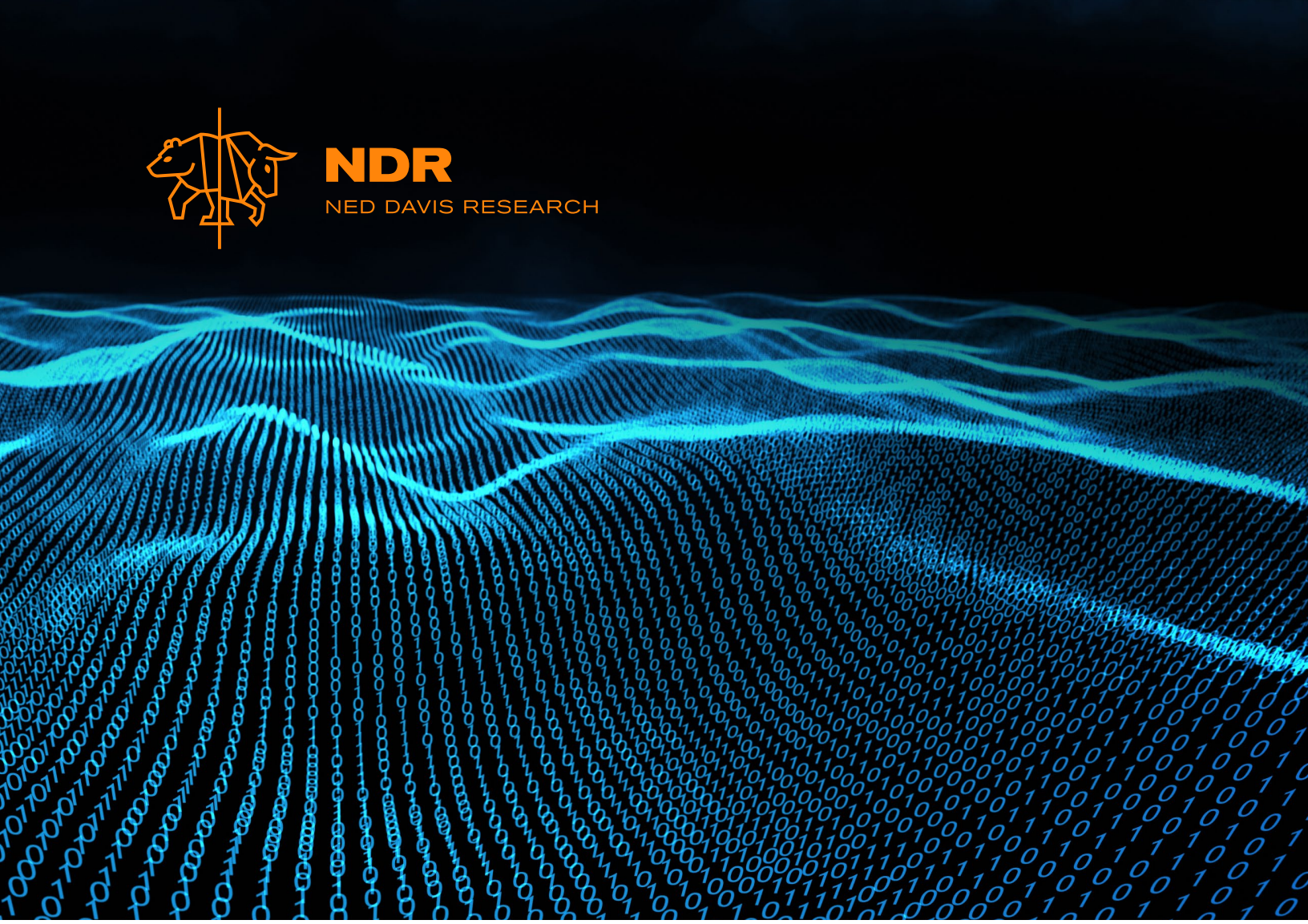


NDR
NED DAVIS RESEARCH

NDR AI Perspectives

JUNE 2022

USING AI/NLP TO TACKLE THEMATIC INVESTING

Author: **Kiersten Engel,** *Senior Research Analyst, NDR*

Executive Summary

Thematic investing is an approach that focuses on long-term trends. At the center of NDR's Thematic Opportunities landscape lie these megatrends : new technologies, demographic shifts, and global shocks. Themes we cover within these megatrends are constantly evolving. They can be highly specific, such as "electric vehicles," or reflect more structural societal shifts, such as "COVID Impact: Work from Home."

