

ANALIZA DANYCH ZE STACJI KONTROLI POJAZDÓW ZA POMOCĄ PROGRAMU STATISTICA DATA MINER

Marek Klimkiewicz, Michał Molenda

Katedra Organizacji i Inżynierii Produkcji, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Streszczenie. W artykule zaprezentowano możliwości jakie daje wykorzystanie programu STATISTICA Data Miner firmy StatSoft w przeprowadzaniu analizy dużych zbiorów danych na przykładzie eksploracji danych gromadzonych w czasie kontroli okresowej pojazdów. Otrzymane wyniki pozwoliły na dokonanie ogólnej charakterystyki kontrolowanych pojazdów. Dokonano klasyfikacji pojazdów pod kątem popularności marek, modeli, oraz ich usterkowości.

Słowa kluczowe: kontrola okresowa pojazdów, data mining, diagnostyka

Wprowadzenie i cel pracy

Badania techniczne są podstawowym elementem systemu zapewnienia właściwego stanu technicznego pojazdów poruszających się po drogach publicznych [Bocheński 2000]. Podczas badań technicznych pojazdów uzyskuje się wiele informacji, które odpowiednio przetworzone, mogą służyć poprawie bezpieczeństwa ruchu drogowego, a także mogą być wykorzystane przez właściciela SKP.

Celem pracy jest wykonanie analizy danych ze stacji kontroli pojazdów z wykorzystaniem programu STATISTICA Data Miner. Zgłębianie danych (ang. Data Mining), jest to proces analityczny przeznaczony do badania dużych zbiorów danych w poszukiwaniu regularnych wzorców oraz systematycznych współzależności pomiędzy zmiennymi, a następnie do oceny wyników poprzez zastosowanie wykrytych wzorców do nowych podzbiorów danych [Tadeusiewicz 2006]. Proces zgłębiania danych, umożliwiający opis zależności znalazł zastosowanie w wielu dziedzinach życia, m.in. w medycynie, bankowości, przemyśle i innych.

Opis metody badawczej

Dane empiryczne gromadzone były na Stacji Kontroli Pojazdów Katedry Organizacji i Inżynierii Produkcji SGGW w Warszawie, jako archiwum programu PATRONAT. Dane dotyczą pojazdów badanych na stacji w okresie od 4 maja 2004 do 31 grudnia 2007 roku. Z danych, które można importować z programu PATRONAT, do analizy wybrano następujące: marka, typ i model pojazdu; rodzaj pojazdu; data pierwszej rejestracji w kraju lub za granicą; data badania; numer rejestracyjny; numer identyfikacyjny (VIN); wynik badania; data następnego badania. Oprócz tego z kart rejestracyjnych wyświetlanych przez

program wybrano takie dane jak: seria i numer dowodu rejestracyjnego, numer silnika, diagnosta, organ wydający, nadawanie numeru lub tabliczka, dokument identyfikacyjny, informacje dotyczące dokonania zmian, zasilanie gazem czy też zatrzymanie dowodu rejestracyjnego. Na tej podstawie została stworzona baza danych w postaci arkusza programu Statistica.

Na etapie przygotowania danych sprawdzono poprawność zapisów, dokonano przekształceń i wybrano zmienne oraz przeprowadzono wstępne analizy. Ujednolicono zapisy, gdyż diagności wprowadzali niektóre zapisy w sposób niejednoznaczny (np. Lanos KAT i Lanos 1,6 występujący w bazie to ten sam model, pozostawiono więc tylko nazwę Lanos 1,6).

Na podstawie danych zawartych w bazie programu PATRONAT dotyczących daty pierwszej rejestracji pojazdu oraz daty ostatniego przeglądu utworzone zostały zmienne: rok produkcji oraz wiek pojazdu podczas badania na Stacji Kontroli Pojazdów. Ponadto na podstawie współwystępowania, w przypadku niektórych pojazdów, daty rejestracji w kraju i za granicą, stworzona została zmienna „CZY POJAZD SPROWADZONY?”.

Do analizy wybrano 27 zmiennych, które można podzielić na trzy grupy:

- zmienne dotyczące pojazdu: marka; model; typ handlowy; rodzaj pojazdu; numer rejestracyjny; data pierwszej rejestracji w kraju; data pierwszej rejestracji za granicą; data pierwszej rejestracji; czy pojazd sprowadzony?; rok produkcji oraz wiek pojazdu w momencie badania;
 - zmienne dotyczące badania: data badania, rodzaj badania, wynik badania oraz termin następnego badania;
 - zmienne dotyczące wykrytych podczas badania usterek: oświetlenie pojazdu, układ wydechowy; układ kierowniczy i zawieszenia; układ hamulcowy; układ paliwowy; instalacja gazowa; uszkodzenia nadwozia; ogumienie; brak trójkąta ostrzegawczego; brak gaśnicy/nie legalizowana; wycieki płynów eksploatacyjnych oraz inne.
- Wśród zmiennych znalazły się zarówno dane ilościowe jak i jakościowe.

