Hinweis: Aufgrund der neuen Längenvorgaben der DFG für Anträge sowie der substanziell detaillierteren Beschreibung des Kapitels 2.3 (Arbeitsprogramm) wurden die Tabellen 1-3 (sowie zugehörige Literaturquellen) ausgelagert. Damit besteht weiterhin die Möglichkeit den Stand der Literatur und die Forschungslücke im Detail nachzuvollziehen (diese Lücke wurde bereits in der ersten Begutachtungsrunde seitens der Gutachten bestätigt).

Tabelle 1. Ausgewählte Ansätze zur DQ-Messung und -Verbesserung bei strukturierten Daten

DQ-Dimension	Literatur (Auszug)	Beschreibung
Korrektheit/ Aktualität	Fisher et al. 2009	Bestimmung der Korrektheit in Datenbanken unter Verwendung von Zufallsmaßen und Wahrscheinlichkeitsverteilungen
	Heinrich und Klier 2015, Heinrich und Hristova 2016, Zak und Even 2017	Ansätze (z. B. probabilistische Metriken basierend auf Wahrscheinlichkeitsverteilungen bzw. stochastischen Prozessen) zur Messung und Verbesserung der Aktualität von Daten
Vollständigkeit	Naumann et al. 2004	Messung der Vollständigkeit von Datenquellen (im Internet) basierend auf der Anzahl fehlender Attributwerte und der Anzahl Referenzen zu anderen Datenquellen
	Heinrich et al. 2008, Blake und Mangiameli 2011	Messung und Verbesserung der Vollständigkeit von Datenbanken anhand der relativen Häufigkeit des Auftretens von NULL-Werten
Konsistenz	Alpar und Winkelsträter 2014	Identifikation und Bereinigung von Inkonsistenzen unter Verwendung von Assoziationsregeln
	Heinrich et al. 2018c	Ansatz zur Messung und Verbesserung der Konsistenz auf der Grundlage probabilistischer Regeln
	Boeckling et al. 2019	Identifikation und Bereinigung von Inkonsistenzen mittels Regeln basierend auf einer verallgemeinerten Lift-Kennzahl
Identität	Weis und Naumann 2005	Hierarchisches Framework zur Identifikation von Duplikaten bei XML Dokumenten
	Heinrich et al. 2018b, Obermeier 2019	Methoden zur wahrscheinlichkeitsbasierten Identifikation und Bereinigung von Duplikaten (basierend auf Realweltereignissen)
	Draisbach et al. 2019	Algorithmische Identifikation von Duplikaten basierend auf Graphstrukturen

Tabelle 2. Ansätze zur direkten DQ-Messung und -Verbesserung bei textuellen NGI

DQ-Dimension	Literatur	Beschreibung	Limitationen
Korrektheit/ Aktualität	Chen und Tseng 2011	Messung der Aktualität von Kundenrezensionen basierend auf dem Vergleich von Termfrequenzen von alten vs. neuen Rezensionen	- Keine Berücksichtigung der Textsemantik - Eingeschränkte Interpretierbarkeit der Messergebnisse
	Li et al. 2017	Bewertung der Korrektheit von (Teil-)Sätzen mithilfe von Naive Bayes-Verfahren und Hidden Markov Models	- Keine Berücksichtigung der Textsemantik - Lediglich Einteilung in die Klassen „sicher“ und „unsicher“ - Eingeschränkte Interpretierbarkeit der Ergebnisse
	Thorne und Vlachos 2018	Bestimmung der Korrektheit durch die Analyse von linguistischen Features und Knowledge Graphen	- Ansätze eingeschränkt auf binäre Klassifikation - Eingeschränkte Interpretierbarkeit der Ergebnisse
	Han et al. 2020	Bestimmung der Korrektheit von Sätzen mithilfe von Neuronalen Netzen mit einem Attention-Mechanismus	- Ansätze eingeschränkt auf binäre Klassifikation - Eingeschränkte Interpretierbarkeit der Ergebnisse
Vollständigkeit	Chen und Tseng 2011	Messung und Verbesserung der Vollständigkeit von Kundenrezensionen durch Auswertung enthaltener Aspekte und Abgleich mit vorher definiertem Schema	- Keine Berücksichtigung der Textsemantik - Definiertes Schema beeinflusst maßgeblich die Messung der DQ
Konsistenz	keine Ansätze bekannt / vorhanden		
Identität			

