



Buy Prediction in E-Commerce with Deep Learning

Tasmin Herrmann, Master Computer Science,
Department of Computer Science, HAW Hamburg



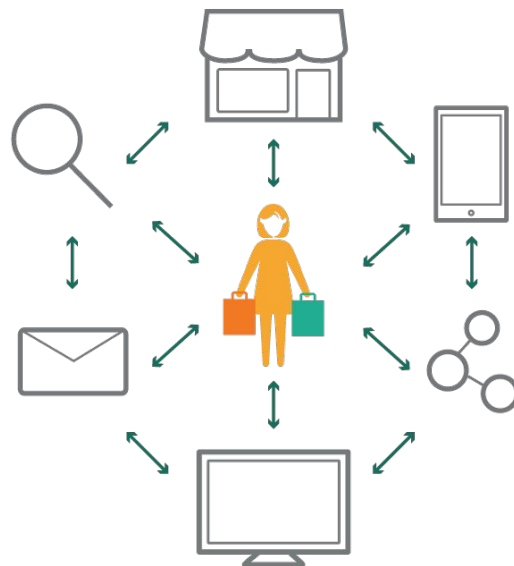
Amazon ships before you order



Outline

1. Introduction & Motivation
2. Literature Review
3. Research Question
4. Methodological Design
5. Summary

Machine Learning in E-Commerce



Customer in E-Commerce



Machine Learning Applications for E-Commerce

- Anticipatory Shipping
- Product Recommendation
- Search
 - search ranking
 - query expansion
- Dynamic Pricing
- Churn Prediction

Amazon's Supply Chain Simplified

Standard Shipping



Same Day



Prime Now



MBA@SYRACUSE

Motivation for Anticipatory Shipping



What buys the user?

When buys the user?



Data Set

- YOOCHOOSE GmbH
 - Personalization Solutions like Recommendation engines
- RecSys Challenge 2015
 - ACM RecSys Conference
- 26 Mio Clicks & 1 Mio Buys
 - Session ID
 - Timestamp
 - Item ID
 - Price
 - Quantity
 - Category



Literature Review

- Peter Romov & Evgeny Sokolov - RecSys Challenge 2015: ensemble learning with categorical features
- Eduard Weigandt - Auf Data-Mining basierende Personalisierung im E-Commerce mit implizitem Feedback



Peter Romov & Evgeny Sokolov

- 1. Which users are going to buy something?
2. What items will users buy?
- 2 staged classification with Gradient Boosting
- Feature extraction: 400 features
- Threshold optimization
- Results:
 - mean Jaccard measure of 0.765



Eduard Weigandt

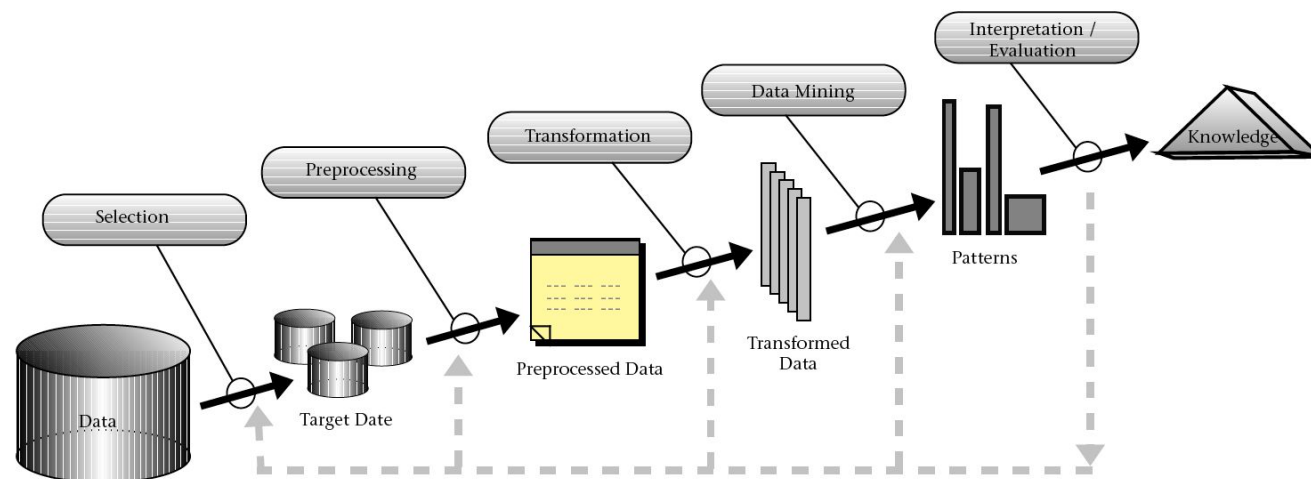
- motivated implicit feedback data for recommendation systems
- predict user behaviour with one model
 - Is the user buying in this session?
 - What is he buying?
- Gradient Boosting with XGBoost
 - accuracy 87%
- Random Forest with Bagging
 - accuracy 88,53%



Research Question

- How does an machine learning algorithm for Anticipatory Shipping look like?
- Is implicit feedback data suitable for building a recommendation systems?
- **Are deep learning models suitable for buy prediction?**
- Is a deep learning model a better buy predictor than an ensemble model?