At Ned Davis Research, we leverage the expertise of our Thematic Strategists along with cutting-edge AI/Natural Language Processing (NLP) algorithms to create scalable, quantitative solutions to thematic investing, including thematic stock selection and ETF investing.

Key Takeaways:

- Investment advisors and wealth managers will continue to see growing client demand for thematic investments.
- Thematic stock selection is labor-intensive, subjective, and not easily scalable.
- Our proprietary "Thematic Opportunities Score" augments NDR's traditional 360° approach with cutting-edge NLP algorithms to quantify alignment to specified themes.
- Constructing a sample portfolio for the theme "Metaverse," we explore the application of our NLP algorithms.

What is NLP?

Natural Language Processing, or NLP, is a form of Artificial Intelligence that enables machines to break down and interpret natural human language. Natural human language is complex, filled with ambiguities, and constantly evolving. These challenges make natural human language difficult to process using rules-based, statistical approaches. Today's NLP algorithms leverage the latest Deep Learning techniques to take unstructured data, like news articles, chats, and tweets, and enable the machine to 'learn' the meaning from the text to complete various tasks like summarizing, predicting the next word in a search, and determining the sentiment of an article.

Why use AI/NLP in the Investments Process?

According to a 2017 Global X survey, 83% of affluent Millennial investors and 61% of Gen X investors expressed they were extremely interested or very interested in thematic investing, versus 31% of the general population. As these investors continue to grow their assets, investment advisors and wealth managers are expected to see growing demand for thematic investments. [9]

The primary focus of thematic investing is to identify trends driven by unique macro drivers and invest in companies poised to profit from those trends. The best thematic investment opportunities are identified early and are unconstrained by traditional investing styles. They span regions, countries, sectors, industries, and market capitalization. While this freedom provides access to some of the best growth opportunities around the world, powerful tools are required to identify them.

Traditional approaches to thematic stock selection are manual, labor intensive, and not easily scaled. Moreover, development work for one theme is not easily transferred to other themes, making it difficult to stay ahead of emerging trends.

To streamline this process, we leverage the expertise of NDR's Thematic Strategists and cutting-edge NLP algorithms to create our proprietary "Thematic Opportunities Score." This score captures how closely a given company aligns to a user-specified investments theme, and when paired with traditional portfolio selection metrics, it provides a scalable, quantitative solution to thematic investing.

In the sections that follow, we explain NDR's "Thematic Opportunities Score" and how to construct a thematic opportunities portfolio using the "Metaverse" theme as an example.

NDR's Thematic Opportunities Score: How It Works

NDR's Thematic Opportunities Score captures how closely a given company aligns to a user-specified theme

As with any AI-powered model, the most successful algorithms combine human expertise and machine processing power to get the best of both worlds. NDR's Thematic Opportunities Score is no exception. Our scoring algorithm is powered by our expert-selected input and cutting-edge NLP algorithms to produce meaningful results.

Challenges of working with text

Humans communicate massive amounts of information via text, making text data a rich source of information. However, conversion of text data into a format that machines can understand requires careful preprocessing to preserve valuable semantic information. Unstructured text data is full of ambiguity due to polysemy (e.g., "get" can mean procure or understand: "go get it" vs. "I get it"), constantly evolving language (especially on social media) and use of idioms (e.g., "raining cats and dogs"). These issues are exacerbated when using text from multiple languages and sources.

NDR's extensive global databases enable seamless incorporation of our company description text into our Thematic Opportunities Score algorithm. Our database contains English business descriptions for over 60,000 global companies, linked to each company's pricing and fundamental data across all NDR's data providers. This standardization and consistency enable us to generate a comprehensive view of each possible constituent for our thematic stock selection.

How it works

The Thematic Opportunities Score is calculated using our proprietary NLP-based algorithm. It scores companies according to how closely their business description aligns with a user-defined target theme. The target theme comprises expert-selected companies, descriptions, or keywords that exemplify the theme. The proprietary scoring algorithm compares the descriptions of over 60,000 global companies to the selected target and scores each company from 0 (not at all related) to 100 (perfectly aligned) with the target theme.

