Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 15. 8. 2019

Základní informace o datech Accidents

Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

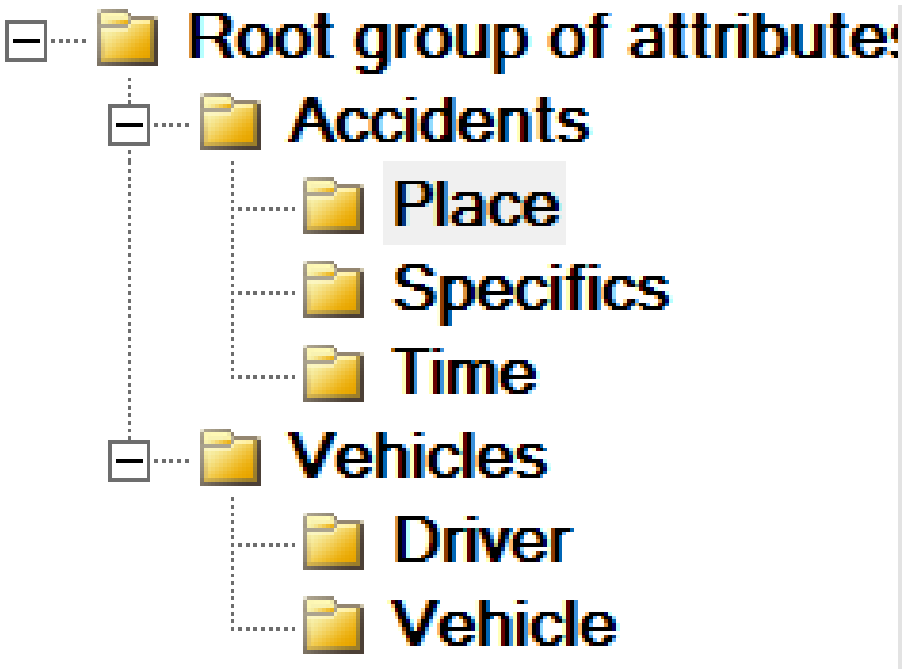
© Jan Rauch

Data Accidents - přehled (1)

- Jedná se o data vzniklá v rámci studentské seminární práce v předmětu *4IZ460 - pokročilé přístupy k DZD* v zimním semestru 2018/2019 na VŠE Praha.
- Výchozí data se skládají ze tří souborů dostupných na https://www.kaggle.com/silicon99/dft-accident-data
- Byly zpracovány pouze nehody jednoho vozidla vzhledem k jasné odpovědnosti řidiče.
- Vzniklá matice dat má 538 989 řádků

Data Accidents - přehled (2)

Vytvořené atributy rozděleny do skupin:



Data Accidents - atributy (1)

Groups of attributes tree

- Root group of attributes
  - Accidents
    - Place

Attribute	Used	DBColumn	Categories	XCat	Sample categories
Area_type	+	Urban_or_Rural_Ar	3		Urban, Rural, Unallocated

Groups of attributes tree

- Root group of attributes
  - Accidents
    - Place
    - Specifics
    - Time
  - Vehicles
    - Driver
    - Vehicle

Attribute	Used	DBColumn	Categories	XCat	Sample categories
1st_Road_Class	+	x_1st_Road_Class	6		Motorway, A(M), A, B, C, Unclassified
Accident_Severity	+	Accident_Severity	3		Fatal, Serious, Slight
Carriageway_Hazards	+	Carriageway_Hazar	6	x	None, Vehicle load on road, Other object on road, Previous accident, Pedestrian
Carriageway_Hazards 5 000+	+	Carriageway_Hazar	3	x	None, Other object on road, Any animal in carriageway
Did_Police_Officer_Attend_Scene_of_Acc		Did_Police_Officer_	3	x	Yes, No, Self rep only
Did_Police_Officer_Attend_Scene_of_Acc	+	Did_Police_Officer_	2	x	Yes, No
Junction_Detail	+	Junction_Detail	9	x	Not at junction or within 20 metres, Roundabout, Mini-roundabout, T or staggere
Light_Conditions	+	Light_Conditions	5		Daylight, Darkness - lights lit, Darkness - lights unlit, Darkness - no lighting, Dark
Number_of_Casualties		Number_of_Casualt	33		1, 2, 3, 4, 5, 6, 7, 8, 9, 10...
Number_of_Casualties_ef5		Number_of_Casualt	5		<1;2), <2;3), <3;4), <4;5), <5;68>
Number_of_Casualties_ef5 5000 +	+	Number_of_Casualt	3		<1;2), <2;3), <3;4)
Road_Surface_Conditions	+	Road_Surface_Con	5	x	Dry, Wet or damp, Snow, Frost or ice, Flood over 3cm. deep
Road_Surface_Conditions 5000 +	+	Road_Surface_Con	3	x	Dry, Wet or damp, Frost or ice
Speed_limit	+	Speed_limit	8		10, 15, 20, 30, 40, 50, 60, 70
Speed_limit _ed3	+	Speed_limit	3		<10;29>, <30;49>, <50;70>
Weather_Conditions	+	Weather_Conditions	9	x	Fine no high winds, Raining no high winds, Snowing no high winds, Fine + high v
Weather_Conditions 5 000 +	+	Weather_Conditions	8	x	Fine no high winds, Raining no high winds, Snowing, Fine + high winds, Raining

