Briefe zur KLASSIFIKATION

Mitgliederbrief der Gesellschaft für Klassifikation e.V. Data Science Society

April 2021

Inhalt

Grußwort des Präsidenten der GfKl 2

ZÄHLEN – UND DANN? Anmerkungen zur COVID-Stellungnahme der DAGStat 3

Mitgliederversammlungen 2020 bis 2022 6

Bericht des Schatzmeisters der GfKl 7

Bericht zu den Advances in Data Analysis and Classification (ADAC) 8

Bericht zu den Archives of Data Science 9

Bericht zu den Studies in Classification, Data Analysis, and Knowledge Organization 10

Information zur ECDA 2021 vom 07. bis 09. Juli 2021 virtuell in Rotterdam 11

Bericht der Arbeitsgruppen Bibliotheken (AG BIB) und Dezimalklassifikation (AG DK) 12

Bericht der Arbeitsgruppe Biostatistik (AG BT) 12

Bericht der Arbeitsgruppe Datenanalyse und Numerische Klassifikation (AG DANK) 12

Arbeitsgruppe Ethik (AG Ethik) (in Gründung) 12

Bericht der Arbeitsgruppe Datenanalyse und Klassifikation im Marketing (AG MARKETING) 13

Personalia 15

© Gesellschaft für Klassifikation e.V. Data Science Society

Grußwort des Präsidenten der GfKl

Liebe Mitglieder und Interessierte,

nun dreht sich seit über einem Jahr (fast) alles um die Corona-Pandemie und eigentlich könnten wir uns als Data Science Society über das nahezu exponentiell gewachsene Interesse an aussagekräftigen Zahlen zur Pandemie, deren Analyse und Interpretation uneingeschränkt freuen. Und doch scheinen nach nunmehr 14 Monaten intensiver Pandemieerfahrung, diversen Lockdowns in gradueller Abstufung und sonstigen aus den Zahlen abgeleiteten Best-Practice-Maßnahmen die Fragezeichen nicht weniger zu werden. Wir scheinen zurückgeworfen auf elementare Grundfragen der Datenanalyse und Statistik, die weiterhin unbeantwortet bleiben: – Wie steht es um die Qualität der verfügbaren Daten? Wie valide sind die eingesetzten Analyse- und Modellierungsverfahren? Wie vergleichbar sind Daten und Situationen aus verschiedenen Ländern/Regionen/Kontexten? Wie steht es um die externe Validität experimenteller Laborergebnisse? Fragen über Fragen, über deren Beantwortung Expert:innen in fachlichen Wettstreit treten und die dennoch als Grundlagen weitreichender politischer Entscheidungen Verwendung finden.

Eine Arbeitsgruppe der DAGStat – darunter auch Vorstandsmitglieder der GfKl – haben im Vorfeld des DAGStat-Symposiums „Zahlen an die Macht? Wie Daten und Statistik Politik beeinflussen“ eine Stellungnahme zum Thema "Daten und Statistik als Grundlage für Entscheidungen: Eine Diskussion am Beispiel der Corona-Pandemie" erarbeitet. Unser geschätzter Kollege Prof. Dr. Ulrich Müller-Funk (Uni Münster) hat diese Stellungnahme einer kritischen Betrachtung unterzogen und seine Anmerkungen in einem Beitrag hier zusammengefasst. Ich würde mich freuen, wenn seine pointierten Aussagen Leser:innen der Briefe zur Klassifikation animieren, hierzu in eine aktive Diskussion einzusteigen. Gerne stelle ich hier eine Sonderausgabe der Briefe zur Klassifikation zusammen oder kümmere ich mich um die Einrichtung einer digitalen Diskussionsplattform.

Für unsere Jahrestagung “European Conference on Data Analysis” haben wir dieses Jahr ein sehr spezielles Format gewählt. Sie findet nicht nur online statt, sondern auch in Kooperation mit der DSSV, die von der International Association for Statistical Computing (IASC) ausgerichtet wird und seit 2017 jährlich die Konferenz-Landschaft bereichert. Die lokalen Organisatoren der Erasmus University in Rotterdam unter der Leitung von Andreas Alfons haben mit Hilfe des Scientific Programme Committees zugkräftige Keynote Speaker und interessante Themen für eingeladene Sektionen zusammengestellt. Durch die Online-Variante werden wir eine geringere Anzahl von Sektionen ermöglichen und dadurch nur eine begrenzte Anzahl von eingereichten Vorträgen für interaktive Sitzungen akzeptieren. dennoch hoffen wir auf zahlreiche Einreichungen, die noch bis 15. Mai möglich sind, Näheres finden Sie in in diesen Briefen zur Klassifikation.

Vor gut zwei Jahren hat sich in München die German Data Science Society (GDS) e.V. gegründet. Die Namenswahl, die Gesellschaftsziele und insbesondere der Antrag zur Aufnahme bei der DAGStat haben damals zu intensiven Diskussionen auf der DAGSTAT-Vertreter- versammlung sowie der GfKl-Mitgliederversammlung 2019 geführt. Im letzten Jahr wurde ich vom Vorstand der GDS gebeten in den dortigen Beirat einzutreten. Diesem Wunsch kam ich gerne nach, ermöglicht es doch, Synergien zwischen den beiden Gesellschaften herzustellen und ein kooperatives Miteinander zu initiieren, zumal sich in der alltäglichen Arbeit der GDS die Unterschiede zur GfKl – berufsständische Vertretung der Data Scientists dort, forschungsorientierte wissenschaftliche Vereinigung hier – immer klarer abzeichnen. Aktuell

möchte ich an dieser Stelle auf die nächste Veranstaltung der GDS zum Thema "Data Science, KI und Machine Learning - was bedeutet das in der Versicherung?" hinweisen.

Als weiter Leseempfehlung möchte ich es nicht versäumen, auf die Data-Literacy-Charta hinzuweisen, die auf Initiative des Stifterverbandes mit zahlreichen Partnern, darunter die GfKl im Rahmen der DAGStat, entstanden ist.