Themes are constantly evolving. They can be highly specific, such as "electric vehicles," or reflect more structural societal shifts such as "COVID Impact: Work from Home." Therefore, careful selection of the target set is critical to generating meaningful results.

Companies selected for the target set do not need to be publicly traded, or even exist. We can create fictitious company descriptions if there are no companies that appropriately capture the theme. This feature allows us to capture emerging trends and technologies before they become mainstream.

For example, if we want to build a portfolio for "Travel to Mars," we may want to use SpaceX as our target company and "Mars," "aerospace," and "spacecraft" as our keywords. Not many publicly traded companies are well aligned with this theme, but the Thematic Opportunities Score provides meaningful results even with a very niche target theme.

Example: "Travel to Mars" Theme

Target Company: SpaceX designs, manufactures and launches the most advanced rockets and spacecraft.

Target Keywords: Space launch, human space flight, aerospace, Mars, spacecraft.

Top Scoring Companies within Global Defense and Aerospace Subindustry

Theme Score	Company Description
100.00	Rocket Lab USA Inc. delivers launch services, satellite manufacture, spacecraft components, and on-orbit management solutions.
99.19	Astro Aerospace Ltd. develops autonomous manned and unmanned and electric vertical take-off and landing aerial vehicles.
96.89	Virgin Galactic Holdings Inc. develops human spaceflight for private individuals and researchers and manufactures air and space vehicles.
95.32	Virgin Orbit Holdings Inc. is a launch and space solutions company. It builds and operates space launch systems and lauches satellites into orbit.
94.75	The Boeing Company designs, develops and supports military aircraft, satellites, human space flight and launch systems and commercial jetliners.

Bottom Scoring Companies within Global Defense and Aerospace Subindustry

Theme Score	Company Description
0.00	POSTD Merchant Banque is a financial services company that offers clients access to growth capital in the private markets.
3.91	Hubei Huaqiang High-Tech Co. Ltd. manufactures rubber products.
7.81	Shanrong Biotechnology Corp. is a project management, executive consulting, and business development firm.
14.76	Vector Aerospace Corporation provides aviation maintenance repair and overhaul services for turbine engine aircrafts.
15.27	Golden Star Enterprises Ltd. provides alternative renewable fuels.

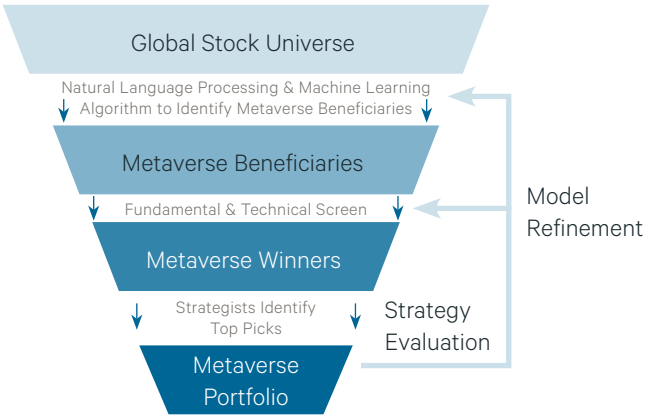
Source: Ned Davis Research

Case Study: Constructing a Thematic Portfolio – Metaverse Example

In this section we construct a "Metaverse" thematic sample portfolio using NDR's proprietary Thematic Opportunities Score.

The "Metaverse" is a network of 3D virtual worlds focused on social connection. The term is often described as a hypothetical iteration of the Internet as a single, universal virtual world that is facilitated using virtual and augmented reality. [8]

Consistent with NDR's proven **360° approach**, we construct our Metaverse portfolio starting with a broad universe of over 40,000 global companies. We refine the universe using our Metaverse Thematic Opportunities Score and fundamental metrics.



Selecting Metaverse Beneficiaries Using Metaverse Thematic Opportunities Score

As discussed above, the first step in creating a Metaverse Thematic Opportunities Score is to select a set of companies and keywords that exemplify "Metaverse" for our theme target. For example, we might include words like "virtual reality," "3D virtual space," "augmented reality," "avatar," etc. We also want to include companies that create content for the Metaverse (like META), companies that support the hardware and infrastructure (like NVIDIA), and those that create and manage digital assets and support commerce (like PayPal). The figure below shows an example target set and the companies with the top Metaverse theme scores.

