

Kamil SIDOR¹

ZASTOSOWANIE METOD EKSPLOKACJI DANYCH DO WSPOMAGANIA PRZYGOTOWANIA PROCESU DYDAKTYCZNEGO

Eksploracja danych dostarcza cennej wiedzy ukrytej w dużych zbiorach danych. Pozwala na odkrywanie zależności niewidocznych gołym okiem. Swoje zastosowanie może znaleźć także w edukacji podczas przygotowywania oferty dydaktycznej. Artykuł przedstawia zastosowanie algorytmów eksploracji danych w przygotowaniu procesu edukacyjnego. W rozważanym zakresie eksploracja danych służy do przekształcania surowych danych w wiedzę, która pozwala na poznanie preferencji studentów. Skupiono się na odkrywaniu grup studentów oraz tworzeniu ich modeli określających style uczenia się. W trakcie budowania grup zastosowano klasyfikację bez nadzoru m.in. metody k-średnich oraz EM. Grupy tworzone były z uwzględnieniem preferencji studentów dotyczących nauki. Pozwoliło to na uzyskanie grup zawierających studentów o podobnych stylach uczenia się. Do zweryfikowania poprawności klasyfikacji wykorzystane zostały indeksy walidacyjne, które pozwoliły na wybranie najbardziej efektywnego podziału studentów. Badania przeprowadzono na danych zebranych wśród studentów Politechniki Rzeszowskiej na podstawie ankiety zawierającej kwestionariusz ILS. Uzyskane podczas badań wyniki pozwoliły na określenie ile różnorodnych materiałów dydaktycznych należy przygotować, aby były dopasowane do preferencji studentów różnych grup. Poznanie stylów uczenia się studentów pozwala nauczycielowi na lepsze zrozumienie upodobań studentów, a samym uczniom na dopasowanie materiałów do własnego stylu uczenia, dzięki czemu łatwiej i szybciej przyswajają wiedzę.

Słowa kluczowe: data mining, ILS, style uczenia się, klasyfikacja, k-means, EM, odkrywanie wiedzy, proces dydaktyczny

1. Wprowadzenie

Jak mówi definicja, „proces dydaktyczny (proces kształcenia, proces nauczania) to intencjonalny, świadomy, prawidłowy i przebiegający regularnie ciąg czynności nauczycieli i uczniów, ukierunkowany na opanowanie przez uczniów

¹ Kamil Sidor, Katedra Automatyki i Informatyki, Politechnika Rzeszowska, ul. W. Pola 2 35-959 Rzeszów, email: ksidor@prz.edu.pl

wiedzy o świecie, kształtowanie umiejętności i nawyków, rozwijanie zdolności i zainteresowań, kształtowanie przekonań i postaw, a także na realizację innych założonych celów dydaktycznych” [10]. Proces dydaktyczny składa się z następujących po sobie ogniw. Jednym z najtrudniejszych jest zaznajamianie studentów z nowym materiałem. Wymaga to od nauczyciela przygotowania materiałów dydaktycznych tak, aby w jasny i zrozumiały sposób przekazywały wiadomości. Nauczyciel powinien być nie tylko ekspertem z dziedziny, której uczy, ale także ekspertem od nauczania – potrafiącym przekazać wiedzę studentom w nieskomplikowany dla nich sposób [11].

Dostosowanie sposobu przekazywania wiedzy uczniom to niezwykle trudne zadanie. Nie można go zrealizować prowadząc wykład, gdzie jest 100 studentów i jeden prowadzący. Realizacja dostosowania sposobu przekazywania wiedzy ma większy sens wraz z zastosowaniem e-learningu, który zyskał dużą popularność. Punktem wyjścia w przygotowaniu materiałów wykorzystywanych w e-learningu jest identyfikacja stylów uczenia się odbiorców. Materiały przygotowane z uwzględnieniem stylów uczenia się są w szybszy sposób przyswajane przez uczniów i nie zniechęcają ich do nauki [3].

