Online Book Rating Prediction

Final Project Report

Yang Zheng, Yu-Hsuan Kuo, Yu-San Lin, Ziyang Qi

{yxz184, yzk5145, yul189, zzq103}@psu.edu

Abstract This project proposed approaches to construct feature sets in order to predict a book's quality based on the data collected from the online bookstore. Review and rating data of books in different categories are collected from the Barnes & Noble website. Various regression and classification algorithms are performed on the constructed feature sets, while the rating data is used as the test set to verify the correctness of the model.

1 Introduction

Predicting the quality of a book is an interesting and important issue among publishers, professional book reviewers, and even expert writers. It is also an interesting question for the public as people prefer to read something interesting and worth their time. Many factors may potentially contribute to the quality of the book, and different categories may have rather different factors. For example, clarity and correctness are crucial for a textbook, while novelty and interestingness are more important for a novel. Among these factors, some concern the intrinsic content and quality of the book, such as the style of writing, while external factors such as social context are important as well.

Several previous studies analyze literary styles of significant authors or consider the content characteristics such as plots, characteristics of characters, action, emotion, of the best-selling novels, which rely on the knowledge and insights of human experts on literature. Other studies investigate the connection between stylistic elements (such as linguistic patterns) and the literary success using mining methods [1]. All these studies analyze the content of the book, which is very complex to build statistical models.

Our project aims to solve this problem from another perspective. Thanks to the online stores, people nowadays can buy books and leave reviews on the website. The reviews can be seen by everyone and definitely have an impact on other consumers' decisions. In this project, we focus on predicting the quality of a book using the data provided on the website, rather than the intrinsic factors of the book. The advantage of using these data lies in the fact that they are easily accessible and of high reliability as only the book buyers will provide the comment.

In this project, the books' basic data is used as the training set. The model is verified using the rating data. Several different classification methods and mining algorithms are evaluated to optimize the prediction accuracy.

In the rest of this report, we firstly introduce how to collect data from the online bookstore, and extract several features that may be important in predicting the popularity of a book. Then the data was processed and analyzed. Several classification methods and text mining algorithms are also introduced.

2 Data Source

Barnes & Noble, Inc. a Fortune 500 company, is the largest retail bookseller in the United States, and the leading retailer of content, digital media and educational products in the country. The company operates 658 retail stores (as of August 2, 2014) in all 50 U.S. states in addition to 705 college bookstores that serve over 5 million students and more than 250,000 faculty members across the country. It contains detailed information of books and authors, so it is one of the best options for our experiments.

We fetch the webpages of books, parse the webpages so that we have fields including Title, Author, Price, Nook, Audio, Hardcover, Subject, Publisher, Published date, Pages, Number of reviews and Rating. These data will be used for feature extraction.

3 Feature Extraction

The dataset produced contains lots of features such as the title, author, hardcover (whether the book comes in hardcover), subjects (categories), number of reviews, and ratings (1 star to 5 stars). Some of the features are not that relevant to the popularity of the book (such as number of pages), while some may be strongly related (such as author). Therefore, it is important to firstly distinguish which features are potentially related to the quality of the book.

In this section, we discuss feature in three groups: basic features, subject features, and title features.

3.1 Basic Features

Author. Books written by famous writers can attract more attention than other books as these writers are expected to publish books with better quality. Thus the possibility of these books becoming popular and having better quality is much higher than any other books.

Price. Although the price is not completely relevant to the quality of a book, it will definitely influence the sales of it. If the price is too high, less people would think it is worth to spend such money on the book; while if the price is reasonable, more people would like to buy one.

However, the price is not strongly related to the popularity of a book. For example, a textbook is necessary for students and thus price is less likely to be affected. People may also find alternatives such as electronic edition for lower prices. In such case, price is not a relevant feature to evaluate the popularity of the book.

Publisher. Intuitively, the publisher of a book can quite affect its success. We therefore include the publisher information of books into the data set. In total, 259 publishers are included into our data set. When analyzing, we encode all the publishers with ids.

Hardcover. This feature describes whether the book comes in hardcover. Hardcover usually means higher price, but some people may be more interested in hardcover books as they are always in high quality and could be used as a gift. However, the hardcover might not have connections with the popularity/quality of a book. The hardcover will only affect the price a little, and people will not become interested in the book just because it comes in hardcover.

Number of Reviews. The number of review reflects how many people bought the book and are interested in leaving a comment on the website. It has more impact on the online buyers as people are attracted if the number of reviews is large, regardless of the reviews being positive or negative. This positive relation will make the book more popular when it already got some attention.

