

A network diagram consisting of numerous yellow circular nodes connected by thin yellow lines. The nodes are scattered across the frame, with some forming a dense central cluster and others more isolated. The lines vary in thickness and opacity, creating a complex web of connections.

Deep Dive Klassifikation

Webinar am 8. August 2023

Das Webinar ist Teil einer Serie

Heutiges Thema: Deep Dive Klassifikation



Neuigkeiten und Erweiterungen für DPS Invoice in der aktuellen Version
am **12. September 2023**

Digitaler Posteingang mit Klassifikation – DPS Mailroom
Oktober 2023

SAVE THE DATE

Treffen Sie uns vor Ort!



Digital Future Congress DFC in München
am **28. September**



Partnerfestival Versicherungsforen in Leipzig We.Xplore
am **28. und 29. September**

Kostenlose [Tickets hier](#) per E-Mail
anfragen an
sara.mazzorana@tcgprocess.de



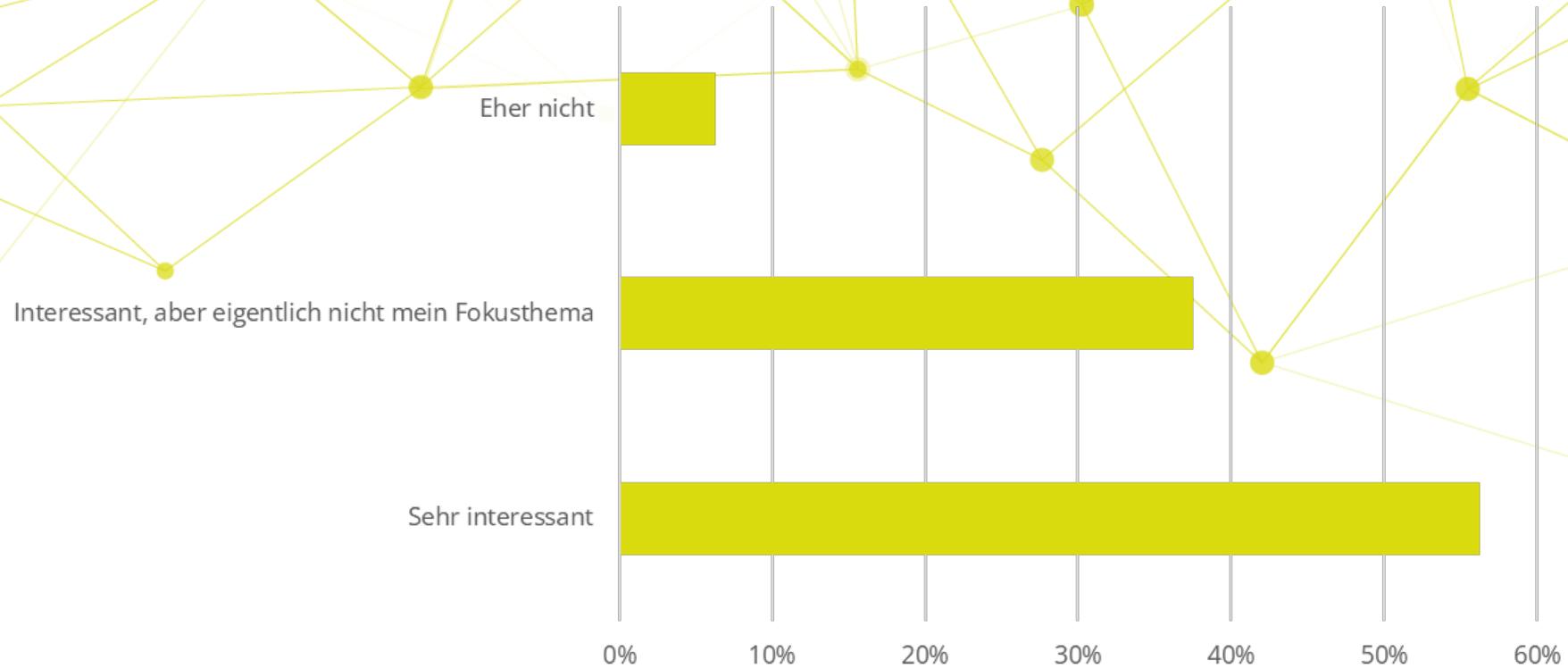
Bleiben Sie auf dem Laufenden und folgen uns auf [LinkedIn](#)

TCGIII

Deep Dive Klassifikation

Umfrageergebnisse aus dem Webinar im Mai 2023

Wie sehr würde Sie ein weiterführendes Webinar mit Deep Dive zur Klassifikation interessieren?