Osoby odpowiedzialne za proces uczenia podkreślają, że istotnym czynnikiem oddziałującym na jakość uczenia jest dotychczas zdobyta wiedza oraz umiejętności ucznia [3]. Jednakże proces uczenia się jest bardzo skomplikowanym zjawiskiem łączącym wiele zależnych od siebie czynników.

2. Style uczenia

Warto rozpatrzeć zagadnienie stylów uczenia się z dwóch perspektyw – nauczyciela i ucznia. Każdy człowiek uczy się w unikatowy dla siebie sposób. Często technika uczenia się jest mieszaniną różnych stylów. Z kolei nauczyciele nie zawsze prezentują wiedzę w najbardziej satysfakcjonujący sposób dla uczniów. Takie niedostosowanie sposobu przekazywania wiedzy do stylu uczenia się studentów nieraz prowadzi do niepowodzeń w kształceniu.

Niejednokrotnie uczniowie twierdzą, że nie potrafią się samodzielnie uczyć. W większości przypadków wynika to z niezajomości własnego stylu uczenia się. W przypadku małych grup, nauczyciel może bliżej poznać uczniów i ich charakter. W przypadku dużych grup studentów, nauczyciel nie ma możliwości poznania studentów, a tym bardziej ich upodobań. Z kolei studenci korzystający z e-learningu są całkowicie nieznani i anonimowi dla nauczyciela. Użyteczne mogą się tutaj wydawać style uczenia się. Tego typu informacja może być pomocna nie tylko dla nauczycieli, ale i dla uczniów, którzy mogą korzystać z najlepiej dopasowanych dla siebie materiałów.

Preferencje i cechy charakterystyczne studentów uzyskuje się zwykle na podstawie wypełnionych przez nich różnego rodzaju kwestionariuszy. Do naj-

częściej używanych atrybutów studentów należą ich indywidualne style uczenia się [13].

Wpływ stylów uczenia się na skuteczność nauczania został przedstawiony przez wielu badaczy. Lee [7] pokazał, że style uczenia się studentów wpływają na osiągnięte przez nich wyniki w nauce. Z kolei Graf i Kinshuk [4] zaprezentowali, że studenci o różnych stylach uczenia mają różne potrzeby i preferencje. Natomiast Beaudoin [1] stwierdził, że style uczenia się określają zachowania studentów. Na tej podstawie można stwierdzić, że style uczenia się mogą w przyszłości stanowić bardzo ważny element podczas przygotowania procesu dydaktycznego.

3. Ankieta stylów uczenia i zbierania danych

Do przeprowadzenia badań wykorzystane zostały dane zebrane poprzez przeprowadzenie ankiety internetowej wśród studentów Politechniki Rzeszowskiej. Ankieta składała się z kwestionariusza Index of Learning Styles (ILS) oraz 14 dodatkowych pytań. Kwestionariusz stosowany jest do określenia preferencji stylu uczenia się. Sformułowany został przez R.M Feldera oraz L. K. Silvermana. Odnosi się do skrajnych stylów uczenia w czterech kategoriach:

- *Aktywny – Refleksyjny*: określa styl myślenia,
- *Sekwencyjny – Globalny*: określa styl przyporządkowania zjawisk,
- *Dotykowy – Intuicyjny*: określa styl postrzegania nowych zjawisk,
- *Wizualny – Werbalny*: określa styl działania.

Kwestionariusz składa się z 44 rozpoczętych zdań, które należy dokończyć poprzez wybór jednego z dwóch podanych wariantów [2,5]. W przypadku gdy brak jest idealnie dopasowanej odpowiedzi, należy wybrać tę najbliższą prawdy. Ze względu na wybór skrajnych odpowiedzi, pojedyncze wyniki testu mają charakter dwudzielny. Każda z odpowiedzi zawartych w kwestionariuszu oznaczona jest jedną z dwóch wartości: -1 lub 1. Dla przykładu mając pytanie odnoszące się do kategorii *aktywny-refleksyjny*, wartość -1 opisuje odpowiedź charakteryzującą *styl aktywny*, z kolei 1 *styl refleksyjny*. Ostatecznie dla każdej z kategorii liczona jest średnia z wybranych przez respondenta odpowiedzi. Indeks, który jest wynikiem wypełnienia kwestionariusza ILS, przyjmuje wartość nieparzystej liczby całkowitej z przedziału [-11,11], przydzielonej do każdego z czterech wymiarów, np. (-3, -7,-9,-1). W ten sposób każdy student, który wypełnił kwestionariusz ILS może być opisany przez wektor o 4 atrybutach. Dla przykładu wartości ujemne oznaczają *styl aktywny*, *sekwencyjny*, *dotykowy*, *wizualny*, zaś wartości dodatnie *styl refleksyjny*, *globalny*, *intuicyjny*, *werbalny* [13].

