



Aplokšņu algu maksātāju atklāšana, izmantojot mašīnmācīšanās algoritmus

Nicolas Gavoille (SSE Riga, BICEPS), Anna Zasova (BICEPS)

12. ikgadējā Ēnu ekonomikas konference

2022. gada 16. maijā

Kāpēc tapa šis pētījums?

- «Aplokšņu» algas – viens no izplatītākajiem nodokļu nemaksāšanas veidiem Latvijā;
- Jautājums: kā identificēt uzņēmumus, kuri maksā algas «aploksnē»?
- VID pielieto dažādas metodes, lai identificētu šādus uzņēmumus;
- Šajā pētījumā: izmantojam mašīnmācīšanās metodes un VID administratīvos datus;
- Piedāvātās metodes ļauj identificēt uzņēmumus, kuri ar lielu varbūtību izvairās no darbaspēka nodokļiem;
- Priekšrocība: šo metožu izmantošanai nepieciešamie dati jau ir VID rīcībā.

Ko mēs darām?

- Izmantojam VID datus par nodokļu auditiem un uzņēmumu finanšu rādītājiem;
- Pielietojam 4 dažādus mašīnmācīšanās algoritmus;
- Identificējam uzņēmumus, kuri ar lielu varbūtību maksā aplokšņu algas.

Dati

No VID tika saņemtas divas (**anonimizētas**) datubāzes:

Datubāze par **auditētajiem** uzņēmumiem:

- Informācija par visiem 2013. – 2020. gadā veiktajiem IIN un VSAOI auditiem;
- Audita rezultāts (aprēķinātie papildu nodokļi un uzliktie sodi);
- Uzņēmuma līmeņa dati: NACE, nodarbināto skaits, vidējā alga uzņēmumā, apgrozījums, aktīvi, peļņa, u.c.

Datubāze par **neauditētajiem** uzņēmumiem tajās pašās NACE nozarēs:

- Uzņēmuma līmeņa dati: NACE, nodarbināto skaits, vidējā alga uzņēmumā, apgrozījums, aktīvi, peļņa, u.c.

Mūsu pieeja

Pielietojam četrus mašīnmācīšanās algoritmus:
Random Forest, Gradient Boosting, Neural Networks un Logit



Trīs galvenie soļi:

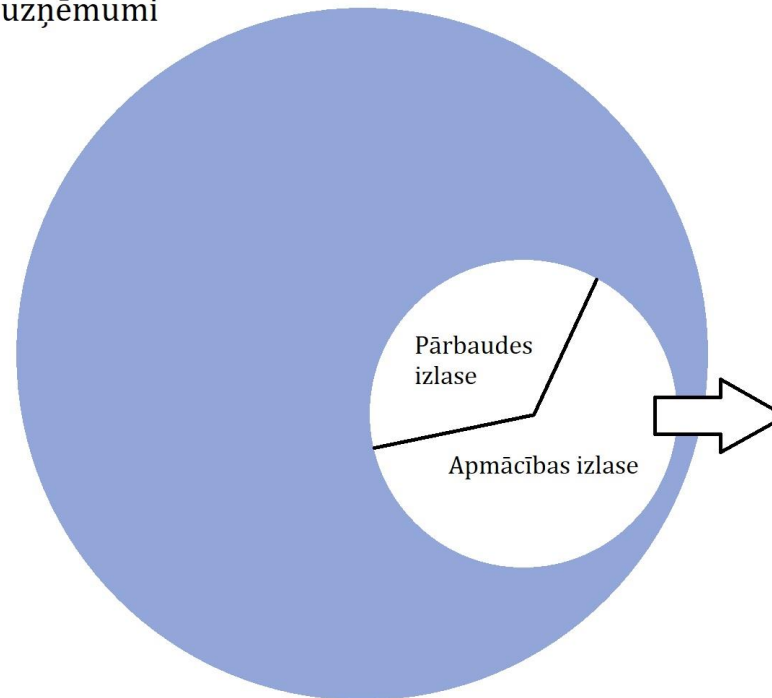
1. solis: izveidojam uzņēmumu izlasi, par kuriem mēs zinām «patiesību» - maksā vai nemaksā aplokšņu algas — auditētie uzņēmumi

2. solis: apmācām algoritmus uz šīs izlases bāzes, pārbaudām algoritmu precizitāti

3. solis: pielietojam algoritmus neauditētajiem uzņēmumiem piecās nozarēs (apstrādes rūpniecība, tirdzniecība, transports, būvniecība, viesnīcas/ēdināšana).

