

mowania ekspresji wrodzonych instynktów i popędów [3]. Znajomość behawioru zwierząt pozwala na przystosowanie ich do życia w środowisku sztucznym, stworzonym przez człowieka. Dlatego też specyficzne zachowania zwierząt powinny być uwzględniane przy wprowadzaniu nowych technologii w hodowli czy też przy opracowywaniu optymalnych czynników środowiskowych, które będą determinowały wysoki poziom dobrostanu zwierząt. „Nowości” te przed wprowadzeniem do praktyki powinny być starannie i sumiennie kontrolowane w fazie eksperymentu pod kątem dobrostanu zwierząt (monitorowanie stanu zdrowia i behawioru).

Przepisy prawne związane z dobrostanem zwierząt

Dobrostan zwierząt zawiera się w obszarze C wymogów wzajemnej zgodności, które obowiązują w Polsce od 1 stycznia 2013 r. Za nieprzestrzeżenie prawa grożą rolnikom kary w postaci zmniejszonych dopłat bezpośrednich. W skrajnych przypadkach ARiMR może nie wypłacić dopłaty. Akty prawne związane z dobrostanem zwierząt to:

– Ustawa z dnia 21 sierpnia 1997 r. o ochronie zwierząt (Dz.U. z 2003 r. nr 106, poz. 1002, z późn. zm.);

– Ustawa z dnia 11 marca 2004 r. o ochronie zdrowia zwierząt oraz zwalczaniu chorób zakaźnych zwierząt (Dz.U. z 2008 r. nr 213, poz. 1342, z późn. zm.);

– Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 28 kwietnia 2004 r. w sprawie zakresu i sposobu prowadzenia ewidencji leczenia zwierząt i dokumentacji lekarsko-weterynaryjnej (Dz.U. nr 100, poz. 1022);

– Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 17 grudnia 2009 r. w sprawie sposobu ustalania poziomu obsady kurcząt brojlerów w kurniku, w którym są one utrzymywane (Dz.U. nr 223, poz. 1784);

– Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 15 lutego 2010 r. w sprawie wymagań i sposobu postępowania przy utrzymywaniu gatunków zwierząt gospodarskich, dla których normy ochrony zostały określone w przepisach Unii Europejskiej (Dz.U. nr 56, poz. 344 z późn. zm.);

– Rozporządzenie Ministra Rolnictwa i Rozwoju Wsi z dnia 28 czerwca 2010 r. w sprawie minimalnych warunków utrzymania

gatunków zwierząt gospodarskich innych niż te, dla których normy ochrony zostały określone w przepisach Unii Europejskiej (Dz.U. nr 116, poz. 778).

