

Scientific Review – Engineering and Environmental Sciences (2017), 26 (2), 226–233
Sci. Rev. Eng. Env. Sci. (2017), 26 (2)
Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska (2017), 26 (2), 226–233
Prz. Nauk. Inż. Kszt. Środ. (2017), 26 (2)
<http://iks.pn.sggw.pl>
DOI 10.22630/PNIKS.2017.26.2.21

Wojciech DROZD, Marcin KOWALIK

Instytut Zarządzania w Budownictwie i Transporcie, Politechnika Krakowska
im. Tadeusza Kościuszki
Institute of Construction and Transportation Engineering & Management,
Tadeusz Kościuszko Cracow University of Technology

Drzewa decyzyjne w analizie wypadkowości w budownictwie Decision trees algorithm in the analysis of the construction site accidents

Słowa kluczowe: teren budowy, bezpieczeństwo pracy, drzewa decyzyjne

Key words: construction site, safety of works, decision trees

syfikatorów statystycznych w postaci drzew decyzyjnych (klasyfikacyjnych).

Metoda badań

Drzewa klasyfikacyjne wykorzystuje się do wyznaczania przynależności przypadków lub obiektów do klas jakościowej zmiennej zależnej na podstawie pomiarów jednej lub więcej zmiennych objaśniających – predyktorów (Breiman, Friedman, Stone and Olshen, 1984; Agrawal, Imielinski i Swami, 1993). Inaczej mówiąc, drzewa decyzyjne starają się podzielić dane na jak najbardziej jednorodne grupy względem zmiennej zależnej. Należy zaznaczyć, że przy drzewach decyzyjnych nie występuje formalne testowanie statystycznej istotności zmiennych. Drzewa klasyfikacyjne dają się prosto przedstawiać graficznie, co sprawia, że są łatwiejsze w interpretacji niż czysto liczbowe wyniki.

Wprowadzenie

Teren budowy i jego elementy stwarzają okoliczności, które sprzyjają powstawaniu zagrożeń bezpieczeństwa pracy w realizacji robót budowlanych. Analizy wskazują na decydujące znaczenie tych czynników w zbiorze cech opisujących przyczyny wypadków w budownictwie (Hoła, 2008; 2016; Drozd, 2015). W niniejszym artykule podjęto próbę analizy czynników związanych z terenem budowy w celu wskazania ich znaczenia w definiowaniu skutków zaistnienia wypadków. W zakresie metodologicznym wykorzystano metodę analiz zasobów danych za pomocą kla-

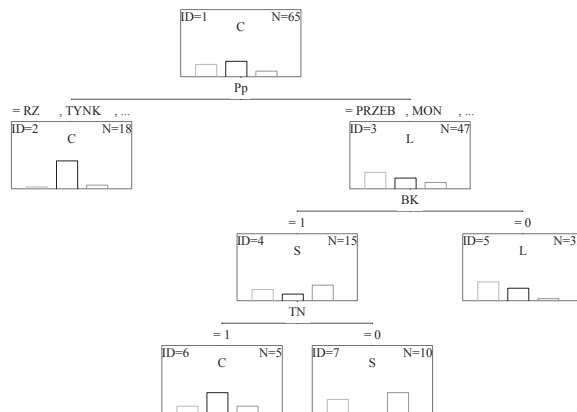
W niniejszym artykule analizę przeprowadzono dla danych o wypadkach budowlanych w Małopolsce, zarejestrowanych przez Okręgowy Inspektorat Pracy w Krakowie w latach 2014–2016. Liczba obserwacji wynosiła 65, zmienną zależną był skutek wypadku (*Sw*), a zmiennymi objaśniającymi (predyktorami): *Mw*, *Pp*, *Cmc*, *Cmo*, *Cmu*, *Cw*, *Wdo*, *Wyu*, oraz zmienne kodujące przyczynę wypadku, gdzie: *Mw* – miejsce powstania wypadku, *Pp* – proces pracy,

Wyu – wydarzenie powodujące uraz.

