



Multivariate Verfahren

MSc Psychologie | MSc Klinische Psychologie und Psychotherapie

WiSe 2024/25

Prof. Dr. Dirk Ostwald

Datum	Einheit	Thema
16.10.2024	Grundlagen	(0) Einführung
23.10.2024	Grundlagen	(1) Matrizen
30.10.2024	Grundlagen	(2) Eigenanalyse
06.11.2024	Grundlagen	(4) Multivariate Normalverteilungen
13.11.2024	Frequentistische Inferenz	(5) Multivariate Deskriptivstatistik
20.11.2024	Frequentistische Inferenz	(6) Multivariate Varianzanalyse
27.11.2024	Frequentistische Inferenz	(7) Kanonische Korrelation
04.12.2024	Prädiktive Modellierung	(6) Prädiktive Modellierung
11.12.2024	Prädiktive Modellierung	(7) Nichtlineare Optimierung
18.12.2024	Prädiktive Modellierung	(8) Dimensionsreduktion
	Weihnachtspause	
08.01.2025	Prädiktive Modellierung	(9) Lineare Diskriminanzanalyse
15.01.2025	Prädiktive Modellierung	(10) Logistische Regression
22.01.2025	Prädiktive Modellierung	(11) Support Vektor Maschinen
29.01.2025	Prädiktive Modellierung	(11) Neuronale Netze
07.02.2025	Klausurtermin	MSc KliPP
Juli 2025	Klausurtermin	MSc Psychologie

(8) Prädiktive Modellierung

Rhetorik

Anwendungsbeispiel

Performanzmetriken

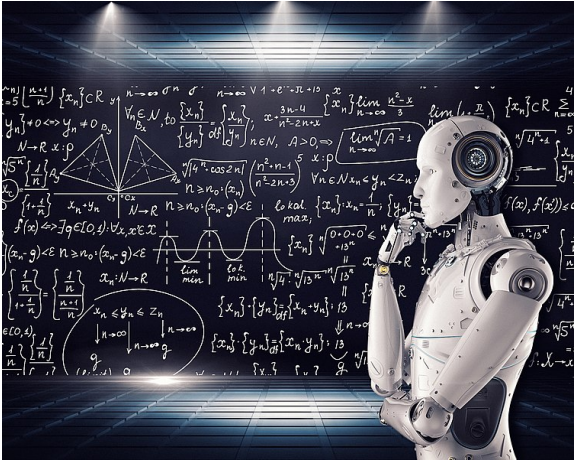
Selbstkontrollfragen

Rhetorik

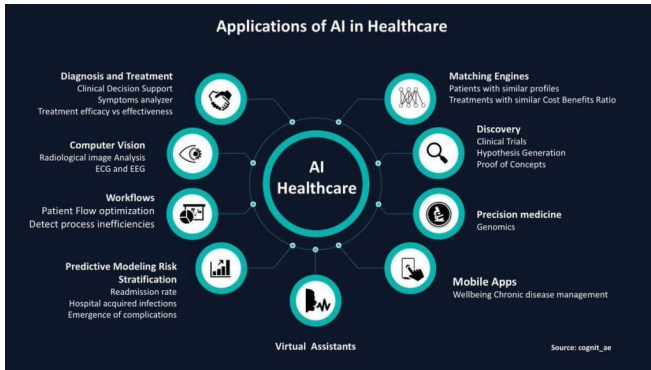
Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen



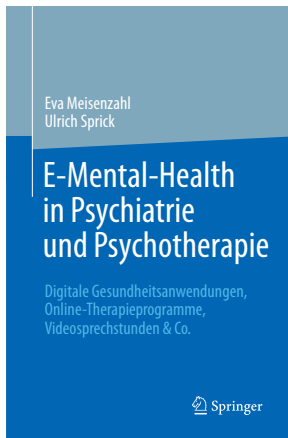
Wikimedia Commons



www.botreetechnologies.com



Pfannstiel (2022)



Meisenzahl and Sprick (2023)

AKTUELLES

[Startseite](#) > [Aktuelles](#) > [DGPs-Kongress 2024: Wie Künstliche Intelligenz die Arbeitswelt, das Bildungswesen und die Psychotherapie verändert](#)

DGPs-Kongress 2024: Wie Künstliche Intelligenz die Arbeitswelt, das Bildungswesen und die Psychotherapie verändert

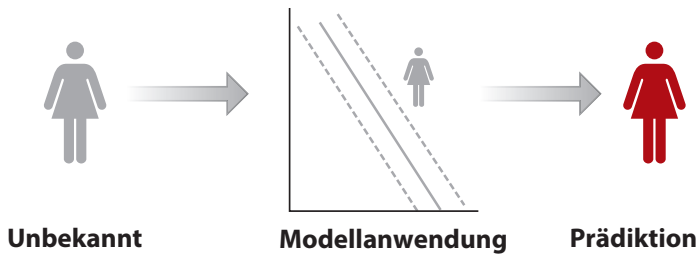
11.09.2024 | Vorstand Pressemitteilung

Wie kann Künstliche Intelligenz (KI) gewinnbringend in der Schule eingesetzt werden? Welche Chancen und Risiken gibt es beim Einsatz von KI in der Arbeitswelt? Wie wird KI psychotherapeutische Diagnostik und Interventionen verändern? Antworten auf diese Fragen und viele mehr werden auf dem 53. Kongress der Deutschen Gesellschaft für Psychologie präsentiert. Der Kongress startet am Montag, dem 16. September 2024 an der Universität Wien. Der Kongress findet im teilhybriden Format statt, für ausgewählte Veranstaltungen ist auch eine Online-Akkreditierung möglich.