Literatura: 1. Appleby M.C., Hughes B.O., 1991 – Word's Poul. Sci. 47, 109-128. 2. Arey D.S., 1993 – Farm Buil. Prog. 114, 18-20. 3. Arnone M., Dantzer R., 1980 – Appl. Anim. Ethol. 6, 351-362. 4. Blokhuis H.J., 1989 – Appl. Anim. Behav. Sci. 22, 65-73. 5. Broom D.M., 1986 – British Vet. J. 142, 524-526. 6. Broom D.M., 1997 – Appl. Anim. Behav. Sci. 54, 21-23. 7. Broom D.M., 1997 – Proc. 9th Intern. Congr. Anim. Hyg., Helsinki, 17-21 August 1997, 371-378. 8. Cronin G.M., Wiepkema P.R., 1984 – Ann. Rech. Vet. 15, 263-270. 9. Dantzer R., 1986 – J. Anim. Sci. 62, 1776-1786. 10. Dantzer R., 1991 – Behav. Processes 25, 95-102. 11. Dantzer R., Mormède P., 1983 – Appl. Anim. Ethol. 10, 233-244. 12. Dawkins M.S., 1983 – Appl. Anim. Behav. Sci. 31, 1195-1209. 13. Dunacan I.J.H., 1978 – Appl. Anim. Ethol. 4, 197-200. 14. Empel W., 1996 – Życie Wet. 71, 65-67. 15. Empel W., 1999 – Życie Wet. 74, 300-301. 16. Fraser A.F., Broom D.M., 1990 – Farm animal behaviour and welfare. Bailliere Tindall, London. 17. Gentle M.J., Waddington D., Hunter L.N., 1990 – Appl. Anim. Behav. Sci. 27, 149-157. 18. Hughes B.O., 1988 – Vet. Res. 123, 33-36. 19. Hughes B.O., Duncan I.J.H., 1988 – Appl. Anim. Behav. Sci. 36, 1696-1707. 20. Janiszewska J., 1999 – Mat. Konf. Nauk. nt. Ochrona zwierząt w świetle prawa i norm etycznych. AR Szczecin, 26-27 marca 1999, 11-14. 21. Jezierski T., Kopowski J., 1997 – Przeg. Hod. 8, 49-53. 22. Kołacz R., 2000 – Życie Wet. 75, 377-380. 23. Kołacz R., 2003 – Życie Wet. 78, 574-577. 24. Kołacz R., 2006 – Przeg. Hod. 9, 8-11. 25. Kołacz R., Bodak E., 1999 – Med. Weter. 55, 147-154. 26. Kowalski A., 2005 – Med. Weter. 61, 1335-1339. 27. Krawczyk J., Wężyk S., 2002 – Polskie Drobniarstwo 8, 24-26. 28. Low M., 2003 – Aust. Vet. J. 81, 192-198. 29. Mal M.E., Friend T.H., Lay D.C., Vogelsang S.G Jenkins O.C., 1991 – Appl. Anim. Behav. Sci. 31, 13-24. 30. Manteca X., 1994 – Bul. Vet. Clin. Ethol. 2, 22-26. 31. Mroczkowski S., Mroczkowska A., 2009 – Przeg. Hod. 9, 29-32. 32. Nowicki B., Zwolińska-Bartczak I., 1993 – Zachowanie się zwierząt gospodarskich. PWRiL Warszawa. 33. Ransom J.I., Cade B.S., 2009 – Quantifying equid behaviour – A research ethogram for free-roaming feral horses. U.S. Geological Survey, Reston, Virginia. 34. Sadowski B., 2001 – Biologiczne mechanizmy zachowania się ludzi i zwierząt. PWN Warszawa. 35. Tinbergen N., 1963 – Z. Tierpsychol. Beih. 20, 410-433. 36. Wechsler B., 1995 – Appl. Anim. Behav. Sci. 43, 123-134. 37. Węgrzynowicz R., 1999 – Mat. Konf. Nauk. nt. Ochrona zwierząt w świetle prawa i norm etycznych. AR Szczecin, 26-27 marca 1999, 7-10. 38. Wiepkema P.R., 1987 – J. Anim. Sci. 65, 1220-1227.

Animal welfare in terms of evaluation criteria and methods

Summary

The objective of the work was to present animal welfare in terms of evaluation criteria and methods. It is difficult to point to a basic criterion of evaluating welfare which is easy to apply. As a result, welfare is assessed by means of many indicators including behavioural, physiological, health and production indicators in addition to supplementary parameters (technical and technological housing parameters and zoohygienic conditions). The more indicators are taken into account, the more precise and objective welfare evaluation is. The most important criterion in animal welfare evaluation is animal behaviour. In practice, the above-mentioned indicators are called objective (clinical and laboratory diagnosis, behavioural and zoohygienic methods, statistical analyses) and subjective methods (current observations of animals). Criteria and methods applied to evaluate animal welfare make it possible to select optimal solutions for animals which assure the highest level of welfare.

KEY WORDS: animal welfare, criteria, methods

Zastosowanie technik *data mining* w zootechnice

Dariusz Piwczyński, Magdalena Kolenda,
Beata Sitkowska, Alicja Czajkowska

Uniwersytet Technologiczno-Przyrodniczy w Bydgoszczy

Dynamiczny rozwój technik komputerowych i metod statystycznych doprowadził w ostatnich latach do powstania nowej dziedzi-

ny analizy danych o nazwie *data mining* (DM), wywodzącej się od wielowymiarowych metod statystycznych i uczenia maszynowego [1, 2, 9, 10, 11, 12]. Ich celem jest poszukiwanie, z wykorzystaniem komputerów, wiedzy zawartej w elektronicznych bazach danych. W skład technik DM wchodzi m.in.: analiza skupień, sztuczne sieci neuronowe i drzewa decyzyjne. Celem niniejszego opracowania jest zaprezentowanie podstawowych założeń techniki drzew decyzyjnych na przykładzie modelowania zawartości laktozy w mleku krowim.

Analiza skupień jest metodą DM, której zadaniem jest tworzenie skupień obiektów (klastrow). Skupienia te są tworzone z wykorzystaniem maksymalizacji podobieństwa wewnątrz klas oraz minimalizacji podobieństwa między klasami. Oznacza to, że obiekty należące do jednego skupienia charakteryzują się dużym

podobieństwem do siebie, różnią się zaś od obiektów z innych skupień. Zagadnienie analizy skupień było już szeroko opisane przez Grużewską na łamach „Przeglądu Hodowlanego” w roku 2009 [4, 5]. Autorka przedstawiała przydatność powyższej metody w analizie wyników lęgow kilku zestawów towarowych kurcząt.