Przy tworzeniu drzewa wykorzystano algorytm CART¹ (Breiman i inni, 1984).

Wyniki

Na rysunku 1 przedstawiono drzewo dla zmiennej zależnej *Sw* (skutek wypadku: L – lekki, C – ciężki, S – śmiertelny) w badanym obszarze. Predyktorami (zmiennymi objaśniającymi) są: *Mw*, *Pp*, *Cmc*, *Cmo*, *Cmu*, *Cw*, *Wdo*, *Wyu*.



RYSUNEK 1. Drzewo dla zmiennej *Sw* (opracowanie własne)
FIGURE 1. Tree for variable *Sw* (own study)

Cmc – czynnik materialny związany z czynnością wykonywaną przez uszkodzowanego w chwili wypadku,
Cmo – czynnik materialny związany z odchyleniem,
Cmu – czynnik materialny będący źródłem urazu,
Cw – czynność wykonywana przez uszkodzowanego w chwili wypadku,
Wdo – wydarzenie będące odchyleniem od stanu normalnego: *TN* – tolerowanie przez nadzór odstępstw od zasad BHP, *BK* – brak wymaganych kwalifikacji (bez badań lekarskich, bez instruktażu),

Litera pojawiająca się w węźle oznacza, jaki wypadek (L – lekki, C – ciężki i S – śmiertelny) został w tym węźle wyznaczony, np. w węźle 6 (ID 6) wybrany został wypadek ze skutkiem ciężkim. Wyniki w węzłach końcowych drzewa zawarto w tabeli 1.

¹CART (ang. *classification and regression tree*). Metoda powstała w 1984 roku. Ma postać ciągu pytań, na które odpowiedzi determinują kolejne pytania, bądź kończą etap. W wyniku otrzymujemy strukturę drzewa, która w węzłach końcowych nie zawiera już pytań, lecz same odpowiedzi.

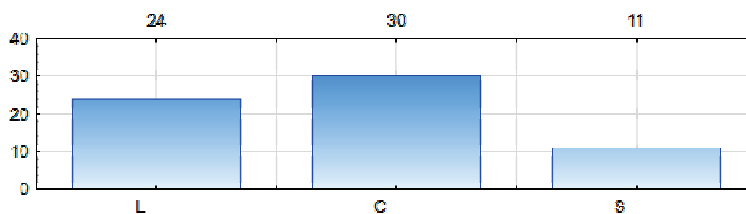
TABELA 1. Wyniki w węzłach końcowych (opracowanie własne)

TABLE 1. The results of the terminal nodes (own study)

ID Węzła Node ID	Klasa – L Class – L	Klasa – C Class – C	Klasa – S Class – S	Łącznie Together
2	1	15	2	18
6	1	3	1	5
7	4	0	6	10
5	18	12	2	32

Węzeł 1 (ID 1). Wśród wszystkich 65 wypadków najczęściej występowały wypadki skutkujące ciężkimi obrażeniami, najrzadziej zaś doszło do wypadków śmiertelnych. Dokładny rozkład przedstawiono na rysunku 2.

wały wypadki ciężkie (15 przypadków). W przypadku robót z węzła 3 (PRZEB – przebudowa, MON – montaż, DESK – deskowanie, SCH – schodzenie, DEM – demontaż, DACH – dachowe, INST – instalacyjne, TRANSP



