

A two dimensional accuracy-based measure for classification performance



Mariano Carbonero-Ruz, Francisco Jose Martínez-Estudillo,
Francisco Fernández-Navarro*, David Becerra-Alonso,
Alfonso Carlos Martínez-Estudillo

Department of Quantitative Methods, Universidad Loyola Andalucía, c/ Escritor Aguayo, 4, Córdoba, Spain

ARTICLE INFO

Article history:

Received 27 June 2016

Revised 30 November 2016

Accepted 4 December 2016

Available online 7 December 2016

Keywords:

Classification metrics

Imbalanced classification

Accuracy

ABSTRACT

Accuracy has been used traditionally to evaluate the performance of classifiers. However, it is well known that accuracy is not able to capture all the different factors that characterize the performance of a multiclass classifier. In this manuscript, accuracy is studied and analyzed as a weighted average of the classification rate of each class. This perspective allows us to propose the dispersion of the classification rate of each class as its complementary measure. In this sense, a graphical performance metric, which is defined in a two dimensional space composed by accuracy and dispersion, is proposed to evaluate the performance of classifiers. We show that the combined values of accuracy and dispersion must fall within a clearly bounded two dimensional region, different for each problem. The nature of this region depends only on the a priori probability of each class, and not on the classifier used. Thus, the performance of multiclassifiers is represented in a two dimensional space where the models can be compared in a more fair manner, providing greater awareness of the strategies that are more accurate when trying to improve the performance of a classifier. Furthermore we experimentally analyze the behavior of seven different performance metrics based on the computation of the confusion matrix values in several scenarios, identifying clusters and relationships between measures. As shown in the experimentation, the graphical metric proposed is specially suitable in challenging, highly imbalanced and with a high number of classes datasets. The approach proposed is a novel point of view to address the evaluation of multiclassifiers and it is an alternative to other evaluation measures used in machine learning.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Comparing learning/classification models is a complex and still open challenge. The first issue is the selection of the property of the model's performance that we want to measure. For example, we could be interested in accuracy, speed, cost or even maybe in readability of the model. Once a performance measure is chosen, the next concern is to estimate it in as unbiased a manner as possible [37]. After that we should test whether the differences in the performance obtained by the

* Corresponding author.

E-mail addresses: i22fenaf@uco.es, fafernandez@uloyola.es (F. Fernández-Navarro).

Proof

The immediate consequence for Property 2 is that both are the only intersection points, while increased and decreased conditions are deduced from the convexity of the function and the sign of its derivative with respect to C :

$$\frac{d\Pi}{dC}(s(I), I, i) = -2s(I) < 0 \quad \frac{d\Pi}{dC}(s(I) + p_i, I, i) = 2(1 - (s(I) + p_i)) > 0$$

Property 4

Given a second set of indexes J with an additional index $j \notin J$, the relative position of the curves $\Pi(c, I, i)$ and $\Pi(c, J, j)$ is exclusively determined by $s(I), p_i, s(j)$ and p_j .

Proof

It is immediate: only these four values define the parabolas.

