Classification Of Oxygen Binding Proteins Using Random Forest Predictive Modeling Utilizing The R Language

Matthew Curcio*

* Department of Bioinformatics, Worcester Polytechnic Institute
100 Institute Rd, Worcester, MA 01609
https://github.com/mccurcio/Oxy-RF

Submitted April 2019, To Dmitri Korkine, Ph.D.,

Abstract- Mention the abstract for the article. An abstract is a brief summary of a research article, thesis, review, conference proceeding or any in-depth analysis of a particular subject or discipline, and is often used to help the reader quickly ascertain the paper's purpose. When used, an abstract always appears at the beginning of a manuscript, acting as the point-of-entry for any given scientific paper or patent application.

Keywords- machine learning; predictive modeling; random forest (RF); R; RStudio; oxygen binding proteins; erythrocruorin; hemerythrin; hemocyanin; hemoglobin; leghemoglobin; myoglobin; classification; support vector machine (SVM)

I. INTRODUCTION

Extensive studies on oxygen-binding proteins have categorized them into six different broad types, including erythrocruorin, hemerythrin, hemocyanin, hemoglobin, leghemoglobin, and myoglobin, each has its own functional characteristics and structure with unique oxygen-binding capacity (1).

Research papers are highly recognized in scholar fraternity and form a core part of PhD curriculum. Research scholars publish their research work in leading journals to complete their grades. In addition, the published research work also provides a big weight-age to get admissions in reputed varsity. Now, here we enlist the proven steps to publish the research paper in a journal.

III. WRITE DOWN YOUR STUDIES AND FINDINGS

Now it is the time to articulate the research work with ideas gathered in above steps by adopting any of below suitable approaches:

A. Bits and Pieces together

In this approach combine all your researched information in form of a journal or research paper. In this researcher can take the reference of already accomplished work as a starting building block of its paper.

Jump Start

This approach works the best in guidance of fellow researchers. In this the authors continuously receives or asks inputs from their fellows. It enriches the information pool of your paper with expert comments or up gradations. And the researcher feels confident about their work and takes a jump to start the paper writing.

Identify the constructs of a Journal – Essentially a journal consists of five major sections. The number of pages may vary depending upon the topic of research work but generally comprises up to 5 to 7 pages. These are:

1) Abstract
2) Introduction
3) Research Elaborations
4) Results or Finding
5) Conclusions

In Introduction you can mention the introduction about your research.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

It's the foremost preliminary step for proceeding with any research work writing. While doing this go through a complete thought process of your Journal subject and research for it's viability by following means:

1) Read already published work in the same field.
2) Goggling on the topic of your research work.
3) Attend conferences, workshops and symposiums on the same fields or on related counterparts.
4) Understand the scientific terms and jargon related to your research work.

B. Use of Simulation software

There are numbers of software available which can mimic the process involved in your research work and can produce the possible result. One of such type of software is Matlab. You can readily find Mfiles related to your research work on internet or in some cases these can require few modifications. Once these Mfiles are uploaded in software, you can get the simulated results of your paper and it easies the process of paper writing.

As by adopting the above practices all major constructs of a research paper can be written and together compiled to form a complete research ready for Peer review.

IV. GET PEER REVIEWED

Here comes the most crucial step for your research publication. Ensure the drafted journal is critically reviewed by your peers or any subject matter experts. Always try to get maximum review comments even if you are well confident about your paper.

V. IMPROVEMENT AS PER REVIEWER COMMENTS

Analyze and understand all the provided review comments thoroughly. Now make the required amendments in your paper. If you are not confident about any review comment, then don't forget to get clarity about that comment. And in some cases there could be chances where your paper receives number of critical remarks. In that cases don't get disheartened and try to improvise the maximum.

After submission IJSRP will send you reviewer comment within 10-15 days of submission and you can send us the updated paper within a week for publishing.

This completes the entire process required for widespread of research work on open front. Generally all International Journals are governed by an Intellectual body and they select the most suitable paper for publishing after a thorough analysis of submitted paper. Selected paper get published (online and printed) in their periodicals and get indexed by number of sources.

Performance Evaluation

To evaluate the feasibility and efficiency of the proposed method, four parameters, the accuracy of prediction (Ac), sensitivity (Sn), precision (Pe), and Matthews' correlation coefficient (Mcc), were computed. They are represented as follows:

$$Ac = \frac{TP+TN}{TP+FP+TN+FN} \quad (20)$$

$$Sn = \frac{TP}{TP+FN} \quad (21)$$

$$Pe = \frac{TP}{FP+TP} \quad (22)$$

$$Mcc = \frac{(TP\,x\,TN)-(FP\,x\,FN)}{\sqrt{(TP+FN)\,x\,(TN+FP)\,x\,(TP+FP)\,x\,(TN+FN)}} \quad (23)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. True positives represent the number of true interacting pairs correctly predicted. True negatives are the number of true non-interacting pairs predicted correctly. False positives stand for the number of true non-interacting pairs falsely predicted, and false negatives are the number of true interacting pairs falsely predicted to be non-interacting pairs. Moreover, we used the Receiver Operating Curve (ROC) to evaluate the performance of our proposed method.

VI. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

APPENDIX

Appendixes, if needed, appear before the acknowledgment.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments.

REFERENCES

[1] S. Muthukrishnan, Aarti Garg, and G.P.S. Raghava, *Oxypred: Prediction and Classification of Oxygen-Binding Proteins*, Geno. Prot. Bioinfo., Vol. 5, No. 3–4, 2007, pp. 250–252.

[2] W.-K. Chen, *Linear Networks and Systems* Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[4] RStudio (2019). RStudio: Integrated development environment for R (Version 1.1.463) [Computer software]. Boston, MA. © 2009-2018 RStudio, Inc., http://www.rstudio.org/

[5] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3):299–314, 1996

[6] R.D. Peng, "Exploratory Data Analysis with R" Leanpub Publishing, https://leanpub.com, © 2015 - 2016

AUTHOR(S)

Correspondence Author– Matthew Curcio, matt.curcio.ri@gmail.com

Conflicts of Interest- The author declares no conflict of interest.