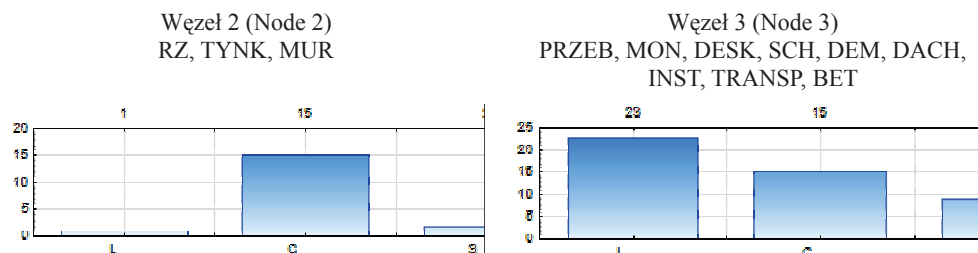
RYSUNEK 2. Rozkład wypadków w węźle 1 (opracowanie własne)

FIGURE 2. Distribution of accidents at Node 1 (own study)

Węzły 2 (ID 2) i 3 (ID 3). Powstały one po podziale węzła 1 (ID 1) ze względu na proces pracy (*Pp*). Podział ten przedstawiono na rysunku 3. Z rysunku 3 możemy odczytać, że przy robotach ziemnych (RZ), tynkarskich (TYNK) i murowych (MUR) najczęściej występo-

– transport ręczny, BET – betonowe) najczęściej było wypadków lekkich (23 przypadki), najmniej śmiertelnych (9 przypadków).

Węzły 4 (ID 4) i 5 (ID 5). Powstały one po podziale węzła 3 (ID 3) ze względu na brak kwalifikacji (zmienna *BK*).



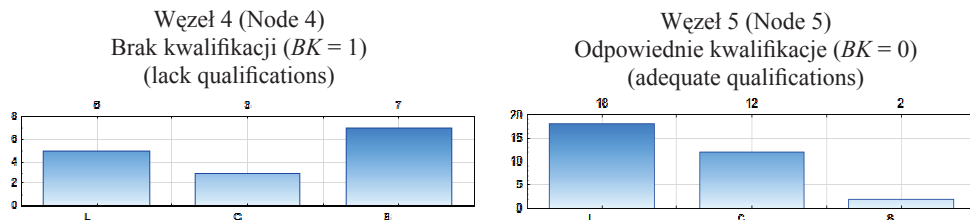
RYSUNEK 3. Podział węzła 1 i rozkład wypadków (opracowanie własne)

FIGURE 3. The division of Node 1 and the distribution of accidents (own study)

Podział ten przedstawiono na rysunku 4, z którego wynika, że osoby wykonujące czynności zakwalifikowane do węzła 3 (PRZEB, MON, DESK, SCH, DEM, DACH, INST, TRANSP, BET) w przypadku braku kwalifikacji najczęściej ulegają wypadkom ze skutkiem śmiertelnym. Z kolei osobom z odpowiednimi kwalifikacjami najczęściej zdrażają się wypadki ze skutkiem lekkim (18 przypadków), najrzadziej ze skutkiem śmiertelnym (2 przypadki).

od zasad BHP, a pracownicy nie posiadają odpowiednich kwalifikacji, najczęściej dochodzi do wypadków śmiertelnych (6 przypadków) i urazów lekkich (sporadycznie).

Do oceny poprawności klasyfikacji zmiennych służy macierz klasyfikacji (tab. 2 i 3). Można ją utworzyć dla dowolnego rodzaju modeli klasyfikacyjnych. Dla przypadków występujących w zbiorze danych porównuje ona klasy wskazywane przez model z rzeczywistością

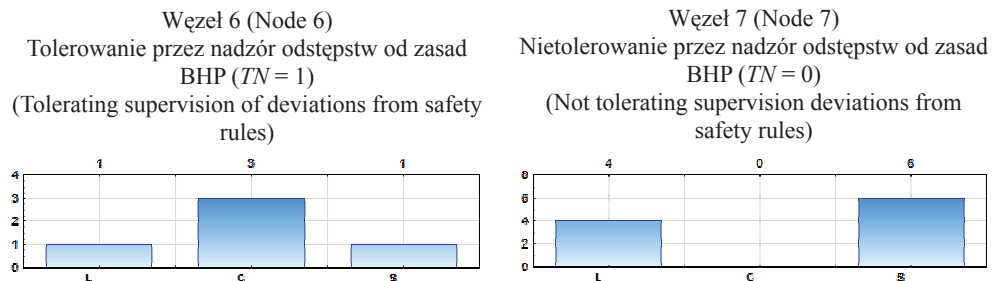


RYSUNEK 4. Podział węzła 3 i rozkład wypadków (opracowanie własne)
 FIGURE 4. The division of Node 3 and the distribution of accidents (own study)