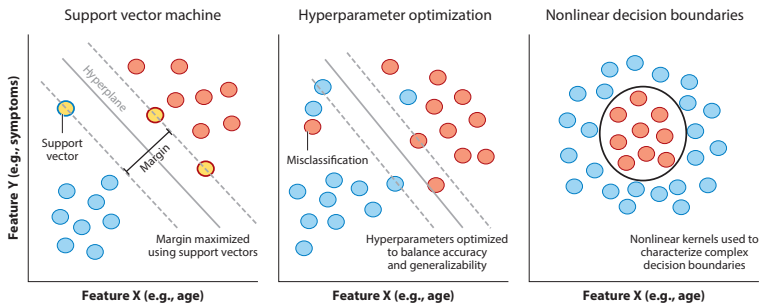
Ähnliche Artikel

53. DGPs-Kongress startet in 10 Tagen: Psychologie in Zeiten gesellschaftlicher Herausforderungen

DGPs-Kongress 2024 in Wien: Menschen | Mitwelt | Medien

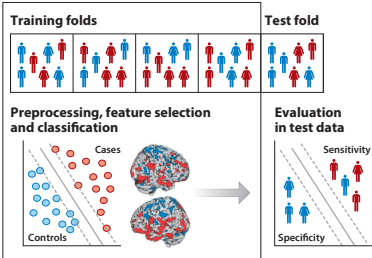


Dwyer, Falkai, and Koutsouleris (2018) *Annual Reviews of Clinical Psychology*

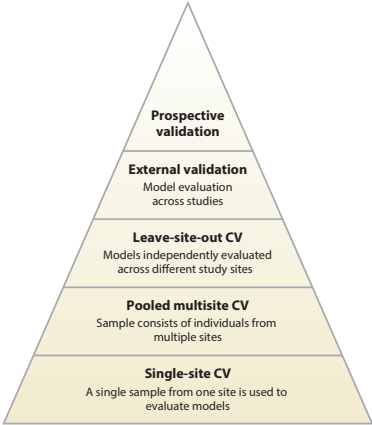


Dwyer, Falkai, and Koutsouleris (2018) *Annual Reviews of Clinical Psychology*

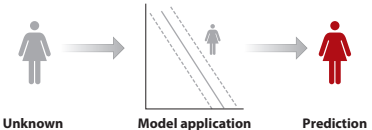
k-fold cross-validation



Generalizability hierarchy

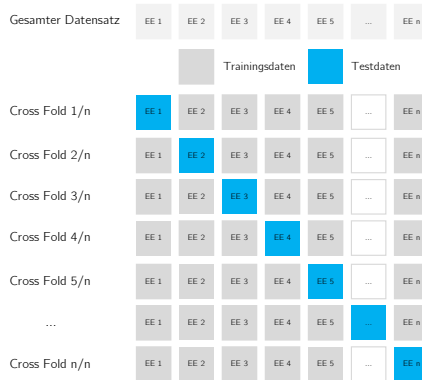


Prospective validation



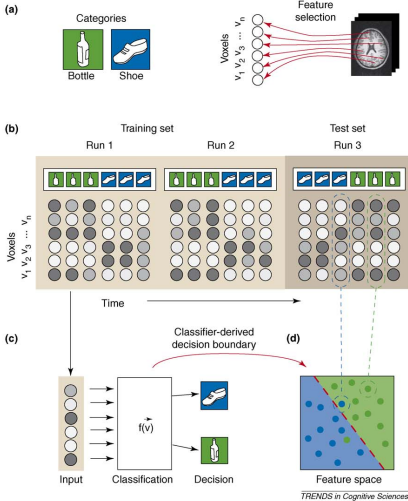
Dwyer, Falkai, and Koutsouleris (2018) *Annual Reviews of Clinical Psychology*

Leave-One-Out-Crossvalidation (LOO-CV)



EE = Experimentelle Einheit (z.B. Proband:in, Studienter, ...)

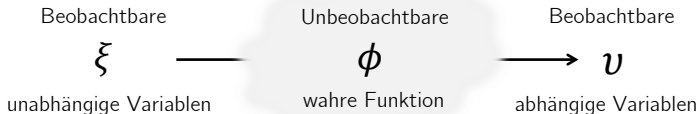
Vgl. Stone (1974)



Norman et al. (2006) *Trends in Cognitive Science*

Explanatorische Modellierung \Leftrightarrow Grundlagenforschung

Bestimmung von $\hat{\phi} := \operatorname{argmin} \|\hat{\phi} - \phi\|$



Bestimmung von $\hat{f} := \operatorname{argmin}_{f \in F} \|\nu - f(\xi)\|$, F beliebig

Prädiktive Modellierung \Leftrightarrow Anwendungsorientierte Forschung

Shmueli (2010), Sainani (2014)

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen



Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach

Nikolaos Koutsouleris, René S Kahn, Adam M Chekroud, Stefan Leucht, Peter Falkai, Thomas Wobrock, Eske M Derks, Wolfgang W Fleischhacker, Alkomiet Hasan

Summary

Background At present, no tools exist to estimate objectively the risk of poor treatment outcomes in patients with first-episode psychosis. Such tools could improve treatment by informing clinical decision-making before the commencement of treatment. We tested whether such a tool could be successfully built and validated using routinely available, patient-reportable information.

Methods By applying machine learning to data from 334 patients in the European First Episode Schizophrenia Trial (EUFEST; International Clinical Trials Registry Platform number, ISRCTN68736636), we developed a tool to predict poor versus good treatment outcome (Global Assessment of Functioning [GAF] score ≥ 65 vs GAF < 65 , respectively) after 4 weeks and 52 weeks of treatment. To enable the unbiased estimation of the predictive system's generalisability to new patients, we used repeated nested cross-validation to prevent information leaking between patients used for training and validating the models. In pursuit of everyday clinical applicability, we retrained the 4-week outcome predictor with only the top ten predictors of the pooled prediction system and then tested this tool in 108 independent patients with 4-week outcome labels. Discontinuation and readmission to hospital events in patients with predicted poor versus good outcomes were assessed with Kaplan-Meier log-rank analyses, whereas generalised linear mixed-effects models were used to investigate the GAF-based predictions against several clinically meaningful outcome indicators, including treatment adherence, symptom remission, and quality of life.

Koutsouleris et al. (2016)



Available online at www.sciencedirect.com



Schizophrenia Research 78 (2005) 147–156

SCHIZOPHRENIA
RESEARCH

www.elsevier.com/locate/schres

The European First Episode Schizophrenia Trial (EUFEST): Rationale and design of the trial

W. Wolfgang Fleischhacker ^{a,*}, Ireneus P.M. Keet ^b, René S. Kahn ^b

EUFEST Steering Committee ¹

^aDepartment of Biological Psychiatry, Medical University Innsbruck, Anichstrasse 35, A-6020 Innsbruck, Austria

^bDepartment of Psychiatry, Radolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands

Received 12 May 2005; received in revised form 2 June 2005; accepted 3 June 2005

Available online 1 August 2005

Abstract

Background: Most studies comparing second generation antipsychotics with classical neuroleptics have been conducted in more or less chronic schizophrenia patients. Such studies were usually conducted in highly selected samples, and were generally designed and financed by the manufacturer of the drug tested. These and other facts have stimulated discussions regarding the effectiveness of the new generation of antipsychotics.