Charakterystyczną cechą kolejnej z technik DM – sieci neuronowych – jest przetwarzanie informacji w sposób analogiczny do procesów zachodzących w ludzkim mózgu [6]. Węzły obliczeniowe, prowadzące obliczenia lub konwersję sygnałów poprzez rząd elementów wykonujących tę samą, podstawową czynność początkową, zwane są neuronami. Interesujący jest fakt, że sieci neuronowe mają zdolność uczenia się i dostosowania do zmiennych warunków. Praktyczne zastosowanie sieci neuronowych w krajowej hodowli bydła opisał w roku 2003 Grzesiak [6] – przykład dotyczył przewidywania wydajności mlecznej krów w okresie standardowej laktacji na podstawie wyników udojów próbnych, przeprowadzonych podczas czterech początkowych miesięcy laktacji.

Technika drzew decyzyjnych wzięła nazwę od sposobu prezentacji statystycznego modelu, który ma kształt podobny do drzewa z wyodrębnionym pnem, gałęziami i liśćmi [1, 9, 12, 13]. Korzeń, składający się z całego zbioru danych, zapoczątkowuje proces tworzenia drzewa decyzyjnego (rys. 1). Kolejne węzły powstałe wskutek dzielenia korzenia zwane są węzłami rodzicielskimi i potomków. Końcowym efektem tworzenia drzewa są podzbiory danych (liście), maksymalnie jednorodnie pod względem wartości zmiennej zależnej. Liście nie podlegają już dalszym podziałom.

W każdym węźle, poza liśćmi, realizowane jest testowanie wartości pewnego atrybutu analizowanego obiektu, a wychodzące z węzła krawędzie są rezultatem powyższego testowania [13].

Z kolei liść zawiera klasę, do której należy obiekt. Wskaźnikiem wielkości drzewa decyzyjnego jest ilość utworzonych liści, zaś o jego głębokości świadczy ilość krawędzi (ilość poziomów) dzielących wierzchołek drzewa od jego najdalszych liści. Przykładowo, drzewo przedstawione na rysunku 1. zawiera 15 liści i jest głębokie na 5 poziomów.

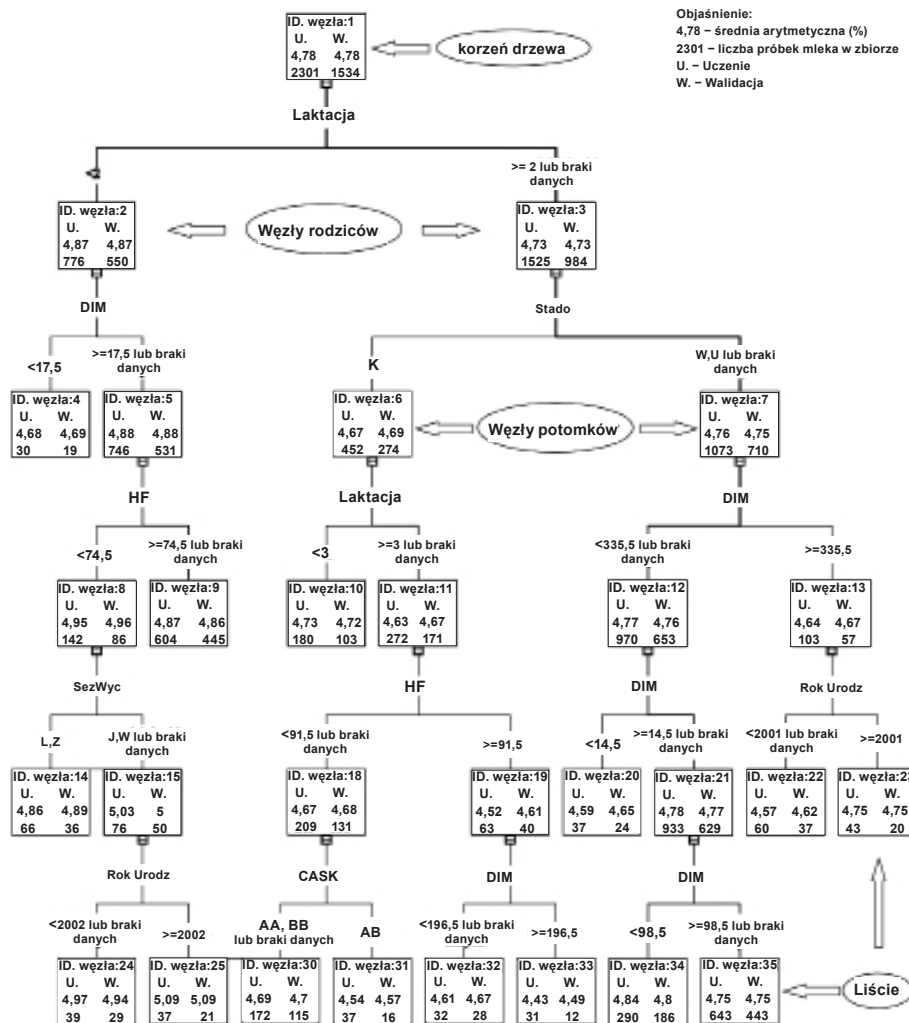
Literatura zootechniczna i weterynaryjna ostatnich lat jest bogata w liczne przykłady praktycznego wykorzystania techniki drzew decyzyjnych (DT) jako metody statystycznej analizy danych [2, 3, 7, 8, 10, 12, 14]. Ebrahimi [3] podaje przykład zastosowania techniki drzew decyzyjnych do porównania różnych właściwości hormonu prolaktyny u kilku gatunków zwierząt. Sawa i wsp. [14] przy użyciu tej metody ustalali genetyczne, fizjologiczne oraz środowiskowe parametry wpływające na wysoki poziom białka oraz na niską zawartość komórek somatycznych w mleku. Z kolei Meyer i wsp. [8], stosując technikę DT, modelowali śmiertelność cieląt, zaś Piwczyński i Sitkowska – liczbę komórek somatycznych [11]. Grzesiak i wsp. [7] na podstawie przeprowadzonych badań stwierdzili, że zastosowanie drzew decyzyjnych pozwoli wskazać hodowcom działania konieczne do podjęcia podczas przygotowywania krów do inseminacji, a w konsekwencji przyczynić się do poprawy rentowności stada.