Groups of attributes tree

- Root group of attributes
  - Accidents
    - Place
    - Specifics
    - Time
  - Vehicles
    - Driver
    - Vehicle

Attribute	Used	DBColumn	Categories	XCat	Sample categories
Day_of_Week	+	Day_of_Week	7		Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
Hour	+	Time_.Hour	24	x	0, 1, 2, 3, 4, 5, 6, 7, 8, 9...
Month	+	Date_.Month	12		January, February, March, April , May, June, July, August , September, October.
Quarter	+	Date_.Quarter	4		Q1, Q2, Q3, Q4
Rush_hours	+	Time_.Hour	2	x	Other, 8-9/16-18
Weekend	+	Day_of_Week	2		Weekend, Weekdays
Year	+	Date_.Year	11		2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014...

Data Accidents - atributy (2)

Groups of attributes tree

- Root group of attributes
  - Accidents
    - Place
    - Specifics
    - Time
  - Vehicles
    - Driver
    - Vehicle

Attribute	Used	DBColumn	Categories	XCal	Sample categories
Age_Band_of_Driver		Age_Band_of_Drive	11	x	0 - 5, 6 - 10, 11 - 15, 16 - 20, 21 - 25, 26 - 35, 36 - 45, 46 - 55, 56 - 65, 66 - 75...
Age_Groups_of_Driver	+	Age_of_Driver	9	x	Under 18, 18 to 25, 26 to 35, 36 to 45, 46 to 55, 56 to 65, 66 to 75, 76 to 85, Ab
Age_of_Driver		Age_of_Driver	98	x	1, 2, 3, 4, 5, 6, 7, 8, 9, 10...
Age_of_Driver_ef20		Age_of_Driver	20		ef_1, ef_2, ef_3, ef_4, ef_5, ef_6, ef_7, ef_8, ef_9, ef_10...
Age_of_Driver_ef5	+	Age_of_Driver	5	x	<1;22), <22;30), <30;40), <40;52), <52;98>
Age_of_Driver_mod		Age_of_Driver	98		1, 2, 3, 4, 5, 6, 7, 8, 9, 10...
Driver_Home_Area_Type	+	Driver_Home_Area_	3	x	Urban area, Small town, Rural
Driver_IMD_Decile	+	Driver_IMD_Decile	10	x	Most deprived 10%, More deprived 10-20%, More deprived 20-30%, More depriv
Journey_Purpose_of_Driver	+	Journey_Purpose_o	5	x	Journey as part of work, Commuting to/from work, Taking pupil to/from school, P
Sex_of_Driver	+	Sex_of_Driver	2	x	Male, Female

Groups of attributes tree

- Root group of attributes
  - Accidents
    - Place
    - Specifics
    - Time
  - Vehicles
    - Driver
    - Vehicle

Attribute	Used	DBColumn	Categories	XCal	Sample categories
Age_of_Vehicle	+	Age_of_Vehicle	82	x	1, 2, 3, 4, 5, 6, 7, 8, 9, 10...
Age_of_Vehicle_ed5	+	Age_of_Vehicle	21	x	<0;4>, <5;9>, <10;14>, <15;19>, <20;24>, <25;29>, <30;34>, <35;39>, <40;44>,
Age_of_Vehicle_ef	+	Age_of_Vehicle	9	x	<2;3), <3;5), <5;6), <6;7), <7;8), <8;10), <10;11), <11;13), <13;105>
Age_of_Vehicle_ef20	+	Age_of_Vehicle	20		ef_1, ef_2, ef_3, ef_4, ef_5, ef_6, ef_7, ef_8, ef_9, ef_10...
Age_of_Vehicle_mod	+	Age_of_Vehicle	82		1, 2, 3, 4, 5, 6, 7, 8, 9, 10...
Engine_Capacity__CC_	+	Engine_Capacity__(	5	x	<4;1199), <1199;1398), <1398;1794), <1794;1998), <1998;99999>
Propulsion_Code	+	Propulsion_Code	11	x	Petrol, Heavy oil, Electric, Steam, Gas, Petrol/Gas (LPG), Gas/Bi-fuel, Hybrid ele
Propulsion_Code 5000+	+	Propulsion_Code	2	x	Petrol, Heavy oil
Vehicle_Type	+	Vehicle_Type	20	x	Pedal cycle, Motorcycle 50cc and under, Motorcycle 125cc and under, Motorcyc
Vehicle_Type 5000 +	+	Vehicle_Type	10	x	Pedal cycle, Motorcycle 50cc and under, Motorcycle 125cc and under, Motorcyc
Vehicle_Type_Groups	+	Vehicle_Type	8	x	Pedal cycle, Motorcycle, Car, Bus, Other, Agricultural vehicle, Tram, Van
Vehicle_Type_Groups 5 000 +	+	Vehicle_Type	5	x	Pedal cycle, Motorcycle, Car, Bus, Van
Was_Vehicle_Left_Hand_Drive	+	Was_Vehicle_Left_I	2	x	No, Yes