Algoritmu apmācība

Visi uzņēmumi



Uzņēmumi, par kuriem mēs zinām, vai tie izvairās no darbaspēka nodokļu maksāšanas

Modeļu precizitāte

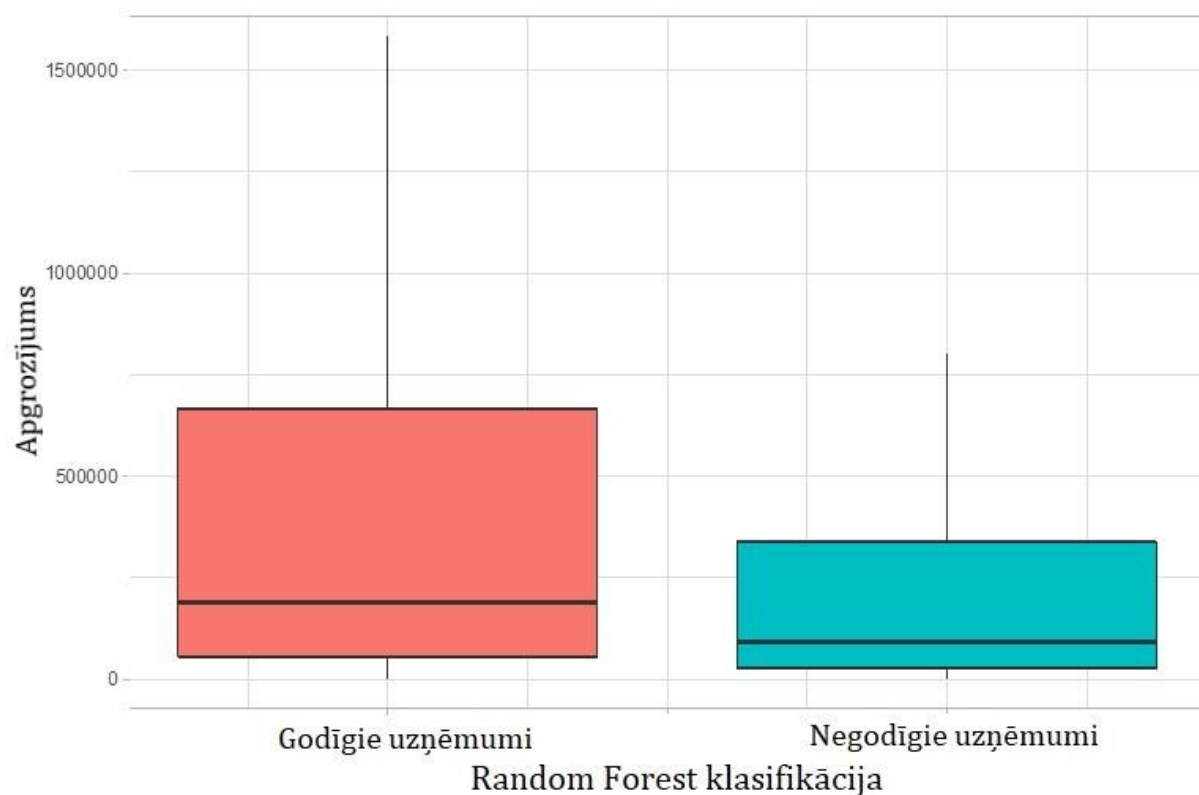
	Random Forest	Gradient Boosting	Neural Network	Logit
Precizitāte*	75,3%	72,0%	66,4%	63,3%

* Parāda, cik % uzņēmumu tiek pareizi klasificēti kā godīgi vai negodīgi.

