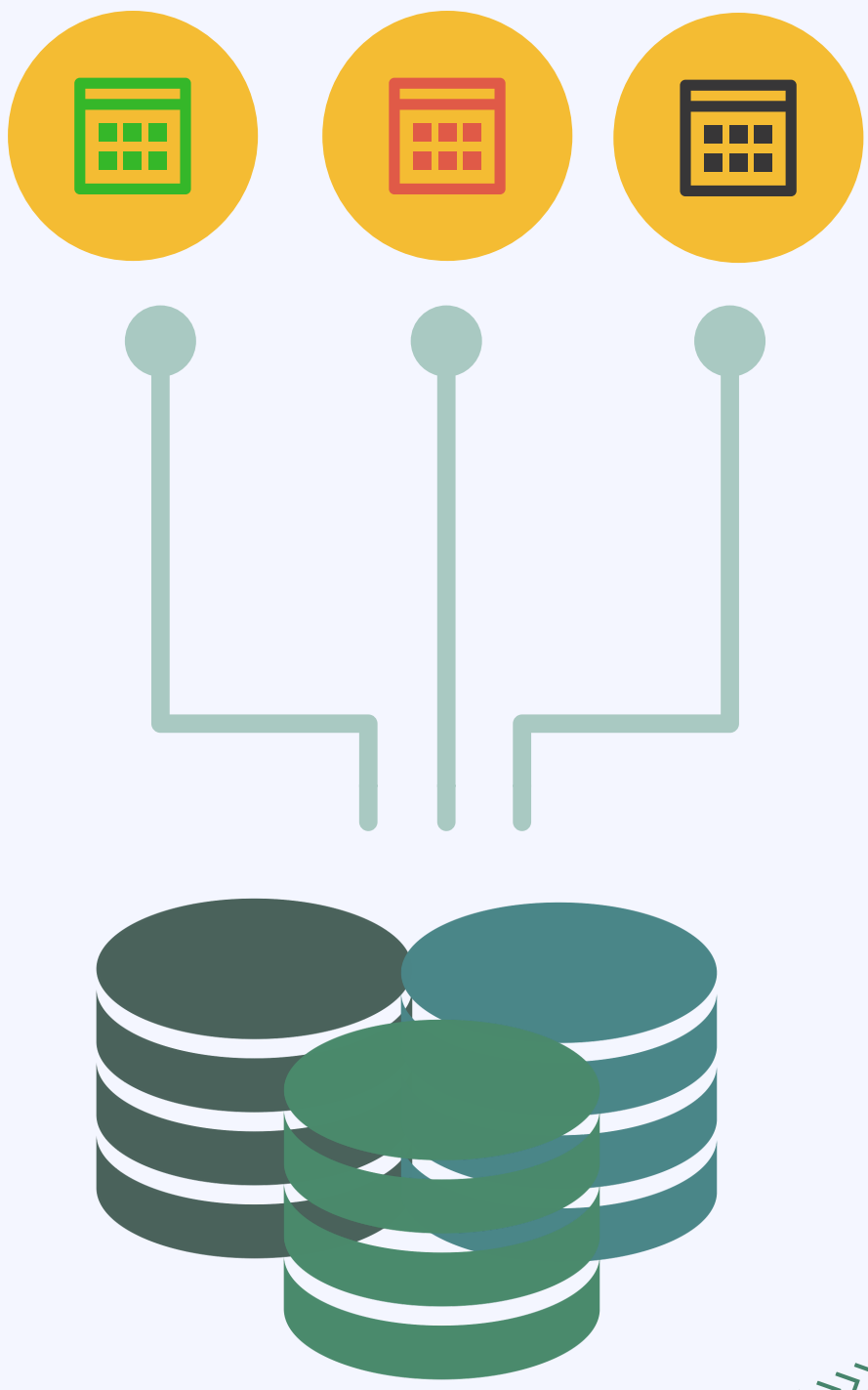


BAYESIAN NETWORKS

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING PROJECT

The Problem

How to automate loading a new data source into the database schema?



