

# 基于机器学习的企业价值影响因素分析

刘玉薇\*, 陈勇明#, 彭毅

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2023年3月24日; 录用日期: 2023年4月18日; 发布日期: 2023年4月28日

## 摘要

本文选取的数据为2009年~2021年创业板企业年度数据, 分为八大类因素搭建了创业板企业价值评估指标体系。综合使用SVR、GBRT、BP神经网络多个机器学习方法, 先验证了机器学习对创业板企业价值预测效果, 然后选择效果最优的机器学习模型对特征重要性进行排序, 最后从机器学习角度综合分析了创业板企业价值的影响因素。结果表明: 企业规模、企业盈利能力、股本扩张能力、发展能力以及创新能力类指标对创业板企业价值有较大影响, 其中代表企业规模的资产总计指标对创业板企业价值影响最大。

## 关键词

机器学习, 创业板企业价值, 影响因素, 特征重要性

# Analysis on Influencing Factors of En-terprise Value Based on Machine Learning

Yuwei Liu\*, Yongming Chen#, Yi Peng

School of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Mar. 24<sup>th</sup>, 2023; accepted: Apr. 18<sup>th</sup>, 2023; published: Apr. 28<sup>th</sup>, 2023

## Abstract

The data selected in this paper is the annual data of GEM (Growth Enterprise Market) enterprises from 2009 to 2021, which is divided into eight categories of factors to build the GEM en-terprise value evaluation index system. Using multiple machine learning methods such as SVR, GBRT and BP neural network, we first verified the effect of machine learning on the value prediction of GEM

\*第一作者。

#通讯作者。

enterprise, and then selected the machine learning model with the best effect to rank the importance of features. Finally, we comprehensively analyzed the factors affecting the value of GEM from the perspective of machine learning. The results show that enterprise scale, enterprise profitability, equity expansion ability, development ability and innovation ability have a great impact on the value of GEM enterprises, among which the total assets index representing enterprise scale has the greatest impact on the value of GEM enterprises.

## Keywords

Machine Learning, Gem Enterprise Value, Influence Factors, Feature Importance

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来随着不断深化改革，资本市场也在不断的升级改造之中。创业板于 2009 年 10 月正式成立，深圳交易所发布的数据显示，截至 2023 年 1 月，创业板上市企业数已达 1233 家，总市值已经高达 12 万亿。其中电子、医药、计算机等类型的高技术企业占比较大，此类企业具有高成长潜力、高风险性，同时，由于创业板成立时间不长，市场相对来说仍然不够成熟，不够稳定，导致一些创业板上市公司股票价格偏离公司真实价值程度较大，从而出现了高估或者低估企业价值现象。由于创业板这些特殊性，找到合适的企业价值评估体系也迫在眉睫。合理地评估创业板企业价值的影响因素，完善创业板企业价值评估体系，可以为创业板企业的估值提供一定的参考，对公司的经营的稳定性，以及引导投资人进行更为理性的投资，都是十分有意义的。

研究创业板上市企业价值以及与其密切相关的多种影响因素的传统分析方法包括主观分析、使用相关系数或是建立多元回归模型。周燕[1]使用了皮尔逊相关系数与多元回归法对创业板企业价值的相关影响因素展开了分析与研究，得到资产负债率、无形资产占比以及企业研发支出比例对企业价值均有正向或负向影响。而高梓棋，张煦[2]从主观分析的角度探讨了企业价值的内涵，结合我国成长型企业发展的特点，研究了成长型企业的内部外部因素对价值评估的影响，探讨了企业发展的前景，支持企业发展的核心和行业的成长环境等六个方面对企业价值评估的影响。罗雨楠[3]使用多元线性回归结合因子分析对当前的创业板企业价值相关影响因素进行了综合研究，其得到结论为相对于非财务因素，财务因素对于创业板上市企业价值的影响较大，其中主要包括盈利能力、营运能力、偿债能力等，而非财务因素方面也同样能够影响企业价值创造与成长质量，如公司规模、高管特征等等。

机器学习突破了传统分析方法大多无法挖掘非线性关系，难以处理高维度数据及对数据约束较多的局限性，有助于更好地识别创业板企业价值的影响因素。大部分的机器学习方法对数据分布或模型形式没有太多且严格要求，并且不需要对数据进行过多严格的前提假设就可以从数据中学习并挖掘创业板企业价值和相关指标之间的各种关系，从而普适性更强。机器学习方法的模型框架也更加灵活，比起传统方法可以考虑更高维度的数据，即可以同时分析更多指标与创业板企业价值之间的关系。另外，机器学习方法需要先建立创业板企业估值模型，使用训练集相关指标进行拟合，并在测试集上进行性能的验证，更容易得到准确科学的结果。最后，机器学习方法本质是数据驱动的，可以比传统估值方法更深入地探寻创业板企业价值以及其相关指标之间的关系。

