

# *Using machine learning to classify theft offences to International Classification of Crimes*

Kimmo Haapakangas, Statistics Finland

# *Contents*

---

- Classifications
  - ICCS
  - National classification
- Prosecutors text data and preprocessing
- Process
- Results
- Discussion

# *Introduction*

---

- Main problem is how to map to offences from one classification to another
- Also to produce more detailed statistics for national purposes
  - → use of text data to cover some gaps in national classification

# ICCS

---

- International Classification of Crime for Statistical Purposes
- Developed by UNODC in 2015
- Based on description of criminal acts
- 11 categories and 4 levels
- Enhances the consistency and international comparability of crime statistics
- Recommendations for counting rules etc.

# ICCS level 1

LEVEL 1 CATEGORIES	
1	Acts leading to death or intending to cause death
2	Acts leading to harm or intending to cause harm to the person
3	Injurious acts of a sexual nature
4	Acts against property involving violence or threat against a person
5	Acts against property only
6	Acts involving controlled psychoactive substances or other drugs
7	Acts involving fraud, deception or corruption
8	Acts against public order, authority and provisions of the State
9	Acts against public safety and state security
10	Acts against the natural environment
11	Other criminal acts not elsewhere classified



0501	Burglary
05011	Burglary of business premises
05012	Burglary of private residential premises
050121	Burglary of permanent private residences
050122	Burglary of non-permanent private residences
05013	Burglary of public premises
05019	Other acts of burglary
0502	Theft
05021	Theft of a motorized vehicle or parts thereof
050211	Theft of a motorized land vehicle
050212	Illegal use of a motorized land vehicle
050213	Theft of parts of a motorized land vehicle
050219	Other theft of a motorized vehicle or parts thereof
05022	Theft of personal property
050221	Theft of personal property from a person
050222	Theft of personal property from a vehicle
050229	Other theft of personal property
05023	Theft of business property
050231	Theft from a shop
050239	Other theft of business property
05024	Theft of public property
05026	Theft of services
05029	Other acts of theft

From UNODC ICCS manual

# *Crime classification in Finland*

---

- Based on sections and paragraphs of criminal code
- Doesn't cover subsections → burglary mentioned in subsection of aggravated theft → no statistical code for burglary
- Police data has some other variables to classify burglary, but we don't know how many are sentenced for burglary
- Changes when legislation changes
  - New law → new code

# Mapping theft and burglary to ICCS

ICCS offence	Number of persons sentenced by principal offence rule in 2017
0501 Burglary	-
05011 Burglary of business premises	-
05012 Burglary of private residential premises	-
050121 Burglary of permanent private residences	-
050122 Burglary of non-permanent private residences	-
05013 Burglary of public premises	-
05019 Other acts of burglary	-
0502 Theft	7574
05021 Theft of a motorized vehicle or parts thereof	343
050211 Theft of a motorized land vehicle	-
050212 Illegal use of a motorized land vehicle	87
050213 Theft of parts of a motorized land vehicle	-
050219 Other theft of a motorized vehicle or parts thereof	-
05022 Theft of personal property	-
050221 Theft of personal property from a person	-
050222 Theft of personal property from a vehicle	-
050229 Other theft of personal property	-
05023 Theft of business property	-
050231 Theft from a shop	-
050239 Other theft of business property	-
05024 Theft of public property	-
05025 Theft of livestock	-
05026 Theft of services	-
05029 Other acts of theft	17

- Situation without machine learning or manually reading text
- No burglaries when using national classification

