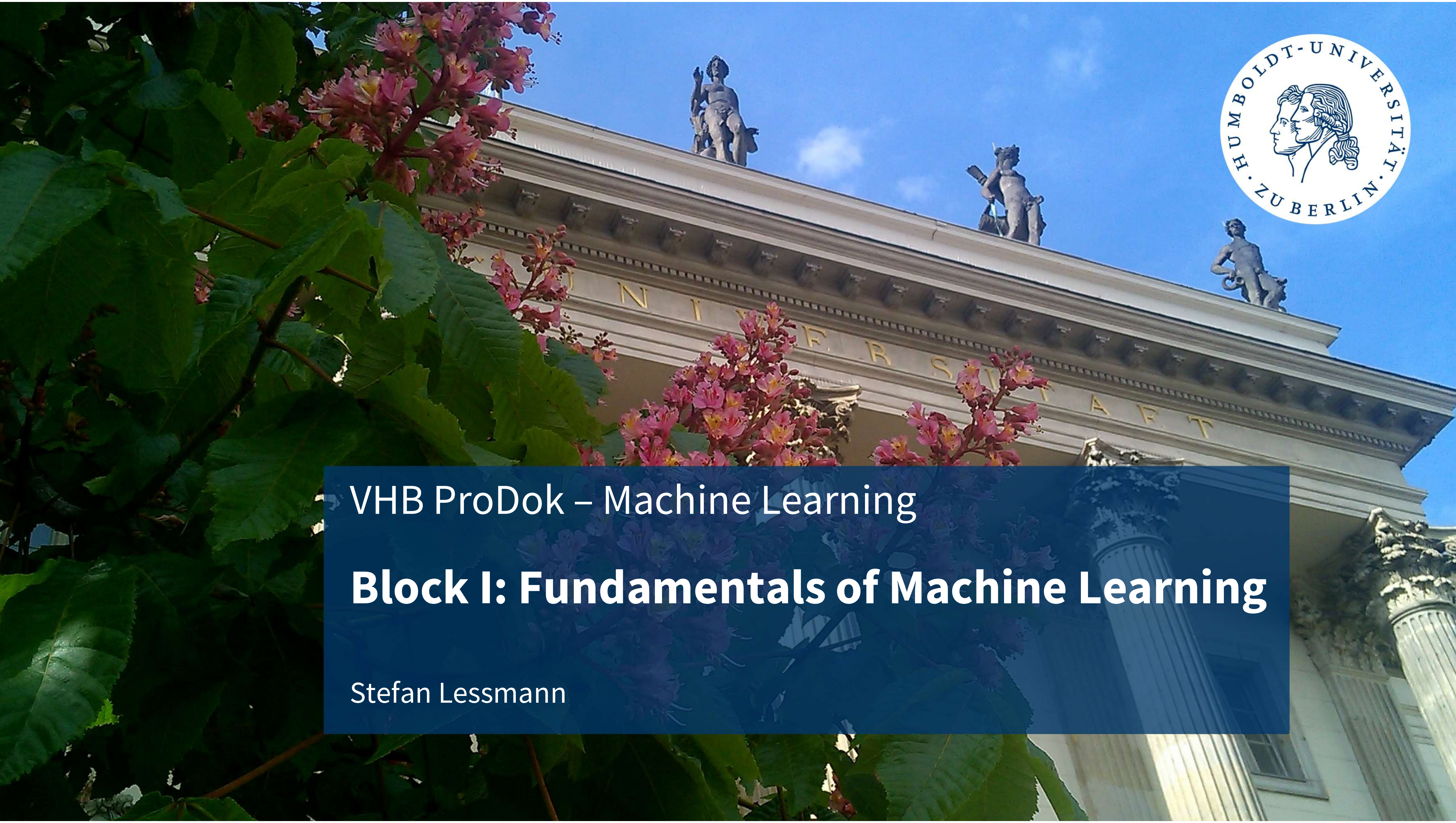
VHB ProDok – Machine Learning

# Closing Sessions

Stefan Lessmann

## Many Open Questions…

- **How use ML for research?**
- **Where/how to learn more?**
- **How about…**
  - LMMs
  - Reinforcement learning
  - Time Series
  - Machine Learning for XYZ



VHB ProDok – Machine Learning

# Block I: Fundamentals of Machine Learning

Stefan Lessmann

Received: 6 July 2021 | Accepted: 22 April 2023

DOI: 10.1111/poms.14021

POMS

ORIGINAL ARTICLE

# Addressing distributional shifts in operations management: The case of order fulfillment in customized production

Julian Senoner[1] | Bernhard Kratzwald[1] | Milan Kuzmanovic[1] | Torbjørn H. Netland[1] | Stefan Feuerriegel[2]

[1]ETH Zurich, Department of Management, Technology and Economics, Zurich, Switzerland

[2]LMU Munich, Institute of Artificial Intelligence (AI) in Management, Munich, Germany

**Correspondence**
Julian Senoner, ETH Zurich. Department of Management, Technology, and Economics. 8092 Zurich, Switzerland.
Email: jsenoner@ethz.ch