Węzły 6 (ID 6) i 7 (ID 7). Powstały one po podziale węzła 4 (ID 4) ze względu na zmienną *TN* (tolerowanie przez nadzór odstępstw od zasad BHP). Podział ten przedstawiono na rysunku 5. Z rysunku 5 wynika, że brak kwalifikacji i tolerowanie przez nadzór odstępstw od zasad BHP najczęściej skutkuje wypadkami o skutku ciężkim. Z kolei w sytuacji, gdy nadzór nie toleruje odstępstw

zaobserwowanymi. Na tej podstawie można uzyskać nie tylko informację o ogólnej trafności przewidywań, ale też o najczęściej popełnianych rodzajach błędów.

Z danych z tabeli 3 wynika, że łącznie pojawiły się 24 wypadki ze skutkiem lekkim, z czego model poprawnie sklasyfikował 18, co stanowi 75%. Wypadki ciężkie wystąpiły 30 razy, z czego



RYSUNEK 5. Podział węzła 4 i rozkład wypadków (opracowanie własne)
 FIGURE 5. The division of Node 4 and the distribution of accidents (own study)

TABELA 2. Macierz klasyfikacji (opracowanie własne)
TABLE 2. Classification matrix (own study)

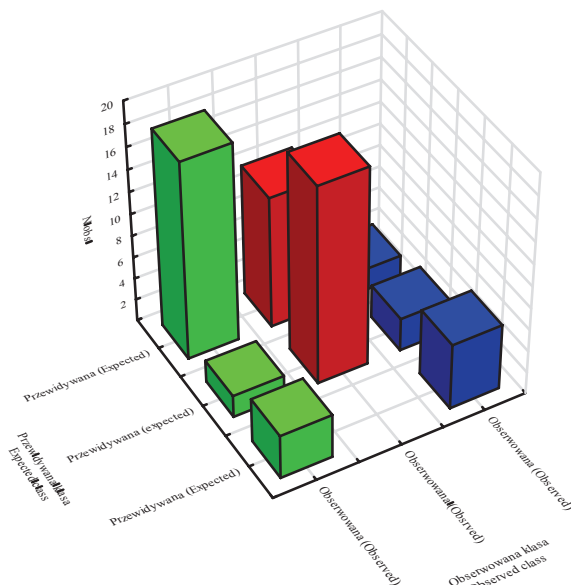
×	Obserwowana L Observed L	Obserwowana C Observed C	Obserwowana S Observed S
Przewidywana L Predicted L	18	12	2
Przewidywana C Predicted C	2	18	3
Przewidywana S Predicted S	4	0	6

TABELA 3. Macierz klasyfikacji (opracowanie własne)
TABLE 3. Classification matrix (own study)

×	Przewidy- wana L Predicted L	Przewidy- wana C Predicted C	Przewidy- wana S Predicted S	Łącznie w wierszu Together
Obserwowana L Observed L	18	2	4	24
Udział z kolumny Share of the column [%]	56,25	8,70	40,00	–
Udział z wiersza Share of the row [%]	75,00	8,33	16,67	–
Udział z ogółu Share of the whole [%]	27,69	3,08	6,15	36,92
Obserwowana C Observed C	12	18		30
Udział z kolumny Share of the column [%]	37,50	78,26	0,00	–
Udział z wiersza Share of the row [%]	40,00	60,00	0,00	–
Udział z ogółu Share of the whole [%]	18,46	27,69	0,00	46,15
Obserwowana S Observed S	2	3	6	11
Udział z kolumny Share of the column [%]	6,25	13,04	60,00	–
Udział z wiersza Share of the row [%]	18,18	27,27	54,55	–
Udział z ogółu Share of the whole [%]	3,08	4,62	9,23	16,92
Liczba Quantity	32	23	10	65
Łącznie Together [%]	49,23	35,38	15,38	–

