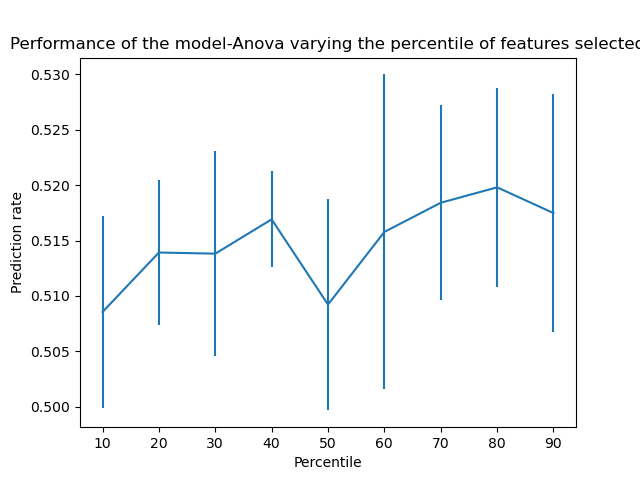
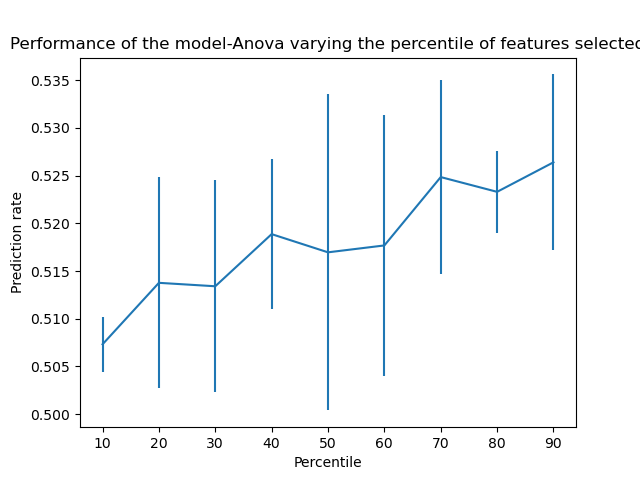
基于随机森林和XGBoost的多因子选股

将处理后的因子数据作为特征，使用ANOVA方差分析选取最有分辨率的百分比因子作为保留特征，分别选取10%到100%的因子特征进行分析，两个模型分别得到的结果如下：



XGBoost



RandomForest

## 参数训练：

在参数训练中，原博客使用了ROC AUC与F1 Score两个评价指标衡量了模型的准确性，F1是Presicion（是正类并且被预测为正类的样本占所有预测为正类的比例）与Recall（实际是正类并且被预测为正类的样本占所有实际为正类样本的比例）的调和均值，ROC AUC在实际数据中出现不平衡现象时也不会改变，两者可以结合使用。

（其实这里感觉用F1有点怪怪的，因为感觉这里的分类问题应该更看重Presicion）

### 随机森林：

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5086 0.5146 0.5313 0.5601 0.5271 0.5283 0.0200

f1 0.4621 0.5068 0.5077 0.4794 0.5037 0.4919 0.0203

------------------------------------------------------------

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5167 0.5142 0.5282 0.5678 0.5230 0.5300 0.0219

f1 0.4627 0.5181 0.4757 0.5151 0.5212 0.4985 0.0273

------------------------------------------------------------

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5103 0.5146 0.5276 0.5734 0.5274 0.5307 0.0251

f1 0.4653 0.5259 0.4774 0.4912 0.5248 0.4969 0.0275

------------------------------------------------------------

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5122 0.5140 0.5259 0.5686 0.5218 0.5285 0.0231

f1 0.4621 0.5349 0.4776 0.4992 0.5254 0.4998 0.0308

------------------------------------------------------------

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5125 0.5168 0.5250 0.5747 0.5231 0.5304 0.0253

f1 0.4641 0.5208 0.4722 0.5006 0.5241 0.4964 0.0274

------------------------------------------------------------

RandomForestClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5131 0.5152 0.5253 0.5712 0.5248 0.5299 0.0237

f1 0.4631 0.5159 0.4730 0.4992 0.5226 0.4947 0.0260

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5158 0.5171 0.5203 0.5703 0.5118 0.5270 0.0243

