

---

# PROGNOZOWANIE CEN ENERGII ELEKTRYCZNEJ NA RYNKU DNIA NASTĘPNEGO METODAMI DATA MINING

**Autor: Kamil Fijorek, Kinga Mróz, Katarzyna Niedziela, Damian Fijorek**

**(„Rynek Energii” – 12/2010)**

**Słowa kluczowe:** prognozowanie cen energii elektrycznej, rynek dnia następnego, lasy losowe

**Streszczenie.** Celem artykułu jest zbadanie potencjału wybranych, współcześnie zaproponowanych metod data mining w kontekście prognozowania cen energii elektrycznej na rynku dnia następnego. Szczególny nacisk został położony na zbadanie algorytmu lasów losowych. Prognozy uzyskane metodą lasów losowych porównano z prognozami wygenerowanymi przez model regresji liniowej, model regresji medianowej oraz regresyjną metodę wektorów nośnych.

## 1. CEL PRZEPROWADZONYCH BADAŃ

Zasadniczym celem artykułu jest zbadanie potencjału wybranych, współcześnie zaproponowanych metod data mining w kontekście prognozowania cen energii elektrycznej na rynku dnia następnego (RDN). Szczególny nacisk został położony na zbadanie algorytmu lasów losowych (*random forests*) [1]. Decyzja ta jest rezultatem dwóch poczynionych przez autorów obserwacji. Pierwsza – jak wynika z przeprowadzonego przeglądu literatury metoda lasów losowych bardzo rzadko jest wykorzystywana jako narzędzie prognozowania zjawisk ekonomicznych, co więcej, nie stwierdzono, aby była ona dotychczas wykorzystana do prognozowania cen energii elektrycznej. Druga – algorytm lasów losowych, w porównaniu do innych zaawansowanych metod analizy regresji, charakteryzuje się przejrzystą zasadą działania, ponadto działanie algorytmu jest uzależnione od niewielkiej liczby parametrów (szerzej o tym aspekcie w części metodologicznej artykułu).

Cel badawczy zostanie zrealizowany poprzez porównanie jakości prognoz uzyskiwanych metodą lasów losowych z metodami klasycznymi, takimi jak regresja liniowa, czy regresja medianowa [6]. Dodatkowo w opracowaniu uwzględniono zdobywającą uznanie wśród praktyków metodę regresyjnych wektorów nośnych (SVR – *Support Vector Regression*) [8].

## 2. DANE

Część empiryczna niniejszego artykułu opiera się na danych dotyczących średnioważonych cen jednej megawatogodziny energii elektrycznej dla każdej godziny doby handlowej na rynku dnia następnego podawanych przez Towarową Giełdę Energii (TGE) [10]. Zgromadzone dane obejmują okres od 01.01.2008 do 31.12.2008 (jest to zbiór uczący służący oszacowaniu modeli prognostycznych) oraz okres dwóch wybranych miesięcy roku 2009 (maj, listopad - jest to zbiór testowy służący ocenie zdolności prognostycznych oszacowanych modeli).

W każdym z modeli prognostycznych zmienną objaśnianą jest cena 1 megawatogodziny energii. Kształtowanie się jej poziomu jest objaśniane przy pomocy wartości cen oraz obrotów sprzed 48, 72, ..., 168 godzin. Dalej uwzględniono szereg zmiennych sztucznych, tzn. zmienne określające godzinę doby, dzień tygodnia, miesiąc oraz zmienne opisujące stan pogody dla każdej godziny doby dnia następnego. W związku z trudnościami w pozyskaniu faktycznych prognoz pogody postanowiono w zamian wykorzystać dane o rzeczywiście zaobserwowanych stanach pogody, symulując w ten sposób prognozę

