

Usage of text mining in auditing process

Ms Meeli Saksing, Auditor
Ms Kaia Philips, Senior adviser
NAO of Estonia

Audit “Publicity efforts of local authorities in municipal newspapers” (2018)





What and Why?

NAO analysed the media coverage of candidates participating in the 2017 local elections across municipal newspapers published from January to October 2017.

The goal was to make sure that newspapers financed by local governments have not been used towards

- personal political interests;
- covert election campaigns;
- other party-political image campaigns.



How?

Outsourced to a Balti Meediamonitooringu Grupp OÜ

- media monitoring and analysis tool called Station
 - created a database of electronically available municipally funded newspapers
 - total of 191 electronically available municipally funded newspapers.
- Keywords: „candidate name“, „political party name“, „elections“, „opening“, „investing“
 - Analysis using the keywords and combinations of keywords incl: sequence, placement, size, visualizations used, overall „emotion“ of the article, connection with elections, author of the article.

How?

- MS Excel spreadsheet:

Stationi link	Date	Name of the newspaper	Emotion of the article	Page nr	Author	Connected with elections	Picture	Size of the picture	Candidate mentioned in the article
http://station.ee/#16999383	2017-01-01 00:00:00	Sindi Sõnumid	Neutral	3	U.S	No	Yes	1/8	M.H
http://station.ee/#15970845	2017-01-01 00:00:00	Haljala Valla Sõnumid	Positive	1	M.T	No	No	1/8	M.T
http://station.ee/#16358417	2017-01-01 00:00:00	Tamsalu Ajaleht	Positive	7	T	No	Yes	-	T.V
http://station.ee/#16045313	2017-01-01 00:00:00	Koduvald (Häädemeeste valla leht)	Neutral	10	K.V	No	No	1/16	K.S

- The search returned 24 500 results/mentions of keywords (in total 6000 articles).

Station?

STATION



CATEGORIES

- Baltic Maritime Association
- Competitors
- Maritime news
- Legislation
- Logistics sector
- Spokespersons

FILTERS

- Real estate
- BNS biz news
- Campaign
- Maritime news
- Spokesperson: Bo Smith
- Spokesperson: Brent Lloyd
- Trade unions
- Port of Klaipeda
- Official announcements

AK



Open Search

Search results (4841)

Vikerraadio 02.01.2017 07:35

Vikerraadio 02.01.2017 07:35

Aktuaalne kaamera (Tags:Lennujaam)

ETV 01.01.2017 21:00

Olukorrast riigis

Raadio2 01.01.2017 12:05

Tagasivaade 2016. aastale vaimuliku pilgu läbi

postimees.ee 31.12.2016 18:16

Keskpäevatund

Raadio Kuku 31.12.2016 12:05

Rahva teenrid

Vikerraadio 31.12.2016 11:05

REPLIIK: Päike ja Saaremaa

saartehaal.ee 31.12.2016 00:03

REPLIIK: Päike ja Saaremaa

Saarte Hääl 31.12.2016

In memoriam 2016

Postimees Arter 31.12.2016

Totaalne ärkamine 8:30

Raadio MyHits 30.12.2016 08:30

Ants Mailend: ühtsust takistab visiooni puudumine

Postimees 30.12.2016

Aktuaalne kaamera



Play all Stop



Lennujaam



AK-le

kilomeetrit rannajoont on pidevas muutumises on kohti kus meri muutub maaks ja teises kohas jälle vastupidi aga kuna maa kerkib tsipa rohkem