Über die TCG Process

TCG Process Gruppe Weltweit

Seit 2006 am Markt mit inzwischen 14 Standorten weltweit



Schweiz
Deutschland
Chile
Brasilien
Portugal
Spanien
Frankreich
Polen
USA
Kanada
Japan
Australien
Italien
UK

Ihre heutigen Referenten



Erwin Schwarzl
Geschäftsführer

Tel. +49 751 568 4989 12
E-Mail: erwin.schwarzl@tcgprocess.de
LinkedIn: [Erwin Schwarzl | LinkedIn](#)



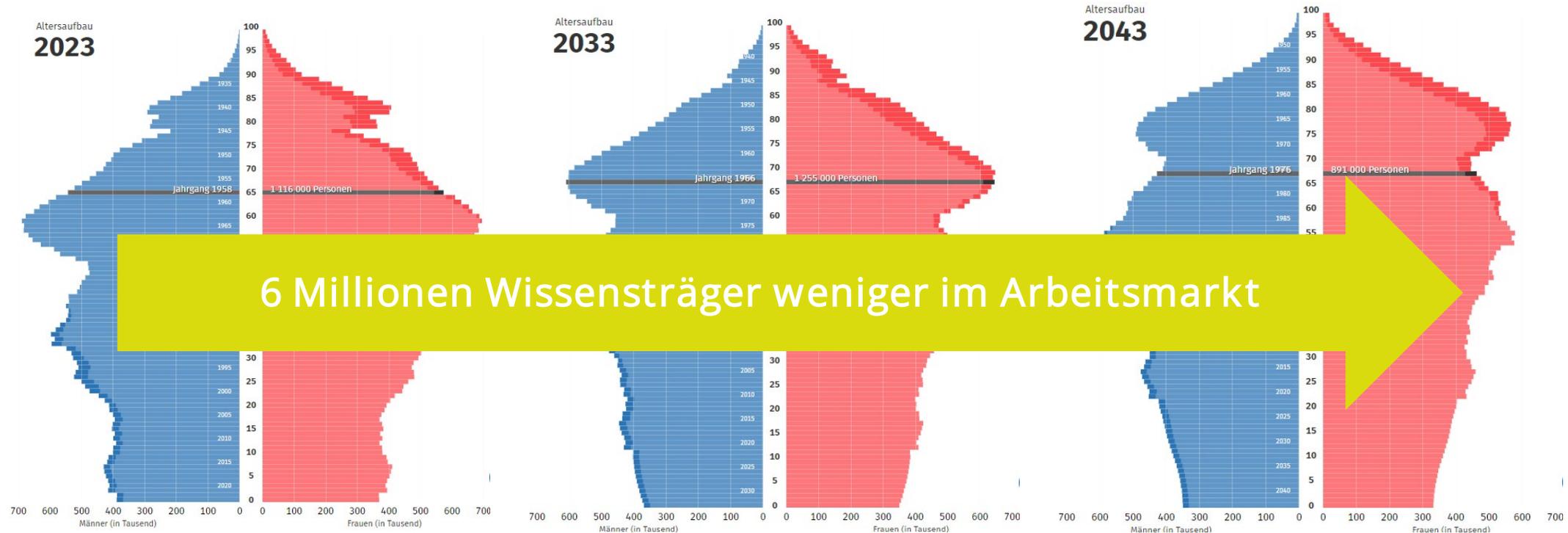
Dr. Oliver Wendel
Professional Services

Tel. +49 751 568 4989 19
Mobil: +49 151 56 14 14 62
E-Mail: oliver.wendel@tcgprocess.de
LinkedIn: [Oliver Wendel | LinkedIn](#)

DocProStar und KI

Warum Automatisierung (mithilfe von KI) ?

KI wird ein wichtiger Wettbewerbsfaktor zur Kompensation des demografisch bedingten Verlustes von Fachkräften und der steigenden Komplexität in unserer Gesellschaft



Quelle Destatis, [Bevölkerungspyramide: Altersstruktur Deutschlands von 1950 - 2060 \(destatis.de\)](https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/07/2307011.html)

DocProStar und KI in der Praxis

Beispiele für erfolgreiche KI-Anbindungen mit DocProStar



DocProStar Classification Suite zum Beispiel für:

- Posteingang in großen Mengen
- Absendererkennung durch Layoutklassifikation
- Personalausweis-Erkennung



