSCE<sup>1</sup>**

---

## **THE MULTIPLE-MODEL APPROACH IN THE PREDICTION OF COMPANY BANKRUPTCY RISK IN POLAND**

---

DOI: 10.15611/pn.2017.468.17

JEL Classification: C380, C520, G330

**Streszczenie:** Tematyka pracy wpisuje się w nurt badań naukowych nad ważnym problemem ekonomicznym, jakim jest upadłość przedsiębiorstw. W zakresie badania zagrożenia przedsiębiorstw upadłością kontynuowane są prace nad metodologią przewidywania pierwszej i kolejnych upadłości przedsiębiorstw. Celem artykułu jest przedstawienie wyników badań empirycznych nad przydatnością stosowania podejścia wielomodelowego w prognozowaniu zagrożenia upadłością przedsiębiorstw w Polsce. Analizę przeprowadzono dla zbilansowanych i niezbilansowanych zbiorów przedsiębiorstw z sektora przetwórstwa przemysłowego. W badaniu zastosowano metody *bagging*, *boosting*, *random subspaces* i *random forests*. Przydatność podejścia wielomodelowego dla prognozowania upadłości oceniono na podstawie wartości mierników skuteczności klasyfikacyjnej przedsiębiorstw ze zbioru testowego. Wnioski z przeprowadzonych badań empirycznych wskazują na przydatność stosowania podejścia wielomodelowego w prognozowaniu zagrożenia przedsiębiorstw upadłością.

**Słowa kluczowe:** podejście wielomodelowe, upadłość, prognozowanie.

**Summary:** The subject matter of the paper fits into the mainstream of scientific research on an important economic problem, which is the bankruptcy of companies. Within the study of company bankruptcy risk, research is continued on the methodology for predicting the first and subsequent bankruptcies of companies. The aim of the paper is to present the results of empirical research on the usefulness of using the multiple-model approach in the prediction

---

<sup>1</sup> Publikacja została dofinansowana ze środków przyznanych Wydziałowi Zarządzania Uniwersytetu Ekonomicznego w Krakowie w ramach dotacji na utrzymanie potencjału badawczego.

of company bankruptcy risk in Poland. The analysis was performed for balanced and unbalanced sets of enterprises from the industrial processing sector. The bagging, boosting, random subspaces and random forests methods were used in the research. The usefulness of the multiple-model approach for the prediction of bankruptcy was evaluated based on values of the classification efficiency measures for companies from the test set. The conclusions from the empirical research carried out indicate the usefulness of using the multiple-model approach in predicting the company bankruptcy risk.

**Keywords:** multiple-model approach, bankruptcy, prediction.

## 1. Wstęp

Tematyka artykułu wpisuje się w nurt badań naukowych nad ważnym problemem ekonomicznym, jakim jest upadłość przedsiębiorstw. Zagadnienia związane z oceną stopnia zagrożenia przedsiębiorstw upadłością były i są szeroko omawiane w literaturze ekonomicznej (np. [Altman 1968; Ohlson 1980; Frydman i in. 1985; Pociecha i in. 2014]). W zakresie badania zagrożenia przedsiębiorstw upadłością kontynuowane są prace nad metodologią przewidywania pierwszej i kolejnych upadłości przedsiębiorstw (np. [Altman, Branch 2015; Platt, Platt 2002]) oraz oceny kondycji finansowej przedsiębiorstw po ogłoszeniu upadłości (np. [Kostrzewska i in. 2016]). Do najważniejszych problemów metodologicznych występujących w prognozowaniu upadłości przedsiębiorstw można zaliczyć: wybór typu zbioru badawczego (zbilansowany lub niezbilansowany), dobór obiektów do zbioru zbilansowanego (nielosowy lub losowy), wybór metody prognozowania upadłości, dobór zmiennych objaśniających, występowanie braków danych, ocena dokładności predykcji rozważanych metod itp. (np. [Pociecha i in. 2014; Sun i in. 2014]).

