**EBERHARD KARLS UNIVERSITÄT TÜBINGEN**

**Wirtschafts- und Sozialwissenschaftliche Fakultät**

**LEHRSTUHL FÜR MARKETING**
**Prof. Dr. Stefan Mayer**

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

Telefon +49 7071 29-76979
stefan.mayer@uni-tuebingen.de
www.uni-tuebingen.de/wiwi/marketing

## Seminar "New developments at the intersection of Data Science and Marketing"

## I. Type of seminar

In this seminar, students will work on selected topics that involve modern tools for data analysis, e.g., from the domain of Machine Learning or Causal Inference, or at the intersection of these two.

The topics can be chosen either from the list of suggested topics, or students propose their own topics. In the latter case, the suitability of the topic will be discussed with the supervisors.

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 **at least very solid skills in statistical software (preferably R or Python)**, equivalent to, e.g., a successful completion of DS400 Data Science Project Management. In addition, we expect that students are willing to **familiarize themselves** with new methods and approaches as well as new tools in R or Python. The respective supervisor will support students in this.

## II. Topics and introductory reading material

| | |
|---|---|
| **Topic 1** | **Analysis of purchasing behavior and shopping baskets** |
| | The personalization of marketing tools has long been a topic in marketing research and is becoming increasingly important also in practice. Personalized targeting is not only intended to make the purchasing process easier for customers and thus increase long-term customer loyalty, but is also an important factor for possible profit optimization from a firm perspective. Basis for personalized targeting is a detailed analysis of customers and their shopping behavior. |
| | The aim of this thesis is therefore to generate valuable insights about the purchasing behavior of customers using the freely accessible transaction data set of "Instacart". This can include various aspects: from a segmentation of households, to product affinity analyses in shopping baskets, to a forecast of which products will end up in a customer's next purchase. All these topics can be tackled with deep learning and machine learning methods. |
| **Literature** | Coussement, K., Harrigan, P., & Benoit, D. F. (2015). Improving direct mail targeting through customer response modeling. *Expert Systems With Applications*, 42(22), 8403-8412. https://doi.org/10.1016/j.eswa.2015.06.054 |
| | Gabel, S., Guhl, D., & Klapper, Daniel (2019). P2V-Map: Mapping Market Structures for Large Retail Assortments. *Journal of Marketing Research*, 56(4), 557-580. https://doi.org/10.1177/0022243719833631 |
| | Reutterer, T., & Dan, D. (2021). Cluster Analysis in Marketing Research. In Handbook of Market Research (pp. 221-249). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57413-4_11 |
| **Data** | Own simulations & https://www.kaggle.com/c/instacart-market-basket-analysis |

| | |
|---|---|
| **Topic 2** | **Sharing is Caring: Understanding key drivers that impact the marketing and commercial success of sharing business models by means of Divvy from Chicago** |
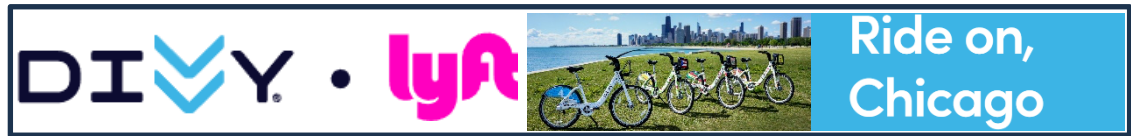


In many metropolitan and urban areas bike sharing services such as Divvy in Chicago have been noticeably gaining momentum in the last couple of years. For instance, thanks to the independence of tense traffic conditions, an inherently sustainable usage, the increased digitization and, thus, the more convenient use, many consumers heavily rely on this sort of service offerings. Divvy provides its clients with different ridable types at various fee conditions so that they may tailor the services according to their personal requirements and preferences. However, in spite of the popularity of bike sharing services, the current marketing literature considerably lacks insights about the key drivers which might be pivotal to these providers' commercial success.

The objective of this project is to generate marketing insights for research and practice, whereas the research focus can be directed on one or several marketing-relevant drivers. For instance, the student could map individual riding routes and detect riding patterns, identify/analyze the customer segments' usage preferences, and investigate the impact of local conditions on the providers' sales, such as hourly traffic conditions, season, or crime rates. Within the scope of this project, the student may deploy statistical modelling such as regression analyses or machine/deep learning techniques.

| | |
|---|---|
| **Literature** | Chu, J., Duan, Y., Yang, X., & Wang, L. (2021). The Last Mile Matters: Impact of Dockless Bike Sharing on Subway Housing Price Premium. *Management Science*, *67*(1), 297–316. https://doi.org/10.1287/mnsc.2019.3550 |
| | Eckhardt, G. M., Houston, M. B., Jiang, B., Lamberton, C., Rindfleisch, A., & Zervas, G. (2019). Marketing in the Sharing Economy. *Journal of Marketing*, *83*(5), 5–27. https://doi.org/10.1177/0022242919861929 |
| | Zhang, S., Lee, D., Singh, P., & Mukhopadhyay, T. (2022). Demand Interactions in Sharing Economies: Evidence from a Natural Experiment Involving Airbnb and Uber/Lyft. *Journal of Marketing Research*, *59*(2), 374–391. https://doi.org/10.1177/00222437211062172 |
| **Data** | https://www.kaggle.com/datasets/grigoriosmiliadis/divvy-trips-202209-202308 ; https://divvybikes.com/ & further complementary publicly available data |

