# 一、重要性分析（开发大表.xlsx）

* **数据描述**

**样本量**：242（Y选择不同，数量可能不同）

**Y**: 90天累产、首年产气量、第二年累产气量、第三年累产气量、第四年累产气量、第五年累产气量、第六年累产气量

**X:** ['I类储层厚度(m)','I类储层连续厚度(m)','龙一11厚度\n(m)','龙一11 (1类厚度)(m)','龙一11水平长度(m)','龙一11压力系数','龙一11钻遇率(%)','龙一11底以上4米箱体钻遇率（%）','水平\n段长\n(m)','实际压\n裂段长(m)','主体单段簇数','单段主体孔数','簇间距(m)','主体射孔密度(P/m)','加砂\n强度(t/m)','用液\n强度(m3/m)','分段段长(m)','是否\n套变', '合压\n长度\n(m)','丢段\n长度(m)','优质页岩钻遇率（%）','设计压\n裂段长(m)','实际\n射孔\n簇数','折算\n有效\n段数','100目石英砂(t)','40/70目石英砂(t)','40/70目陶粒(t)','30/50目陶粒(t)','总砂量(t)','酸液\n(m3)','滑溜水\n(m3)','线性胶\n(m3)','弱凝胶\n(m3)','总液量(m3)','平均单\n段砂量\n(t)','平均单\n段液量\n(m3)','陶粒\n比例\n（%）','平均停\n泵压力\n(MPa)','排量\n（m3/min）','压裂结束停泵压力(MPa)','焖井时间(天)','开始排液时井口套压\n(MPa)','投产前累计\n排液量(m3)','投产前返排率(%)','井间距(m)']

**去除特征：**'方案','统计时间','投产时间','目前累产气量\n万方（23.6.30）', '生产时间','首年日产（104m3/d）','平均首年日产(104m3/d)','平均首年日产(104m3/d)', '气井分类','甜点分区','研究所复核储量丰度-2023.6.30(108m3/km2)','研究所复核井控储量23.5(108m3)', '研究所复核采收率','测试产量(104m3/d)','EUR(亿方）\n(研究所-23.6)'

* **方法：**通过树模型分析特征重要性

**原理：**机器学习中的树模型分析特征重要性的原理主要是基于特征在决策过程中的作用。

在决策树模型中，每个特征的重要性可以通过计算特征在决策树中的使用次数来衡量。具体来说，如果一个特征被选为分割点的次数越多，那么这个特征的重要性就越强。这个理念可以被推广到集成算法中，只要将每棵树的特征重要性进行简单的平均即可。

另外，随机森林模型也是一种常用的决策树模型，它通过构建多个决策树并取其输出的平均值来提高预测精度。在随机森林中，每个特征的重要性可以通过计算每个特征在所有决策树中的使用次数并取平均值来得到。这个使用次数可以是通过计算特征在决策树中的分裂次数、信息增益等指标来得到的。