Prognozowanie zagrożenia przedsiębiorstw upadłością może być prowadzone na podstawie: jednego modelu (podejście tradycyjne) lub wielu modeli (podejście wielomodelowe). Podejście wielomodelowe polega na łączeniu  $M$  modeli bazowych  $D_1, \dots, D_M$  w jeden model zagregowany  $D^*$  w celu poprawy dokładności prognozowania [Gatnar 2008, s. 62]. Jednym z warunków osiągnięcia wspomnianego celu stosowania podejścia wielomodelowego jest dysponowanie zróżnicowanymi modelami bazowymi. Zwiększenie stopnia zróżnicowania modeli bazowych można uzyskać w wyniku np. [Gatnar 2008, s. 138-169]:

- losowego doboru obserwacji, np. metoda agregacji bootstrapowej (*bagging*), metoda wzmacniania (*boosting*),
- losowego doboru zmiennych, np. metoda losowych podprzestrzeni (*random subspaces*),
- losowego doboru obserwacji i zmiennych, np. metoda lasów losowych (*random forests*).

Próba zastosowania podejścia wielomodelowego jest aktualnym i coraz częściej podejmowanym przedmiotem badań z zakresu prognozowania upadłości przedsię-

biorstw. Najczęściej stosowanymi metodami są metody *bagging* i *boosting* (np. [Min 2016b; Virág, Nyitrai 2014]) oraz *random subspace* (np. [Min 2016a; Li i in. 2011]). Autorzy porównują między sobą metody zaliczane do podejścia wielomodelowego, ale przede wszystkim porównują je z podejściem tradycyjnym.

Celem artykułu jest przedstawienie wyników badań empirycznych nad przydatnością stosowania podejścia wielomodelowego w prognozowaniu zagrożenia przedsiębiorstw upadłością w Polsce.

W pracy przedstawiono wyniki analizy przeprowadzonej dla drzew klasyfikacyjnych. Drzewo klasyfikacyjne (*classification tree*) jest metodą dokonującą rekurencyjnego podziału badanego zbioru na rozłączne podzbiory, aż do uzyskania homogeniczności ze względu na wyróżnioną cechę. W wyniku zastosowania tej metody uzyskuje się reguły decyzyjne pozwalające przyporządkować obiekty do rozważanych podzbiorów (np. [Breiman i in. 1984]).

Obliczenia wykonano w programie R, wykorzystując zmodyfikowane na potrzeby prowadzonych badań empirycznych skrypty autorstwa E. Gatnara [2008].

## 2. Baza danych i procedura badawcza

Podstawę badań stanowił zbiór 7223 przedsiębiorstw działających w sektorze przetwórstwa przemysłowego w Polsce. Pod uwagę wzięto 42 przedsiębiorstwa, w przypadku których sąd ogłosił upadłość w latach 2014 i 2015 (bankruci, B). Dane finansowe dotyczyły lat 2013 i 2014, czyli pochodziły ze sprawozdań finansowych opublikowanych na rok przed ogłoszeniem upadłości. W bazie danych uwzględniono także 7181 przedsiębiorstw „zdrowych”, tj. kontynuujących działalność gospodarczą w latach 2014 i 2015 (nie-bankruci, NB). Dobierając przedsiębiorstwa „zdrowe” do zbioru danych, kierowano się m.in. głównym obszarem działalności oraz dostępnością danych finansowych za lata 2013 i 2014. Dane pobrano ze strony Emerging Markets Information Service. Przedsiębiorstwa, w przypadku których ogłoszono upadłość, stanowiły około 0,6% wszystkich przedsiębiorstw w bazie danych. Zagrożenie przedsiębiorstw upadłością prognozowano z rocznym wyprzedzeniem.

W oparciu o wejściowy zbiór danych utworzono dwa zbiory badawcze:

- wariant I: zbiór zbilansowany: 42 B –  $\frac{1}{2}$  ogółu i 42 NB –  $\frac{1}{2}$  ogółu (ogółem 84 obiekty),
- wariant II: zbiór niezbilansowany: 42 B –  $\frac{1}{4}$  ogółu i 126 NB –  $\frac{3}{4}$  ogółu (ogółem 168 obiektów)<sup>2</sup>.

