Big Data and Marketing – From Marketing Analytics to Customer Orientation

In today's digital world marketers have access to millions of data points which are helpful to understand and predict consumer and customer behavior. With the help of social media data, clickstream and web journey data, as well as consumption data marketers can better segment and target consumers, enhance cross-selling through recommendations, predict customer churn, optimize consumer communication and marketing budget allocation. This said, the many chances through data availability come at a price: they require marketers to understand complex analytical methods that all rely on artificial intelligence and machine learning.

In this seminar students will have the opportunity to familiarize themselves with machine learning in general and will learn how to use common standard methods in AI and machine learning to address common marketing tasks as described above.

Students will have the opportunity to gain first-hand experience by applying different ML algorithms to real corporate data sets with the help of R.

The seminar will be accompanied by the opportunity to familiarize yourself prior to the seminar with R through different web courses, which will allow students to develop a first understanding of the software and coding environment.

Content

The seminar intends to introduce students to the latest developments in marketing with a strong emphasis on big data analytics. Students shall acquire hands on experience by applying machine learning algorithms to customer data to understand and learn how such algorithms can help managers making better marketing decisions. Topics will cover fields such as customer segmentation, targeting, market-basket analysis and product recommendations, fraud detection, churn- and conversion-prediction, as well as marketing budget allocation and marketing mix modeling.

General Remarks

Students are required to develop own insights into how different approaches of unsupervised and supervised machine learning algorithms work. The provided chapter

Kübler, Wieringa, and Pauwels (2017) Machine Learning and Big Data. In: Leeflang P., Wieringa J., Bijmolt T., Pauwels K. (eds) Advanced Methods for Modeling Markets. International Series in Quantitative Marketing. Springer, Cham

will help as a starting point to get a basic understanding of what machine learning means and which different main fields in machine learning exist. Beside this, students may perceive the following two readings as helpful:

Provost, F. and T. Fawcett (2013): Data Science for Business – What you need to know about data mining and data-analytic thinking, O'Reilly, Cambridge

Lantz, B. (2015): Machine Learning with R, 2nd edition, Packt Publishing, Birmingham

For each topic a starting list of literature is provided that shall help students to acquire a basic overview of how a method works and how the specific method has been applied in marketing research. Students are required to broaden that list by identifying further work in their particular field (with a strong focus on marketing applicability).

Seminar Thesis

For the seminar thesis students are required to hand in a 15-page (Times New Roman, double spaced, 12 point) essay.

Students will require to carefully assess the data set provided. Like in marketing practice you shall not take it for granted that the data will be in perfect order. Observations may be missing or suitable variables may need to be calculated by the students (data/feature engineering). It may also be a good idea to use all sorts of available descriptive tools to get more familiar with the data set. This includes plots as well as the use of descriptive analytics. The descriptive work may already point you on first insights from the data that may guide your further analysis!

Students shall then use R (we strongly recommend to use R together with R-Studio which tremendously increases usability) to apply their specific machine learning method to the data set to answer the specific research questions provided with each topic. Beside a short list of starting literature specific to each method, the references also contain information on which packages in R are most helpful for a specific topic. Still, you may discover that there are other packages available in R, which are equally (or even more) helpful. You are free to use any package (presumed that it applies your machine learning method).

Students need to (again individually) document their approach in their seminar thesis, and report and discuss their empirical findings (with respect to the provided research question).

Furthermore, students are required to present their seminar thesis in a 60-minute presentation (45 minutes presentation time, 15 minutes discussion).

Topics

1. **Customer Segmentation with Unsupervised Learning Algorithms**
   Customers are heterogeneous with respect to preferences and shopping behavior. To be able to address consumers adequately, marketers need to understand how to properly separate consumers in homogenous subgroups. Different unsupervised algorithms such as k-mean or n-neighbor are suitable to split shopping data or other behavioral observations into meaningful groups. You will have the opportunity to use a large set of survey data from the airline industry to identify different clusters of airline

customers to give airlines recommendations which clusters deserve more or less attention and how to optimally address the different clusters.

2. **Product Recommendations with Unsupervised Learning Algorithms**
   Not only consumers are heterogenous. Products are too. Movies, beverages, tv-shows, food, and many other products come at very different tastes and sizes, feature different content and flavors. Marketers therefore need to understand which product meets which needs and preferences to make suitable recommendations. No matter if you sell books or food, whether you want to suggest a movie or a music piece, modern recommender systems require you to match a consumer's taste with specific product attributes. Recommender systems can use basic unsupervised algorithms such as e.g., k-mean clustering or other distance-based methods to make quiet powerful suggestions. Relying on a whiskey and wine data set, you will have the opportunity to develop your very own recommender system.

3. **Market Basket Analysis Tools with Unsupervised Learning Algorithms**
   Customers who bought this, also bought that… We are all used to these types of cross-selling mechanisms when it comes to online shopping. And the days recommenders were limited to Amazon are long gone. Using market basket analysis, companies can enhance cross-selling by observing which products are frequently bought together. In this topic, you will have the opportunity to develop your own market-basket recommender with the help of a large shopping data set from one of Europe's largest online retailers.

4. **Churn Prediction with Supervised Learning Algorithms**
   Acquiring customers is quite costly. Loosing and re-acquiring is expensive! Therefore, it is important for marketers to predict which consumers are likely to quit your company before they do so. Recent years showed that supervised learning algorithms are very helpful to achieve this. With the help of support vector machines and neural networks, you can use previous churn data to prevent future customers from leaving the company. You will have access to real data from a large German telecommunication company to train your own model and your own prediction skills.

5. **Marketing Mix Modeling with Supervised Learning Algorithms**
   Let me guess, you already know that half of your ad budget is wasted? But you don't know which one? Marketing Mix Models can be quite helpful to understand how specific channels and ads are convincing customers to behave in your intended way so that you can kill the non-working one. You will have access to click-stream data from an online retailer that uses social media ads, search engine advertising and banner advertising and will need to rely on different regression models to filter out the non-working channels.

6. **Predictive Analytics with Supervised Learning Algorithms**
   In sales and B2B marketing a common problem is to predict which customers are more likely to be interested in your product than others. Only if you understand where to put effort and where to leave early, you will be able to operate efficiently. Supervised learning algorithms such as Naïve Bayes, Logistic Regressions, or CART models have been shown helpful for predicting customer behavior and interest. You will have access to a large sales data base and will be responsible for determining which customers deserve more or less attention.

7. **Preference Mining with Semi-Supervised Learning and User Generated Content**
   Consumers leave a lot of information online on social media or on review sites. Marketers can use this information to gain a better understanding of how they can

improve. In this topic you will get access to a review crawler, that will enable you to download thousands of online reviews. With the help of topic models, you will learn how to make sense of the obtained data and how to deliver specific recommendations for companies.

Schedule

Access to Training Material (DataCamp)	April 1, 2021
Kick-Off:	May 10, 2021
Selection of Topics (until):	May 13, 2021
Submission of written report	July 22, 2021
Presentations	August 13 and August 14, 2021

The syllabus as the well as the schedule are subject to change according to the preferences of the instructor and the development of the Covid-19 situation in Münster.