# *Prosecutors text data and preprocessing 1/2*

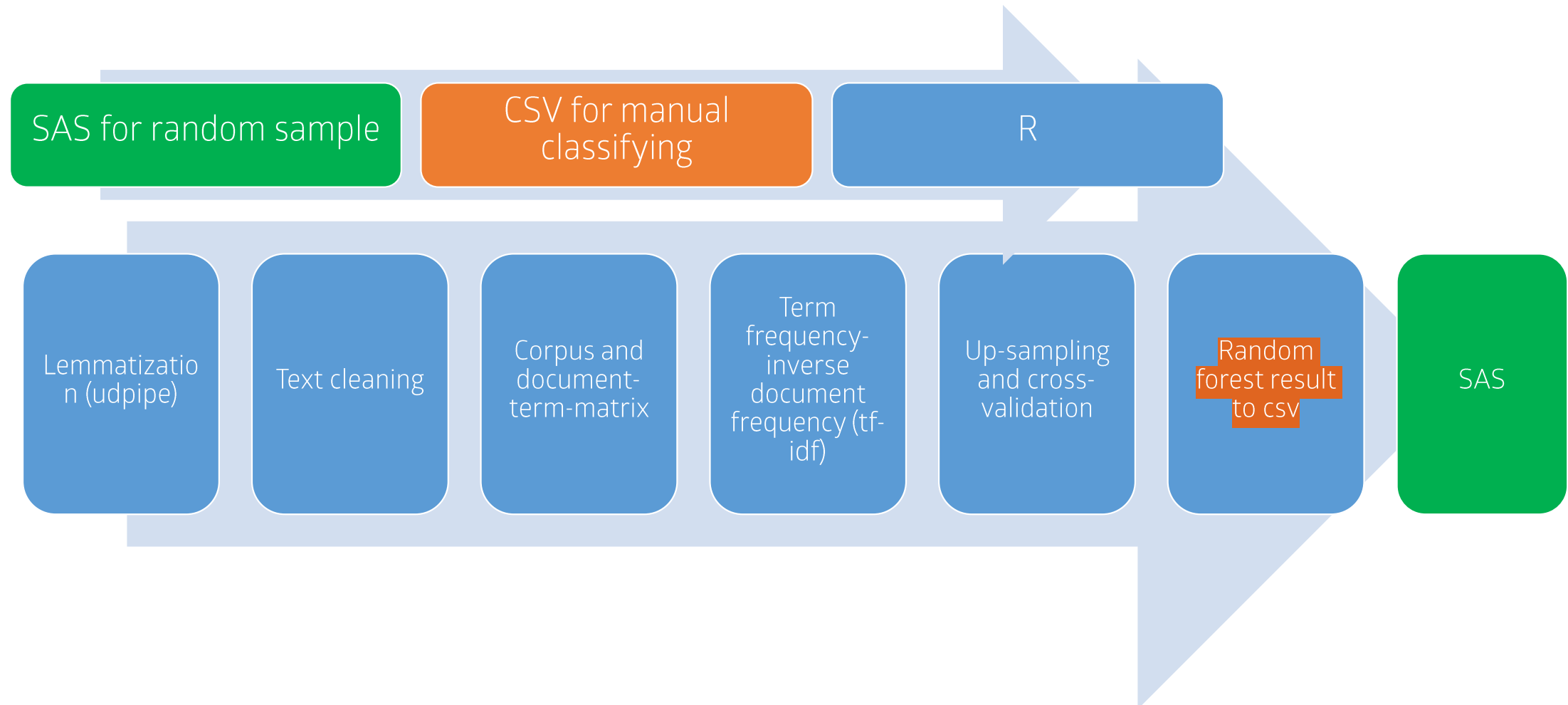
- Short description of what happened by the prosecutors point of view
- Usually less than 10 sentences, but sometimes longer and can include more than one offence
- In Finnish or Swedish
- In 2019, total 23.600 text for theft, petty theft, aggravated theft and their attempts
- Weighted random sample of 2.000 text was manually classified to ICCS 0501 Burglary and 0502 Theft and subcategories



# Prosecutors text data and preprocessing 1/2

- “Person A has stolen property of plaintiff’s B and C by force entry to their dwelling from open window located in address X and taking an iPad tablet from there. The stolen property has been returned to the owners”.
- → ICCS 050121 burglary of private residential premises
- Only few observations in some categories like 050130 burglary of public premises or 050260 theft of services
  - keyword search to find some more observations
- Final data of 2.100 observations, only Finnish text
  - Split for 75/25 train/test data

# Process



## Results

- Small n in some classes in train/test data
- Overall accuracy 78.8 %
- Some classes have bad sensitivity or precision values
- Balanced accuracy still good
- Larger n in train data → better results