总的来说，机器学习中的树模型分析特征重要性的原理主要是基于特征在决策过程中的作用，通过计算特征的使用次数、使用深度以及信息增益等指标来评估特征的重要性。

* **结果**

**90天累产(242)**

龙一11 (1类厚度)(m): 0.23735086574438005

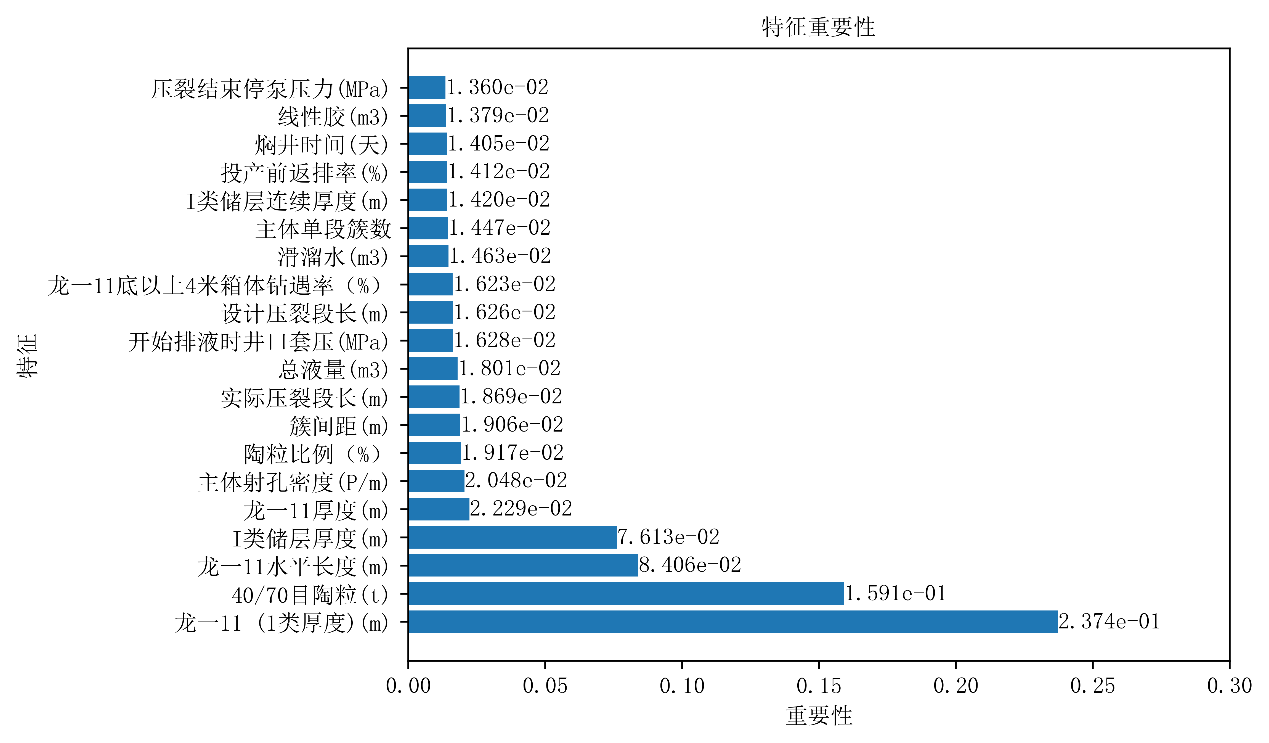
40/70目陶粒(t): 0.15913630947448792

龙一11水平长度(m): 0.08405732198728065

I类储层厚度(m): 0.07612976098668854

龙一11厚度(m): 0.022287257898480458

……



**首年产气量(215)**

龙一11 (1类厚度)(m): 0.35544895381672675

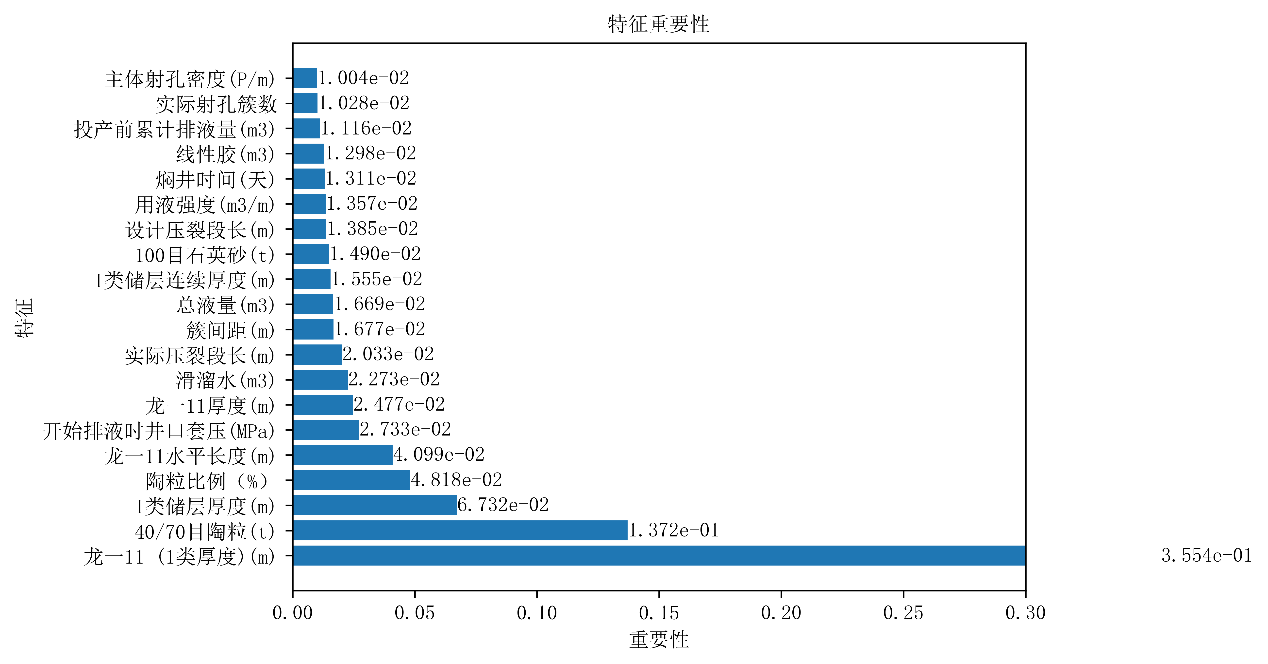
40/70目陶粒(t): 0.13718305604328285

I类储层厚度(m): 0.0673207997275141

陶粒比例（%）: 0.04817523336088908

龙一11水平长度(m): 0.04099069430632559

……



**第二年累产气量(173)**

龙一11 (1类厚度)(m): 0.35870169334188146

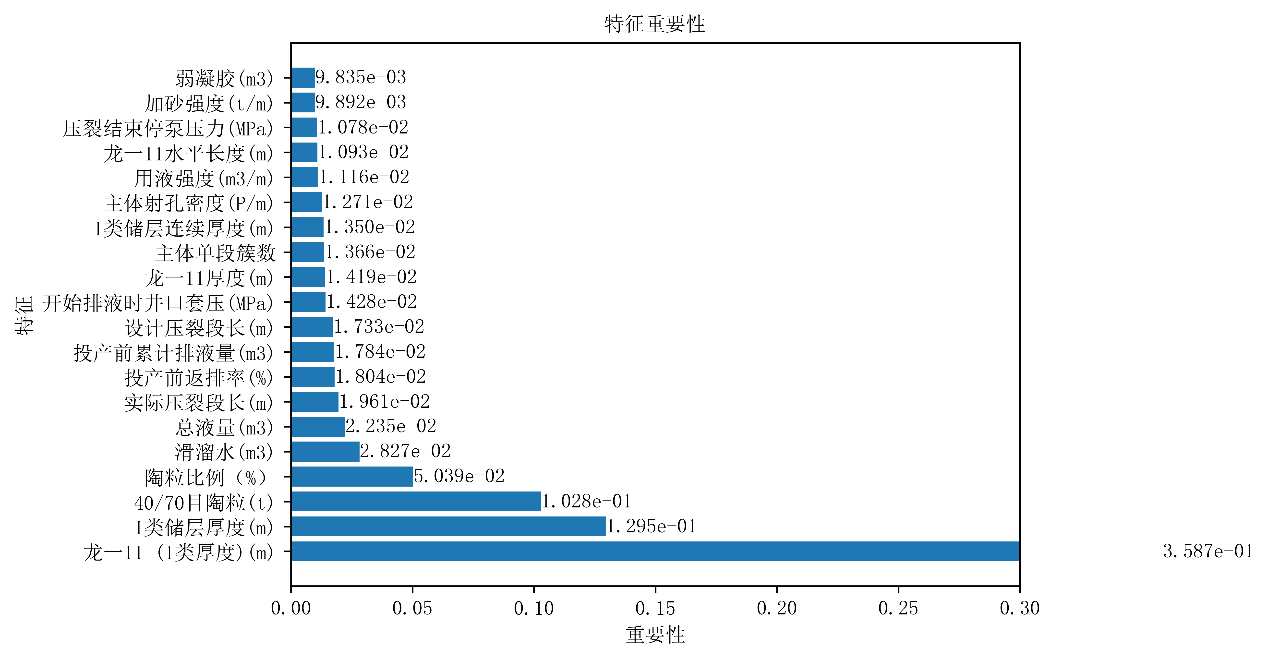
I类储层厚度(m): 0.12953480066690787

40/70目陶粒(t): 0.10283223683951917

陶粒比例（%）: 0.05039231798664311

滑溜水(m3): 0.02827021549492475

……



