

BOOK RECOMMENDED

BY MICHELLE HOANG - 2021



EXECUTIVE SUMMARY

Reading books is one of the most effective ways to enhance your vocabulary – the linguistic aspect of the left brain and nurture and enrich our imagination for our right brain. Due to these advantages, many teachers encourage students to read to improve their academic and personal development. However, choosing the book suitable for their level of understanding is a hassle: no 'one book suits all'. Hence, we explored several methods to analyse the readability based on paragraphs and recommended books based on their readability score for secondary students (year 9 to 12)

For further study, we recommend looking into the relationship between the number of downloads with the readability to answer the question if we can rely on the books' popularity to access their readability

INTRODUCTION

Natural Language Processing (NLP) is the interdisciplinary field combining the knowledge of linguistic, computer science and artificial intelligence. An international company which specialises in developing the English language curriculums for educational institutions worldwide seek to optimise the development of their curriculum using NLP. We aim to predict the complexity of the books by analyzing their paragraphs. In this report, the target/score number represents how hard it is for students to read.

It is notable that the methods used to produce the target value, which was conducted by Georgia State University's researchers, were only applied for year 4-13 students. It should not be applied to a wider audience.

METHODS

We scraped paragraphs from the 25 most popular books from a free e-book website called Project Gutenberg (<https://www.gutenberg.org/>). The paragraphs were then cleaned and vectorized, an NLP process converting word to number so the machine can understand.

We considered and evaluated three models. The model with the best performance was used to predict the target value of the paragraphs.

It is unclear how the scoring (target) is corresponding to certain age ranges, we are only certain that the higher the score, the easier it is to read. To determine which books to recommend for students, we used percentile values of the target to decide ranges of books with scores higher than -0.87 for junior students (year 9 to 10), books with scores lower than -1.47 for seniors (year 12 to 13) and the rest for year 10 to 11.

RESULTS

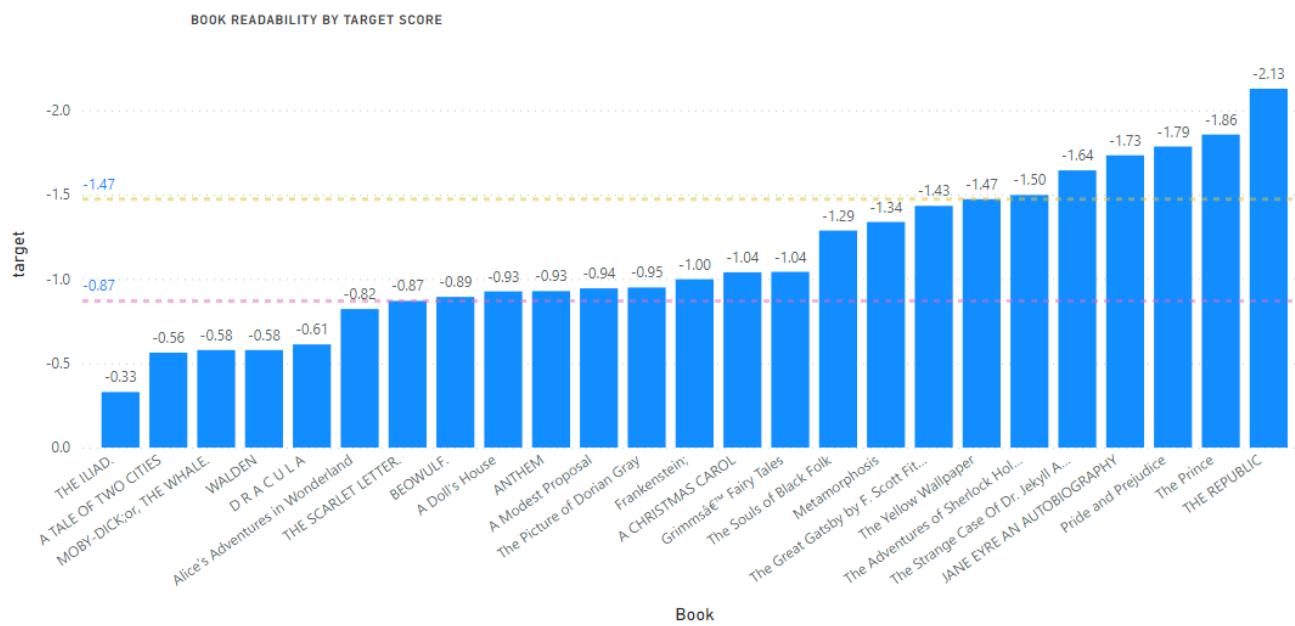


Figure 1. Book readability is represented by the target score. 25th percentiles (pink) indicates recommended books for juniors, while 75th percentile (yellow) indicates books for seniors

Surprisingly, the most famous book, "Pride and Prejudice" by Jane Austen, was classified as one of the most difficult to read. In contrast, "The Iliad" by Homer, with only a third of downloads compared to the former, was the easiest to read.

LIST OF BOOK RECOMMENDATION

Book for juniors	Book for seniors
THE SCARLET LETTER.	THE REPUBLIC
Alice's Adventures in Wonderland	The Prince
D R A C U L A	Pride and Prejudice
MOBY DICK; or, THE WHALE.	JANE EYRE AN AUTOBIOGRAPHY
WALDEN	The Strange Case Of Dr. Jekyll And Mr. Hyde
A TALE OF TWO CITIES	The Adventures of Sherlock Holmes
THE ILIAD.	The Yellow Wallpaper

Books for year 10 - 11	Books for year 10 - 11
The Great Gatsby by F. Scott Fitzgerald	The Picture of Dorian Gray
Metamorphosis	A Modest Proposal
The Souls of Black Folk	ANTHEM
Grimms' Fairy Tales	A Doll's House
A CHRISTMAS CAROL	BEOWULF.
Frankenstein;	