References

- [1] R. Alaiz-Rodríguez, N. Japkowicz, P. Fischer, Visualizing classifier performance on different domains, in: Tools with Artificial Intelligence, 2008. ICTAI'08. 20th IEEE International Conference on, vol. 2, 2008, pp. 3–10.
- [2] C. Apte, E. Bibelnicks, R. Natarajan, E. Pednault, F. Tipu, D. Campbell, B. Nelson, Segmentation-based modeling for advanced targeted marketing, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2001, pp. 408–413.
- [3] A. Asuncion, D. Newman, UCI machine learning repository, 2007. URL <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [4] R. Baeza-Yates, B. Ribeiro-Neto, et al., Modern Information Retrieval, vol. 463, ACM press New York, 1999.
- [5] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159.
- [6] T. Byrt, J. Bishop, J.B. Carlin, Bias, prevalence and kappa, J. Clin. Epidemiol. 46 (5) (1993) 423–429.
- [7] R. Caruana, A. Niculescu-Mizil, Data mining in metric space: an empirical analysis of supervised learning performance criteria, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 69–78.
- [8] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46.
- [9] M. Cruz-Ramirez, C. Hervás-Martínez, J. Sánchez-Monedero, P. Gutiérrez, A preliminary study of ordinal metrics to guide a multi-objective evolutionary algorithm, in: Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on, 2011, pp. 1176–1181.
- [10] J.F.P. da Costa, H. Alonso, J.S. Cardoso, The unimodal model for the classification of ordinal data, Neural Netw. 21 (2008) 78–91.
- [11] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, 2006, pp. 233–240.
- [12] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
- [13] C. Drummond, R. Holte, Cost curves: an improved method for visualizing classifier performance, Mach. Learn. 65 (1) (2006) 95–130.
- [14] R.M. Everson, J.E. Fieldsend, Multi-class ROC analysis from a multi-objective optimisation perspective, Pattern Recognit. Lett. 27 (8) (2006) 918–927.
- [15] T. Fawcett, ROC graphs: notes and practical considerations for researchers, Mach. Learn. 31 (2004) 1–38.
- [16] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
- [17] J.C. Fernández Caballero, F.J. Martínez, C. Hervás, P.A. Gutiérrez, Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks, Neural Netw. IEEE Trans. 21 (5) (2010) 750–770.
- [18] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15 (1) (2014) 3133–3181.
- [19] F. Fernández-Navarro, C. Hervás-Martínez, C. García-Alonso, M. Torres-Jimenez, Determination of relative agrarian technical efficiency by a dynamic over-sampling procedure guided by minimum sensitivity, Expert Syst. Appl. 38 (10) (2011) 12483–12490.
- [20] F. Fernández-Navarro, C. Hervás-Martínez, P.A. Gutiérrez, A dynamic over-sampling procedure based on sensitivity for multi-class problems, Pattern Recognit. 44 (8) (2011) 1821–1833.
- [21] C. Ferri, J. Hernández-Orallo, R. Modroiu, An experimental comparison of performance measures for classification, Pattern Recognit. Lett. 30 (1) (2009) 27–38.
- [22] P. Flach, H. Blockeel, C. Ferri, J. Hernández-Orallo, J. Struyf, Decision support for data mining, in: Data Mining and Decision Support, in: The Springer International Series in Engineering and Computer Science, vol. 745, 2003, pp. 81–90.
- [23] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, Ordering-based pruning for improving the performance of ensembles of classifiers in the framework of imbalanced datasets, Inf. Sci. 354 (2016) 178–196.
- [24] L.A. Goodman, W.H. Kruskal, Measures of association for cross classifications*, J. Am. Stat. Assoc. 49 (268) (1954) 732–764.
- [25] P.A. Gutiérrez, C. Hervás-Martínez, F.J. Martínez-Estudillo, M. Carbonero, A two-stage evolutionary algorithm based on sensitivity and accuracy for multi-class problems, Inf. Sci. 197 (0) (2012) 20–37.
- [26] D.J. Hand, R.J. Till, A simple generalisation of the area under the ROC curve for multiple class classification problems, Mach. Learn. 45 (2) (2001) 171–186.
- [27] J. Hernández-Orallo, P. Flach, C. Ferri, Brier curves: a new cost-based visualisation of classifier performance, in: Proceedings of 28th International Conference on Machine Learning (ICML 2011), 2011, pp. 585–592.
- [28] J. Hernández-Orallo, P. Flach, C. Ferri, A unified view of performance metrics: translating threshold choice into expected classification loss, J. Mach. Learn. Res. 13 (1) (2012) 2813–2869.
- [29] G. Hripcsak, A.S. Rothschild, Agreement, the F-measure, and reliability in information retrieval, J. Am. Med. Inform. Assoc. 12 (3) (2005) 296–298.
- [30] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybern. Part B 42 (2) (2012) 513–529.
- [31] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: IEEE International Conference on Neural Networks - Conference Proceedings, vol. 2, 2004, pp. 985–990.
- [32] N. Japkowicz, M. Shah, Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, 2011.
- [33] L.A. Jeni, J.F. Cohn, F. De La Torre, Facing imbalanced data—recommendations for the use of performance metrics, in: Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, IEEE, 2013, pp. 245–251.
- [34] T.C.W. Landgrebe, R.P.W. Duin, Approximating the multiclass ROC by pairwise analysis, Pattern Recognit. Lett. 28 (13) (2007) 1747–1758.
- [35] C. Liu, P. Frazier, L. Kumar, Comparative assessment of the measures of thematic classification accuracy, Remote Sens. Environ. 107 (4) (2007) 606–616.
- [36] V. López, A. Fernández, S. García, V. Palade, F. Herrera, An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics, Inf. Sci. 250 (2013) 113–141.
- [37] M. Markatou, H. Tian, S. Biswas, G.M. Hripcsak, Analysis of variance of cross-validation estimators of the generalization error, J. Mach. Learn. Res. 6 (2005) 1127–1168.
- [38] T.M. Mitchell, Machine Learning, WCB McGraw-Hill, 1997.
- [39] PASCAL, Pascal (Pattern Analysis, Statistical Modelling and Computational Learning) machine learning benchmarks repository, 2011. <http://mldata.org/>.
- [40] L. Prechelt, PROBEN1: A set of Neural Network Benchmark Problems and Benchmarking Rules, Technical Report, Fakultät für Informatik (Universität Karlsruhe), 1994.