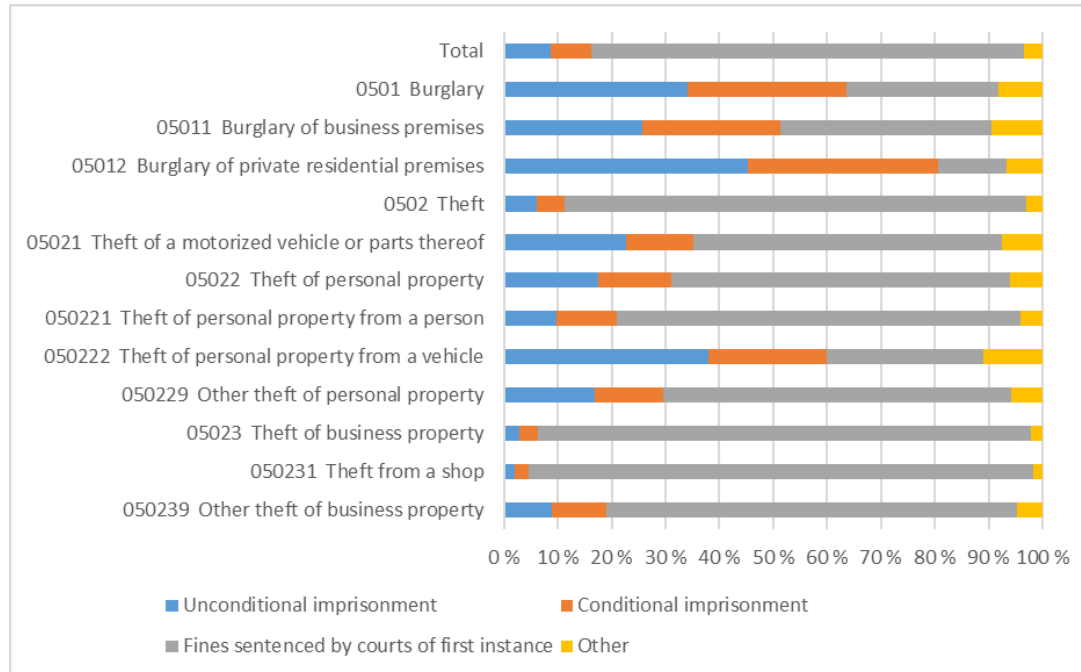
cat	Train (n)	Test (n)	Sensitivity/ Recall	Specificity	Precision	Balanced Accuracy
iccs_5011	148	49	0,897	0,963	0,721	0,93
iccs_50121	123	40	0,85	0,991	0,894	0,92
iccs_50122	56	18	0,666	1	1	0,833
iccs_5013	13	4	0,25	0,992	0,2	0,621
iccs_5019	22	7	0,142	1	1	0,571
iccs_50211	22	7	0,857	0,998	0,857	0,927
iccs_50213	22	7	0,428	1	1	0,714
iccs_50219	20	6	0,666	0,996	0,666	0,831
iccs_50221	90	29	0,931	0,965	0,613	0,948
iccs_50222	72	24	0,708	0,971	0,548	0,84
iccs_50229	235	78	0,743	0,923	0,63	0,833
iccs_50231	513	170	0,976	0,982	0,965	0,979
iccs_50239	200	66	0,484	0,982	0,8	0,733
iccs_5024	22	7	0,142	0,996	0,333	0,569
iccs_5026	10	3	0,333	0,998	0,5	0,665
iccs_5029	15	5	0,6	1	1	0,8
Total	1583	520				0,79

# Results (2/3)

- In SAS unauthorized use (28:7-9) and stealing of a motor vehicle for temporary use (28:9a-9c) classified with keyword search

	Offences total	Principal offence
Total	23380	8790
0501 Burglary	2170	810
05011 Burglary of business premises	1150	430
050121 Burglary of permanent private residences	750	310
05013 Burglary of public premises	70	20
0502 Theft	21220	7980
050211 Theft of a motorized land vehicle	980	270
050212 Illegal use of a motorized land vehicle	180	70
050213 Theft of parts of a motorized land vehicle	60	20
050221 Theft of personal property from a person	1080	290
050222 Theft of personal property from a vehicle	690	140
05023 Theft of business property	16110	6390
050231 Theft from a shop	13810	5630
05024 Theft of public property	60	30
05026 Theft of services	20	20

# Results (3/3)



- Fines given from shoplifting, imprisonment from theft of personal property
- Previous crime history also effects overall sentence

# *Discussion*

- Model not yet in production and not fully automated
- Some more train data could be usefull
- Could benefit from more advanced text processing (Google Bert/FinBert)
- Heavy for personal laptop → need for server environment
- Still very good results and has lots of potential

*Thank you*

kimmo.haapakangas@stat.fi