Aims: The aim of the European First Episode Schizophrenia Trial (EUFEST) is to compare treatment with amisulpride, quetiapine, olanzapine and ziprasidone to a low dose of haloperidol in an unselected sample of first episode schizophrenia patients with minimal prior exposure to antipsychotics.

Fleischhacker, Keet, and Kahn (2005)

Predictor variables and outcome target

Definition of outcome targets, predictor variables, and analysis cohorts

Global Assessment of Functioning (GAF) scores between 61 and 70 have been proposed as thresholds between disease states and at-risk syndromes,¹⁸ indicators of clinically relevant functional impairment,¹⁹ and markers of recovery as part of more complex criteria.²⁰ Based on these publications, the patients' skewed 1-year follow-up GAF scores, and the resulting low sensitivity of a dimensional prediction model (appendix), we chose a classification approach to differentiate between so-called good (GAF ≥ 65) and poor (GAF < 65) outcome endpoints. An alternative GAF threshold of 60 was assessed by the machine learning methods described below (appendix).

By following a data-driven, multivariate analysis strategy, we deliberately did not pre-select a small set of statistically independent predictor variables. Instead, we set up a rigorously cross-validated learning strategy that autonomously identified the most accurate and parsimonious predictor patterns (figure 1). We initially

Koutsouleris et al. (2016)

Anwendungsbeispiel

Predictor variables and outcome target

Sociodemographic variables			
1 Sex	52 Obsessive-compulsive disorder	100 CAN_204 Minirey user Need	190 Num_igbtg_Raw-Wrong
2 Age	53 Alcohol dependence Past 6 months	101 CAN_210 Minirey staff Need	180 Num_pmpg_Placed-Correct-Don't-need
3 Education	52 Alcohol dependence Lifetime	102 CAN_214 Benefits user Need	181 Num_pmpg_Placed-Correct-Not-Don't-need
4 Weight	53 Alcohol above Past 12 months	103 CAN_220 Benefits staff Need	180 Num_pmpg_Placed-Correct-Both-Hands
5 BMI	54 Alcohol above Lifetime	109 CAN Summary score (No & Low Need)	183 RAWL_T0_Num_Correct_Responses
6 ECG abnormal	55 Substance dependence Past 12 months	110 CAN Summary score (Moderate Needs)	184 RAWL_T0_Num_Requests
7 Diastolic blood pressure	56 Substance dependence Lifetime	111 CAN Summary score (Urgent Need)	185 RAWL_T0_Num_Requests
8 Systolic blood pressure	57 Substance above Past 12 months	112 CGI	186 RAWL_T02_Num_Correct_Responses
9 Pulse	58 Substance abuse Lifetime	113 GAF	187 RAWL_T02_Num_Mistrusts
10 AD23 Marital status patient married	59 Psychotic disorder Lifetime	114 CDB6	188 RAWL_T02_Num_Requests
11 AD30 Current occupation patient	60 Mood disorder with psychotic features	Positive and Negative Symptoms Scale	189 RAWL_T02_Num_Correct_Responses
12 AD32 Previous occupation patient	61 Schizophrenia Current	115 PANSS P1 Delusions	170 RAWL_T03_Num_Mistrusts
13 AD34 Highest level occupation father	62 Schizophrenia Lifetime	116 PANSS P2 Conceptual disorganization	171 RAWL_T03_Num_Requests
14 AD34 Highest level occupation mother	63 Schizophrenia disorder Current	117 PANSS P3 Hallucinations	172 RAWL_T04_Num_Correct_Responses
15 AD37 Highest degree patient	64 Schizophrenia disorder Lifetime	118 PANSS P4 Inappropriately	173 RAWL_T04_Num_Mistrusts
16 AD38 Educational problems	65 Schizophreniform disorder Current	119 PANSS P5 Grandiosity	174 RAWL_T04_Num_Correct_Responses
17 AD39 Education father	66 Schizophreniform disorder Lifetime	120 PANSS P6 Suspiciousness/persecution	175 RAWL_T05_Num_Mistrusts
18 AD40 Education mother	67 Substance induced psychotic disorder Lifetime	121 PANSS P7 Irritability	176 RAWL_T05_Num_Requests
19 AD41 Living alone	Caribemef Assessment of Needs	122 PANSS N1 Somatic affect	177 RAWL_T05_Num_Requests
20 AD50 Types of dwelling	68 CAN_061 Accommodation user Need	123 PANSS N2 Emotional withdrawal	178 RAWL_T05-Num_Correct_Responses
21 AD63 Living environment	69 CAN_067 Accommodation staff Need	124 PANSS N3 Poor rapport	180 RAWL_T05-Num_Requests
Medication areas	70 CAN_071 Food user Need	125 PANSS N4 Frustration/irritability social withdrawal	181 RAWL_DR
22 Haloperidol treatment	71 CAN_077 Food staff Need	126 PANSS N5 Difficulty in alcohol drinking	182 RAWL_DOSG-7
23 Clozapine treatment	72 CAN_082 Self-care user Need	127 PANSS N6 Lack of spontaneity and flow of conversation	183 RAWL_DR_Num_Mistrusts
24 Quetiapine treatment	73 CAN_088 Self-care staff Need	128 PANSS N7 Disorganized thinking	184 RAWL_DR_Num_Requests
25 Aripiprazole treatment	74 CAN_040 Day time activities user Need	129 PANSS G1 Somatic concern	185 ADAMSA score
26 Ziprasidone treatment	75 CAN_052 Physical health user Need	130 PANSS G2 Anxiety	186 Systems score
27 Ziprasidone treatment	76 CAN_058 Physical health staff Need	131 PANSS G3 Guilt feelings	187 Parkinsonism score
MMI Plus interview	77 CAN_060 Psychotic symptoms user Need	132 PANSS G4 Tension	188 Active-Cyberlapse score
27 Former substance abuse	78 CAN_068 Psychotic symptoms staff Need	133 PANSS G6 Manicness and posturing	189 Passive-Cyberlapse score
28 Persistent substance abuse	79 CAN_066 Psychotic symptoms staff Informal help given	134 PANSS G8 Depression	
29 Disorganized schizophrenia	80 CAN_070 Psychotic symptoms staff Formal help given	135 PANSS G7 Motor retardation	
30 Catatonic schizophrenia	81 CAN_076 Psychotic symptoms staff Formal help given	136 PANSS G8 Unusual perceptions	
31 Paranoid schizophrenia	82 CAN_071 Psychotic symptoms staff Formal help needed	137 PANSS G9 Unusual thought content	
32 Schizophreniform disorder	83 CAN_078 Information user Need	138 PANSS G10 Disorientation	
33 Residual state	84 CAN_078 Information staff Need	139 PANSS G11 Poor attention	
34 Schizophrenia disorder	85 CAN_082 Psychological distress user Need	140 PANSS G12 Lack of judgment and insight	
35 Unaffected/schizophrenia	86 CAN_080 Psychological distress staff Need	141 PANSS G13 Disorientation of volition	
36 Major depressive episode Current	87 CAN_082 Safety to self user Need	142 PANSS G14 Poor impulse control	
37 Major depressive episode Recurrent	88 CAN_088 Safety to self staff Need	143 PANSS G15 Preoccupation	
38 Substance induced mood disorder Past	89 CAN_102 Safety to others user Need	144 PANSS G16 Active social avoidance	
39 Major depressive episode with melancholic features Current	90 CAN_138 Safety to others staff Need	145 PANSS Positive score	
40 Dysthymia Current	91 CAN_138 Accura staff Need	146 PANSS Negative score	
41 Dysthymia Past	92 CAN_122 Drugs user Need	147 PANSS General score	
42 Succulity Current	93 CAN_139 Drugs staff Need	148 PANSS Total score	
43 Hypomanic episode Past	94 CAN_130 Company user Need	Neuropsychological test battery	
44 Substance induced manic episode Past	95 CAN_138 Company staff Need	149 TMT-A	
45 Panic disorder Current past month	96 CAN_142 Intimate relationships user Need	150 TMT-B Errors	
46 Panic disorder Lifetime	97 CAN_140 Intimate relationships staff Need	151 TMT-A Pop	
47 Agoraphobia Lifetime	98 CAN_133 Sexual expression user Need	152 TMT-B	
48 Social phobia	99 CAN_130 Sexual expression staff Need	153 TMT-B Errors	
49 Specific phobia	100 CAN_174 Education user Need	154 TMT-B Pop	
	101 CAN_180 Education staff Need	155 TMT time (B-A)	
	102 CAN_184 Telephone user Need	156 TMT error (B-A)	
	103 CAN_194 Transport user Need	157 Num_igbtg_Raw-Correct	
	104 CAN_200 Transport staff Need	158 Num_igbtg_Raw-Skipped	

