

# Supplementary Material

## Data processing

We exclude special books that are not representative of the general market. To do this, we filter the books based on the following criteria:

- Filter out books of a genre that has less than 100 books over the 8 year period from 2008 to 2015. Books in those small genres are usually not very recognizable to the general market.
- Filter out new editions of old books. Books may have special editions after the original publication or are reprinted if there is demand. Such reprints behave differently from new releases.

## AUC (Area Under Curve) and ROC (Receiver Operating Characteristic) curves

In our experiments, we employ  $k$ -fold cross-validation, for each fold, we train a model using the  $k - 1$  pieces and test on the remaining set of instances. The AUC score and the ROC curve are calculated based on the predictions on the test set. L2P outputs predicted outcomes and their rank within the test set can be computed. To compute the true-positive rate (TPR) and false-positive rate (FPR), we transform continuous outcomes and target variables into binary values by considering various thresholds between each test instance. Any instance predicted below and above the threshold considered as low- and high-sale books. Using the binary definitions, we can compute TPR and FPR for each threshold and obtains ROC curve. Then, we calculate the area under ROC as the AUC score (diagram see Figure 1).

## Model with More Complex Features

We also extract hybrid feature group for imprint and publication month. The additional feature groups are:

- **Month and Genre Cluster:** For each genre cluster, we obtain the book sales distribution for each month and include statistics of each distribution in the features.
- **Month and Topic:** For each topic, we obtain the book sales distribution for each month and corresponding statistics for each distribution. Then for each book, represented as a linear combination of several topics, the features are calculated as a weighted average of each statistic of each topic.
- **Imprint and Genre cluster:** For each genre cluster, we obtain the book sales distribution for each imprint and include statistics of each distribution in the features.

With those complex feature groups, we can obtain even better results compare to simple feature groups (see Table 1). In Figure 2 and Figure 3, we show the prediction scatter plot and ROC curve for the model with complex features. We can see that *Learning to Place* outperforms other methods and avoids underprediction for high-selling books.

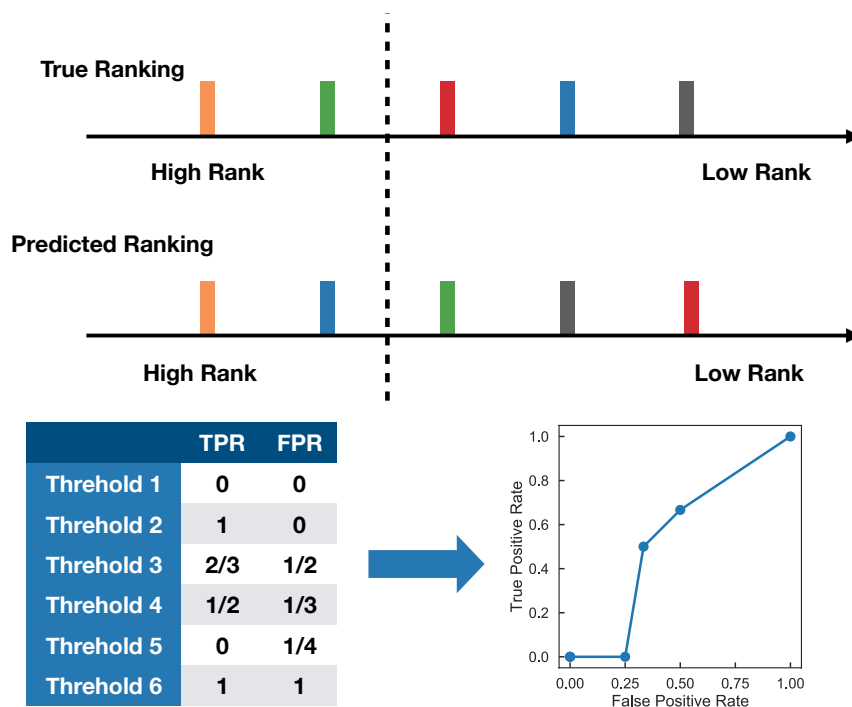


Figure 1: **The diagram demonstrating ROC calculation.** Setting thresholds between each rank, we calculate how many true positive and false positive on both side of the threshold. After obtaining the true positive rate and false positive rate for each threshold, we plot the ROC curve.