Tabelle 3. Ansätze zur Diskussion und Berücksichtigung von DQ bei maschinellen Lernverfahren

DQ-Dim.	Literatur	Beschreibung	Limitationen			
			L1	L2	L3	Weitere spezifische Limitationen
Korrektheit/ Aktualität	Quinlan 1986	Anpassung von Entscheidungsbaumverfahren zur Vermeidung einer Überanpassung aufgrund von Ausreißern mit inkorrekten Datenwerten	X	X		- Ansatz nur bei speziellen DQ-Defekten valide
	Tresp et al. 1994, Wright 1999, Svensson 2002	Modellierung der Abweichungen der Inputdatenwerte von den tatsächlichen Datenwerten durch eine bedingte Wahrscheinlichkeitsverteilungsfunktion	X	X		- Bedingte Verteilungsfunktionen müssen bekannt sein - Ansätze sind nur für spezielle Daten/DQ-Defekte valide
	Pierce et al. 2006	Nutzung der Information Gap Theory, um Abweichungen der Inputdatenwerte von den tatsächlichen Werten zu modellieren und deren Effekte auf die Prognoseergebnisse zu quantifizieren	X	X		
	Blake und Mangiameli 2011	Analyse der Effekte einer höheren Korrektheit/Aktualität auf ausgewählte Lernverfahren im Decision Support-Bereich	X	X	X	
	Hristova 2014	Ansatz zum Einbezug Aktualitäts-annotierter Inputdaten bei Entscheidungsbaumverfahren (zur Klassifikation)	X			- DQ-annotierte Inputdaten werden nicht für die Erstellung der Bäume, sondern nur bei der Nutzung genutzt
	Heinrich und Hristova 2016	Ansatz zur Modellierung des Einflusses von Aktualität auf Entscheidungen und Entscheidungsgüte basierend auf Konzepten des Informationswerts	X		X	- Fokus auf die Effekte einer verbesserten DQ hinsichtlich Entscheidungsgüte
	Dallachiesa et al. 2019	Techniken, um potenziell fehlerhafte Teilbereiche von Netzwerken für die Klassifikation von Netzwerkknoten zu berücksichtigen	X	X		- Spezieller Fokus auf die Klassifikation von Knoten in einem Netzwerk
Vollständigkeit	Quinlan 1986, Nowicki et al. 2018, Śmieja et al. 2018, Che et al. 2018, Elaidi et al. 2018	Vorschläge, unvollständige Inputdaten durch Erwartungswerte, geschätzte Auftrittsmuster, Approximationen (Rough Set Theory) o. ä. zu adressieren	X	X		- Ansätze nur unter restriktiven Annahmen valide
	Tresp et al. 1994, Williams et al. 2007	Nutzung eines parametrischen Schätzverfahrens, um Klassenwahrscheinlichkeiten für unvollständige Daten zu bestimmen	X	X		- Ansätze nur für wenige Anwendungsfälle valide (z. B. Annäherung der Integrale sehr rechenaufwendig)
	Blake und Mangiameli 2011	Analyse der Effekte einer höheren Vollständigkeit auf ausgewählte Lernverfahren im Decision Support-Bereich	X	X	X	
	Struski et al. 2016	Vorschlag einer Trainingsstrategie für Klassifikationsverfahren, welche die Trainingsdaten in Segmente unterteilt, die keine fehlenden Werte enthalten	X	X		- Ansatz nur bei bestimmter Verteilung und Anzahl fehlender Werte valide
	Struski et al. 2017	Verwendung von Vektoren mit der Dichtefunktion der Normalverteilung zur Repräsentation von unvollständigen Daten für Regressions-basierte SVM	X	X		- Ansatz ist nur für Regressions-basierte SVM anwendbar
	Feldman et al. 2018	Vorschlag eines analytischen Rahmenwerks zur Untersuchung des Einflusses unvollständiger Daten auf binäre Klassifikationen	X	X	X	
	Binder et al. 2019	Analyse zu Effekten der Vollständigkeit von Aspekt-basierten Sentiments (in textuellen Rezensionen) hinsichtlich der Erklärung der Gesamtbewertungen		X	X	
	Heinrich et al. 2019, Heinrich et al. 2020	Analyse der Effekte einer höheren Vollständigkeit auf die Precision Accuracy eines Matrixfaktorisierungs-Empfehlungssystems	X	X	X	
	Taguchi et al. 2021	Nutzung von unabhängigen Gauß-Verteilungen zur Verarbeitung von fehlenden Werten in Graphen als Input für Graph Convolutional Networks (GCN)	X			- Ansatz ist nur für GCN und fehlende Werte mit unabhängigen Gauß-Verteilungen anwendbar
Konsistenz	Blake und Mangiameli 2011	Analyse der Effekte einer höheren Konsistenz auf ausgewählte Lernverfahren im Decision Support-Bereich	X	X	X	
	Favieiro und Balbinot 2019	Verwendung von Parakonsistenz-Logic für die Klassifikation mit Entscheidungsbäumen basierend auf inkonsistenten Daten	X	X		
Identität	keine Ansätze bekannt / vorhanden					