**第三年累产气量(102)**

I类储层厚度(m): 0.36140690327690217

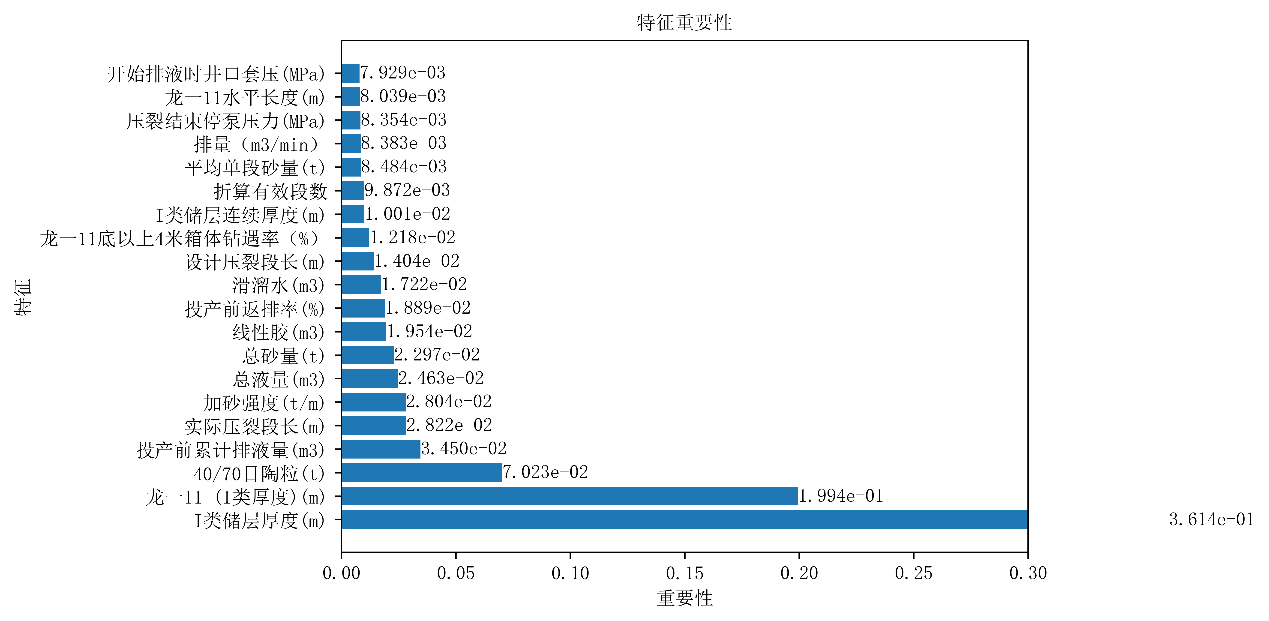
龙一11 (1类厚度)(m): 0.19938108269626584

40/70目陶粒(t): 0.07022592895916914

投产前累计排液量(m3): 0.034504794769419325

实际压裂段长(m): 0.02822196841158229

……



**第四年累产气量(53)**

I类储层厚度(m): 0.27207928964214995

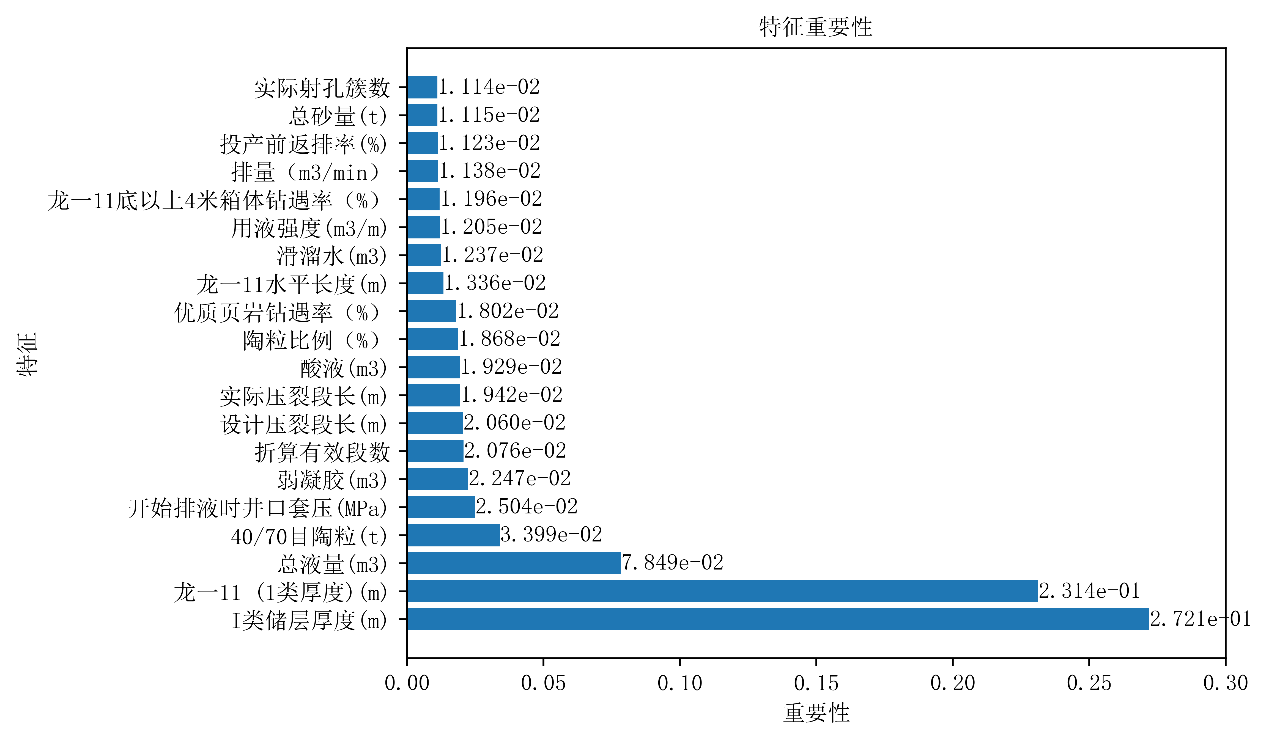
龙一11 (1类厚度)(m): 0.23135482190121187

总液量(m3): 0.07848895305942448

40/70目陶粒(t): 0.033992516547666725

开始排液时井口套压(MPa): 0.025044246890937413

……



**第五年累产气量(36)**

龙一11 (1类厚度)(m): 0.44597699516462014

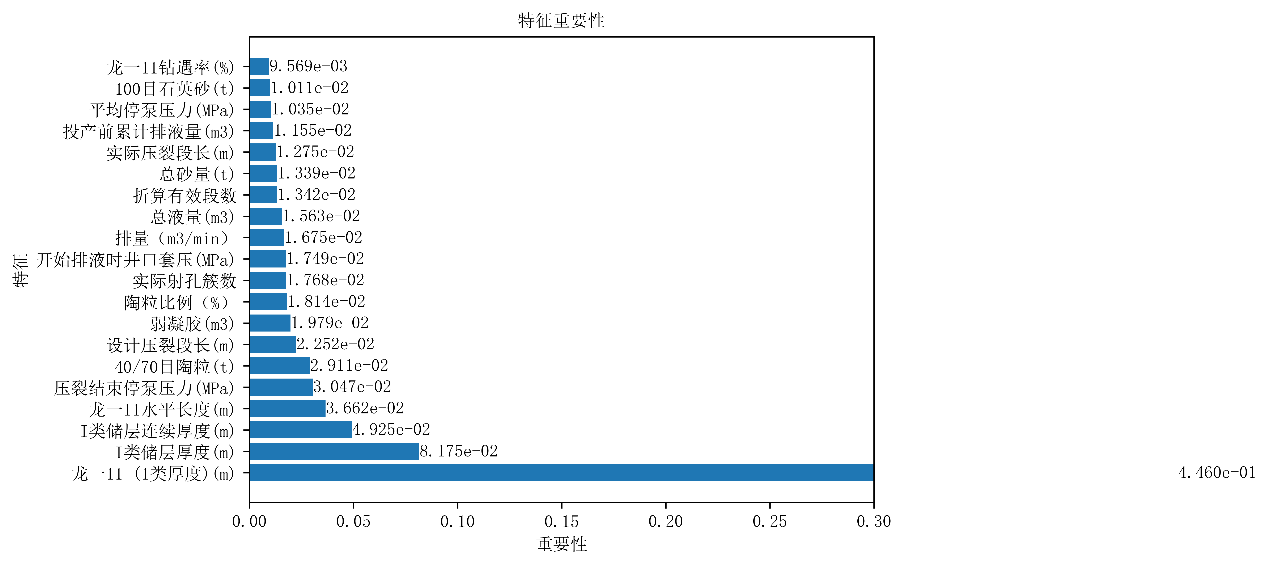
I类储层厚度(m): 0.0817481764563274

I类储层连续厚度(m): 0.04925091165542873

龙一11水平长度(m): 0.0366222817993451

压裂结束停泵压力(MPa): 0.030467925261212794

……



**第六年累产气量(22)**

I类储层连续厚度(m): 0.2021825954133162

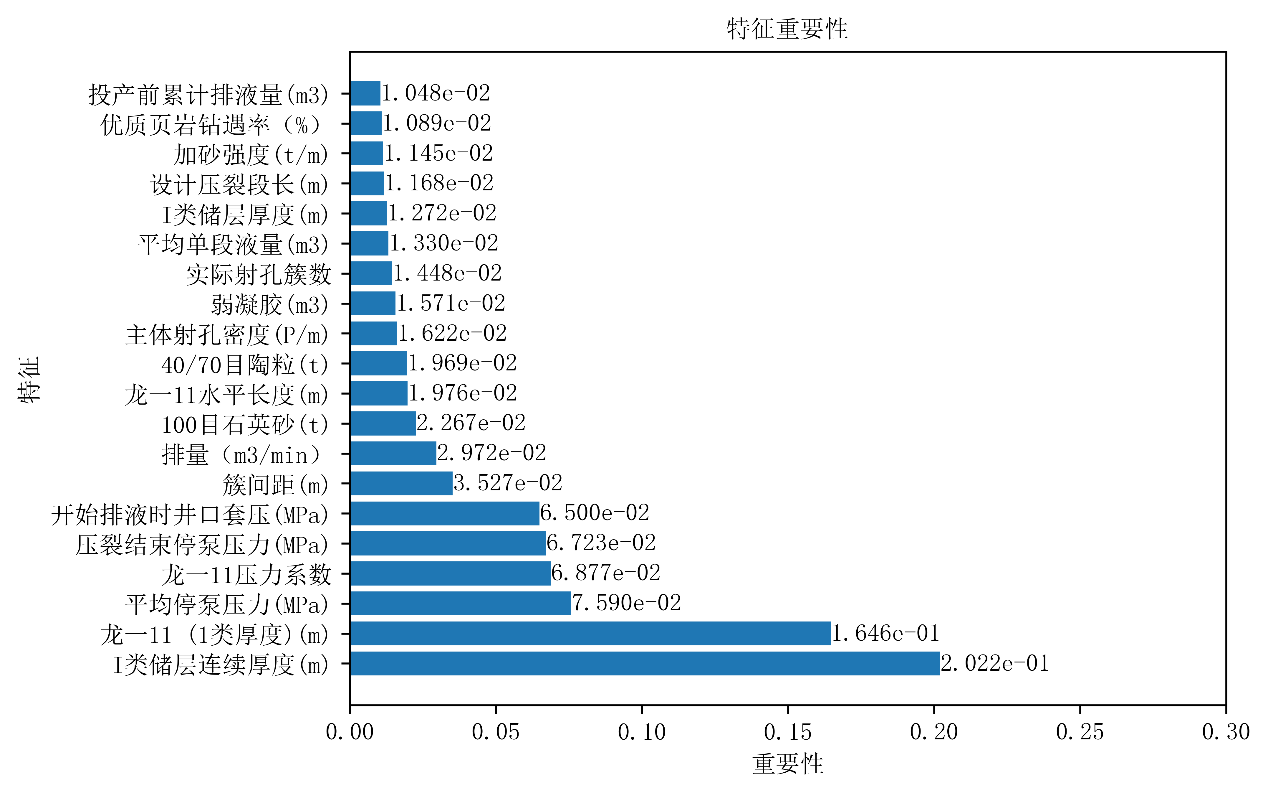
