Analysis of Financial Metrics of Major US Companies Using Machine Learning Techniques for Stock Recommendation

Sergio Beamonte González

UPM

November 4, 2025

1. Problem: Analyst Recommendations availability

- Investors rely on expert analyst recommendations to decide **when to buy or sell companies stocks**.
- These recommendations are rarely free and are even less accessible for **small or non-indexed companies**.

1. Problem: Analyst Recommendations availability

- Investors rely on expert analyst recommendations to decide **when to buy or sell companies stocks**.
- These recommendations are rarely free and are even less accessible for **small or non-indexed companies**.

- **The goal**: infer analysts strategy for recommendations.

2. Can fundamentals predict stock-market behavior?

Question: Can we predict stock recommendations using only public financial fundamentals?

Financial statements and ratios reflect company performance, but can they anticipate how analysts rate a stock?

3. Dataset Overview

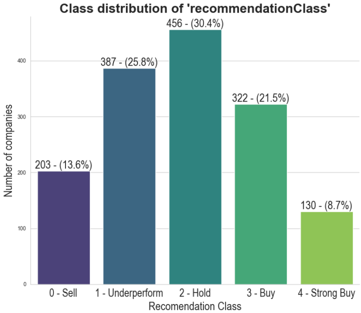
- Dataset: 49 variables from **Russell 3000** companies (Yahoo Finance API).
- Represents companies across all US sectors and market caps.
- Variables grouped into 5 major categories:
  1. **Company Info:** sector, employees, SP500 membership.
  2. **Market Metrics:** marketCap, floatShares, shortRatio.
  3. **Stock Performance:** currentPrice, beta, dividendYield.
  4. **Financial Health:** revenue, ebitda, totalDebt.
  5. **Valuation Ratios:** PER, priceToBook, ROE.

4. Consensus and Dataset Challenges

- We establish an **analyst consensus** based on multiple ratings per company.
- Classes are grouped into 5 categories: **Sell**, **Underperform** (company is expected to do worse than index), **Hold** (equal), **Buy**, and **Strong Buy** (do better).

4. Consensus and Dataset Challenges

- We establish an **analyst consensus** based on multiple ratings per company.
- Classes are grouped into 5 categories: **Sell**, **Underperform** (company is expected to do worse than index), **Hold** (equal), **Buy**, and **Strong Buy** (do better).
- Several **issues**. For instance:
  - Analyst A $\rightarrow$ Buy
  - Analyst B $\rightarrow$ Sell
  - Consensus $\rightarrow$ Hold?
- Highlights the inherent subjectivity and noise in analyst data.



Distriburion of `recommendationClass` variable.

4. Data Preprocessing

- The dataset was tested under six different configurations:
  1. **Raw data (no preprocessing)**.
  2. **Normalized data** (mean 0, variance 1).
  3. **Transformed + normalized data.**
- In each dataset **SMOTE** was also applied to balance target variable.

4. Data Preprocessing

- The dataset was tested under six different configurations:
  1. **Raw data (no preprocessing)**.
  2. **Normalized data** (mean 0, variance 1).
  3. **Transformed + normalized data.**
- In each dataset **SMOTE** was also applied to balance target variable.
- The transformation involved either:
  - Removing the top and bottom 2% of outliers, or
  - Applying a **logarithmic transformation** when it corrected skewness more effectively.
- Each variable was evaluated independently to determine which correction improved symmetry.

4. Data Preprocessing

- The dataset was tested under six different configurations:
  1. **Raw data (no preprocessing)**.
  2. **Normalized data** (mean 0, variance 1).
  3. **Transformed + normalized data.**
- In each dataset **SMOTE** was also applied to balance target variable.
- The transformation involved either:
  - Removing the top and bottom 2% of outliers, or
  - Applying a **logarithmic transformation** when it corrected skewness more effectively.
- Each variable was evaluated independently to determine which correction improved symmetry.
- These preprocessing choices significantly influenced model accuracy and stability.

