Predictive Analytics

Pune, September 26, 2023 (SIT)

Prof. Dr.-Ing. Markus Schaal

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Dr. Markus (Schaal) – Data Science History 1

- 1990: Start of Study, first Email sent and received
- 1993: Exchange student in Uppsala, Sweden (ERASMUS)
- 1995: Practical Traineeship with Modi Olivetti, New Delhi, India
- 1996: Diplom-Informatiker (MSc in Computer Science), University of Stuttgart
- 1996-1999: Worked with Prof. Radermacher, a member of the Club of Rome, an organisation of individuals who share a common concern for the future of humanity and strive to make a difference.
- 2000-2004: Dr.-Ing. (PhD in Engineering) at the Graduate School for Distributed Information Systems
- 2005-2010: Teaching at Bilkent University, Ankara, Turkey – Research Work on Types and Quality of Users providing Feedback, Object Rankings, or Evaluations
- 2010-2011: Teaching at METU-NCC, North Cyphrus

Dr. Markus (Schaal) – Data Science History 2

- 2011-2013: Post-Doc at UCD in Dublin, Ireland – Working with Prof. Barry Smyth and Ruihai Dong et al. on Reviews and Review Writing
- Since 2013: First Data Science Lead, then Head of Data Science, and now Data Science Advisor - Enabling Companies to use the Oil of the 21st century – their Data - to increase their business value!
- Since 2017: Professor at BSEL – Berlin School of Economics and Law – Research Work on Topic Mining for Multi-National ParlSpeech Transparency
- Fields of Application:
  1. Perfect Market Transparency - Artificial Intelligence as a tool to guide providers and consumers
  2. Objectivity and Participation - Artificial Intelligence for political Transparency

Background

- Company Webtrekk/ Mapp: Tracking Web Shopping data for e.g. Deutsche Bank, Tchibo, Esprit, etc.
- Customers are not aware of the potential of their data, at the same time huge money is spent by marketing for data scientists
- Predictions for Marketing as an off-the-shelf product of the company

Use Cases (Industry perspective)

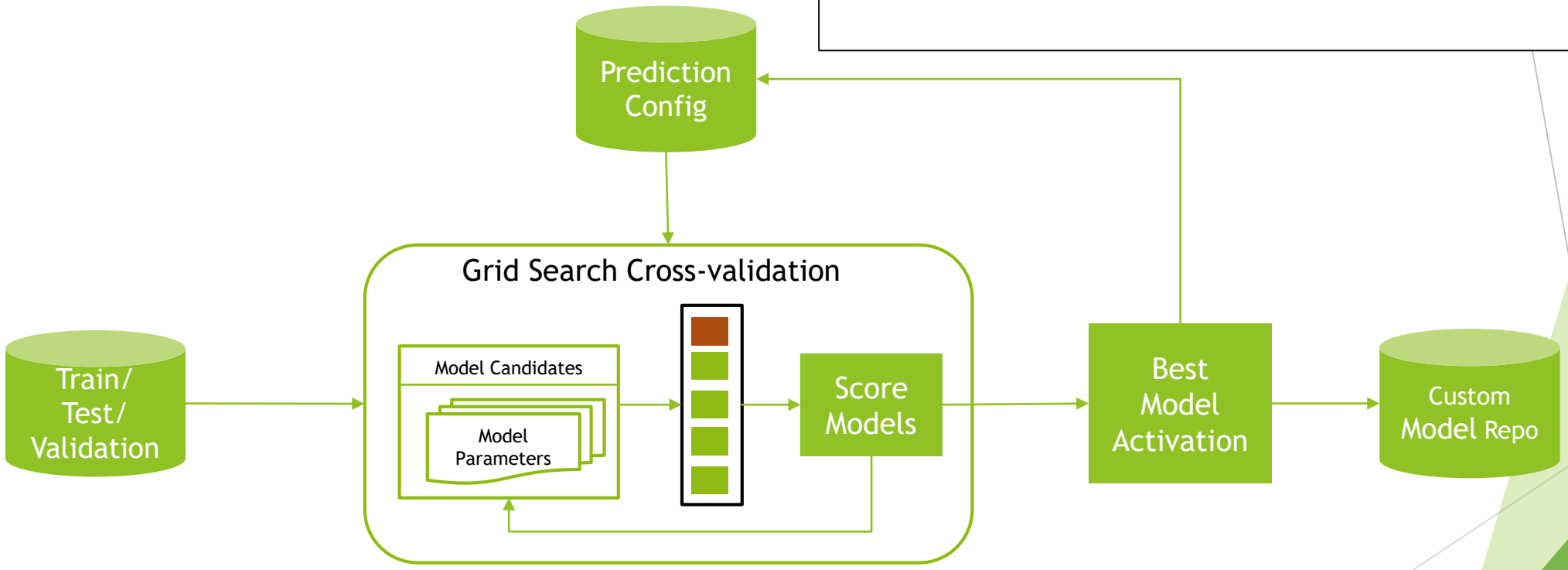
- Goal UC1: Convert high-value, low-conversion-probability users with discounts and minimal opportunity costs
- Needed UC1: Next Basket Value, Conversion Probability
- Goal UC2: Strengthen the loyalty of valuable users who are at massive risk of termination
- Needed UC2: Customer Lifetime Value, Churn Probability