- [41] F. Provost, T. Fawcett, Robust classification for imprecise environments, *Mach. Learn.* 42 (3) (2001) 203–231.
- [42] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manage.* 45 (4) (2009) 427–437.
- [43] W. Waegeman, K. Dembczyński, A. Jachnik, W. Cheng, E. Hüllermeier, On the Bayes-Optimality of F-Measure maximizers, *J. Mach. Learn. Res.* 15 (2014) 3333–3388.
- [44] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.
- [45] J. Wu, H. Xiong, J. Chen, Cog: local decomposition for rare class analysis, *Data Min. Knowl. Discov.* 20 (2) (2010) 191–220.

2017 Journal Performance Data for: INFORMATION SCIENCES

ISSN	0020-0255	EISSN	1872-6291
JCR ABBREVIATION	INFORM SCIENCES	ISO ABBREVIATION	Inf. Sci.
Journal Information			
EDITION	Science Citation Index Expanded (SCIE)	CATEGORY	COMPUTER SCIENCE, INFORMATION SYSTEMS - SCIE
LANGUAGES	English	REGION	USA
		1ST ELECTRONIC JCR YEAR	1997

Publisher Information

PUBLISHER	ADDRESS	PUBLICATION FREQUENCY
ELSEVIER SCIENCE INC	360 PARK AVE SOUTH, NEW YORK, NY 10010-1710	36 issues/year

Journal's Performance

Journal Impact Factor

The Journal Impact Factor (JIF) is a journal-level metric calculated from data indexed in the Web of Science Core Collection. It should be used with careful attention to the many factors that influence citation rates, such as the volume of publication and citations characteristics of the subject area and type of journal. The Journal Impact Factor can complement expert opinion and informed peer review. In the case of academic evaluation for tenure, it is inappropriate to use a journal-level metric as a proxy measure for individual researchers, institutions, or articles. [Learn more](#)