1 Input

Global Universe

Raw text describing business and operations for all MSCI ACWI companies from NDR's Global Database

2 Build Pipeline/Corpus

NLP

Example: Entity Recognition

Larry Page person and Sergey Brin person founded Google org in September 1998 date .

Since then the company has grown to more then 130 000 quantity employees worldwide...

3 Compute Customized Metaverse Theme Score

Tokenized Text

Quintessential Metaverse Companies

Known "Metaverse" Keywords

Proprietary Algorithm

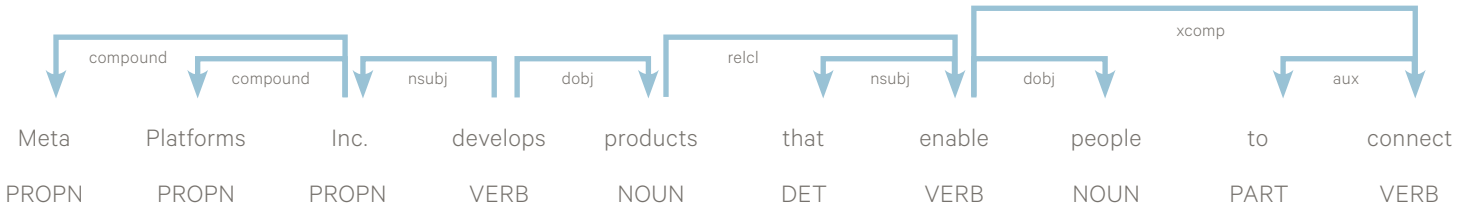
4 Select Universe of Metaverse Beneficiaries

Thematic Opportunities Score

Metaverse Beneficiaries Universe

Top Scores

Example: Text Relationships and POS Tags



Now that we've selected our target set, we can create our Metaverse Thematic Opportunities Score and select the top scoring companies to create a Metaverse Beneficiaries sub-universe. The Metaverse Beneficiaries are a subset of our Global Universe related to the Metaverse theme.

Top Thematic Opportunity Scores for "Metaverse" Theme

Theme Score	Company Description
100.00	Unity Software Inc.
93.43	Meta Platforms Inc.
92.55	Matterport Inc.
91.95	ON24 Inc.
91.57	Roblox Corporation

Selecting Metaverse Winners Using Fundamental and Technical Metrics

Although the theme and potential investment candidates have been identified, our job is not done yet. We must determine which companies are best positioned to profit from the Metaverse investment theme and build a robust portfolio.

NDR has a proven track record of selecting predictive fundamental and technical indicators that we use in our sector, stock market, and asset allocation models. We review our indicators on a regular basis for effectiveness and are continually incorporating new ideas as markets evolve.

After we determine Metaverse Beneficiaries using our Thematic Opportunities Score, we further refine our selection, using more traditional factor analysis to assess the quality of prospective companies. This analysis applies fundamental and technical metrics to differentiate between high- and low-quality companies. High-quality companies tend to offer more stable returns and outperform low-quality companies over long-term holding periods.

We consult our NDR Thematic Opportunities Strategists to identify a pool of metrics that are likely to detect high-quality companies within our Metaverse Beneficiaries universe. Consistent with NDR's 360 approach, we evaluate company fundamentals that indicate how markets *should* be acting (such as Earnings, Book Value, etc.) and technical metrics (such as 1-year momentum, etc.) that reveal how markets *are* acting. Combining fundamentals and technical metrics provides a balanced view of the market and identifies return drivers for high-quality businesses. Over 100 metrics are considered, based on their ability to identify the best performing companies within the Metaverse Beneficiaries universe.

Only the most informative metrics are included in our screen. For a metric to be included, it must meet the following criteria:

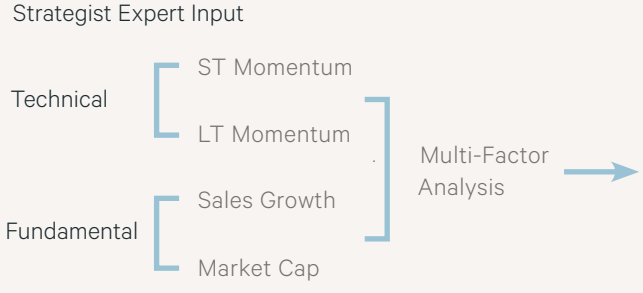
- **Metric is proven to differentiate between outperforming companies and underperforming companies** (See appendix for details.)
- **Metric signal is consistent with economic rationale** (e.g., The top "sales growth" quartile should outperform the bottom "sales growth" quartile, because increasing sales benefits the company. If the bottom quartile outperformed the top, this would contradict the rationale and "Sales Growth" would be excluded.)
- **Metric is relevant to the theme** (e.g., "R&D spending" is relevant for themes like "electric cars," but likely less relevant for themes like "vacation.")