Modeļu validācija: godīgo un negodīgo uzņēmumu salīdzinājums pēc apgrozījuma

Apgrozījums

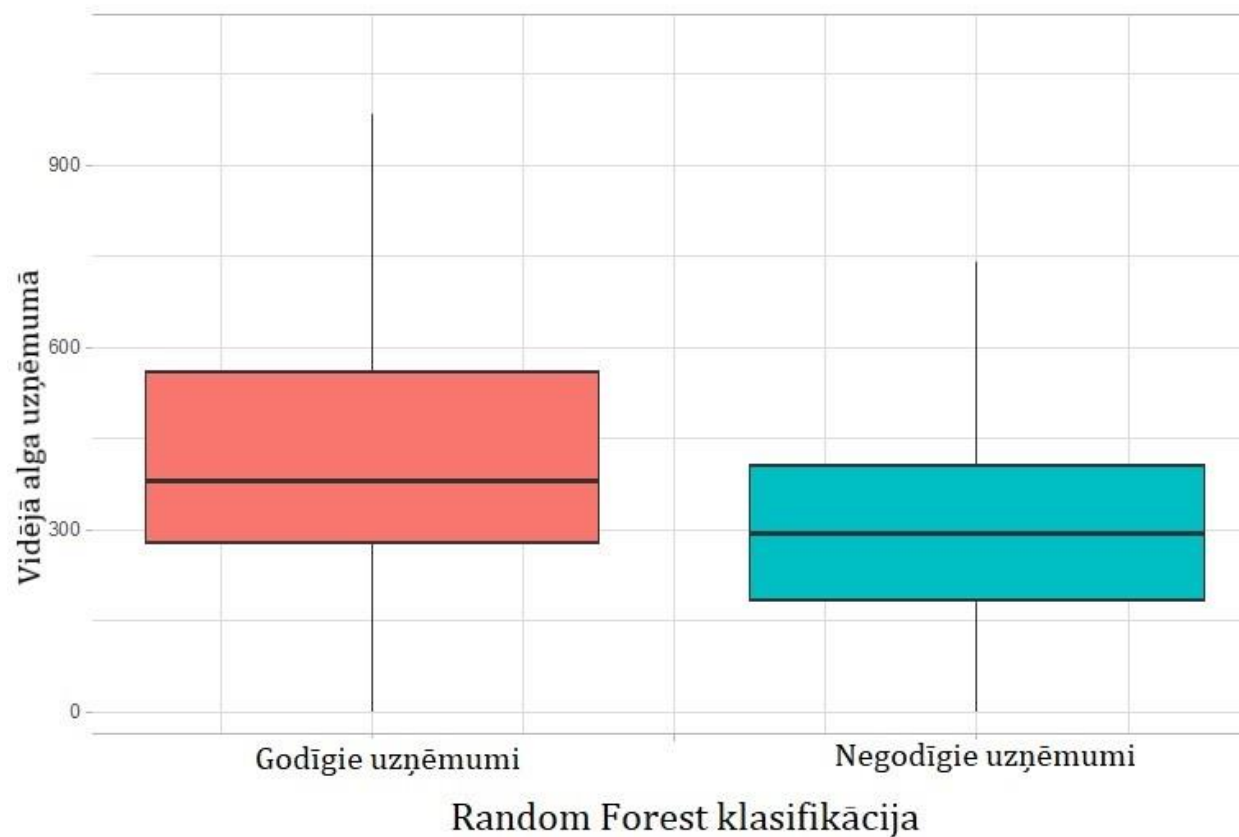
Godīgie uzņēmumi vidēji uzrāda lielāku apgrozījumu



