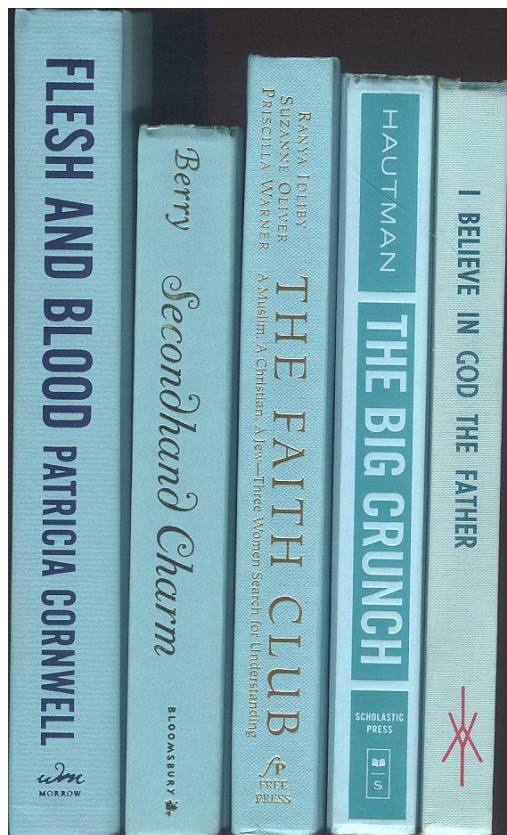


Machine Learning Approach to Book Recommendation.

Prepared by:

Fuki Babasaki

20/9/2021



Executive Summary

This report explores a way to determine the readability of books based on their paragraphs as well as showcasing the findings on the relationship between the books and the readability. We will cover introduction, methods used and conclusion. The books analysed are top 100 books at <https://www.gutenberg.org/>.

We conclude that there are 29 books to recommend for year 11-13 students (senior students) and 38 books to recommend for year 9-10 students (junior students) from the top 100 books. in the process, we have also found that there is no difference in ease of reading between year 9 and 10 students, and year 11, 12, and 13 students. The recommendation is included in Appendix B.

Table of Contents

Executive Summary 1

Introduction..... 2

 Preview of Background..... 2

 Limitations 2

Methods 2

Results 3

 Interesting findings..... 3

 Book Recommendation 3

Conclusion 3

Appendices 3

 Appendix A 3

 Appendix B..... 4

Introduction

PREVIEW OF BACKGROUND

Natural Language Processing (NLP) is a branch of linguistic studies that enables analysis on text. An international company is looking to this technology to optimise English language curriculums for educational institutions. More specifically, we are aiming to predict the complexity of books given their paragraphs to recommend books for students of a specific age. In this report, target value is referred to numbered complexity of a book.

LIMITATIONS

Limitations include that the target value of a books is determined for year 4-13 which means that the model should not be used to recommend books for readers aged out of the range. It's also worth noting that the complexity is determined by teachers, not year 4-13 students so it might not be the perfect measure of how difficult the students feel reading it.

Methods

Paragraphs of books to predict the target value of are obtained from the website <https://www.gutenberg.org/>. The paragraphs are then cleaned using one of the NLP technique to make it understandable by the machine.

Three models used to predict were considered and evaluated. The one that performed the best was picked as the final model and is used for predictions of target values.

Since the definition of the target value is vague, there is no clear answer as to how it can be used to determine the year level. One thing confirmed is that the higher target value indicates the easier the book is to read. Therefore, I decided to find the relationship between the target values and the target year level range using an additional dataset. The additional dataset contains information of a few books, their targeted year ranges (Year 9-10, 11 and 12-13) and the predicted target values (see Appendix A for more information). From analysis, I found that the target value for year 11 students is not significantly different than year 12-13 students, and the target value for year 9-10 students is significantly different from that for other year levels at 95% confidence. I have obtained the range of target values for senior students and that for junior students which can be used for recommending books (see Appendix A for table of target values).

Results

Interesting findings

“The War of the Worlds” is the most difficult book to read among the top 100 books, whereas “Don Quixote” is the easier to read. It also seems that there are more books suitable for junior students than for senior students in the top 100 books.