龙一11 (1类厚度)(m): 0.16461695525230033

平均停泵压力(MPa): 0.07589797213674584

龙一11压力系数: 0.06877326193662138

压裂结束停泵压力(MPa): 0.06722891682716116

……



# 二、重要性分析（所有井的每日累产数据）

**Y:** 生产周期内每日排名

**X:** ['I类储层厚度(m)','I类储层连续厚度(m)','龙一11厚度\n(m)','龙一11 (1类厚度)(m)','龙一11水平长度(m)','龙一11压力系数','龙一11钻遇率(%)','龙一11底以上4米箱体钻遇率（%）','水平\n段长\n(m)','实际压\n裂段长(m)','主体单段簇数','单段主体孔数','簇间距(m)','主体射孔密度(P/m)','加砂\n强度(t/m)','用液\n强度(m3/m)','分段段长(m)','是否\n套变', '合压\n长度\n(m)','丢段\n长度(m)','优质页岩钻遇率（%）','设计压\n裂段长(m)','实际\n射孔\n簇数','折算\n有效\n段数','100目石英砂(t)','40/70目石英砂(t)','40/70目陶粒(t)','30/50目陶粒(t)','总砂量(t)','酸液\n(m3)','滑溜水\n(m3)','线性胶\n(m3)','弱凝胶\n(m3)','总液量(m3)','平均单\n段砂量\n(t)','平均单\n段液量\n(m3)','陶粒\n比例\n（%）','平均停\n泵压力\n(MPa)','排量\n（m3/min）','压裂结束停泵压力(MPa)','焖井时间(天)','开始排液时井口套压\n(MPa)','投产前累计\n排液量(m3)','投产前返排率(%)','井间距(m)']

