

UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Laura Ruusmann

Comparison of category-level, item-level and general sales forecasting models

Master's Thesis (30 ECTS)

Supervisor: Marlon Dumas, PhD

Supervisor: Eerik Muuli, MSc

Tartu 2020

Comparison of category-level, item-level and general sales forecasting models

Abstract:

Sales forecasting is the process of estimating future sales. In this thesis, multiple methods are tested out for achieving best forecasting accuracy with lowest computational requirements.

Three families of methods are investigated: a traditional statistical forecasting approach (ARIMA), classical machine learning techniques (specifically ensemble methods) and a third one based on deep learning methods (specifically recurrent neural networks with LSTM architectures).

The study uses real-world sales transaction data from a large retail company in a Baltic country and the aim of this thesis is to improve their current sales forecasting system.

Here we show that improving on their current sales forecasting is possible and additionally analyse the influence of promotional sales to prediction accuracy. The results show that using a combination of multiple item-level decision tree-based ensemble models yields the best prediction accuracy with regard to training complexity. Additionally, when comparing accuracy of forecasts for promotional sales and non-promotional sales, a variant of ARIMA achieves the most accurate results when forecasting promotional sales.

Keywords: machine learning, regression, time series analysis, sales forecasting, retail sales

CERCS: P175 informatics, systems theory

Kategooria-, toote- ning kogu kataloogi põhjal müüke prognoosivate mudelite võrdlemine

Lühikokkuvõte:

Müügi prognoosimine on tulevase müügi hindamise protsess. Selles lõputöös on katsetatud mitut meetodit parima prognoosimise täpsuse saavutamiseks madalaimate arvutuslike nõuetega.

Uuritakse kolme meetodite klassi: traditsiooniline statistiline prognoosimismeetod (ARIMA), klassikalised masinõppe tehnikad (täpsemalt ansamblimeetodid) ja kolmas põhineb tehisnärvivõrkudel (konkreetselt rekurrentsed närvivõrgud LSTM-i arhitektuuriga).

Uuringus kasutatakse Baltimaade suure jaemüügiettevõtte ajaloolisi andmeid müügitehingute kohta ja käesoleva lõputöö eesmärk on täiustada nende praegust müügi prognoosimissüsteemi.

Näitame siin, et nende praeguse müügi prognoosimise parandamine on võimalik, ja lisaks analüüsime kampaaniamüügi mõju ennustuste täpsusele. Tulemused näitavad, et mitmete üksiktoote tasemel treenitud otsustuspuul põhinevate ansamblimudelite kombineerimise kasutamine annab parima ennustustäpsuse, säilitades samal ajal arvutusliku lihtsuse. Kui võrrelda prognooside täpsust, kampaaniamüükide ja tavamüükide lõikes, siis saavutab variant ARIMA-st kõige täpsema tulemuse kampaaniamüükide puhul.

Võtmesõnad:

masinõpe, regressioon, aegridade analüüs, müügi prognoosimine, jaemüük

CERCS: P175 informaatika, süsteemiteooria

II. Licence

Non-exclusive licence to reproduce thesis

I, Laura Ruusmann,
(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to
 - (a) reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, and
 - (b) make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from 10/08/2025 until the expiry of the term of copyright,

Comparison of category-level, item-level and general sales forecasting models,

(title of thesis)

supervised by Marlon Dumas and Eerik Muuli.

(supervisor's name)

Publication of the thesis is not allowed.

2. I am aware of the fact that the author retains the right specified in p. 1.
3. This is to certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Laura Ruusmann
10/08/2020