Knowledge Discovery in Databases





Selection

- Data in HDFS
- Balance Data for product buy prediction
- Sample Data
- Select Rows/Columns



Dataset - Clicks

Session ID	Timestamp	Item ID	Category
1	2014-04-07T10:51:09.277Z	214536502	0
1	2014-04-07T10:54:09.868Z	214536500	0
2	2014-04-07T13:56:37.614Z	214662742	0



Dataset - Buys

Session ID	Timestamp	Item ID	Price	Quantity
420374	2014-04-06T18:44:58.314Z	214537888	12462	1
420374	2014-04-06T18:44:58.325Z	214537850	10471	1
281626	2014-04-06T09:40:13.032Z	214535653	1883	1



Preprocessing

- Removal of noise or outliers
- Collecting necessary information to model or account for noise
- Strategies for handling missing data fields



Transformation

- Feature Extraction
 - Features by Session
 - Features by Item in Session
- Dimensionality Reduction

Features by Session

Session time in seconds

Average time between two clicks

Day of the week

Month of the year

Features by Item in Session

Whether the item appears more than once in the session

Whether the item was clicked first in the session

Number of appearances in the session

Data Mining

- Deep learning
 - Feedforward Networks
 - Convolutional Neural Networks
 - Recurrent Neural Networks
- build prediction models with TensorFlow in Python
- other libraries: Pandas, Numpy, Scikit-Learn





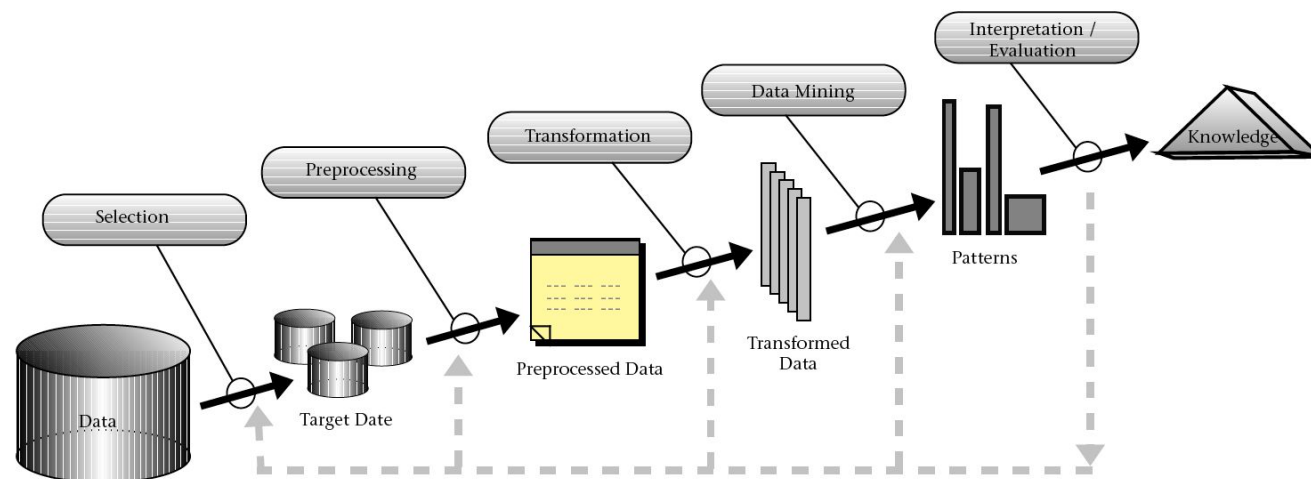
Interpretation/Evaluation

Confusion matrix & Jaccard Index

Compare to other results:

- Romov & Sokolov:
 - mean Jaccard measure of 0.765
- Weigandt:
 - Random Forest with Bagging: accuracy 88,53%
 - Gradient Boosting with XGBoost: accuracy 87%

Knowledge Discovery in Databases



Amazon's Supply Chain Simplified

Standard Shipping



Same Day



Prime Now



MBA@SYRACUSE

Motivation for Anticipatory Shipping



Risks of work

- Data Set is not enough to answer the questions
- Quality of the data set
- Deep learning model is not better than Ensemble Learning Models



Summary

What?

- Analyse:
 - Are deep learning models suitable for buy prediction?
 - Is a deep learning model a better buy predictor than an ensemble model?

How?

- Knowledge Discovery in Databases with Python Libraries like TensorFlow



Sources I

Papers:

- Peter Romov & Evgeny Sokolov - RecSys Challenge 2015: ensemble learning with categorical features
- Eduard Weigandt - Auf Data-Mining basierende Personalisierung im E-Commerce mit implizitem Feedback



Sources II

Pictures:

- Customer in E-Commerce:
<https://techblog.commercetools.com/top-5-machine-learning-applications-for-e-commerce-268eb1c89607> (03.05.2018)
- Motivation for Anticipatory Shipping: <https://blog.kinaxis.com/2016/04/17617/> (03.05.2018)
- Knowledge Discovery in Databases:
http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html (03.05.2018)