**Handling Editor**: Panos Kouvelis

**Funding information**
Norges Forskningsråd, Grant/Award Number: 309810

**Abstract**

To meet order fulfillment targets, manufacturers seek to optimize production schedules. Machine learning can support this objective by predicting throughput times on production lines given order specifications. However, this is challenging when manufacturers produce customized products because customization often leads to changes in the probability distribution of operational data—so-called *distributional shifts*. Distributional shifts can harm the performance of predictive models when deployed to future customer orders with new specifications. The literature provides limited advice on how such distributional shifts can be addressed in operations management. Here, we propose a data-driven approach based on adversarial learning, which allows us to account for distributional shifts in manufacturing settings with high degrees of product customization. We empirically validate our proposed approach using real-world data from a job shop production that supplies large metal components to an oil platform construction yard. Across an extensive series of numerical experiments, we find that our adversarial learning approach outperforms common baselines. Overall, this paper shows how production managers can improve their decision making under distributional shifts.

**KEYWORDS**

adversarial learning, distributional shifts, machine learning, manufacturing, order fulfillment

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor


Innovative Applications of O.R.

# A transformer-based model for default prediction in mid-cap corporate markets

Kamesh Korangi[a,b,*], Christophe Mues[a,b], Cristián Bravo[c]

[a] Department of Decision Analytics and Risk, Southampton Business School, University of Southampton, University Road, SO17 1BJ, United Kingdom
[b] Centre for Operational Research, Management Sciences and Information Systems (CORMSIS), University of Southampton, University Road, SO17 1BJ, United Kingdom
[c] Department of Statistical and Actuarial Sciences, The University of Western Ontario, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

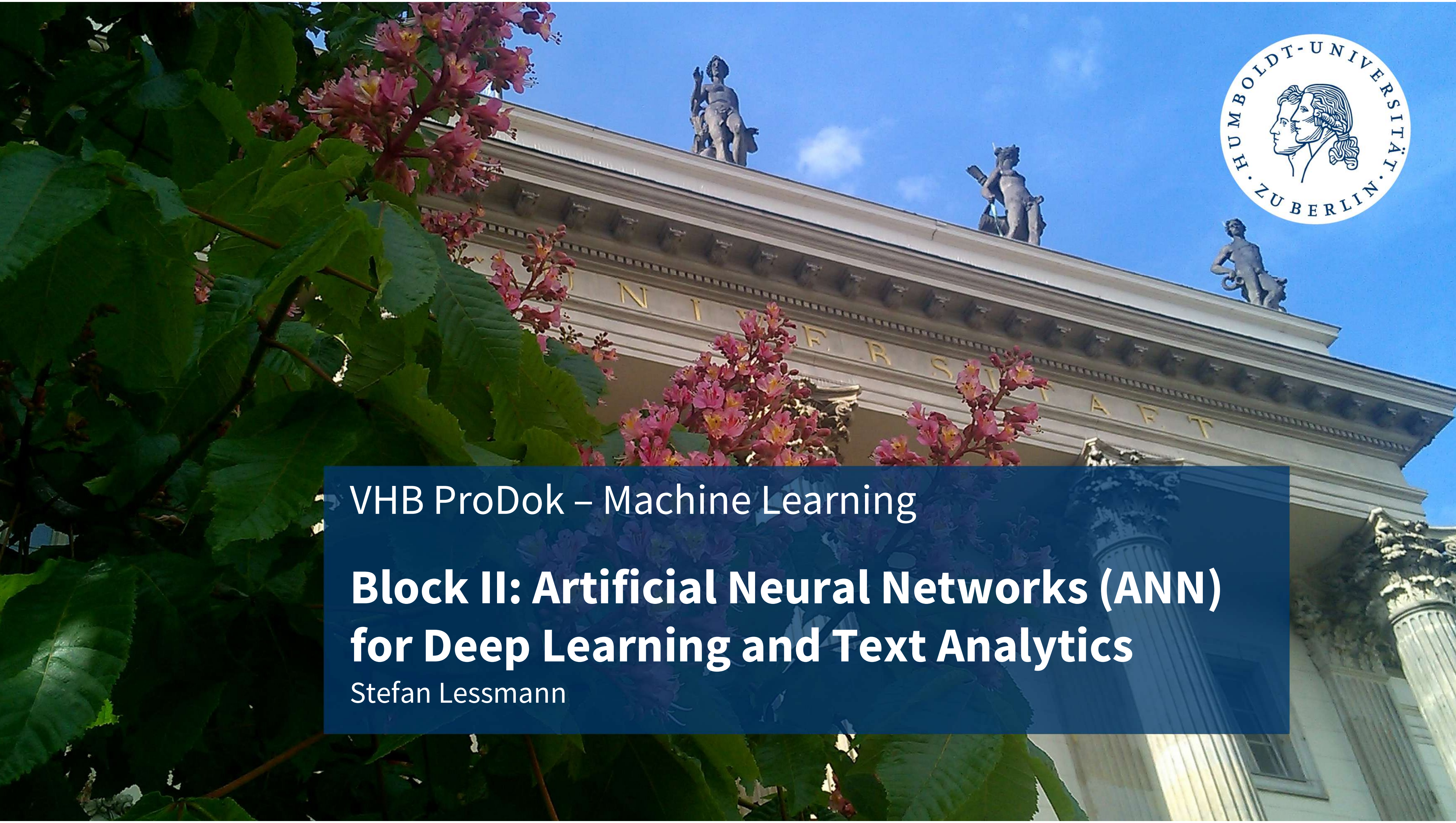
ARTICLE INFO

*Article history:*
Received 15 November 2021
Accepted 19 October 2022
Available online 27 October 2022