W badaniach wykorzystano własne tłumaczenie kwestionariusza, które wykonane zostało z możliwie najwyższą starannością. W tabeli 1. przedstawione

zostały przykładowe pytania kwestionariusza odnoszące się do poszczególnych skrajnych kategorii stylu uczenia.

Tabela 1. Przykłady pytań kwestionariusza ILS

Table 1. Examples of questions questionnaire ILS

Styl	Kategoria stylu	Przykład
Myślenia	Aktywny (akt)	<i>Pracując w grupie nad trudnym materiałem będę proponował rozwiązania.</i>
	Refleksyjny (ref)	<i>Pracując w grupie nad trudnym materiałem będę słuchał rozwiązań innych.</i>
Przyporządkowania zjawisk	Globalny (glo)	<i>Łatwiej mi zrozumieć całą strukturę tematu, nie pamiętać o szczegółach.</i>
	Sekwencyjny (sek)	<i>Łatwiej mi zrozumieć szczegóły tematu, i mieć rozmyty ogół.</i>
Postrzegania nowych zjawisk	Dotykowy (dot)	<i>Uważam, że łatwiej uczyć się faktów.</i>
	Intuicyjny (int)	<i>Uważam, że łatwiej uczyć się pojęć.</i>
Działania	Wizualny (wiz)	<i>Kiedy podczas zajęć widzę diagram, najprawdopodobniej zapamiętam obraz diagramu.</i>
	Werbalny (wer)	<i>Kiedy podczas zajęć widzę diagram, najprawdopodobniej zapamiętam to, co mówił na jego temat prowadzący zajęcia.</i>

4. Analiza danych

Badania przeprowadzone zostały na 621 próbkach, gdzie każda z próbek reprezentowała jednego studenta. Podczas badań skupiono się na odkrywaniu grup studentów. Podział studentów przeprowadzony został z wykorzystaniem klasyfikacji bez nadzoru. Wykorzystane zostały następujące algorytmy:

- k-średnich z euklidesową funkcją odległości,
- k-średnich z miejską funkcją odległości,
- EM (Expectation Maximization) [6,8,14].

Algorytm k-średnich należy do najbardziej popularnych algorytmów grupowania. W grupowaniu wykorzystuje miarę odległości pomiędzy obiektami (zazwyczaj euklidesową lub miejską). W przeciwieństwie do algorytmu

k-średnich, algorytm EM przydziela każdy obiekt do klastra zgodnie z prawdopodobieństwem przynależności [8,14].

Liczba zróżnicowanych materiałów dla studentów tworzona przez nauczycieli nie może być bardzo duża. Z tego względu studenci byli dzieleni każdym algorytmem na 2, 3, 4 i 5 grup. Oznacza to, że maksymalna liczba zróżnicowanych materiałów do przygotowania przez nauczyciela nie będzie większa od 5.

W badaniach wykorzystane zostało środowisko WEKA (Waikato Environment for Knowledge Analysis). WEKA zawiera pakiet algorytmów uczenia maszynowego przeznaczony do zastosowań związanych z eksploracją danych. Oprogramowanie to jest dostępne na licencji GNU General Public License.

4.1. Algorytm k-średnich z euklidesową funkcją odległości

Wyniki grupowania algorytmem k-średnich z euklidesową funkcją odległości zostały przedstawione w tabeli 2. Tabela przedstawia liczebność każdej z grup oraz średnie wartości atrybutów w poszczególnych grupach.