Microsoft Cognitive Services

- Read API – auch lokal
- Bildanalyse z.B. für Bildbetrugserkennung



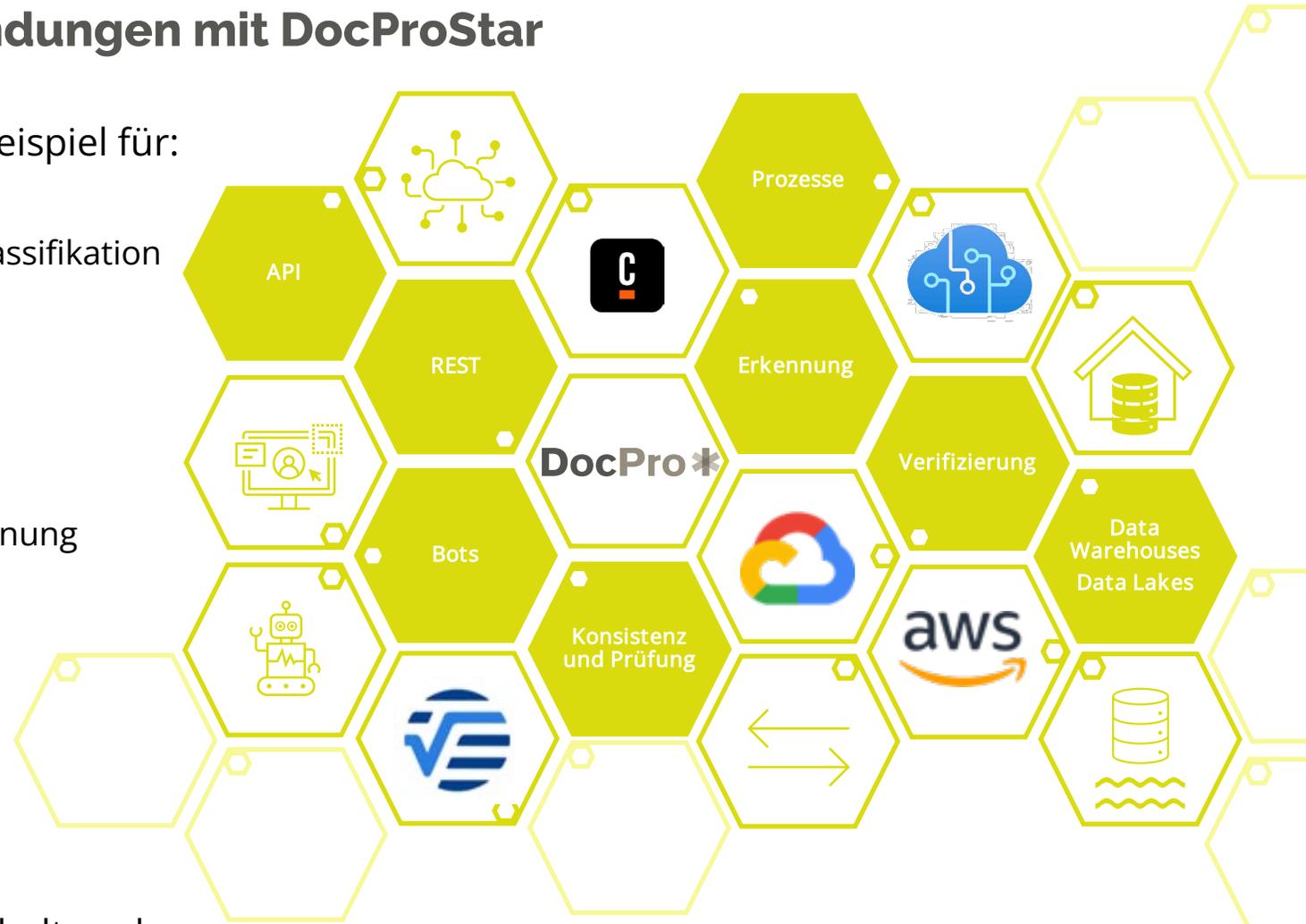
Google Vision OCR Engine



Amazon Rekognition Services –
Gesichtserkennung / Ausweisvergleich



Actineo -Prüfung von Vollständigkeit, Inhalt und
Konsistenz von Vorgängen (KI)



Grundlagen

Was heißt Klassifikation?

Begriffe

Klassifikation (auch: *Typifikation, Systematik*)

Planmäßige Sammlung von abstrakten *Klassen* (auch *Konzepten, Typen* oder *Kategorien*), die zur Abgrenzung und Ordnung verwendet werden

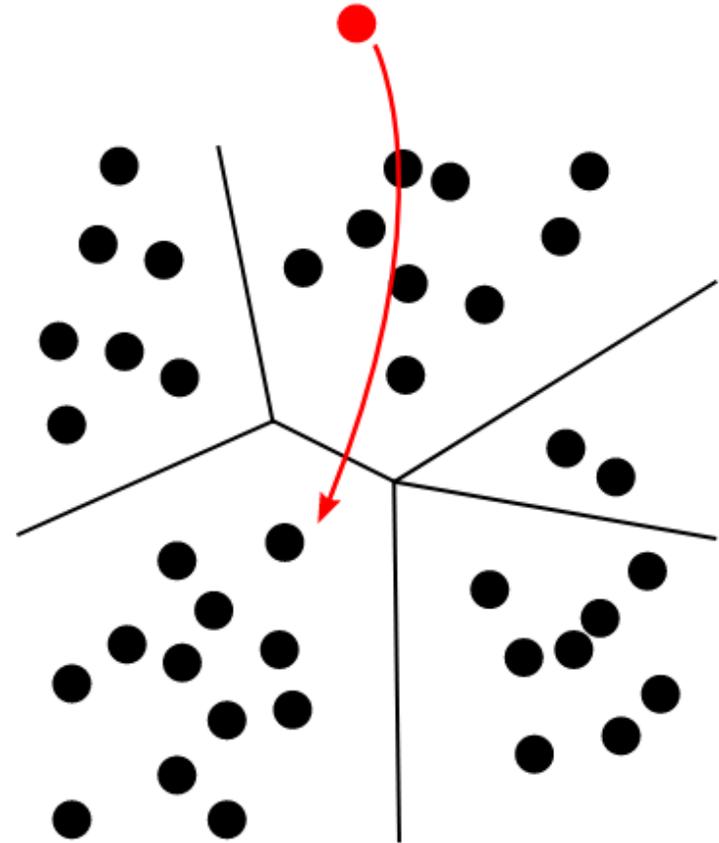
Oft hierarchische Klassifikationssysteme

