本项目使用Anaconda3提供的虚拟环境作为开发环境，前、后端所需要的环境依赖相差较大。若只要运行前端，则只需满足前端所需的环境依赖即可。

# 前端

如何运行前端：

在终端中定位至源代码文件夹，运行

python ./SearchSystem/Main\_1.py

环境要求如下：

python=3.6.8

wxpython=4.0.4

pandas=0.25.0

numpy=1.16.4

pickle

# 后端

说明：

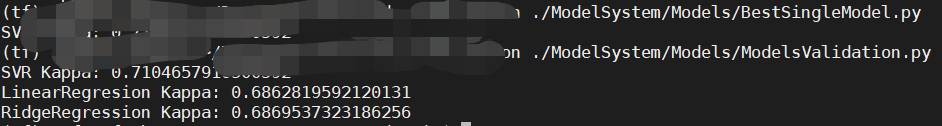
由于特征提取之后的文件较大，故提交物中已删去所有特征数据文件，仅保留根据特征生成的最后的训练、验证和测试集结果提交文件。

如何查看模型在验证集上的性能表现（Kappa系数）：(运行时间约10-15分钟)

在终端中定位至源代码文件夹，运行

python ./ModelSystem/Models/ModelsValidation.py

运行结果截图：



如何生成提交文件：（运行时间约20-30分钟）

在终端中定位至源代码文件夹，运行

python ./ModelSystem/Models/GenerateNormalSubmission.py.

如何查看原始的数据文件：

训练集：./ModelSystem/RawData/train.csv

测试集：./ModelSystem/RawData/test.csv

如何查看生成的提交文件（即测试集输出结果）：

./ModelSystem/Models/NormalSubmission\_final.csv

./ModelSystem/Models/NormalSubmission\_lir.csv

./ModelSystem/Models/NormalSubmission\_ridge.csv

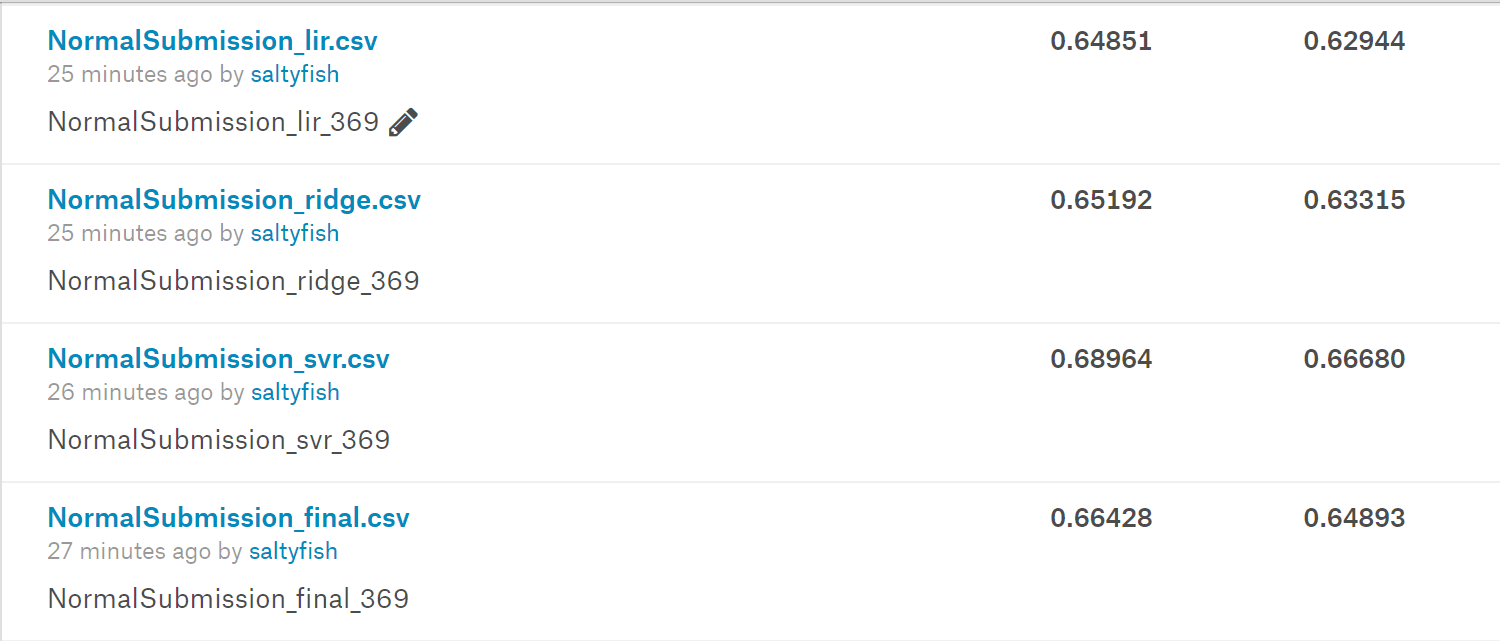
./ModelSystem/Models/NormalSubmission\_svr.csv

如何查看模型在测试集上的性能表现（Kappa系数）：

提交生成的提交文件至网站

<https://www.kaggle.com/c/crowdflower-search-relevance/>

这里附上网站的返回结果截图



若要运行上述.py，所需要的环境有：

python=3.6.8

pandas=0.25.0

numpy=1.16.4

pickle

scikit-learn=0.21.2

若要运行.py文件，进行数据预处理、特征工程、模型训练等整个完整的流程，则需时间4-5小时左右，其需要的环境配置如下，具体运行操作在**《用户手册》第三章 后端** 部分给出。

