Robert Kozakiewicz Gdańsk, 30.05.2015

Dawid Piasecki

Eksploracja danych

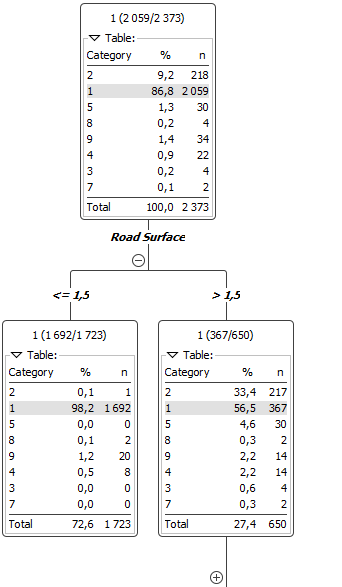
Zrozumienie warunków biznesowych i identyfikacja źródeł danych

**Cele biznesowe eksploracji danych.**

a) Jaka jest charakterystyka kierowcy, który uczestniczy w wypadku?

b) Jaka jest charakterystyka warunków pogodowych podczas wypadków?

W rozwiązaniu tego celu wykorzystaliśmy danę ze źródła o wypadkach. Kolumny "Road Surface" i "Weather Conditions". Wybrany model to drzewo decyzyjnę zależności warunków pogodowych od powierzchni. Atrybutem nominalnym uczyniliśmy tylko jeden z dwóch a mianowicie "Weather Conditions", po to by można było uzależnić warunki pogodowe od kilku rodzaji dróg.



|  |  |  |  |
| --- | --- | --- | --- |
| Road Surface: | | | |
| 1 - Dry |  |  |  |
| 2 - Wet / Damp | |  |  |
| 3 - Snow |  |  |  |
| 4 - Frost / Ice | |  |  |
| 5 - Flood (surface water over 3cm deep) | | | |
| |  |  | | --- | --- | | Weather Conditions: | | | 1 - Fine without high winds | | | 2 - Raining without high winds | | | 3 - Snowing without high winds | | | 4 - Fine with high winds | | | 5 - Raining with high winds | | | 6 - Snowing with high winds | | | 7 - Fog or mist – if hazard | | | 8 - Other |  | | 9 - Unknown | | | | | |

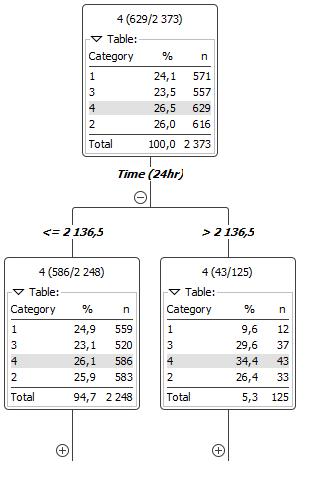
c) Jaka jest charakterystyka pojazdu, który uczestniczy w wypadku?

d) Jak pora dnia i data w roku wpływa na ilość wypadków?

Danę, które przyczyniły się do rozwiązania tego problemu to: kwartał roku, otrzymany za pomocą "Date Field Extractor" z kolumny "Accident Date" oraz godzina w której miał miejsce wypadek z kolumny "Time (24hr)" ze źródła o wypadkach. Wybrany model to drzewo decyzyjne zależności kwartału od godziny.

Godzina w formacię hhmm, gdzie minuty to koniecznie 2 cyfry np. 06, a godzina może być tylko jedną np. 152 oznacza 1:52.

Kwartały od 1-4.



e) Jak charakterystyka drogi(rodzaj drogi i oświetlenie) wpływa na ilość wypadków?