Rating. Our model is built based on the review data, and then rating data is used to verify the correctness of the prediction result of the model. The relation between rating and popularity may not be that strong, since a popular book will have more reviews and thus is more possible to receive negative reviews than other books.

Below we summarize the above-mentioned basic features in a table, along with the data types used when constructing our data set.

Table 1. Summary of basic features with feature descriptions and data types

Feature name	Description	Type
Price	price of the book sold online	numeric (float)
Nook	whether this book comes with Nook version	boolean
Audio	whether this book comes with Audio version	boolean
Hardcover	whether this book comes with Hardcover	boolean
Publisher	publisher of the book	categorical
Until 1206	days been published until December 6th, 2014	numeric (integer)
Pages	number of pages in the book	numeric (integer)
Number of reviews	number of reviews posted for this book on Barnes and Noble	numeric (integer)

3.2 Book subject

On Barnes and Nobles, books are classified into different subjects based on their contents, purposes, and audiences. We believe that due to the fact that for different type of books, the attracted audiences are different, how harsh they are when it comes to rating books will be different. For example, for recipes, most readers will enjoy the book as long as it is presented well with nice pictures and famous cooking show hosts. And this will lead to higher ratings, even if the recipes turn out not ideal (the readers might blame their own cooking skills). However, for academic books, the readers tend to be critical scholars, who will rate the books with more background knowledge and harsher criteria. Therefore, even though academic books in general take a long time to publish and have very high quality, they receive lower ratings.

We therefore are interested in looking into how books' subjects contribute to predicting books' ratings. However, we face two challenges when dealing with books' subjects: (1) for a single book, Barnes and Noble classifies it into multiple small categories, and (2) Barnes and Nobles include a large amount of subjects in its collection. These two facts will lead to an extremely long feature vector when considering the books' subjects. Because of this, filtering out minor subjects is necessary.

Before filtering, our data set includes 517 book subjects. We computed the number of books each subject has, and filter out all those subjects with less than 10 books. This results in total 122 subjects left in our data set, which are presented in the Appendix.

3.3 Book title

The title of a book is extremely important and thus writers tend to spend lots of time on thinking of an attractive title. Potential buyers can be attracted by the title either because of the title is funny or because it contains some keyword that they may be interested in.

However, the relation between the popularity and the book title might not be that straightforward as most will definitely go over the content of the book before they actually decide to buy the book.

To transform the book titles into features, we conduct the following steps:

1. Each book title is transformed into a bag-of-words (BOW) by applying the `WhiteSpaceTokenizer` from Natural Language Toolkit.
2. All letters in the bag-of-words are converted to lowercase to remove unnecessary confusion.
3. Each word is stemmed with `PorterStemmer` from Natural Language Toolkit.
4. Finally, all stop words are removed.

Fig. 1. Top 20 occurred words in collected book titles with corresponding occurrences

Word	Count	Word	Count
series	144	**sign**	34
life	91	**war**	32
book	87	**america**	29
story	66	**year**	28
guide	52	**live**	26
world	52	**true**	25
edit	51	**love**	25
bible	50	**history**	22
one	40	**day**	22
new	36	**study**	21

3.4 Preliminary feature selection

Before we start running machine learning algorithms, we want to conduct a simple test for checking the importance of features in terms of predicting ratings. To achieve this, a well known algorithm, ReliefF, is chosen. Due to the large dimension of subject and title features, at this moment we only examine the basic features. We use Weka to run the ReliefF test, and generate the following results:

Table 2. ReliefF results on basic features

Rank	Weights	Feature
1	0.0022093	Publisher
2	0.001711	Pages
3	0.0013039	Price
4	0.0007116	Hardcover
5	0.0000431	Audio
6	-0.0006221	Nook
7	-0.001979	Until 1206
8	-0.0031981	Number of reviews

It is interesting to see that for now, Publisher is the most important feature in terms of determining a book's rating. On the other hand, number of reviews is the least important.

4 Data Analysis

In this section, the features are extracted from the dataset, and their relations are analyzed. Relative features are collected for further analysis as they are more likely to affect the popularity of a book.

4.1 Number of books vs. Category

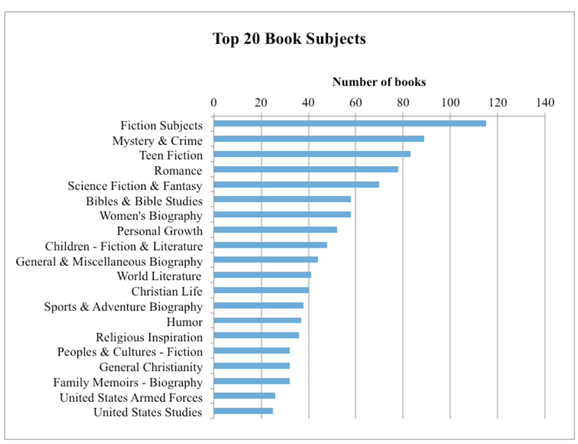


Fig. 2. Number of books vs. Subject.