* **结果**

**90天累产每日排名(21780)**

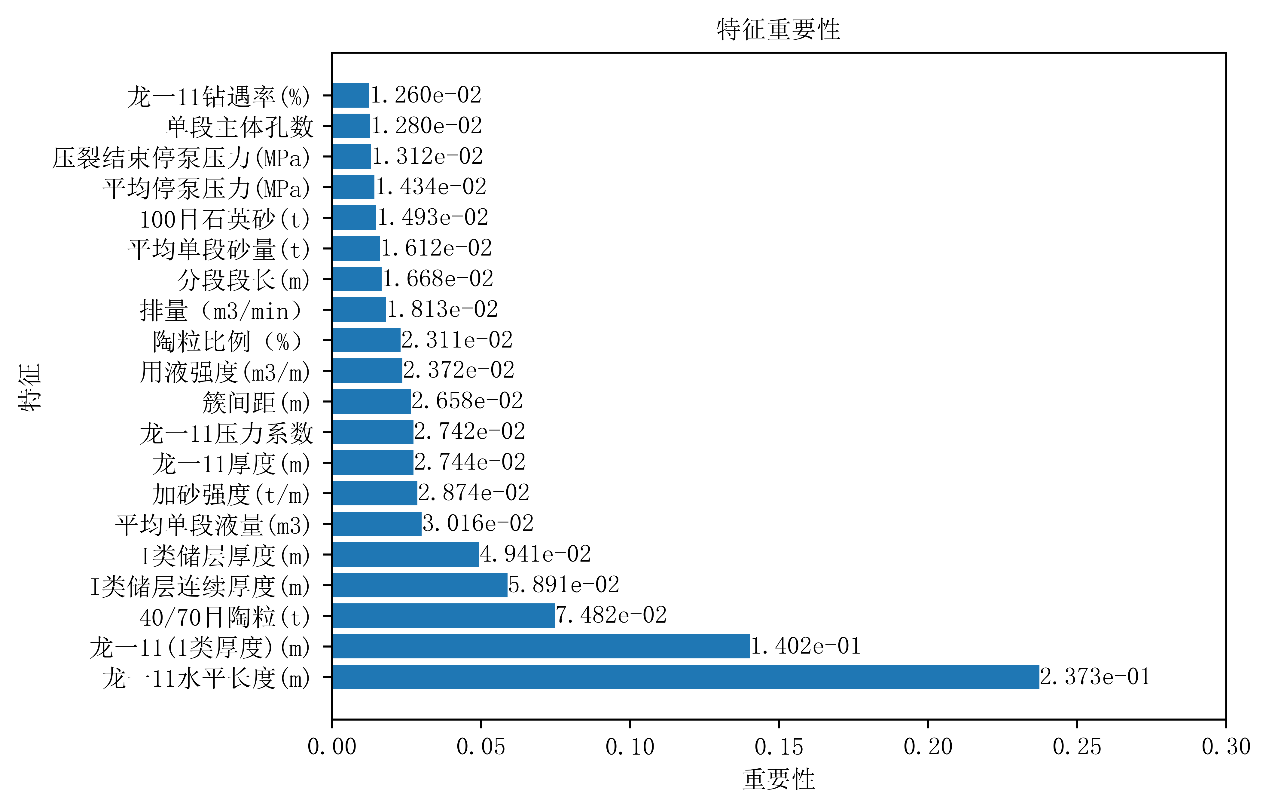
龙一11水平长度(m): 0.23734945064939278

龙一11(1类厚度)(m): 0.14018673071960416

40/70目陶粒(t): 0.07481956305041661

I类储层连续厚度(m): 0.058908168029538105

I类储层厚度(m): 0.049414190145500456



**两个季度累产每日排名(43560)**

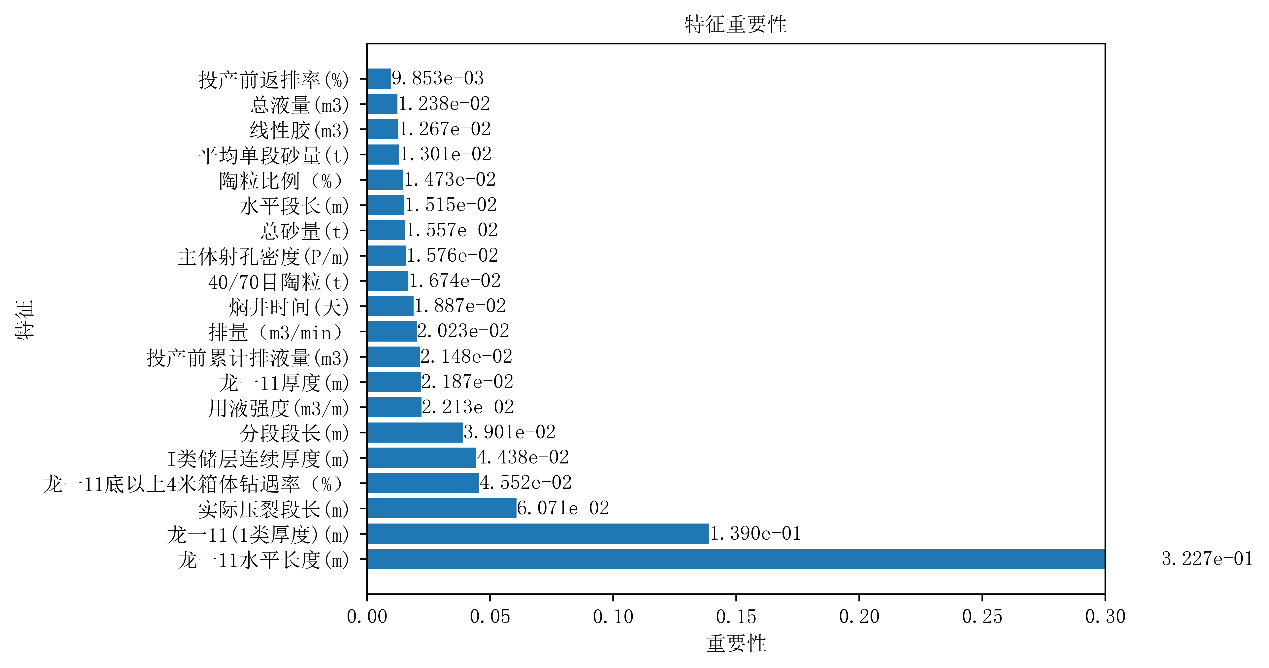
龙一11水平长度(m): 0.32267689229120766

龙一11(1类厚度)(m): 0.1390024437959226

实际压裂段长(m): 0.060714657586270565

龙一11底以上4米箱体钻遇率（%）: 0.045517448038062736

I类储层连续厚度(m): 0.044382661842750734



**三个季度累产每日排名(65340)**

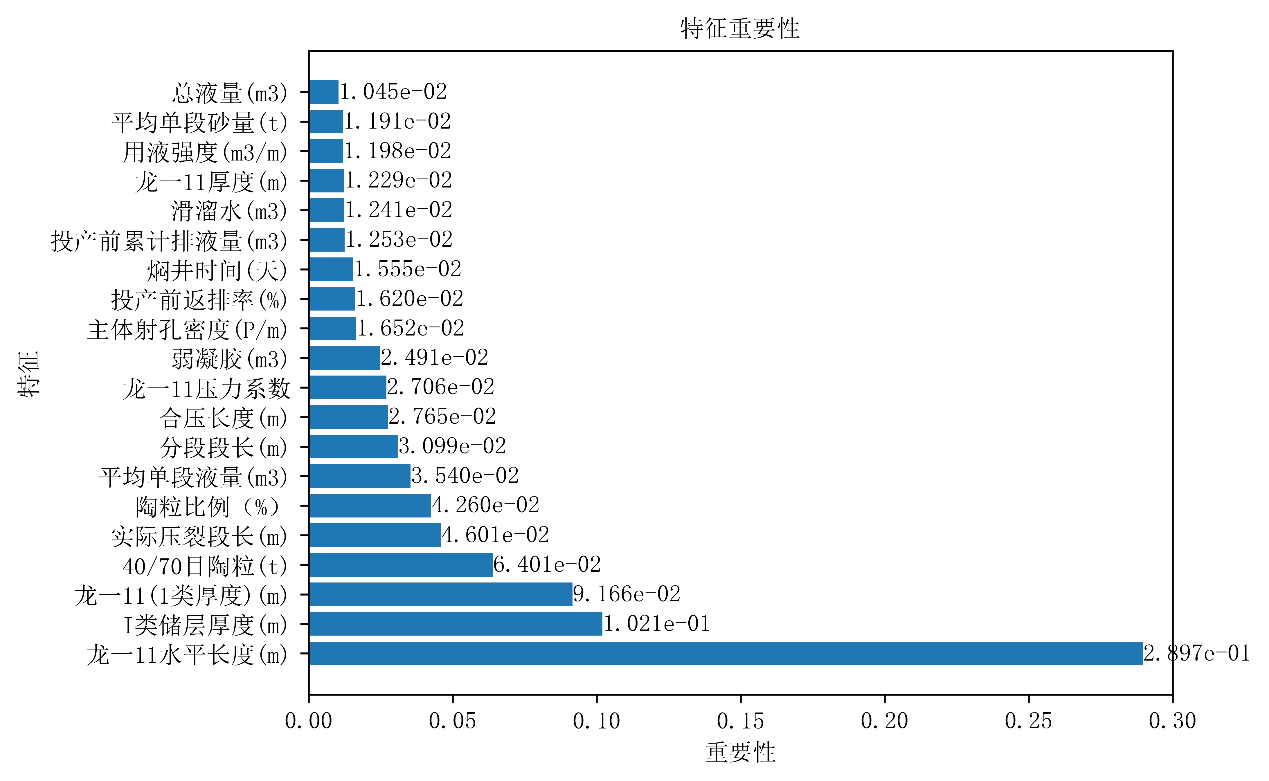
龙一11水平长度(m): 0.28969139645816905

I类储层厚度(m): 0.10206464048641004

龙一11(1类厚度)(m): 0.09165774035981597

40/70目陶粒(t): 0.06400802634356949

实际压裂段长(m): 0.04601468623504882



**首年累产每日排名(88330)**

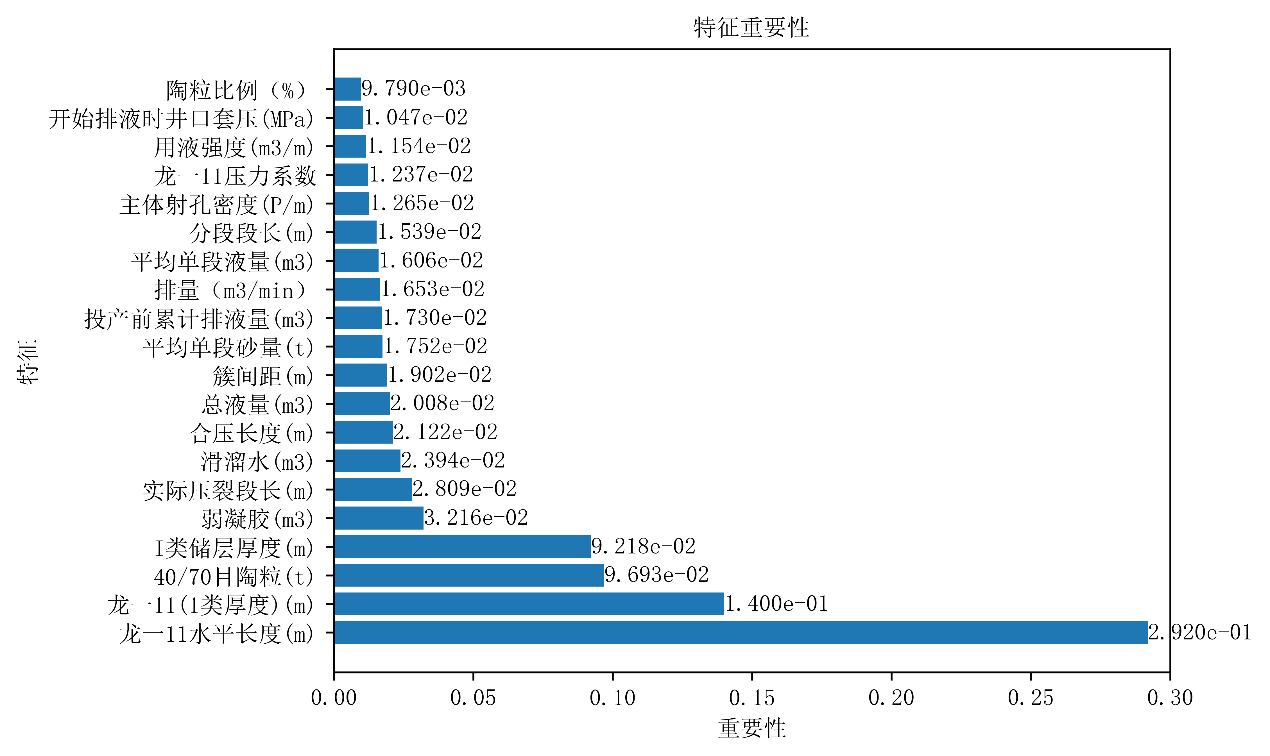
