

In our thesis, we examine the book-sharing and review site Goodreads, using a sample of 6,539 items. We focus on the relationship between author point scores and literary significance. We are looking for an answer to the question of whether the scoring system of the author is formed only based on the entertainment value of the author's books or whether the literary value also plays a role in it. We draw attention to the consequences of these effects on the book market.

In our opinion, even rating the books on a scale of 1-5 is an oversimplification of the problem, which is further strengthened by the fact that the author cannot be rated separately. The average rating is automatically calculated from the ratings given to the author's books. In our opinion, the entertainment - and literary factor of the book should be distinguished. The first should appear in the point score of the book, while the second in the author's point. According to our hypothesis, entertainment is more important for the website's users. One of the reasons for this may be the age distribution among users. The consequence of this factor is that the highly rated and often recommended books represent this quality.

We examine the questions using the methodology of k-centred clustering, sentiment -, principal component - and network analysis. Based on our results, well-known awards do not significantly increase the point score of books, nor their frequency of recommendation. The most frequently used vocabulary does not differ significantly in the highest-rated books and award-winning works. Most of the authors have ten times as many books on the bookshelves as their popular competitors. The recommendation system does not only work based on the values of the average evaluation variables. The system also takes into account the variables describing the evaluation numbers. Designing the author evaluation in this way gives false results since it is given to the greatest extent by the point score of a work. The recommendation network is low-density and has small-world properties. The system is not random, there lies an algorithm behind it that does not overrepresent the award-winning works.

Keywords: Goodreads, Principal component analysis, Clustering, Sentiment analysis, Network analysis

JEL classification: C55, C38, D83