Koutsouleris et al. (2016)

Predictor variables and outcome target

Global Assessment of Functioning

Das **Global Assessment of Functioning** (kurz **GAF**) ist eine **Bewertungsskala** aus der klinischen Psychologie und Psychiatrie. Sie wird verwendet, um das allgemeine Funktionsniveau einer Person zu erfassen. Körperliche Erkrankungen und Umgebungsbedingungen sollen hierbei explizit nicht berücksichtigt werden. Die psychischen, sozialen und beruflichen Funktionen werden dabei auf einem hypothetischen Kontinuum von psychischer Gesundheit bis Krankheit gedacht.^[1]

Die Skala findet Anwendung innerhalb der vierten Ausgabe des *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* und innerhalb der *Operationalisierten Psychodynamischen Diagnostik (OPD)*. Der Zeitraum der Beurteilung im OPD 2 beschränkt sich auf die letzten 7 Tage. Die GAF-Skala wurde ursprünglich 1989 von der *American Psychiatric Association* entwickelt.^{[1][2]}

Aufbau | Bearbeiten | Qualität bearbeiten |

Die GAF-Skala ist in 10 Funktionsniveaus mit je 10 Punkten unterteilt. Sie reicht von 100 (höchstes Leistungsniveau) bis zu 1 (niedrigstes Leistungsniveau).^[1]

Der Beurteiler muss nun einen einzelnen Wert angeben, der das allgemeine Funktionsniveau des Patienten am treffendsten wiedergibt. Dabei sollen sowohl die Symptomschwere als auch das Ausmaß der Beeinträchtigung berücksichtigt werden. Wenn Symptomausprägung und Beeinträchtigungsgrad sich in verschiedenen 10er-Kategorien befinden, wird das schlechtere Niveau gewählt. Beispiel: Bei einer Person, die ernsthaft eigengefährdet ist, im Übrigen jedoch nicht beeinträchtigt ist, liegt der GAF-Wert unter 20.^[2]

Funktion in %	Beschreibung ^[1]
100–91	Hervorragende Leistungsfähigkeit in einem breiten Spektrum von Aktivitäten. Schwierigkeiten im Leben scheinen nie außer Kontrolle zu geraten, keine Symptome.
90–81	Keine oder nur minimale Symptome (z. B. leichte Angst vor einer Prüfung), gute Leistungsfähigkeit in allen Gebieten, interessiert und eingebunden in ein breites Spektrum von Aktivitäten, sozial effektiv im Verhalten, im Allgemeinen zufrieden mit dem Leben, übliche Alltagsprobleme oder -sorgen (z. B. nur gelegentlicher Streit mit einem Familienmitglied).
80–71	Wenig Symptome vorliegen, sind dies vorübergehende oder zu erwartende Reaktionen auf psychosoziale Belastungsfaktoren (z. B. Konzentrationserschwerungen nach einem Familienstreit), höchstens leichte Beeinträchtigung der sozialen beruflichen und schulischen Leistungsfähigkeit (z. B. zeitweises Zurückbleiben in der Schule).
70–61	Einige leichte Symptome (z. B. depressive Stimmung oder leichte Schlaflosigkeit), ODER einige leichte Schwierigkeiten hinsichtlich der sozialen, beruflichen oder schulischen Leistungsfähigkeit (z. B. gelegentliches Schneeschaufen oder Diebstahl im Haushalt), aber im Allgemeinen relativ gute Leistungsfähigkeit; Bestehen einiger wichtiger zwischenmenschlicher Beziehungen.
60–51	Mäßig ausgeprägte Symptome (z. B. Aktivitätsverfälschung, weichesprachige Sprache, gelegentliche Panikattacken) ODER mäßig ausgeprägte Schwierigkeiten bezüglich der sozialen, beruflichen oder schulischen Leistungsfähigkeit (z. B. wenige Freunde, Konflikte mit Arbeitskollegen, Schulkameraden oder Bezugspersonen).
50–41	Ermittelte Symptome (z. B. Suizidgedanken, schwere Zwangsstörungen, häufige Lasterdebatte) ODER eine Beeinträchtigung der sozialen, beruflichen und schulischen Leistungsfähigkeit (z. B. keine Freunde, Unpünktigkeit, eine Arbeitsstelle zu behalten).
40–31	Einige Beeinträchtigungen in der Realitätskontrolle oder der Kommunikation (z. B. Sprache zeitweise unlogisch, unverständlich oder belanglos) ODER starke Beeinträchtigung in mehreren Bereichen, z. B. Arbeit oder Schule, familiäre Beziehungen, Unhöflichkeit, Denken oder Stimmung (z. B. ein Mann mit einer Depression vermisst Freunde, vernachlässigt seine Familie und ist unfähig zu arbeiten, ein Kind schlägt häufig jüngere Kinder, ist zu Hause hitzig und versagt in der Schule).
30–21	Das Verhalten ist ernsthaft durch Wahrnehmungen oder Halluzinationen beeinflusst ODER ernsthafte Beeinträchtigung der Kommunikation und des Urteilsvermögens (z. B. manchmal inkohärent, handelt grob inadäquat, starkes Eigenmessen von Selbstmordgedanken) ODER Leistungsunfähigkeit in fast allen Bereichen (z. B. bleibt den ganzen Tag im Bett, hat keine Arbeit, kein Zuhause und keine Freunde).
20–11	Selbst- und Fremdgefährdung (z. B. Selbstmordversuche ohne eindeutige Todesabsicht, häufig gewalttätig, marasche Erregung) ODER ist gelegentlich nicht in der Lage, die geringste Hygiene aufrechtzuerhalten (z. B. schmiert mit Kot) ODER grobe Beeinträchtigung der Kommunikation (größtenteils inkohärent oder stumm).
10–1	Ständige Gefahr, sich oder andere schwer zu verletzen (z. B. wiederholte Gewaltanwendung) ODER anhaltende Unfähigkeit, die minimale persönliche Hygiene aufrechtzuerhalten ODER ernsthafter Selbstmordversuch mit eindeutiger Todesabsicht.
0	Unzureichende Information.

