Machine Learning facilitates bibliometric analyses of **emerging fields** with **inconsistent terminology** and enables **mapping of unexplored fields**

	Inclusion probability ≥ .9			Inclusion probability < .9 ∩ ≥ .7*		
Round of Active Learning	*n* Predicted	*n* Included	**Precision**	*n* Predicted	*n* Included	**Precision**
1	211	107	**.51**	76	43	**.57**
2	74	74	**1**	62	59	**.95**

Inconsistent Terminology

↓

Broader Search

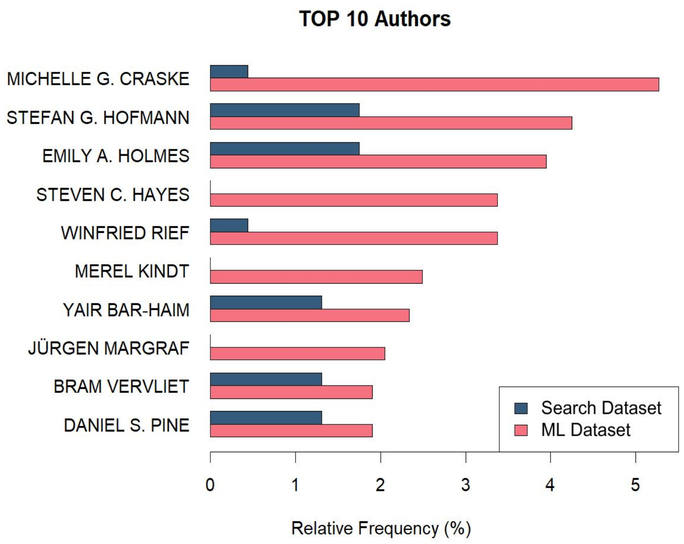
↓

False Positives

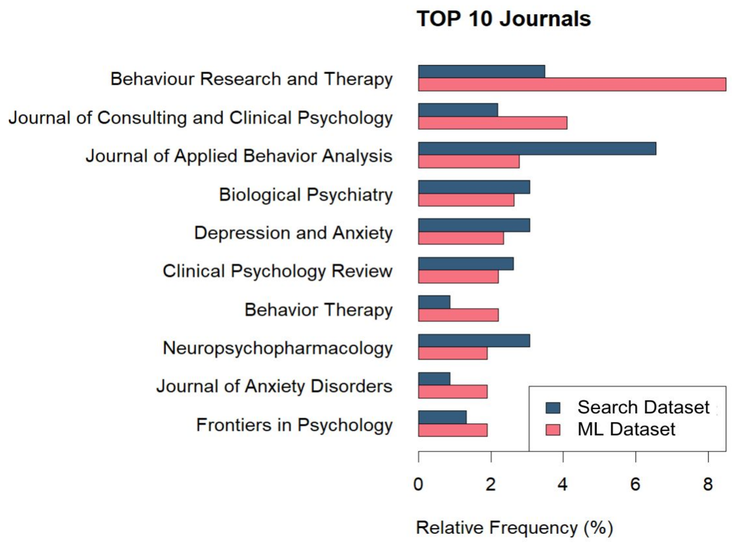
↓

ML-Assisted Screening

Criteria	Search Dataset	ML Dataset
N. of publications	**229**	**683**
Annual growth rate	**9.33**	**13.07**
M citations	**21.56**	**67.38**
Median citations	**17**	**21**
M Twitter mentions	**4.86**	**20.58**
Median Twitter mentions	**1**	**2**
Open Access	**34.74%**	**37.21%**



The most productive authors are either absent (n = 3) or poorly represented in the search dataset (TOP 10 Authors in ranking order according to ML Dataset).



A similar trend is also observed in the journal distribution, where 5 of the Top 10 journals differ in distribution by 50%.

Support Vector Machine

Citation Mining

Active Learning

Background

- Traditional bibliometric methods can be limited in their ability to analyze **emerging** fields where the object of study resists **clear delineations**, the **boundaries between subfields** are porous or the relationships between different subfields are complex.
- **Inconsistent terminology** prevents a satisfactory coverage of the construct of interest and leads to biased and distorted representations.
- **Broadening the search** dramatically increases the noise and the amount of publications needed to be screened, but ML helps semi-automatizing the screening process.

Aim

Mapping the emerging research landscape of translational psychotherapy and **comparing** the ML-augmented results with those based on a typical search query.

Methods

- PsycInfo & PSYNDEX databases
- APA-Thesaurus
- 200 000 + records
- Citation Mining
- Support Vector Machine for Screening Automation
- OpenScience Tools

zpid Leibniz-Institut für Psychologie

Leveraging machine learning for bibliometric analysis of emerging fields

Claudiu Petrule[1], André Bittermann[1], Viktoria Ritter[2], Anke Haberkamp[3], Winfried Rief[3]

[1] ZPID – Leibniz Institute for Psychology, Trier, Germany
[2] Goethe-University of Frankfurt, Dept. of Psychology
[3] Philipps-University of Marburg, Dept. of Psychology

Correspondence: cp@leibniz-psychology.org