Fig. 2 shows the number of books in each category. Some categories have relatively large quantities of books such as ‘Public Health & Safety’ and ‘Children History’, while some contain only several books such as ‘Espionage’ and ‘Asian American Studies’. The large difference of the number of books between different categories might results from the unbalanced dataset, while another more possible reason is that some categories do have much larger quantities of publications as these books are more popular. For example, the public may be more interested in Entertainment Biography than Military, and thus the number of Entertainment Biography books is much larger than that of Military books.

A larger and fairer dataset is made and analyzed as shown in Figure 2. The unbalanced distribution of total number of books demonstrates that people tend to be more interested in some categories.

4.2 Number of Reviews

Table 3. Summary of feature number of reviews

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.0	6.0	19.0	257.5	99.0	27760.0

The number of reviews of a book will seriously influence the prediction accuracy. More reviews eliminate the impact of biased comments and contains more minable words. **Fig. 3** shows the distribution of the number of reviews for each book. Most books have relatively small quantities of reviews, while some have extremely large amount of reviews. It is reasonable to conclude that a book is more popular if the book receives more reviews. Once a book attracts the public attention, it becomes much easier to get even more reviews. This positive correlation results in the exponential distribution of the reviews.

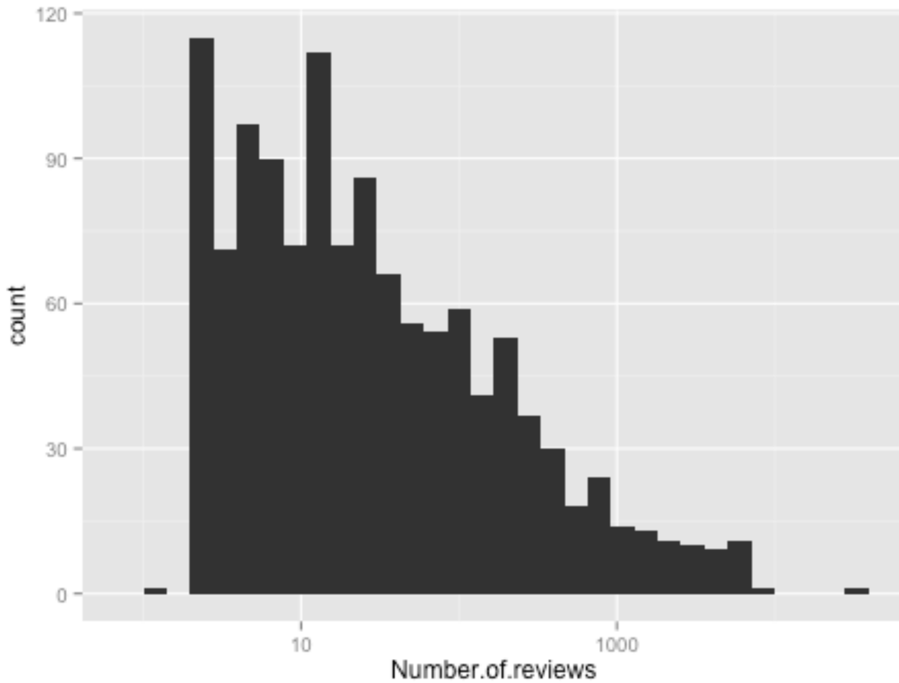


Fig. 3. Histogram of $\log_{10}$ Number of Reviews

4.3 Number of Reviews vs. Rating

Fig. 4 demonstrates the above conclusion that more reviews usually mean higher popularity. Most books that have large amount of reviews receive a rating of higher than 4 stars. It is not surprise that the number of reviews and rating are not strictly positive correlated as more reviews are more likely to contain more negative comments and ratings as well. Those books who receive only a few reviews are more likely to receive all 5 stars as only those people who are really interested in the book will buy and leave reviews.

Fig. 5 shows the distribution of ratings of the books in the dataset. 56% books receive a rating of equal to or higher than 4 stars, which makes the dataset not that balanced. However, this is still reasonable as the rating is too general. Most people will give a rating of around 4 stars even if they think the book is only around average. Extreme ratings such as 1 or 2 stars will only appear when the reviewer has a strong negative opinion of a book.