Anwendungsbeispiel

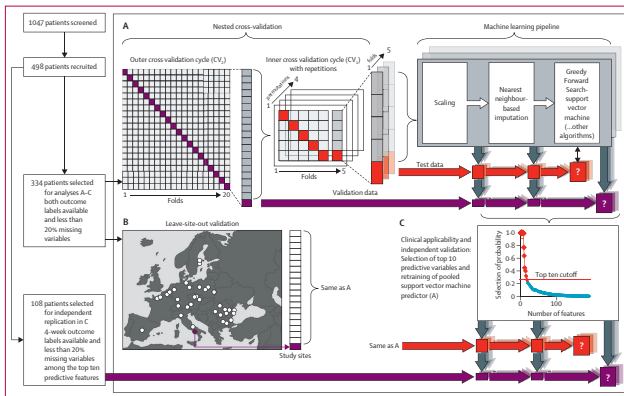


Figure 1: Machine learning analysis design

(A) We used nested, repeated cross-validation to train and validate the 4-week and 52-week outcome predictors, consisting of an outer 20-fold cross-validation cycle (CV_o), which provided validation participants for computing an unbiased estimate of predictor generalisability to new patients, and an inner 4×5-fold cross-validation cycle (CV_i), which delivered training participants to the multivariate pattern analysis pipeline as well as test participants for feature and parameter optimisation. (B) Generalisability of predictors to new sites was assessed by iteratively leaving out once each of 44 the EUFEST study sites included in the analysis, training the predictive system as in A and applying the trained predictor to the left-out site. (C) Using the top ten predictive features of the pooled 4-week outcome classifier, the prediction system was retrained and then applied to 108 patients, who were excluded from the discovery and cross-validation analyses.

Koutsouleris et al. (2016)

Findings The generalisability of our outcome predictions were estimated with cross-validation (test-fold balanced accuracy [BAC] of 75.0% for 4-week outcomes and 73.8% for and 52-week outcomes), and leave-site-out validation across 44 European sites (BAC of 72.1% for 4-week outcomes and 71.1% for 52-week outcomes). We identified a smaller group of ten predictors still providing a BAC of 71.7% in 108 patients never used for model discovery. Unemployment, poor education, functional deficits, and unmet psychosocial needs predicted both endpoints, whereas previous depressive episodes, male sex, and suicidality additionally predicted poor 1-year outcomes. 52-week predictions identified patients at risk for symptom persistence, non-adherence to treatment, readmission to hospital and poor quality of life. Specifically among these patients, amisulpride and olanzapine showed superior efficacy versus haloperidol, quetiapine, and ziprasidone.

Interpretation Our results suggest that prognostic models operating on brief, patient-reportable pre-treatment data might provide useful insight into individualised outcome trajectories, optimising treatment selection, and targeted clinical trial designs. To embed these tools into real-world care, replication is needed in external first-episode samples with overlapping variables, which are not available in the field at present.

Koutsouleris et al. (2016)