Techniki drzew decyzyjnych posiadają szereg zalet. Jedną z nich jest fakt, że graficzne modele drzew są relatywnie zrozumiałe i łatwe w interpretacji, w szczególności gdy uwarunkowanie interesującej nas zmiennej zależnej, tzw. zmiennej celu, jest efektem wielu czynników, między którymi istnieją istotne interakcje [10, 11, 12, 13]. Kolejną zaletą drzew decyzyjnych jest możliwość analizy cech mierzonych w skali nominalnej, porządkowej i interwałowej. Drzewa decyzyjne pozwalają również na prowadzenie wstępnej selekcji cech, które mogą mieć statystyczny wpływ na zmienną celu.

W zależności od jakościowego lub ilościowego rodzaju zmiennej celu wyróżnia się drzewa klasyfikacyjne (zmienna celu wyrażona w skali nominalnej) oraz regresyjne (zmienna celu wyrażona w skali przedziałowej) [13].

Zatem w przypadku modelowania przeżycia lub upadku jagnięcia będzie tworzone drzewo klasyfikacyjne, zaś w przypadku zawartości laktozy w mleku – drzewo regresyjne. Algorytm odpowiedzialny za tworzenie drzewa może podzielić zbiór danych na dwa lub więcej podzbiorów. W pierwszym przypadku podziały zwane są binarnymi (rys. 1), w drugim zaś wieloosobowymi.

Proces tworzenia drzewa jest wieloetapowy [13]. Na każdym etapie podziału węzłów są wykorzystywane różne zmienne, tj. takie, których użycie zapewni najlepszy z możliwych podział węzła, czyli w konsekwencji utworzy najbardziej homogeniczne podzbiory. Podczas tworzenia drzewa brane są pod uwagę różne kryteria podziału, zwłaszcza w sytuacji, gdy zmienna celu wyrażona jest w skali nominalnej czy porządkowej: test χ^2 Pearsona, indeks Gini oraz funkcja entropii [11, 12, 13]. Obliczane wartości indeksu Gini, jak i funkcji entropii posiada wspólną cechę: im wyższą wartość przyjmują, tym jest większe zróżnicowanie populacji pod względem zmiennej celu. Wartość zero przyjmują, gdy rozkład cechy jest skupiony na pojedynczej wartości. W przypadku budowy drzew regresyjnych (zmienna celu w skali przedziałowej) kryterium podziału zbioru jest test F lub tzw. redukcja wariancji [13]. W praktyce oznacza to, że na każdym etapie podziału węzłów przeprowadzana jest jednoczynnikowa analiza wariancji, w której rolę czynnika doświadczalnego pełnią kolejno wyła-



Rys. 1. Podziały zbioru danych

niane zmienne niezależne. W konsekwencji oblicza się wartości statystyki F lub bada redukcję zmienności, które są podstawą do wspomnianego podziału węzła.