Tabela 2. Wyniki grupowania studentów metodą k-średnich z euklidesową funkcją odległości

Table 2. The results of grouping students k-means (Euclidean distance)

Liczba grup	Liczba studentów	Średnie			
		akt_ref	dot_int	wiz_wer	sek_glo
2	378	-5,11	-5,16	-8,33	-2,58
	243	-0,46	-3,90	-3,32	-1,45
3	209	-5,06	-1,95	-7,27	0,17
	177	0,06	-4,60	-2,48	-2,24
	235	-4,24	-7,14	-8,51	-4,11
4	187	-5,97	-2,10	-7,32	-0,02
	119	1,71	-3,13	-6,19	-1,07
	206	-4,57	-7,26	-8,45	-4,61
	109	-1,71	-5,84	-1,02	-2,27
5	166	-5,90	-3,02	-7,93	0,83
	110	-0,18	-1,71	-6,69	-3,18
	141	-5,54	-7,07	-9,07	-5,57
	135	-1,35	-7,59	-4,64	-2,57
	69	-1,14	-2,71	0,01	0,25

Na podstawie grupowania możemy stwierdzić, że algorytm k-średnich z euklidesową funkcją odległości utworzył grupy o różnej liczebności. Niezależnie od liczby tworzonych grup, liczebność poszczególnych klastrów jest zróżnicowana.

4.2. Algorytm k-średnich z miejską funkcją odległości

Kolejną wykorzystaną metodą klasyfikacji był algorytm k-średnich z miejską funkcją odległości. W tabeli 3 przedstawione zostały wyniki grupowania.

Tabela 3. Wyniki grupowania studentów metodą k-średnich z miejską funkcją odległości

Table 3. The results of grouping students k-means (Manhattan distance)

Liczba grup	Liczba studentów	Średnie			
		akt_ref	dot_int	wiz_wer	sek_glo
2	407	-5	-5	-7	-1
	214	-1	-3	-5	-5
3	226	-3	-5	-7	1
	170	-1	-3	-3	-5
	225	-5	-7	-9	-3
4	197	-3	-5	-7	1
	161	-1	-3	-5	-5
	197	-5	-7	-9	-3
	66	-3	-9	-1	-3
5	166	-5	-3	-7	1
	115	-1	-3	-5	-5
	169	-5	-7	-9	-3
	96	-3	-9	-5	-3
	75	-1	-5	-1	-1

Podobnie jak w przypadku budowania grup z euklidesową funkcją odległości, utworzone grupy posiadają zróżnicowaną liczebność studentów.

4.3. Algorytm maksymalizacji wartości oczekiwanej

Ostatnim wykorzystanym algorytmem klasyfikacji był algorytm EM, czyli algorytm maksymalizacji wartości oczekiwanej. Wyniki grupowania zostały przedstawione w tabeli 4.

Podobnie jak w dwóch poprzednich algorytmach, także i w tym występuje duże zróżnicowanie liczebności studentów w grupach.

W kolejnym kroku jakość otrzymanych klastrów została porównana z wykorzystaniem indeksów walidacyjnych.

Tabela 4. Wyniki grupowania studentów metodą EM

Table 4. The results of grouping students by EM

Liczba grup	Liczba studentów	Średnie			
		akt_ref	dot_int	wiz_wer	sek_glo
2	328	-4,70	-5,19	-8,80	-2,55
	293	-2,04	-4,20	-4,22	-1,77
3	331	-1,94	-4,43	-4,31	-1,93
	134	-5,34	-2,71	-8,29	-0,66
	156	-4,47	-7,18	-9,38	-4,16
4	296	-1,99	-4,25	-4,05	-2,04
	108	-2,94	-6,71	-8,30	-1,11
	135	-5,58	-2,62	-8,37	-0,76
	82	-4,96	-7,45	-9,80	-6,35
5	217	-2,49	-5,64	-4,80	-3,53
	88	-4,90	-7,33	-10,04	-6,08
	150	-5,88	-2,85	-8,29	-0,93
	96	-1,24	-2,39	-3,30	0,27
	70	-2,24	-6,48	-8,70	-0,14

4.4. Oszacowanie poprawności wyników grupowania

W celu oszacowania poprawności klasyfikacji każdego z zastosowanych algorytmów klasteryzacji zostały wyznaczone indeksy pozwalające ocenić efektywność przeprowadzonej analizy. Były to indeksy: Gamma, Silhouette, Dunna.