Doboru przedsiębiorstw „zdrowych” do zbiorów badawczych dokonano w sposób losowy, powtarzając go 10 razy. Każdy z dwudziestu zbiorów badawczych

---

<sup>2</sup> Rozważania zaprezentowane w niniejszej pracy są wprowadzeniem w szersze i głębsze badania nad przydatnością podejścia wielomodelowego w prognozowaniu upadłości przedsiębiorstw. Przyjęto arbitralnie strukturę zbioru niezbilansowanego, proporcje w podziale zbioru na część uczącą i testową oraz liczbę powtórzeń. Dalsze kierunki badań nakreślono w zakończeniu pracy.

(10 zbiorów zbilansowanych i 10 zbiorów niezbilansowanych) został podzielony losowo (po 10 razy) na część uczącą (2/3 ogółu) i część testową (1/3 ogółu).

W badaniu wykorzystano:

- zmienną zero-jedynkową, która przyjmowała kategorię „1” w przypadku przedsiębiorstw, które ogłosiły upadłość w 2014 lub 2015 roku i kategorię „0” dla przedsiębiorstw „zdrowych”,
- 16 wskaźników finansowych podzielonych na grupy: wskaźniki płynności (3 zmienne), wskaźniki zadłużenia (4 zmienne), wskaźniki rentowności (5 zmiennych), wskaźniki sprawności działania (4 zmienne).

Badanie przeprowadzono z wykorzystaniem drzewa klasyfikacyjnego CART. Modele bazowe były budowane na podstawie zbioru uczącego w oparciu o losowo dobrane obserwacje lub losowo dobrane zmienne objaśniające, z wykorzystaniem metod: *bagging* (np. [Breiman 1996]), *boosting* (np. [Freund, Schapire 1997]), *random subspaces* (np. [Ho 1998]) i *random forests* (np. [Breiman 2001]). Na podstawie zbudowanych modeli bazowych przeprowadzono klasyfikację przedsiębiorstw do dwóch grup: jednostek niezagrażonych upadłością za rok (grupa „NB”) i jednostek zagrożonych upadłością za rok (grupa „B”). Agregację modeli bazowych, tj. łączenie wyników prognozowania na podstawie modeli bazowych przeprowadzono zgodnie z metodą głosowania większościowego [Gatnar 2008, s. 63].

Oceny przydatności podejścia wielomodelowego dla prognozowania zagrożenia przedsiębiorstw upadłością w Polsce dokonano na podstawie wartości następujących mierników skuteczności klasyfikacyjnej przedsiębiorstw należących do zbioru testowego:

- błąd ogółem (odsetek przedsiębiorstw, które zostały niepoprawnie zaklasyfikowane do grupy „B” lub grupy „NB”);
- błąd I typu (odsetek bankrutów, którzy zostali zaklasyfikowani do grupy „NB”);
- błąd II typu (odsetek przedsiębiorstw „zdrowych”, które zostały zaklasyfikowane do grupy „B”).

### 3. Wyniki badań empirycznych

Cztery metody rozważane w ramach podejścia wielomodelowego zostały wykorzystane do prognozowania zagrożenia upadłością przedsiębiorstw z sektora przetwórstwa przemysłowego w Polsce. W celu sprawdzenia, czy liczba agregowanych modeli bazowych ma wpływ na skuteczność klasyfikacyjną rozważanych metod w przypadku analizowania zjawiska upadłości przedsiębiorstw, badaniu poddano modele agregatowe zbudowane na podstawie 50 lub 100 modeli bazowych. Dla każdego z 200 zbiorów badawczych (po 10 zbiorów zbilansowanych i niezbilansowanych, które zostały 10 razy podzielone na część uczącą i testową) obliczono odsetek błędnie zaklasyfikowanych obiektów ogółem i – ze względu na charakter badanego zjawiska – w podziale na bankrutów i przedsiębiorstwa „zdrowe”. Następnie dla każdej metody osobno uśredniono uzyskane wyniki, uwzględniając zarówno liczbę

agregowanych modeli bazowych, jak i typ zbioru badawczego (zbilansowany lub niezbilansowany). Podobnie postąpiono z wynikami uzyskanymi w podejściu tradycyjnym, czyli bazującym na pojedynczym drzewie klasyfikacyjnym.