Analiza statystyczna zmiennej celu przy użyciu różnych kryteriów podziału może prowadzić do otrzymania różnych modeli drzew klasyfikacyjnych. Jakość informacji zawartych w utworzonych modelach można porównywać za pomocą różnych miar. Należą do nich m.in.: przeciętny błąd kwadratowy, statystyka Kołmogorowa-Smirnowa, stopień błędnych klasyfikacji oraz obszar pod krzywą ROC [11, 12, 13]. Obniżenie wartości przeciętnego błędu kwadratowego, stopnia błędnych klasyfikacji oraz zwiększenie wartości statystyki Kołmogorowa-Smirnowa i obszaru znajdującego się pod krzywą ROC wskazuje na wyższą jakość danego modelu.

Budowane modele drzew decyzyjnych są efektem wielu podziałów, powstałych przy udziale wielu zmiennych. Dogodną miarą pozwalającą na ustalenie rankingu zmiennych, jakie odgrywają rolę w czasie realizowanych podziałów jest miara „Importance” [11, 12, 13]. Miara ta przyjmuje wartości z zakresu od 0 do 1 – im wyższe są wartości dla konkretnej zmiennej, tym jej rola w budowie drzewa jest większa.

Tworzenie drzewa decyzyjnego poprzedzane jest zwyczajowo procesem partycjonowania wejściowego zbioru danych. W czasie tego procesu tworzony jest na ogół treningowy i walidacyjny zbiór danych [13]. Zbiór treningowy wykorzystywany jest do wstępnego dopasowania modelu. Z kolei zbiór walidacyjny pozwala na porównanie jakości tworzonych modeli, ale również zapobiega przetrenowaniu modelu. To niepożądane zjawisko, występujące w trakcie tworzenia drzewa, prowadzi do powstania dużego drzewa decyzyjnego, które jest dobrze dopasowane do danych, na których zostało zbudowane, ale jednocześnie jest nieprzydatne podczas klasyfikacji nowych przypadków. O przetrenowaniu drzewa mogą świadczyć duże rozbieżności w statystykach węzłów, między danymi treningowymi a walidacyjnymi. Najczęściej stosowanymi metodami partycjonowania danych są: losowanie proste i warstwowe. W przypadku zmiennej celu kategoryzującej następuje warstwowanie próby w docelowej klasie, w przeciwnym razie powinno być zastosowane losowanie proste.

Aplikacja SAS Enterprise Miner, za pomocą której dokonano opracowania statystycznego opisanego w dalszej części pracy, w celu stworzenia modelu drzewa decyzyjnego wykorzystuje sprawdzone algorytmy decyzyjne, takie jak: CHAID, CART, C4.5 [13]. Zastosowany algorytm zwykle tworzy drzewo w dwóch fazach. W pierwszej fazie, zwanej fazą budowania, tworzy drzewo z treningowej bazy danych, a w drugiej, zwanej fazą obcinania lub redukcji (pruning), następuje obcinanie drzewa w celu poprawy jego dokładności i ułatwienia interpretacji. W trakcie przycinania drzewa usuwane są zawodne gałęzie – reprezentują one punkty osobliwe i szumy. Zabiegi te pozwalają również uniknąć efektu przetrenowania drzewa. Istnieją dwie strategie przycinania: postpruning i prepruning. W przypadku postpruning budowane jest maksymalnie duże drzewo, a następnie usuwane są „zawodne” liście. Strategia prepruning wstrzymuje zaś rozrost drzewa w momencie, gdy gałęzie zaczynają być niepewne.

Tworzenie drzew, jakie odbywa się przy wsparciu programów komputerowych [13], może być realizowane podczas treningu automatycznego i interaktywnego. W pierwszym przypadku następuje automatyczne szeregowanie zmiennych wejściowych, na podstawie wielkości udziału w tworzeniu drzewa. Z kolei tryb interaktywny pozwala użytkownikowi indywidualnie przeglądać odpowiednie wyniki lub też samodzielnie wykonywać trening danych. Trening interaktywny może obcinać drzewo w celu uniknięcia przetrenowania modelu, osiągnięcia stabilności w szacunkach oraz zwiększenia uogólnienia modelu.

