

March 2024

機械学習を用いた企業の特徴分析
**Analysis of Company Characteristics using
Machine Learning**

山口 健二
Kenji Yamaguchi

Research Institute of Economic Science

College of Economics, Nihon University

機械学習を用いた企業の特性分析
Analysis of Company Characteristics using Machine Learning

山口 健二
Kenji Yamaguchi

要旨

本研究の目的は、個々の企業の株価と財務データとの関係を、機械学習を用いて評価する手法を構築することである。例えば、売上高が上がれば株価が上がりやすいといったことは、誰しも思い浮かぶことであるが、在庫回転率など、他の財務データの変動と比べて、どの程度株価に影響があるのかといったことについては、評価が難しく、企業によっては、それが必ずしも成り立たないこともありえる。すなわち、企業規模や業種によって株価と財務データの関連性が大きく異なるため、一義的な関係性の抽出は困難であった。しかし、協力ゲーム理論の Shapley 値を機械学習に応用した SHAP 値を使うことで、個々の企業で株価と財務データとの関係性を抽出することが可能となる。その結果から、企業ごとの売上高や在庫回転率、営業利益率といった財務データの目標値を設定することができるのではないかと考えた。本研究で、重回帰と SHAP 値による分析をそれぞれ行ったので結果を報告する。

1. はじめに

株価と企業の財務データは、その企業の健全性、成長性、収益性を示す重要な指標である。例えば、売上高は、企業が製品やサービスの販売から得た総収入であり、企業の基本的な収益性と市場でどの程度のシェアを持っているか見ることができる。従来から「製造業では在庫回転率が高い企業の方が、営業利益率が高い」と言われている。これは、効率的な在庫管理により、キャッシュフローの改善や保管コストの低減、柔軟な市場対応ができ、それにより営業利益率の向上に繋がると言う考え方である。また、企業の業績が好調で、売上高が上がった場合、株価も上がると予想されるが、これは、投資家が、企業が成長していると認識し、将来的な収益増加の可能性があると見なされるため、投資家が株を購入することで株価を押し上げることによるものである。ただし、これには、営業利益や経常利益といった利益率にも着目しなければならず、また、企業が属する業界の状況も加味する必要がある。即ち、いくら、売上高が好調だといっても、業界全体の売り上げが下降している場合は、株価を押し上げるには難しいと考えられる。これは、投資家が将来の成長見通しを重視するためである。また、それ以外にも、政治的リスクなども考慮する必要がある。このような、定性的な要因が存在するものの、定量的な評価を避けるのは好ましくなく、株価と企業の財務データの間をなんかの手法で分析する研究は重要であると考えられる。

2. 既存研究

ここでは、株価や財務データの予測に関する既存研究について、いくつか説明する。

株価を決定する要因については、数多くの研究が行われている。ファイナンス関連の研究による決定要因としては、企業財務に関する指標、配当、企業規模、未分配利益、ROA、ROE、負債比率などが挙げられる[1、2]。伊藤レポートでは、日本企業のROEは世界平均のROEよりも低く、日本企業の目標はROE8%の達成であるとしている[3]。このように、投資家はROEを重要なファクターとして捉えている。また、マクロ経済の観点からは、長期金利と短期金利の乖離、インフレ率、鉱工業生産などの指標も関連要因として研究されている[4]。

また、株価の変動を予測するために、様々な時系列分析モデルが用いられてきた。株価のボラティリティや分散は重要な研究テーマであり、ARCHモデル[5]をはじめとする多くの時系列分析モデルが開発されています。株価予測のために時系列分析モデルが構築されてきた。しかし、時系列モデルを予測に用いる場合、分散に影響を与える要因を特定する必要があるが、一般的にはそのような要因を特定することは困難である。例えば、1970年代には石油危機、金融政策の変更、ブレトンウッズ体制の崩壊など様々な要因が投資のボラティリティを高めたが、どの要因が最も重要であったかの判断は研究者によって異なっている