poprawnie sklasyfikowanych zostało 18, co stanowi 60%. Z kolei z 11 wypadków śmiertelnych poprawnie rozpoznanych zostało 6, co stanowi 54,55%. Na histogramie, na rysunku 6 przedstawiono wartości obserwowane i przewidywane dla danego modelu.

Drzewa klasyfikacyjne są dobrym narzędziem do wyznaczania przynależności przypadków lub obiektów do klas jakościowej zmiennej zależnej na podstawie pomiarów zmiennych objaśniających (predyktorów) i podziału ich na jak najbardziej jednorodne grupy względem



RYSUNEK 6. Macierz klasyfikacji. Wartości obserwowane i przewidywane (opracowanie własne)
 FIGURE 6. Classification matrix. The values of the observed and predicted (own study)

Podsumowanie

Obszarem badań objęto tereny budów Małopolski, na których w latach 2014–2016 miały miejsce wypadki przy pracy. Zarejestrowane przez Okręgowy Inspektorat Pracy w Krakowie obserwacje pozwoliły dokonać analizy, której celem była eksploracja wpływu wybranych czynników terenu budowy na skutki zaistnienia wypadków przy pracy. Wykorzystano w tym celu klasyfikator statystyczny w postaci drzewa decyzyjnego (klasyfikacyj-

tej zmiennej. Identyfikacja i profilowanie wzorców wypadków oraz prognozowanie czynników zagrożenia bezpieczeństwa pracy pozwoliły osiągnąć założony cel. Podjęte w artykule analizy wpływu czynników związanych z terenem budowy na bezpieczeństwo realizacji robót stanowią propozycję powiększenia zbioru wiedzy w zakresie rozwoju naukowych metod oceny zagrożeń na budowach oraz w zakresie możliwości wykorzystania ich w praktyce dla poprawy warunków bezpiecznej realizacji robót budowlanych.

Reguły wynikające ze sporządzonego drzewa przedstawiają się następująco:

- Jeżeli zmienna Pp (proces pracy) przyjmuje wartość ze zbioru {RZ, TYNK, MUR}, to najczęściej skutek wypadku jest ciężki (ID 2).
- Jeżeli zmienna Pp przyjmuje wartość ze zbioru {PRZEB, MON, DESK, SCH, DEM, DACH, INST, TRANSP, BET} i pracownik posiada odpowiednie kwalifikacje, to najczęściej skutek wypadku jest lekki (ID 5).
- Jeżeli zmienna Pp przyjmuje wartość ze zbioru {PRZEB, MON, DESK, SCH, DEM, DACH, INST, TRANSP, BET} i pracownik nie posiada odpowiednich kwalifikacji, a zmienna TN (tolerowanie przez nadzór odstępstw od zasad BHP) przyjmuje wartość 1, to najczęściej skutek wypadku jest ciężki (ID 6).
- Jeżeli zmienna Pp przyjmuje wartość ze zbioru {PRZEB, MON, DESK, SCH, DEM, DACH, INST, TRANSP, BET} i pracownik nie posiada odpowiednich kwalifikacji, a zmienna TN przyjmuje wartość 0, to najczęściej skutek wypadku jest śmiertelny (ID 7).