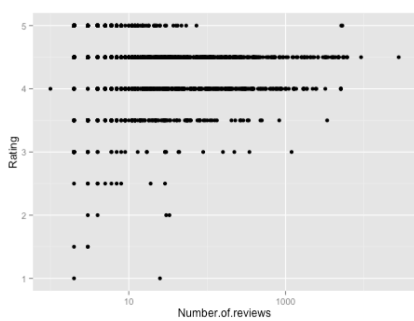


Fig. 4. $\log_{10}$ Number of Reviews vs. Ratings

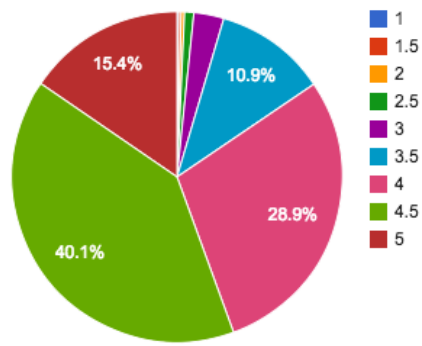


Fig. 5. Distribution of Ratings.

5 Regression and Classification Methods

This section briefly introduced the four methods we use in this project: Linear Regression, Decision Tree, Support Vector Machine, Random Forest and Multiclass SVMs.

5.1 Linear Regression

Linear regression is a statistical method to model the relationship between a dependent (or, target) variable y and one or more explanatory variables (data features) denoted X. It can be trained as a predictive model by an observed data set of y and X values. The predictive model has two parameters: weight w and offset b. Once, the weight w and offset b are learned, the fitted model can be used to make a prediction of the value of y given the associated X values.

Besides, given the variable y (target) and $X_1, \ldots, X_p$ (feature vectors) that may be related to y, linear regression analysis can be applied to measure the strength of the relationship between y and the X_j. It is often fitted using the least square approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit". In our project, because of not enough data records, we choose the ridge regression which minimize a penalized version of the least squares objective function (L2-norm penalty).

Therefore, the objective function used in this project is $\min_w \left||X \cdot w - y|\right|_2^{\ 2} + \propto \left||w|\right|_2^{\ 2}$. The regularization $\propto$ is determined by the cross validation. The coefficient of determination R^2 is computed for how well the data fit. Noted that R^2 is defined as $(1 - \frac{u}{v})$, where u is regression sum of squares and v is the residual sum of squares. R^2

measures the correlation between the features X and the target value y (book rating in our case).

Also, we try 4 sets of features: group1, group 2, group 3, and group 4. The evaluation is based on the mean absolute error (denoted as MAE, which is average of the difference between the predicted value y' and the target y.) Moreover, we proposed baselines model which predicts the mean value of the y for each prediction, where the y is the rating of the book. The baseline error to compare against is $mean(|y_i - mean(y)|)$.

5.2 Decision Tree

Decision tree is widely used as a predictive model which maps observations about an item to conclusions about the item's target value. A decision tree is a simple representation for classifying examples. Decision tree learning is one of the most successful techniques for supervised classification learning. For this section, assume that all of the features have finite discrete domains, and there is a single target feature called the classification. Each element of the domain of the classification is called a class. A decision tree or a classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes.

C4.5 algorithm is used to build the decision tree. It uses the concept of information entropy. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (which is the difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sublists.

As we learned in the class, the information gain uses the entropy to measure the node impurity. The entropy at a given node t can be computed using the following formula:

$$Entropy(t) = -\sum_{j} \mathrm{p(j|t)} \log \mathrm{p(j|t)}$$

The information gain is therefore computed using the following formula:

$$GAIN_{split} = Entropy(p) - \sum_{i=1}^{k} \frac{n_i}{n} Entropy(i)$$

The pseudocode of the C4.5 algorithm can be illustrated as follows:

```
Check for base cases
For each attribute a
    Find the normalized information gain ratio from splitting on a
```

4 Let a_best be the attribute with the highest normalized information gain
5 Create a decision node that splits on a_best
6 Recurse on the sublists obtained by splitting on a_best, and add those nodes as children of node

5.3 Random Forest

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. It is considered as one of the most powerful and accurate classification algorithms. The famous Netflix competition winner user this algorithm to win the first price. However, because of its sophisticated design, the running time is longer than most of the classification algorithms, which leads to the consequence that it is not suitable for real-time applications. Since in this project, our data set is relatively small compare to real-world data, we still choose Random Forest as one of the testing algorithms. And we expect to see it outperform all the other previous algorithms.