---

pogody (podobnie [9]). Dane pogodowe (godzinowa temperatura oraz prędkość wiatru) zostały pobrane ze strony internetowej „National Severe Storms Laboratory Historical Weather Data Archives” [11]. Stan pogody jest reprezentowany przez temperaturę oraz prędkość wiatru zaobserwowane w ośmiu polskich miastach (Gdańsk, Szczecin, Poznań, Wrocław, Warszawa, Katowice, Kraków, Rzeszów) w każdej godzinie doby. Ze względu na znaczną liczbę zmiennych pogodowych postanowiono dokonać redukcji ich liczby metodą odpornych (na obserwacje odstające) składowych głównych (RPCA – *Robust Principal Component Analysis* [3, 4]). Zarówno dla temperatury jak i prędkości wiatru pozostawiono pierwsze 2 składowe główne. Zachowują one 98% informacji tkwiącej w zbiorze 8 temperatur oraz 87% informacji zawartej w zbiorze 8 prędkości wiatru. Jak wynika z eksperymentów przeprowadzonych przez autorów zabieg ten spowodował niewielkie polepszenie się zdolności prognostycznych rozważanych modeli. Taki efekt jest często opisywany w literaturze i uzasadnia się go eliminacją szumu zawartego w danych oryginalnych.

### 3. WPROWADZENIE DO LASÓW LOSOWYCH

Podstawowym elementem budującym algorytm lasów losowych jest regresyjne drzewo decyzyjne. Tym samym zasadne jest rozpoczęcie opisu metody lasów losowych od tej właśnie techniki.

Niech zbiór danych (próbka ucząca) składa się z  $n$  obserwacji na  $p$  zmiennych objaśniających oraz jednej zmiennej objaśnianej. Algorytm regresyjnych drzew decyzyjnych w pierwszym kroku określa regułę podziału zbioru  $n$  obserwacji na dwa (niekoniecznie równoliczne) podzbiory. Dla ciągłych (ilościowych) zmiennych objaśniających oznacza to wybranie tej z  $p$  zmiennych (wraz z odpowiednim punktem odcięcia), która prowadzi do podziału (w pewnym sensie) najlepszego, tzn. podziału minimalizującego wariancję zmiennej objaśnianej w podzbiorach. W następnym kroku rozumowanie przedstawione powyżej jest powtarzane w wygenerowanych podzbiorach, aż do spełnienia warunku zatrzymania procedury.

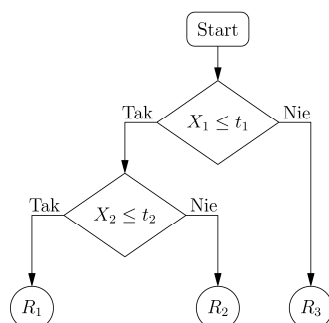
Omówiony pokrótce mechanizm działania metody zostanie teraz poparty prostym przykładem. Przedstawione na rysunku 1 regresyjne drzewo decyzyjne prezentuje sytuację, w której próbka ucząca została podzielona na trzy rozłączne podzbiory (obszary) oznaczone  $R_1$ ,  $R_2$  oraz  $R_3$ . Na drzewo decyzyjne składają się dwie reguły decyzyjne skonstruowane w oparciu o zmienne  $X_1$  oraz  $X_2$ . Pierwsza reguła decyzyjna bada czy wartość zmiennej  $X_1$  jest mniejsza lub równa od pewnej stałej  $t_1$ . W przypadku, gdy ten warunek nie jest spełniony należy podążać w prawo (jest to przyjęta konwencja). Ta reguła decyzyjna definiuje obszar  $R_3$  (znajdują się tam wszystkie obserwacje spełniające regułę decyzyjną). Każdej obserwacji w obszarze  $R_3$  zostanie przypisana prognoza równa średniej wartości zmiennej objaśnianej w tym obszarze. Analogicznej interpretacji można dokonać dla obszarów  $R_1$  oraz  $R_2$ .

Dodatkowych wyjaśnień wymaga wspomniany powyżej warunek zatrzymania procedury kolejnych podziałów zbioru uczącego. Powszechnie stosowaną praktyką jest (celowe) dokonywanie na tyle dużej liczby podziałów, aby w obszarach końcowych (liściach drzewa) zostało co najwyżej kilka obserwacji. Następnie wykonywane jest tzw. „przycinanie drzewa”, które oznacza usuwanie z drzewa tych podziałów, które w małym stopniu pogarszają optymalność drzewa. Jest to postępowanie lepsze od bieżącej kontroli optymalności drzewa, gdyż często okazuje się, że dany podział nie wydaje się być dobrym, lecz podział następny takim już jest.