従来の株価変動に関する研究は、株価下落のパターンと投資家の行動との関係を説明することに着目するもある。例えば、天災・人災を問わず何らかの危機が発生すると、投資家（自動売買プログラムも含む）が反応し、経済指標や企業業績の指標を参考にしながら株式の売買を決定する。したがって、なんらかの危機が株価変動に及ぼす影響に関する研究は、投資家の動機や行動に及ぼす影響を測定して、市場全体への影響を分析するものと考えられる。

3. データと分析手法

本研究で用いたデータは次のとおりである。日経 NEEDS Financial QUEST で企業の株価および財務データを取得した。対象企業は、日本の製造業の自動車部門に属する自動車会社9社、自動車部品会社50社、自動車車体会社4社である。便宜上、財務データが3月末日でない企業は除外した。

目的変数と説明変数として、それぞれ2パターンのデータを用意した。

目的変数の1つ目は、短期的株価変動を見るための2022年2月の終値を1としたときの2022年3月の株価指数、2つ目は中期的株価変動を見るための、2021年3月の終値を1としたときの、2022年3月の株価指数である。

説明変数の1つ目は、2018年から2022年の5年間の財務データ6項目の各平均値、2つ

目は、2013年から2022年の10年間の財務データ6項目の各平均値である。財務データの6項目は、以下のとおりである（表1）。ただし、「(前年比)売上高・営業収益 [累計]」は平均値として幾何平均を使用した。また、分析上、表1にあるように用語を英語で置き換えている。

目的変数と説明変数それぞれ2パターンを組み合わせて、ケース1からケース4で分析を行った（表2）。

表1：説明変数として使用した財務データと分析上の名称

日経NEEDS Financial QUEST上の名称	分析上の名称 (重回帰分析)	分析上の名称 (SHAP値による分析)
(前年比)売上高・営業収益 [累計]	(Ave5) Net Sales	SHAP-(Ave5)SG
(売上高に対する)税金等調整前当期純利益 [累計]	(Ave5) Net Income before Income Taxes and Others / Net Sales	SHAP-(Ave5)NI
(売上高に対する)棚卸資産	(Ave5) Inventories / Net Sales	SHAP-(Ave5)IVT
(売上高に対する)有形固定資産	(Ave5) Tangible Fixed Assets / Net Sales	SHAP-(Ave5)TFA
使用総資本事業利益率	(Ave5) ROA	SHAP-(Ave5)ROA
ROE (自己資本利益率) (A)	(Ave5) ROE	SHAP-(Ave5)ROE

(10年間の財務データの平均値を使用した場合は、(Ave10)となる)

表2：目的変数と説明の変数の組み合わせと各ケース

	ケース1	ケース2	ケース3	ケース4
目的変数	(短期的株価変動) 2022年2月の終値を1としたときの 2022年3月の株価指数		(中期的株価変動) 2021年3月の終値を1としたときの、 2022年3月の株価指数	
説明変数	2018年から2022年の5年間の財務データ6項目の各平均値	2013年から2022年の10年間の財務データ6項目の各平均値	ケース1と同じ	ケース2と同じ
対象企業	自動車メーカー9社 自動車部品会社50社 自動車車体会社4社 計63社			

4. 重回帰による分析

まず、Excel で重回帰分析の結果を表 3 から表 10 に記載する。ケース 1 からケース 4 すべてにおいて、重相関 R は高くなく、また、得られた P-値から有意な係数を見つけることはできなかった。

表 3：ケース 1 の重回帰分析の回帰統計

回帰統計	
重相関 R	0.308127
重決定 R2	0.094942
補正 R2	-0.00203
標準誤差	0.073343
観測数	63

表 4：ケース 1 の重回帰分析の係数

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	0.577488	0.218805	2.639283	0.010741	0.139169	1.015806
(Ave5) Net Sales	0.435424	0.233605	1.863931	0.067577	-0.03254	0.903391
(Ave5) Net Income before Income Taxes and Others / Net Sales	0.425188	0.653276	0.650855	0.517802	-0.88348	1.733858
(Ave5) Inventorie s / Net Sales	0.039423	0.192005	0.205321	0.838066	-0.34521	0.424055
(Ave5) Tangible Fixed Assets / Net Sales	-0.03455	0.07067	-0.48893	0.6268	-0.17612	0.107016
(Ave5) ROA	-0.01003	0.008403	-1.19343	0.237732	-0.02686	0.006805
(Ave5) ROE	-0.00076	0.002904	-0.26169	0.794522	-0.00658	0.005057