Wyniki badań

Analizy wykonywane były pod kątem następujących pytań:

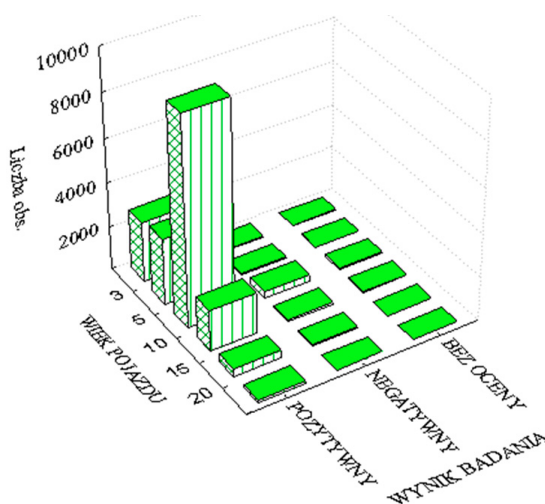
- Czy wiek pojazdu wpływa na wynik okresowego badania technicznego?
- Czy importowane używane pojazdy charakteryzują się większą usterkowością niż nabyte w kraju?
- Jakie są najczęstsze przyczyny wydania przez stację kontroli pojazdów negatywnej opinii dotyczącej stanu technicznego?
- Czy istnieje zależność między wynikiem badania technicznego, a marką i modelem badanego pojazdu?

Z szerokiego spektrum metod dostępnych w STATISTICA Data Miner wybrano te, które pozwoliły dać odpowiedź na powyższe pytania.

Na rysunku 1 przedstawiono zależność między wiekiem pojazdu a wynikiem badania stanu technicznego. Średnia wieku dla wszystkich samochodów przebadanych na SKP SGGW wynosiła 7,25 a odchylenie standardowe 3,56 lat, natomiast średnia wieku dla samochodów uznanych za spełniające wymagania wynosiła 7,2 a odchylenie standardowe 3,51 lat. Dla samochodów nie spełniających wymagań średnia wynosiła 8,82 a odchylenie

Analiza danych...

standardowe 4,28. Tak jak się można było spodziewać, starsze pojazdy charakteryzowały się większą usterkowością niż pojazdy eksploatowane krócej. Z analizy wynika, że wśród badanych pojazdów dominowały te z przedziału wiekowego 5-10 lat.



Źródło: opracowanie własne autorów

Rys. 1. Zależność między wiekiem pojazdu a wynikiem badania stanu technicznego

Fig. 1. Dependence between vehicle age and the result of technical condition check

W tabeli 1 przedstawiono wyniki oceny pojazdów przeprowadzonej przez SKP dla pojazdów zakupionych w kraju i pojazdów używanych, sprowadzonych z zagranicy. Jak widać w tabeli 3, udział pojazdów sprowadzonych z zagranicy stanowi zaledwie nieco ponad 7% wszystkich pojazdów, które poddane były ocenie stanu technicznego przez SKP SGGW w omawianym okresie czasu. W grupie samochodów używanych sprowadzonych z zagranicy ocenę negatywną uzyskało około 3,6% pojazdów, co stanowi zaledwie 0,26% ogółu samochodów, natomiast wśród grupy samochodów nabytych w kraju ocenę negatywną uzyskało 2,75%, co stanowi około 2,55% ogółu badanych pojazdów.

W tabeli 2 przedstawiono usterki i niedomagania wykryte w badanych pojazdach uznanych za niespełniające wymagania technicznie, ich licznosc oraz udział procentowy w ogólnej liczbie pojazdów uznanych za niesprawne.

Biorąc pod uwagę procentowy udział pojazdów z różnymi usterkami w ogólnej liczbie pojazdów, które uzyskały negatywną ocenę podczas badania ich stanu technicznego, można zauważyć, że najliczniej występują usterki układów wpływających na bezpieczeństwo ruchu drogowego. Niemal co drugi pojazd niesprawny technicznie miał usterkę układu hamulcowego lub układu oświetlenia pojazdu. Zjawisko to może napawać niepokojem, gdyż pojazdy takie stwarzają duże zagrożenie ruchu drogowego.