Regresyjne drzewa decyzyjne w swojej podstawowej formie charakteryzują się szeregiem zalet, ale również istotnymi wadami. Do najważniejszych zalet należy łatwość interpretacji drzewa decyzyjnego oraz możliwość graficznej prezentacji zasady jego działania. Inną zaletą drzew decyzyjnych jest to, że

---

radzą sobie one z problemem braków danych poprzez tworzenie dla każdej reguły podziału reguły zastępczej (drugi w kolejności najlepszy podział), która jest stosowana w miejsce reguły najlepszej w przypadku napotkania wspomnianego braku danych. Mechanizm ten można uogólnić tworząc kilka podziałów zastępczych.



Rys. 1. Przykładowe regresyjne drzewo decyzyjne

Zasadniczą wadą regresyjnych drzew decyzyjnych jest ich duża niestabilność. Ogólnie mówiąc oznacza to, że nawet stosunkowo niewielka zmiana zbioru uczącego, np. polegająca na dodaniu lub usunięciu kilku obserwacji, może prowadzić w konsekwencji do całkowicie innej sekwencji podziałów. W świetle tej obserwacji zaprezentowana powyżej zaleta w postaci łatwej interpretacji może zostać poddana w wątpliwość. Drugą wadą jest to, że generowane przez drzewa decyzyjne prognozy są natury skokowej. Innymi słowy płynnej zmianie wartości zmiennej objaśniającej będą towarzyszyć skokowe zmiany prognozowanych wartości.

### 3.1. Algorytm lasów losowych

Algorytm lasów losowych jest techniką rozwiązującą zaprezentowane wcześniej problemy związane z regresyjnymi drzewami decyzyjnymi. Ogólna idea lasów losowych polega na stworzeniu wielu różnych drzew regresyjnych, stąd w nazwie słowo „las”. W ten sposób ograniczone zostaje zjawisko niestabilności oraz braku ciągłości prognoz. Natomiast słowo „losowe” bierze swój początek w tym, że każde drzewo jest tworzone na losowej próbie  $n$  obserwacji pobieranych ze zwracaniem ze zbioru uczącego (próba bootstrapowa). Drugi element losowości polega na tym, że w przeciwieństwie do regresyjnych drzew losowych, gdzie każdy kolejny podział jest wyłaniany na podstawie wszystkich zmiennych objaśniających, w algorytmie lasów losowych najlepszy podział jest określany w oparciu o próbę losową zmiennych. Standardowym zaleceniem dla problemów regresyjnych jest losowanie  $p/3$  zmiennych. Istotna jest również uwaga, że każde z drzew składowych lasu losowego nie podlega przycinaniu. Prognozowanie na podstawie modelu lasu losowego polega na określeniu prognoz dla każdego drzewa wchodzącego w skład lasu oraz wyznaczeniu średniej arytmetycznej tych prognoz indywidualnych jako prognozy całego modelu.

Najważniejszymi zaletami lasów losowych jest ich empirycznie wykazana efektywność [1] oraz mała liczba parametrów sterujących ich działaniem. Analityk musi jedynie zdecydować ile drzew ma wchodzić w skład lasu oraz ile powinna wynosić minimalna liczba obserwacji w każdym liściu drzewa. Jest to istotna zaleta w porównaniu do sieci neuronowych czy regresyjnej metody wektorów nośnych, gdzie liczba parametrów (a tym samym liczba możliwych wariantów ich ustawień) jest duża. Optymalne wartości opisanych powyżej parametrów metody lasów losowych mogą być określone na zbiorze uczącym z wykorzystaniem techniki sprawdzianów krzyżowych, o których szerzej w książce [7].