5. Experimental Setup and Evaluation Metrics

- Each configuration was evaluated using 10-fold cross-validation repeated 5 times.

5. Experimental Setup and Evaluation Metrics

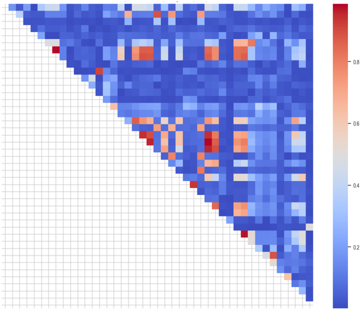
- Each configuration was evaluated using 10-fold cross-validation repeated 5 times.
- **Evaluation metrics:**
  - **Accuracy:** proportion of correct predictions (PC).
  - **AUC (Area Under ROC Curve):** measures ranking quality and class separability.
  - **Root Squared Mean Error (RMSE):** to try to minimize big errors (i.e. Strong Buy predicted as a Sell worse than predicted as a Buy).

5. Experimental Setup and Evaluation Metrics

- Each configuration was evaluated using 10-fold cross-validation repeated 5 times.
- **Evaluation metrics:**
  - **Accuracy:** proportion of correct predictions (PC).
  - **AUC (Area Under ROC Curve):** measures ranking quality and class separability.
  - **Root Squared Mean Error (RMSE):** to try to minimize big errors (i.e. Strong Buy predicted as a Sell worse than predicted as a Buy).
- Best models were selected based on this three and interpretability.

6. Feature Selection

- **Methods used:** Univariate FSS (Mutual Information), Multivariate FSS (Correlation Based FS), and Wrapper methods.
- **Consensus variables among univariate and multivariate FSS:**
  - totalRevenue
  - grossProfits
  - PER (Price-to-Earnings Ratio)
  - priceToBook
  - sector
  - numberOfAnalystOpinions
- These features capture profitability, valuation, and market perception.



(heatmap with variable correlation)

Overview of Tested Models

Non-Probabilistic Models

- Decision Trees (J48)
- Induction Rules (RIPPER)
- **SVM (Polynomial Kernel)**
- ANN (MLP and **RBF Classifier**)
- k-Nearest Neighbors (kNN)

Probabilistic Models

- Naïve Bayes classifiers (A1DE, **A2DE**, TAN)
- **Logistic Regression**
- Lineal Discriminant Analysis (LDA, QDA)

Meta-Models / Ensembles

- **Bagging**
- AdaBoost
- Random Forest
- **Logistic Tree Model (LMT)**
- **Stacking (SVM + A1DE + Logistic Regression)**

7. Model Overview and Results

Model	Accuracy	AUC	RMSE
SVM (Polynomial) (P)	0.365	0.65	0.37
RBF Classifier (P)	0.367	0.67	0.37
Logistic Regression (PW)	0.372	0.67	0.36
A2DE (Bayesian) (PCS)	0.350	0.65	0.38
LMT (PCS)	0.370	0.67	0.36
Bagging RIPPER	0.364	0.68	0.36
Stacking (SVM, A1DE, Log) (P)	0.378	0.69	0.36

Comparison of best models under different configurations.

Key insight

Although these models are the present the best metrics, many of them act like black boxes and we cannot extract conclusions or rules.

8. Logistic Regression – Odds Ratio Interpretation

Variable	Strong Buy	Buy	Hold	Underperform
EBITDA	7.7×	7.0×	5.1×	4.7×
Current Price	5.6×	3.3×	2.2×	1.5×
Revenue Per Share	5.2×	8.2×	4.7×	3.3×
Free Cash Flow	0.8×	0.9×	1.0×	1.1×
PER (P/E Ratio)	0.6×	0.8×	0.8×	1.1×
Price-to-Book	0.5×	0.4×	0.4×	0.7×
SP500 Membership	0.02×	0.11×	0.22×	0.5×

Odds ratios (relative to the reference class “Sell”). Values >1 increase probability of recommendation improvement.

