

Data mining model for identifying high-quality journals

***J. K. D. B. G. Jayaneththi and B. T. G. S. Kumara**

Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

**bgayashani@gmail.com*

Abstract

The focus in local universities over the last decade, have shifted from teaching at undergraduate and postgraduate levels to conducting research and publishing in reputed local and international journals. Such publications will enhance the reputation on the individual and the university. The last two decades has seen a rapid rise in open access journals. This has led to quality issues and hence choosing journals for publication has become an issue. Most of these journals focus on the monetary aspect and will publish articles that previously may not have been accepted. Some of the issues include design of the study, methodology and the rigor of the analysis. This has great consequences as some of these papers are cited and used as a basis for further studies. Another cause for concern is that, the honest researchers are sometimes duped, into believing that journals are legitimate and may end up by publishing good material in them. In addition, at present, it is very difficult to identify the fake journals from the legitimate ones. Therefore, the objective of the research was to introduce a data mining model which helps the publishers to identify the highest quality and most suitable journals to publish their research findings. The study focused on the journals in the field of Computer Science. Journal Impact Factor, H-index, Scientific Journal Rankings, Eigen factor Score, Article Influence Score and Source Normalized Impact per Paper journal metrics were used for building this data mining model. Journals were clustered into five clusters using K-Means clustering algorithm and the clusters were interpreted as excellent, good, fair, poor and very poor based on the results.

Keywords: Data mining, K-Means clustering, Journal ranking

Introduction

Research is a self-motivated creative work undertaken by researchers on a systematic basis in order to seek answers to questions that arise in their minds. The results of such studies are published in journals primarily to share the new findings with a larger peer group. By 2015, more than half a million papers had been published in predatory journals, and at the end of 2016, the number of predatory journals on Beall's list (about 10,000) approached the number indexed by the Directory of Open Access Journals (DOAJ) and Journal Citation Reports (JCR). Most of these are hosted by publishing companies (including some industry giants). Therefore, predatory publishing is becoming an organized industry (Sorokowski et al., 2017). Beall (2015) coined the term 'predatory publishers' to describe publishers in the scholarly publishing business who collect article processing charges and provide rapid publishing without a proper peer-review process. He has listed a list of predatory publishers. This list is rather long with 48 criteria, which directly or indirectly indicate whether the publisher and individual journal is predatory.

Accordingly, measuring the quality of the journals and identifying the journals with high-quality has become a vital need in the current society. There are many factors to measure the quality of electronic journals, such as quality of content standardization, purpose and coverage, periodicity and continuity, timeliness and