Modeļu validācija: godīgo un negodīgo uzņēmumu salīdzinājums pēc vidējās algas

Vidējā alga uzņēmumā

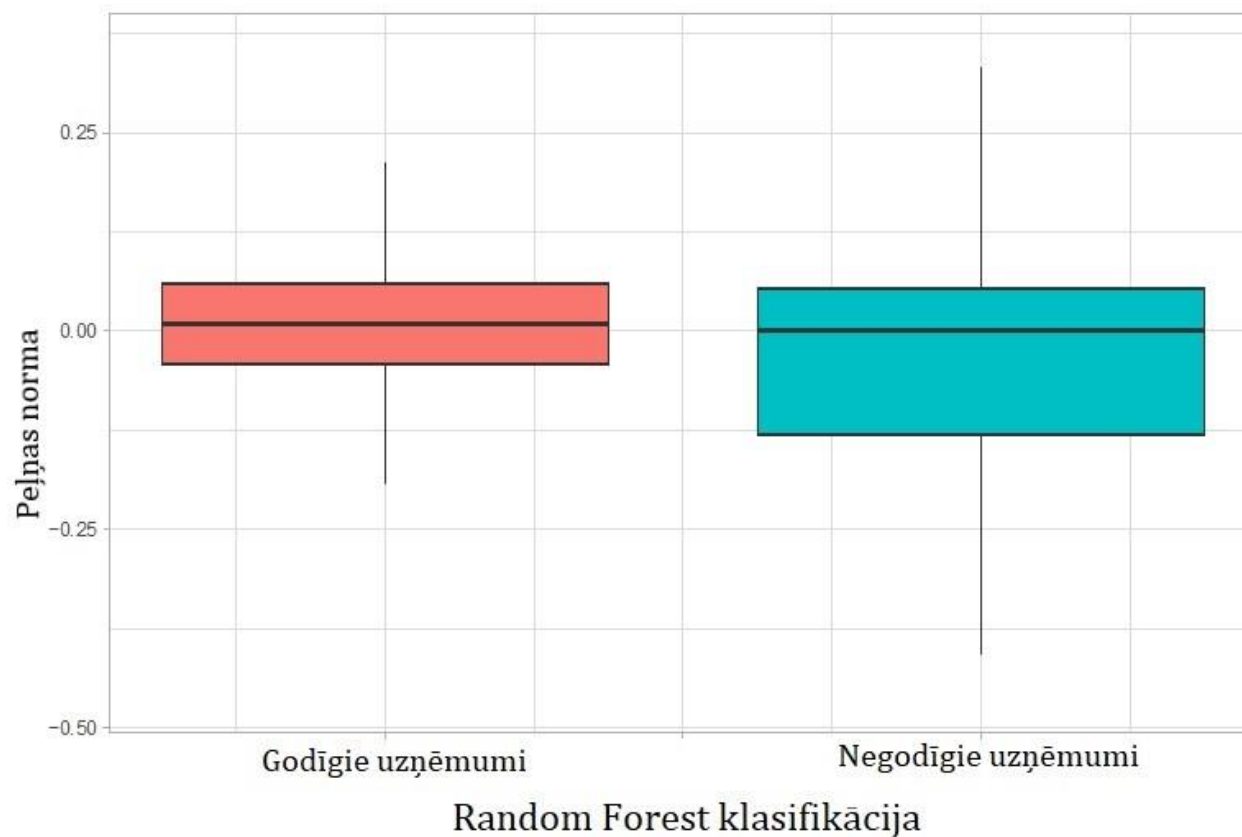
Godīgie uzņēmumi vidēji uzrāda lielākas algas



