EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Wirtschafts- und Sozialwissenschaftliche Fakultät

LEHRSTUHL FÜR MARKETING
Prof. Dr. Dominik Papies

Universität Tübingen · LS für Marketing · Nauklerstr. 47 · 72074 Tübingen

Telefon +49 7071 29-76977
Telefax +49 7071 29-5078
dominik.papies@uni-tuebingen.de
www.uni-tuebingen.de/wiwi/marketing

Seminar „Exploring the Value of Machine Learning Methods for Marketing“

I. Type of seminar

“10 Ways Machine Learning Is Revolutionizing Marketing” – Headlines like this example from forbes.com[1] are omnipresent in the business press today. The reason is that Machine Learning methods have seen a boost in popularity in recent years, and Machine Learning methods are now widely applied in many companies across different industry sectors and business functions. Some managers even seem to believe that Machine Learning and Artificial Intelligence can solve important business problems, and for the domain of marketing, many analysts believe that Machine Learning (ML) will redefine how marketing is conducted.

The question that arises is whether this is truly a revolution that happens because Machine Learning methods can do things that we were unable to achieve with traditional statistical methods, or whether this is a hype that will soon be over.

To shed some light on the potential and the capabilities of ML relative to traditional statistical methods, students will explore and apply ML methods to marketing data sets and discuss the advantages and disadvantages of these methods, and compare them to traditional methods where applicable.

In this seminar, students will also acquire relevant tools to be prepared for writing a research based master thesis. This will be supported by an obligatory workshop on academic research as well as an obligatory workshop on presentation skills, which includes a short presentation of each student’s current state of the thesis (“research plan presentation”). On top of that, we expect and encourage active participation and interaction between students.

It is expected that students have some experience with statistical software (preferably R), and we expect that students are willing to **familiarize themselves** with new methods and approaches as well as new tools in R. The respective supervisor will support students in this.

[1] https://www.forbes.com/sites/louiscolumbus/2018/02/25/10-ways-machine-learning-is-revolutionizing-marketing/

II. Topics and introductory reading material

Topic 1 **Classifying Image Data Using Neural Networks**

Nowadays, large amounts of image data are available. Neural networks are a powerful tool to make use of the rich information contained for example in image data. They can perform tasks like classification, i.e. assigning an observation to predefined categories based on training data. The aim of this thesis is to use/build neural networks that classify image data. The first challenge is to acquire image data and to process it in an adequate format that is suitable for neural networks. In a second step, the category predictions for test images should be verified and improved. Furthermore, possible applications of image classification in a marketing context should be explored.

Literature

Levine, Barry (2017): Are neural networks the next marketing creatives? https://martechtoday.com/are-neural-networks-the-next-marketing-creatives-201643

Zwebner, Yonat, Anne-Laure Sellier, Nir Rosenfeld, Jacob Goldenberg, and Ruth Mayo (2017), We Look Like Our Names: The Manifestation of Name Stereotypes in Facial Appearance, *Journal of Personality and Social Psychology Vol. 112(4),* 527–554.

Chollet, F. & Allaire, J.J. (2018). *Deep Learning with R*. Manning Publications Company.

Data

There are various freely available datasets (e.g. http://deeplearning.net/datasets/). Students can analyze image data in a context of their choice.

Topic 2 **What Songs Tell Us About: Text Mining with Lyrics**

With the growing availability of unstructured text data, marketing analysts increasingly focus on making sense of and creating knowledge from texts. Using different text mining techniques, it is for example possible to classify texts according to characteristics like the sentiment they carry, the topics they discuss, or their lexical complexity, to name but a few.

Song lyrics are one example case of where text data occurs. Musical lyrics are omnipresent and influence our thoughts with subtle ubiquity. Messages in songs are not always obvious but are often conveyed “between the lines”, which makes this an interesting application of text mining techniques.

As a core aspect of this thesis, it is required that students gather necessary data from the Internet by accessing APIs and/or scraping websites. The thesis can further investigate questions like: Did the sentiment of chart song lyrics change over the past decades? Can we predict whether a song is sung by a woman or a man, using only the lyrics? What are the topics that artists sing about? Can song lyrics be a predictor of commercial success?

This topic represents a fun and easily accessible entrance into the world of text mining with R, and upon successful completion, students will be able to apply this text mining knowledge to a wide range of different topics and problems.

Literature

Kwartler, Ted (2017). *Text Mining in Practice with R.* 1st ed. John Wiley & Sons Ltd. (https://onlinelibrary.wiley.com/doi/book/10.1002/9781119282105)

Büschken, J., & Allenby, G. M. (2016). Sentence-Based Text Analysis for Customer Reviews. *Marketing Science 35(6)*, 953–975. (https://doi.org/10.1287/mksc.2016.0993)

Netzer, O. et al. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science 31(3)*, 521–543. (https://doi.org/10.1287/mksc.1120.0713)

Data A comprehensive list of song lyrics APIs with different scope and functionality: https://www.programmableweb.com/category/lyrics/api

---

Topic 3 **Predictive Power of Machine Learning Methods**

In many different business problems, managers and analysts have to make predictions. When will a given customer churn? Will a given loan be paid back or not? Will a customer buy product A or product B? How many streams of a given song will I be able to attract? How many units of my product will I sell next month? When making predictions in these settings, one can either rely on traditional statistical methods (e.g., regression analysis, logit models), or Machine Learning. Machine Learning methods in particular have attracted a lot of attention in recent years due to their presumed superior predictive performance. However, it is not clear whether these machine learning methods such as Support Vector Machines or Random Forest Regressions are actually superior to traditional models that predict using the same set of variables and information.