Anwendungsbeispiel

	True negatives	True positives	False positives	False negatives	Sensitivity	Specificity	Balanced accuracy	Positive predictive value	Negative predictive value	Prognostic summary index	Positive likelihood ratio	Diagnostic odds ratio	Number-needed-to-diagnose	κ^*
4-week outcome predictor: pooled cross-validation analyses														
RBF-SVM	165	84	26	59	73.7%	76.4%	75.0%	86.4%	58.7%	44.1%	3.1	9.7	2.00	-
Linear SVM	170	75	35	54	75.9%	68.2%	72.0%	82.9%	58.1%	41.1%	2.4	5.7	2.27	0.724
Univariate LOGREG	197	55	55	27	88.0%	50.0%	69.0%	78.2%	67.1%	45.2%	1.8	3.1	2.64	0.541
Multivariate LOGREG	197	59	51	27	88.0%	53.6%	70.8%	79.4%	68.6%	48.0%	1.9	3.6	2.40	0.589
Decision Trees	197	54	56	27	88.0%	49.1%	68.5%	77.9%	66.7%	44.5%	1.7	3.0	2.70	0.302
4-week outcome predictor: leave-site-out analyses														
RBF-SVM (10% of top-ranked features)	147	81	29	77	65.6%	73.6%	69.6%	83.5%	51.3%	34.8%	2.5	6.2	2.55	0.814
RBF-SVM (20%)	152	80	30	72	67.9%	72.7%	70.3%	84.4%	54.7%	39.1%	2.5	6.2	2.46	0.837
RBF-SVM (30%)	151	82	28	73	67.4%	74.5%	71.0%	84.0%	54.8%	38.9%	2.6	7.0	2.38	0.807
RBF-SVM (40%)	158	80	30	66	70.5%	72.7%	71.6%	84.4%	52.9%	37.3%	2.6	6.7	2.31	0.837
RBF-SVM (50%)	157	81	29	67	70.1%	73.6%	71.9%	83.5%	52.6%	36.1%	2.7	7.1	2.29	0.873
RBF-SVM (All selected variables)	156	82	28	68	69.6%	74.5%	72.1%	84.8%	54.7%	39.4%	2.7	7.5	2.26	0.909
4-week outcome predictor: condensed battery (top ten features)														
RBF-SVM in 108 new patients	64	13	5	26	71.1%	72.2%	71.7%	92.8%	33.3%	26.1%	2.6	6.6	2.31	-
52-week outcome predictor: pooled cross-validation analyses														
RBF-SVM	52	207	49	26	66.7%	80.9%	73.8%	51.5%	88.8%	40.3%	3.5	12.1	2.10	-
Linear SVM	49	189	67	29	62.8%	73.8%	68.3%	42.2%	87.7%	28.9%	2.4	5.8	2.73	0.734
Univariate LOGREG	13	241	15	65	16.7%	94.1%	55.4%	46.4%	78.8%	25.2%	2.8	8.1	9.25	0.282
Multivariate LOGREG	18	245	11	60	23.1%	95.7%	59.4%	62.1%	80.3%	42.4%	5.4	28.8	5.32	0.331
Decision trees	27	234	21	51	34.6%	91.8%	63.2%	56.3%	82.1%	38.4%	4.2	17.7	3.79	0.429
52-week outcome predictors: leave-site-out analyses														
RBF-SVM (10% of top-ranked features)	49	186	70	29	62.8%	72.7%	67.7%	41.2%	86.5%	27.7%	2.3	5.3	2.82	0.649
RBF-SVM (20%)	52	193	63	26	66.7%	75.4%	71.0%	46.0%	88.2%	34.3%	2.7	7.3	2.38	0.768
RBF-SVM (30%)	51	189	67	27	65.4%	73.8%	69.6%	45.3%	88.5%	33.8%	2.5	6.2	2.55	0.777
RBF-SVM (40%)	53	192	64	25	67.9%	75.0%	71.5%	43.2%	87.5%	30.7%	2.7	7.4	2.33	0.810
RBF-SVM (50%)	52	195	61	26	66.7%	76.2%	71.4%	45.2%	88.1%	33.3%	2.8	7.8	2.33	0.808
RBF-SVM (All selected variables)	50	200	56	28	64.1%	78.1%	71.1%	47.2%	87.7%	34.9%	2.9	8.6	2.37	0.797

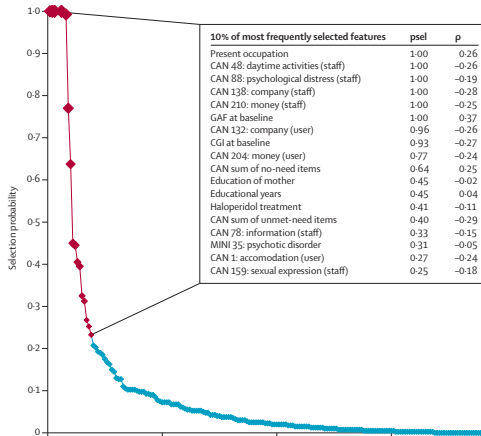
Based on the established sensitivity and specificity nomenclature, we label predicted poor outcomes as positive predictions and predicted good outcomes as negative predictions, hence sensitivity measures the classifiers' capacity to correctly identify patients with poor outcomes as such. Methodological validation of SVM was done against univariate and L2-regularised multivariate logistic regression models as well as decision tree ensembles in both 4-week and 52-week pooled prediction analyses. Prediction performance was further validated in leave-site-out analyses without clinical variable confinement (all selected variables) as well as with different degrees of top-variable selection (10-50%). An additional validation step was done for the pooled 4-week SVM predictor, which was trained on the top ten most predictive features (figure 2) and then applied to a cohort of 108 EUFEST patients never included in any analysis procedure.

SVM= support Vector Machine. LOGREG= logistic regression. RBF= non-linear radial basis function kernel. SVM= support vector machine. * κ =Cohen's kappa measuring predictor agreement between the Pooled SVM and the respective predictor.

Table 2: Validated predictive performance of classifiers trained on the 4-week and 52-week outcome endpoints

Koutsouleris et al. (2016)

Anwendungsbeispiel



Koutsouleris et al. (2016)

Research in context

Evidence before this study

We systematically searched PubMed for all papers published in English up to April 7, 2016, using the following search terms “(prediction) AND (psychosis OR schizophrenia) AND (machine learning OR multivariate pattern analysis OR support vector machine)”. We identified 31 unique records, none of which presented a thoroughly validated approach to predict the treatment outcomes of individual patients with first-episode psychosis using broadly accessible clinical baseline data. Although many group-level sociodemographic, clinical, and neurocognitive variables are known to be moderately associated with therapeutic outcomes, it is unclear how and which of these variables are effectively combined into prognostic models so that accurate and generalisable outcome predictions can be generated on a case-by-case level. Furthermore, how these predictions could inform therapeutic decision making so that the right patients receive the most appropriate treatments is unknown.

Added value of this study

This the first study to show accurate and generalisable individual-patient outcome predictions after 4 and 52 weeks of treatment for first-episode psychosis. Our machine learning models used only clinical baseline information, particularly

psychosocial, sociodemographic and psychometric variables, and as few as ten questions. Notably, the resulting predictions tracked an array of clinically meaningful outcome indicators, such as treatment adherence, readmission to hospital, functioning, symptom burden, and quality of life, thus approximating the more complex phenotype of recovery from first-episode psychosis. The present approach might provide a more than 40% gain in correctly predicting outcomes across geographically distinct health-care settings and patient populations. Importantly, our models revealed that the efficacy differences previously found among the five European First Episode Schizophrenia Trial antipsychotic treatments, and those reported in recent meta-analytic comparisons, might be driven by the group of patients with poor prognoses, as predicted by our analytical tools.

Implications of all the available evidence

Our study exemplifies methodological advances crucial to the development of reliable outcome prediction models in the psychosis field. In this context, the evidence generated by our study strongly suggests that measuring individual risk by prognostic modelling could inform everyday clinical care as well as new clinical trials designs in the era of digital mental health.