---

Potencjalnie algorytm lasów losowych może się nie sprawdzić w przypadku dużej liczby zmiennych objaśniających, z których tylko niewielka liczba jest znacząca dla prognozowania. Sytuacja ta powoduje, że w momencie, gdy następuje losowanie ich podzbioru może się okazać, że nie została wylosowana żadna ważna zmienna objaśniająca. W rozważanym w artykule przypadku zmiennych objaśniających jest niewiele toteż zjawisko to nie stanowi realnego ryzyka.

Inną interesującą obserwacją jest ta płynąca z eksperymentów przeprowadzonych przez [2], wskazująca, że nawet znaczna liczba drzew oraz niska minimalna liczba obserwacji w liściach drzew nie prowadzi do dużego pogorszenia się zdolności prognostycznych lasów losowych. Stwierdzenie to łagodzi obawy związane z konsekwencjami stosowania nieoptymalizowanych wartości parametrów sterujących.

Kolejną pozytywną cechą lasów losowych jest wbudowany mechanizm określania ważności poszczególnych zmiennych objaśniających umożliwiający stworzenie wykresu ważności zmiennych.

## **4. WYNIKI BADAŃ EMPIRYCZNYCH**

### **4.1. Metody oceny jakości prognoz**

Ocena zdolności prognostycznych modeli została wykonana na testowym zbiorze danych, czyli na danych, które nie uczestniczyły w procesie wyznaczania modeli. Zabieg ten jest niezwykle ważny, aby uzyskać wiarygodne oszacowanie jakości generowanych prognoz. Ponadto etap ten nabiera dodatkowego znaczenia w kontekście obiektywnej oceny stosowanych algorytmów, które mogą mieć tendencję do przeuczania się. Zjawisko to występuje, gdy model odznacza się bardzo dobrym dopasowaniem do danych ze zbioru uczącego, ale bardzo słabym do danych ze zbioru testowego.

W literaturze przedmiotu można odnaleźć niezmierną mnogość miar dobroci dopasowania. Z praktycznego punktu widzenia nie jest wygodne raportowanie wielu wyników, gdyż automatycznie pojawiają się problemy z interpretacją różnorodnych miar oraz zasadnicze pytanie, o to która z nich jest najlepsza. W celu uniknięcia tego stanu autorzy zdecydowali się podawać jedynie trzy miary dobroci dopasowania, tzn. miary najbardziej popularne – odchylenie standardowe składnika resztowego ( $S_e$ ), średni bezwzględny błąd procentowy – Mean Absolute Percentage Error (MAPE) oraz miarę dopasowania odporną na obserwacje odstające – Median Absolute Deviation (MAD, mediana bezwzględnych odchyleń). W przypadku prognozowania cen energii elektrycznej odporność MAD jest cechą niezwykle pożądaną, gdyż w badanym materiale liczbowym występowały nadspodziewanie wysokie (w porównaniu do wartości przeciętnej) poziomy cen.

### **4.2. Ocena jakości prognoz**

W tabeli 1 przedstawiono wartości odchylenia standardowego składnika resztowego, średniego bezwzględnego błędu procentowego oraz mediany bezwzględnych odchyleń dla poszczególnych modeli obliczone na zbiorze uczącym oraz zbiorze testowym. W przypadku zbioru uczącego, biorąc pod uwagę  $S_e$ , najlepszych prognoz dostarczają lasy losowe (25,2 PLN), a najgorszych model regresji medianowej (36,2 PLN). W przypadku rankingu modeli względem MAD najlepszy okazał się model regresyjnych wektorów nośnych (6,5 PLN), a najgorszy model regresji liniowej (12,1 PLN). Na zbiorze testowym różnice pomiędzy konkurującymi modelami stały się znacznie mniejsze, jednak nadal (pod względem  $S_e$ ) model lasów losowych dostarcza najlepszych prognoz. Pewną niespodzianką jest to, że w oparciu o MAD najlepszy okazał się model regresji medianowej, dodatkowo model ten nieznacznie tylko odbiega od lasów losowych pod względem  $S_e$ .