Klassifizierung

Gewinnung der einzelnen *Klassen* durch die Einteilungen von *Objekten* anhand bestimmter übereinstimmender *Merkmale*

Klassenzuordnung

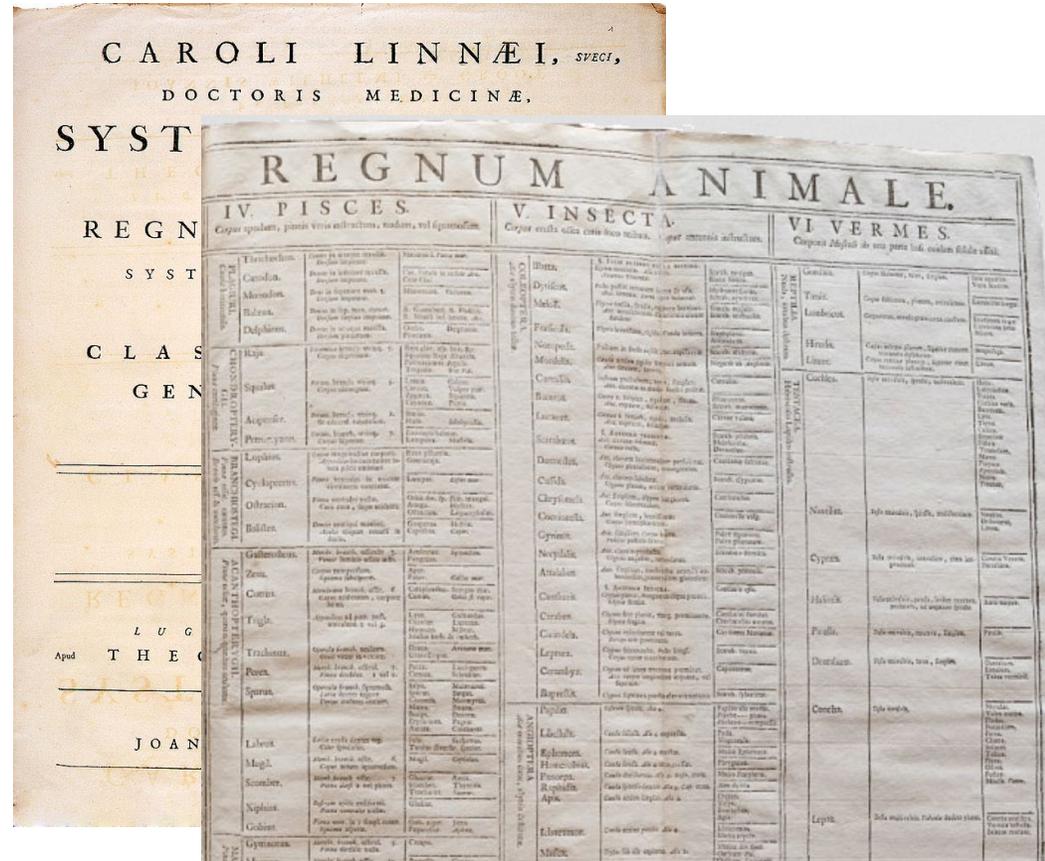
Anwendung einer Klassifikation auf ein Objekt durch Auswahl einer passenden Klasse der gegebenen Klassifikation



Quelle: <https://de.wikipedia.org/wiki/Klassifikation>

Beispiele für Klassifikation - Systematiken

Biologische Systematik *Systema Naturae*, Carl von Linné, 1735



Quelle: v.1 - Caroli Linnaei...Systema naturae per regna tria naturae - Biodiversity Heritage Library (biodiversitylibrary.org)

ICD - *International Statistical Classification of Diseases and Related Health Problems*

