



Machine Learning Based Approach for Future Prediction of Authors in Research Academics

Sovan Bhattacharya^{1,2} · Ayan Banerjee¹ · Abhik Goswami¹ · Subrata Nandi² · Dinesh Kumar Pradhan¹

Received: 30 September 2022 / Accepted: 9 January 2023
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

Abstract

In the present time, it is important to predict superior authors from the huge research community, which determines their performance with their future prospects and opportunities, in the field of scientific research. There is a need to solve the authors' future prediction problem for equally assessing the performance of the researchers that helps to improve research quality highly, influence other types of research, that will also help to identify the research carrier and other research parameters, that will further influence budding researchers. In our proposed model, to solve the authors' future prediction problems, five machines learning models i.e., SVM, Logistic Regression, Naive Bayes, Decision Tree, and Random Forest are implemented on a data set of 2750 authors where 1000 authors are the ranked authors. So, we have collected 2750 authors' data from their bibliographic data sets. Then used feature engineering and feature scaling to build the desired model and prepare the final data set used to build the model. There after varied data sizes and training ratios against the five ML models and measured all the evaluation metrics in different cases. We have seen that the Decision Tree and Random Forest model outperform the other models with an accuracy score of 1.0 and the other scores also performed well.

Keywords Machine learning · Citation networks · Bibliographic information · Author impact evaluation · Author future prediction · Rising star prediction

Sovan Bhattacharya, Ayan Banerjee, Abhik Goswami, Subrata Nandi and Dinesh Kumar Pradhan contributed equally to this work.

This article is part of the topical collection "Enabling Innovative Computational Intelligence Technologies for IOT" guest edited by Omer Rana, Rajiv Misra, Alexander Pfeiffer, Luigi Troiano and Nishtha Kesswani.

✉ Sovan Bhattacharya
sovan.cse@gmail.com; sovan.bhattacharya@bcrec.ac.in

Ayan Banerjee
ayanbanerjee905@gmail.com

Abhik Goswami
gos24abhik@gmail.com

Subrata Nandi
subrata.nandi@cse.nitdp.ac.in

Dinesh Kumar Pradhan
dineshkrp@gmail.com

¹ Department of CSE, Dr. B. C. Roy Engineering College, Jemua Road, Durgapur, West Bengal 713206, India

² Department of CSE, National Institute of Technology, Mahatma Gandhi Road, Durgapur, West Bengal 713209, India

Introduction

The number of researchers, scientists, and authors are steadily expanding in the modern day, and the same is true for their research publications. Now, it is bewildering who is the best author and how their future will be clocking. All of them may not be as experienced as their seniors but the grim reality is that they can develop themselves to be at the top 1 day. Hence, this is very important to analyze and evaluate the citation count productivity amongst the academicians to determine the speculate future publications. Predictions play a major role in the emergence of an increase in research papers and authors in the research field. This will help us in having great research and also help us to work with excellent researchers and build their careers. Although a scientist's citations may have been plentiful in the past, his or her success may have waned since then. As a result, using the predicted citation count of scientists to make decisions about grant awards, application acceptance, referee assignments, and so on is a valuable endeavor. The researchers have discovered a dynamic problem. The influence of scientists'