While most metrics are designed to work across all themes, some may not be transferable. Therefore, it is critical to assess the predictive power of potential metrics for each theme. (See the appendix for metric evaluation methodology.)

5 Input Metaverse Beneficiaries

Metaverse Beneficiaries Sub-Universe

6 Fundamental/Technical Screen



7 Select Top 50 for Portfolio

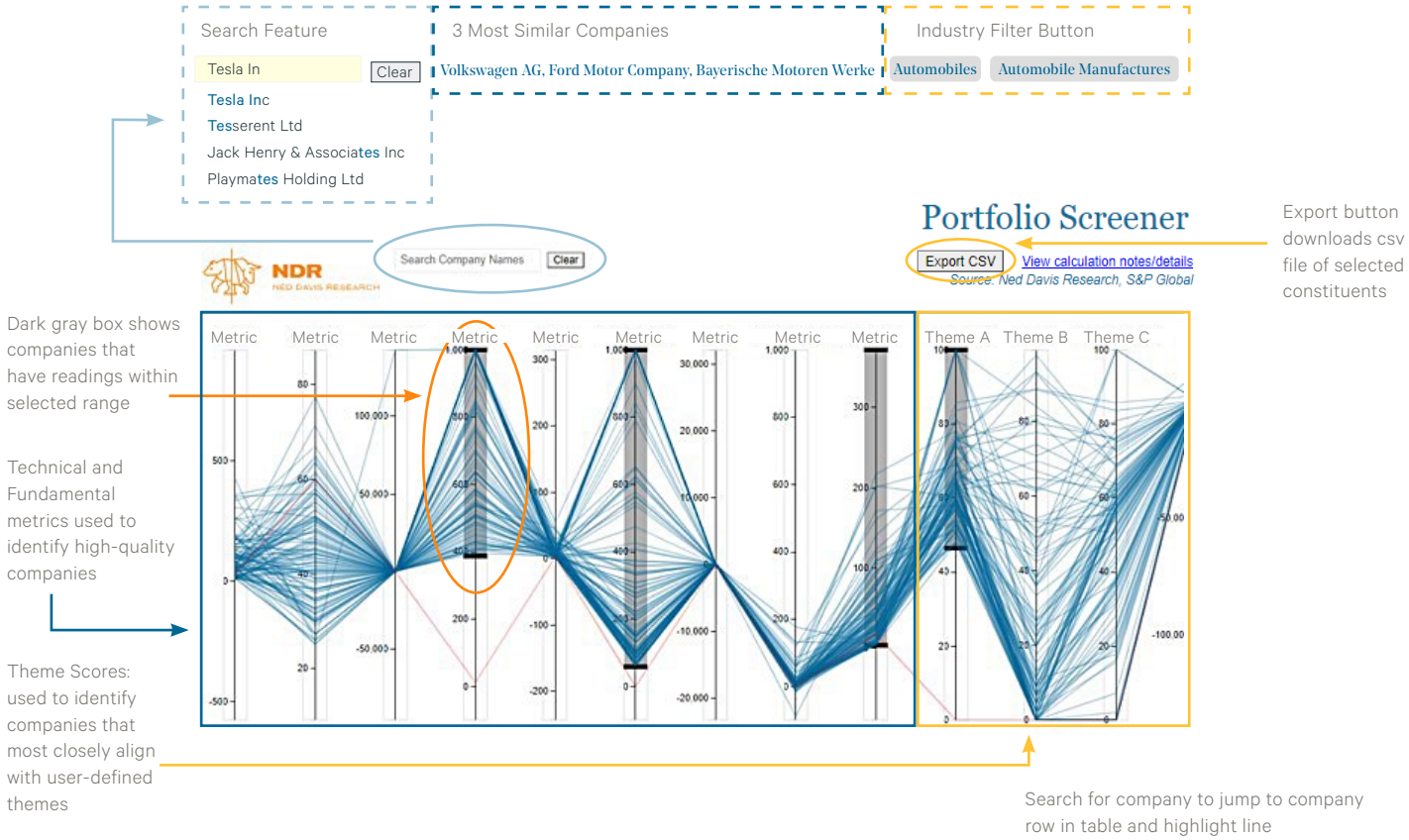
Metaverse Portfolio

Final Portfolio Selection

After constructing NDR's Thematic Opportunities Score and selecting the most meaningful fundamentals and technical metrics, we combine everything into a Parallel-Coordinates Framework with proprietary enhancements designed for final selection.

Using this framework (shown below), we can instantly filter our universe according to metrics that are best aligned with our Metaverse theme. This framework allows us flexibility to identify results for any combination of criteria, including one or more of NDR's Thematic Opportunities Scores. This framework not only maximizes efficiency and flexibility at the portfolio inception phase but also during reselection and rebalance cycles, and it can easily be extended to include NDR's Thematic Opportunities Theme Score for other themes.

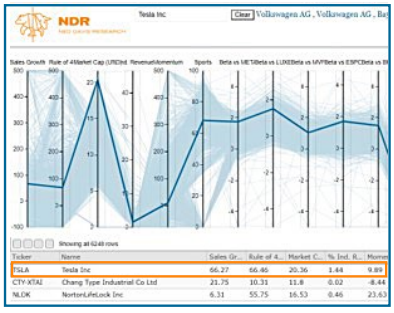
Parallel-Coordinates Portfolio Selection Tool



Ticker	Name	Metric	Metric	Metric	Metric	Metric	Metric	Metric	Theme A	Theme B	Theme C
AMZN	Amazon com Inc	18.62	-2.94	21.27	46.08	7.77	0.69	0.43	0.26	0.67	0.4
NET	CloudFlare	36.58	28.82	17.63	0.19	87.16	1.22	0.54	0.22	1.24	0.42
PYPL	PayPal Holdings Inc	14.52	21.28	19.35	7.87	-22.34	1.06	0.66	0.53	0.94	0.79
FUBO	fuboTV In	135.31	16.12	15.04	0.07	-29.84	0.84	0.97	0.59	0.38	0.98