ETV

01.01.2017 21:00

126500

Rannajoon

Add new tag

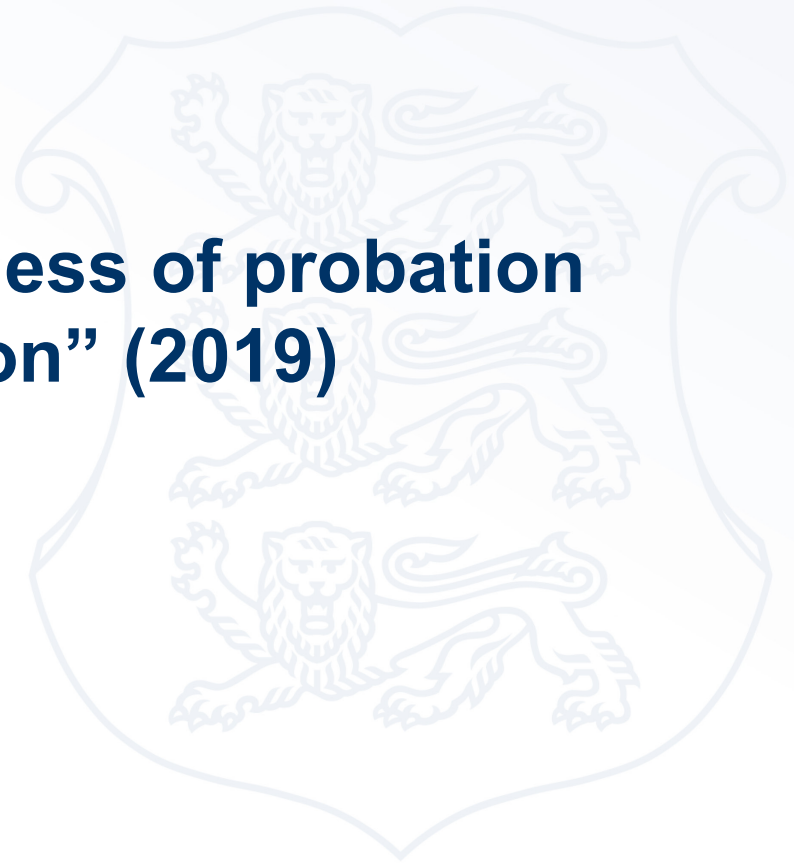
Edit comment

Star

Send



Audit “Effectiveness of probation supervision” (2019)





What and Why?

- Community sanctions such as probation is considered to be more efficient than imprisonment.
- Efficiency of probation system is measured by the rate of recidivism.

Audit question was focussing on possible ways to improve the effect of probation or lower the recidivism.

Offenders on probation are affected by:

- the service offered by probation system: e.g. professionalism of the probation officers, readiness of the care system (material and skill-related);
- many other factors - such as addictions, employment, residence characteristics, other kind of services, debts etc.





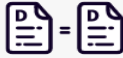




How? – Data from 8 public sector databases

1. Prisoners and probationers register -> probationers' personal data
2. Criminal Records Database -> criminal record 2 years after probation
3. E-dossier system -> criminal record 2 years after probation
4. Enforcement procedure register -> debts, claims by bailiffs
5. Social services and benefits register -> social services and benefits
6. Income tax register -> income and the regularity of income
7. Unemployment and labour market services register -> unemployment status and services provided by Unemployment Fund
8. Population register -> deaths and moving abroad

How? - Analysis of textual data with help of TEXTA (texta.ee)

TEXTATEXTA ToolkitDemoGrantsContacts

With the help of TEXTA Toolkit and support from our analysts, our partners can solve most relevant tasks related to text analytics.



CUSTOMER SUPPORT AUTOMATION

INFORMATION EXTRACTION

DMS AUDIT

DOCUMENT RECOMMENDATION

DETECT TEXT FROM IMAGES

DOCUMENT CLASSIFICATION

SENTIMENT ANALYSIS

DATA VISUALIZATION

WITH TEXTA TOOLKIT YOU:

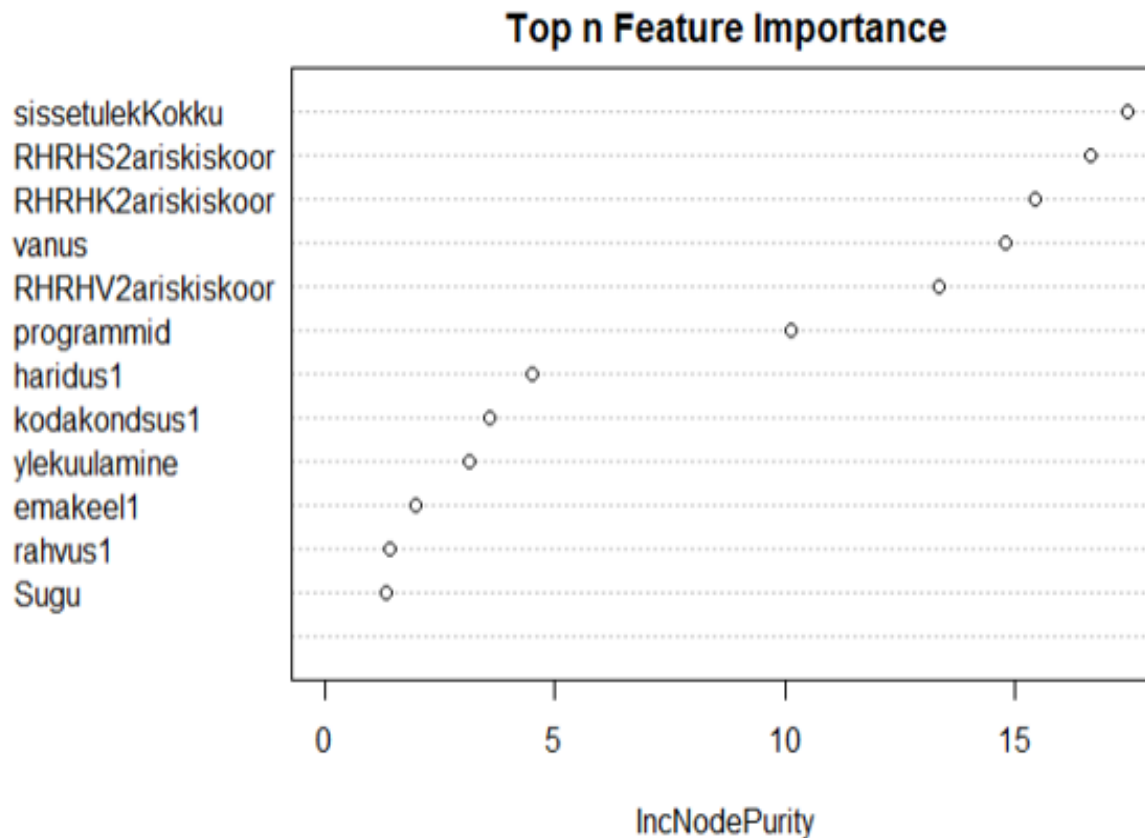
- Save in labor cost
- Understand better your products and performance of different units
- React quicker to business events
- Gain control over your DMS
- Less routine tasks for your workforce