Predictions: PCLV, NOV, Churn, Conv

- **Churn Probability:** The likelihood of a visitor not returning.
- **Conversion Probability:** The likelihood of a visitor to place an order within a certain time limit.
- **Next Order Value (NOV):** The expected value of the next order, which – together with the conversion probability – is a perfect proxy for the expected order value in the near future.
- **Potential Customer Lifetime Value (PCLV):** The potential customer lifetime value is defined here as the expected order value in the next year, under the condition that the customer does not churn.

Off-the-Shelf Rollout

REST Interface: The relevant Python libraries (scikit-learn, pandas, etc.) have been encapsulated to allow for algorithm choice, meta-learning, automated feature selection, and Import/ Export functionalities.



Churn Probability

The **churn probability** represents the visitor's likelihood of not returning.

- **Features:** previous visits, orders, order value, orders-per-visit, value-per-order, etc.
- **Learning:** Calibrated for a sample of all visitors based on their churn behavior in the past, using **logistic regression** by default*)
- **Quality Measurement:** e.g. MCC (Matthews Correlation Coefficient)

Conversion Probability

The **conversion probability** represents the visitor's likelihood of conversion.

- **Features:** previous visits, orders, order value, orders-per-visit, value-per-order, etc.
- **Learning:** Calibrated for a sample of all visitors based on their conversion behavior in the past, using **logistic regression** by default*)
- **Quality Measurement:** e.g. MCC (Matthews Correlation Coefficient

Next Order/ Basket Value (NOV/ NBV)

The **next order value** represents the visitor's expected next order value.

- **Features:** previous visits, orders, order value, orders-per-visit, value-per-order, etc.
- **Learning:** Calibrated for a sample of buyers based on their next order values in the past, using a weighted RFM model **with linear regression** (and sanity checks) by default*)
- **Quality Measurement:** R^2 (Coefficient of Determination) + MSE

Customer Lifetime Value

The **customer lifetime value** gives an estimate of the **expected conversion value of a buyer** in the coming year.

- **Features:** previous visits, orders, order value, orders-per-visit, value-per-order, etc.
- **Learning:** Calibrated for a sample of returning (non-churned) buyers based on their next month conversion values in the past, using **linear regression** (and sanity checks) by default*)
- **Quality Measurement:** R^2 (Coefficient of Determination) + MSE

Deep Learning vs. Regression

- Deep Learning can learn generalized models ...
- but it requires much data for doing so
- If the data is less, the feature space must be reduced ...
- so deep learning has less or no advantage over regression
- Therefore and because deep learning provides less explanation
- ... simple algorithms which also require less energy are usually better ...
- but here is an investigation on deep learning nonetheless:

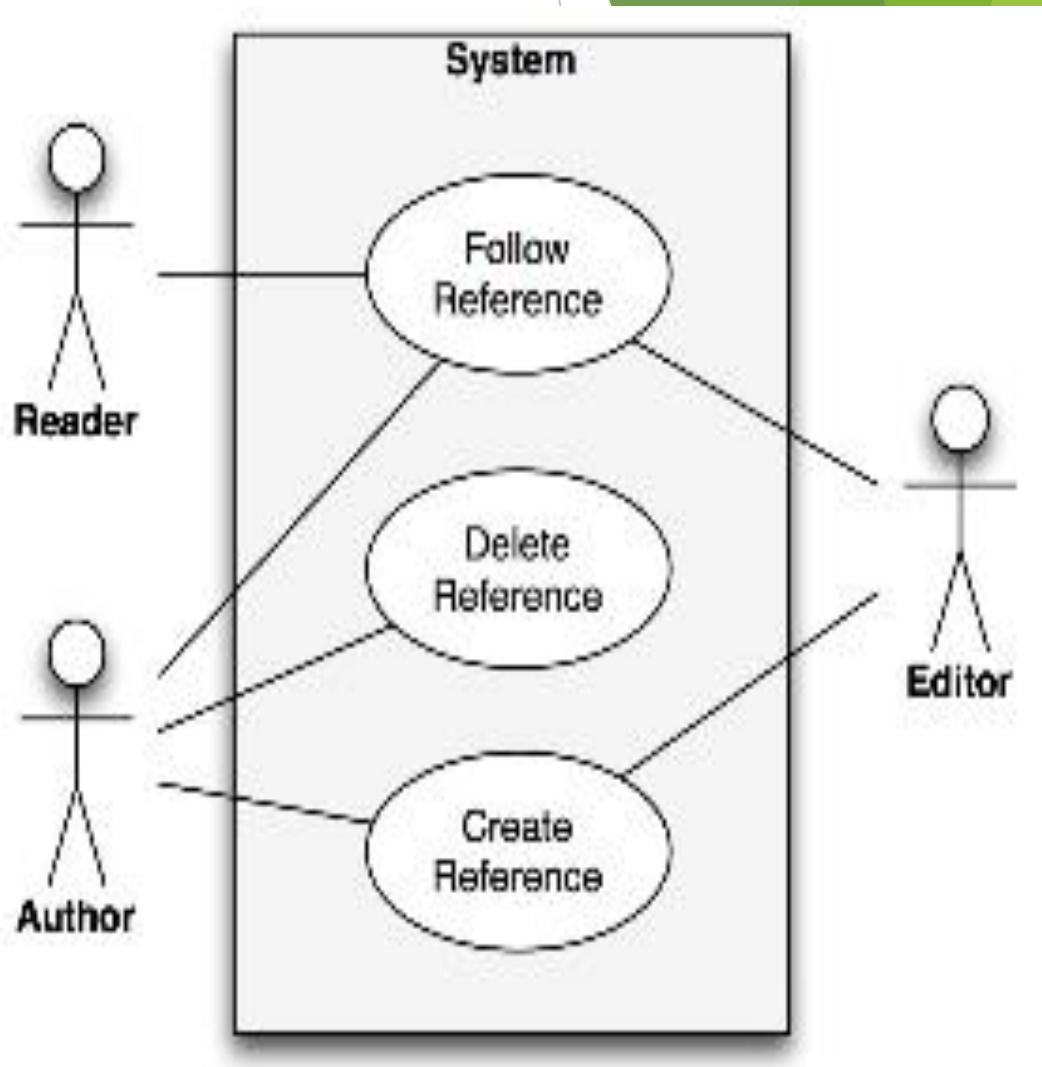
Dennis Koehn, Stefan Lessmann, Markus Schaal:
Predicting online shopping behaviour from clickstream data using deep learning. Expert Syst. Appl. 150: 113342 (2020)

Use Cases re-visited – can we design a solution? If not, why not?

- Goal UC1: Convert high-value, low-conversion-probability users with discounts and minimal opportunity costs
- Needed UC1: Next Basket Value, Conversion Probability
- Goal UC2: Strengthen the loyalty of valuable users who are at massive risk of termination
- Needed UC2: Customer Lifetime Value, Churn Probability

Use Case Modeling by Example

- Description: The understanding/ comprehension of paper-based documents is often aided by the inclusion of margin notes and by highlighting important sections using fluorescent marker pens. Since many people are now reading texts in electronic form, a program which can provide such aids to understanding would be a useful tool.
- Use Cases: Margin Notes (skipped), Highlighting (skipped), References (create, delete, follow)
- Use Case „Create Reference“ on next slide …



Use Case Example „Create Reference“

Name: Create Reference
Participating Actors: Author, Editor
Entry Condition: Text is opened and Editor has navigated to the source area of the planned reference.
Exit Condition: Reference is placed and can be followed.
Event Flow:

1. Editor marks text fragment (Use case: Mark Text) as source for the reference.
2. Editor selects function "Create Reference".
3. Editor navigates to destination of reference (Use case: Navigate).
4. Editor marks text fragment (Use case: Mark Text) as destination of reference.
5. Editor completes the creation of the reference and adds an explanatory text.
6. System shows the reference.