Nun wünsche ich Ihnen allen angenehme Lektüre, hoffe auf eine aktive Diskussion der Themen und auf zahlreiche Teilnahme an unserer Jahrestagung bei der DSSV-ECDA 2021.

Mit herzlichen Grüssen und bleiben Sie gesund,

Ihr

Adalbert Wilhelm

Vorsitzender der Gesellschaft für Klassifikation e.V. Data Science Society

ZÄHLEN – UND DANN? Anmerkungen zur COVID-Stellungnahme der DAGStat

U. Müller-Funk, Universität Münster

Die Analyse von COVID- Daten hat Schwachstellen deutlich werden lassen: Die Datenlage, die Kommunikation statistischen Argumentierens und die Prognose-Modelle selbst. Eine Äußerung seitens der geballten deutschen Statistik hierzu ist angebracht. Deren in der Einleitung formulierte Fokus geht über diese Punkte weit hinaus: „Die Entscheidungsunterstützung in Krisensituationen".

Man fragt sich schnell, an wen der Artikel eigentlich gerichtet ist, wer diesen überhaupt zur Kenntnis nehmen soll. Im Militär - Funk gibt es den Aufruf „Charly an alle, kommen!", der scheint hier Pate gestanden zu haben. Die Sprache wirkt teils unangenehm belehrend, ohne immer argumentativ zu sein, weckt dank zahlreicher Soll-Aufrufe Assoziationen an Ratgeber zur Gesundheit. Das Fehlen einer Ansprechgruppe geht einher mit der Flucht ins Allgemeine - bis hin zum Banalen. Der Abschnitt„Daten" liest sich wie die ersten Seiten eins Skripts für Statistik-Einsteiger. Zum Datenproblem der Corona-Krise wäre Kluges zu Schlagworten wie „Metadaten", „Schichtung", „räumliche Zuordnung", „Dunkelziffern" hilfreich gewesen. Auf den Aspekt „Datenauswertung" wurde verzichtet, da moderne Krisen sich wohl den Gesetzmäßigkeiten mechanischer Systeme gebeugt haben, wie der nächste Teil suggeriert. Was soll aber eine Taxonomie blutleerer Validierungsbegriffe, ohne die Aufgabe selbst zu beleuchten? Wer soll aus den Erklärungsversuchen zur politischen Entscheidungsfindung Honig saugen? Glaubt man, Herr Spahn würde, beglückt durch diese Lektüre, Entscheidungsbäume malen? Unter der Überschrift „Internationale Initiativen" hätte man mehr erwartet als nur weitere Daten-Gebote. Gemeinsam signalisieren die Ausführungen ein Desinteresse an jeder Art von Krise. Über diese Belanglosigkeiten kann bedenkenlos der abgewetzte Mantel der Nächstenliebe gebreitet werden. Ein einziger Abschnitt, der zu Kommunikation und Visualisierung, liefert einen inhaltlichen Beitrag, fokussiert sich auf das Thema COVID. Die folgenden Anmerkungen betreffen daher allein die „Literacy" und Aspekte der Modellierung und Prognose.

Stichwort: „Statistical Literacy"

Zu Ende des Abschnitts machen die Verfasser einen Schlenker zur Initiative des Stifterverbandes. Er betont die Herausforderung „... Menschen zu befähigen Sinn, aus diesen Daten zu erschließen, richtet sich gleichermaßen an Bildungseinrichtungen (z.B. Schulen, Hochschulen, Volkshochschulen) wie auch an Medien und Produzenten von öffentlichen Statistiken". Der Wunsch ist schön, sei es nur um belanglose Studien wie „Dicke/Optimisten/Polygame/Dirigenten leben länger" überflüssig zu machen. Führen breit gestreute Statistik-Impfaktionen aber wirklich zur Herdenkompetenz, ist die überhaupt erreichbar? Wer und mit welcher Qualifikation ausgestattet soll das umsetzen?

Datenpicken ist ein bewährtes Instrument zur Durchsetzung von Interessen. Gesellschaftlichen Gruppierungen sind „Stiftungen" und „Institute" zuzuordnen, deren „Experten" helfen Ansprüche zu belegen, Positionen pseudowissenschaftlich zu veredeln. „Data analytics" ist ein Geschäft für unterschiedlichste Escort Services, wie etwa Beratungsfirmen und Fraunhofer-Institute, aber auch für Hochschullehrer, in Form kommerzieller Weiterbildung. Ein Blick auf Inhalt und Umfang zeigt, dass zumeist etwas deskriptive Statistik, garniert mit aktuellen Modeworten, als „machine learning" angepriesen wird. Die schnelle Nummer führt eher nicht zu „statistical literacy".

Hochschulen sind auch kein Quell der Hoffnung. Statistik - als Teil realwissenschaftlicher Studiengänge - gibt es umfänglicher in der Psychologie und in eher ärmlicher Form in Medizin und Wirtschaft - und Sozialwissenschaften. Eine adäquate Ausbildung, hierzu wie auch in praktischer Informatik, ist gegenwärtig nicht durchsetzbar. Ein Promovend aus diesen Bereichen ist oft auf ein statistisches Thema angesetzt, für das ihm Methodenwissen fehlt- wie auch dem Doktorelter.

Bootcamp-Veranstaltungen schaffen dadurch Abhilfe, dass sich Beinahe -Statistiker an den Modulen von Packages wie „Statlab“ entlanghangeln, um mittels Clickstreams statistische Zusammenhänge zu vermitteln, ganz praxisnah. In Fortsetzung führt diese erfolgreiche statistische Alphabetisierung zu der exzellenten, trotzdem umfangreichen, empirischen Forschung, die auf universitären Homepages poliert und hervorgehoben wird. Wer soll aber eigentlich die Missionsarbeit in anderen Fakultäten leisten, etwa Zusatzveranstaltungen für Lehramtskandidaten philologischer Fächer anbieten? Die Grausamkeiten zu schildern, die der Stochastik in Gymnasien widerfährt, soll allen erspart bleiben.

