The final python notebook is homedepotsearch-final

The report for submission is report.pdf

To run the code in the notebook below prerequisites need to be installed.

1. Installation steps for genism

<https://radimrehurek.com/gensim/install.html>

easy\_install -U gensim

1. Install Inflect

Pip install inflect

1. import nltk

nltk.download()

1. Install fuzzywuzzy

pip install fuzzywuzzy

The train, test, product\_descriptions, attributes for this project is from the below kaggle competition

<https://www.kaggle.com/c/home-depot-product-search-relevance/data>

The files mentioned in file names that are output of different algorithms and feature combinations is in folder resultsouput

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Tuned Prameters | File name | Score | Rank |
| Benchmark |  |  | 0.5105 | 1681 |
| Cosine features with linear regression |  | python\_submisstion\_lr\_justcosinescores | 0.49293 | 1456 |
| Cosine features with decision tree |  | python\_submisstion\_dt\_justcosinescores | 0.49865 | 1528 |
| Cosine features with random forest tuning | {'max\_depth': 14,  'max\_features': 'auto',  'min\_samples\_split': 250,  'n\_estimators': 300} | python\_test\_submisstion\_rf\_justcosinescores | 0.48899 | 1377 |
| Cosine features and fuzzy wuzzy with linear regression |  | python\_submisstion\_lr\_cosine\_fuzzywuzzy | 0.48152 | 990 |
| Cosine features and fuzzy wuzzy with decision tree |  | python\_submisstion\_dt\_cosine\_fuzzywuzzy | 0.48571 | 1142 |
| Cosine features and fuzzy wuzzy with Random forest tuning | {'max\_depth': 10,  'max\_features': 'log2',  'min\_samples\_split': 500,  'n\_estimators': 200} | python\_test\_submisstion\_rf\_cosine\_fuzzywuzzy | 0.4699 | 342 |
| Linear Regression with all features including material |  | python\_submisstion\_lr\_material | 0.47901 | 932 |
| Decision tree with all features including material |  | python\_submisstion\_dt\_material | 0.48461 | 1091 |
| Random forest with more parameters to tune with all features excluding material features |  | python\_test\_submisstion\_rf\_moreparm\_nowordvecsim | 0.4699 | 342 |
| Gradient descent with more parameters to tune  with all features excluding material features |  | python\_test\_submisstion\_gd\_nowordvecsim | 0.4691 | 313 |
| Random forest with less parameters to tune  but with wordvecsim with all features excluding material features |  | python\_test\_submisstion\_rf\_nowordvecsim | 0.46954 | 330 |
| Random forest with less parameters to tune  but with wordvecsim with all features excluding material features |  | python\_test\_submisstion\_rf\_wordvecsim | 0.46932 | 321 |
| Random forest with more parameters to tune  with all features including material features | {'max\_depth': 14,  'max\_features': 'auto',  'min\_samples\_split': 250,  'n\_estimators': 400} | python\_test\_submisstion\_material\_rf\_wordvecsim | 0.46858 | 297 |
| Gradient descent with more parameters to tune and with all features including material features | {'max\_depth': 10,  'max\_features': 'log2',  'min\_samples\_split': 500,  'n\_estimators': 200} | python\_test\_submisstion\_gd\_material\_wordvecsim | 0.46751 | 264 |