龙一11水平长度(m): 0.2919741751407319

龙一11(1类厚度)(m): 0.13998992825153797

40/70目陶粒(t): 0.09693230701836167

I类储层厚度(m): 0.09218132941288963

弱凝胶(m3): 0.032156556112713876



**二年累产每日排名(149650)**

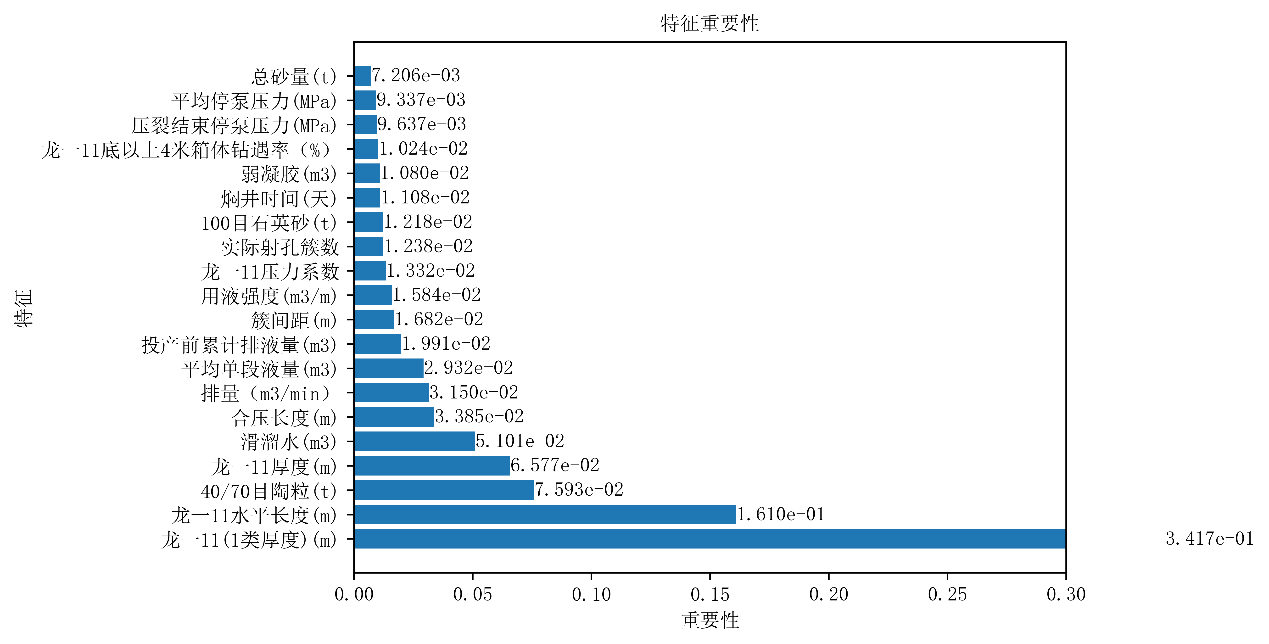
龙一11(1类厚度)(m): 0.34165844510903426

龙一11水平长度(m): 0.16102872240719063

40/70目陶粒(t): 0.07593363285665024

龙一11厚度(m): 0.06577312310180343

滑溜水(m3): 0.05100994836511564



**三年累产每日排名(185055)**

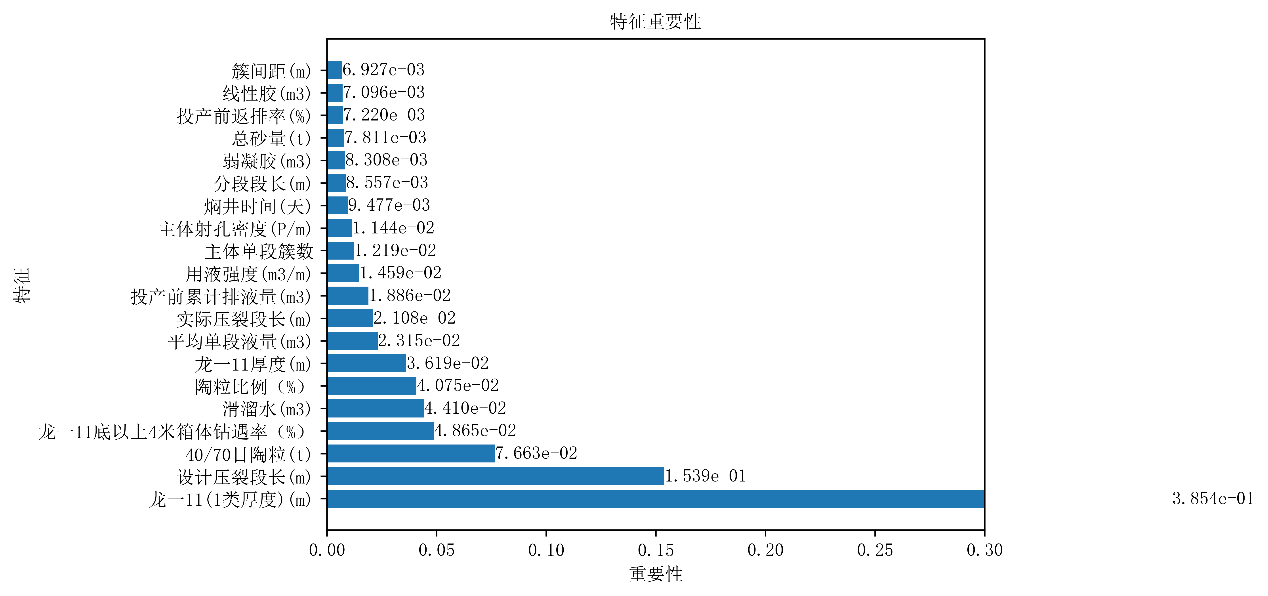
龙一11(1类厚度)(m): 0.3854038645416289

设计压裂段长(m): 0.1538516118051495

40/70目陶粒(t): 0.07662913806910315

龙一11底以上4米箱体钻遇率（%）: 0.04864799864292277

滑溜水(m3): 0.044100679122902324



**四年累产每日排名(163520)**

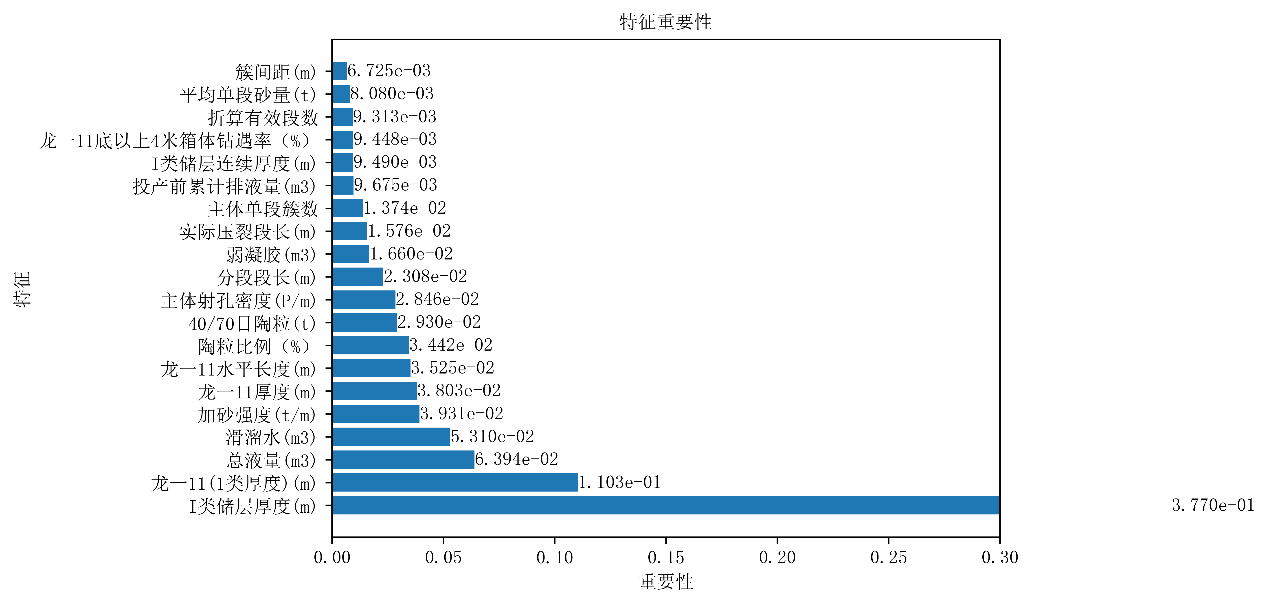
I类储层厚度(m): 0.37702053950604575

龙一11(1类厚度)(m): 0.11033704423244109

总液量(m3): 0.06393858383822132