Die erhofften Nothelfer entpuppen sich eher als „super spreader“ statistischen Analphabetentums und es braucht auch keine Bewegung „Thursdays for Data“. Wesentlich wäre es, gezielt spezifische statistische Expertise anzubieten und einzufordern, etwa in der professionellen Ausbildung für Journalisten oder für Daten-Verantwortliche in der Verwaltung.

Stichwort: „Modelle und Prognosen“

Der Begriff „Modell“ durchzieht die ganze Stellungnahme, primär im Zusammenhang von Prognose und Simulation – ohne eine Konkretisierung. Anforderungen an ein Prognosemodell werden nicht diskutiert, der Begriff wird „irgendwie“ rein technisch verwendet. Ausgangspunkt sollte die Binsenweisheit sein, dass ein Modell in diesem Zusammenhang einerseits eine mathematische Karikatur des realen Sachverhaltes sein soll, andererseits aber nur Größen aufweisen darf, über die die vorhandenen Daten auch Auskunft geben können.

Ein statistischer Blick auf das SIR-Modell identifiziert S(t), I(t), R(t) als Zählprozesse - und sonst nichts. Deren Zusammenhang kann mit 0-1 Variablen erfasst werden, Zeitangaben durch Markierungen. Ein Ansatz wurde von D.G. Kendall 1956 publiziert, im 3. Berkeley Symposium. Die Theorie von Punkt- und Zählprozessen ist allerdings erst in den 70ern und 80ern entwickelt worden, ihre Anwendungen laufen unter der Überschrift „Time-to-Event Analysis“. Viele Beiträge im Bereich Epidemiologie scheinen da nicht publiziert worden zu sein. Ohne ein solches Modell ist allerdings auch eine stochastische Simulation nicht möglich. Auf welcher Basis sonst sollen die Wartezeiten für die Ereignisliste generiert werden? Für einen Nicht-Biometriker bleibt unklar, dass solche stochastischen Modellierungen in der Analyse von COVID-Daten keine Rolle spielen. Eine Antwort hätte man in der Stellungnahme zwingend erwartet. Stochastische Sprung - Prozesse sind ja auch zentrale Elemente für die Modellierung diverser Szenarien im Rahmen des Krisenmanagements, Stichwort Logistik.

Ein stieftöchterliches Dasein fristet in der Stellungnahme die Reproduktionszahl, obwohl die ein Indikator für die Dynamik sein könnte. Ihre Schätzung nach RKI-Rezept, auf zwei Nachdezimalstellen genau, leidet unter der Datenlage, u.a. den Dunkelziffern, ist aber auch methodisch nicht wirklich überzeugend. Ihre stereotyp wiedergegebene Interpretation wird bislang nicht hinterfragt. Allen Prognoseversuchen von Corona-Verläufen sind ebenfalls durch die Datenlage enge Grenzen gesetzt, nämlich durch die phasenweise hohe Nicht-Stationarität. Trotzdem wird die Öffentlichkeit mit teils mehrmonatigen Prognosen konfrontiert. Öffentliche Einwände von Biometern sind nicht hörbar.

Motiviert durch die physikalische Epidemiologie, hat ein Exkurs über dynamische Modelle Raum in der Stellungnahme gefunden. Auch Laien stellen sich da wohl die Frage: Was hat das Abzählen von Ereignissen - die zu isolierten, zufälligen Momenten auftreten – mit einem deterministischen, dynamischen System in einem Kontinuum zu tun? Die Antwort können die sich selbst geben: Nichts. Die im Text angeführten deterministischen DGL- Modelle aus Biologie, Physik, VWL und eben auch SIR sind sämtlich für die mathematische Deduktion von Mustern entwickelt worden, für die Präzisierung gedanklicher Experimente. Für eine Induktion, das Lernen aus Daten, sind die Lösungskurven der DGL völlig ungeeignet, weil diese eine synthetische Struktur aufgepfropft bekommen. Auf der baut dann die anschließende Analyse auf. Eingespeiste reale Daten können sich leider auf die für sie ungewohnten Verhältnisse nicht richtig einstellen und daher fehlt ihnen die Aussagekraft für die Parameter der DGL. Der Verweis auf eine „Approximation" greift nicht. Dazu müsste man zumindest präzisieren können, wie diese deterministischen Kurven sich denn zu den stochastischen Zählprozessen faktisch verhalten, was da eigentlich prognostiziert wird, was ein

„coronamässiger Trend" bedeuten soll. Der physikalische Experte kann glücklicherweise beruhigen: In der „Praxis" funktioniere das bestens - wie immer das ex post auch gemessen wird.

Resümee

Aus öffentlicher Sicht hat die Medizinstatistik/Biometrie nichts zur Analyse und Prognose von COVID-Daten beigetragen. Die Datenkulisse hinter den Diskussionen von Virologen, Politikern, Ethikern und Journalisten ist im Wesentlichen das Werk von Physikern. Das verwundert mit Blick auf das für diese Gruppe übliche Arbeitswissen „Mathematik". Diffusionsprozesse etwa kommen dort schon mal vor, die physikalische Denkweise ist aber bestimmt durch Einsteins These: Der Alte würfelt nicht. Viren widersprechen erst und mutieren dann. Quereinsteiger haben die Statistik immer wieder bereichert, mit der Betonung auf „einsteigen". Davon kann hier keine Rede sein. Mit einer schon kindlichen Naivität wird Methodik und Intuition eines anderen Faches einfach negiert und unreflektiert durch Ansätze ersetzt, zu denen vertrautes physikalisches Lehrbuchwissen den Zugang erleichtert. Eine Diskussion oder gar Begründung der Vorgehensweise unterbleibt, ein Vergleich mit alternativen Ansätzen ebenfalls - vermutlich schon aus Unwissenheit. Diese Attitude offenbart auch einen schon erstaunlichen Mangel an Selbst-Reflexion. Richard Feynman hat die Vorstellungswelt hierzu sehr schön in Form der „cargo cult science" beschrieben.