Legende der Limitationen L1 bis L3: L1: Keine Betrachtung von textuellen NGI; L2: Keine Berücksichtigung von DQ-annotierten Inputdaten; L3: Ausschließliche Betrachtung des Effekts von (verbesserter) DQ auf die Ergebnisgüte maschineller Lernverfahren, d. h. kein Ansatz zur Berücksichtigung von DQ-annotierten Inputdaten bei maschinellen Lernverfahren

Hinweis: Für den Antrag abzugrenzen und demnach nicht in Tabelle 3 sind Simulationsansätze zur Berücksichtigung von DQ in maschinellen Lernverfahren (vgl. z. B. (Tachioka und Watanabe) 2015). Diese sind auf DQ-Defekte bei textuellen NGI nicht übertragbar: Hier wäre eine extrem hohe Anzahl an Simulationsläufen erforderlich, v. a. wenn auf Wort-, Aussagen- oder Satzbasis analysiert wird. Ein solches Vorgehen ist hinsichtlich Berechenbarkeit nicht erfolgsversprechend, gerade vor dem Hintergrund, dass bei modernen maschinellen Lernverfahren oft schon wenige Durchläufe mit enormem Rechen- und Speicheraufwand verbunden sind. Auch ist das Forschungsfeld „Uncertainty Quantification“ abzugrenzen, bei dem allgemein versucht wird, die (Un-)Sicherheit einer Schlussfolgerung aus Daten zu bemessen ((Sullivan) 2015). Hierbei wird die Unsicherheit als unveränderbar angesehen (Aleatoric Uncertainty) bzw. als Modell-inhärent (Epistemic Uncertainty). Es wird nicht betrachtet, wie Unsicherheiten durch DQ-Defekte in den Inputdaten entstehen, wie die Daten mit DQ-(Metrik-)Werten annotiert werden können und ob man diese reduzieren bzw. eliminieren kann.

3 Literaturverzeichnis

Acharya, S.; Fung, G. (2020): Using Optimal Embeddings to Learn New Intents with Few Examples: An Application in the Insurance Domain. In: CEUR Workshop Proceedings.

Alpar, P.; Winkelsträter, S. (2014): Assessment of data quality in accounting data with association rules. In: Expert Systems with Applications 41 (5), S. 2259–2268.

Alt, R.; Reinhold, O. (2020): Social CRM: Challenges and Perspectives. In: Social Customer Relationship Management, Bd. 4. Cham: Springer (Management for Professionals), S. 81–102.

Barbado, R.; Araque, O.; Iglesias, C.A. (2019): A framework for fake review detection in online consumer electronics retailers. In: Information Processing & Management 56 (4), S. 1234–1244.

Batini, C.; Barone, D.; Cabitza, F.; Grega, S. (2011): A data quality methodology for heterogeneous data. In: International Journal of Database Management Systems 3 (1), S. 60–79.

Batini, C.; Scannapieco, M. (2016): Data and Information Quality. Dimensions, Principles and Techniques: Springer.

Beduè, P.; Graef, R.; Klier, M.; Zolitschka, J.F. (2018): A Novel Hybrid Knowledge Retrieval Approach for Online Customer Service Platforms. In: Proc. of 26th ECIS.