## 2. 基础指标体系构建

### 2.1. 创业板企业价值指标体系框架

在过去的研究中，对创业板企业价值的影响因素分析集中在财务与非财务两个角度，其中关于财务指标与企业价值的关系，已经取得了丰硕的研究成果。Rappaport [4]发现公司价值增长会在很大程度上受所得税率、企业价值增长的时期、资本成本及毛利率等财务指标影响。赵宇龙[5]和陈晓等[6]研究均显示公司股价的波动与财务指标有很强的关联。Lewellen [7]研究发现市盈率等财务指标可以为预测股票价格提供有效信息。文海涛等[8]分行业进行了分析，研究发现对于不同的细分行业，上市公司的股价与公司

**Table 1.** GEM enterprise value index system

**表 1.** 创业板企业价值指标体系

	指标类型	指标名称
财务因素	营运能力	流动资产周转率
		总资产周转率
		应收账款周转率
		存货周转率
	盈利能力	主营业务利润占比
		每股经营活动产生的现金流量净额
		成本费用利润率
		净资产收益率
		总资产净利润率(ROA)
		每股净资产
		营业利润率
	发展能力	营业利润增长率
		营业收入增长率
		总资产增长率
偿债能力	速动比率	
	流动比率	
	资产负债率	
非财务因素	股权结构	长期资产适合率
		第一大股东持股比率(%)
	股本扩张能力	前十大股东持股比例(%)
		每股未分配利润
	企业规模	每股资本公积
		资产总计
创新能力	员工人数	
	机构投资者持股比例	
		无形资产比率
		研发投入占营业收入比例

的财务指标之间有明显关联，而不同的细分行业两者之间的相关程度又有一定的差异。胡逸凡[9]认为中小企业的 ROA、固定资产周转率和资产负债率与企业价值都有显著关联。

类似地对于非财务指标与企业价值的关系，随着市场的发展，学者们发现仅依靠财务指标信息解释企业价值是不全面的，非财务指标同样是企业价值重要的影响因素。McConnell 和 Servaes [10]对股权结构和企业价值之间的关系进行了实证分析，结果显示两者有曲线相关关系。Amir 和 Lev [11]研究了财务信息和非财务信息与企业价值相关性，强调了财务数据和非财务数据在解释企业价值中的互补性。胡奕明[12]研究发现非财务指标可以从不同角度评价企业的经营活动，很好地弥补大部分财务指标只能反映企业历史的经营状况的缺陷。

考虑到创业板上市企业相较主板有一定的差异，于静静[13]研究后认为科技型企业的财务价值与企业的营业收入净利润率、流动资产周转率显著正相关，另外与企业资产负债率、流动资产周转率、研发投入强度等指标存在不显著的相关关系。陈小平[14]提出创业板企业高管持股与公司价值存在一定的正相关。而余思佳[15]在对创业板企业 R&D 投入进行研究后，验证了研发强度对企业价值有正向的调节作用。常晓辉[16]认为企业内机构投资者的数量、机构投资持股状态对创业板企业价值也有一定的影响。战星羽[17]提出投资者关注度对创业板企业价值的影响呈现倒 U 型，而投资者情绪对企业价值存在正向促进效应。

## 2.2. 创业板企业价值指标体系

如前所述，企业价值的影响因素主要包括财务价值因素与非财务价值因素，影响创业板企业价值的因素较多，本文根据付扬扬[18]提出的评估指标体系并参照赖艳丽[19]、胡杨[20]等的研究做出一些改进，从运营能力、盈利能力、发展能力、偿债能力、股权结构、股本扩张能力、企业规模、创新能力八大类指标选取企业价值的可能影响因素，其中前四种为财务指标，后四种为非财务指标，尽可能涵盖以往文献的研究结果，并考虑数据的可得性，共提炼出 27 个指标进行衡量。具体如表 1 所示。

## 3. 研究方法 with 模型构建

本文利用支持向量回归机、梯度提升树、BP 神经网络这三种常用的非线性机器学习方法来研究创业板企业价值问题，能更深入地挖掘变量间的关系。最后使用排列重要性方法分析各影响因素在模型中的重要性。

### 3.1. 支持向量机

SVR (Support Vector Regression)原理为在线性函数的两侧创造“间隔带”，对落入到间隔带外的样本计算损失，最后使总损失最小化和间隔最大化得出最优模型。优点在于计算复杂度较低，可以解决高维非线性问题，并且对样本量较小，维度较高的问题适应良好。