| Topic 3 | **Investigating Similarity Measures for Embedding Spaces and their Alignment with Human Perception** |
|---|---|
| | Recent advances in deep learning have enabled the transformation of discrete entities such as text, images, audio, and video into dense, real-valued vector representations, or embeddings. These embeddings offer several advantages, most notably the ability to capture semantic similarities between entities that were previously considered atomic or categorical. One of the most widely used metrics for assessing similarity between embeddings is cosine similarity. Its popularity stems from the fact that it emphasizes the directional alignment between vectors rather than their magnitude, which is often seen as less relevant in many practical applications. |
| | While the use of cosine similarity has been extensively reported as effective in numerous studies, recent evidence suggests that other similarity measures, such as the (unnormalized) dot product, may outperform cosine similarity in certain contexts (e.g., see [1]). This raises important questions about the general efficacy of cosine similarity, especially when compared to alternative metrics. |
| | The aim of this project is to systematically evaluate different similarity measures in terms of how well they correlate with human perception of similarity. We will gather human judgment data through a survey-based approach and compare these assessments with the output of various similarity measures, such as cosine similarity, dot product, and others, across diverse datasets. By conducting this study on multiple domains, we aim to produce findings that are both robust and generalizable, thereby providing insights into the alignment between mathematical similarity metrics and human semantic perception. |
| | This research has the potential to enhance our understanding of how well current embedding-based similarity measures capture human-like notions of similarity, with possible implications for improving the design and selection of these measures in real-world applications. |
| **Literature** | [1] K. Zhou, K. Ethayarajh, D. Card, and D. Jurafsky. Problems with cosine as a measure of embedding similarity for high frequency words. In 60th Annual Meeting of the Association for Computational Linguistics, 2022. |
| **Data** | Publicly available datasets |

| | |
|---|---|
| **Topic 4** | **Evaluating Disentangled Representations for Product Embeddings and their Alignment with Human Perception** |

Deep learning models have demonstrated significant effectiveness in generating dense, real-valued vector representations, or embeddings, for products based on their images. However, recent research on shortcut learning has revealed that neural networks often learn to represent objects in ways that differ from human perception. These networks tend to focus on spurious correlations or features that may not align with the attributes humans consider important.

A potential solution to this problem is to guide the model to learn disentangled representations - representations that capture distinct, human-relevant attributes of the products. Such disentanglement could lead to more interpretable and human-like embeddings, which better reflect the attributes that matter most to users.

In this project, we aim to investigate whether embeddings that explicitly represent human-centric attributes of products are more aligned with human perception of similarity compared to embeddings generated by standard pre-trained models. We will conduct a field study where users will be asked to assess the similarity of various image pairs. In parallel, we will compute the similarity between these image pairs using both the disentangled embeddings and embeddings from an out-of-the-box pre-trained network.

By comparing the correlation between the human judgments and the model-generated similarities from both types of embeddings, we hope to determine the potential benefits of disentangled representations in providing a more human-aligned understanding of products. This research could inform the design of more interpretable and effective deep learning models for real-world applications, particularly in domains such as e-commerce, recommendation systems, and image-based search.

| | |
|---|---|
| **Literature** | Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard S. Zemel, Wieland Brendel, Matthias Bethge, Felix A. Wichmann: Shortcut Learning in Deep Neural Networks. |
| **Data** | Publicly available datasets |

- **III. Dates**

| | |
|---|---|
| October 13, 2024 | Online Application via Alma |
| October 18, 2024 | 9:00 a.m s.t. – 1:00 p.m. s.t.– Room SR 331 (Mohlstr. 36)<br>Kick-off and topic assignment<br>Workshop „Academic Writing" |
| November 6, 2024 | 9:00 a.m. s.t. – 1:00 p.m. s.t. – Room FSR (Nauklerstr. 47)<br>Workshop "Presentation Skills" |
| November 27, 2024 | 8:00 a.m. s.t. – 2:00 p.m. s.t. – Room 063 (Neue Aula)<br>Research plan presentation |
| December 20, 2024 | Term paper is due by noon (12 p.m. s.t.)<br>(You can drop your term paper in the letterbox outside<br>the faculty (addressed to Chair of Marketing - Nauklerstr. 47) or send it by post (postmark date is relevant).)<br>2 copies of the term paper with a filing<br>clip (https://de.wikipedia.org/wiki/Heftstreifen) must be submitted.<br><br>Submit the electronic version (pdf) of the term paper incl. analysis scripts as file upload in ILIAS. |
| January 13, 2025 | Feedback Session – meeting schedule will be coordinated individually |
| January 23, 2025 | 8:00 p.m.<br>Upload Presentation in ILIAS |
| *January 24, 2025* (tentative, subject to change) | All day Seminar – Room FSR (Nauklerstr. 47) |

## 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

Approx. 50% of the final grade will be the thesis, and 50% of the final grade will be the presentation and the participation in the seminar.

**Please note:**

Topics are subject to change - Students are invited to propose their own topics that fit under the general theme of the seminar.

Tübingen, October 2024