*Keywords:*
OR in banking
Mid-cap credit risk
Default prediction
Deep learning
Transformers

ABSTRACT

In this paper, we study mid-cap companies, i.e. publicly traded companies with less than US$10 billion in market capitalisation. Using a large dataset of US mid-cap companies observed over 30 years, we look to predict the default probability term structure over the short to medium term and understand which data sources (i.e. fundamental, market or pricing data) contribute most to the default risk. Whereas existing methods typically require that data from different time periods are first aggregated and turned into cross-sectional features, we frame the problem as a multi-label panel data classification problem. To tackle it, we then employ transformer models, a state-of-the-art deep learning model emanating from the natural language processing domain. To make this approach suitable to the given credit risk setting, we use a loss function for multi-label classification, to deal with the term structure, and propose a multi-channel architecture with differential training that allows the model to use all input data efficiently. Our results show that the proposed deep learning architecture produces superior performance, resulting in a sizeable improvement in AUC (Area Under the receiver operating characteristic Curve) over traditional models. In order to interpret the model, we also demonstrate how to produce an importance ranking for the different data sources and their temporal relationships, using a Shapley approach for feature groups.



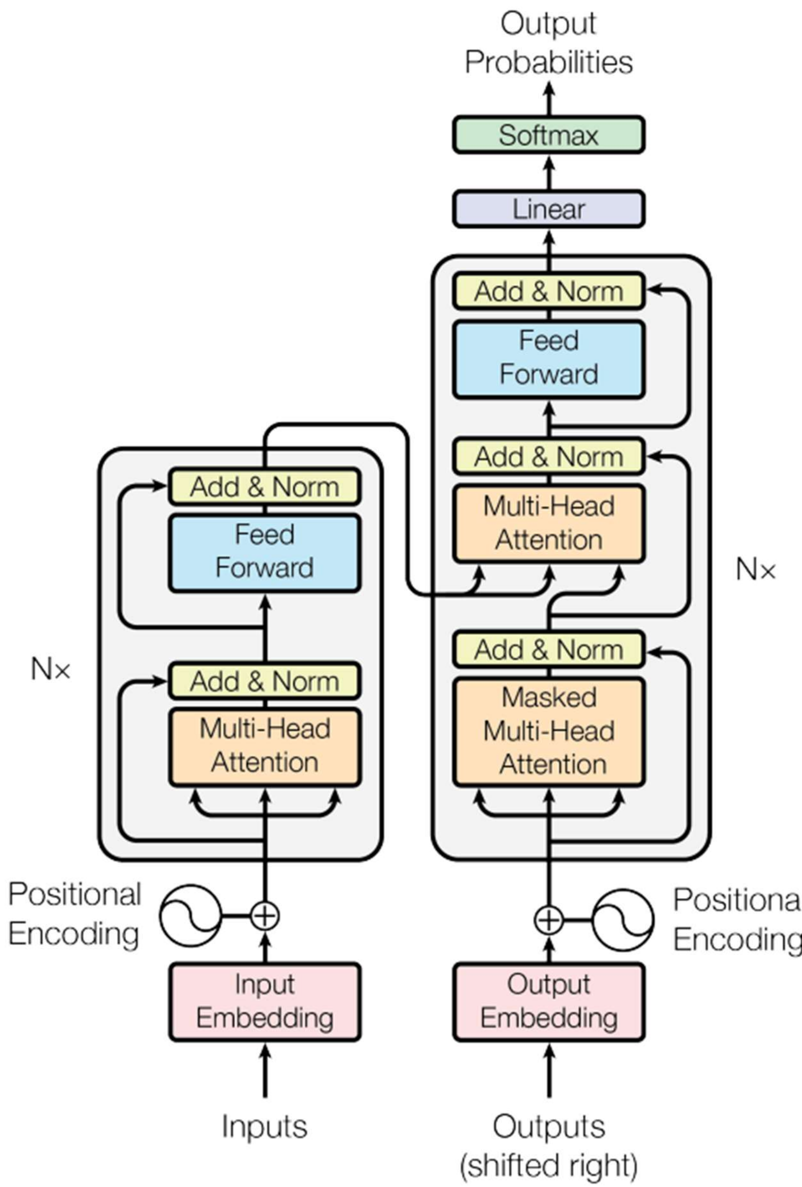
VHB ProDok – Machine Learning

# Block II: Artificial Neural Networks (ANN) for Deep Learning and Text Analytics

Stefan Lessmann

## Key Concepts in NLP

- Text is a sequence of discrete symbols (aka tokens)
- Tokens are represented (aka embedded) as vectors
- Embeddings encapsulate latent characteristics of tokens and their relationships with other tokens
  - e.g., syntactic and semantic characteristics of a word
- Deep networks learn these characteristics by solving upstream NLP tasks (e.g., language modeling) using large volumes of data
- These pretrained networks provide a starting point for your analysis
  - They don't know your specific task but "understand" language
  - You may need to adjust them a to your setting