表 5：ケース 2 の重回帰分析の回帰統計

回帰統計	
重相関 R	0.249874
重決定 R ²	0.062437
補正 R ²	-0.03802
標準誤差	0.074649
観測数	63

表 6：ケース 2 の重回帰分析の回帰統計

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	0.461946	0.381064	1.212253	0.230505	-0.30142	1.225308
(Ave10) Net Sales	0.525196	0.387283	1.356104	0.180509	-0.25063	1.301018
(Ave10) Net Income before Income Taxes and Others / Net Sales	0.402139	0.713224	0.563833	0.575119	-1.02662	1.830899
(Ave10) Inventorie s / Net Sales	0.065983	0.216829	0.304309	0.762021	-0.36838	0.500343
(Ave10) Tangible Fixed Assets / Net Sales	-0.03943	0.080391	-0.49044	0.62574	-0.20047	0.121615
(Ave10) ROA	-0.00436	0.007807	-0.55885	0.578495	-0.02	0.011276
(Ave10) ROE	-0.00353	0.003992	-0.88315	0.380931	-0.01152	0.004472

表 7：ケース 3 の重回帰分析の回帰統計

回帰統計	
重相関 R	0.225919
重決定 R2	0.051039
補正 R2	-0.05064
標準誤差	0.205352
観測数	63

表 8：ケース 3 の重回帰分析の回帰統計

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	0.151472	0.612625	0.247251	0.805618	-1.07576	1.378707
(Ave5) Net Sales	0.649228	0.654064	0.992606	0.325172	-0.66102	1.959475
(Ave5) Net Income before Income Taxes and Others / Net Sales	-1.31536	1.829088	-0.71914	0.475048	-4.97947	2.348742
(Ave5) Inventorie s / Net Sales	0.30927	0.537588	0.575291	0.567399	-0.76765	1.386189
(Ave5) Tangible Fixed Assets / Net Sales	0.127253	0.197867	0.643124	0.522769	-0.26912	0.523628
(Ave5) ROA	0.014672	0.023527	0.623643	0.535394	-0.03246	0.061803
(Ave5) ROE	-0.00352	0.00813	-0.43236	0.667139	-0.0198	0.012772

表 9：ケース 4 の重回帰分析の回帰統計

回帰統計	
重相関 R	0.17337
重決定 R2	0.030057
補正 R2	-0.07387
標準誤差	0.207609
観測数	63

表 10：ケース 4 の重回帰分析の回帰統計

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	1.593231	1.059796	1.503337	0.138371	-0.5298	3.716258
(Ave10) Net Sales	-0.81331	1.077093	-0.7551	0.453353	-2.97099	1.344363
(Ave10) Net Income before Income Taxes and Others / Net Sales	-0.80578	1.983585	-0.40623	0.686125	-4.77938	3.167815
(Ave10) Inventorie s / Net Sales	0.46528	0.603033	0.771566	0.443615	-0.74274	1.673301
(Ave10) Tangible Fixed Assets / Net Sales	0.137457	0.223578	0.614804	0.541175	-0.31042	0.585338
(Ave10) ROA	0.000683	0.021712	0.03148	0.974998	-0.04281	0.044177
(Ave10) ROE	0.006171	0.011103	0.555772	0.580581	-0.01607	0.028412