Tabela 1. Wpływ pochodzenia pojazdu na wynik badania stanu technicznego
Table 1. Influence of vehicle origin on the result of investigation of technical state

Pochodzenie pojazdu		Wynik badania			Razem
		pozytywny	negatywny	bez oceny	
Nabyty w kraju	[szt.]	1259	48	18	1325
	[%]	95,02	3,62	1,36	100
	% ogółu	6,88	0,26	0,10	7,24
Sprowadzony z zagranicy	[szt.]	16318	467	181	16966
	[%]	96,18	2,75	1,07	100
	% ogółu	89,21	2,55	0,99	92,76
Badane pojazdy ogółem	[szt.]	17577	515	199	18291
	[%]	96,10	2,82	1,09	100,00

Źródło: obliczenia własne autorów

Aby ocenić popularność poszczególnych marek i modeli samochodów dane uzyskane ze stacji kontroli pojazdów zostały poddane analizie pod kątem liczebności poszczególnych marek i modeli, które były badane na SKP SGGW.

Tabela 2. Rodzaje usterek oraz ich udział w ogólnej liczbie pojazdów niesprawnych technicznie
Table 2. The kinds of defects and their share in general number of faulty vehicles

Lp.	Rodzaj usterki	Liczba odnotowanych przypadków	Udział pojazdów z wynikiem negatywnym
1	Oświetlenie pojazdu	253	49,1
2	Układ hamulcowy	253	46,8
3	Układ wydechowy	159	30,9
4	Układ kierowniczy i zawieszenia	140	27,2
5	Uszkodzenia nadwozia	82	16,9
6	Inne	70	13,6
7	Brak gaśnicy/nie legalizowana	43	8,4
8	Instalacja gazowa	40	7,8
9	Wycieki płynów eksploatacyjnych	31	6
10	Układ paliwowy	13	2,5
11	Ogumienie	10	1,9
12	Brak trójkąta ostrzegawczego	5	0,97

Źródło: obliczenia własne autorów

Na podstawie tej analizy wybrano 10 marek i modeli pojazdów występujących najczęściej, które przedstawiono w tabeli 3. Najczęściej badanymi markami samochodów były Fiat i Opel. Marki te stanowią ponad 20% wszystkich pojazdów badanych na stacji kontroli pojazdów, w rozpatrywanym okresie. Dziesięć najliczniej odnotowywanych marek pojazdów stanowi około 75% badanych pojazdów, przy czym należy zauważyć iż podczas analizy wyróżnione zostały pojazdy aż 83 marek. W przypadku modeli nie widać aż tak dużej dominacji jakiegokolwiek z nich, jednak ponad 25 procentowy udział dziesięciu modeli spośród 672 modeli wyróżnionych podczas analizy, świadczy o wysokim zainteresowaniu

Analiza danych...

klientów właśnie tą grupą samochodów. Na podstawie liczby poszczególnych modeli pojazdów badanych na SKP można stwierdzić, że największy jest udział aut miejskich lub klasy kompakt.

Tabela 3. Najpopularniejsze marki i modele pojazdów
Table 3. The most popular vehicle makes and models

Lp.	Marka	Najczęściej badane marki		Model	Najczęściej badane modele	
		Liczba	Udział procentowy		Liczba	Udział procentowy
1	Fiat	2164	11,83	Corolla	815	4,46
2	Opel	1967	10,75	Astra	702	3,84
3	Toyota	1628	8,90	Civic	526	2,88
4	Daewoo	1520	8,31	Lanos	498	2,72
5	Ford	1467	8,02	Felicia	481	2,63
6	Skoda	1173	6,41	Corsa	463	2,53
7	Renault	1087	5,94	Seicento	439	2,40
8	Volkswagen	935	5,11	Almera	389	2,13
9	Nissan	862	4,71	Punto	352	1,92
10	Peugeot	760	4,15	Matiz	351	1,92
	Ogółem	13563	74,15	Ogółem	5016	27,43

Źródło: obliczenia własne autorów

Oznacza to, że w okręgu działania SKP ten typ pojazdów najczęściej znajduje nabywców. Nieco dalsze miejsca w tej klasyfikacji zajęły samochody klasy średniej wyższej, takie jak Ford Mondeo, Toyota Avensis czy też Opel Vectra. Są to modele, które upodobali sobie pracownicy kadr kierowniczych wielu firm. Wskazuje to, że w tej części miasta, w której działa SKP SGGW zamieszkuje dość licznie właśnie ta grupa zawodowa.