Pierwszy z zastosowanych wskaźników to indeks Gamma. Podczas jego obliczania porównuje się odległości wewnątrzklasowe z wszystkimi odległościami międzyklasowymi. Liczba tych porównań to iloczyn liczby odległości wewnątrzklasowych oraz liczby odległości międzyklasowych. Jeśli odległość wewnątrzklasowa jest mniejsza od odległości międzyklasowej to para taka uznawana jest jako zgodna, w przeciwnym wypadku jako niezgodna. W przypadku równości odległości nie są uwzględniane. Wartość indeksu Gamma obliczana jest wg. wzoru:

$$\text{Gamma} = \frac{s(+)-s(-)}{s(+)+s(-)} \quad (1)$$

gdzie:

$s(+)$ – liczba par zgodnych,

$s(-)$ – liczba par niezgodnych .

Indeks może przyjmować wartości z przedziału $\langle -1,1 \rangle$. Większa wartość indeksu odpowiada lepszemu grupowaniu [12].

Kolejnym zastosowanym wskaźnikiem był indeks Silhouette. Obliczany jest na podstawie średnich odległości każdego obiektu od obiektów klasy do której należy oraz obiektów z klasy położonej najbliżej. Indeks Silhouette obliczany jest według wzoru:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i); b(i)\}}, \quad (2)$$

gdzie:

$a(i)$ – średnia odległość obiektu i od pozostałych obiektów klasy P_r ,

$b(i)$ – średnia odległość obiektu i od obiektów należących do klasy P_s .

Używane średnie odległości $a(i)$ i $b(i)$ obliczane są na podstawie wzorów:

$$a(i) = \frac{\sum_{k \in P_r} d_{ik}}{n_r - 1}, \quad (3)$$

gdzie:

d_{ik} – odległość obiektu i od obiektu k należącego do klasy P_r ,

n – liczba obiektów należących do klasy P_r .

$$b(i) = \min \left(\frac{\sum_{k \in P_s} d_{ik}}{n_s} \right), \quad (4)$$

gdzie:

d_{ik} – odległość obiektu i od obiektu k należącego do klasy P_s ,

n – liczba obiektów należących do klasy P_s .

Indeksy $S(i)$ są następnie uśredniane dla klas, a następnie uśredniane są dla całego zbioru. Wartość średnia indeksu dla klasy P oznacza się jako $S(P)$ i obliczana jest według wzoru [12]:

$$S(P) = \frac{\sum_{i \in K} S(i)}{n_k}. \quad (5)$$

Uśredniony indeks Silhouette dla całego zbioru oblicza się według wzoru:

$$S = \frac{\sum S(P)}{n}. \quad (6)$$

Ostatnim wykorzystanym wskaźnikiem był indeks Dunna, który definiowany jest przez dwa parametry:

- średnicę $diam(S_i)$ klastra S_i – odległość między najdalszymi obiektami w danym klastrze,
- zbiór odległości $\delta(S_i, S_j)$ – minimalna odległość pomiędzy obiektami klastra K_i i K_j .

Obliczany jest na podstawie wzoru:

$$Dunn = \min_{1 \leq i \leq m} (\min_{1 \leq j \leq m, i \neq j} (\frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq m} diam(S_k)})) \quad (7)$$

Większe wartości wskaźnika Dunna odpowiadają lepszemu grupowaniu, gdzie klastry są zwarte i oddzielone od siebie [9].

Wyniki wyznaczonych indeksów walidacyjnych dla poszczególnych algorytmów zostały przedstawione w tabeli 5.