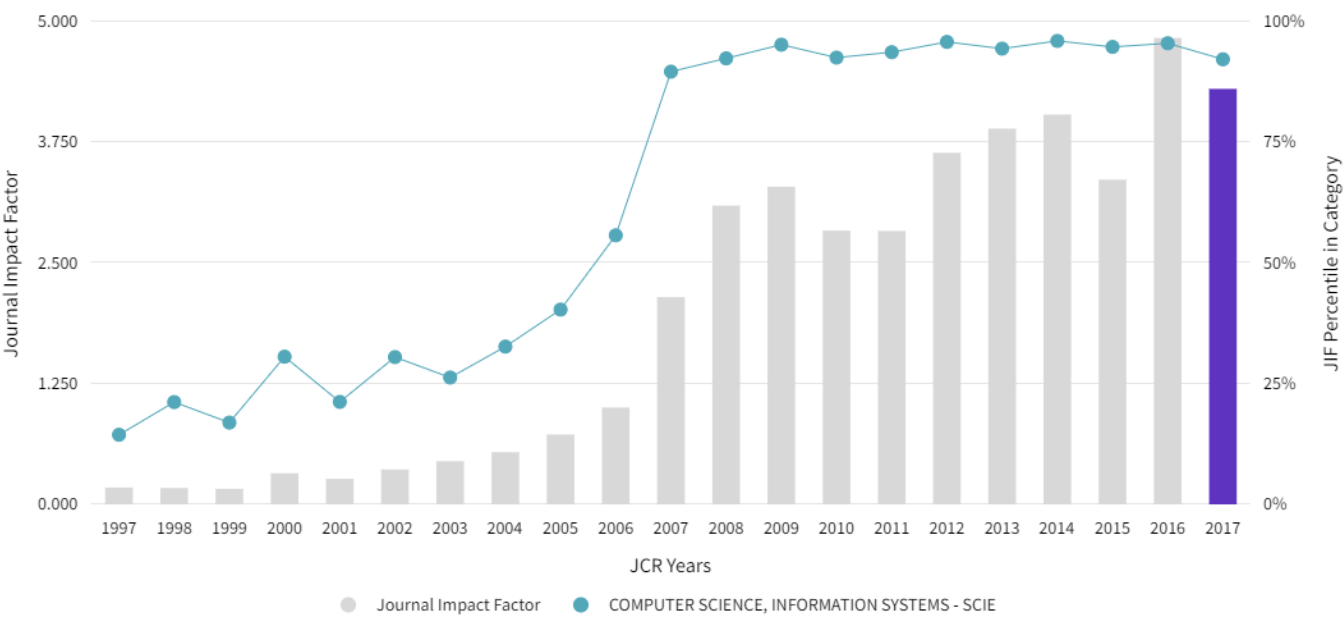
2017 JOURNAL IMPACT FACTOR

4.305

2017 JOURNAL IMPACT FACTOR WITHOUT SELF CITATIONS

3.690

Journal Impact Factor Trend 2017



Journal Impact Factor is calculated using the following metrics

Citations in 2017 to items published in 2015 (3,322) - 2016 (2,774)		6,096		
	=		=	4.305
Number of citable items in 2015 (615) + 2016 (801)		1,416		

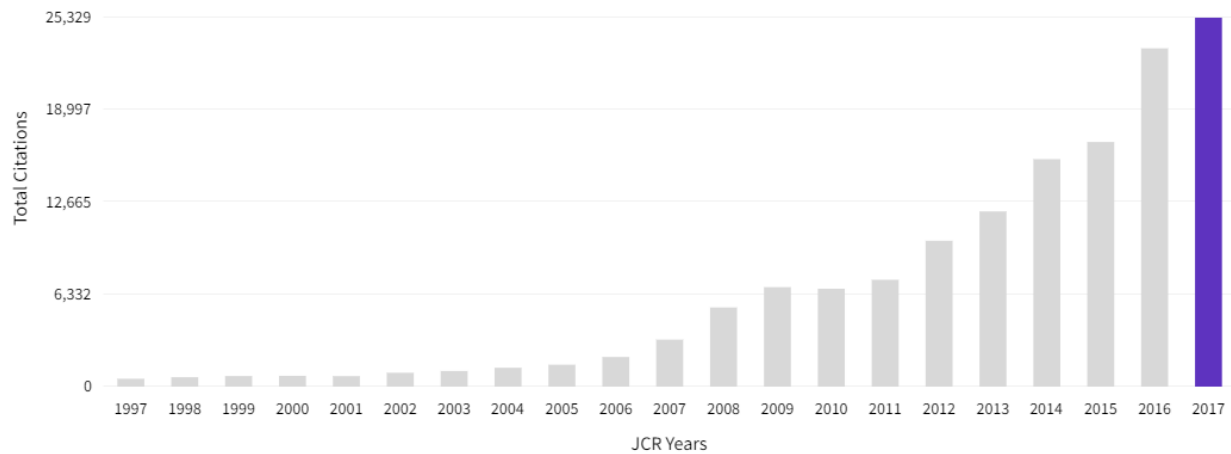
Journal Impact Factor without self cites is calculated using the following metrics