Wyniki otrzymane w podejściu tradycyjnym i podejściu wielomodelowym opartym na agregacji 50 modeli bazowych zamieszczono w tabeli 1, natomiast wyniki uzyskane w rezultacie połączenia 100 modeli bazowych przedstawiono w tabeli 2.

**Tabela 1.** Uśredniony odsetek błędnych klasyfikacji (błąd ogółem, błąd I typu – w grupie bankrutów, błąd II typu – w grupie przedsiębiorstw „zdrowych”) dla metod: pojedyncze drzewo, *bagging*, *boosting*, *random subspaces* i *random forests* przy agregacji 50 modeli bazowych

| Błąd    | Zbiór           | <i>CT</i> * | <i>Bagging</i> * | <i>Boosting</i> * | <i>RSM</i> ** | <i>RSM</i> *** | <i>RF</i> ** | <i>RF</i> *** |
|---------|-----------------|-------------|------------------|-------------------|---------------|----------------|--------------|---------------|
| Ogółem  | zbilansowany    | 0,355       | 0,295            | 0,341             | 0,319         | 0,284          | 0,292        | 0,277         |
|         | niezbilansowany | 0,279       | 0,218            | 0,232             | 0,223         | 0,216          | 0,225        | 0,216         |
| I typu  | zbilansowany    | 0,348       | 0,291            | 0,362             | 0,313         | 0,279          | 0,303        | 0,286         |
|         | niezbilansowany | 0,577       | 0,611            | 0,566             | 0,574         | 0,610          | 0,571        | 0,567         |
| II typu | zbilansowany    | 0,361       | 0,298            | 0,320             | 0,326         | 0,289          | 0,281        | 0,269         |
|         | niezbilansowany | 0,179       | 0,087            | 0,121             | 0,106         | 0,085          | 0,109        | 0,099         |

*CT* – pojedyncze drzewo klasyfikacyjne, *Bagging* – metoda agregacji bootstrapowej, *Boosting* – metoda wzmacniania, *RSM* – metoda losowych podprzestrzeni, *RF* – metoda lasów losowych, \* – wejściowy zbiór zawierał 16 wskaźników finansowych, \*\* – z wejściowego zbioru 16 wskaźników finansowych było losowanych 8 wskaźników, \*\*\* – z wejściowego zbioru 16 wskaźników finansowych były losowane 4 wskaźniki.

Źródło: obliczenia własne.

**Tabela 2.** Uśredniony odsetek błędnych klasyfikacji (błąd ogółem, błąd I typu – w grupie bankrutów, błąd II typu – w grupie przedsiębiorstw „zdrowych”) dla metod: pojedyncze drzewo, *bagging*, *boosting*, *random subspaces* i *random forests* przy agregacji 100 modeli bazowych

| Błąd    | Zbiór           | <i>CT</i> * | <i>Bagging</i> * | <i>Boosting</i> * | <i>RSM</i> ** | <i>RSM</i> *** | <i>RF</i> ** | <i>RF</i> *** |
|---------|-----------------|-------------|------------------|-------------------|---------------|----------------|--------------|---------------|
| Ogółem  | zbilansowany    | 0,355       | 0,294            | 0,328             | 0,298         | 0,263          | 0,273        | 0,255         |
|         | niezbilansowany | 0,279       | 0,211            | 0,223             | 0,219         | 0,212          | 0,219        | 0,214         |
| I typu  | zbilansowany    | 0,348       | 0,290            | 0,344             | 0,296         | 0,261          | 0,275        | 0,263         |
|         | niezbilansowany | 0,577       | 0,583            | 0,540             | 0,570         | 0,619          | 0,562        | 0,564         |
| II typu | zbilansowany    | 0,361       | 0,298            | 0,311             | 0,300         | 0,264          | 0,271        | 0,247         |
|         | niezbilansowany | 0,179       | 0,087            | 0,117             | 0,102         | 0,076          | 0,105        | 0,097         |

Uwagi: jak pod tabelą 1.