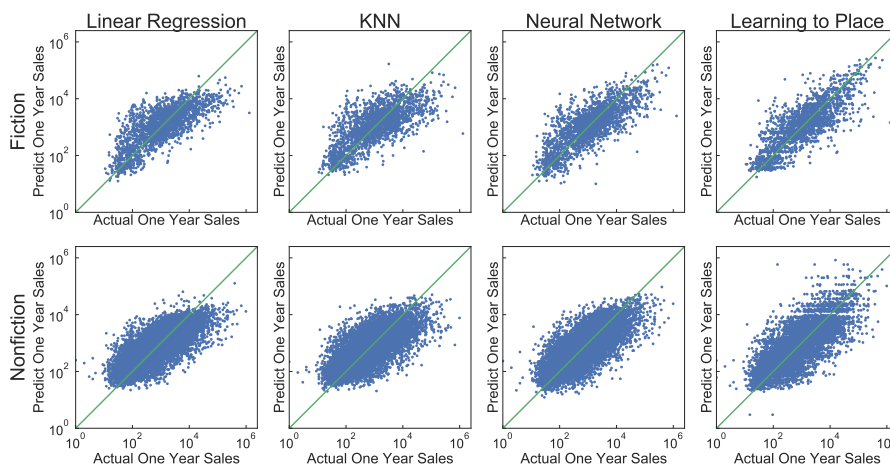


Figure 2: **Actual vs predicted One Year Sales for fiction and nonfiction books using complex features.** We observe similar underprediction on high-end for *Linear Regression*, *KNN* and *Nueral Network*, while *Learning to Place* has underprediction largely reduced.

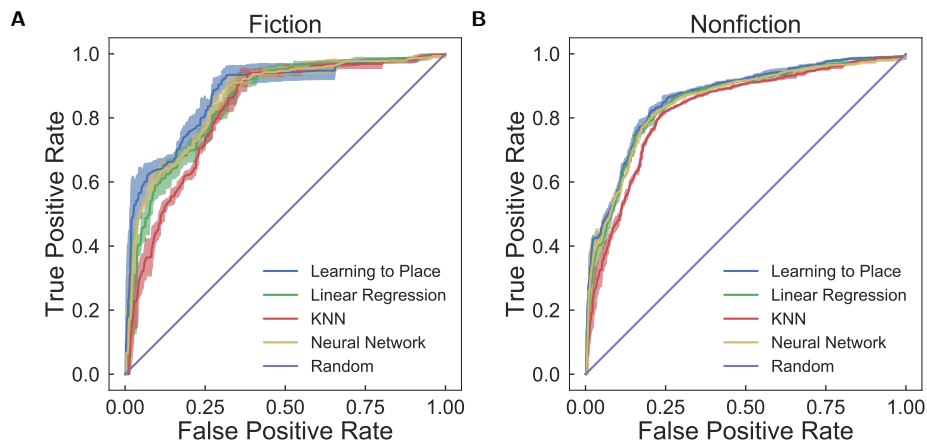


Figure 3: Models result of One Year Sale for fiction and nonfiction books, evaluated with ROC curve, using complex features. ROC curves for (A) fiction and (B) nonfiction books. Our Learning to Place approach and the various baseline methods all perform better than Random, and Learning to Place performs better than the KNN, Neural Network and Linear Regression. Band around the curve represents the standard deviation of the score across 5-fold cross validation.

		Method	AUC score	High-end RMSE	RMSE
Simple Features	Fiction	KNN	$0.83 \pm 0.03$	$0.71 \pm 0.07$	$0.38 \pm 0.02$
		LR	$0.85 \pm 0.01$	$0.60 \pm 0.08$	$0.34 \pm 0.02$
		NN	$0.86 \pm 0.03$	$0.50 \pm 0.12$	$0.33 \pm 0.02$
		L2P	<b><math>0.88 \pm 0.01</math></b>	<b><math>0.46 \pm 0.13</math></b>	<b><math>0.28 \pm 0.02</math></b>
	Nonfiction	KNN	$0.82 \pm 0.01$	$0.62 \pm 0.02$	$0.37 \pm 0.01$
		LR	$0.85 \pm 0.01$	$0.52 \pm 0.03$	$0.31 \pm 0.01$
		NN	$0.85 \pm 0.01$	$0.49 \pm 0.05$	$0.32 \pm 0.01$
		L2P	<b><math>0.86 \pm 0.01</math></b>	<b><math>0.43 \pm 0.01</math></b>	<b><math>0.30 \pm 0.01</math></b>
Complex Features	Fiction	KNN	$0.83 \pm 0.03$	$0.75 \pm 0.04$	$0.40 \pm 0.02$
		LR	$0.85 \pm 0.02$	$0.58 \pm 0.08$	$0.34 \pm 0.03$
		NN	$0.86 \pm 0.02$	$0.54 \pm 0.11$	$0.35 \pm 0.03$
		L2P	<b><math>0.88 \pm 0.04</math></b>	<b><math>0.44 \pm 0.10</math></b>	<b><math>0.27 \pm 0.02</math></b>
	Nonfiction	KNN	$0.83 \pm 0.01$	$0.62 \pm 0.05$	$0.36 \pm 0.01$
		LR	$0.85 \pm 0.01$	$0.50 \pm 0.02$	$0.31 \pm 0.01$
		NN	$0.85 \pm 0.01$	$0.50 \pm 0.06$	$0.34 \pm 0.01$
		L2P	<b><math>0.87 \pm 0.01</math></b>	<b><math>0.43 \pm 0.05</math></b>	<b><math>0.30 \pm 0.01</math></b>