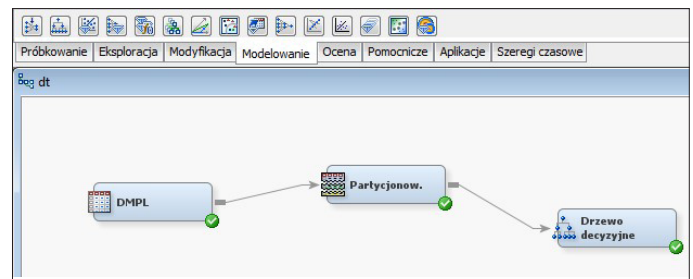
Algorytm budujący drzewo musi zdecydować, kiedy zaprzestać budowania dalszych struktur [13]. Podział zbioru danych zostaje zatrzymany, gdy wszystkie próby należą do jednej klasy, gdy nie ma dalszych cech, względem których można byłoby dokonywać podziałów lub też gdy nie ma już dalszych prób, które mogłyby stworzyć nową gałąź.

Przykład zastosowania drzewa decyzyjnego

Przykład użycia drzewa decyzyjnego (regresyjnego) dotyczy zawartości laktozy w mleku krowim (zmienna celu). Badania, któ-

rych wyniki zaprezentowano, dotyczyły 150 krów rasy polskiej holsztyńsko-fryzyskiej (HF) odmiany czarno-białej użytkowanych w jednym ze stad na terenie Pomorza i Kujaw. Na podstawie przeprowadzonych próbnych dojów uzyskano informacje o zawartości laktozy w 3835 próbach mleka. Średni poziom tej cechy wyniósł 4,782%, a odchylenie standardowe 0,227%. Zmienność zawartości laktozy badano w zależności od następujących czynników: udział rasy HF, stada (K, U, W), kolejnej laktacji (1, 2, 3, ≥ 4), roku urodzenia (1996...2004), sezonu wycielenia, sezonu doju próbnego (wiosna, lato, jesień, zima), dnia doju próbnego oraz genotypu białka CASK (AA, AB, BB). Analizę statystyczną zebranego materiału liczbowego przeprowadzono równoległe za pomocą dwóch metod: wieloczynnikowej analizy wariancji oraz metody drzew decyzyjnych. W obydwu przypadkach posłużono się aplikacją komputerową Enterprise Miner 12.1, wchodzącą w skład pakietu SAS [13]. Budując drzewo decyzyjne założono, że spośród przebadanych 3835 próbek 60% zostanie wykorzystanych do utworzenia zbioru treningowego, a pozostałe 40% – walidacyjnego. Założono również, że minimalna wielkość końcowego węzła wynosić będzie 50 prób mleka.

Na rysunku 2. zaprezentowano diagram przetwarzania danych wygenerowany przez program Enterprise miner 12.1. Węzeł „DMPL” odpowiedzialny jest za wprowadzanie danych, „Partycjonow.” pełni zadanie podziału obserwacji na zbiory treningowy i walidacyjny. Węzeł „Drzewo decyzyjne” służy do tworzenia modelu drzew decyzyjnych.



Rys. 2. Diagram przetwarzania danych

Zastosowany algorytm budowy drzew decyzyjnych wyłonił następujące czynniki odpowiedzialne za zmienność zawartości laktozy, według znaczenia (Importance) w porządku malejącym: kolejna laktacja, dzień doju próbnego, stado, udział rasy HF, sezon wycielenia, rok urodzenia, genotyp białka CASK (tab. 1). Podkreślenia wymaga fakt, że te same czynniki zostały wskazane również jako źródło zmienności zawartości laktozy za pomocą analizy wariancji (tab. 2). Wykonana analiza wariancji wykazała wpływ wielu statystycznych interakcji. Autorzy uważają, że ich złożoność może być interpretowana za pomocą diagramu drzewa decyzyjnego.

Powstały diagram drzewa jest głęboki na 5 poziomów i składa się z 15 liści (rys. 1). W poszczególnych węzłach umieszczono informację dotyczącą średniej arytmetycznej oraz liczebności zbiorów uczących i walidacyjnych. Interpretację wyników powstałego drzewa ograniczono do zbioru treningowego (Uczenie) – na podstawie którego drzewo zostało zbudowane.