Kapitel	Notation	Bezeichnung
I	A00–B99	Bestimmte infektiöse und parasitäre Krankheiten
II	C00–D48	Neubildungen (beispielsweise Tumore u. A.)
III	D50–D89	Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems
IV	E00–E90	Endokrine, Ernährungs- und Stoffwechselkrankheiten
V	F00–F99	Psychische und Verhaltensstörungen
VI	G00–G99	Krankheiten des Nervensystems
VII	H00–H59	Krankheiten des Auges und der Augenlider
VIII	H60–H95	Krankheiten des Ohres und des Innenohrs
IX	I00–I99	Krankheiten des Kreislaufsystems
X	J00–J99	Krankheiten des Atmungssystems
XI	K00–K93	Krankheiten des Verdauungssystems
XII	L00–L99	Krankheiten der Haut und der Unterhaut
XIII	M00–M99	Krankheiten des Muskel-Skelett-Systems
XIV	N00–N99	Krankheiten des Urogenitalsystems
XV	O00–O99	Schwangerschaft, Geburt und Wochenbett
XVI	P00–P96	Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben
XVII	Q00–Q99	Angeborene Fehlbildungen, Deformationen und abnorme klinische Erscheinungen
XVIII	R00–R99	Symptome und abnorme klinische Erscheinungen
XIX	S00–T98	Verletzungen, Vergiftungen und bestimmte andere Folgen von Ursachen
XX	V01–Y98	Außere Ursachen von Morbidität und Mortalität
XXI	Z00–Z99	Faktoren, die den Gesundheitszustand beeinflussen
XXII	U00–U65	Schlüsselnummern für besondere Zustände

INTERNATIONAL STATISTICAL CLASSIFICATION OF DISEASES AND RELATED HEALTH PROBLEMS

INFORMATION SUPPORT TO PRIMARY HEALTH CARE

- Lay reporting
- Other community-based information schemes in health

ICD 3-CHARACTER CORE CLASSIFICATION

- Diagnoses
- Symptoms
- Abnormal laboratory findings
- Injuries and poisonings

SPECIALTY-BASED ADAPTATIONS

- Oncology
- Dentistry and stomatology
- Dermatology
- Psychiatry
- Neurology
- Obstetrics and gynaecology
- Rheumatology and orthopaedics
- Paediatrics, etc.
- General medical practice

Quelle: International statistical classification of diseases and related health problems (who.int), <https://de.wikipedia.org/wiki/Klassifikation>

Klassifikation von Dokumenten

Worum geht es?

- Geschäftsprozessrelevante Informationen stecken verborgen in **unstrukturierten** Dokumentformaten
- Für die **Automatisierung** werden aber strukturierte Daten benötigt
- **Dokumentklassifikation und Inhaltsanalyse** können diese unstrukturierten Daten in strukturierte Formate abbilden und die Prozess-relevanten Informationen extrahieren oder ableiten bzw. erzeugen
- Manuelle Klassifikation ist keine Option

Was braucht es dazu?

- Ein zuverlässiges, performantes, **automatisches Klassifikationssystem** mit reproduzierbaren, konsistenten und optimalen Ergebnissen



DocProStar Classification Suite

Technologien

Integration verschiedener AI- und ML-Ansätze

- Neuronale Netze
- Statistische Verfahren
- Pattern Matching
- Supervised & Unsupervised Learning
- Explizite Regeln

Grundlegende Qualitätsmaße: Recall und Precision

Unterstützung des kompletten Entwicklungs-Zyklus für Klassifikations-Modelle:



Recall und Precision

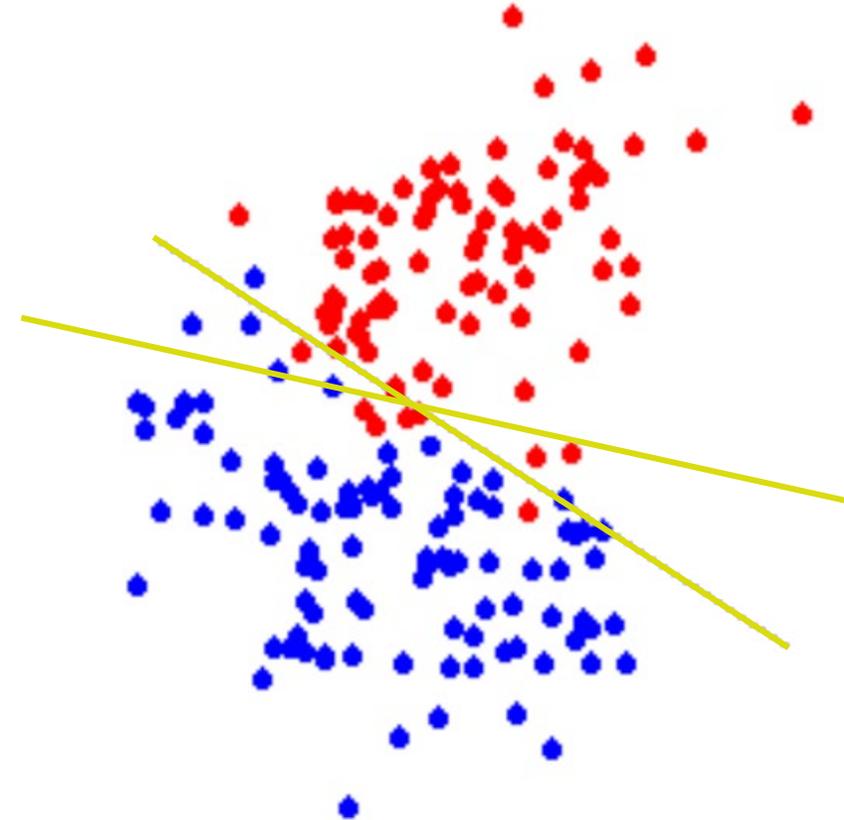
Trefferquote vs. Genauigkeit

Beispiel

- Jeder Punkt entspreche einem Dokument
- Rot und blau repräsentieren 2 Dokumentklassen
- Der Merkmalsraum hat nur 2 Dimensionen

Aufgabe:

- Wohin wäre eine Trennlinie am besten zu platzieren?
- Welche „Ausreißer“ nimmt man in Kauf?



Quelle: <http://www.statistics4u.info/>

Überwachtes Lernen (Supervised Learning)

Merkmale	Content Classifier	Layout Classifier
Welche Klassifikationselemente werden verwendet?	 Verwendung von Text-Elementen der Dokumente zur Klassifikation	 Verwendung von Layout-Elementen der Dokumente zur Klassifikation
Ist OCR erforderlich?	<input checked="" type="checkbox"/> ja	<input type="checkbox"/> nein
Ermittlung von Klassenzugehörigkeit und Ähnlichkeit	 durch Text -Vergleich	 durch Layout -Vergleich
Besonderheiten	 Text-Features sind fehlertolerant bezüglich Schreibvarianten oder -fehlern	 Muster oder Pattern können sein: Linien, Logos, Textblöcke, Zusammenhangskomponenten
Sprachunabhängig	<input checked="" type="checkbox"/> ja	<input checked="" type="checkbox"/> ja (keine OCR)
Regeln	 Textregeln anwendbar	 geometrische Bereiche definierbar

Erstellung eines Klassifikationsmodells

Zyklus für den Content und Layout Classifier



Informationen zur Demo

Infos	Content Classifier	Layout Classifier
Datensatz	<p>Newsgroups - Diskussionsforen</p> <ul style="list-style-type: none">• 20 Newsgroups• 1000 Dokumente pro Newsgroup	<p>Ghega-Dataset: 2 Dokumentkategorien</p> <ul style="list-style-type: none">• 110 Datenblätter elektronischer Bauteile• 136 Patentschriften
Quelle	<p>Tom Mitchell, Carnegie-Mellon Univ., USA, 1999</p> <p>Referenz: https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups</p>	<p>Machine Learning Lab, University of Trieste, Italien</p> <p>Referenz: http://machinelearning.inginf.units.it/data-and-tools/ghega-dataset</p>
Demo	<p>20 Klassen 2267 Dokumente ca. 100 Dokumente pro Klasse</p>	<p>8 Klassen 130 Dokumente ca. 10-30 Dokumente pro Klasse</p>

- alt.atheism
- comp.graphics
- comp.os.ms-windows.misc
- comp.sys.ibm.pc.hardware
- comp.sys.mac.hardware
- comp.windows.x
- misc.forsale
- rec.autos
- rec.motorcycles
- rec.sport.baseball
- rec.sport.hockey
- sci.crypt
- sci.electronics
- sci.med
- sci.space
- soc.religion.christian
- talk.politics.guns
- talk.politics.mideast
- talk.politics.misc
- talk.religion.misc

- diodes-zener
- espacenet-nv
- rohm-zener-1
- sharp-led-1
- slovenia
- spagna
- taiwan-switching
- wipo-de