Źródło: obliczenia własne.

- W ramach przeprowadzonych analiz zrealizowano pięć zadań badawczych:
- zadanie I: porównanie błędów klasyfikacji uzyskanych w podejściu wielomodelowym w wyniku agregacji 50 modeli bazowych lub 100 modeli bazowych;
  - zadanie II: porównanie błędów klasyfikacji uzyskanych w podejściu jednomodelowym z wynikami otrzymanymi w podejściu wielomodelowym;
  - zadanie III: porównanie błędów klasyfikacji uzyskanych w podejściu wielomodelowym opartym tylko na losowym doborze obserwacji (*bagging* i *boosting*) z wynikami otrzymanymi w podejściu wielomodelowym wykorzystującym losowy dobór zmiennych (*random subspaces* i *random forests*);
  - zadanie IV: porównanie błędów klasyfikacji uzyskanych w podejściu wielomodelowym opartym tylko na losowym doborze obserwacji w wyniku zastosowania metody *bagging* lub metody *boosting*;
  - zadanie V: porównanie błędów klasyfikacji uzyskanych w podejściu wielomodelowym wykorzystującym losowy dobór zmiennych w wyniku zastosowania metody *random subspaces* lub metody *random forests*.

Porównując wyniki zamieszczone w tabelach 1 i 2 (zadanie I), można zauważyć, że wzrost liczby agregowanych modeli bazowych sprzyjał na ogół zmniejszeniu wartości uśrednionego błędu klasyfikacji (33 na 36 przypadków). Można zatem przypuszczać, że – zgodnie z literaturą przedmiotu (np. [Gatnar 2008]) – zwiększanie liczby agregowanych modeli bazowych będzie sprzyjać dalszemu zmniejszeniu błędu klasyfikacji aż do uzyskania stabilizacji na pewnym poziomie. Ustalenie optymalnej liczby łączonych modeli bazowych i określenie poziomu stabilizacji błędu klasyfikacji dla rozważanych metod wymaga dalszych badań empirycznych. Trzeba jednak zaznaczyć, że zwiększanie liczby agregowanych modeli bazowych wiąże się z wydłużaniem czasu potrzebnego do przeprowadzenia obliczeń.

Na podstawie analizy wyników uzyskanych w podejściu tradycyjnym (jednomodelowym) i podejściu wielomodelowym (zadanie II) można stwierdzić, że błąd klasyfikacji był na ogół mniejszy w podejściu wielomodelowym (33 na 36 przypadków przy agregacji 50 modeli bazowych i 34 na 36 przypadków przy agregacji 100 modeli bazowych). Nieliczne odstępstwa od tej prawidłowości występowały w grupie bankrutów i przede wszystkim w przypadku zbioru niezbilansowanego. Największą poprawę skuteczności klasyfikacyjnej ogółem i w grupie przedsiębiorstw „zdrowych” w porównaniu z pojedynczym drzewem (*CT*) uzyskano w wyniku zastosowania metody *random forests* (*RF*) w zbiorze zbilansowanym przy losowaniu 4 z 16 wejściowych wskaźników finansowych i agregacji 100 modeli bazowych. Z kolei, w grupie bankrutów największą poprawę w stosunku do podejścia tradycyjnego odnotowano w wyniku zastosowania metody *random subspaces* (*RSM*) w zbiorze zbilansowanym przy losowaniu 4 z 16 wejściowych wskaźników finansowych i agregacji 100 modeli bazowych.

Na wyniki uzyskane w ramach podejścia wielomodelowego można spojrzeć pod kątem sposobu zwiększania stopnia zróżnicowania modeli bazowych, przyjętego w rozważanych metodach (zadanie III). Można zauważyć przewagę, rozumianą jako