Pierwszy podział wejściowego zbioru danych został wykonany z uwzględnieniem kolejności laktacji, wyłaniając dwa podzbiory prób mleka: pochodzące od pierwiastek (węzeł 2) i wieloródek (węzeł 3). Średnia zawartość laktozy w mleku pierwiastek wyniosła 4,87%, a wieloródek 4,73% (węzeł 3). Zmienną najsilniej różnicującą zawartość laktozy u pierwiastek był dzień doju próbnego. Stwierdzono, że w początkowych 17 dniach laktacji (węzeł 4) zawartość laktozy była o 0,22 punktu procentowego (p.p.) niższa aniżeli w czasie dalszego okresu laktacji (węzeł 5). Podkreślenia wymaga fakt, że podzbiór złożony z prób mleka pierwiastek do 17. dnia laktacji nie podlegał już dalszym podziałom – stał się liściem. Różnicujące podziały oparte na udziale rasy HF były natomiast realizowane na próbach uzyskanych od 18. dnia laktacji (węzeł 5). Wykazano, że pierwiastki o udziale HF poniżej 74,5% posiadały w mleku o 0,08 p.p. więcej laktozy niż te o udziale równym lub wyż-

Tabela 1
Znaczenie zmiennych

Wyszczególnienie	Liczba reguł podziału	Importance
Laktacja	2	1,000
Dzień doju próbnego (DIM)	5	0,728
Stado	1	0,477
HF	2	0,398
Sezon wycielenia	1	0,305
Rok urodzenia	2	0,296
CASK	1	0,253
Sezon doju próbnego	0	0,000

Tabela 2
Rezultaty analizy wariancji

Źródło zmienności	Wartość F	Pr > F
Stado	46,97	<,0001
RokUrodz	2,38	0,0148
Stado x RokUrodz	8,12	<,0001
Laktacja	25,14	<,0001
Stado x Laktacja	13,95	<,0001
RokUrodz x Laktacja	4,84	<,0001
SezTD	8,31	<,0001
Stado x SezTD	1,84	0,0867
RokUrodz x SezTD	2,35	0,0002
Laktacja x SezTD	0,88	0,5430
SezWyc	2,32	0,0735
Stado x SezWyc	3,72	0,0011
RokUrodz x SezWyc	2,78	<,0001
Laktacja x SezWyc	4,48	<,0001
SezWyc x SezTD	6,61	<,0001
CASK	4,60	0,0101
Stado x CASK	1,42	0,2360
RokUrodz x CASK	6,99	<,0001
CASK x Laktacja	1,96	0,0678
CASK x SezTD	1,23	0,2871
CASK x SezWyc	3,67	0,0012
DIM	5,69	0,0171
DIM x DIM	11,14	0,0009

SezWyc – sezon wycielenia, SezTD – sezon doju próbnego, RokUrodz – rok urodzenia

szym od 74,5%. Czynnikiem, który różnicował poziom laktozy u pierwiastek o niższym udziale HF był sezon wycielenia. Stwierdzono, że zawartość laktozy w mleku pierwiastek wycielonych latem (L) i zimą (Z) była o 0,17 p.p. niższa aniżeli w pozostałych porach roku (J – jesień, W – wiosna) (węzeł 15). Podzbiór prób mleka pierwiastek wycielonych wiosną i jesienią (węzeł 15) uległ kolejnemu podziałowi na podstawie roku urodzenia. Zaobserwowano, że zawartość laktozy w mleku była o 0,12 p.p. wyższa wśród pierwiastek urodzonych po 2001 roku (węzeł 25) niż we wcześniejszych latach (węzeł 24).

W zbliżony sposób należy interpretować część drzewa dotyczącą wieloródek (węzeł 3). Podkreślenia wymaga jednak fakt, że

ważnym czynnikiem wpływającym na zawartość laktozy w mleku był efekt stada (węzły 6 i 7). Próby pochodzące ze stada K (węzeł 6) charakteryzowały się o 0,09 p.p. niższym poziomem laktozy niż z pozostałych stad (węzeł 7). Uwzględnienie w algorytmie podziału zbioru danych czynnika stada świadczyć może o zróżnicowanych w nich warunkach utrzymania i żywienia zwierząt. Zaobserwowano też, że krowy pochodzące ze stada K i będące w trzeciej i dalszych laktacjach (węzeł 11) produkowały mleko o niższej zawartości laktozy niż będące w laktacji drugiej (węzeł 10). Spośród wymienionych podzbiorów dalszym podziałem poddany był zbiór starszych zwierząt. Czynnikiem różnicującymi w tym przypadku były: udział rasy HF (węzły 18 i 19), genotyp białka CASK (węzły 30 i 31) i dzień doju próbnego (węzły 32 i 33). Na podstawie przeprowadzonych podziałów zbioru krów będących w trzeciej i dalszych laktacjach (węzeł 11) wynika, że wyższy udział rasy HF obniżał o 0,15 p.p. zawartość laktozy. Zaobserwowano, że zwierzęta o genotypie AB białka CASK zawierały w mleku o 0,15 p.p. mniej laktozy niż o genotypie AA i BB. Z kolei bardziej zaawansowana faza laktacji (od 197. dnia) w porównaniu do prób z okresu poprzedzającego skutkowało obniżeniem zawartości laktozy o 0,18 p.p. W analogiczny sposób można interpretować podziały wieloródek z innych niż K stad (węzeł 7). Podziały te opierały się przede wszystkim na dniu doju próbnego (3 podziały) i roku urodzenia (1 podział).