5. 機械学習回帰と SHAP 値による分析

次に、機械学習回帰と SHAP 値による分析を行った結果を記載する。まず機械学習による回帰を行い、その結果から SHAP 値を求め、どの財務データが株価に影響を与えているのかを考察する。機械学習回帰の手法として、XGBoost を使用した。XGBoost (Extreme Gradient Boosting) は機械学習回帰で用いられるアルゴリズムの一つで、決定木をベースにしており、勾配ブースティングという手法を用いて、複数の決定木を組み合わせて回帰予測を行う。勾配ブースティングは、前の決定木の誤差を次の決定木が修正することで、モデルの性能を向上させる手法である。また、SHAP 値は、各企業の各目的変数に対して数値 (SHAP 値) が得られる (表 11)。SHAP 値が正值だった場合、目的変数にプラスの影響を与えたものと判断される。逆に負値だった場合、目的変数にマイナスの影響を与えたものと判断される。SHAP 値は、各企業の各目的変数に対して数値 (SHAP 値) が算出されるため、例えば、ある企業では、売上高の増加が株価の上昇に最も影響を与えるといった、企業の特徴を調べることが可能である。実際には、ある説明変数に対して、企業ごとの SHAP 値と株価指標をプロットすることで、各企業の特徴を見ることが可能となる (図 1~図 4)。ケース 1 からケース 4 の各説明変数の SHAP 値と目的変数の相関係数を算出した結果を表 12 に示す。

表 11 : SHAP 値 (一部のみ抜粋)

Industry	Target1	(Ave5)S G	(Ave5)N I	(Ave5)IV T	(Ave5)TF A	(Ave5)RO A	(Ave5)RO E
VEHICLE S	1.00458 5	-0.00275	-0.01566	0.004543	0.004235	0.026016	0.00785
VEHICLE S	1.01991	0.015357	0.006599	0.004096	0.008898	0.011405	0.005756
VEHICLE S	1.03928	0.014985	0.005305	0.006294	0.010569	0.014321	0.004736
VEHICLE S	0.66852 4	-0.09365	-0.08899	-0.00182	-0.05217	-0.03439	-0.00074
VEHICLE S	1.10333 3	0.036121	-0.01162	0.007005	0.019981	0.066777	0.009758
VEHICLE S	1.06191 6	0.020526	0.008748	0.007637	0.026004	0.024155	0.00023
VEHICLE S	0.98922	0.010305	0.003637	0.0038	0.001381	0.001442	-2.6E-05
VEHICLE S	0.92127 7	0.007565	0.002584	-0.00523	-0.00488	-0.02648	-0.0011