ADSK	Autodesk Inc	15.44	36.52	17.98	0.69	1.46	0.89	0.75	0.55	0.73	0.79
APPN	Applan Corporation -CL A	13.57	4.23	15.01	0.06	-50.95	1.56	1.12	0.98	1.36	1.18
AKAM	Akamai Technologies Inc	6.38	35.82	16.7	1.09	14.86	0.48	0.35	0.29	0.53	0.41
ADBE	Adobe Inc	22.67	61.78	19.51	2.6	19.29	0.85	0.56	0.38	0.77	0.62
LUMN	CenturyLLink Inc	-3.61	39.56	16.41	25.34	-0.44	0.41	0.63	0.64	0.28	0.68
SPT	Sprout Social Inc Com CL A	29.33	16.71	15.46	0.03	57.01	1.63	1.21	0.97	1.34	1.35

Hover over company in table to view description and highlight line in chart

Search for company to jump to company row in table and highlight line



After applying selected filters, our strategists review the selected top constituents across all categories, and we create our final Metaverse portfolio. The companies can be weighted using measures such as market capitalization, fundamentals, or the theme score.

Metaverse Winners (representative example of constituents)

Meta Platforms Inc.	Cloudflare	Alphabet Inc.
Snap Inc.	Twilio Inc.	Autodesk Inc.
Twitter Inc.	Unity Software Inc.	Microsoft Corp.
NAVER Corp.	VMware Inc.	NVIDIA Corp.
Apple Inc.	Block Inc.	Roblox Corp.

SAMPLE

Conclusion

NDR's Thematic Opportunities Score and AI/NLP-powered thematic stock selection process outlined in this paper were developed for clients seeking to build thematic investment portfolios. Combining traditional technical and fundamental metrics and our Thematic Opportunities Score within a Parallel-Coordinates Framework reduces stock selection time and scales to incorporate additional themes with maximum efficiency.

Glossary

Artificial Intelligence (AI)
The use of algorithms, designed to imitate human intelligence, programmed to discover patterns or structure in data.