In our experiment, sci-kit learn from Python is used for running Random Forest classification.

5.4 Multiclass Support Vector Machine

In machine learning, support vector machines (SVMs) are supervised learning models that analyze data and recognize patterns. When given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.
However, in the scenario of our dataset, there are more than one classes for us to classify, so we are doing implementing multiclass SVMs, with no kernel. (i.e. linear classification)

In our experiments, we use svmtrain function provided by Matlab. And we implemented a new svmclassify_ function base on the svm regression values.

During the training process, we utilize the Divide and Conquer Strategy, according to different classes, and train different sub models.

During our classification, we utilize the Divide and Conquer Strategy, according to different classes. And after calculating the regression results of all sub models. We select the optimized classification results.

6 Experiments

In this section, we present our experiments of four algorithms and four feature sets. The four algorithms we use are: Linear Regression, Decision Tree, Support Vector Machine, and Random Forest. As for the feature sets, we use the following two combinations:

Group 1: basic features
Group 2: basic features + subject features
Group 3: basic features + title features
Group 4: basic features + subject features + title features

6.1 Linear Regression

The parameter $\propto$ is derived from cross validation is 1.0 in our model. The possible values of $\propto$ are [0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0]. Among these, the cross validation decides 1.0 as the result parameter value.

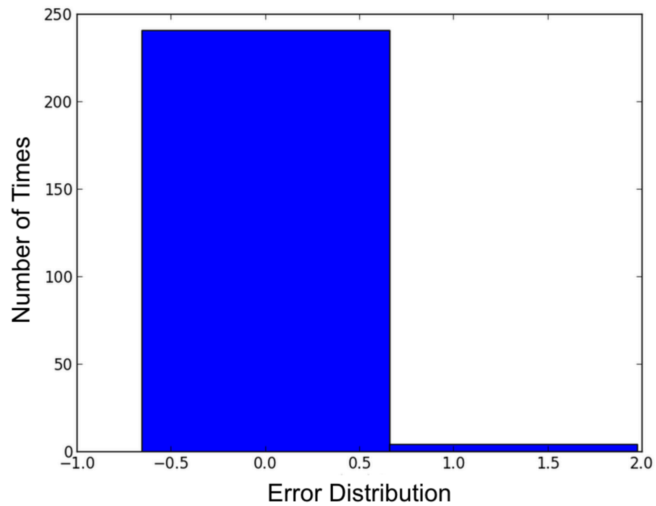
The R^2 is defined as (1-u/v), where u is regression sum of squares and v is the residual sum of squares. It can be served as the training score. The range of R^2 is between 0 and 1. This is similar to the correlation measurement. $R^2 = 1$ tells the features X has significant linear correlation with target y. On the other hand, $R^2 = 0$ tells the features X has no visible correlation with target y. The results are shown in table 1. The error distribution is presented in **Fig. 6**. The errors fall in the range of -0.5~ 0.5. The can be inferred that the trained model does not learn the from the rating differs in 0.5.

Our model has slightly smaller MAE as compared with the baseline error. However, the R^2 is around 0.14, and this tells that there is no significant linear relationship between our features and target. That is why the linear regression does not get good results.

Table 4. Linear regression result on four feature sets

Feature sets	R^2	MAE (Linear Regression)	MAE (baseline)
Group 1	0.01431	0.435200172	0.4450645564
Group 2	0.14612	0.439952463	0.4450645564
Group 3	0.17761	0.440125462	0.4450645564
Group 4	0.14832	0.439820017	0.4450645564

Fig. 6. Error distribution generated by Linear Regression



6.2 Decision Tree

In the first step, we need to handle several base cases. For example, if all the samples in the list belong to the same class, the algorithm simply creates a leaf node for the decision tree saying to choose that class. Also, if none of the features provide any information gain, C4.5 creates a decision node higher up the tree using the expected value of the class. Moreover, we need to decide when to terminate the building process of the decision tree. Since the rating is always the target we would like to predict, the rating feature is always at the leaf of the tree. If all the other features have been evaluated, or if the rest features are all in the same class, the program can simply terminates.

Although there are implementations in Weka and other data mining tools, we decide to write the program by ourselves so that the above-mentioned cases can be handled as well. The program is written in Python. It creates the decision tree using recursion method. Each time it selects the attribute with the highest information gain. If the attribute is rating, we select the second highest one as the rating should always be the last attribute to deal with.