Bertoldi, N.; Cettolo, M.; Federico, M. (2010): Statistical Machine Translation of Texts with Misspelled Words. In: Human Language Technologies: Annual Conference of North American Chapter of the ACL: ACL, S. 412–419.

Bertrand, J.W.M.; Fransoo, J.C. (2002): Operations management research methodologies using quantitative modeling. In: International Journal of Operations & Production Management 22 (2), S. 241–264.

Binder, M.; Heinrich, B.; Klier, M.; Obermeier, A.; Schiller, A.P.R. (2019): Explaining the stars: Aspect-based sentiment analysis of online customer reviews. In: Proceedings of the 27th European Conference on Information Systems.

Blake, R.; Mangiameli, P. (2011): The Effects and Interactions of Data Quality and Problem Complexity on Classification. In: Journal of Data and Information Quality 2 (2), S. 1–28.

Boeckling, T.; Bronselaer, A.; Tré, G. de (2019): Mining data quality rules based on T-dependence. In: Proc. of 11th Conf. of the European Society for Fuzzy Logic and Technology, Bd. 1: Atlantis, S. 184–191.

Bundesministerium der Justiz und für Verbraucherschutz (2021): Bundesgesetzblatt 2021 Teil I Nr. 19.

Cardoso, E.F.; Silva, R.M.; Almeida, T.A. (2018): Towards automatic filtering of fake reviews. In: Neurocomputing 309, S. 106–116.

Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. (2018): Recurrent Neural Networks for Multivariate Time Series with Missing Values. In: Scientific reports 8 (1).

Chen, C.C.; Tseng, Y.-D. (2011): Quality evaluation of product reviews using an information quality framework. In: Decision Support Systems 50 (4), S. 755–768.

Chen, P.; Sun, Z.; Bing, L.; Yang, W. (2017): Recurrent attention network on memory for aspect sentiment analysis. In: Proc. of 22nd Conf. on Empirical Methods in Natural Language Processing, S. 452–461.

Dallachiesa, M.; Aggarwal, C.C.; Palpanas, T. (2019): Improving Classification Quality in Uncertain Graphs. In: Journal of Data and Information Quality 11 (1), S. 1–20.

Das, S.; Datta, S.; Chaudhuri, B.B. (2018): Handling data irregularities in classification: Foundations, trends, and future challenges. In: Pattern Recognition 81, S. 674–693.

Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. (2018): Bert: Pre-training of deep bidirectional transformers for language understanding. In: arXiv preprint arXiv:1810.04805.

Dey, L.; Haque, S.K.M. (2009): Studying the effects of noisy text on text mining applications. In: Proc. of the 3rd Workshop on Analytics for Noisy Unstructured Text Data. New York City, USA: ACM Press, S. 107–114.

Draisbach, U.; Christen, P.; Naumann, F. (2019): Transforming Pairwise Duplicates to Entity Clusters for High-quality Duplicate Detection. In: Journal of Data and Information Quality 12 (1).

Elaidi, H.; Benabbou, Z.; Abbar, H. (2018): Using Game Theory to Handle Missing Data at Prediction Time of ID3 and C4. 5 Algorithms. In: International Journal of Advanced Computer Science and Applications 9 (12), S. 218–224.

Fan, W. (2015): Data quality: from theory to practice. In: ACM SIGMOD Record 44 (3), S. 7–18.

Favieiro, G.W.; Balbinot, A. (2019): Paraconsistent Random Forest: An Alternative Approach for Dealing With Uncertain Data. In: IEEE Access 7, S. 147914–147927.

Feldman, M.; Even, A.; Parmet, Y. (2018): A methodology for quantifying the effect of missing data on decision quality in classification problems. In: Communications in Statistics - Theory and Methods 47 (11), S. 2643–2663.

Fisher, C.W.; Lauria, E.J.M.; Matheus, C.C. (2009): An Accuracy Metric. In: Journal of Data and Information Quality 1 (3), S. 1–21.

García Lozano, M.; Brynielsson, J.; Franke, U.; Rosell, M.; Tjörnhammar, E.; Varga, S.; Vlassov, V. (2020): Veracity assessment of online data. In: Decision Support Systems 129.