滑溜水(m3): 0.053101010738928825

加砂强度(t/m): 0.03930765542904596



**五年累产每日排名(114975)**

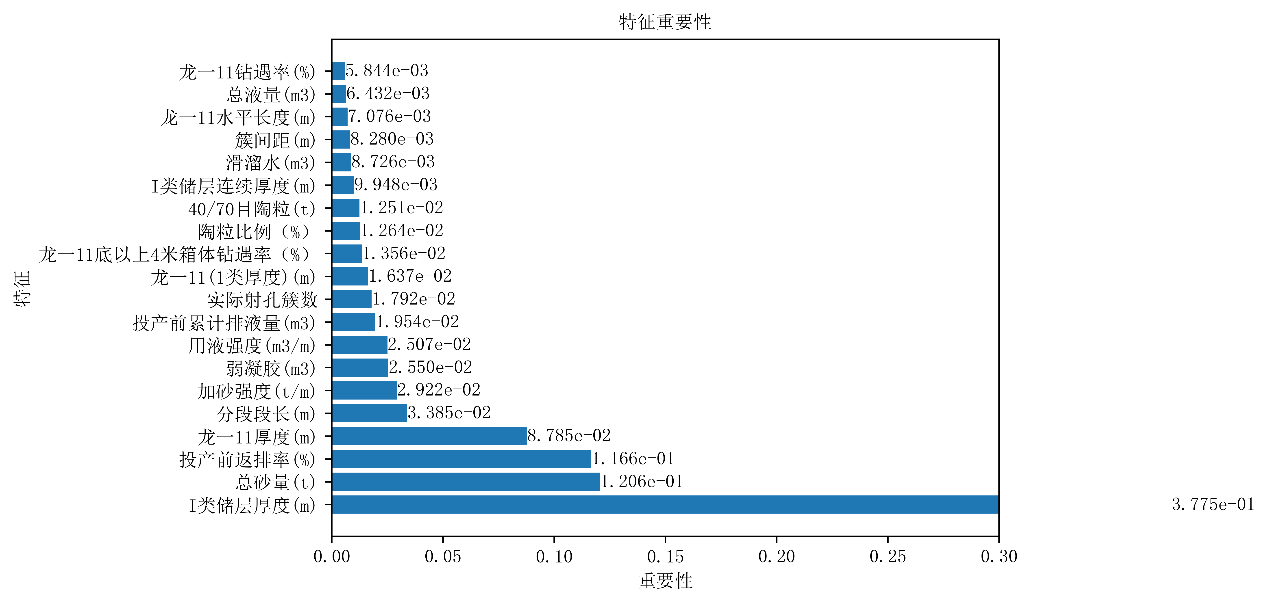
I类储层厚度(m): 0.3775192990640617

总砂量(t): 0.12060437164587035

投产前返排率(%): 0.11663394346313426

龙一11厚度(m): 0.08784634084415197

分段段长(m): 0.033850396638890175



**六年累产每日排名(100740)**

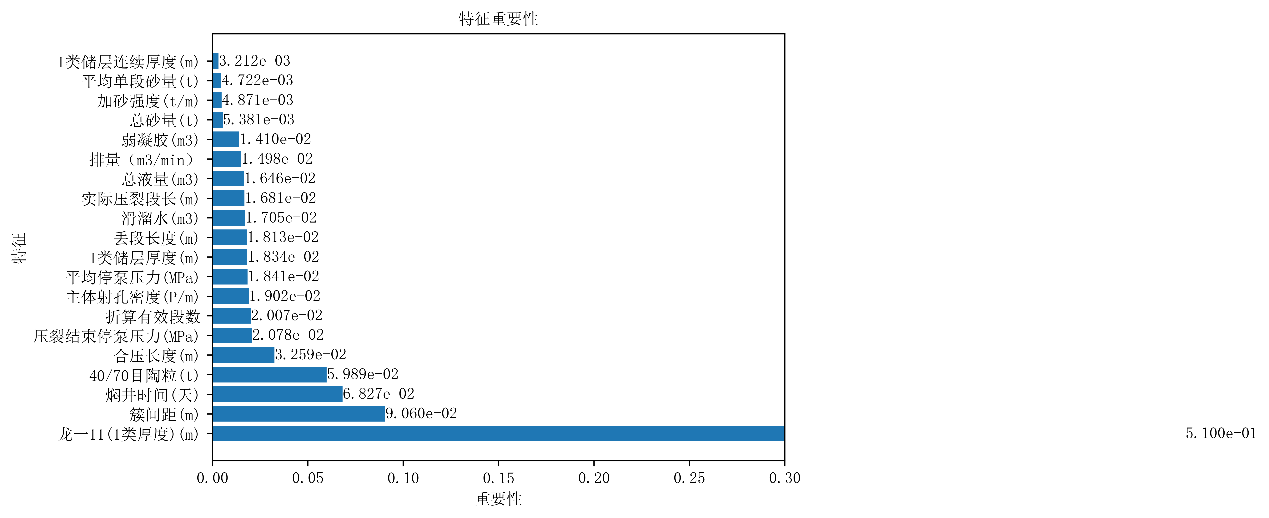
龙一11(1类厚度)(m): 0.5100093949868797

簇间距(m): 0.09059785437899473

焖井时间(天): 0.06827031849793805

40/70目陶粒(t): 0.059886514663299896

合压长度(m): 0.03259142822386644



# 三、最优参数仿真调优（开发大表.xlsx or 所有井的每日累产数据）

* **方法**

接收模型 (​model​)、特征 (​X​)、目标变量 (​y​) 和交叉验证的折数 (​k​) 作为输入，并返回交叉验证得到的均方根误差 (​rmse​)🡪均方根误差 (​rmse​)比较🡪选择最优模型🡪对指定参数进行特征值的搜索🡪得到最优参数值和最佳收益值（对应的累产）

* **示例结果**

数据：**90天累产每日排名(21780)**

交叉验证模型结果对比：