# A Step Toward Consolidating Modern NLP

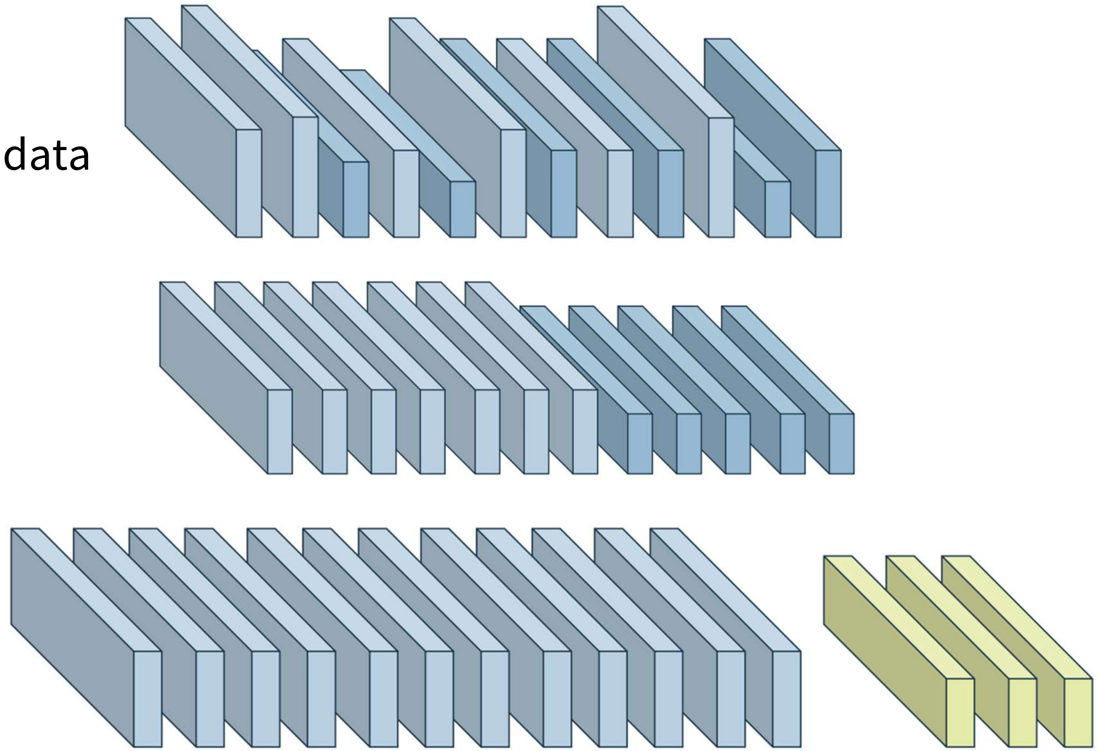
## Some considerations and take-aways

- **When working on a NLP task, start from a pre-trained model**
  - The pre-training was done on corpus far larger than what we can access or process
  - Therefore, a pre-trained model possesses an advanced understanding of language
  - Even if your text data is very specific, the general understanding of language in a pre-trained model will almost certainly be valuable (exception: unsupported languages)
- **Pre-trained (transformer) models are available for many tasks**
  - Text classification, sentiment analysis, topic modeling, keyword extraction, summarization, etc.
  - How to adjust such models to your data?

# A Step Toward Consolidating Modern NLP

## Strategies to adjust a pre-trained model to your text

- **Start from a pre-trained model**
- **Fine-tuning**
  - ☐ Adjust parameters of all layers based on target data
  - ☐ Initialize training with pre-trained weights
- **Freezing**
  - ☐ Freeze earlier layers
  - ☐ Retrain later layers based on your data
- **Progressive learning**
  - ☐ Keep all layers unchanged
  - ☐ Add more layers (progression)



# A Step Toward Consolidating Modern NLP

## Is adjustment (aka fine-tuning) still state-of-art?

- **Zero-shot learning**
  - NLP models build for generality (e.g., BART-MNLI)
  - NLI = Natural language inference is the task of determining whether a "hypothesis" is true (entailment), false (contradiction), or undetermined (neutral) given a "premise".
  - No fine-tuning
- **General-purpose text classifier**
  - Output is a probability
  - Financial forecasting example
    - Premise: some text (e.g., tweet)
    - Hypothesis: "This example is bullish for XYZ"
    - Different from sentiment analysis!
- **And how about prompting an LLM?**

| Premise | Label | Hypothesis |
|---|---|---|
| A man inspects the uniform of a figure in some East Asian country. | contradiction | The man is sleeping. |
| An older and younger man smiling. | neutral | Two men are smiling and laughing at the cats playing on the floor. |
| A soccer game with multiple males playing. | entailment | Some men are playing a sport. |

Source: Natural language inference | NLP-progress (nlpprogress.com)

Journal of Marketing Research

AMERICAN MARKETING ASSOCIATION

Impact Factor: **5.1**
5-Year Impact Factor: **7.4**

Jou

Restricted access | Research article | First published online June 6, 2019

# P2V-MAP: Mapping Market Structures for Large Retail Assortments

Sebastian Gabel, Daniel Guhl, and Daniel Klapper View all authors and affiliations