Gast, J.; Roth, S. (2018): Lightweight Probabilistic Deep Networks. In: Proc. of IEEE Conf. on CVPR.

Graef, R.; Klier, M.; Kluge, K.; Zolitschka, J.F. (2020): Human-machine collaboration in online customer service – a long-term feedback-based approach. In: Electronic Markets.

Han, X.; Li, B.; Wang, Z. (2020): An attention-based neural framework for uncertainty identification on social media texts. In: Tinshhua Sci. Technol. 25 (1), S. 117–126.

Hefny, A.H.; Dafoulas, G.A.; Ismail, M.A. (2020): Intent Classification for a Management Conversational Assistant. In: Proc. of 15th Intern. Conf. on Computer Engineering and Systems (ICCES): IEEE, S. 1–6.

Heinrich, B.; Hopf, M.; Lohninger, D.; Schiller, A.; Szubartowicz, M. (2019): Data quality in recommender systems: the impact of completeness of item content data on prediction accuracy of recommender systems. In: Electron Markets 23 (2–3), S. 169.

Heinrich, B.; Hopf, M.; Lohninger, D.; Schiller, A.; Szubartowicz, M. (2020): Something's Missing? A Procedure for Extending Item Content Data Sets in the Context of Recommender Systems. In: Inf Syst Front.

Heinrich, B.; Hristova, D. (2016): A quantitative approach for modelling the influence of currency of information on decision-making under uncertainty. In: Journal of Decision Systems 25 (1), S. 16–41.

Heinrich, B.; Hristova, D.; Klier, M.; Schiller, A.; Szubartowicz, M. (2018a): Requirements for Data Quality Metrics. In: Journal of Data and Information Quality 9 (2), S. 1–32.

Heinrich, B.; Kaiser, M.; Klier, M. (2007): How to measure Data Quality? A Metric-based Approach. In: Proc. of 28th ICIS.

Heinrich, B.; Kaiser, M.; Klier, M. (2008): Does the EU Insurance Mediation Directive Help to Improve Data Quality? - A Metric-based Analysis. In: Proc. of 16th ECIS.

Heinrich, B.; Klier, M. (2006): Ein Optimierungsansatz für ein fortlaufendes Datenqualitätsmanagement und seine praktische Anwendung bei Kundenkampagnen. In: Zeitschrift für Betriebswirtschaft 76 (6), S. 559–587.

Heinrich, B.; Klier, M. (2015): Metric-based data quality assessment — Developing and evaluating a probability-based currency metric. In: Decision Support Systems 72, S. 82–96.

Heinrich, B.; Klier, M.; Obermeier, A.; Schiller, A. (2018b): Event-driven duplicate detection: a probability-based approach. In: Proceedings of the 26th European Conference on Information Systems.

Heinrich, B.; Klier, M.; Schiller, A.; Wagner, G. (2018c): Assessing data quality – A probability-based metric for semantic consistency. In: Decision Support Systems 110, S. 95–106.

Helversen, B. von; Abramczuk, K.; Kopeć, W.; Nielek, R. (2018): Influence of consumer reviews on online purchasing decisions in older and younger adults. In: Decision Support Systems 113, S. 1–10.

Hristova, D. (2014): Considering Currency in Decision Trees in the Context of Big Data. In: Proceedings of the 22nd International Conference on Information Systems.

Hu, D. (2020): An Introductory Survey on Attention Mechanisms in NLP Problems. In: Intelligent Systems and Applications, Bd. 1038. Cham: Springer (Advances in Intelligent Systems and Computing), S. 432–448.

Jones, Z.; Linder, F. (2015): Exploratory data analysis using random forests. In: 73rd annual MPSA conference.

Kassner, N.; Schütze, H. (2020): Negated and Misprimed Probes for Pretrained Language Models: Birds Can Talk, But Cannot Fly. In: Proc. of 58th Annual Meeting of ACL: Association for Computational Linguistics, S. 7811–7818.

Krishnan, S.; Franklin, M.J.; Goldberg, K.; Wu, E.: BoostClean: Automated Error Detection and Repair for Machine Learning. In: arXiv preprint arXiv:1711.01299.

Kumar, S.; West, R.; Leskovec, J. (2016): Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In: Proc. of the 25th international conference on World Wide Web, S. 591–602.