f1 0.4671 0.4999 0.5000 0.4919 0.5335 0.4985 0.0238

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5202 0.5163 0.5228 0.5664 0.5277 0.5307 0.0204

f1 0.4712 0.5179 0.4870 0.5053 0.5256 0.5014 0.0223

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5138 0.5128 0.5229 0.5715 0.5275 0.5297 0.0242

f1 0.4782 0.5128 0.4807 0.4905 0.5165 0.4957 0.0179

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5158 0.5172 0.5298 0.5730 0.5258 0.5323 0.0235

f1 0.4887 0.5199 0.4680 0.5242 0.5253 0.5052 0.0257

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5159 0.5161 0.5266 0.5717 0.5284 0.5318 0.0231

f1 0.4730 0.5044 0.4763 0.5143 0.5252 0.4986 0.0231

------------------------------------------------------------

RandomForestClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5165 0.5158 0.5269 0.5748 0.5264 0.5321 0.0245

f1 0.4812 0.5171 0.4791 0.5064 0.5246 0.5017 0.0207

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5146 0.5120 0.5279 0.5716 0.5242 0.5300 0.0241

f1 0.4979 0.5158 0.4806 0.5070 0.5268 0.5056 0.0176

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5155 0.5175 0.5248 0.5718 0.5286 0.5316 0.0231

f1 0.4816 0.4900 0.4755 0.5148 0.5176 0.4959 0.0193

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5206 0.5152 0.5248 0.5723 0.5268 0.5319 0.023

f1 0.4923 0.5104 0.4761 0.5020 0.5167 0.4995 0.016

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5137 0.5126 0.5236 0.5719 0.5274 0.5298 0.0244

f1 0.4996 0.5127 0.4810 0.5071 0.5160 0.5033 0.0139

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5196 0.5162 0.5269 0.5755 0.5262 0.5329 0.0242

f1 0.4997 0.5132 0.4784 0.5073 0.5237 0.5045 0.0170

------------------------------------------------------------

RandomForestClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5175 0.5161 0.5247 0.5735 0.5267 0.5317 0.0238

f1 0.4856 0.5062 0.4791 0.5115 0.5191 0.5003 0.0172

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5135 0.5138 0.5236 0.5596 0.5266 0.5274 0.0189

f1 0.4924 0.4864 0.4830 0.5158 0.5156 0.4986 0.0159

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5213 0.5119 0.5292 0.5697 0.5310 0.5326 0.0221

f1 0.4921 0.5147 0.4848 0.5257 0.5164 0.5067 0.0174

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5200 0.5094 0.5269 0.5710 0.529 0.5313 0.0235

f1 0.5102 0.5056 0.4889 0.5141 0.507 0.5052 0.0097

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5213 0.5141 0.5236 0.5726 0.5306 0.5325 0.0232

f1 0.4983 0.5090 0.4874 0.5099 0.5268 0.5063 0.0147

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5166 0.5147 0.5257 0.5720 0.5288 0.5316 0.0234

f1 0.5000 0.5007 0.4785 0.5062 0.5106 0.4992 0.0123

------------------------------------------------------------

RandomForestClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5191 0.5171 0.5228 0.5741 0.5288 0.5324 0.0238

f1 0.5018 0.5007 0.4840 0.5137 0.5212 0.5043 0.0142

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5192 0.5201 0.5161 0.5583 0.5213 0.5270 0.0176

f1 0.5127 0.4898 0.4834 0.5140 0.5345 0.5069 0.0206

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.518 0.5219 0.5192 0.5663 0.529 0.5309 0.0203

f1 0.505 0.5087 0.4868 0.5163 0.504 0.5042 0.0108

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5228 0.5149 0.5230 0.5703 0.5292 0.5320 0.0220

f1 0.4982 0.5188 0.4889 0.5249 0.5201 0.5102 0.0157

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5197 0.5128 0.5227 0.5756 0.5313 0.5324 0.0251

f1 0.4974 0.5079 0.4782 0.5076 0.5170 0.5016 0.0148

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5211 0.5161 0.5246 0.5739 0.5288 0.5329 0.0234

f1 0.5056 0.5009 0.4839 0.5150 0.5206 0.5052 0.0142

------------------------------------------------------------

RandomForestClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5215 0.5182 0.5244 0.5735 0.5300 0.5336 0.0228