Tabela 5. Wartości indeksów walidacyjnych

Table 5. Validation index values

Algorytm grupowania	2 grupy	3 grupy	4 grupy	5 grup
Indeks Gamma				
k-średnich (f. Euklides)	0,4122	0,4811	0,5392	<u>0,5441</u>
k-średnich (f. miejska)	0,4009	0,4340	0,4278	0,4934
EM	0,3926	0,4464	0,4427	0,4802
Indeks Silhouette				
k-średnich (f. Euklides)	0,1991	0,2234	0,2453	<u>0,2471</u>
k-średnich (f. miejska)	0,2083	0,1951	0,2238	0,2133
EM	0,2138	0,2106	0,1963	0,2141
Indeks Dunna				
k-średnich (f. Euklides)	0,0807	0,0848	<u>0,0871</u>	<u>0,0871</u>
k-średnich (f. miejska)	0,0407	0,0495	0,0521	0,0521
EM	0,0428	0,0525	0,0550	0,0548

Zastosowane algorytmy walidacji klastrów jednoznacznie wskazały że najbardziej efektywny podział występuje przy klasyfikacji na 5 grup. Jedyne Indeks Dunna był identyczny dla 4 i 5 klastrów. Jednak z zastosowanych algorytmów klasyfikacji najlepszy okazał się algorytm k-średnich z euklidesową funkcją odległości, stąd też podział ten został przyjęty jako najlepszy sposób grupowania spośród wszystkich przeprowadzonych badań.

Tabela 6 przedstawia cechy charakterystyczne dla każdej z grup. Dla większej czytelności wartości numeryczne zostały przedstawione jako wartości kategoryczne.

Tabela 6. Otrzymane cechy charakterystyczne grup

Table 6. The resulting characteristics groups

Grupa	akt_ref	dot_int	wiz_wer	sek_glo
1	akt	dot_int	wiz	sek_glo
2	akt_ref	dot_int	wiz	sek_glo
3	akt	dot	wiz	sek
4	akt_ref	dot	wiz	sek_glo
5	akt_ref	dot_int	wiz_wer	sek_glo

Przedstawione w tabeli 6 wartości kategoryczne pozwalają na szybkie określenie preferencji studentów znajdujących się w danej grupie. Warto zwrócić uwagę na to, że grupy różnią się od siebie przynajmniej jedną charakterystyczną cechą.

Przeprowadzone badania pozwoliły na określenie charakterystycznych cech dla utworzonych grup. W celu wykorzystania powyższych wyników w przygotowaniu materiałów dydaktycznych należałoby wykonać pięć rodzajów materiałów dostosowanych do preferencji studentów znajdujących się w poszczególnych grupach. Dla przykładu materiały dydaktyczne dla pierwszej grupy powinny umożliwiać pracę w grupie, posiadać dużo ilustracji, wykresów, tabel i diagramów.

5. Podsumowanie

W artykule przedstawione zostały badania dotyczące klasyfikacji studentów do grup na podstawie ich stylów uczenia się. Style uczenia się studentów zostały wyznaczone na podstawie danych zebranych z wykorzystaniem ankiety ILS, przeprowadzonej wśród studentów uczelni.

Na podstawie uzyskanych wyników wykonana została klasyfikacja z wykorzystaniem algorytmów klasyfikacji bez nadzoru. Zastosowano algorytmy k-średnich z euklidesową funkcją odległości, k-średnich z miejską funkcją odległości oraz EM. Do oceny jakości klasyfikacji zostały użyte indeksy walidacyjne takie jak: Dunna, Silhouette oraz Gamma. Podczas klasyfikacji najskuteczniejszy okazał się podział studentów na 5 grup z wykorzystaniem algorytmu k-średnich z euklidesową funkcją odległości.

Na podstawie uzyskanych wyników zostały określone cechy charakterystyczne utworzonych grup. Cechy te pozwoliły na poznanie preferencji studentów w konkretnych grupach. Bazując na preferencjach studentów istnieje moż-

liwość przygotowania materiałów dydaktycznych dopasowanych do stylów uczenia się danej grupy. Dzięki temu uczniowie otrzymają materiały, które będą odpowiadały ich stylom uczenia, a co za tym idzie nie zniechęcą się do nauki. Korzyści z poznawania stylów uczenia się studentów może odnieść także nauczyciel. Znając preferencje studentów będzie mógł dostosować swój sposób przekazywania wiedzy.