Li, X.; Bing, L.; Li, P.; Lam, W. (2019): A Unified Model for Opinion Target Extraction and Target Sentiment Prediction. In: AAAI 33, S. 6714–6721.

Li, X.; Gao, W.; Shavlik, J.W. (2017): Detecting Semantic Uncertainty by Learning Hedge Cues in Sentences Using an HMM. In: Social media content analysis: Natural language processing and beyond: World Scientific Publishing Co., Inc., S. 89–107.

Liu, W.; Zhou, P.; Zhao, Z.; Wang, Z.; Ju, Q.; Deng, H.; Wang, P. (2020): K-bert: Enabling language representation with knowledge graph. In: Proc. of the AAAI Conference on Artificial Intelligence, Bd. 34, S. 2901–2908.

Louppe, G. (2014): Understanding random forests: From theory to practice. In: arXiv preprint arXiv:1407.7502.

Lukyanenko, R.; Parsons, J. (2015): Information Quality Research Challenge: Adapting Information Quality Principles to User-Generated Content. In: Journal of Data and Information Quality 6 (1), S. 1–3.

Madlberger, M. (2011): Can data quality help overcome the penguin effect? The case of item master data pools. In: Proc. of 19th ECIS.

Manning, C.D. (2011): Part-of-speech tagging from 97% to 100%: is it time for some linguistics? In: International conference on intelligent text processing and computational linguistics. Springer, S. 171–189.

Meredith, J.R.; Raturi, A.; Amoako-Gyampah, K.; Kaplan, B. (1989): Alternative research paradigms in operations. In: Journal of Operations Management 8 (4), S. 297–326.

Müller, O.; Junglas, I.; Debortoli, S.; Vom Brocke, J. (2016): Using text analytics to derive customer service management benefits from unstructured data. In: MIS Quarterly Executive 15 (4), S. 243–258.

Naumann, F.; Freytag, J.-C.; Leser, U. (2004): Completeness of integrated information sources. In: Information Systems 29 (7), S. 583–615.

Nolte, F.; Guhr, N.; Breitner, M.H.; Badtke, L.; Göing, K. (2019): Enterprise Social Media Moderation and User Generated Content Qualilty: a Critical Discussion and New Insights. In: Proc. of 27th ECIS.

Nowicki, R.K.; Korytkowski, M.; Scherer, R. (2018): Rough Neural Network Ensemble for Interval Data Classification. In: 2018 IEEE International Conference on Fuzzy Systems: IEEE, S. 1–7.

Obermeier, A. (2019): Anomaly-Based Duplicate Detection: A Probabilistic Approach. In: 14th Intern. Conf. on Design Science Research in Information Systems and Technology, 221–236 (Publ. wurde an der Professur Klier erstellt).

Orr, K. (1998): Data quality and systems theory. In: Communications of the ACM 41 (2), S. 66–71.

Ostendorff, M.; Bourgonje, P.; Berger, M.; Moreno-Schneider, J.; Rehm, G.; Gipp, B. (2019): Enriching bert with knowledge graph embeddings for document classification. In: arXiv preprint arXiv:1909.08402.

Otto, B.; Legner, C. (2016): Master Data erfolgreich managen. In: Controlling & Management Review 60 (3), S. 6–17.

Parssian, A.; Sarkar, S.; Jacob, V.S. (2004): Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. In: Management Science 50 (7), S. 967–982.

Pentek, T.; Legner, C.; Otto, B. (2017): Towards a reference model for data management in the digital economy. In: 12th Intern. Conf. on Design Science Research in Information Systems and Technology, S. 73–82.

Pierce, S.G.; Ben-Haim, Y.; Worden, K.; Manson, G. (2006): Evaluation of neural network robust reliability using information-gap theory. In: IEEE transactions on neural networks 17 (6), S. 1349–1361.

Quinlan, J.R. (1986): Induction of decision trees. In: Machine Learning 1 (1), S. 81–106.

Ro, Y.; Lee, Y.; Kang, P. (2020): Multi^ 2OIE: Multilingual Open Information Extraction based on Multi-Head Attention with BERT. In: arXiv preprint arXiv:2009.08128.