W tabeli 1 umieszczono również przybliżony czas pracy komputera potrzebny do wyznaczenia poszczególnych modeli. W przypadku prostych modeli regresyjnych (regresja liniowa, medianowa) czasy te są zanedbywalnie małe. Jednak w przypadku lasów losowych oraz SVR są one znacznie większe. Mogłoby się wydawać, że nadal są one obiektywnie małe. Jednak w ocenie tego zjawiska należy wziąć pod uwagę bardzo ważny czynnik. Otóż te dwa modele (jak większość zaawansowanych metod data mining) są uzależnione od wartości parametrów sterujących. W przypadku lasów losowych, tych parametrów jest mało więc wyszukanie ich optymalnych wartości nie nastręcza znacznych trudności. Inna sytuacja występuje w przypadku metody SVR, tutaj parametrów sterujących algorytmem jest znacznie więcej. Powoduje to, że przestrzeń, którą trzeba przeszukać w celu odnalezienia ich najlepszych wartości jest bardzo duża (niejednokrotnie oznacza to konieczność zbadania setek modeli). W rezultacie czas potrzebny do znalezienia optymalnego modelu SVR jest znaczną wielokrotnością czasu przedstawionego w tabeli 1. Ponadto w praktyce badacz może chcieć oszacować opisane modele na danych pochodzących z długiego okresu czasu. Tutaj należy mieć na uwadze, że czas obliczeń większości zaawansowanych metod data mining rośnie w stopniu więcej niż proporcjonalnym (nierzadko w tempie kwadratowym) wraz ze wzrostem liczby obserwacji i zmiennych. Ważna staje się więc możliwość wyznaczania modeli w oparciu o obliczenia równoległe, np. przy wykorzystaniu coraz bardziej popularnych procesorów wielordzeniowych. W tym kontekście ujawnia się kolejna zaleta lasów losowych, które umożliwiają stosunkowo łatwą implementację na procesorach wielordzeniowych (fakt ten empirycznie zweryfikowali autorzy w środowisku obliczeń statystycznych R).

**Tabela 1** Jakość prognoz generowanych przez poszczególne modele prognostyczne

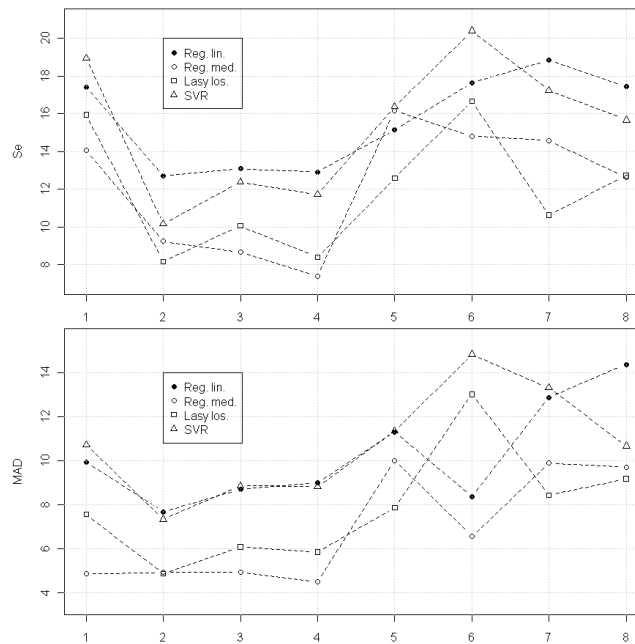
Model	Zbiór uczący			Zbiór testowy			Czas obliczeń, sek.
	S <sub>e</sub>	MAD	MAPE	S <sub>e</sub>	MAD	MAPE	
<b>Regresja liniowa</b>	34,5	12,1	10,9%	15,8	10,0	8,1%	~ 1
<b>Regresja medianowa</b>	36,2	9,2	9,9%	12,3	6,5	5,6%	~ 6
<b>Lasy losowe<sup>1</sup></b>	25,2	7,4	7,2%	12,1	7,5	5,9%	~ 300
<b>SVR<sup>2</sup></b>	27,8	6,5	7,0%	15,6	10,4	7,9%	~ 70