Wykorzystanie metod analizy danych w przygotowaniu procesu dydaktycznego pozwala na lepsze przygotowanie pomocy naukowych oraz uatrakcyjnienie oferty dydaktycznej. Dziedzina ta jest w fazie intensywnego rozwoju, ale z biegiem czasu może zyskać na popularności i być powszechnie wykorzystywana.

Literatura

- [1] Beaudoin M.F.: Learning of lurking? Tracking the „invisible” on-line student, *Internet & Higher Educ.* 5, 2002, p.147-155.
- [2] Felder R.M., Silverman L.K.: Learning and teaching styles in engineering education, *Eng. Educ.* 1988, p.674-681.
- [3] Gajewski R.: O stylach uczenia się w I-edukacji, <http://www.e-mentor.edu.pl> [dostęp: 15.08.2015].
- [4] Graf S., Kinshuk: Considering learning styles in learning managements systems: investigation the behaviour of students in an online course. In: *Proc. of the 1st IEEE Int. Workshop on Semantic Media Adaptation and Personalization*, Athens 2006.
- [5] Kotarska-Lewandowska B.: Programy komputerowe a style uczenia się, XIV Konferencja Diagnostyki Edukacyjnej, Opole, 2008.
- [6] Larose D.T.: *Metody i modele eksploracji danych*, Wydawnictwo Naukowe PWN 2008.
- [7] Lee M. : Profiling students adaptation styles in web-based learning, *Comput. Educ.* 36, 2003, p. 121-132.
- [8] Morzy T.: *Eksploracja danych: metody i algorytmy*, Wydawnictwo Naukowe PWN, Warszawa 2013.
- [9] *Popularna Encyklopedia Powszechna*, Wydawnictwo FOGRA, 1994-1998
- [10] Taraszkiewicz M., *Nowa szkoła... wspieranie kariery ucznia*, WSiP, Warszawa 1997, s. 20.
- [11] Zakrzewska D.: *Eksploracja danych w modelowaniu użytkowników edukacyjnych i systemów internetowych*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2012.
- [12] Zakrzewska D.: *Cluster Analysis for Users’s Modeling in Intelligent E-Learning Systems*, Springer Heidelberg 2008, s. 209-214.
- [13] Oszust M.: *Rozprawa doktorska – Zastosowanie grupowania szeregów czasowych do rozpoznawania wypowiedzi w języku migowym na podstawie sekwencji wizyjnych*, Kraków 2013, s. 35-36.

- [14] Walesiak M., Dudek A.: Symulacyjna optymalizacja wyboru procedury klasyfikacyjnej dla danego typu danych – charakterystyka problemu, Zeszyty naukowe Uniwersytetu Szczecińskiego, 2006.

APPLICATION DATA MINING METHODS TO SUPPORT THE PREPARATION OF TEACHING

S u m m a r y

Data mining provides valuable knowledge hidden in large data sets. It allows to explore depending invisible to the naked eye. It has been used in education while preparation educational offer.

The article shows the application of data mining algorithms in the preparation of the educational process. In the considered range, data mining is used to transform raw data into knowledge, which allows to know the students' preferences. It has been focused on discovering groups of students and the development of models for the assessment of their learning styles. It has been applied unsupervised classification during process build groups. Groups have been created taking into account the preferences of students in science. It has been allowed get the groups consisting of students with similar learning styles. To verify the accuracy of the classification has been used indexes validation that allowed you to select the most efficient distribution of students. The study was conducted on data collected among students of Rzeszow University of Technology based on a survey questionnaire containing the ILS. Obtained during the studies results allowed to determine what materials teaching should be prepared to be tailored to the preferences of different groups of students. Understanding the learning styles of students allows teachers to better understand the preferences of students and the students to tailor materials to their own learning style, making it easier and faster to acquire knowledge.

Keywords: data mining, ILS, learning styles, classification, k-means, EM, knowledge discovery, teaching process

DOI: 10.7862/re.2016.3

Tekst złożono w redakcji: styczeń 2016

Przyjęto do druku: marzec 2016