Interpretation

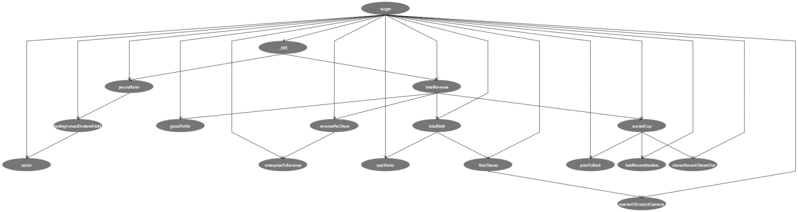
Higher profitability (EBITDA) and reasonable valuation (PER, Price/Book) increase the odds of a positive recommendation. Companies already in the S&P500 with excessive free cash flow are more likely to be rated as “Sell”.

9. SVM Model – Coefficients Between Classes

Variable	0 vs 4	1 vs 4	2 vs 4	3 vs 4
In_SP500	1.6843	0.3209	0.0281	0.0005
EBITDA	-1.7028	-0.2493	-0.1551	-0.0013
TotalCashPerShare	0.6885	0.4936	0.0649	0.0010
TotalDebt	0.3484	-0.1863	-0.0521	-0.0012
QuickRatio	-1.5648	-0.9446	-0.1671	-0.0027
TrailingAnnualDividendYield	2.5395	2.5780	0.1506	0.0015
PayoutRatio	-0.0774	-0.9044	-0.1446	-0.0023
SharesOutstanding	0.2114	-0.1112	0.1631	0.0117
FloatShares	0.1143	-0.6611	-0.1560	-0.0134
ShortPercentOfFloat	1.3889	1.6547	0.1840	0.0013
HeldPercentInstitutions	-0.2445	-0.8532	-0.0538	-0.0030
Intercept	-1.1164	-0.7462	-0.9864	-0.9954

Note: Positive values push the decision toward the second class of the pair (e.g., 0 vs 4: positive values favor 'Sell' over 'Strong Buy'). **Brief interpretation:** DividendYield and shortPercent show strong positive coefficients towards 'Sell'; EBITDA and QuickRatio are negative (better fundamentals → favor Buy").

10. TAN Model – Structure of Dependencies



The Tree-Augmented Naive Bayes (TAN) structure reveals the relationships between key variables: **PER, totalRevenue, and marketCap** emerge as the main parents of *recommendationClass*, confirming that **profitability and valuation** drive most analyst decisions.

11. Conclusions

- **Main Findings:**
  - The **Stacking model (SVM + A1DE + Logistic)** achieved the best performance whereas simpler models like SVM, Logistics or Trees win in interpretability.
  - Data **transformation and normalization** were key to stability and accuracy.
  - Feature importance confirms analysts focus on **profitability, valuation, and market confidence**.

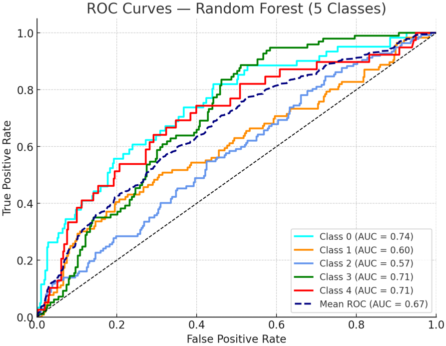
Rules of Thumb

For a _PER $>=$ 30.967348 is a hold, underperform or sell and for a _PER $<$ 30.967348 is a buy or strong buy (with $\approx$ 63% of accuracy).

12. Further Steps

- **Further Steps:**
  - Expand dataset to include **international markets** and longer time series.
  - Integrate other type of metrics.

- **Further Steps:**
  - Expand dataset to include **international markets** and longer time series.
  - Integrate other type of metrics.
  - More conservative approach.



ROC curves for each class.

Thank You!

Questions?

UNIVERSIDAD
POLITÉCNICA
DE MADRID