charakteryzowanie się mniejszymi błędami klasyfikacji, podejścia wykorzystującego losowy dobór zmiennych. Przy czym przewaga metod *RSM* i *RF* występowała głównie w stosunku do metody *boosting* (39 na 48 przypadków). Odstępstwa od tej prawidłowości odnotowano głównie w grupie bankrutów w przypadku zbioru niezbilansowanego. Natomiast w porównaniu z metodą *bagging* większa skuteczność klasyfikacyjna metod wykorzystujących losowy dobór zmiennych wystąpiła w 28 na 48 przypadków. Zastosowanie metody *bagging* prowadziło do trafniejszej klasyfikacji głównie w przypadkach, gdy w metodach *RSM* i *RF* dobierano losowo 8 z 16 wejściowych wskaźników finansowych. Przewaga metody *bagging* nad metodami *RSM* i *RF* często występowała w przypadku zbiorów niezbilansowanych (błąd ogółem i w grupie przedsiębiorstw „zdrowych”).

W grupie metod bazujących tylko na losowym doborze obserwacji (zadanie IV) zaobserwowano na ogół większą skuteczność klasyfikacyjną metody *bagging* w stosunku do metody *boosting*. Wyjątek stanowiła grupa bankrutów w zbiorze niezbilansowanym. W tym przypadku występowała przewaga metody *boosting* nad metodą *bagging*.

Zestawiając wyniki uzyskane metodami wykorzystującymi losowy dobór zmiennych i uwzględniając liczbę zmiennych losowanych spośród 16 wejściowych wskaźników finansowych (zadanie V), można stwierdzić, że w przypadku losowania 8 zmiennych ( $\frac{1}{2}$  liczby wejściowych zmiennych) przewagę pod względem skuteczności klasyfikacyjnej miała metoda *RF*. Nieliczne wyjątki od tej prawidłowości miały miejsce w przypadku zbiorów niezbilansowanych. Z kolei, w przypadku losowania 4 zmiennych ( $\frac{1}{4}$  liczby wejściowych zmiennych) większa skuteczność klasyfikacyjna metody *RF* w porównaniu z metodą *RSM* wystąpiła w 6 na 12 przypadków. Porównując wyniki uzyskane dla 4 zmiennych z wynikami otrzymanymi dla 8 zmiennych, można stwierdzić, że w 21 na 24 przypadki metody *RSM* i *RF* charakteryzowały się mniejszymi błędami klasyfikacji przy losowaniu 4 zmiennych.

#### 4. Zakończenie

Zastosowanie podejścia wielomodelowego w prognozowaniu zagrożenia przedsiębiorstw upadłością w Polsce może przyczynić się do poprawy dokładności prognozowania na podstawie drzew klasyfikacyjnych.

Na podstawie wyników przeprowadzonych badań empirycznych nie można wskazać jednej metody wśród rozważanych czterech metod w ramach podejścia wielomodelowego, która charakteryzowałaby się największą skutecznością klasyfikacyjną ogółem oraz w grupie bankrutów i przedsiębiorstw „zdrowych” zarówno w zbiorze zbilansowanym, jak i niezbilansowanym.

W zbiorze zbilansowanym największą skutecznością klasyfikacyjną ogółem i w zbiorze przedsiębiorstw „zdrowych” charakteryzowała się metoda *random forests*, wykorzystująca losowy dobór obserwacji i zmiennych. W przypadku grupy bankrutów najlepsza okazała się metoda *random subspaces*, opierająca się tylko na

losowym doborze zmiennych. Przy czym zastosowanie metody *random forests* prowadziło do klasyfikacji bankrutów obciążonej niewiele większym błędem.

W zbiorze niezbilansowanym metody *bagging* (wykorzystująca tylko losowy dobór obserwacji), *random subspaces* i *random forests* charakteryzowały się skutecznością klasyfikacyjną ogółem na podobnym poziomie. Metoda *random subspaces* dominowała pod względem skuteczności klasyfikacyjnej przedsiębiorstw „zdrowych”. Natomiast zastosowanie metody *boosting*, opierającej się tylko na losowym doborze obserwacji, umożliwiło uzyskanie trafniejszej klasyfikacji bankrutów niż przy użyciu innych z rozważanych metod.

Zwiększenie liczby agregowanych modeli bazowych na ogół poprawia skuteczność klasyfikacyjną rozważanych metod, ale zwiększa czas wykonywania obliczeń.