Die Statistik hat das bislang stoisch registriert. Eine DAGStat-Stellungnahme hätte dazu mal Position beziehen können. Stattdessen zeichnet diese der Öffentlichkeit ungewollt ein Bild des Faches und des Statistikers, das wenig schmeichelhaft ist. Ein Romancier des 19.Jahrhunderts hat einer seiner Figuren, einer englischen Lady, den Satz in den Mund gelegt:"It never came to my mind that statisticians enter ordinary social relationships". Die Scheu vor der Welt scheint mittlerweile auch die vor Daten, Modellen und Prognosen zu umfassen.

Was nun? Vielleicht eine Stellungnahme 2.0 und eine Arbeitsgruppe „Medizinstatistik", die eine Perspektive für einen statistischen Beitrag zur Corona-Krise erarbeitet. Wir wollen Virologen und andere doch nicht allein lassen – und schon gar nicht unsere Kollegen aus der Physik.

Autor:	Ulrich Müller-Funk

Mitgliederversammlungen 2020 bis 2022

Im Jahr 2020 ist nicht nur zum ersten Male in der Geschichte der GfKl die Jahrestagung ausgefallen, sondern in einer etwas gedehnten Auslegungen unserer Satzung haben wir auf die Durchführung einer Mitgliederversammlung verzichtet. Da es für 2021 ebenfalls unwahrscheinlich erscheint, eine Mitgliederversammlung in Präsenz durchzuführen, hat der Vorstand in seiner letzten Sitzung beschlossen, die nächste Mitgliederversammlung virtuell im Rahmen der DSSV-ECDA 2021 anzusetzen. Eine entsprechende Einladung folgt separat zu gegebener Zeit.

Für 2022 stehen wieder Vorstandswahlen an. Die ECDA 2022 ist für Neapel geplant, wir wollen dann die 2020 ausgefallene Ausgabe der ECDA zum Thema *Avoiding drowning in the data: recent trends and future challenges in learning from data* nachholen. Die offizielle Jahrestagung der GfKl und die Mitgliederversammlung 2022 werden im Rahmen der DAGSTAT-Tagung in Hamburg stattfinden, die vom 28. März bis zum 01. April 2022 geplant ist. Für die ordentliche Gültigkeit der Wahlen und sonstiger Beschlüsse der Mitgliederversammlung ist die Durchführung der Mitgliederversammlung innerhalb Deutschlands erforderlich. Zudem liegt die DAGSTAT terminlich passend zur Laufzeit der aktuellen Vorstandschaft.

Autor:	Adalbert Wilhelm

Bericht des Schatzmeisters der GfKl

Bericht des Schatzmeisters zum Vereinsjahr 2020

Die Gesellschaft für Klassifikation e.V. (GfKl) verzeichnet am 31.12.2020 einen positiven Vermögensstand von 53.093,85 €. Dieser setzt sich aus dem Guthaben auf dem Girokonto (32.642,17 €), dem Festgeldkonto1 (10.225,84 €) und dem Festgeldkonto2 (10.225,84 €) zusammen. Das Vereinsvermögen ist im Vergleich zum Vereinsjahr 2019 gestiegen. Am 31.12.2019 lag der Vermögensstand bei 50.106,98 €.

Die Mitgliederzahl blieb im Vereinsjahr 2020 relativ stabil. Am 05.02.2021 zählt der Verein 195 Mitglieder (02/20:197; 03/19: 198; 07/18: 201; 09/17: 206; 03/16: 202. Erfreulicherweise stieg die Zahl der ordentlichen Mitglieder leicht an. Diese lag am 05.02.21 bei 168 Mitgliedern (02/20: 166; 03/19: 167; 07/18: 169; 09/17: 175; 03/16: 174).

Der Einzug der Mitgliederbeiträge lief im Jahr 2020 erfolgreich ab. Wie bereits 2018 angekündigt, werden wir in Zukunft den Einzug nicht mehr am Ende des Kalenderjahres durchführen, sondern in den Herbst verlegen. Sollten Sie den Betrag selbst überweisen, bitten wir Sie dies ebenfalls bis zum 01. September des jeweiligen Kalenderjahres zu tun.

Die Kontodaten (bei der Postbank) der GfKl e.V. lauten:
Gesellschaft für Klassifikation e.V.
IBAN: DE31 5001 0060 0316 2686 08
BIC: PBNKDEFF

Gerne stellen wir Sie auch auf Lastschrift um. Das entsprechende Formular finden Sie unter http://www.gfkl.org/wp-content/uploads/2014/04/Einzugsermächtigung_DE.pdf.
Bitte senden Sie dieses ausgefüllt und unterschrieben an schatzmeister@gfkl.de.

Aufgrund einer Änderung im Einzugsverfahren können wir nun nur noch von deutschen Konten (IBAN beginnt mit DE) den Betrag direkt einziehen. Sollten Sie über ein ausländisches Konto verfügen, bitten wir Sie auf die Zahlungsart „Überweisung“ umzustellen und sich bei uns zu melden. Herzlichen Dank!

Und noch eine Bitte:
Leider verliert die Gesellschaft für Klassifikation e.V. Geld, wenn die Mitgliedsgebühr nicht eingezogen werden kann. Gemeinsam können wir diese Rückbuchungen verhindern: Sollten Sie Änderungen Ihres Kontos, postalische Adresse oder Ihrer Email-adresse haben wenden Sie sich bitte formlos per E-Mail an schatzmeister@gfkl.de.

Vielen Dank für Ihre Unterstützung!