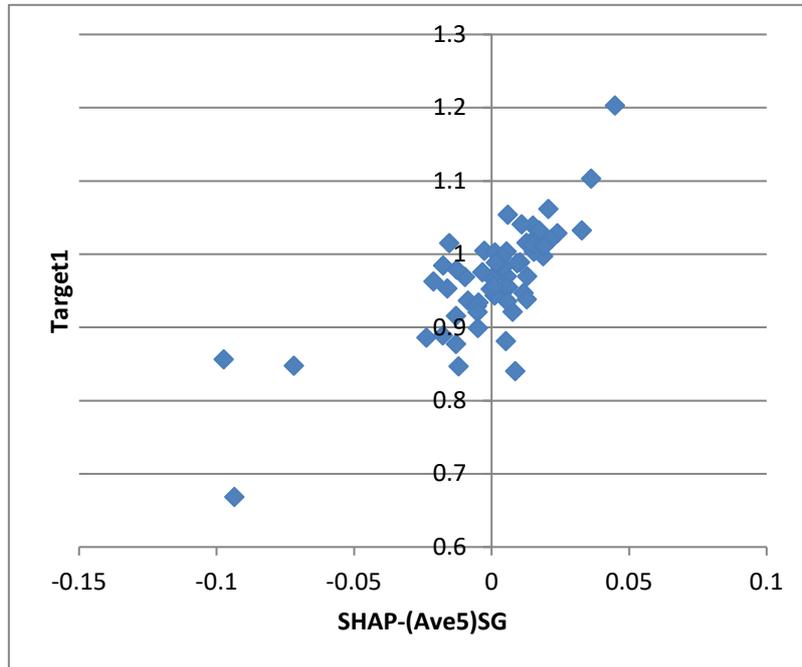


図 1 : ケース 1 における SHAP-(Ave5)SG と株価指標 (Target1) の散布図

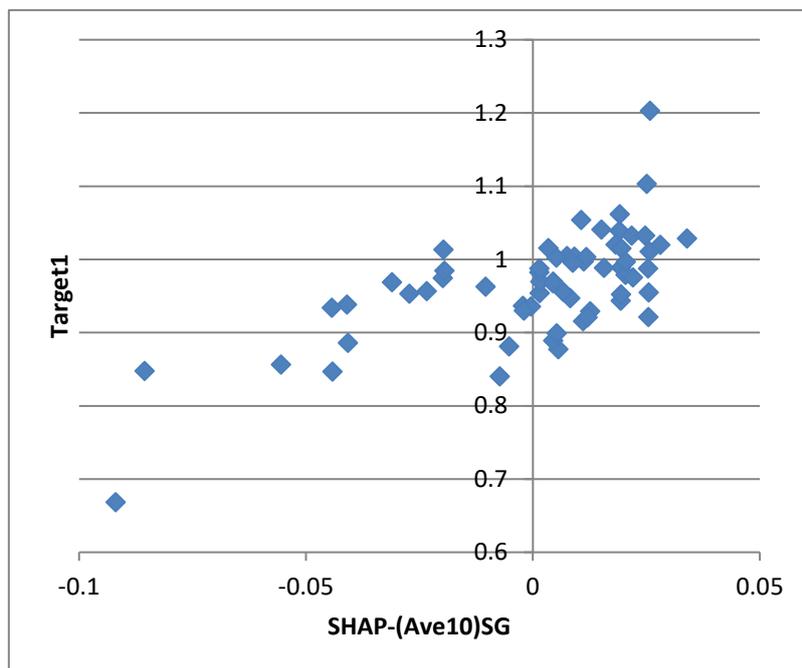


図 2 : ケース 2 における SHAP-(Ave5)SG と株価指標 (Target1) の散布図

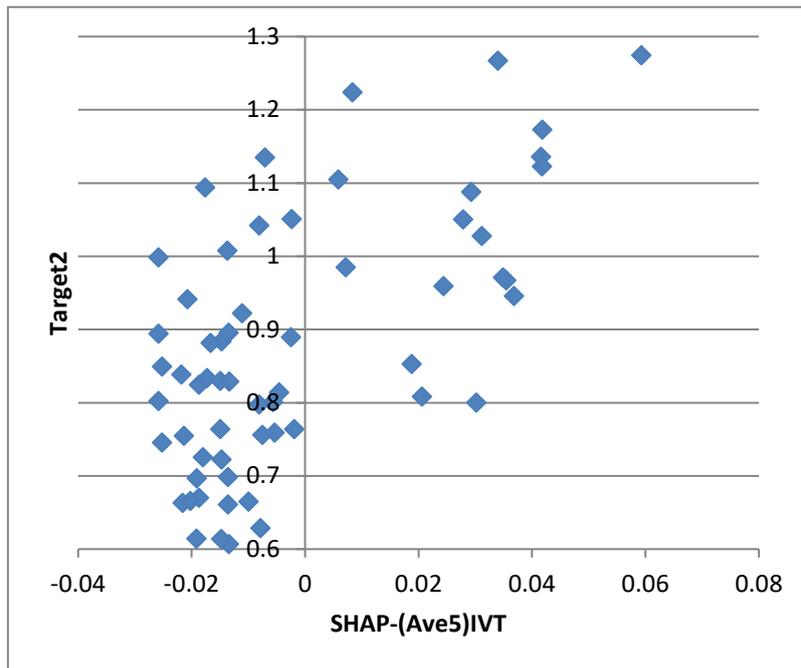


図 3 : ケース 3 における SHAP-(Ave5)IVT と株価指標 (Target2) の散布図

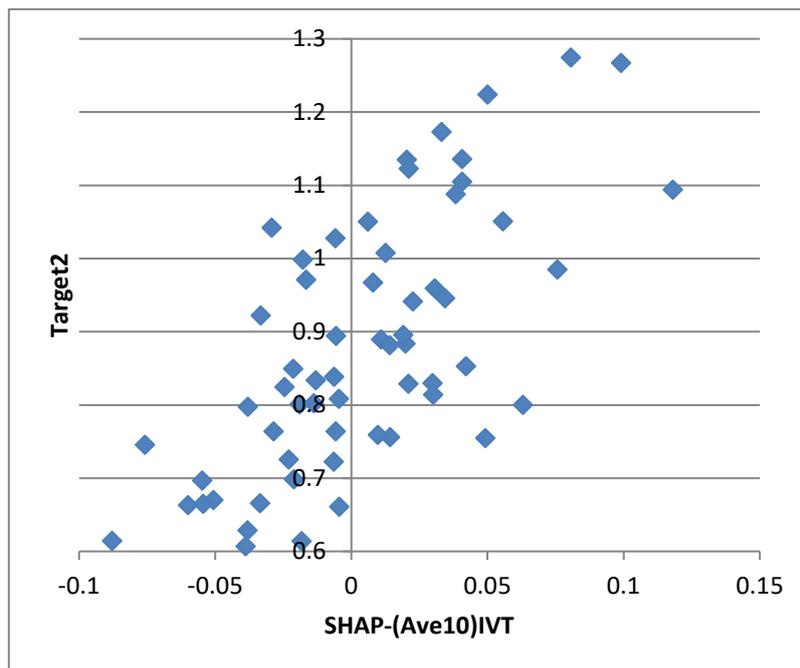


図 4 : ケース 4 における SHAP-(Ave10)IVT と株価指標 (Target2) の散布図

表 12：目的変数と SHAP 値の相関係数の一覧

	ケース 1	ケース 2	ケース 3	ケース 4
SHAP-(Ave5)SG	0.7551	-	0.7343	-
SHAP-(Ave10)SG	-	0.6694	-	0.4203
SHAP-(Ave5)NI	0.5631	-	0.6681	-
SHAP-(Ave10)NI	-	0.5012	-	0.5370
SHAP-(Ave5)IVT	0.4916	-	0.7276	-
SHAP-(Ave10)IVT	-	0.5776	-	0.6392
SHAP-(Ave5)TFA	0.7319	-	0.6472	-
SHAP-(Ave10)TFA	-	0.6354	-	0.2163
SHAP-(Ave5)ROA	0.6498	-	0.2759	-
SHAP-(Ave10)ROA	-	0.6865	-	0.3394
SHAP-(Ave5)ROE	0.5007	-	0.3645	-
SHAP-(Ave10)ROE	-	0.5177		0.5547

表 12 の結果によれば、目的変数である株価指数に対して、影響の度合いが高い説明変数は以下のものとなった。

- 短期的な株価変動に対する分析（ケース 1、ケース 2）では、「売上高・営業収益(SG)」と「ROA」