Since decision tree only allows finite discrete domains for each attribute, some attributes for predicting book rating need to be preprocessed. For example, the 'Until1206' feature describes the number of dates from the publishing date to 12/06/2014. Since seldom books in the dataset are published in the same day, this attribute needs to be modified to be 'if the number of dates from the publishing date to 12/06/2014 is larger than a?', where a is a classification variable we need to define. In the first set of experiments, the decision tree we build is always a binary tree and thus we may set a to be the mean value of the 'Until1206' attribute. In the third set of experiment, the 'Until1206' attribute is divided into 4 ranges and thus half of the mean value is used to divide the attribute.

Similarly, several other attributes such as 'Price', 'Number of reviews', and 'Number of pages' are also processed in this way. As mentioned above, we encoded publishers into numbers and thus the 'Publisher' attribute should also be splitted into 2 (or 4) ranges.

Another problem we need to address is that there are some branches that might not be covered by the dataset. In such case, some nodes in the tree might only have one branch instead of two. When a test data arrives at this node and would like to go to the other branch (which does not exist), the model will know that the other branch does not exist and thus it will continue on the existing branch (Otherwise it will stuck on that node and the algorithm cannot terminate). The incompleteness of the training data introduces prediction error in the way that in the above case some attribute in the test data may be classified in a wrong way.

Fig. 7 illustrates a part of the decision tree we produced using the Group 1 features. Each node in the tree corresponds to a feature. Different nodes on the same level may differ (e.g. when pages=1, the node 'Nook' has two different children), and some nodes may have only one child (e.g. the leftmost Price node). Each attribute is divided into 2 groups. When a test data is evaluated by the model, it simply starts from the root node and goes down until it reaches a leaf node, which is the prediction rating.

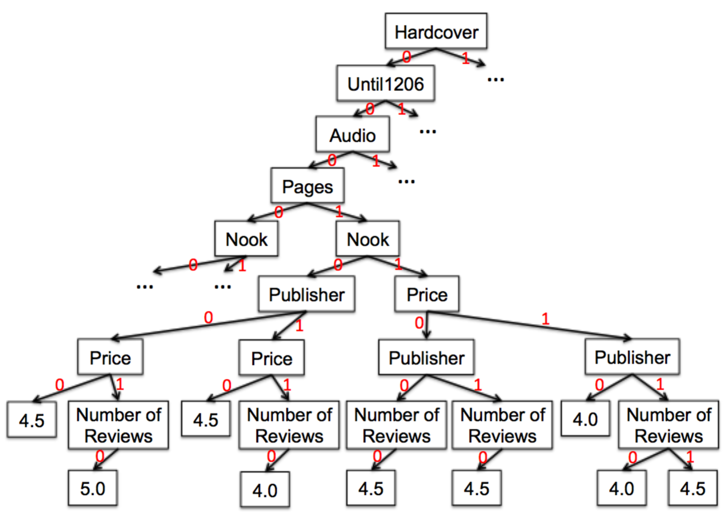


Fig. 7. A sample of (part of) the produced decision tree.

Four groups of attributes are evaluated using our model. **Table 5** shows the prediction result. Group 1 generates the best prediction accuracy and the MAE. Surprisingly, the model gives worse accuracy when adding more features into the data. As shown in Table I, the accuracy of Group 2 to Group 4 is slightly worse than that of Group 1. The reason might be due to the disturbance of irrelevant features. Different from Random Forest and many other methods, more features make the tree structure much more complicated and thus it will not perform well when the dataset is not big enough. Another reason might be the vagueness of data with middle rating (i.e. data with a rating of 3 or 3.5). Due to the subjectivity of people's review, the middle rating does not provide any useful information. If the rating is 1 (or 5), the confident of stating the book is not good (or good) is much higher than that of 3 and 3.5. For this reason, another set of experiment is performed to demonstrate the disturbance of vague data. All the data with rating of 3.5 is removed from the dataset, and Group 1 has been used to test the accuracy. As shown in **Table 6**, the prediction accuracy has been improved from 81.97% to 93.01%, demonstrating that the vagueness of the middle rating data hurts the prediction accuracy.

Table 5. I. Prediction result for Set 1.

	Prediction Accuracy		Mean Absolute Error (MAE)
	Direct	Fuzzy	
Group 1	40.98%	81.97%	0.431147540984
Group 2	33.44%	77.38%	0.501639344262
Group 3	34.31%	72.22%	0.513071895425
Group 4	32.68%	72.55%	0.553921568627

Table 6. II. Prediction result for Set 1 and Set 2.

Group 1	Prediction Accuracy		Mean Absolute Error (MAE)
	Direct	Fuzzy	
Set 1	40.98%	81.97%	0.431147540984
Set 2	44.12%	93.01%	0.356617647059

Table 7. III. Prediction result for Set 1 and Set 3.