Podsumowując wyniki przeprowadzonych analiz należy stwierdzić, że zastosowane równolegle dwie metody statystyczne, tj. drzew decyzyjnych oraz wieloczynnikowa analiza wariancji, wskazały te same czynniki warunkujące zawartość laktozy w mleku krowim: kolejna laktacja, dzień doju próbnego, stado, udział rasy HF, sezon wycielenia, rok urodzenia i genotyp białka CASK. Na podstawie skonstruowanego modelu drzewa decyzyjnego można wnioskować, że wysokiej zawartości laktozy należy oczekiwać w mleku pierwiastek, powyżej 17. dnia doju, o udziale HF co najmniej 74,5%. Z kolei dalszy okres produkcyjny, powyżej 2. laktacji, sprzyja niższej zawartości laktozy, zwłaszcza gdy próby mleka, od krów o udziale HF co najmniej 91,5%, pochodzą ze stad o warunkach utrzymania i żywienia zbliżonych do panujących w stadzie K. Tworzone na zbliżonej zasadzie, jak w przeprowadzonych badaniach, graficzne modele drzew mogą być opracowywane praktycznie w odniesieniu do każdej cechy użytkowej. Mogą stanowić tym samym narzędzie wskazujące dogodny układ poziomów czynników genetycznych i środowiskowych, warunkujące wysoki poziom interesujących hodowcę cech.

Literatura: 1. Austin P.C., 2007 – Statistics Med. 26, 2937-2957. 2. Dawson S., Else R.W., Rhind S.M., Collie D.D.S., 2005 – Vet. Record 157, 433-436. 3. Ebrahimi M., 2011 – Iranian J. Vet. Res. 12, 205-213. 4. Grużewska A., 2009 – Przegląd Hod. 10, 28-30. 5. Grużewska A., 2009 – Przegląd Hod. 12, 13-14. 6. Grzesiak W., 2003 – Pr. Mat. Zoot. 61, 71-89. 7. Grzesiak W., Zaborski D., Sablik P., Pilarczyk R., 2011 – Anim. Sci. Pap. Rep. 29, 293-302. 8. Meyer C.L., Berger P.J., Koehler K.J., 2000 – J. Dairy Sci. 83, 2657-2663. 9. Myles J.A., Feudale R.N., Liu Y., Woody N.A., Brown S.D., 2004 – J. Chemometrics 18, 275-285. 10. Piwczyński D., 2009 – J. Central Europ. Agricult. 10(3), 303-310. 11. Piwczyński D., Sitkowska B., 2012 – Arch. Tierzucht 55, 4, 332-345. 12. Piwczyński D., Sitkowska B., Wiśniewska E., 2012 – Small Rum. Res. 103, 225-231. 13. SAS Institute Inc., 2012. Getting started with SAS® Enterprise Miner 12.1TM. Cary, NC: SAS Institute Inc. 14. Sawa A., Kowaliszyn B., Piwczyński D., 2004 – Zesz. Nauk. Przeg. Hod. 72, z. 1, 41-49.

Application of data mining techniques in animal husbandry

Summary

While summing up the results of the analyzes it should be noted that parallel application of two statistical methods, the method of decision trees and multivariate analysis of variance, have indicated the same factors that determine the lactose content in cow's milk: successive lactation, day of milking testing, herd, the participation of HF breed, calving season, year of birth, CASK protein genotype. Based on the constructed model of the decision tree, it can be concluded that a high lactose content may be expected in the primiparas milk, after 17 days of milking, with the participation of HF at the level of at least 74.5%. However, a subsequent production period of more than 2 lactations, promotes the lower lactose content in milk, especially when milk samples, collected from cows with the participation of HF of at least 91.5%, are obtained from flocks with living and feeding conditions similar to those in the herd K. The graphical models of trees, created in the similar way to those described in the research, may be developed for virtually every feature. Therefore, they may become the tool indicating a good system of levels of genetic and environmental factors and provide the high level of features that are useful for breeders.

KEY WORDS: classification trees, modeling, performance traits