### 3.2. GBRT

GBRT (Gradient Boost Regression Tree)梯度提升回归树算法是一种集成学习方法，其基础学习器是回归树。其有效克服了单个决策树为提高训练准确度而导致的过拟合，有训练时间少、收敛的速度较快及对数据量纲不敏感的优点。

### 3.3. BP 神经网络

BP (Back Propagation)算法是一种基于误差反向传播算法训练的多层前馈神经网络。本文的创业板企业价值模型由经典的三层全连接神经网络构建，三层神经网络已被证实可逼近任意具有有限个间断点的连续非线性函数，因此在解决回归预测问题上有很好的表现，能很好的挖掘非线性关系。

### 3.4. 特征重要性评估方法

本研究需要的各特征重要性与模型紧密相关，并且需要探索全局特征的重要性，基于这种考虑使用排列重要性方法(permutation importance)。原理为对于每个特征逐一对其数据进行随机打乱，其余特征保持不变，对预测精度进行计算与对比，模型表现评分下降越多，特征就越重要。

## 4. 实证及结果分析

### 4.1. 数据说明及预处理

本文选取 2009 年~2021 年创业板企业年度数据，数据来源为国泰安与锐思数据库。在建模前对所有指标进行正态化处理及 Z-score 标准化处理(其中 GBRT 算法使用的数据不需要进行标准化处理)，以保证训练模型时不受量纲和偏态分布影响。

### 4.2. 模型训练及效果评价

将数据分别按照 70%和 30%的比例随机分为训练集和测试集，训练集用于训练模型，测试集来评估模型的预测精度，将均方误差(MSE)与拟合优度(R2)作为模型预测精度的评价准则，并且使用三折交叉检验和随机网格法确定大致参数范围，最后再使用人工调整法对范围附近参数逐一尝试，精确定合适的参数。主要参数如表 2 所示。

Table 2. Main parameters of the model

表 2. 模型主要参数

模型	主要参数
SVM	C=4, gamma=0.02
BP 神经网络	activation = 'relu', alpha = 0.01, hidden_layer_sizes = (8,), learning_rate = 'adaptive', learning_rate_init = 0.007, max_iter = 350000, solver = 'sgd', tol = 0.001
GBRT	criterion = 'mse', learning_rate = 0.01, loss = 'ls', max_depth = 4, min_samples_split = 8, n_estimators = 900, random_state = 42, subsample = 0.5

对创业板企业价值建立预测模型，模型训练后，使用 MSE、R2 对比模型预测值和真实值在测试集上的效果，并将传统企业价值评估方式 EVA 值与真实市值的 MSE 做对比，结果如表 3 所示。

Table 3. Comparison of model effects

表 3. 模型效果对比

	SVR	BP 神经网络	GBRT	EVA 值
MSE	3.86E+20	3.32E+20	2.27E+20	9.25E+20
R2	0.755	0.744	0.779	

可以看到模型效果最好的是 GBRT，其次为 BP 神经网络，三种非线性机器学习模型效果均优于 EVA 法计算出的企业价值。综合三个指标来看，GBRT 效果最好。

### 4.3. 各类因素对企业价值的影响重要性分析

分析本文建立的机器学习创业板企业价值模型效果，选择其中表现最好的 GBRT 进行特征重要性的评估，评估方法为排列重要性，用 eli5.sklearn 中的 Permutation Importance 函数进行，模型特征重要性排序及权重结果如表 4。

**Table 4.** GBRT feature importance ranking and weight results  
**表 4.** GBRT 模型特征重要性排序及权重结果

指标	重要性排序	指标	重要性排序
资产总计	1	营业利润增长率	14
总资产净利润率(ROA)	2	应收账款周转率	15
每股资本公积	3	存货周转率	16
总资产增长率	4	流动资产周转率	17
研发投入占营业收入比例	5	营业收入增长率	18
机构投资者持股比例	6	资产负债率	19
每股净资产	7	总资产周转率	20
净资产收益率	8	速动比率	20
每股未分配利润	9	无形资产比率	22
员工人数	10	每股经营活动产生的现金流量净额	23
前十大股东持股比例	11	营业利润率	24
主营业务利润占比	12	流动比率	25
成本费用利润率	13	第一大股东持股比率	25

具体来看, 其中资产总计在所有特征最为重要, 资产总计是企业规模的一个重要体现, 企业总资产水平最能影响企业价值。其他排名前五的特征: 总资产净利润率(ROA), 每股资本公积, 总资产增长率, 研发投入占营业收入比例同样是创业板企业价值的重要影响因素, 分别代表企业盈利能力、股本扩张能力、发展能力以及创新能力, 排名前 5 的特征中有 3 个特征为非财务指标。另外机构投资者持股比例、每股净资产, 净资产收益率, 每股未分配利润, 员工人数也在模型中重要性排名为前十, 其中每股净资产、净资产收益率为财务指标, 代表企业的盈利能力, 而每股未分配利润、机构投资者持股比例、员工人数为非财务指标, 每股未分配利润代表了企业的股本扩张能力, 而后两个代表了企业规模。