Anaconda3（Windows端）的环境配置如下：

name: kaggle

channels:

- https://mirrors.ustc.edu.cn/anaconda/cloud/msys2

- conda-forge

- https://mirrors.ustc.edu.cn/anaconda/pkgs/free

- https://mirrors.ustc.edu.cn/anaconda/cloud/conda-forge

- defaults

dependencies:

- absl-py=0.7.1=py36\_0

- asn1crypto=0.24.0=py36\_0

- attrs=19.1.0=py\_0

- backcall=0.1.0=py\_0

- blas=1.0=mkl

- bleach=3.1.0=py\_0

- boto=2.49.0=py36\_0

- boto3=1.9.162=py\_0

- botocore=1.12.194=py\_0

- bz2file=0.98=py36\_1

- ca-certificates=2019.5.15=1

- certifi=2019.6.16=py36\_1

- cffi=1.12.3=py36h7a1dbc1\_0

- chardet=3.0.4=py36\_1

- colorama=0.4.1=py\_0

- cryptography=2.7=py36h7a1dbc1\_0

- decorator=4.4.0=py\_0

- defusedxml=0.5.0=py\_1

- docutils=0.14=py36h6012d8f\_0

- entrypoints=0.3=py36\_1000

- gensim=3.4.0=py36hfa6e2cd\_0

- grpcio=1.16.1=py36h351948d\_1

- icc\_rt=2019.0.0=h0cc432a\_1

- idna=2.8=py36\_0

- intel-openmp=2019.4=245

- ipdb=0.12=py\_0

- ipykernel=5.1.1=py36h5ca1d4c\_0

- ipython=7.6.1=py36h5ca1d4c\_0

- ipython\_genutils=0.2.0=py\_1

- jedi=0.14.1=py36\_0

- jinja2=2.10.1=py\_0

- jmespath=0.9.4=py\_0

- joblib=0.13.2=py36\_0

- jsonschema=3.0.1=py36\_0

- jupyter\_client=5.3.1=py\_0

- jupyter\_core=4.4.0=py\_0

- kiwisolver=1.1.0=py36he980bc4\_0

- libprotobuf=3.8.0=h7bd577a\_0

- libsodium=1.0.17=h2fa13f4\_0

- m2w64-gcc-libgfortran=5.3.0=6

- m2w64-gcc-libs=5.3.0=7

- m2w64-gcc-libs-core=5.3.0=7

- m2w64-gmp=6.1.0=2

- m2w64-libwinpthread-git=5.0.0.4634.697f757=2

- markdown=3.1.1=py36\_0

- markupsafe=1.1.1=py36hfa6e2cd\_0

- mistune=0.8.4=py36hfa6e2cd\_1000

- mkl=2019.4=245

- mkl-service=2.0.2=py36he774522\_0

- mkl\_fft=1.0.12=py36h14836fe\_0

- mkl\_random=1.0.2=py36h343c172\_0

- msys2-conda-epoch=20160418=1

- nbconvert=5.5.0=py\_0

- nbformat=4.4.0=py\_1

- notebook=6.0.0=py36\_0

- numpy-base=1.16.4=py36hc3f5095\_0

- openssl=1.1.1c=he774522\_1

- pandas=0.25.0=py36ha925a31\_0

- pandoc=2.7.3=0

- pandocfilters=1.4.2=py\_1

- parso=0.5.1=py\_0

- pickleshare=0.7.5=py36\_1000

- pip=19.1.1=py36\_0

- prometheus\_client=0.7.1=py\_0

- prompt\_toolkit=2.0.9=py\_0

- protobuf=3.8.0=py36h33f27b4\_0

- pycparser=2.19=py36\_0

- pygments=2.4.2=py\_0

- pyopenssl=19.0.0=py36\_0

- pyrsistent=0.15.3=py36hfa6e2cd\_0

- pysocks=1.7.0=py36\_0

- python=3.6.8=h9f7ef89\_7

- python-dateutil=2.8.0=py\_0

- pytz=2019.1=py\_0

- pywinpty=0.5.5=py36\_1000

- requests=2.22.0=py36\_0

- s3transfer=0.2.0=py36\_0

- scikit-learn=0.21.2=py36h6288b17\_0

- scipy=1.3.0=py36h29ff71c\_0

- send2trash=1.5.0=py\_0

- setuptools=41.0.1=py36\_0

- six=1.12.0=py36\_1000

- smart\_open=1.8.4=py\_0

- sqlite=3.28.0=he774522\_0

- terminado=0.8.2=py36\_0

- testpath=0.4.2=py\_1001

- tornado=6.0.3=py36hfa6e2cd\_0

- traitlets=4.3.2=py36\_1000

- urllib3=1.24.2=py36\_0

- vc=14.1=h0510ff6\_4

- vs2015\_runtime=14.15.26706=h3a45250\_4

- wcwidth=0.1.7=py\_1

- webencodings=0.5.1=py\_1

- werkzeug=0.15.4=py\_0

- wheel=0.33.4=py36\_0

- win\_inet\_pton=1.1.0=py36\_0

- wincertstore=0.2=py36h7fe50ca\_0

- winpty=0.4.3=4

- wxpython=4.0.4=py36ha925a31\_0

- zeromq=4.3.2=h6538335\_2

- zlib=1.2.11=h62dcd97\_3

- pip:

- keras==2.2.4

- matplotlib==3.1.0

- numpy==1.16.4

- pillow==6.0.0

- pyqt5==5.13.0

- pyqt5-sip==4.19.18

- pyzmq==18.0.2

- tensorboard==1.12.2

prefix: D:\Anaconda3\envs\kaggle