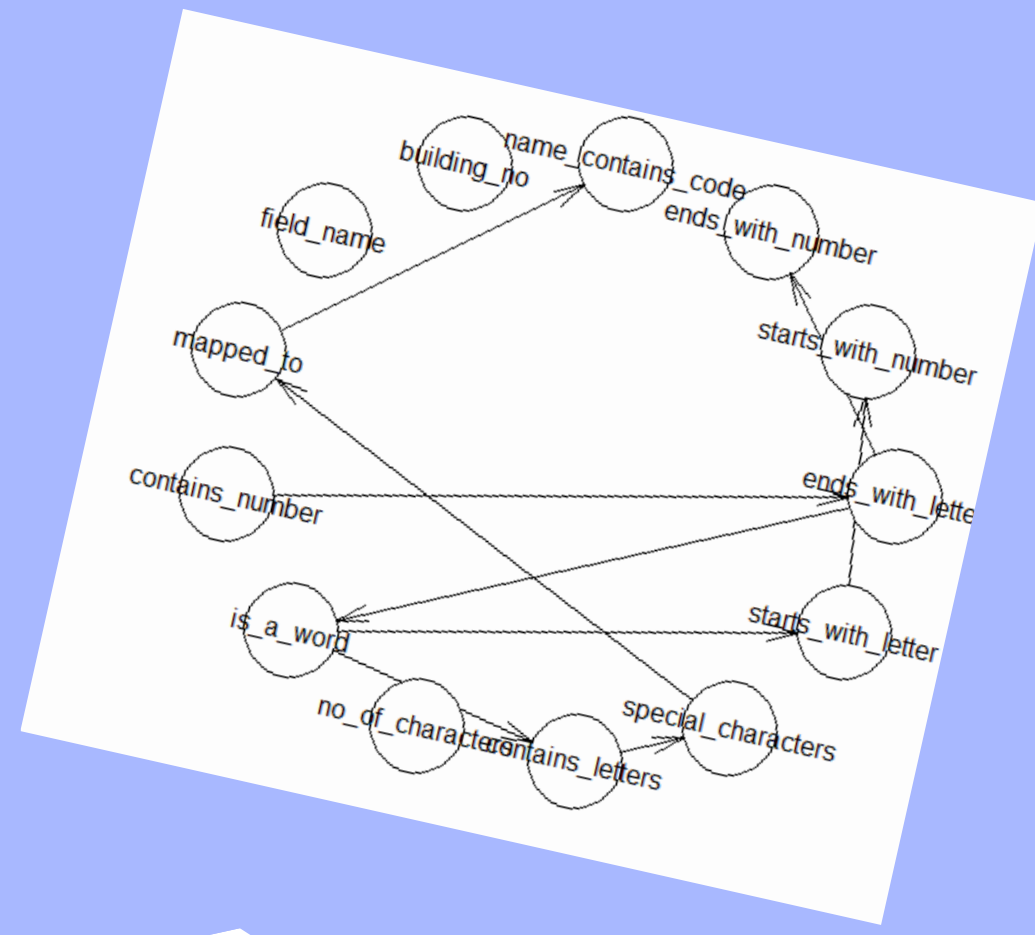
The Model

training dataset

field_name	building_no	mapped_to	contains_number	is_a_word	no_of_characters	contains_letters	special_characters	starts_with_letter	ends_with_letter	starts_with_n
reading_date	B001	date	yes	no	10	no	yes	no	no	yes
reading_time	B001	time	yes	no	5	no	yes	no	no	yes
device_no	B001	device_id	yes	yes	6	yes	no	yes	no	no
status_code	B001	alert_code	no	yes	03-May	yes	no	yes	yes	no
date	B002	date	yes	no	10	no	yes	no	no	yes
time	B002	time	yes	no	5	no	yes	no	no	yes
dev_id	B002	device_id	no	no	5	no	no	yes	yes	no
message_code	B002	alert_code	no	yes	1	yes	no	yes	yes	no
date_of_reading	B003	date	yes	yes	Nov-17	yes	no	no	no	yes
time_of_reading	B003	time	yes	no	5	no	yes	no	no	yes
code_alert	B003	device_id	yes	no	6	yes	no	no	no	yes
code_alert	B003	alert_code	no	yes	3	yes	no	yes	yes	no
readDate	B004	date	yes	no	06-Jul	no	yes	no	no	yes
readTime	B004	time	yes	no	5	no	yes	no	no	yes
machineID	B004	device_id	yes	no	5	yes	no	no	no	yes
code	B004	alert_code	yes	no	1	no	no	no	no	yes