W ostatniej części analizy oceniono najczęściej badane marki i modele samochodów ze względu na ich usterkowość. Wyniki oceny przedstawione zostały w tabeli 4. Zarówno marki jak i modele pojazdów, które najczęściej były badane w SKP uszeregowano według rosnącego procentowego udziału pojazdów o negatywnej ocenie, uzyskanej podczas badania, w całkowitej liczbie samochodów danej marki lub danego modelu. Wśród marek samochodów najmniej usterkowymi okazały się Nissan, Skoda oraz Volkswagen. W tej grupie udział pojazdów z usterkami nie przekroczył 2%. Pozostałe siedem wyszczególnionych marek uzyskało wyniki z przedziału 2-3,5%, co również oznacza niski poziom usterkowości. W przypadku modeli, najmniej awaryjnymi okazały się modele Nissan Almera oraz Skoda Felicia, wśród których w przybliżeniu tylko jeden na sto pojazdów nie przeszedł pomyślnie badania okresowego. Wśród najczęściej badanych modeli najbardziej awaryjnym okazał się Daewoo Matiz. W przypadku modeli Matiz pojazdy niesprawne stanowiły ponad 5% ogólnej liczby badanych aut tego modelu.

Tabela 4. Awaryjność najpopularniejszych marek i modeli pojazdów
 Table 4. The faultability of the most popular makes and the models of vehicles

Lp.	Marka	Wyniki oceny marki [%]			Model	Wyniki oceny modelu [%]		
		Negatywny	Pozytywny	Bez oceny		Negatywny	Pozytywny	Bez oceny
1	Nissan	1,39	97,68	0,93	Almera	1,03	98,71	0,26
2	Skoda	1,62	97,18	1,19	Felicia	1,04	98,13	0,83
3	Volkswagen	1,93	97,54	0,53	Civic	1,71	98,29	0,00
4	Toyota	2,15	96,74	1,11	Corolla	1,84	97,30	0,86
5	Opel	2,69	96,34	0,97	Corsa	2,16	97,19	0,65
6	Peugeot	2,76	95,79	1,45	Lanos	2,21	97,39	0,40
7	Citroen	2,88	95,78	1,34	Punto	2,56	97,16	0,28
8	Renault	2,94	96,23	0,83	Astra	2,71	96,58	0,71
9	Fiat	3,37	95,75	0,88	Seicento	3,42	95,67	0,91
10	Ford	3,41	95,43	1,16	Matiz	5,13	94,87	0,00

Źródło: obliczenia własne autorów

Podsumowanie

Wyniki analizy danych zgromadzonych podczas okresowych badań kontrolnych pojazdów na SKP SGGW przeprowadzonej za pomocą programu Statistica Data Miner, pozwoliły na uzyskanie cennych informacji o stanie pojazdów samochodowych eksploatowanych w okręgu działania SKP. Na podstawie przedstawionych wyników można stwierdzić, iż wśród eksploatowanych pojazdów silnikowych większość to samochody osobowe w wieku od 5 do 10 lat. Tak jak można było oczekiwać pojazdy starsze charakteryzowały się dużą usterkowością. Również usterkowość pojazdów używanych sprowadzonych z zagranicy była wyższa niż nabytych w kraju. Wśród usterek, które wpływały na negatywną ocenę podczas kontroli na SKP SGGW najliczniejsze były usterki mające wpływ na bezpieczeństwo ruchu drogowego.

Przedstawione wyniki mogą być wykorzystane przez ośrodki decyzyjne nadzorujące stacje kontroli pojazdów, ośrodki odpowiedzialne za bezpieczeństwo ruchu drogowego, właściciela SKP, a także użytkowników nabywających pojazdy. Wykonanie takiej analizy w skali całego kraju na podstawie danych ze wszystkich SKP pozwoliłoby uzyskać opinię o stanie polskiej motoryzacji.

Bibliografia

- Bocheński C.** (red.). 2000. Badania kontrolne Samochodów. WKŁ, Warszawa. ISBN 83-206-1349-3.
Tadeusiewicz R. 2006. Data Mining jako szansa na relatywnie tanie dokonywanie odkryć naukowych poprzez przekopywanie pozornie całkowicie wyeksploatowanych danych empirycznych. Statystyka i Data Mining w badaniach naukowych. StatSoft, Kraków. ISBN-13: 978-83-88724-31-2.

ANALYSIS OF DATA FROM VEHICLE INSPECTION STATION USING THE STATISTICA DATA MINER APPLICATION

Abstract. The article presents potential of using the STATISTICA Data Miner application from StatSoft to carry out analyses of large data sets on the example of exploring data being acquired during routine inspection of vehicles. Obtained results allowed to make general characteristic of inspected vehicles. The vehicles were classified according to popularity of makes, models, and their defect rates.

Key words: routine inspection of vehicles, data mining, diagnostics

Adres do korespondencji:

Marek Klimkiewicz; e-mail: marek_klimkiewicz@sggw.pl
Katedra Organizacji i Inżynierii Rolnictwa
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
ul. Nowoursynowska 166
02-787 Warszawa