Book's readability recommendation list from a Machine Learning product

Sky Nguyen
Department of Electrical and Computer Engineering
University of Auckland
Auckland, New Zealand
pugn057@aucklanduni.ac.nz

***Abstract*— This document is a reflection on the approaches and decisions made for the design and implementation process of the machine learning model to get the complexity or readability score of any given passage string. Thus, included the list of book recommendations for high school English teachers range from Year 9 to Year 13 in each year level.**

Keywords—Machine learning, Readability, English.

I. INTRODUCTION

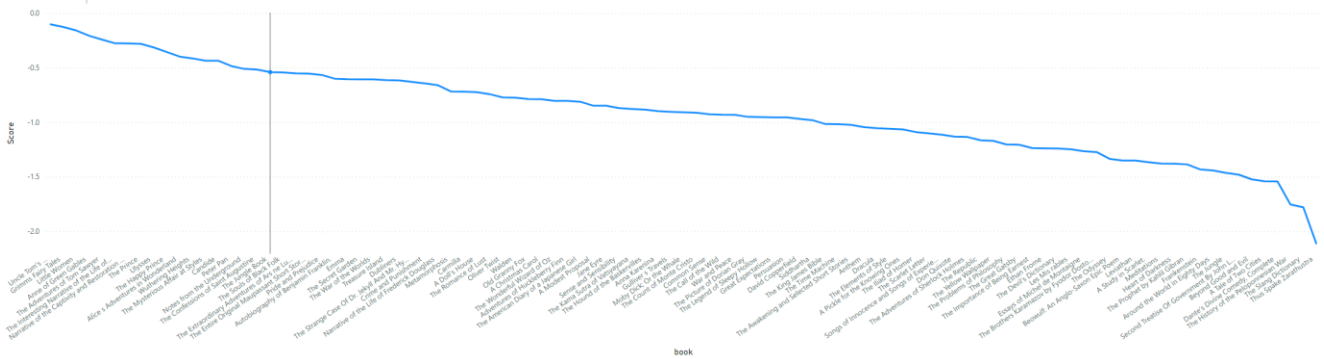
With a business case of participated in a competition for a company that targets toward English language curriculums by using NLP to optimise the development, the goal is to establish a machine learning model to calculate the readability or complexity of any given books. Thus, this result can be used to provide suitable books for different English year level. This project is achieved via the use of Google Collaboration service and Microsoft Azure service. All written code will be provided in GitHub library.

The rest of the paper is organized as follows; Section II will be discussed about the machine learning model approach. Before concludes this paper with Section IV, a list of book recommendations will be presented in Section III.

II. MACHINE LEARNING MODEL

The machine learning model is created to enter the CommonLit, Inc. Readability competition through Kaggle. The company is a nonprofit education technology organization serving over twenty million teacher and students with free digital reading and writing lessons. The model is trained with the training dataset from the competition to achieve the ability to score the complexity or readability of those paragraphs. However, the model also contains the ability to predict the score of any unfamiliar passage input from the user. This feature is used to calculate the readability score of the top 100 books according to the Project Gutenberg. The model will examine through each individual book and retrieve the longest paragraph to decide the complexity of that book. The higher the readability score given by the model indicates that book is easier to read compare to the one which has lower score. Furthermore, based on this result, the model produced a sorted list of top 100 books from easy to hard for reader.

III. LIST OF BOOK RECOMMENDATIONS



book	Score
Uncle Tom's Cabin	-0.10
Grimms Fairy Tales	-0.13
Little Women	-0.16
Anne of Green Gables	-0.21
The Adventures of Tom Sawyer	-0.24
The Interesting Narrative of the Life of Olaudah Equiano, or Gustavus Vassa, The African	-0.28
Narrative of the Captivity and Restoration Of Mrs. Mary Rowlandson	-0.28
The Prince	-0.28
Ulysses	-0.31
The Happy Prince	-0.36
Alice s Adventures in Wonderland	-0.40
Wuthering Heights	-0.41
The Mysterious Affair at Styles	-0.44
Candide	-0.44
Peter Pan	-0.49
Notes from the Underground	-0.51
The Confessions of Saint Augustine	-0.52
The Jungle Book	-0.54
The Souls of Black Folk	-0.54
The Extraordinary Adventures of Ars ne Lupin, Gentleman-burglar	-0.55
The Entire Original Maupassant Short Stories	-0.55
Pride and Prejudice	-0.57
Autobiography of Benjamin Franklin.	-0.60

Figure 1. Visualization of ranked list books

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

From figure 1, a table of book recommendations for different year level can be form. With the list of top 100 books, the table will include 20 books for each year level. The top 20 books with highest score will be suggested for Year 9. Meanwhile, the bottom 20 books with lowest score will suggest for Year 13 with general better literature understanding. The table below will be an example of top 5 books recommended for each level. For more book recommendation, user can access through the visualization of ranked list books (Figure 1)

Year 9 (Score: -0.1 => -0.55)	**Year 10 (Score: -0.55 => -0.8)**	**Year 11 (Score: -0.8 => -0.98)**	**Year 12 (Score: -1.02 => -1.25)**	**Year 13 (Score: -1.27 => -2.11)**
1. Uncle Tom's Cabin	1. The Entire Original Maupassan Short Stories	1. Adventures of Huckleberry Finn	1. The King James Bible	1. Essays of Michel de Montaigne
2. Grimms Fairy Tales	2. Pride and Prejudice	2. The American Dairy of a Japanese Girl	2. The Time Machine	2. The Brothers Karamazov by Fyodor Dostoyevsky
3. Little Women	3. Autobiography of Benjamin Franklin	3. A Modest Proposal	3. The Awakening and Selected Short Stories	3. The Odyssey
4. Anne of Green Gables	4. Emma	4. Jane Eyre	4. Anthem	4. Beowulf: An Anglo-Saxon Epic Poem
5. The adventures of Tom Sawyer	5. The Secret Garden	5. Sens and Sensibility	5. Dracula	5. Leviathan

IV. CONCLUSION

With the aid of machine learning, this report provides a solution for English teacher, who is looking for a new book recommendation for their year level classes bases on their readability. This report also included a list of book recommendation in a table form. This table do not include all the selected books; however, it does include the suggested readability score for different level. Base on that score, teachers can use the model to check any book before use it for their classes or also get other books in the top 100 list accordingly.