Koutsouleris et al. (2016)

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

Binärer Klassifikationstrainingdatensatz

Ein *binärer Klassifikationsdatensatz*

$$\mathcal{D} := \{(x_1, y_1), \dots, (x_n, y_n)\} = \{(x_i, y_i)\}_{i=1}^n \quad (1)$$

ist eine Menge von n *Trainingsdatenpunkten*

$$(x_i, y_i) \text{ mit } x_i \in \mathbb{R}^m \text{ und } y_i \in \{0, 1\} \text{ for } i = 1, \dots, n, \quad (2)$$

wobei x_i *m-dimensionaler Featurevektor* und y_i *Label* genannt wird. Üblicherweise werden die Trainingsdatenpunkte dabei als unabhängige und identische Realisierungen eines $m + 1$ -dimensionalen Zufallsvektors $\zeta := (\xi, v)$ verstanden.

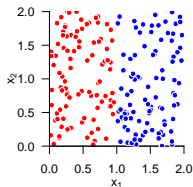
Bemerkungen

- $y_i \in \{0, 1\}$ bezeichnet die Klassenzugehörigkeit des Featurevektors $x_i \in \mathbb{R}^m$.
- Ein Beispiel für y_i ist "GAF < 65" (0) vs. "GAF ≥ 65" (1).
- Beispiele für die m Komponenten der x_i sind Testscores, soziodemographische Daten, ...

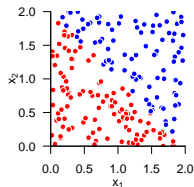
Beispiele bivariater Featureplotszenarien

$$x_i \in \mathbb{R}^2, y_i \in \{0, 1\}, i = 1, \dots, n, \bullet y_i = 0, \bullet y_i = 1$$

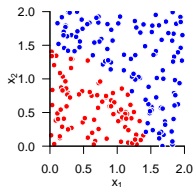
$$x_{i_1} > 1 \Leftrightarrow y_i = 1$$



$$x_{i_1} + x_{i_2} > 2 \Leftrightarrow y_i = 1$$



$$x_{i_1}^2 + x_{i_2}^2 > 2 \Leftrightarrow y_i = 1$$



Sprache der prädiktiven Modellierung

Daten	Trainingsdaten und Testdaten
Statistisches Modell	Modell, Algorithmus, "Die KI"
Schätzen von Parametern	Trainieren des Modells, Lernen, Supervised Learning

Workflow der prädiktiven Modellierung

Featureselektion

- Auswahl von möglichen prädiktiven Variablen
- Dimensionreduktion zur Verringerung des Curse of Dimensionality

Kreuzvalidierung

- Wiederholtes Trainieren und Testen eines Modells an einem Datensatz
- Einsatz zur Modelloptimierung
- Einsatz zur Messung der probabilistischen Assoziation von Features und Labels

Konfusionsmatrix bei LOO-CV mit binärem Label für (x_i, y_i)

		Prädiktion		
		$f(x_i) = 0$	$f(x_i) = 1$	
Fall	$y_i = 0$	True Negative TN	False Positive FP	Negative $N := TN + FP$
	$y_i = 1$	False Negative FN	True Positive TP	Positive $P := FN + TP$
		Predicted Negative $PN := TN + FN$	Predicted Positive $PP := TP + FP$	Total $T := N + P$

True Positive = Hit, False Positive = False alarm, True Negative = Correct rejection, False Negative = Miss

Prädiktionsgüte

$$\text{Accuracy (ACC)} = \frac{\text{Anzahl richtiger Prädiktionen}}{\text{Anzahl aller Prädiktionen}} = \frac{TN + TP}{T} \quad (3)$$

Prädiktionsraten mit Bezug zu tatsächlichen Fällen

True Positive Rate (aka Hit Rate, Sensitivity, Detection probability, Recall, Power)

$$\text{TPR} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{TP}}{\text{P}} \quad (4)$$

False Positive Rate (aka Type I Error Rate, False Alarm Probability, Fall-Out)

$$\text{FPR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{FP}}{\text{N}} \quad (5)$$

True Negative Rate (aka Specificity, Selectivity)

$$\text{TNR} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{TN}}{\text{N}} \quad (6)$$

False Negative Rate (aka Type II Error Rate, Miss Rate)

$$\text{FNR} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{FN}}{\text{P}} \quad (7)$$

Performanzmetriken

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

		Prädiktion		
		$f(x_i) = 0$	$f(x_i) = 1$	
Fall	$y_i = 0$	True Negative TN = 165	False Positive FP = 26	Negative N = 191
	$y_i = 1$	False Negative FN = 59	True Positive TP = 84	Positive P = 143
		Predicted Negative PN = 224	Predicted Positive PP = 110	Total T = 334

Prädiktionsgüte

$$\text{Accuracy (ACC)} = \frac{\text{Anzahl richtiger Prädiktionen}}{\text{Anzahl aller Prädiktionen}} = \frac{\text{TN} + \text{TP}}{\text{T}} = \frac{165 + 84}{334} = 74.6\% \quad (8)$$

Performanzmetriken

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

True Positive Rate (aka Hit Rate, **Sensitivity**, Detection probability, Recall, Power)

$$\text{TPR} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{TP}}{\text{P}} = \frac{84}{143} = 58.7\% \quad (9)$$

False Positive Rate (aka Type I Error Rate, False Alarm Probability, Fall-Out)

$$\text{FPR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{FP}}{\text{N}} = \frac{26}{191} = 13.6\% \quad (10)$$

True Negative Rate (aka **Specificity**, Selectivity)

$$\text{TNR} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Fälle}} = \frac{\text{TN}}{\text{N}} = \frac{165}{191} = 86.4\% \quad (11)$$

False Negative Rate (aka Type II Error Rate, Miss Rate)

$$\text{FNR} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl positiver Fälle}} = \frac{\text{FN}}{\text{P}} = \frac{59}{143} = 41.3\% \quad (12)$$

Prädiktionsraten mit Bezug zu Prädiktionsanzahlen

Positive Predictive Value / Precision

$$\text{PPV} = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{TP}}{\text{PP}} \quad (13)$$

False Discovery Rate

$$\text{FDR} = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{\text{FP}}{\text{PP}} \quad (14)$$

Negative Predictive Value

$$\text{NPV} = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{TN}}{\text{PN}} \quad (15)$$