Autor:	Andreas Geyer-Schulz
Datum:	09.02.2021

Bericht zu den Advances in Data Analysis and Classification (ADAC)

Mit vier Heften und 982 Seiten konnte der letztjährige Band 14 (2020) der Zeitschrift *Advances in Data Analysis and Classification (ADAC)* die quantitativen Höchstmarken der Vorjahre nicht erreichen, qualitativ muss sich dieser Jahrgang aber nicht hinter den Vorgängern verstecken. Die Bearbeitungszeit bis zu einer ersten Entscheidung ist angestiegen und beträgt im Mittel 58 Tage. Die Zeit bis zur finalen Akzeptanz eines Manuskriptes hat sich hingegen im Mittel auf 379 Tage reduziert. Danach sind die Manuskripte als „OnlineFirst" auf der Springer-Website gelistet und werden zügig in Papierform produziert. Typischerweise umfasst jedes Heft zehn Beiträge.

Das erste Heft des Bandes 14 (2020) enthält neun Artikel, die sich mit Regressionsbäumen, Clustering mit Constraints, Clustering von genomischen Wörtern, Clustering auf Basis von Hauptkurven, Ensemble von Bäumen, Clustering von Big-Data-Zeitreihen, SEM-Algorithmen, SVM-basierter Multiklassenklassifikation und Klassifikation mit sequenzieller Ordnungsstatistik beschäftigen.

Heft 2 ist die vierte Sonderausgabe der ADAC, die den neuesten Entwicklungen im Bereich des modellbasierten Clustering und der Klassifikation gewidmet ist, einem Gebiet, das sowohl in der Theorie als auch in der Anwendung zunehmend aktive Forschung bietet. Der Call for Papers für dieses Sonderheft führte zur Einreichung von 46 Manuskripten, von denen schließlich zehn nach dem üblichen Peer-Review-Verfahren zur Veröffentlichung angenommen wurden. Dieses Sonderheft, herausgegeben von den Gasteditoren Christophe biernacki, Luis angel García-Escudero und Salvatore Ingrassia, enthält Beiträge mit einem breiten Themenspektrum. Die ersten drei Papiere beleuchten das Gebiet der Mischungen von Regressionsmodellen Regressionen aus unterschiedlichen Perspektiven. Die drei folgenden Beiträge befassen sich mit Mischungsmodellen für robustes modellbasiertes Clustering, Datenmodellierung mit mehreren partiellen rechtsseitigen Zensierungspunkten bzw. Modellierung unter Messinkonsistenzen. Die nächsten beiden Artikel betrachten Mischungen von schiefen Verteilungen. Die beiden letzten Beiträge befassen sich mit Partikel-Monte-Carlo-Methoden und stochastischen Blockmodellen.

Die acht Artikel des dritten Heftes widmen sich den Themen Entscheidungsbäume, Clustering von Punktprozessen, Klassifikation mit gepaarten Kovariaten, einer Variante der kanonischen Analyse, topologische Klassifikation, Chi-Quadrat-Zerlegung, Rangtests für funktionale Daten und Gruppenvergleiche mit dem Heterogeneous-Choice-Modell.

Das vierte Heft von Band 14 ist wieder ein Sonderband, diesmal herausgegeben von Daniel Baier, Berthold Lausen, Angela Montanari und Ute Schmid, und umfasst 13 Artikel, die aus der ECDA-Tagung 2019 in Bayreuth resultieren und thematisch die ganze Bandbreite der Tagung abdecken: von theoretischen Betrachtungen zusammengesetzter Strukturgleichungsmodelle über Mischungen von Dirichlet-Multinomialverteilungen, Quantilregression, interpretierbaren Machine-Learning-Modellen, hinzu Variablenselektion und Fragen der Markerpositionierung basierend auf Werbebildanalysen bietet dieses Heft Einblicke in theoretische und angewandte Forschung der Datenanalyse und Klassifikation.

Bei den Publikationsmetriken divergiert die Entwicklung, der 5-Jahres-Impact-Faktor sank ab und liegt nun bei 1.888 (2019) während Google's h5-Index auf 25 angestiegen ist. Zudem erfahren die Publikation in ADAC große Beachtung durch zahlreiche Downloads.

Mehr zur Zeitschrift Advances in Data Analysis and Classification finden Sie unter www.gfkl.de oder http://www.springer.com/statistics/journal/11634.

Autor:	Adalbert Wilhelm

Bericht zu den Archives of Data Science

Archives of Data Science, Series A.
Unter http://www.archivesofdatascience.org/journals/series_a/publications?c=MBG21
können Sie zurzeit 111 Artikel abrufen. Weitere Artikel folgen in den nächsten Wochen.

Bereits abgeschlossen sind
*Vol. 1,1 (Band zur GPSDAA2013, 10 Artikel) und
*Vol. 2,1 und Vol. 2,2 (Bänder zur ECDA2015, 29 Artikel).

Teilweise erschienen sind
*Vol. 3,1 (reguläre Einreichungen bis 2019, zz. 3 Artikel),
*Vol. 4,1 (Band zur ECDA2017, zz. 22 Artikel),
*Vol. 5,1 (Band zur ECDA2018, zz. 28 Artikel) und
*Vol. 6,1 und Vol. 6,2 (Bänder zur ECDA2019 und GPSDAA2019, zz. 15 Artikel),
*Vol. 7,1 (reguläre Einreichungen ab 2020, zz. 4 Artikel).

Der Submissionprozess für die Bänder Vol.3,1-6,2 ist abgeschlossen. Weitere Artikel befinden sich noch im Copy-Editing und werden bald online verfügbar sein.
Für Vol. 7,1 nehmen wir gerne weiterhin Einreichungen entgegen.

Archives of Data Science, Series B (AoDSB)
Sie finden unter http://www.archivesofdatascience.org/journals/series_b/publications?c=MBG21B
die Artikel des Journals AoDSB.
Der Band Vol. 1,1 (Band zur IFCS17; 8 Artikel) ist abgeschlossen.
Der Submissionprozess für Vol. 2,1 (reguläre Einreichungen ab 2021) ist ab sofort eröffnet.