Live Demo

Content- und Layout-Klassifikation

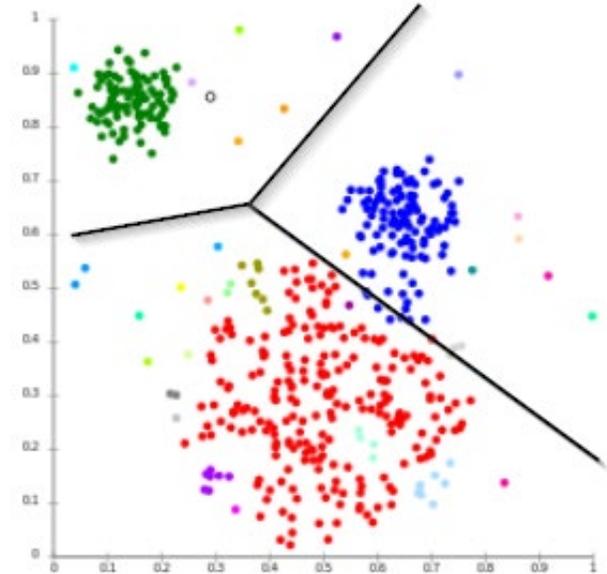
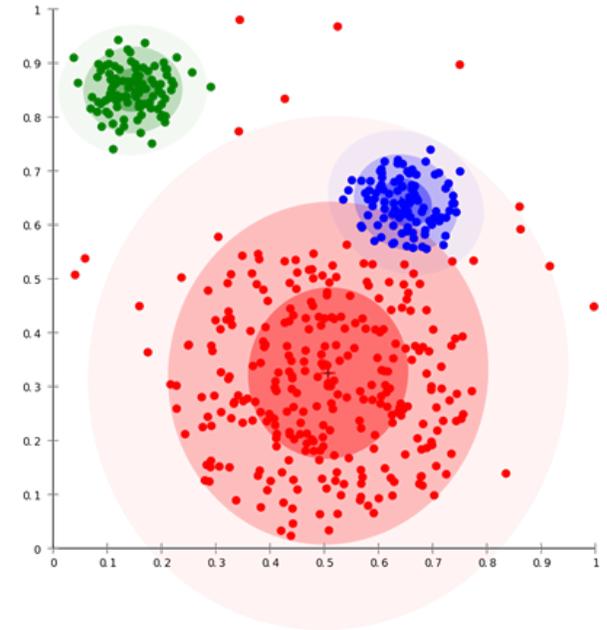
Clusteranalyse (Unsupervised Learning)

Data-Mining, Knowledge Discovery

Was ist Clusteranalyse?

Ein Verfahren zur Entdeckung von Ähnlichkeitsstrukturen in (meist relativ großen) Datenbeständen

Der Algorithmus selbst liefert keine Interpretation der gefundenen Gruppen ! Hierzu ist eine separate Analyse der Gruppen durch einen Domänen-Experten notwendig



Clustering (Unsupervised Learning)

Clusteranalyse: Topic- und Layout-Detection

- Clustering
 - Input: Menge unsortierter Dokumente
 - Output: Ordner mit ähnlichen Dokumenten
- Anwendung
 - Cluster als Ausgangspunkte zum Aufbau von Trainingsmengen für ein Klassifikationsprojekt
 - Verwenden bei bestehenden Trainingsmengen, um unbekannte Themencluster oder bessere Klassen-Partitionierungen zu entdecken

The screenshot displays the DocProStar interface for 'Image Auto Clustering'. The left sidebar shows a list of clusters with their respective document counts:

Cluster	Count
Cluster 11	13
Cluster 3	12
Cluster 12	9
Cluster 1	9
Cluster 7	8
Cluster 8	7
Cluster 6	7
Cluster 9	7

The main window shows a document preview for 'BZX84C2V4S - BZX84C39S'. The document content includes:

- Features:** Plastic Die Construction, 200W Power Dissipation, Zero Voltage from 2.4V - 30V, Ultra-Easy Surface Mount Package.
- Mechanical Data:** Case: SOT-233, Molded Plastic, Case material: 10, Flame-retardant (UL94V-0), Minimum order quantity: 1,000 (1,000 in 1000s), Packaging: Substrate per MIL-STD-202, Marking: 016, Die Design: 1, Marking Size: Table page 2, Weight: 0.004 grams (approx.).
- Maximum Ratings:** A table with columns for Parameter, Symbol, Rating, and Unit.
- Ordering Information:** A table with columns for Part, Packaging, and Shipping.

Performance metrics for the clustering process are shown in the top right:

- Number of clusters: 36
- Average classification time (ms): 52
- Total processing time (sec): 15.60
- Minimum classification time (ms): 18
- Average training time (ms): 38
- Maximum classification time (ms): 149
- Total number of documents: 130

Live Demo

Clustering

Vortrainierte Klassifikatoren

Classification Knowledge Bases

Was ist eine Knowledge Base (KB)?

- Konversion von Klassifikatoren einer Domäne in eine Wissensbasis
- KB enthält keine Klartextinformationen, nur den Klassifikator mit abstrakten Features

Anwendung:

- Domänenspezifische KB kann in anderen Projekten wiederverwendet werden
- Kombination von KBs: Untereinander und mit eigenen Trainingsmengen

Beispiel: Human Resources

- KB: Standardklassen (Lebenslauf, Bescheinigungen, Zeugnisse)
- Projekt-/Kunden-spezifische Klassen als zusätzliche Trainingsmengen