Group 1	Prediction Accuracy		Mean Absolute Error (MAE)
	Direct	Fuzzy	
Set 1	40.98%	81.97%	0.431147540984
Set 3	41.64%	79.34%	0.444262295082

To evaluate if the binary split is good enough for the prediction model, another set of experiments is performed. In Set 3, for all the continuous attributes (or those with extremely large amount of distinct values), we divide each attribute into 4 ranges instead of 2. In this way, some of the nodes in the decision tree may have 4 branches instead of 2. **Table 7** shows the prediction result. For the direct way, the prediction accuracy increases slightly while it decreases for the fuzzy way. The MAE also slightly increases, meaning that the model generates slightly more errors. But generally speaking, these two decision trees perform similarly, and thus we may conclude that the binary split is good enough for the prediction model.

6.3 Random Forest

As seen in **Table** , when applying Random Forest on Group 2 feature set, we achieve the lowest MAE. However, the highest accuracy is achieved when applying Random Forest on Group 4 feature set.

Table 8. Result accuracy of Random Forest on four feature sets

	Accuracy		Mean Absolute Error (MAE)
	Direct	Fuzzy	
Group 1	36.41%	82.61%	0.44
Group 2	42.12%	82.07%	0.06
Group 3	43.21%	81.52%	0.44
Group 4	87.77%	97.55%	0.08

6.4 Multiclass Support Vector Machines

As seen in **Table** , when applying Multiclass SVMs on Group 2 feature set, we achieve the lowest MAE and accuracy (0.73 & 62.62) of Multiclass SVMs.

Table 9. Result accuracy of Multiclass SVM on four feature sets

	Accuracy		Mean Absolute Error (MAE)
	Direct	Fuzzy	
Group 1	18.03%	48.20%	1.02
Group 2	19.02%	55.08%	0.86
Group 3	18.36%	54.75%	0.86
Group 4	21.31%	62.62%	0.73

6.5 Comparison

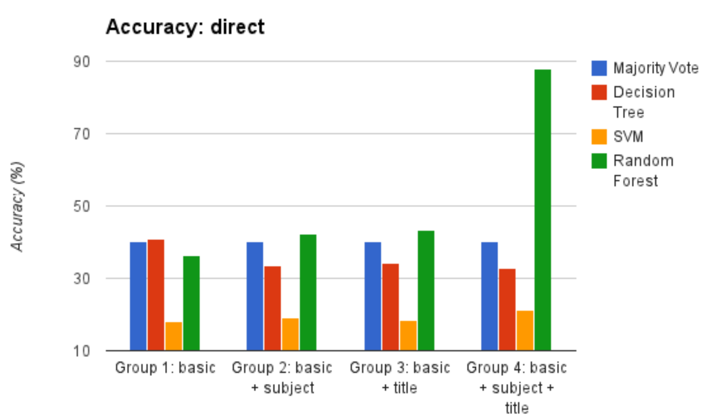


Fig. 8. Comparison on direct accuracy of three classification algorithms: Majority Vote, Decision Tree, and Random Forest

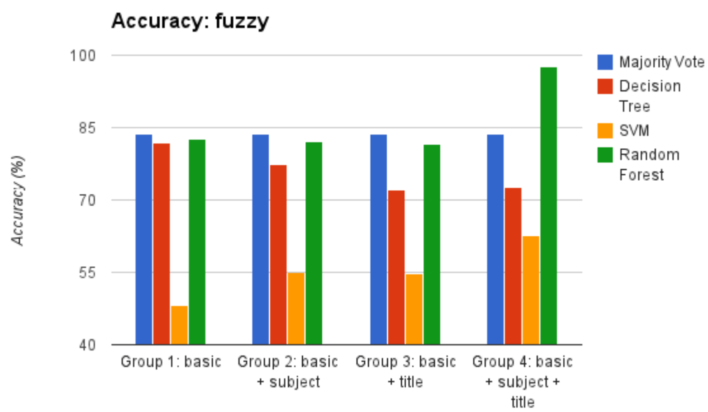


Fig. 9. Comparison on fuzzy accuracy of three classification algorithms: Majority Vote, Decision Tree, and Random Forest