Archives of Data Science, Series A und Series B

Da alle unsere Beiträge in AoDSA und AoDSB bereits in der Online First Version mit einer eindeutigen DOI versehen sind und in den permanenten Speicher des Karlsruher Institut für Technologie (KIT) eingepflegt werden, sind diese ab dem ersten Erscheinen bereits voll zitierfähig. Als Extra werden wir in der gedruckten Form spezielle doppelte Seitenzahlen anbieten. Weitere Informationen und Zusatzmaterialen zu den jeweiligen Artikeln finden Sie unter dem Button „Details“ auf der Seite „Published Issues and Articles“. Auch im Jahr 2021 freuen wir uns über zahlreiche reguläre sowie special issue bezogene Einreichungen.
Die Autorenguidelines finden Sie unter
http://www.archivesofdatascience.org/journals/series_a/author-guidelines?c=MBGAu21 [AoDSA]
bzw. http://www.archivesofdatascience.org/journals/series_b/author-guidelines?c=MBGAu21 [AoDSB].
Alle weiteren Informationen entnehmen Sie bitte unserer Webseite.
Gerne können Sie uns auch bei Twitter folgen. Sie finden alle Neuigkeiten rund um das Journal unter: https://twitter.com/AoDataScience
Wir freuen uns über die stetig steigende Zahl an Einreichungen und Lesern. Aufgrund der vielen Einreichungen und unseren Bemühungen den Autoren hilfreiche Reviews zukommen zu lassen, benötigen wir eine große Anzahl an Reviewern. Sollten Sie sich vorstellen können uns als Reviewer unterstützen zu können, wären wir Ihnen sehr dankbar. Bitte melden Sie sich zu diesem Zweck formlos bei Victoria-Anne Schweigert (v.schweigert@kit.edu).
Vielen Dank an alle Leser, Autoren, Reviewer und Editoren für Ihre Unterstützung und Mithilfe!

Autor:	Andreas Geyer-Schulz

Bericht zu den Studies in Classification, Data Analysis, and Knowledge Organization

Die Studies in Classification, Data Analysis, and Knowledge Organization ist eine Buchreihe, die ständige und aktuelle Informationen zu den neuesten Entwicklungen und Methoden in den Bereichen statistische Datenanalyse, explorative Statistik, Klassifikation und Clustering, Umgang mit Informationen und Bestellung von Wissen veröffentlicht. In einer breiten Palette von theoretischen, methodischen sowie anwendungsorientierten Beiträgen, Studien und Diskussionen einer internationalen Autorenschaft werden Bereiche wie Computational Statistics, Mustererkennung, biologische Taxonomie, DNA- und Genomanalyse, Marketing, Finanzen und andere Bereiche der Wirtschaftswissenschaften, Datenbanken und das Internet umfasst. Ein Hauptziel ist es, das Zusammenspiel zwischen verschiedenen, scheinbar nicht verwandten Bereichen aufzuzeigen und die Zusammenarbeit zwischen Mathematikern, Statistikern, Informatikern, Natur-, Technik- und Wirtschaftswissenschaftlern sowie Praktikern zu fördern, indem fundierte und innovative Lösungen für drängende Probleme der Praxis angeboten werden.

2020 sind die folgenden drei Bände in der Buchreihe neu erschienen:

- Jajuga, Krzysztof, Batóg, Jacek, Walesiak, Marek (Eds.) (2020): Classification and Data Analysis Springer, 335 Seiten
- Imaizumi, Tadashi, Okada, Akinori, Miyamoto, Sadaaki, Yamamoto, Yoshiro und Vichi, Maurizio (Eds.) (2020): Advanced Studies in Classification and Data Science, Springer, 524 Seiten.
- Iezzi, Domenica Fioredistella, Mayaffre, Damon, Misuraca, Michelangelo (Eds.) (2020):Text Analytics, Springer, 302 Seiten.

Mehr zur Buchreihe finden Sie unter www.gfkl.de und www.springer.com/series/1564.

Autor:	Adalbert Wilhelm

Information zur ECDA 2021 vom 07. bis 09. Juli 2021 virtuell in Rotterdam

Nach den vorangegangenen Veranstaltungen in Luxemburg (2013), Bremen (2014), Colchester (2015), Wroclaw (2017), Paderborn (2018) und Bayreuth (2019) wird die European Conference on Data Analysis (ECDA) dieses Jahr – **zusammen mit der 44. Jahreskonferenz der GfKl** – vom 07. bis 09. Juli zusammen mit der DSSV als virtuelle Konferenz in Rotterdam stattfinden (siehe https://iasc-isi.org/dssv-ecda2021/).

Gemeinsame Ausrichter der DSSV-ECDA 2021 sind

- die International Association for Statistical Computing (IASC)
- die **Gesellschaft für Klassifikation (GfKl),**
- die **European Association for Data Science (EuADS),**
- die niederländische/flämische Klassifikationsgesellschaft (VOC).

Die DSSV-ECDA 2021 wird von der Erasmus Universität Rotterdam, Niederlande, ausgerichtet und als Satellitenkonferenz des World Statistics Congress der ISI durchgeführt.

Ziel der gemeinsamen Ausrichtung von Data Science, Statistics & Visualisation (DSSV) und der European Conference on Data Analysis (ECDA) ist es, Forscher:innen und Praktiker:innen, die sich für das Zusammenspiel von Statistik, Informatik und Visualisierung interessieren, zusammenzubringen und Brücken zwischen diesen Bereichen für interdisziplinäre Forschung zu bauen.

Als Keynote-Speaker konnten
Daniela Witten (U Washington), Jack van Wijk (Technische Universität Eindhoven), Michel Wedel (Universität von Maryland) und Peter Bühlmann (ETH Zürich) gewonnen werden.