PRICE:

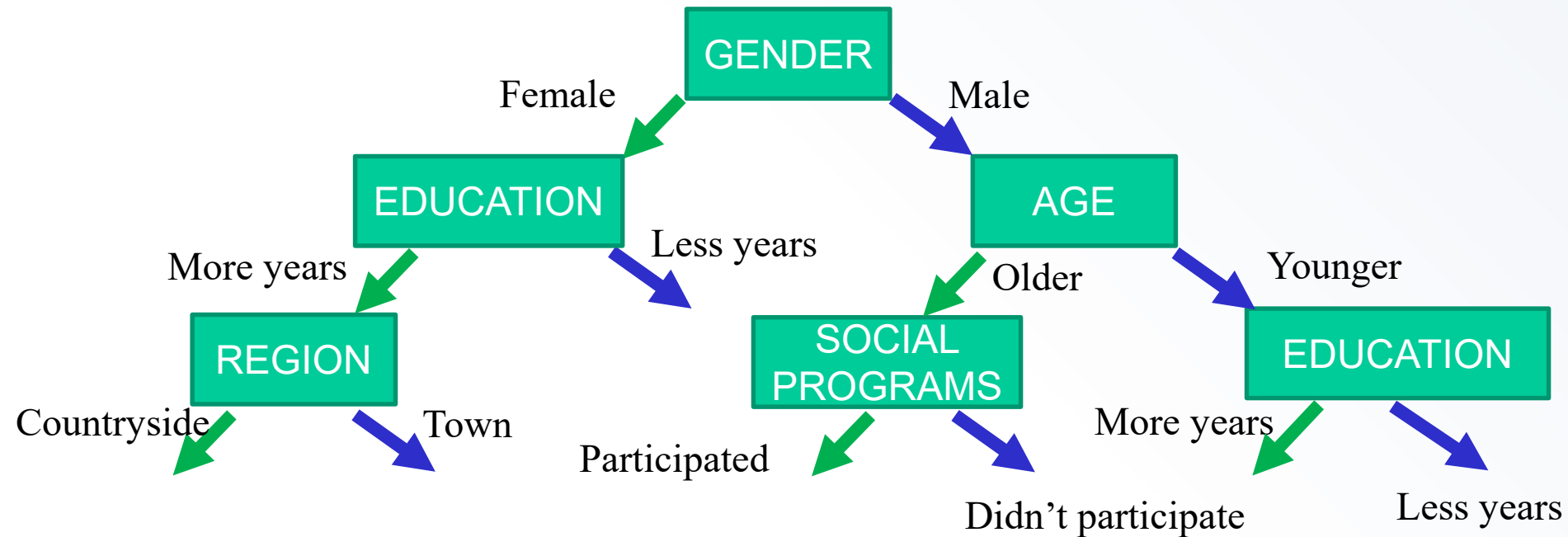
TEXTA Toolkit is open source and freely available.

[Download from Github](#)

How? – Regression analysis and experiments of using random forest and decision tree



Decision tree – who is more likely to reoffend?



Legend:



More likely to
reoffend



Less likely to
reoffend



Some findings - which factors lead to success?

Probation related

- The length of the probation officer's responsibility period (+)
- Change of probation officers (-)
- Social and rehabilitation programs (+)

Person related

- Addiction problems (drugs and alcohol) (-)
- Unemployment (-)
- Living conditions and cohabitation (+ / -)



Thank you

Questions?