Online Learning Classification

Kontinuierliche Verbesserung der Klassifikation mit Machine Learning Ansatz

- Kontinuierliches Lernen und Optimieren der Klassifikatoren
- Prozessschritt speichert die manuell vorgenommenen Klassifikationskorrekturen
- Background Service erzeugt neue Versionen des Klassifikationsprojekts in definierten Zeitabständen
- Die neue Version wird automatisch getestet
- Nur wenn die neue Version besser ist, wird sie automatisch für Produktion freigegeben



Klassifikation wird kontinuierlich besser & verhält sich dynamisch adaptiv gegenüber dem Beleggut, das sich über die Zeit verändert

Dokument-Klassifikation in Projekten

Empfehlungen und Hinweise

Vorbereitung



- Sichtung des Schriftguts
- Identifikation geschäftsprozess-relevanter Dokumente und Klassen

Klassifikationsstruktur und Trainingsmenge



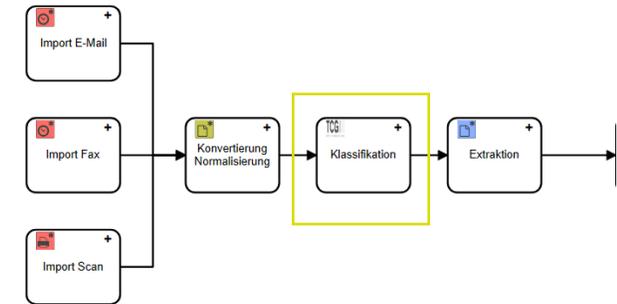
- Reduktion der Klassen, Harmonisierung
- Repräsentative, gute Trainingsdokumente
- Trennschärfe
- Klasse „Sonstige“ vermeiden

Kontinuierliche Pflege



- Dokumente und Klassen ändern sich dynamisch

Integration in DPS



- Einfache Integration des Klassifikationsmodells über Prozessaktivität in einen DPS-Prozess

Praxisbeispiele

Nutzungsbeispiele in 2 Branchen

Versicherung

- Pro Sparte (GI, Leben, Schaden) eigene Klassifikatoren
- Innerhalb von Schaden eigene Klassen für Kfz, Haft, Sach, Rechtsschutz, Unfall
- Document vs. Page Classifier

Energieversorger

- Digitalisierung der Hauspost
- 10 Dokumenttypen werden automatisch klassifiziert
- der Rest mittels maschinell gefundener Schlagworte zu bestätigen
- 40 E-Mail Postfächer
- Portal

Für gute Dunkelverarbeitungsrate müssen die Klassifikatoren laufend gepflegt werden.

Ausblick: Digitaler Posteingang mit DPS Mailroom Standardlösung

Mehrstufige Standardlösung zur Steigerung der Automatisierung

DocPro*

MAILROOM

1

Einfache kostengünstige Lösung mit **datenbankgestützter manueller Erfassung des Bearbeiters** und anschließender **Verteilung** per E-Mail oder unserer Postkorblösung

2

Erweiterung mit **schlagwortbasierter Klassifikation** zur automatischen **Bestimmung des Anliegens** und des **Bearbeiters**

3

Erweiterung um **KI-basierte Klassifikation** mit **automatischem Training** im Hintergrund

4

Individuelle projektspezifische Mailroomlösung mit kundenspezifischem Design, Training und Optimierung der Automatisierungsraten

ab Q4 2023



Vielen Dank für Ihre Aufmerksamkeit!

Gibt es Fragen?

Webinar verpasst?

Unsere Aufzeichnungen gibt es jetzt auch auf Youtube, die Links dazu gibt es auf unserer Website www.tcgprocess.de **Download – TCG Process GmbH:**



[DOWNLOAD](#) | [PLATTFORM](#) | [KONTAKT](#)

DEMO BUCHEN

Webinare

Jetzt kostenfrei für unser nächstes Webinar anmelden zum Thema

Deep Dive Klassifikation

JETZT REGISTRIEREN

Business Process Management Plattform DocProStar

Webinar



DPS als Business Process Management Plattform

AUFZEICHNUNG

SLIDES

Eingehende E-Mails & Posteingang digital verarbeiten

Webinar

Speaker: Lutz Schäfer
Moderation: Sara Mazzorana



Eingehende E-Mails und Posteingang digital verarbeiten

AUFZEICHNUNG

SLIDES

Auftragsbestätigungen digital verarbeiten

mit Marcus Kern

Webinar



Wir stellen

Webinar



Business Intelligence Tools von DocProStar

mit Marcus Kern

Webinar



Follow-up

Alle wichtigen Infos zum Webinar erhalten Sie per E-Mail.

- PPT Slides
- Aufzeichnung
- Fragen und Antworten



Bleiben Sie auf dem Laufenden und folgen uns auf [LinkedIn](#)

SAVE THE DATE



12.09

Neuigkeiten DPS Invoice



28.09

DFC München



28.09

Partnerfestival Leipzig

Kostenlose [Tickets hier](#) per E-Mail anfragen an sara.mazzorana@tcgprocess.de