Po uzyskaniu potwierdzenia empirycznego przydatności podejścia wielomodelowego w prognozowaniu zagrożenia przedsiębiorstw upadłością w Polsce z rocznym wyprzedzeniem, autorki zamierzają kontynuować podjęte badania. Planowana jest m.in. analiza szczegółowa zależności między poziomem błędu klasyfikacji a liczbą agregowanych modeli bazowych dla poszczególnych metod. Celem tej analizy będzie próba ustalenia optymalnej liczby modeli bazowych w podejściu wielomodelowym w przypadku stosowania go w prognozowaniu zagrożenia przedsiębiorstw upadłością w Polsce z rocznym wyprzedzeniem. W dalszych badaniach planowane jest uwzględnienie m.in.: zbiorów niezbilansowanych o innej strukturze bankrutów i przedsiębiorstw „zdrowych” niż w zaprezentowanych rozważaniach, różnych proporcji w podziale zbioru badawczego na część uczącą i testową, większej liczby powtórzeń, różnych liczb agregowanych modeli bazowych. Na kolejnym etapie badań planowane jest powtórzenie rozważań w przypadku prognozowania zagrożenia przedsiębiorstw upadłością w Polsce z dwuletnim wyprzedzeniem.

## Literatura

- Altman E.I., 1968, *Financial ratios, discriminant analysis and prediction of corporate ankrupcty*, The Journal of Finance, vol. 23, no. 4, s. 589-609.
- Altman E.I., Branch B., 2015, *The bankruptcy system's chapter 22 recidivism problem: How serious is it?*, The Financial Review, vol. 50, s. 1-26.
- Breiman L., 1996, *Bagging predictors*, Machine Learning, vol. 24, s. 123-140.
- Breiman L., 2001, *Random Forests*, Machine Learning, vol. 45, s. 5-32.
- Breiman L., Friedman J., Stone C.J., Olshen R.A., 1984, *Classification and Regression Trees*, CRC Press, Boca Raton.
- Freund Y., Schapire R.E., 1997, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences, vol. 55, s. 119-139.
- Frydman H., Altman E.I., Kao D., 1985, *Introducing recursive partitioning for financial classification: The case of financial distress*, Journal of Finance, vol. 40, no. 1, s. 269-291.
- Gatnar E., 2008, *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.



- Ho T.K., 1998, *The random subspace method for constructing decision forests*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, s. 832-844.
- Kostrzewska J., Kostrzewski M., Pawełek B., Gałuszka K., 2016, *The Classical and Bayesian Logistic Regression in the Research on the Financial Standing of Enterprises after Bankruptcy in Poland*, [w:] Papież M., Śmiech S. (red.), *The 10th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków, s. 72-81.
- Li H., Lee Y.-Ch., Zhou Y.-Ch., Sun J., 2011, *The random subspace binary logit (RSBL) model for bankruptcy prediction*, Knowledge-Based Systems, vol. 24, s. 1380-1388.
- Min S.-H., 2016a, *A genetic algorithm-based heterogeneous random subspace ensemble model for bankruptcy prediction*, International Journal of Applied Engineering Research, vol. 11, no. 4, s. 2927-2931.
- Min S.-H., 2016b, *Integrating instance selection and bagging ensemble using a genetic algorithm*, International Journal of Applied Engineering Research, vol. 11, no. 7, s. 5060-5066.
- Ohlson J., 1980, *Financial ratios and the probabilistic prediction of bankruptcy*, Journal of Accounting Research, vol. 18, no. 1, s. 109-131.
- Platt H.D., Platt M.B., 2002, *A re-examination of the effectiveness of the bankruptcy process*, Journal of Business Finance & Accounting, vol. 29, no. 9-10, s. 1209-1237.
- Pociecha J. (red.), Pawełek B., Baryła M., Augustyn S., 2014, *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Sun J., Li H., Huang Q.-H., He K.-Y., 2014, *Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches*, Knowledge-Based Systems, vol. 57, s. 41-56.
- Virág M., Nyitrai T., 2014, *The application of ensemble methods in forecasting bankruptcy*, Financial and Economic Review, vol. 13, no. 4, s. 178-193.