Home Depot Competition

Zhaoxuan Ren
Yumeng Lu
Dan Sun
Tao Li
Yue Su

Overview

TASK: Predict the relevance of search results on homedepot.com

In the term project, the purpose is to build an appropriate model using certain methods and tools to predict the relevance of search results and improve the shopping experience of customers.

Method: Because of its relevance continuity,we decide to use regression analysis to deal with this problem.

Tools: R

Dataset Description

●**train.csv** - the training set, contains products, searches, and relevance scores

●**test.csv** - the test set, contains products and searches. You must predict the relevance for these pairs.

●**product_descriptions.csv** - contains a text description of each product.

●**attributes.csv** - provides extended information about a subset of the products (typically representing detailed technical specifications). Not every product will have attributes.

●**product_uid** - an id for the products

●**product_title** - the product title

●**product_description** - the text description of the product (may contain HTML content)

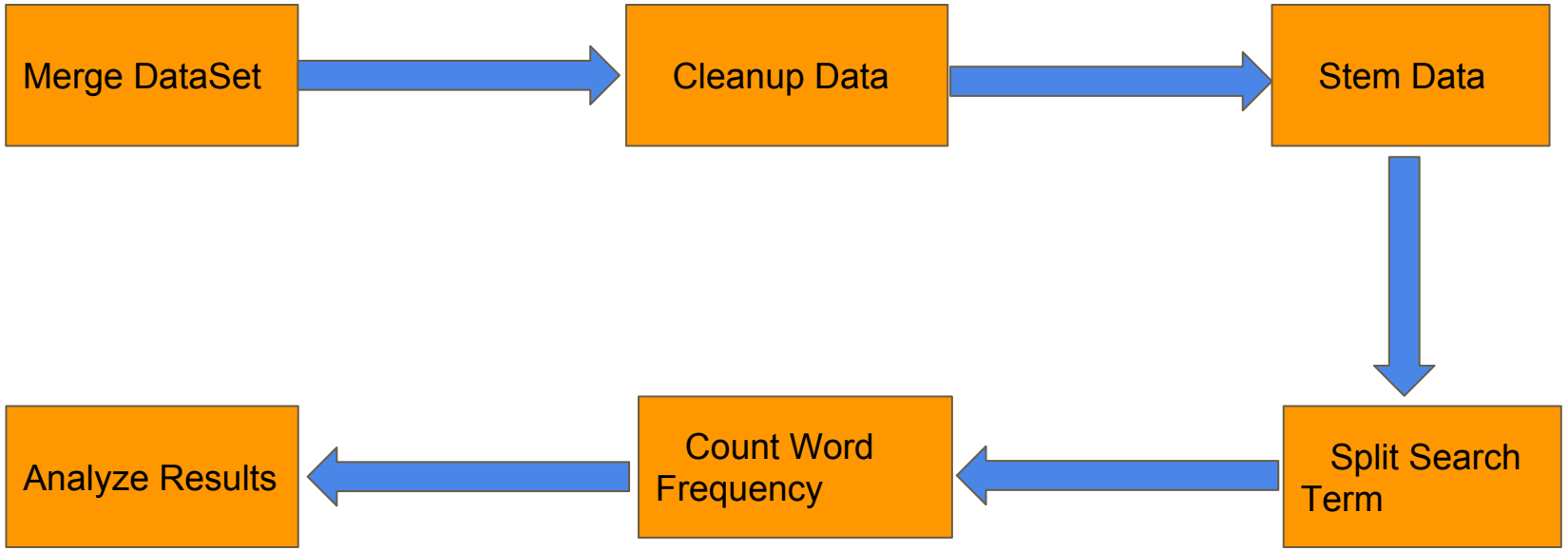
●**search_term** - the search query

●**material -** a kind of attributes

●**brand -** a kind of attributes

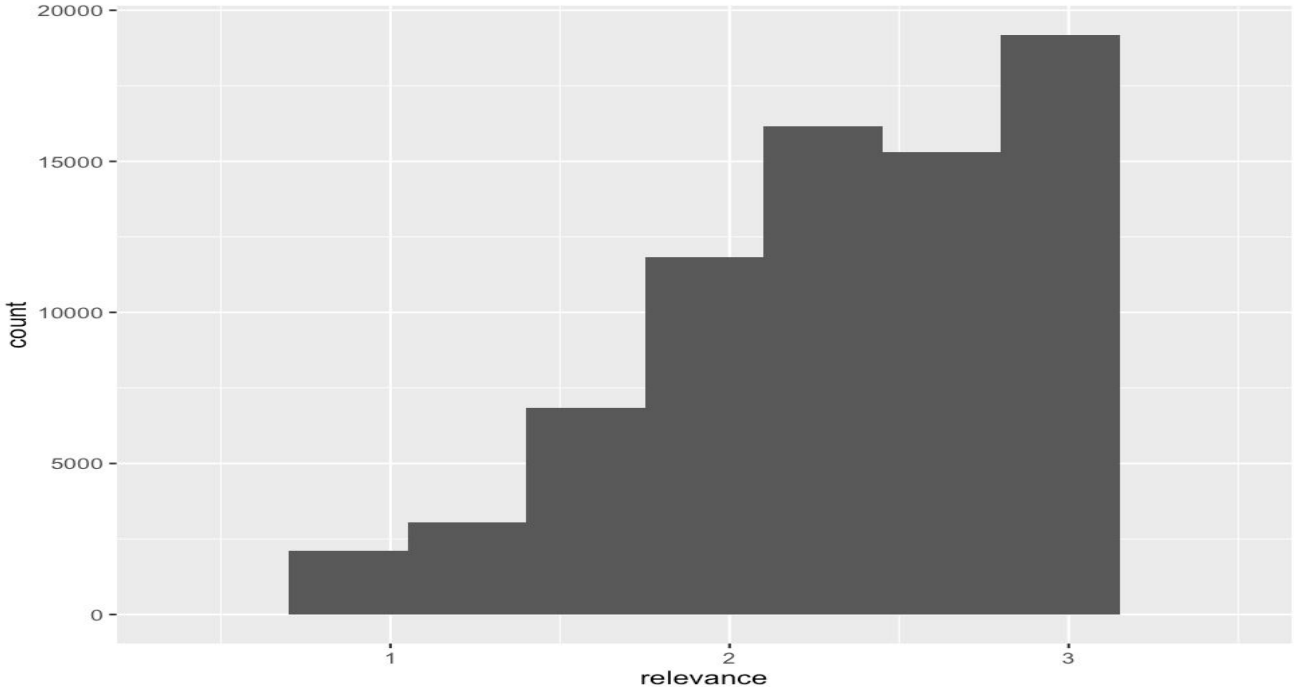
●**relevance** - the average of the relevance ratings for a given id

Data Process



Relevance Distribution

This picture indicates the relevance distribution.We can find out that according to the increase of relevance, the count number continually increase.



Analysis&Conclusion

This is the dataset after data processing

```
> train = read.csv('./finalTrain(wordRoot).csv')
> train[1,]
  X.1 nwords n_title n_desc n_brand n_material X product_uid id                    product_title   search_term
1   1      2       1      1       0          0 1      100001  2 Simpson Strong-Tie 12-Gauge Angle angle bracket
  relevance
1         3

product_description
1 Not only do angles make joints stronger, they also provide more consistent, straight corners. Simpson Strong-