Citations in 2017 to items published in 2015 (3,322) + 2016 (2,774) - Self Citations in 2017 to items published in 2015 (415) + 2016 (456)		6,096 - 871		
	=		=	3.690
Number of citable items in 2015 (615) + 2016 (801)		1,416		

Total Citations

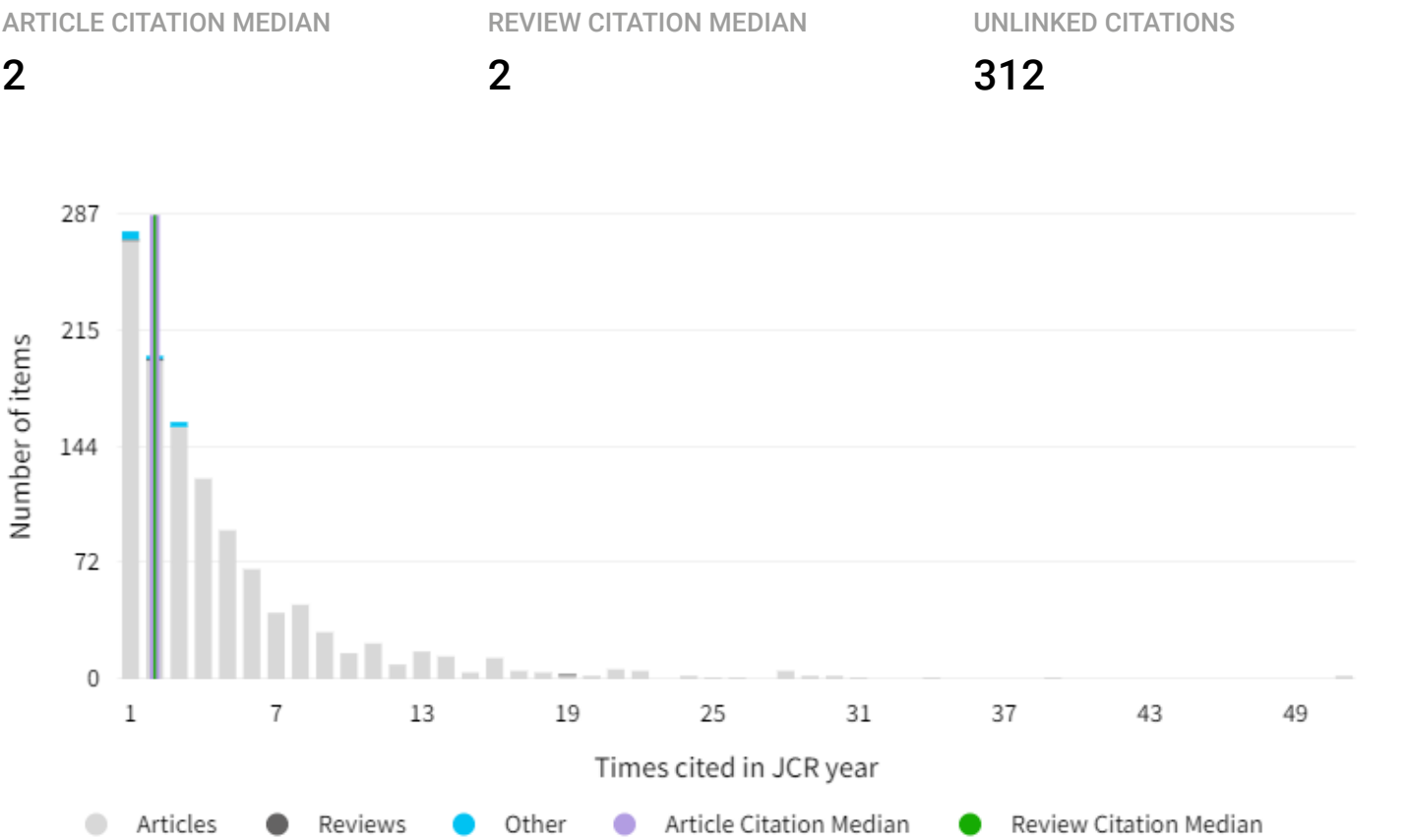
25,329

The total number of times that a journal has been cited by all journals included in the database in the JCR year. Citations to journals listed in JCR are compiled annually from the JCR years combined database, regardless of which JCR edition lists the journal.



Citation Distribution

The Citation Distribution shows the frequency with which items published in the year or two years prior were cited in the JCR data year (i.e., the component of the calculation of the JIF). The graph has similar functionality as the JIF Trend graph, including hover-over data descriptions for each data point, and an interactive legend where each data element's legend can be used as a toggle. You can view Articles, Reviews, or Non-Citable (other) items to the JIF numerator. [Learn more](#)



0 times cited

ARTICLES	REVIEWS	OTHER
252	0	13

Rank by Journal Impact factor

Journals within a category are sorted in descending order by Journal Impact Factor (JIF) resulting in the Category Ranking below. A separate rank is shown for each category in which the journal is listed in JCR. Data for the most recent year is presented at the top of the list, with other years shown in reverse chronological order. [Learn more](#)