Volume 56, Issue 4 | https://doi.org/10.1177/0022243719833631

Contents | Get access | Cite article | Share options | Information, rights and permissions

## Abstract

The authors propose a new, exploratory approach for analyzing market structures that leverages two recent methodological advances in natural language processing and machine learning. They customize a neural network language model to derive latent product attributes by analyzing the co-occurrences of products in shopping baskets. Applying dimensionality reduction to the latent attributes yields a two-dimensional product map. This method is well-suited to retailers because it relies on data that are readily available from their checkout systems and facilitates their analyses of cross-category product complementarity, in addition to within-category substitution. The approach has high usability because it is automated, is scalable and does not require a priori assumptions. Its results are easy to interpret and update as new market basket data are collected. The authors validate

**Figure 2.1.** Overview of P2V-MAP.

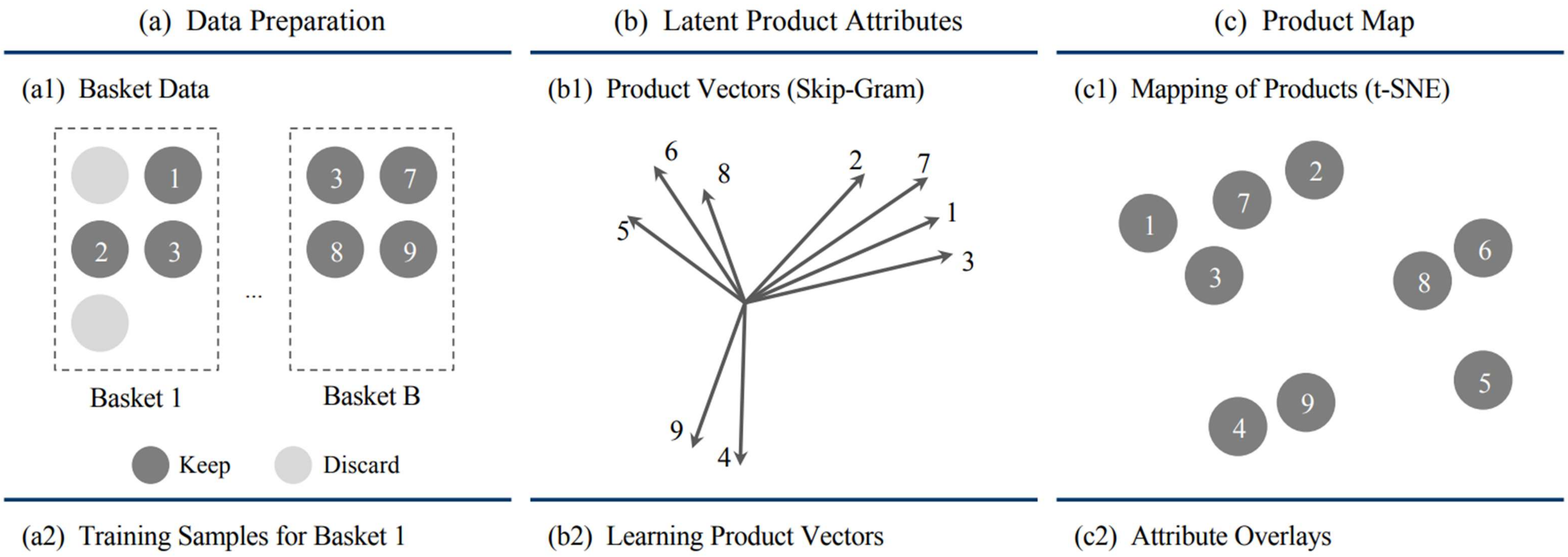
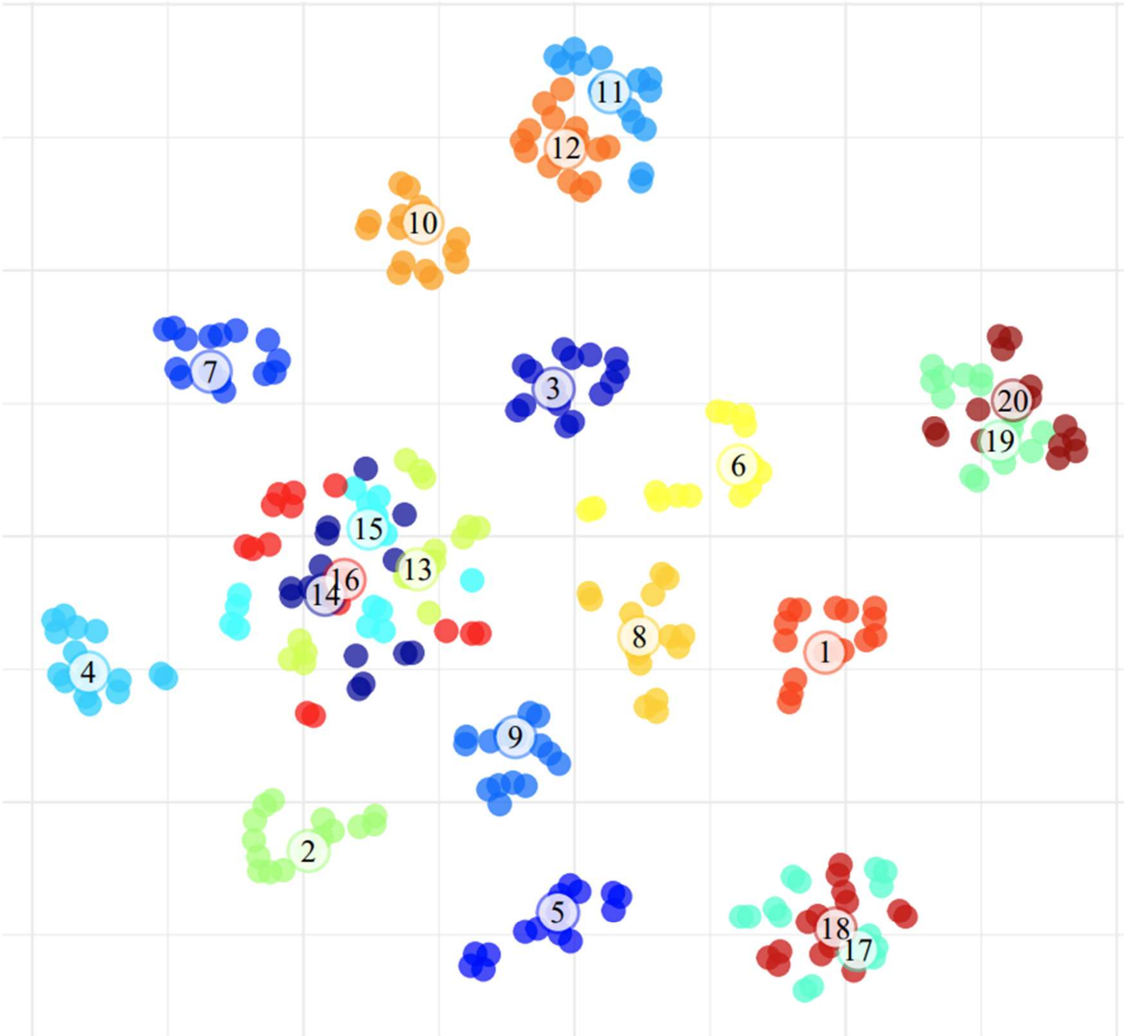