f1 0.5128 0.5026 0.4924 0.5238 0.5142 0.5092 0.0120

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5128 0.5151 0.5254 0.5474 0.5331 0.5268 0.0141

f1 0.5089 0.5125 0.4796 0.5162 0.5214 0.5077 0.0164

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5217 0.5166 0.5249 0.5685 0.5240 0.5312 0.0211

f1 0.5032 0.5141 0.4760 0.5120 0.5185 0.5047 0.0170

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5209 0.5187 0.5242 0.5653 0.5329 0.5324 0.0192

f1 0.5069 0.5040 0.4927 0.5199 0.5246 0.5096 0.0128

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5186 0.5133 0.5187 0.5676 0.5333 0.5303 0.0221

f1 0.5090 0.5059 0.4829 0.5168 0.5157 0.5061 0.0137

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5197 0.5157 0.5237 0.5723 0.5321 0.5327 0.0229

f1 0.5141 0.4996 0.4908 0.5193 0.5176 0.5083 0.0125

------------------------------------------------------------

RandomForestClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5238 0.5191 0.5241 0.5745 0.5348 0.5353 0.0227

f1 0.5098 0.5049 0.4866 0.5197 0.5187 0.5079 0.0134

------------------------------------------------------------------------------------------------------------------------

### XGBoost

XGBClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5114 0.5024 0.5153 0.5464 0.5357 0.5223 0.0182

f1 0.5079 0.5023 0.5123 0.5107 0.5180 0.5102 0.0058

------------------------------------------------------------

XGBClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5147 0.4999 0.5152 0.5425 0.5319 0.5209 0.0165

f1 0.5040 0.5112 0.5109 0.5049 0.5211 0.5104 0.0068

------------------------------------------------------------

XGBClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.4979 0.5140 0.5214 0.5611 0.5338 0.5256 0.0237

f1 0.5255 0.5242 0.5116 0.5357 0.4962 0.5186 0.0152

------------------------------------------------------------

XGBClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5058 0.5133 0.5190 0.5549 0.5227 0.5231 0.0189

f1 0.5137 0.5264 0.5112 0.5186 0.4849 0.5109 0.0157

------------------------------------------------------------

XGBClassifier set max\_depth: 3

0 1 2 3 4 mean std

roc\_auc 0.5053 0.5064 0.5173 0.5524 0.5374 0.5238 0.0205

f1 0.5227 0.5029 0.5071 0.5227 0.5097 0.5130 0.0092

------------------------------------------------------------

XGBClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5050 0.5106 0.5176 0.5502 0.5231 0.5213 0.0175

f1 0.4986 0.5163 0.5064 0.5241 0.5119 0.5115 0.0097

------------------------------------------------------------

XGBClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5151 0.5016 0.5150 0.5510 0.5248 0.5215 0.0184

f1 0.5193 0.5174 0.5067 0.5026 0.5044 0.5101 0.0077

------------------------------------------------------------

XGBClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5102 0.5172 0.5159 0.5390 0.5269 0.5218 0.0113

f1 0.5094 0.5113 0.5079 0.5056 0.4873 0.5043 0.0097

------------------------------------------------------------

XGBClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5063 0.5159 0.5275 0.5480 0.5320 0.5259 0.0159

f1 0.5158 0.5137 0.5141 0.5099 0.4918 0.5091 0.0099

------------------------------------------------------------

XGBClassifier set max\_depth: 4

0 1 2 3 4 mean std

roc\_auc 0.5106 0.5081 0.5208 0.5445 0.5380 0.5244 0.0162

f1 0.5052 0.5060 0.5046 0.5138 0.5031 0.5065 0.0042

------------------------------------------------------------

XGBClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5179 0.4955 0.5292 0.5341 0.5365 0.5227 0.0168

f1 0.5201 0.5020 0.5196 0.5072 0.5212 0.5140 0.0088

------------------------------------------------------------

XGBClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5045 0.5040 0.5177 0.5446 0.5292 0.52 0.0173

f1 0.5096 0.5119 0.5013 0.5125 0.5145 0.51 0.0052

------------------------------------------------------------

XGBClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5126 0.5016 0.5411 0.5400 0.5284 0.5247 0.0173