- 中長期的な株価変動に対する分析（ケース 3、ケース 4）では、「(売上高に対する) 棚卸資産(IVT)」

まとめと今後の課題

株価と財務データの関係性を見るために、重回帰による分析と、機械学習回帰と SHAP 値を用いて分析をそれぞれ行った。説明変数の候補が多い場合、重回帰分では難しいデータに対して、機械学習回帰と SHAP 値を用いて分析を行うことで、株価と財務データの関係性について、業種の全体として傾向と、個別の企業の特徴をある程度見ることが可能となった。

ただし、財務データの選択には注意する必要がある。例えば、説明変数同士で高い相関があった場合、多重共線性が起きてしまうこともある。

参考文献

- [1] 松井久勝、音川和久、会計情報のファンダメンタル分析、中央経済社、2013
- [2] 永島正康、グローバル・サプライチェーンにおける新しい製販協働のかたち、2021
- [3] 経済産業省、持続的成長への競争力とインセンティブ～ 企業と投資家の望ましい関係構築～、プロジェクト最終報告書（伊藤レポート）、2014
- [4] 松井美樹、サプライチェーン・マネジメント、放送大学教育振興会、2021
- [5] 渡部敏明、佐々木浩、ARCH 型モデルと Realized Volatility によるボラティリティ予測と Value-at-Risk、日本銀行金融研究所、Discussion Paper No.2006-J-13、2006

Research Institute of Economic Science
College of Economics, Nihon University

1-3-2 Kandamisaki-cho, Chiyoda-ku, Toyko 101-8360 JAPAN
Phone: 03-3219-3309 Fax: 03-3219-3329
E-mail: keikaken.eco@nihon-u.ac.jp
<http://www.eco.nihon-u.ac.jp/research/economic/>