**Figure 2.3.** Product map for simulated data.




*Notes*: Replication 1 of 30. Colors indicate the true (simulated) product categories. Best viewed in color.

**Figure 2.4.** Product maps for method benchmarking.

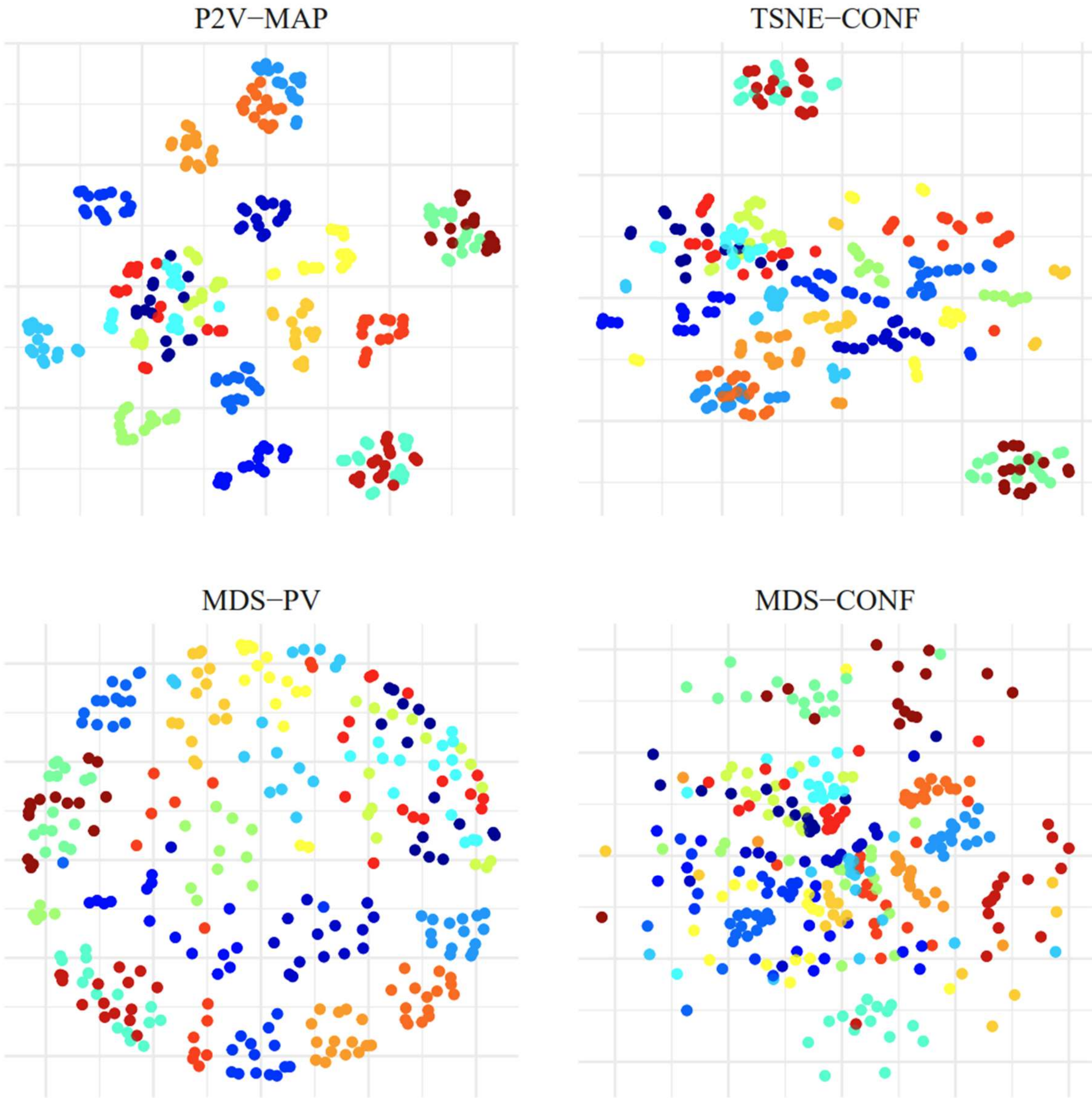
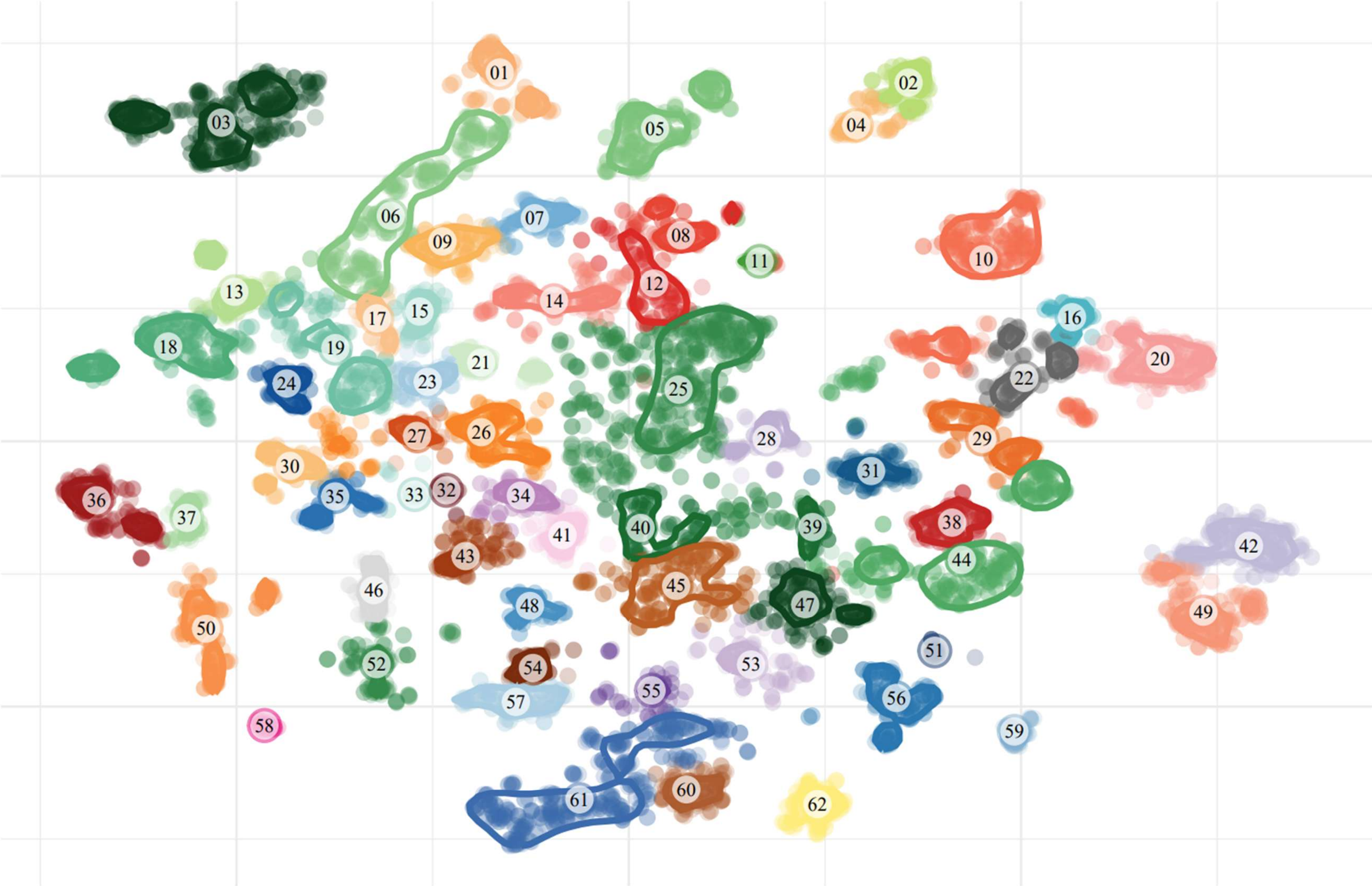


**Figure 2.5.** Product map for empirical application.




*Notes*: The contours mark cluster regions that contain 90% of all products in the respective cluster. Appendix 2.M describes the creation of the overlay and contains the cluster names. Best viewed in color.