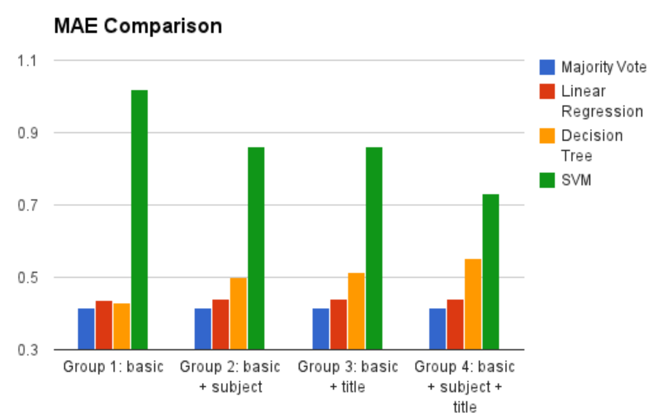


Fig. 10. Comparison on Mean Average Error of four algorithms: Majority Vote, Linear Regression, Decision Tree, and Random Forest

7 Conclusion

We collected data of 1224 books from one of the largest online bookstores, Barnes and Noble. Various experiments are run on different feature sets to see whether we can predict/capture books' ratings by looking at data other than the content of the books.

We found that it is possible to predict a book's rating on the online store without knowing the content. Four combinations of feature sets are tested: When considering the basic profile, the title, and the subjects of a book, Random Forest gives the best prediction of the book's rating

8 Reference

[1] Ashok V G, Feng S, Choi Y. Success with style: Using writing style to predict the success of novels [J]. Poetry, 2013, 580(9): 70.

[2] T. Hastie, R. Tibshirani and J. Friedman. Elements of Statistical Learning, Springer, 2009.

[3] Dumont M, Marée R, Wehenkel L, et al. Fast multi-class image annotation with random windows and multiple output randomized trees[C]//Proc. International Conference on Computer Vision Theory and Applications (VISAPP) Volume. 2009, 2: 196-203.

[4] Crammer, Koby, and Yoram Singer. "On the algorithmic implementation of multiclass kernel-based vector machines." The Journal of Machine Learning Research 2 (2002): 265-292.

9 Appendix

9.1 Part of the book subjects included in the data set

African American History
Americas - Travel Essays & Descriptions
Astronomy
Education Biography
Ethnic & Minority Studies - United States
Europe - Travel Essays & Descriptions
Historical Biography - United States
Latinos/Hispanics - Biography
Reference - Medicine
Sewing Quilting & Textile Arts
Sports - General & Miscellaneous
United States History - General & Miscellaneous
Film Biographies & Interviews
Football & Rugby
House & Home
Japanese History
Social Sciences - General & Miscellaneous
General & Miscellaneous
Crafts & Hobbies
Comics & Graphic Novels
Television Biography
General & Miscellaneous
Cooking
Teens - Biography
General & Miscellaneous
Religion

Politics & Government - General & Miscellaneous
Religious Biography
School Guides & Test Preparation
Science - General & Miscellaneous
Television
United States History - African American History
Educational Theory Research & History
Espionage
General Reference
Health - Diseases & Disorders
Political Theory & Ideology
U.S. Politics - History
Children - Biography
Computer Business & Culture
Dogs
Education - Social & Political Aspects
Emotional Healing
Jewish History
Music Biography
Parapsychology
Photography - History Criticism & Collections
Teaching & Teacher Training
United States History - 19th Century - General & Miscellaneous
Business - General & Miscellaneous
Family & School
Psychological Self-Help
Strategy & Weapons of War
Travel - General & Miscellaneous
U.S. Cooking
Biology & Life Sciences
Cooking for Special Diets
Peoples & Cultures - Biography
Roman Catholicism
Historical Biography
Pets
Business Life & Skills
Gardening
General & Miscellaneous
World History
Travel Essays & Descriptions
True Crime
African Americans - General & Miscellaneous
Aviation - Military
Baseball & Softball
Health
Legal Figures Law Enforcers & Criminals
Outdoor & Adventure Sports
Scientists - Biography
Scientists Inventors & Naturalists
American Fiction
Christian Biography
Clinical Medicine
Medical Figures
Fiction & Literature Classics
Needlework & Fiber Arts
Basketball
Business Biography
Comedy
Diet & Nutrition
US & Canadian Literary Biography
United States History - 20th Century - Wars & Conflict
Americans - Regional Biography
Military Biography
Patient Narratives
U.S. Politics in the Post Cold-War Era
United States History - 20th Century - 1945 to 2000
African American Biography
African American Biography & Memoir
Nature
United States Studies
Family - Assorted Topics
General & Miscellaneous
Military History
World War II
Peoples & Cultures - Fiction
Executive Branch
General Christianity
Religious Inspiration
World Literature
Decorating
Educational Settings
Biography
Inspiration