False Omission Rate

$$\text{NPV} = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{\text{FN}}{\text{PN}} \quad (16)$$

Performanzmetriken

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

Positive Predictive Value / Precision

$$PPV = \frac{\text{Anzahl richtiger Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{TP}{PP} = \frac{84}{110} = 76.4\% \quad (17)$$

False Discovery Rate

$$FDR = \frac{\text{Anzahl falscher Positivprädiktionen}}{\text{Anzahl positiver Prädiktionen}} = \frac{FP}{PP} = \frac{26}{110} = 23.6\% \quad (18)$$

Negative Predictive Value

$$NPV = \frac{\text{Anzahl richtiger Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{TN}{PN} = \frac{165}{224} = 73.7\% \quad (19)$$

False Omission Rate

$$NPV = \frac{\text{Anzahl falscher Negativprädiktionen}}{\text{Anzahl negativer Prädiktionen}} = \frac{FN}{PN} = \frac{59}{224} = 26.3\% \quad (20)$$

Balanced Accuracy

Bei ungleichen Verhältnissen von Fallzahlen ist Accuracy (ACC) leicht missverständlich

Seien bspw. $N = 95$ und $P = 5$ und alle Prädiktionen negativ. Dann gilt

$$\text{ACC} = \frac{\text{TN} + \text{TP}}{T} = \frac{95 + 0}{100} = 95\% \quad (21)$$

Die Balanced Accuracy ist der Mittelwert der True Positive Rate und der True Negative Rate

$$\text{Balanced Accuracy (BAC)} = \frac{\text{TPR} + \text{TNR}}{2} \quad (22)$$

Seien $N = 95$ und $P = 5$ und alle Prädiktionen negativ. Dann gilt

$$\text{TPR} = \frac{\text{TP}}{P} = \frac{0}{5} = 0\%, \quad \text{TNR} = \frac{\text{TN}}{N} = \frac{95}{95} = 100\%, \quad \text{BAC} = \frac{\text{TPR} + \text{TNR}}{2} = \frac{0\% + 100\%}{2} = 50\%$$

Seien $N = 95$ und $P = 5$ und alle Negativprädiktionen und alle Positivprädiktionen richtig. Dann gilt

$$\text{TPR} = \frac{\text{TP}}{P} = \frac{5}{5} = 100\%, \quad \text{TNR} = \frac{\text{TN}}{N} = \frac{95}{95} = 100\%, \quad \text{BAC} = \frac{\text{TPR} + \text{TNR}}{2} = \frac{100\% + 100\%}{2} = 100\%$$

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

$$\text{BAC} = \frac{\text{TPR} + \text{TNR}}{2} = \frac{58.7\% + 86.4\%}{2} = 72.6\% \quad (23)$$

F1 Score

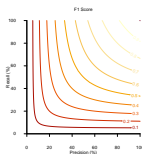
F1 Score ist ein integriertes Maß von Precision und Recall aus der Information Retrieval Literatur

$$\text{Precision} = \frac{TP}{PP} = \text{PPV} \quad \text{Recall} = \frac{TP}{P} = \text{TPR} \quad (24)$$

Es gilt

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (25)$$

Der F1 Score ist groß, wenn sowohl Precision als auch Recall groß sind.



Ist Precision oder Recall oder beide gering, so ist der F1 Score niedrig

Beispiel (Koutsouleris et al. (2016), Table 1, 4-week outcome prediction, RBF-SVM)

$$\text{F1} = 2 \cdot \frac{76.4\% \cdot 58.7\%}{76.4\% + 58.7\%} = 66.4\% \quad (26)$$

Rhetorik

Anwendungsbeispiel

Performanzmetriken

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Rhetorik der Prädiktiven Modellierung.
2. Erläutern Sie Unterschiede und Gemeinsamkeiten der explanatorischen und prädiktiven Modellierung.
3. Erläutern Sie die Idee der Leave-One-Out-Crossvalidation (LOO-CV).
4. Erläutern und definieren Sie die Konfusionsmatrix bei LOO-CV mit binärem Label.
5. Geben Sie die Definitionen von True/False Positive Rate und True/False Negative Rate wieder.
6. Geben Sie die Definitionen von Positive/Negative Prediction Value und False Discovery/Omission Rate wieder.
7. Erläutern und definieren Sie den Begriff der Balanced Accuracy.

Referenzen I

- Dwyer, Dominic B., Peter Falkai, and Nikolaos Koutsouleris. 2018. "Machine Learning Approaches for Clinical Psychology and Psychiatry." *Annual Review of Clinical Psychology* 14 (1): 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>.
- Fleischhacker, W. Wolfgang, Ireneus P. M. Keet, and René S. Kahn. 2005. "The European First Episode Schizophrenia Trial (EUFEST): Rationale and Design of the Trial." *Schizophrenia Research* 78 (2-3): 147–56. <https://doi.org/10.1016/j.schres.2005.06.004>.
- Koutsouleris, Nikolaos, René S Kahn, Adam M Chekroud, Stefan Leucht, Peter Falkai, Thomas Wobrock, Eske M Derks, Wolfgang W Fleischhacker, and Alkomiet Hasan. 2016. "Multisite Prediction of 4-Week and 52-Week Treatment Outcomes in Patients with First-Episode Psychosis: A Machine Learning Approach." *The Lancet Psychiatry* 3 (10): 935–46. [https://doi.org/10.1016/S2215-0366\(16\)30171-7](https://doi.org/10.1016/S2215-0366(16)30171-7).
- Meisenzahl, Eva, and Ulrich Sprick. 2023. *E-Mental-Health in Psychiatrie und Psychotherapie: Digitale Gesundheitsanwendungen, Online-Therapieprogramme, Videosprechstunden & Co.* Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-64457-7>.
- Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby. 2006. "Beyond Mind-Reading: Multi-Voxel Pattern Analysis of fMRI Data." *Trends in Cognitive Sciences* 10 (9): 424–30. <https://doi.org/10.1016/j.tics.2006.07.005>.
- Pfannstiel, Mario A., ed. 2022. *Künstliche Intelligenz im Gesundheitswesen: Entwicklungen, Beispiele und Perspektiven.* Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-33597-7>.
- Sainani, Kristin L. 2014. "Explanatory Versus Predictive Modeling." *PM&R* 6 (9): 841–44. <https://doi.org/10.1016/j.pmrj.2014.08.941>.
- Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3). <https://doi.org/10.1214/10-STS330>.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2): 111–47. <https://www.jstor.org/stable/2984809>.