由此可见, 财务指标与非财务指标对创业板企业价值均有一定程度的影响, 其中财务指标主要是代表盈利能力与发展能力的指标在影响企业价值。而非财务指标中的企业规模、企业的股本扩张能力与创新能力对创业板企业价值均有较大影响。而在所有类型的指标中, 影响最大的又为企业规模。

## 5. 结语

本文搭建了创业板企业价值评估指标体系, 综合使用多种非线性机器学习方法, 对创业板企业价值进行了预测, 将预测结果与传统方法的估值结果进行对比, 并选择效果最优的模型机器学习模型对特征重要性进行排序, 并对影响因素进行了综合分析。

由研究结果可知, 财务与非财务指标对创业板企业价值均有影响, 并且企业规模、企业盈利能力、股本扩张能力、发展能力以及创新能力类指标对创业板企业价值有较大影响, 其中代表企业规模的资产总计指标最能影响创业板企业价值。本研究对进一步研究创业板企业价值的影响因素建立以及对创业板企业价值的预测模型研究均有一定的参考价值。

## 基金项目

四川省科技厅项目资助(编号: 2022JDR0043)。

## 参考文献

- [1] 周燕. 创业板企业价值影响因素研究[J]. 广西质量监督导报, 2020(1): 9.
- [2] 高梓棋, 张煦. 成长型企业的价值评估影响因素分析[J]. 知识经济, 2018(7): 75-77.
- [3] 罗雨楠. 创业板上市企业价值影响因素研究[D]: [硕士学位论文]. 上海: 上海财经大学, 2020.
- [4] Rappaport, A. (1986) *Creating Shareholder Value: The New Standard for Business Performance*. Free Press, New York.
- [5] 赵宇龙. 会计盈余披露的信息含量——来自上海股市的经验证据[J]. 经济研究, 1998(7): 42-50.
- [6] 陈晓, 陈小悦, 刘钊. A 股盈余报告的有用性研究——来自上海、深圳股市的实证证据[J]. 经济研究, 1999(6): 21-28.
- [7] Jonathan, L. (2004) Predicting Returns with Financial Ratios. *Journal of Financial Economics*, **74**, 209-235. <https://doi.org/10.1016/j.jfineco.2002.11.002>
- [8] 文海涛, 倪晓萍. 我国上市公司财务指标与股价相关性实证分析[J]. 数量经济技术经济研究, 2003(11): 118-122.
- [9] 胡逸凡. 中小企业板块中小企业价值的影响因素[J]. 金融经济, 2018(22): 167-168.
- [10] Mcconnell, J.J. and Servaes, H. (1990) Additional Evidence on Equity Ownership and Corporate Value. *Journal of Financial Economics*, **27**, 595-612. [https://doi.org/10.1016/0304-405X\(90\)90069-C](https://doi.org/10.1016/0304-405X(90)90069-C)
- [11] Amir, E. and Lev, B. (1996) Value-Relevance of Nonfinancial Information: The Wireless Communications Industry. *Journal of Accounting and Economics*, **22**, 3-30. [https://doi.org/10.1016/S0165-4101\(96\)00430-2](https://doi.org/10.1016/S0165-4101(96)00430-2)
- [12] 胡奕明. 非财务指标的选择——价值相关分析[J]. 财经研究, 2001(5): 44-49.
- [13] 于静静. 科技型企业价值的影响因素研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2017.
- [14] 陈小平. 创业板上市公司高管持股与公司价值——基于区间效用论的实证研究[J]. 会计之友, 2015(21): 59-62.
- [15] 余思佳. 创业板企业 R&D 投入与企业价值相关性研究[D]: [硕士学位论文]. 贵阳: 贵州财经大学, 2017.
- [16] 常晓辉. 基于创业板的机构持股对企业价值影响研究[D]: [硕士学位论文]. 北京: 华北电力大学, 2018.
- [17] 战星羽. 投资者关注度、投资者情绪对创业板企业价值影响的实证分析[D]: [硕士学位论文]. 上海: 上海财经大学, 2021.
- [18] 付扬扬, 仇晓洁. 基于 BP 神经网络的创业板上市公司价值评估[J]. 中国资产评估, 2019(2): 35-40.
- [19] 赖艳丽. 创业板股票内在价值多元回归模型的研究[D]: [硕士学位论文]. 上海: 复旦大学, 2013.
- [20] 胡杨. 企业长期价值影响因素研究[D]: [硕士学位论文]. 成都: 四川大学, 2021