f1 0.5097 0.5050 0.5197 0.5127 0.5118 0.5118 0.0053

------------------------------------------------------------

XGBClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5054 0.5140 0.5264 0.5297 0.5316 0.5214 0.0113

f1 0.5131 0.5312 0.5175 0.5121 0.4993 0.5146 0.0115

------------------------------------------------------------

XGBClassifier set max\_depth: 5

0 1 2 3 4 mean std

roc\_auc 0.5114 0.5099 0.5254 0.5491 0.5339 0.5259 0.0163

f1 0.5120 0.5186 0.5194 0.5259 0.4972 0.5146 0.0109

------------------------------------------------------------

XGBClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5089 0.5205 0.5201 0.5343 0.5241 0.5216 0.0091

f1 0.4934 0.5262 0.5122 0.5049 0.5206 0.5115 0.0130

------------------------------------------------------------

XGBClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5233 0.5057 0.5314 0.5461 0.5270 0.5267 0.0146

f1 0.5263 0.5093 0.5153 0.5158 0.5138 0.5161 0.0062

------------------------------------------------------------

XGBClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5156 0.5149 0.5294 0.5330 0.5214 0.5229 0.0081

f1 0.5211 0.5218 0.5120 0.5206 0.5109 0.5173 0.0054

------------------------------------------------------------

XGBClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5094 0.5118 0.5319 0.5410 0.5372 0.5263 0.0147

f1 0.5011 0.5120 0.5216 0.5062 0.5169 0.5115 0.0082

------------------------------------------------------------

XGBClassifier set max\_depth: 6

0 1 2 3 4 mean std

roc\_auc 0.5115 0.5177 0.5345 0.5355 0.5393 0.5277 0.0123

f1 0.5062 0.5236 0.5169 0.5101 0.5174 0.5148 0.0068

------------------------------------------------------------

XGBClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5181 0.5154 0.5136 0.5384 0.5303 0.5232 0.0107

f1 0.5140 0.5097 0.5072 0.5111 0.5067 0.5098 0.0030

------------------------------------------------------------

XGBClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5190 0.5061 0.5084 0.5460 0.5242 0.5207 0.0160

f1 0.5213 0.5141 0.4940 0.5016 0.4933 0.5049 0.0124

------------------------------------------------------------

XGBClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5145 0.5083 0.5206 0.5437 0.5324 0.5239 0.0142

f1 0.5025 0.5051 0.4981 0.5206 0.5016 0.5056 0.0087

------------------------------------------------------------

XGBClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5159 0.5063 0.5294 0.5433 0.5233 0.5236 0.0139

f1 0.5202 0.5063 0.5126 0.5201 0.4887 0.5096 0.0130

------------------------------------------------------------

XGBClassifier set max\_depth: 7

0 1 2 3 4 mean std

roc\_auc 0.5191 0.5142 0.5186 0.5336 0.5303 0.5231 0.0083

f1 0.5150 0.5119 0.5095 0.5057 0.5049 0.5094 0.0042

------------------------------------------------------------

XGBClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.4985 0.4964 0.5232 0.5377 0.5121 0.5136 0.0173

f1 0.4928 0.5024 0.5141 0.5194 0.4980 0.5053 0.0111

------------------------------------------------------------

XGBClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5146 0.5114 0.5207 0.5296 0.5279 0.5208 0.008

f1 0.5000 0.5101 0.4997 0.5056 0.5265 0.5084 0.011

------------------------------------------------------------

XGBClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5114 0.5175 0.5285 0.5413 0.5337 0.5265 0.0121

f1 0.4982 0.5156 0.5182 0.5176 0.5111 0.5121 0.0083

------------------------------------------------------------

XGBClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5029 0.5119 0.5252 0.5454 0.5263 0.5223 0.0161

f1 0.4837 0.5072 0.5224 0.5184 0.5014 0.5066 0.0154

------------------------------------------------------------

XGBClassifier set max\_depth: 8

0 1 2 3 4 mean std

roc\_auc 0.5179 0.5108 0.5315 0.5479 0.5371 0.5290 0.0148

f1 0.5173 0.5209 0.5212 0.5225 0.5115 0.5187 0.0045

------------------------------------------------------------