<sup>1</sup> Parametry modelu: liczba drzew = 500, minimalna liczba przypadków w liściu drzewa = 30

<sup>2</sup> Parametry modelu:  $v = 0,5$ ,  $C = 0,1$ , funkcja jądra typu RBF,  $\gamma = 0,04$

Rysunek 2 przedstawia graficzną prezentację jakości prognoz (mierzonych przy pomocy odpowiednio S<sub>e</sub> i MAD) dla kolejnych tygodni okresu testowego. W obu przypadkach klasyczna regresja liniowa oraz metoda regresyjnych wektorów nośnych dostarczają najgorszych prognoz, natomiast lasy losowe oraz regresja medianowa dostarczają prognoz porównywalnej jakości.

W celu zbadania stabilności uzyskanych wyników omawiane modele oszacowano również na danych z roku 2009 oraz początku roku 2010. Generalne wnioski płynące z tego doświadczenia można sformułować w następujący sposób: model regresji liniowej ponownie charakteryzował się najgorszymi prognozami, jakość prognoz generowanych przez SVR była wyraźnie lepsza, a tych generowanych przez regresję medianową wyraźnie gorsza. Jednak najistotniejsza jest uwaga, że lasy losowe dostarczyły najlepszych prognoz zarówno pod względem S<sub>e</sub>, jak i MAD.

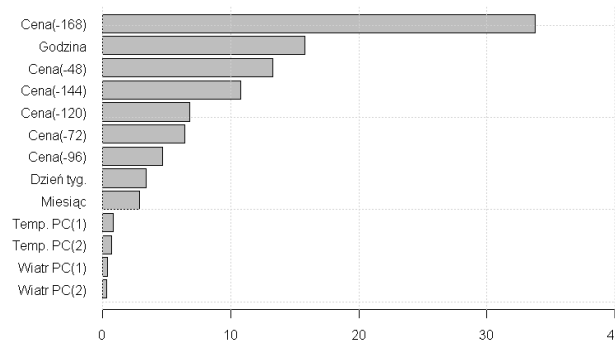


Rys. 2.  $S_e$  oraz MAD dla poszczególnych modeli w kolejnych tygodniach okresu testowego

### 4.3. Ocena mocy prognostycznej zmiennych objaśniających

Zdolności prognostyczne badanych modeli są niewątpliwie najważniejszym czynnikiem decydującym o ich późniejszym zastosowaniu w praktyce. Często jednak w rzeczywistości gospodarczej zdarza się, że odbiorca modelu żąda opisanego sposobu, w jaki poszczególne zmienne uwzględnione w modelu wpływają na kształtowanie się zmiennej prognozowanej. Wynika to z chęci chociaż częściowego poznania reguł, jakimi kieruje się model prognostyczny w generowaniu prognoz, odbiorca pragnie ocenić, czy te reguły są zgodne z oczekiwaniami, doświadczeniem, wiedzą ekspercką. Jeżeli model jest z nimi zgodny, powoduje to pewne uwiarygodnienie uzyskiwanych prognoz. Z drugiej strony, gdy wykryte zostanie pewne zaskakujące zachowanie się modelu, może to poddać w wątpliwość jego zasadność lub może rzucić cień podejrzeń na jakość danych wejściowych. Alternatywnym rezultatem zauważenia takiej anomalii może być odkrycie pewnej nowości, z której istnienia ekspert nie zdawał sobie sprawy, prowadząc tym samym do lepszego poznania mechanizmu generującego prognozowaną zmienną.