Śmieja, M.; Struski, L.; Tabor, J.; Zieliński, B.; Spurek, P. (2018): Processing of missing data by neural networks. In: Advances in Neural Information Processing Systems 31: Curran Associates, Inc, S. 2719–2729.

Stieglitz, S.; Dang-Xuan, L.; Bruns, A.; Neuberger, C. (2014): Social Media Analytics. In: Business & Information Systems Engineering 6 (2), S. 89–96.

Struski, L.; Śmieja, M.; Tabor, J. (2016): Incomplete data representation for SVM classification. In: Computing Research Repository.

Struski, Ł.; Śmieja, M.; Zieliński, B.; Tabor, J. (2017): Regression SVM for Incomplete Data. In: SI (26).

Sullivan, T. J. (2015): Introduction to uncertainty quantification. Cham: Springer (63).

Svensson, J. (2002): An MLP training algorithm taking into account known errors on inputs and outputs. In: International journal of neural systems 12 (5), S. 369–379.

Tachioka, Y.; Watanabe, S. (2015): Uncertainty training and decoding methods of deep neural networks based on stochastic representation of enhanced features. In: Nuclear Physics A 2015-January, S. 3541–3545.

Taguchi, H.; Liu, X.; Murata, T. (2021): Graph convolutional networks for graphs containing missing features. In: Future Generation Computer Systems 117, S. 155–168.

Thorne, J.; Vlachos, A. (2018): Automated Fact Checking: Task Formulations, Methods and Future Directions. In: Proc. of 27th Intern. Conf. on Computational Linguistics. Santa Fe, New Mexico, USA: ACL, S. 3346–3359.

Tilly, R.; Posegga, O.; Fischbach, K.; Schoder, D. (2017): Towards a Conceptualization of Data and Information Quality in Social Information Systems. In: Business & Information Systems Engineering 59 (1), S. 3–21.

Tresp, V.; Ahmad, S.; Neuneier, R. (1994): Training neural networks with deficient data. In: Advances in Neural Information Processing, S. 128–135.

Tsang, S.; Kao, B.; Yip, K.Y.; Ho, W.-S.; Lee, S.D. (2011): Decision Trees for Uncertain Data. In: IEEE Trans. Knowl. Data Eng. 23 (1), S. 64–78.

Tunkiel, A.T.; Sui, D.; Wiktorski, T. (2020): Data-driven sensitivity analysis of complex machine learning models: A case study of directional drilling. In: Journal of Petroleum Science and Engineering 195, S. 107630.

Wand, Y.; Wang, R.Y. (1996): Anchoring data quality dimensions in ontological foundations. In: Communications of the ACM 39 (11), S. 86–95.

Wang, R.Y.; Strong, D.M. (1996): Beyond accuracy: What data quality means to data consumers. In: JMIS, S. 5–33.

Wechsler, A.; Even, A. (2012): Using a Markov-Chain model for assessing accuracy degradation and developing data maintenance policies. In: Proc. of 18th Americas Conf. on Inf. Sys.

Weis, M.; Naumann, F. (2005): DogmatiX tracks down duplicates in XML. In: Proc. of 2005 ACM SIGMOD: ACM.

Williams, D.; Liao, X.; Xue, Y.; Carin, L.; Krishnapuram, B. (2007): On classification with incomplete data. In: IEEE transactions on pattern analysis and machine intelligence 29 (3), S. 427–436.

Wright, W.A. (1999): Bayesian approach to neural-network modeling with input uncertainty. In: IEEE transactions on neural networks 10 (6), S. 1261–1270.

Xu, H.; Liu, B.; Shu, L.; Yu, P. (2019): BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In: North American Chapter of ACL: Human Language Technologies: ACL (1), S. 2324–2335.

Zak, Y.; Even, A. (2017): Development and evaluation of a continuous-time Markov chain model for detecting and handling data currency declines. In: Decision Support Systems 103, S. 82–93.

Zhang, S.; Zhang, X.; Chan, J.; Rosso, P. (2019): Irony detection via sentiment-based transfer learning. In: Information Processing & Management 56 (5), S. 1633–1644.

Zhou, X.; Zafarani, R. (2020): A survey of fake news: Fundamental theories, detection methods, and opportunities. In: ACM Computing Surveys (CSUR) 53 (5), S. 1–40.