Table 1: Measurement Table comparing performance of K-nearest neighbor (KNN), Linear Regression (LR), Neural Networks (NN), and Learning to Place (L2P) using complex features. We see that Learning to Place outperforms in every measure. We also report the usual RMSE scores and we still observe that Learning to Place has the best performance.

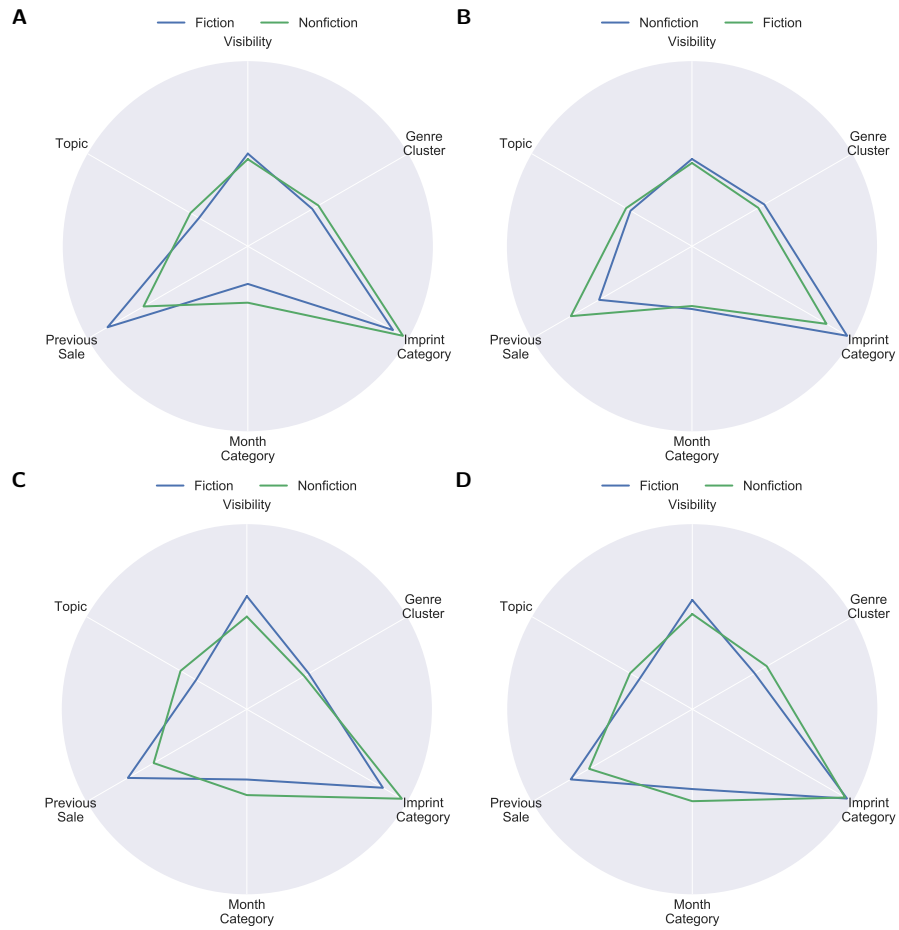


Figure 4: **Feature importance of Fiction and Nonfiction across different quarters.** Results of first quarter, second quarter, third quarter and fourth quarter respectively. We observe that feature importance is generally stable across different quarters.