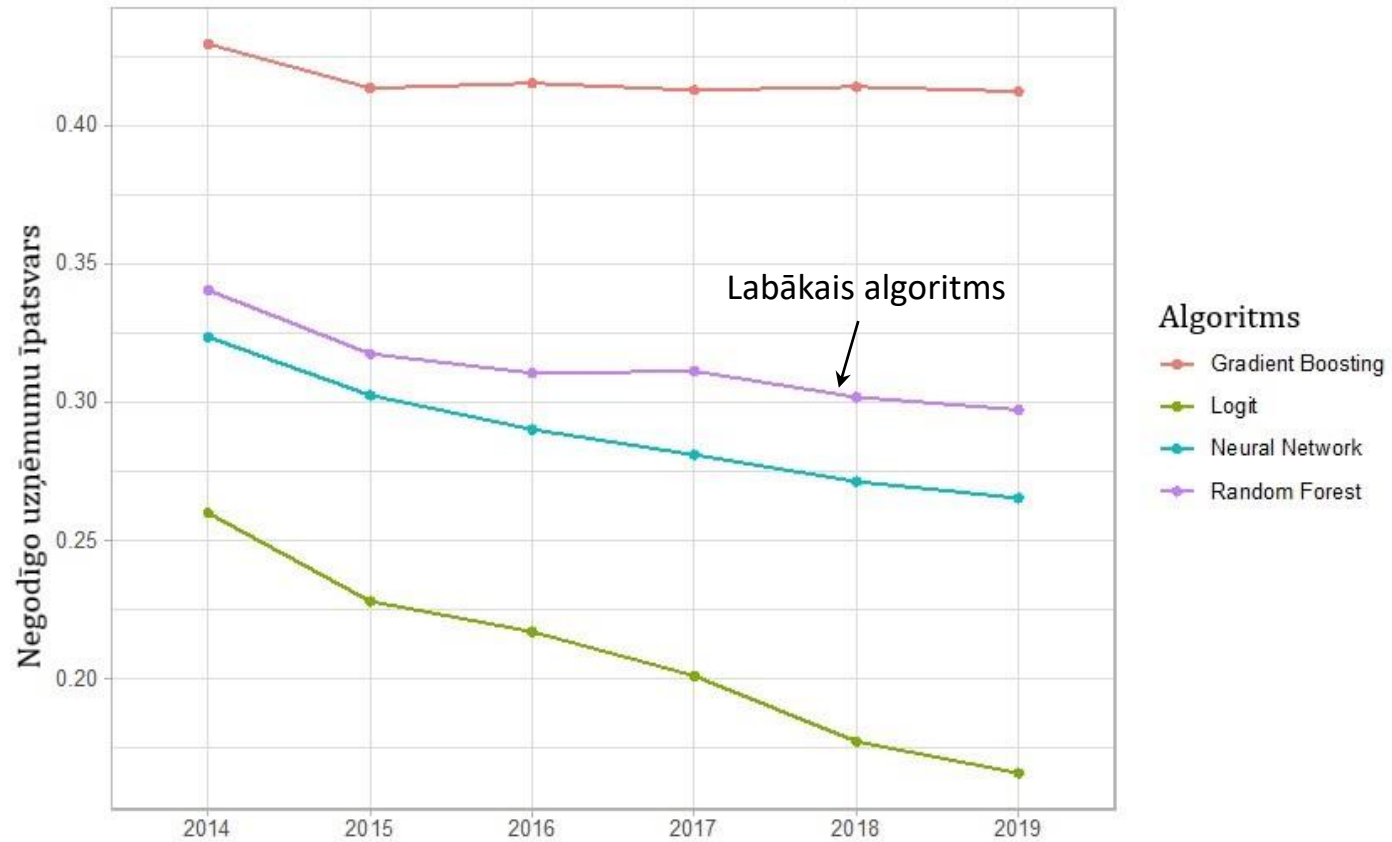
Modeļu validācija: godīgo un negodīgo uzņēmumu salīdzinājums pēc peļņas normas

Peļņas norma (peļņa/apgrozījums)

