1

Mining Data From Ravelry to Determine the Important Attributes for a Highly Rated Yarn

Alexander Hall

***Abstract*—Various machine learning methods were applied to the entire yarn database from Ravelry.com in an attempt to identify attributes that influence the average user-rating for a given yarn. The dataset was used to train Support Vector Machines, Deep Neural Networks and Random Forests; with various parameters. 5 fold cross validation was used to evaluate each model, with the Root Mean Squared Error (RMSE) used as the validation metric. The Random Forest method outperformed the other two methods, with an RMSE of 0.551. Attributes relating to the yarn producer (for example, number of yarns previously produced) and physical dimensions of the yarn (grams, gauge etc) were found to be more important than the physical compisition of the yarn when predicting user rating. The time a yarn has been present on the Ravelry database was found to be the most significant predictor for a yarn's rating.**

I. Introduction

THIS demo file is intended to serve as a "starter file" for IEEE journal papers produced under LaTeX using IEEEtran.cls version 1.8a and later. I wish you the best of success.

mds
September 17, 2014

A. Subsection Heading Here

Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

II. Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

Department of Computing and Digital Media,Robert Gordon University, Aberdeen, UK

PLACE PHOTO HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.