RandomForestRegressor交叉验证的均方根误差： 71.5604022225756

MLPRegressor交叉验证的均方根误差： 131.34617405640958

LinearRegression交叉验证的均方根误差： 129548221910.4449

SVR交叉验证的均方根误差： 67.68693064677504

DecisionTreeRegressor交叉验证的均方根误差： 79.24683559412021

KNeighborsRegressor交叉验证的均方根误差： 96.55065052882361

GradientBoostingRegressor交叉验证的均方根误差： 65.70290918960153

**AdaBoostRegressor交叉验证的均方根误差： 64.61475767568265**

BaggingRegressor交叉验证的均方根误差： 70.91247591740516

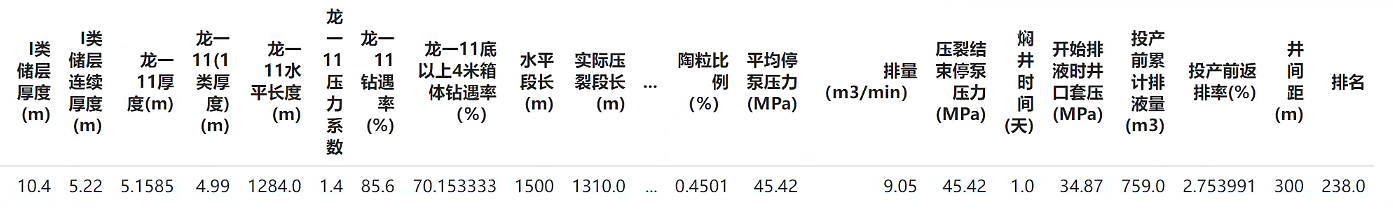
Lasso交叉验证的均方根误差： 74.30404038350635

Ridge交叉验证的均方根误差： 92.56372503209178

ElasticNet交叉验证的均方根误差： 74.90451282467359

选取模型：**AdaBoostRegressor**

特征值的搜索：固定其他基础参数不变，搜索实施过程中的可调参数（如‘40/70目陶粒(t)‘），得到该模型下对应最优参数，为实施提供指导。  
**例**：



40/70目陶粒(t)：524.47

start\_value = 10.0；end\_value = 3000.0；step = 10

特征值搜索下的最优值：**Optimal feature value: 1200.0**