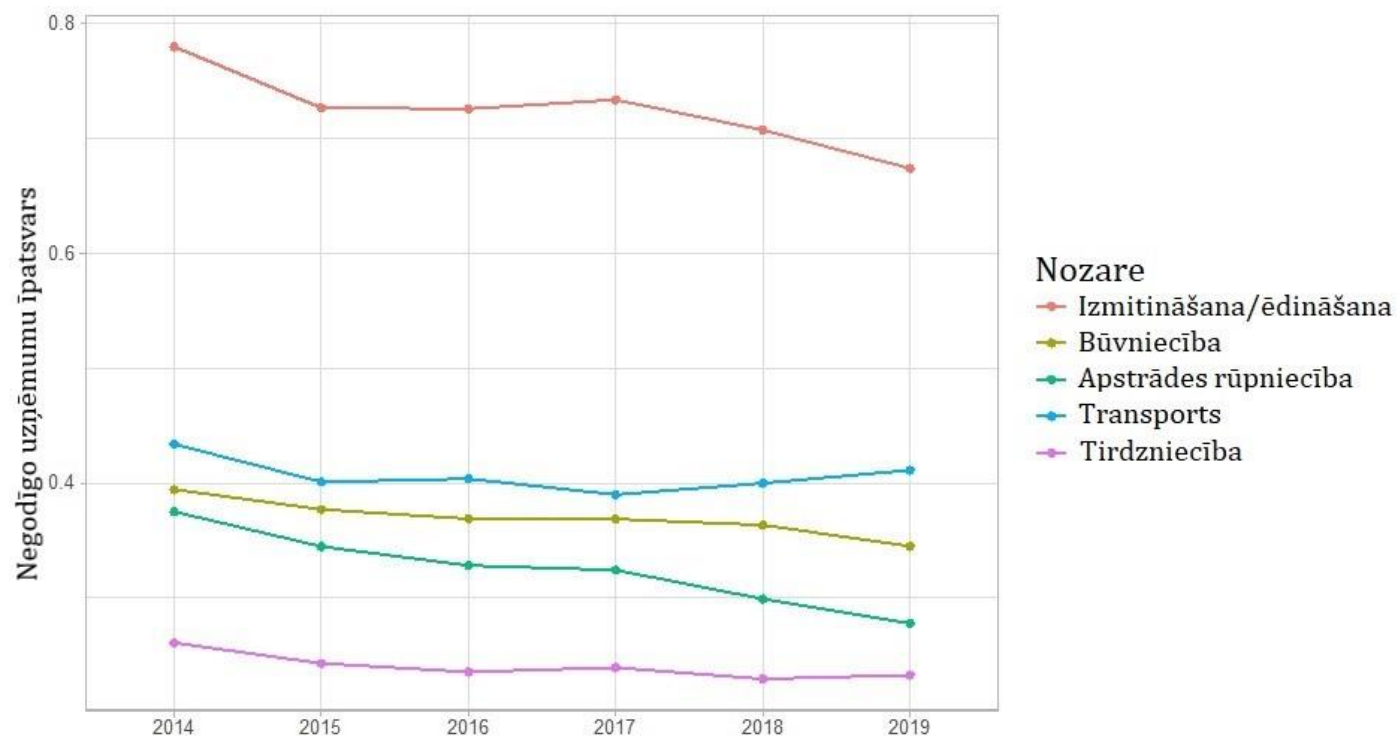
Vidējā peļņas norma godīgo un negodīgo uzņēmumu vidū būtiski neatšķiras, bet negodīgie uzņēmumi biežāk norāda ļoti zemu vai ļoti augstu peļņas normu.



Rezultāti: negodīgo uzņēmumu īpatsvars



Rezultāti: negodīgo uzņēmumu īpatsvars nozaru griezumā (Random Forest)



Secinājumi (1)

- Mašīnmācīšanās metodes var tikt veiksmīgi pielietotas, lai identificētu uzņēmumus, kuri izvairās no (darbaspēka) nodokļu maksāšanas;
- Negodīgo uzņēmumu īpatsvars piecās apskatītajās nozarēs 2014. – 2019. gadā bija 30-40%;
- Šis īpatsvars nedaudz samazinājās kopš 2014. gada;
- NB: tas *nenozīmē*, ka samazinājās «aplokšņu» algu apjoms – iespējams, ka tie uzņēmumi, kuri izvairās no nodokļiem, vidēji izvairās vairāk.

Secinājumi (2)

- Nozare ar lielāko negodīgo uzņēmumu īpatsvaru – izmitināšana un ēdināšana (~ 80%);
- Nozares ar mazāko negodīgo uzņēmumu īpatsvaru – apstrādes rūpniecība un tirdzniecība.

Rīcībpolitikas rekomendācijas

- Mašīnmācīšanās metodes «aplokšņu» algu identificēšanai var tikt pielietotas kā papildu rīks VID modeļu klāstā;
- Izstrādātos modeļus ir iespējams vēl uzlabot, jo VID rīcībā ir detalizētāki dati par uzņēmumiem (tie nevar tikt nodoti trešajām personām konfidencialitātes apsvērumu dēļ).

Paldies par uzmanību!

Nicolas Gavoille (Nicolas.Gavoille@sseriga.edu)

Anna Zasova (Anna@biceps.org)

The background features a dark blue, textured pattern of overlapping lines and shapes, resembling a stylized architectural or technical drawing. A prominent yellow rectangle is centered on the page, containing the main text. A faint, light-colored line chart with square markers is visible behind the text, extending across the width of the yellow area.

12. IKGADĒJĀ ĒNU EKONOMIKAS KONFERENCE