Im Rahmen der Konferenz findet der LIS-Workshop zum Thema "Classification and Subject Indexing in Library and Information Science" (LIS 2021) der AG Bibliotheken und AG Dezimalklassifikation der GfKl statt sowie ein Workshop der AG Marketing.

Wichtige Termine für die DSSV-ECDA2021

- Einreichung von Abstracts: 15. Mai 2021
- Information über Vortragsannahme: 30. Mai 2021
- Stichtag für die Early Bird Registration: **05. Juni 2021 (23:59)**
- Konferenz: **07. - 09. Juli 2021**

Wir freuen uns auf Ihren virtuellen Besuch in Rotterdam! Weitere Informationen finden Sie auf der Veranstaltungswebsite https://iasc-isi.org/dssv-ecda2021/ und der Homepage der GfKl www.gfkl.org.

Autor:	Adalbert Wilhelm

Bericht der Arbeitsgruppen Bibliotheken (AG BIB) und Dezimalklassifikation (AG DK)

Ulrike Junger hat den Vorsitz der AGs von Frank Scholze übernommen. Die endgültige Integration der beiden AGs unter einem Namen wird weiterverfolgt. Die AGs haben beschlossen den LIS-Workshop zukünftig im zweijährigen Rhythmus nach Möglichkeit im Rahmen der ECDA auszurichten. Der nächste LIS-Workshop wird virtuell in Rotterdam als Teil der DSSV-ECDA 2021 Konferenz stattfinden.

Autor:	Ulrike Junger

Bericht der Arbeitsgruppe Biostatistik (AG BT)

Die Statistical Computing 2020 auf der Reisensburg entfiel corona-bedingt. Für 2021 liefen die Vorbereitungen für die Statistical Computing 2021 in Präsenz auf der Reisensburg und mussten mittlerweile aufgrund der Pandemiesituation wieder eingestellt werden. Es wird angedacht einen kleineren Workshop zusammen mit der AG Statistical Computing der GMDS/IBS-DR zu veranstalten.

Autor:	Ludwig Lausser und Hans Kestler

Bericht der Arbeitsgruppe Datenanalyse und Numerische Klassiifkation (AG DANK)

Aufgrund der aktuellen Situation wurde die Entscheidung getroffen, in 2020 keine Herbsttagung der AG DANK stattfinden zu lassen. Prof. Matthias Gehrke von der FOM in Frankfurt hat jedoch bereits signalisiert, auch in 2021 als Ausrichter zur Verfügung zu stehen, wofür ich mich schon jetzt einmal bedanken möchte.

Autor:	Gero Szepannek

Bericht der Arbeitsgruppe Ethik (AG Ethik) (in Gründung)

Coronabedingt kam der Aufbau dieser Arbeitsgruppe leider ins Stocken. Die für Mai 2020 geplante Auftaktveranstaltung in Hamburg musste leider abgesagt werden. Das allgemeine Ziel der Arbeitsgruppe ist es, Datenwissenschaftler zu inspirieren, sich an der Debatte über die Auswirkungen ihrer beruflichen Arbeit auf die Gesellschaft zu beteiligen und als Datenwissenschaftlerin in der öffentlichen Debatte über die digitale Welt aktiv zu werden.
Alle, die Interesse an dieser Thematik haben, sind aufgerufen sich unter ag-ethics-gfkl+subscribe@groups.io für die Mailingliste der AG anzumelden.

Autorin:	Ursula Garczarek

Bericht der Arbeitsgruppe Datenanalyse und Klassifikation im Marketing (AG MARKETING)

Die AG Marketing wächst weiter und wir freuen uns neue AG Mitglieder zu begrüßen.
Beim 2. Treffen der AG Marketing konnten wir wie beim letzten Mal spannenden Vorträgen lauschen und rege diskutieren. Aufgrund der angespannten Corona-Lage musste das 2. Treffen der AG Marketing in diesem Jahr virtuell stattfinden. Das tat jedoch weder der Qualität der Vorträge noch regen Diskussionen noch dem sozialen Austausch der Teilnehmer einen Abbruch.

Insgesamt 15 Teilnehmer nahmen am 17. und 18. August 2020, jeweils nachmittags, am 2. Treffen der AG Marketing teil. Die inhaltliche und technische Leitung übernahmen die AG Vorsitzenden Friederike Paetz (Technische Universität Clausthal) und Daniel Guhl (Humboldt-Universität zu Berlin).
Insgesamt 8 Beiträge von Vortragenden aus der Praxis und von internationalen Universitäten wurden präsentiert. Die Vorträge befassten sich thematisch mit unterschiedlichen Anwendungen von Discrete Choice Modellen (Künstliche Intelligenz, Umweltbewusstsein, Produktlinienoptimierung, Attribute-Non-Attendance), Referenz- preisen (interne und externe), Reabilitäts- sowie Validitätsanalysen (Best-Worst Scaling, Vorhersagen in Choice-Based Conjoint Modellen) und Content Marketing (review contents).

Vortragende waren Benedikt Martin Brand (Universität Bayreuth), Ossama Elshiewy (Universität Göttingen), Peter Kurz (bms marketing research + strategie), Marcel Lichters (Otto-von-Guericke-Universität Magdeburg), Friederike Paetz (Technische Universität Clausthal), Vlada Pleshcheva (Statis Information GmbH), Nadine Schröder (WU Wien) und Narine Yegorian (Humboldt-Universität zu Berlin).

Ein Kurzbeitrag zum 2. Treffen der AG Marketing wird demnächst in den Archives of Data Science, Series A erscheinen.

Das nächste Treffen der AG Marketing wird voraussichtlich im Rahmen der DSSV-ECDA 2021 vom 7. bis 9. Juli 2021 (virtuell) stattfinden. Wir freuen uns sehr, dass wir über die Kontakte der AG Marketing den renommierten quantitativen Marketingwissenschaftler *Michel Wedel* als Keynote-Speaker für die DSSV-ECDA 2021 gewinnen konnten!