Book Recommendation

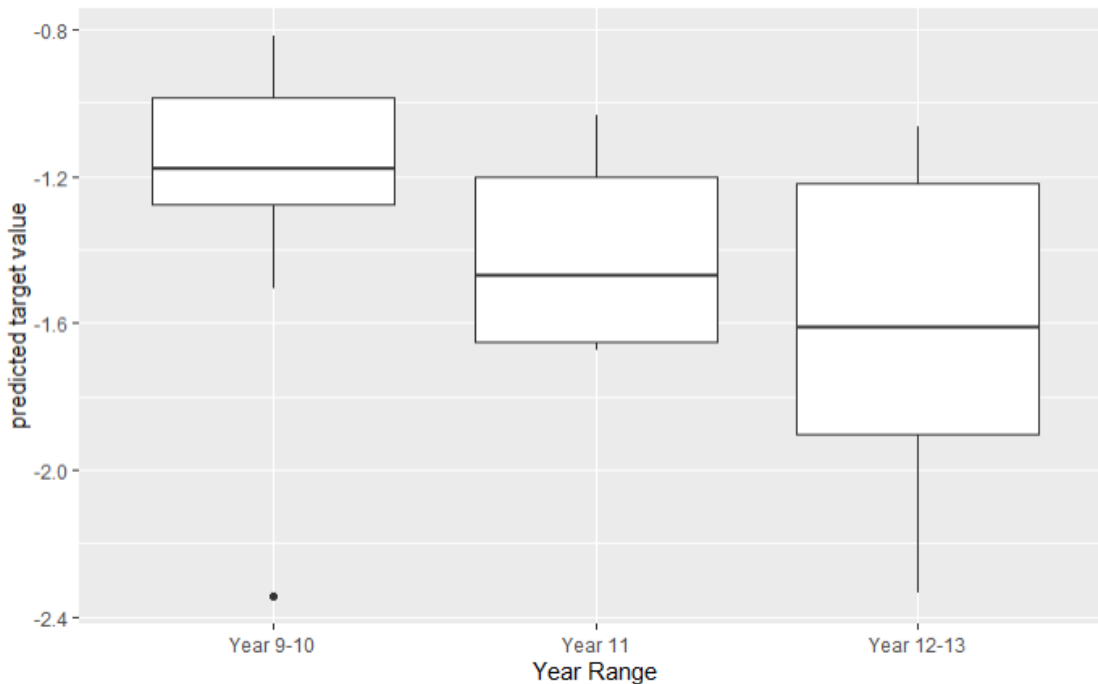
Please see Appendix B in appendices section for book recommendation.

Conclusion

Finding the right book(s) for students in a specific age group can often be challenging. We have studied a way of recommending books using machine learning approach and NLP techniques. In doing so, we have found that there is difference in ease of reading when it comes to books for junior students and senior students, but no difference between books for year 9 and 10, and that for year 11, 12 and 13 students. Further work includes extending the recommendation to all age groups by gathering more data and automating the process of recommendation.

Appendices

APPENDIX A



Target Value range	Year Level
$-1.87 < x < -1.35$	Senior
$-1.35 < x < -0.98$	Junior

APPENDIX B

Recommended books for senior students	Recommended books for junior students
Pride and Prejudice	Alice's Adventures in Wonderland
Moby Dick Or, The Whale	Metamorphosis
The Scarlet Letter	A Tale of Two Cities
The Yellow Wallpaper	A Modest Proposal
The Picture of Dorian Gray	Grimms' Fairy Tales
Dracula	Anthem
Walden, and On The Duty Of Civil Disobedience	The Republic
Jane Eyre: An Autobiography	The Odyssey
The Souls of Black Folk	The Happy Prince, and Other Tales
Crime and Punishment	The Importance of Being Earnest: A Trivial Comedy for Serious People
War and Peace	Anne of Green Gables
Little Women	Dubliners
The Awakening, and Selected Short Stories	Autobiography of Benjamin Franklin
Peter Pan	The Interesting Narrative of the Life of Olaudah Equiano, Or Gustavus Vassa, The African
Old Granny Fox	The Hound of the Baskervilles
The Legend of Sleepy Hollow	The Time Machine
The Call of the Wild	The Romance of Lust: A classic Victorian erotic novel
Emma	Uncle Tom's Cabin
Oliver Twist	Ethan Frome
The Kama Sutra of Vatsyayana	The Secret Garden
Essays of Michel de Montaigne " Complete	The Brothers Karamazov
The History of the Peloponnesian War	A Pickle for the Knowing Ones

David Copperfield	The Slang Dictionary: Etymological, Historical and Andecdotal
Carmilla	Notes from the Underground
The Works of Edgar Allan Poe â€” Volume 2	Thus Spake Zarathustra: A Book for All and None
The Jungle	The Extraordinary Adventures of Arsene Lupin, Gentleman-Burglar
Japanese Girls and Women	Anna Karenina
Persuasion	The Mysterious Affair at Styles
The History of the Decline and Fall of the Roman Empire	Complete Original Short Stories of Guy De Maupassant
	The King James Version of the Bible
	Pygmalion
	On Liberty
	Le Morte d'Arthur: Volume 1
	The Iliad
	The Art of War
	Frankenstein Or, The Modern Prometheus
	Three Men in a Boat (To Say Nothing of the Dog)
	Through the Looking-Glass