Tie offers a wide variety of angles in various sizes and thicknesses to handle light-duty jobs or projects wher
e a structural connection is needed. Some can be bent (skewed) to match the project. For outdoor projects or th
ose where moisture is present, use our ZMAX zinc-coated connectors, which provide extra resistance against corr
osion (look for a "Z" at the end of the model number).Versatile connector for various 90 connections and home r
epair projectsStronger than angled nailing or screw fastening aloneHelp ensure joints are consistently straight
and strongDimensions: 3 in. x 3 in. x 1-1/2 in.Made from 12-Gauge steelGalvanized for extra corrosion resistanc
eInstall with 10d common nails or #9 x 1-1/2 in. Strong-Drive SD screws
              Brand         Material
1 Simpson Strong-Tie Galvanized Steel
```

Analysis&Conclusion

```
> weights <- chi.squared(relevance~., dataset)
> weights
             attr_importance
nwords            0.08322153
n_title           0.14742180
n_desc            0.10800737
n_brand           0.03715768
n_material        0.03926718
id                0.20753466
search_term       0.49586038
```

Analysis&Conclusion

Model1: Generalized Boosted Regression Models

RMSE: 0.49

```
#gbm model
gbm_model <- gbm.fit(train.set[,1:5],train.set$relevance,distribution = "gaussian",
test_relevance <- predict(gbm_model,test.set[,1:5],n.trees=600)
test_relevance <- ifelse(test_relevance>3,3,test_relevance)
test_relevance <- ifelse(test_relevance<1,1,test_relevance)
error1 = c(error1,(test_relevance-test.set$relevance)^2)
```

```
> sqrt(mean(error1))
[1] 0.488607
```

Model2: Random forest

RMSE:0.50

```
#rf model
rf_model <- randomForest(relevance~nwords + n_title + n_desc, data=train.set)
test_relevance <- predict(rf_model,test.set[,2:4])
test_relevance <- ifelse(test_relevance>3,3,test_relevance)
test_relevance <- ifelse(test_relevance<1,1,test_relevance)
error2 = c(error2,(test_relevance-test.set$relevance)^2)
```

```
> sqrt(mean(error2))
[1] 0.4955083
```

Thanks !