An dieser Stelle möchten die AG Vorsitzenden auch noch einmal auf einen gemeinsamen CfP für eine Special Issue im OR Spectrum zusammen mit Udo Wagner und Michel Wedel aufmerksam machen (Deadline: 31.12.2021).

Autorin:	Friederike Paetz

Springer

springer.com

CALL FOR PAPERS

OR Spectrum
Quantitative Approaches in Management

http://www.springer.com/291

Special Issue on
Predicting and Optimizing Marketing Performance in Dynamic Markets

Special Issue Editors:
Friederike Paetz, Daniel Guhl, Udo Wagner, and Michel Wedel

Submission deadline: December 31th, 2021

Our world is turbulent: ecological, social, economic, and business environments change continuously and sometimes rapidly. Consumers have changing preferences, learn from experiences and interactions with other consumers, build (or lose) trust in brands, and adopt new technologies. Firms innovate and introduce new products, grow, fail, merge with other firms, and respond to competition. Exogenous events such as dramatic changes in economic conditions (e.g., stock market crash of 2008), regulations (e.g., GDPR in Europe, US-Chinese import tariffs), man-made crises (e.g., VW diesel emissions scandal), as well as natural disasters (e.g., Hurricane Katrina in the US or the COVID19 pandemic), cause major shifts in markets. Market turbulence has not only immediate effects on manufacturing, distribution, product portfolios, pricing, competition, and consumer behavior but has long-term strategic consequences for firms having to anticipate and address the uncertainty caused by market dynamics. To address dynamics and shifts in markets, firms need to be able to analyze and predict changes and their impact on market behaviors and reallocate resources accordingly. To this end modern technologies (e.g., Internet of Things, mobile and GPS technologies, virtual and other digital environments) enable firms to faster identify market shifts and better track and capitalize on dynamics.

The special issue focuses on novel marketing data and modern methodologies from OR, statistics, econometrics, and/or computer science that help companies respond to market dynamics and market shifts more efficiently. In particular, we encourage new research incorporating big, high velocity, or unstructured data, as well as further development of models and methods that account for dynamics (e.g., dynamic linear models, vector autoregression, or hidden Markov models) and market shifts (e.g., regime switching models), uncertainty (e.g., Bayesian learning, dynamic discrete choice models, dynamic programming), as well as heterogeneity (e.g., Bayesian hierarchical modeling). We encourage research that involves large scale applications that have direct practical relevance in business practice in industries such as retailing, transportation, inventory management, entertainment and tourism. Boundary spanning applications of machine learning combined with more classical approaches from marketing and OR, are of great interest to the special issue. The specific topics of interest include but are not limited to

- Dynamic multimedia budget allocation
- Product-harm crisis
- Consumer learning and behavior shifts
- Dynamic pricing
- Brand marketing policy shifts
- Dynamic customer relationship management
- Response dynamics in direct marketing
- Brand and corporate equity development
- Marketing and dynamic competition
- Advertising new products or events (e.g., movies)
- Dynamics of viral marketing
- Demand under dynamic uncertainty
- Real-time recommendations
- New product innovation and introduction

Submission Guidelines and Review Process: Papers must be submitted at http://www.editorialmanager.com/orsp/ under the category "S.I. Predicting and Optimizing Marketing Performance in Dynamic Markets" by December 31, 2021. All papers submitted to this special issue should report original work and make a contribution to the journal OR Spectrum by using a quantitative research paradigm and OR methodologies. According to the aims of OR Spectrum, high quality papers are wanted that match the scope of the journal, show rigor in applying state-of-the-art OR techniques, and promise to have an impact on future work of the OR community.

Papers will be screened by the Editor-in-Chief and one special issue editor. If the paper is deemed to be of sufficient quality, it will be reviewed according to the standards of OR Spectrum by at least two experienced reviewers. We will adopt a rapid and fair review process striving to provide reviews within three months of submission. Accepted papers will be available online prior to the publication of the special issue.

Special Issue Editors

Friederike Paetz
TU Clausthal,
Germany
friederike.paetz@tu-clausthal.de

Daniel Guhl
HU Berlin,
Germany
daniel.guhl@hu-berlin.de

Udo Wagner
University of Vienna,
Austria
udo.wagner@univie.ac.at

Michel Wedel
University of Maryland,
United States
mwedel@umd.edu

Personalia

Am 12.05.2020 verstarb Prof. Dr. Erhard Godehardt, (geb. 17.07.1949) ehemaliger Leiter der EDV und Statistik an der Klinik für Kardiovaskuläre Chirurgie der Heinrich-Heine Universität Düsseldorf. Prof. Godehardt war der GfKl über Jahrzehnte verbunden und regelmäßig auf den Jahrestagungen präsent. Als Mathematiker und Statistiker war er an vielen Forschungsarbeiten beteiligt und hatte sich bis zuletzt wissenschaftlichen Fragestellungen, etwa der Untersuchung von Clustern in komplexen Netzwerken, gewidmet. Weitere Informationen zum Leben udn Wirken von Prof. Godehardt finden Sie hier.

Nach langer schwerer Krankheit verstarb am 08. November 2020 Prof. Dr. Rüdiger Klar, geb. am 30.07.1942, ehemaliger Leiter der Abteilung Medizinische Informatik an der Albert-Ludwigs Universität Freiburg. Als langjähriges Mitglied der GfKl hat Prof. Klar 1996 die Jahrestagung in Freiburg ausgerichtet, war lange Zeit Beirat und Mitglied des GfKl-Vorstandes und ein starkes Bindeglied zwischen GfKl und GMDS. Ein ausführlicher Nachruf auf Prof. Klar findet sich hier auf den Seiten der GMDS

Den Hinterbliebenen der Verstorbenen gilt unser tiefes Mitgefühl.

Autor:	Adalbert Wilhelm