EDITION

Science Citation Index Expanded (SCIE)

CATEGORY

COMPUTER SCIENCE, INFORMATION SYSTEMS

12/148

JCR YEAR	JIF RANK	QUART ILE	JIF PERCENTILE	
2022	13/158	Q1	92.1	
2021	16/164	Q1	90.55	
2020	18/161	Q1	89.13	
2019	9/156	Q1	94.55	
2018	9/155	Q1	94.52	
2017	12/148	Q1	92.23	
2016	7/146	Q1	95.55	
2015	8/144	Q1	94.79	
2014	6/139	Q1	96.04	
2013	8/135	Q1	94.44	
2012	6/132	Q1	95.83	
2011	9/135	Q1	93.70	
2010	10/128	Q1	92.58	
2009	6/116	Q1	95.26	
2008	8/99	Q1	92.42	
2007	10/92	Q1	89.67	
2006	39/87	Q2	55.75	
2005	50/83	Q3	40.36	
2004	53/78	Q3	32.69	
2003	58/78	Q3	26.28	
2002	54/77	Q3	30.52	
2001	58/73	Q4	21.23	
2000	47/67	Q3	30.60	
1999	52/62	Q4	16.94	
1998	47/59	Q4	21.19	
1997	45/52	Q4	14.42	

Rank by Journal Citation Indicator (JCI)

Journals within a category are sorted in descending order by Journal Citation Indicator (JCI) resulting in the Category Ranking below. A separate rank is shown for each category in which the journal is listed in JCR. Data for the most recent year is presented at the top of the list, with other years shown in reverse chronological order.[Learn more](#)

CATEGORY

COMPUTER SCIENCE, INFORMATION SYSTEMS

10/216

JCR YEAR	JCI RANK	QUART ILE	JCI PERCENTILE	
2022	16/251	Q1	93.82	
2021	12/246	Q1	95.33	
2020	9/223	Q1	96.19	
2019	12/223	Q1	94.84	
2018	7/220	Q1	97.05	
2017	10/216	Q1	95.60	