Wyniki badań uwidaczniają, że najczęściej do wypadków ze skutkiem ciężkim dochodzi przy robotach ziemnych (RZ), tynkarskich (TYNK) i murowych (MUR). Niekorzystny skutek wypadków potęgują brak kwalifikacji zatrudnionych oraz tolerowanie przez nadzór odstępstw od zasad BHP. Do grupy najsilniej powiązanych ze skutkiem wypadku zmiennych należą: Cmo (czynnik materialny związany z odchyleniem), Cmu (czynnik materialny będący źródłem urazu), Cmc (czynnik materialny związany z czyn-

nością wykonywaną przez poszkodowanego w chwili wypadku), Wdo (wydarzenie będące odchyleniem od stanu normalnego) i Wyu (wydarzenie powodujące uraz). Wśród nich istotne czynniki to: brak środków ochrony zbiorowej i indywidualnej, samowolne zachowanie pracownika, utrata kontroli nad obsługiwaną maszyną, wykonywanie prac bez usunięcia zagrożeń i stan psychofizyczny pracownika.

Literatura

- Agrawal, R., Imielinski, T. i Swami, A. (1993). Mining association rules between sets of items in large database. *Proceedings of the ACM SIGMOD Conference on Management of Data*, 207-216.
- Breiman, L., Friedman, J., Stone, C.J. i Olshen, R.A. (1984). Classification and Regression Trees. *Wadsworth Statistics/Probability*.
- Drozd, W. (2015). Regression analysis of accident absenteeism and variables describing working conditions. W: E. Plebankiewicz (red.) *Recent advances in civil engineering*. Monography 480. Construction Management – Inżynieria Łądowa, Politechnika Krakowska (s. 13-27).
- Hoła, B. (2008). *Modelowanie jakościowe i ilościowe wypadkowości w budownictwie*. Wrocław: Oficyna Wydawnictwo Politechniki Wrocławskiej.
- Hoła, B. (2016). *Bezpieczeństwo pracy w procesach budowlanych*. Wrocław: Wydawnictwo Politechnika Wrocławska.

Streszczenie

Drzewa decyzyjne w analizie wypadkowości w budownictwie. Teren budowy i jego elementy stwarzają okoliczności, które sprzyjają powstawaniu zagrożeń bezpieczeństwa pracy w realizacji robót budowlanych. Analizy wskazują na decydujące znaczenie

tych czynników w zbiorze cech opisujących przyczyny wypadków w budownictwie. W artykule podjęto próbę analizy cech terenu budowy w celu wskazania ich znaczenia w definiowaniu okoliczności wypadku przy pracy. Badania zrealizowano na podstawie danych z rejestru prowadzonego przez Okręgowy Inspektorat Pracy w Krakowie. Zasadnicze zadania merytoryczne obejmują wyodrębnienie wzorców wypadków na terenie budowy i wskazanie na te spośród analizowanych cech, które mają istotne znaczenie w definiowaniu tych wzorców. W zakresie metodologicznym wykorzystano analizę zasobów danych za pomocą klasyfikatorów statystycznych – drzew decyzyjnych.

Summary

Decision trees algorithm in the analysis of the construction site accidents. The construction site and its elements create circumstances that are conducive to the formation of risks to work safety during the execution of works. Analysis indicates

the critical importance of these factors in the set of characteristics that describe the causes of accidents in the construction industry. This article attempts to analyse the characteristics of the construction site to indicate their importance in defining the circumstances of an accident at work. The research was carried out on the basis of data from the register kept by the District Labour Inspectorate in Kraków. Main substantive tasks include isolating patterns of accidents on the site and identifying those of the analysed characteristics that are important in defining these patterns. In terms of methodology, the dissertation presents method of analysing data resources: using statistical classifiers – decision trees algorithm.

Authors' address:

Wojciech Drozd, Marcin Kowalik
Politechnika Krakowska
Instytut Zarządzania w Budownictwie
i Transporcie
ul. Warszawska 24
31-155 Kraków, Poland
e-mail: wdrozd@zob.pk.edu.pl
mkowalik@zob.pk.edu.pl