

TARTU ÜLIKOOL
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikutele õppekava

Sven Kuning

**Klientide lahkumise ennustamine
masinõppega SEB pensionifondide näitel**

Magistritöö (15 EAP)

Juhendaja: Anna Leontjeva, PhD

Tartu 2018

Klientide lahkumise ennustamine masinõppega SEB pensionifondide näitel

Lühikokkuvõte:

Masinõppe meetodeid on tihti kasutatud klientide lahkumise ennustamiseks teenindussektoris. Käesolev töö annab ülevaate meetodikatest ja keskendub lähemalt juhumetsade ja närvivõrgu mudelile ja nende rakendamisele reaalsele kliendiandmetele. Lisaks tavapärastele profiiliandmetele püütakse klientidevahelise tehinguinfo põhjal leida, kas selliste andmete lisamine parandab esialgse mudeli karakteristikuid. Parimaks mudeliks osutub profiiliandmetele tuginev juhumetsade mudel saagisega 59%, täpsusega 62% ja esitustäpsusega 4% mida võib pidada ebapiisavaks, et seda reaalses äritegevuses rakendada. Võrgustikuandmed, mille kaasamiseks mudelitesse kasutatakse node2vec algoritmi ja kliendi egograafis aset leidnud varasemaid klientide lahkumisi ei aidanud kaasa esialgse mudeli parandamisele.

Võtmesõnad:

Masinõpe, klientide lahkumise ennustamine, juhumetsad, närvivõrgud

CERCS: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Client churn prediction with machine learning based on SEB pension fund client data

Abstract:

Machine learning methods have often been used to predict client churn in the service sector. This paper gives an overview of some methods used and focuses on the random forest and neural network models and their application on real-life client data. In addition to using the ordinary profile data, it seeks to find out whether the amendment of the data set with inter-client transaction data helps to improve the characteristics of the original model. The best model proposed is one based on random forests with recall of 59%, accuracy of 62% and precision of 4%, which can be regarded as insufficient to deserve a real-life business implementation. Node2vec algorithm and previous churn cases in each client's ego graph were used to process network data to add it to the usual profile data for

model training. Unfortunately, the addition of such data did not help to improve the original model.

Keywords:

Machine learning, client churn prediction, random forests, neural networks

CERCS: P160 Statistics, operation research, programming, actuarial mathematics

Lihtlitsents lõputöö reprodutseerimiseks

Mina, **Sven Kuning**,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
Klientide lahkumise ennustamine masinõppega SEB pensionifondide näitel,

(lõputöö pealkiri)

mille juhendaja on **Anna Leontjeva, PhD,**

(juhendaja nimi)

reprodutseerimiseks ainult säilitamise, sealhulgas digitaalarhiivis DSpace säilitamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni. **Lõputöö avaldamine ei ole lubatud.**

2. olen teadlik, et punktis 1 nimetatud reprodutseerimise õigus jääb alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tallinnas, **09.08.2018**