VHB ProDok – Machine Learning

# Block III: Selected Topics in Machine Learning Research

Stefan Lessmann

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 68, No. 8, August 2022, pp. 5704–5723
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

# Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing

**Julian Senoner,**[a] **Torbjørn Netland,**[a] **Stefan Feuerriegel**[a]

[a] Department of Management, Technology, and Economics, ETH Zurich, 8092 Zurich, Switzerland

**Contact:** jsenoner@ethz.ch, https://orcid.org/0000-0001-6633-5602 (JS); tnetland@ethz.ch, https://orcid.org/0000-0001-7382-1051 (TN); sfeuerriegel@ethz.ch, https://orcid.org/0000-0001-7856-8729 (SF)

**Received:** April 15, 2020
**Revised:** December 22, 2020; May 4, 2021
**Accepted:** June 16, 2021
**Published Online in Articles in Advance:** December 9, 2021

**https://doi.org/10.1287/mnsc.2021.4190**



**Abstract.** We develop a data-driven decision model to improve process quality in manufacturing. A challenge for traditional methods in quality management is to handle high-dimensional and nonlinear manufacturing data. We address this challenge by adapting explainable artificial intelligence to the context of quality management. Specifically, we propose the use of nonlinear modeling with Shapley additive explanations to infer how a set of production parameters and the process quality of a manufacturing system are related. Thereby, we contribute a measure of process importance based on which manufacturers can prioritize processes for quality improvement. Grounded in quality management theory, our decision model selects improvement actions that target the sources of quality variation. The decision model is validated in a real-world application at a leading manufacturer of high-power semiconductors. Seeking to improve production yield, we apply our decision model to select improvement actions for a transistor chip product. We then conduct a field experiment to confirm the effectiveness of the improvement actions. Compared with the average yield in our sample, the experiment returns a reduction in yield loss of 21.7%. Furthermore, we report on results from a postexperimental rollout of the decision model, which also resulted in significant yield improvements. We demonstrate the operational value of explainable artificial intelligence by showing that critical drivers of process quality can go undiscovered by the use of traditional methods.

Cornell University

arXiv > cs > arXiv:2303.02186

Computer Science > Machine Learning

[Submitted on 3 Mar 2023 (v1), last revised 14 Feb 2024 (this version, v2)]

# Causal Deep Learning

Jeroen Berrevoets, Krzysztof Kacprzyk, Zhaozhi Qian, Mihaela van der Schaar

Causality has the potential to truly transform the way we solve a large numbe
unlocked as causality often requires crucial assumptions which cannot be te
thinking about causality -- we call this causal deep learning. Our causal deep
dimension, which incorporates partial yet testable causal knowledge rather t
variables of interest; (2) a parametric dimension, which encompasses param
of interest; and (3) a temporal dimension, which captures exposure times or
Causal deep learning enables us to make progress on a variety of real-world
independencies among variables) and quantitatively characterising causal re
framework clearly identifies which assumptions are testable and which ones
practice. Using our formulation we can combine or chain together causal rep
assumptions are required to build these solutions, pushing real-world impact
education, through causal deep learning.

van_der_Schaar \ LAB

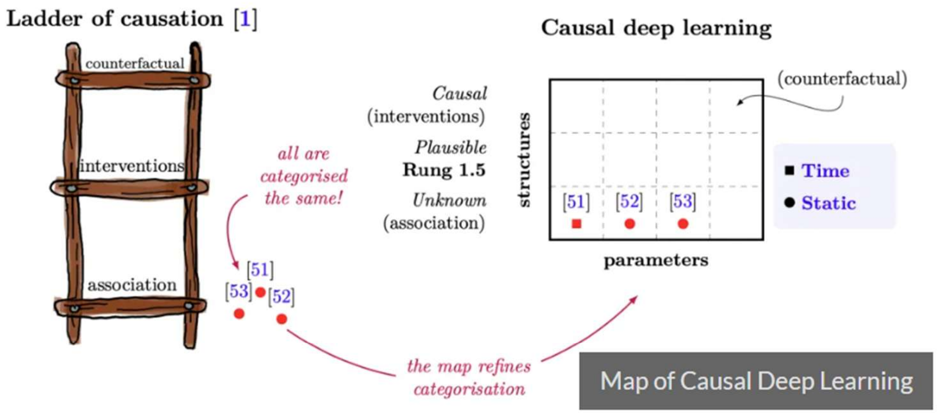
The lab | Publications | Big ideas | News | Videos | Events | Software | Engagement sess
Hub for Healthcare | Connect

BIG IDEAS | NEWS

## Causal deep learning: a new framework

Jeroen Berrevoets | Mihaela van der Schaar | March 6, 2023

2 min read




Map of Causal Deep Learning

# Thank you for your attention!


Stefan Lessmann

Chair of Information Systems
School of Business and Economics
Humboldt-University of Berlin, Germany

Tel. +49.30.2093.5742
Fax. +49.30.2093.5741

stefan.lessmann@hu-berlin.de
http://bit.ly/hu-wi


**www.hu-berlin.de**


Photo: Heike Zappe