Against this background, this thesis will explore the predictive power of popular Machine Learning methods and benchmark them against traditional statistical methods. To this end, the thesis will use publicly available data and established packages in R to tackle this task.

Literature

Kübler, R. V., Wieringa, J. E., & Pauwels, K. H. (2017). Machine Learning and Big Data. In P. S. H. Leeflang, J. E. Wieringa, T. H. . Bijmolt, & K. H. Pauwels (Eds.), Advanced Methods for Modeling Markets (pp. 631–670). Cham: Springer International Publishing. (https://doi.org/10.1007/978-3-319-53469-5_19)

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research 43(2)*, 204–211.

Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications 32(2),* 277–288. (https://doi.org/10.1016/j.eswa.2005.11.037)

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research 42(4)*, 415–430.

Risselada, H., Verhoef, P. C., & Bijmolt, T. H. A. (2010). Staying Power of Churn Prediction Models. *Journal of Interactive Marketing 24(3)*, 198–208. (https://doi.org/10.1016/j.intmar.2010.04.002)

Data Publicly available data sets such as Dominick's Finer Food, Spotify Streams, Kaggle, or other data sets that can be used for prediction (e.g., http://archive.ics.uci.edu/ml/datasets/Bank+Marketing)

Topic 4	**Using Deep Learning to Estimate Design Similarity**
	Given the vast amount of image data that is nowadays available, images are recognized as an important source of valuable information for marketing research and practice, as vision is the key sense that guides us through our life. One area where the analysis of images is particularly helpful is product design. A product's design is an important marketing aspect for most firms, as consumers place great value on how products look like. The thesis' key aspect is to estimate product design similarity (i.e., how similar products look). Possible applications may include: How much did the design of a product change over the past decades? How much design competition does a company face? What firms are a company's closest design competitors? How important is a product's design for its commercial success? How can design be incorporated into product recommendation systems? Students interested in this topic will first have to select an appropriate image dataset of their choice. After preprocessing the data for the deep learning task, the next task is to apply a deep learning model and extract the images' features. These image features can then, for example, be used to visualize the images' similarity on a 2D map or to calculate the images' similarity based on an appropriate distance measure. A possible application in a marketing context should be explored.
Literature	Keydana, S. (2018). *Getting started with deep learning in R*. https://blog.rstudio.com/2018/09/12/getting-started-with-deep-learning-in-r/ RStudio (2018). *R interface to Keras*. https://keras.rstudio.com/ Chollet, F. & Allaire, J.J. (2018). *Deep Learning with R*. Manning Publications.
Data	There are various freely available datasets (e.g. https://www.kaggle.com/datasets, http://deeplearning.net/datasets/). Students can analyze image data in a context of their choice.

III. General recommended literature:

- Kübler, R. V. et al. (2017). Machine Learning and Big Data. In P. S. H. Leeflang, J. E. Wieringa, T. H. Bijmolt, & K. H. Pauwels (Eds.), *Advanced Methods for Modeling Markets* (631–670). Cham: Springer International Publishing. (https://doi.org/10.1007/978-3-319-53469-5_19)
- Lantz, Brett (2015), *Machine learning with R*, Birmingham: Packt Publishing.
- Chintagunta, P. et al. (2016). Editorial - Marketing Science and Big Data. *Marketing Science 35(3)*, 341–342. (https://doi.org/10.1287/mksc.2016.0996)
- Wedel, M. & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing 80(6)*, 97–121. (https://doi.org/10.1509/jm.15.0413)
- https://towardsdatascience.com/beyond-the-hype-the-value-of-machine-learning-and-ai-artificial-intelligence-for-businesses-892128f12dd7 (last retrieved on January 31, 2019)
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science 349(6245)*, 255–260. (https://doi.org/10.1126/science.aaa8415)

IV. Dates

4th February – 8th March 2019 (8 pm s.t.)	Online Application – please see our website for further information.
12th April 2019	9 am s.t. – 12 pm Kick-off, topic assignment and workshop „Academic Writing", room 236 Neue Aula following introduction to data analysis, PC Lab, Nauklerstr. 47 from 12.30 s.t – 4 pm
6th May 2019	9 am s.t. – 1 pm Workshop "Presentation Skills", room 236 Neue Aula
10th or 17th May 2019	All day Research Plan Presentation, room 236 Neue Aula Final date will be announced in March.
21st June 2019	12 noon s.t. Seminar thesis is due (Chair of Marketing Office, Nauklerstr. 47) Containing 2 versions of the seminar paper with adhesive strips (https://de.wikipedia.org/wiki/Heftstreifen) and one electronic version of the seminar paper incl. analyses and files.
1st July 2019	All day (dates will be coordinated individually) Feedback Session (Nauklerstr. 47)
11th July 2019	6 pm s.t. Upload Presentation in ILIAS
12th July 2019	Seminar (all day), room Lessingweg 3, 72076 Tübingen

IV. Course credits

Students can obtain course credit (9 ECTS). To obtain course credit students must meet the following criteria:

- Students participate in all meetings listed above
- Students submit their 12 page thesis on time
- Students present their thesis during the seminar
- Students actively participate during the seminar

Tübingen, February 2019