+ bnlearn package

bayesian net



Tools and software

bnlearn package
• hill climbing score based learning algorithm

data preparation



learning and visualiation of bayesian networks

Bayesian networks
A little bit of theory

- based on the Bayes' theorem
- can be used for statistical classification
- is defined by conditional probability tables for the nodes DAG (directed acyclic graph)

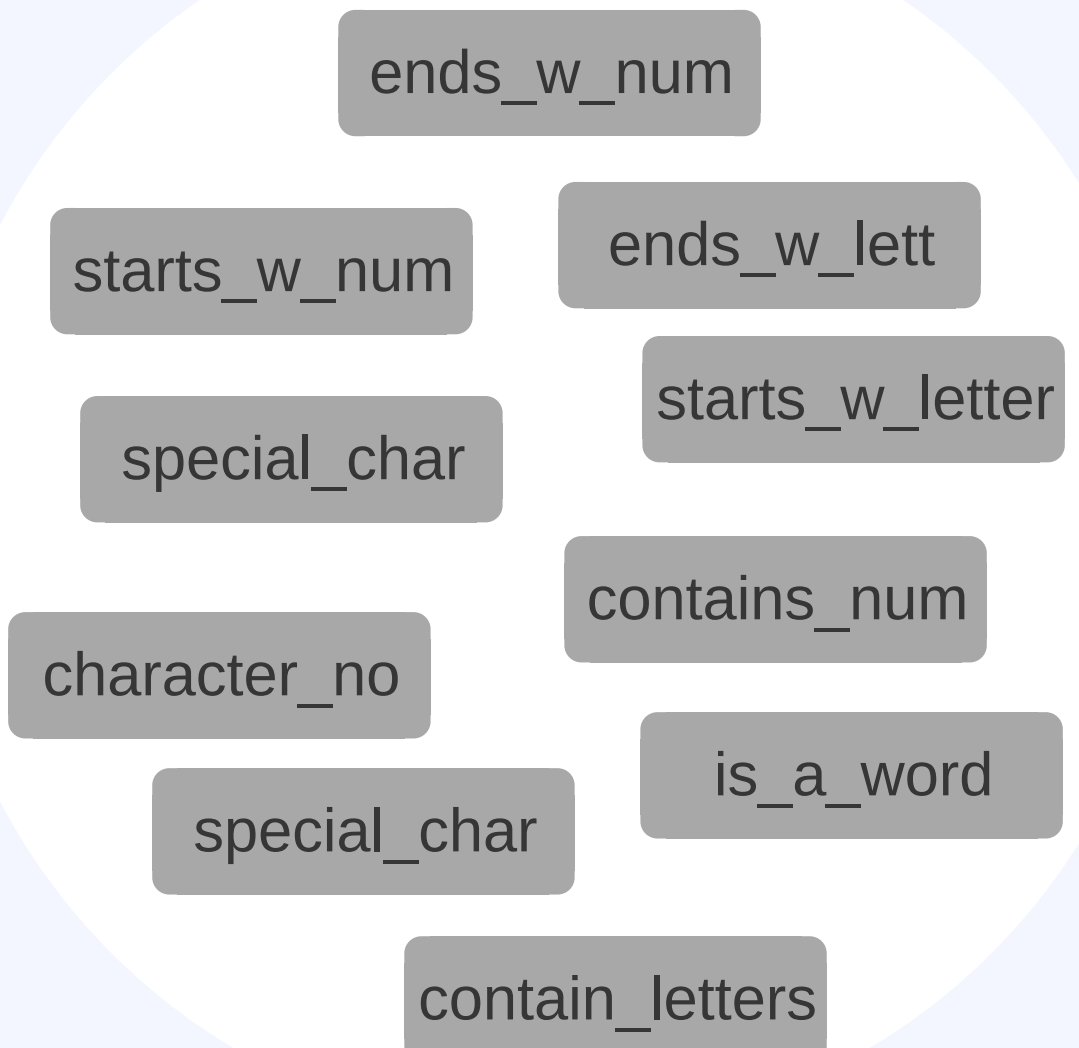
Example:

RAIN	SPRINKLER		GRASS WET
	T	F	
F	0.4	0.6	
T	0.01	0.99	

SPRINKLER	RAIN	GRASS WET	
		T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Result

New dataset



The parameters are put into our model

```
> all_results
      probability
alert_code      1
date            0
time           0
device id       0
```

the model assigns the probability for each schema variable