W przypadku prostych modeli regresji względnie łatwo można określić kierunek i siłę oddziaływania poszczególnych zmiennych objaśniających na zmienną objaśnianą, lecz zazwyczaj odbywa się to kosztem gorszej jakości prognoz. Diametralnie inna sytuacja występuje w przypadku zaawansowanych modeli data mining. Najczęściej traktowane są one jako „czarne skrzynki”, których mechanizm działania jest trudny do precyzyjnego określenia. Taka ich natura wynika głównie z uwzględniania nieliniowych wpływów zmiennych objaśniających oraz złożonych interakcji pomiędzy zmiennymi (charakter oddziaływania danej zmiennej jest uzależniony od poziomu innych zmiennych w modelu). W tym kontekście dobrą wiadomością jest dostępność w przypadku lasów losowych, tzw. wykresów ważności zmiennych. Wykres ten określa moc wyjaśniającą poszczególnych zmiennych, którą w uproszczeniu można rozumieć jako stopień przyrostu błędu prognozy, gdyby daną zmienną wyłączyć z modelu.



**Rys. 3.** Wykres ważności zmiennych w modelu lasów losowych

Na rysunku 3 przedstawiono wykres ważności zmiennych włączonych do modelu lasów losowych. Z rysunku wynika, że najważniejsze są zmienne związane z historycznymi cenami energii elektrycznej oraz zmienne określające godzinę doby, miesiąc, dzień tygodnia. Znaczenie zmiennych informujących o stanie pogody jest bardzo małe.

## 5. PODSUMOWANIE

W artykule w sposób empiryczny wykazano, że stosowanie metod data mining w celu prognozowania cen energii elektrycznej może prowadzić do uzyskania prognoz obciążonych mniejszymi błędami w porównaniu do metod klasycznych. Na podstawie wyników badań można sformułować następujące uwagi:

- bez wyraźnego uzasadnienia model regresji liniowej nie powinien być stosowany,
- modele regresji medianowej oraz regresyjnych wektorów w szczególnych przypadkach dają prognozy wysokiej jakości,
- model lasów losowych dostarczył stabilnych i dobrych prognoz na dwóch niezależnych zbiorach danych w związku z czym można go wskazać jako użyteczne narzędzie wspomagania decyzji.

## LITERATURA

- [1] Breiman L.: Random forests. *Machine learning* 45, Springer, 2001.
- [2] Hastie T., Tibshirani R., Friedman J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.
- [3] Hubert M., Rousseeuw P.J., Vanden Branden K.: ROBPCA: a new approach to robust principal components analysis. *Technometrics* 47, 2005.
- [4] Jolliffe I.T.: *Principal component analysis*. Springer, 2002.
- [5] Kamińska-Chuchmała A., Wilczyński A.: Zastosowanie metod symulacyjnych do przestrzennego prognozowania obciążeń elektrycznych. *Rynek Energii* 2008, nr 1.
- [6] Koenker R., Hallock K.F.: Quantile regression. *Journal of Economic Perspectives*, 2001.
- [7] Koronacki J., Ćwik J.: *Statystyczne systemy uczące się*. WNT, 2005.
- [8] Smola A.J., Schölkopf B.: *A tutorial on support vector regression*. *Statistics and Computing*, Springer, 2004.
- [9] Weron R., Misiorek A.: Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, Elsevier, 2008.
- [10] <http://www.tge.pl/>
- [11] <http://data.nssl.noaa.gov>

---

## **FORECASTING HOURLY ELECTRICITY PRICES IN POLISH DAY-AHEAD MARKET: DATA MINING APPROACH**

**Key words:** electricity prices forecasting, day-ahead market, random forests

**Summary.** The aim of the paper is to study capabilities of selected, recently introduced supervised data mining algorithms in context of day-ahead electricity prices forecasting. Special stress was put on studying random forests. The quality of random forests forecasts was compared to forecasts obtained from multiple regression, median regression and support vector regression.

**Kamil Fijorek**, asystent w Katedrze Statystyki na Wydziale Zarządzania Uniwersytetu Ekonomicznego w Krakowie. Zainteresowania naukowe dotyczą zastosowań metod data mining (szczególnie efektywne metody analizy dużych zbiorów danych), biostatystyki (m.in. analiza czasu trwania, analiza mikromacierzy DNA, metody eksperymentu), odpornych metod statystycznych. E-mail: [kamil.fijorek@uek.krakow.pl](mailto:kamil.fijorek@uek.krakow.pl)