Deep Learning
A subset of Machine Learning and Artificial Intelligence based on Neural Network models, inspired by the structured network of neurons in human brains.

Natural Human Language
Any language that has developed organically through use and repetition. It does not necessarily follow structured rules and is free to evolve naturally.

Natural Language Processing (NLP)
A form of Artificial Intelligence that enables machines to break down and interpret natural human language. NLP algorithms take unstructured data like news articles, chats, and tweets, and enable the machine to "learn" the meaning from the text to complete various tasks like summarizing, predicting the next word in a search, and determining the sentiment of an article.

NDR's 360° Approach
NDR's 360° approach combines fundamental, macroeconomic, technical, and sentiment research disciplines. Fundamentals and macroeconomics tell us how the markets should be acting, while technicals and sentiment reveal how the markets are acting. Truly insightful and timely ideas demand a balance between fundamentals and technicals. Actionable ideas meet balanced, strategic insights through our 360° methodology.

Successful investing is seeing the signals and making informed decisions. To do this, we provide a clear view of the whole investment picture, including a perspective on sentiment. Historically, NDR has relied on quantitative indicators within this framework, but advancements in text classification have created an opportunity to enhance the 360° approach.

Machine Learning
A subset of Artificial intelligence wherein algorithms are designed to learn and uncover patterns in data without being given the underlying structure or rules.

Unstructured data
Data that is not well organized for storage or analysis, such as tweets, online product reviews, news articles, etc.

Parallel-Coordinates
A visualization framework that allows the user to view a dataset across many different dimensions simultaneously.

KIERSTEN ENGEL
Senior Research Analyst, NDR

Kiersten Engel applies quantitative methods to provide clients with a broad scope of analysis, including economic studies, and custom reporting and modeling. Kiersten joined Ned Davis Research Group in 2014.

Kiersten graduated **magna cum laude** from the University of Alabama at Birmingham (UAB) with a Bachelor of Science degree in Applied Mathematics and a minor in Finance. She has also been featured in **UAB Magazine** for her work in mathematics and her career as a ballet dancer with the Alabama Ballet Company.

Kiersten is a **summa cum laude** graduate from the Georgia Institute of Technology (Georgia Tech) with a Master of Science in Analytics focusing on Computational Analytics (machine learning). She continues her research by leading NDR's partnership with Georgia Tech's MS in Analytics Program to sponsor research projects for graduate students. She designs data science projects that apply the latest machine learning techniques to financial markets and serves as an advisor to 50+ graduate students. She also serves as a Head of Teaching Assistant for Georgia Tech's Business Fundamentals for Analytics Masters course.

Contact us

+1-941-412-2300
www.ndr.com

NDR's Custom Research Solutions (CRS) group offers a premium service that builds models and stock portfolios to meet specific client needs. Our analysts apply the techniques outlined in this publication to stock selection while accommodating unique client constraints including turnover, risk tolerance, and sector/country exposures.

Interested in developing customized theme scores to enhance your thematic stock selection process? Please contact NDR's **custom solutions team.**

Appendix

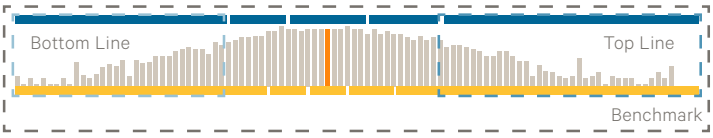
Technical and Fundamental Metrics Evaluation

To determine if selected technical and fundamental metrics are accurately differentiating between high- and low-quality companies, we simulate a stock selection strategy that invests in two portfolios. Each portfolio contains the companies with metric readings in the top and bottom quintiles, respectively.

For example, consider the readings for the metric "Momentum" (shown below). The Momentum back-test compares the performance of two portfolios: the Bottom Quintile portfolio, which invests only in companies with momentum readings in the bottom quintile, and the Top Quintile portfolio, which invests only in companies with momentum readings in the top quintile.

To simulate performance, we compute momentum readings for each company at the end of every quarter, then rebalance the "Top" and "Bottom" strategy portfolios to invest in the companies assigned to their respective quintiles. We repeat this process every quarter and compound the returns to generate a simulated performance. To round out our assessment, we also create an equal-weighted benchmark that invests in all companies every quarter to simulate the "buy-hold" scenario (or the control group).

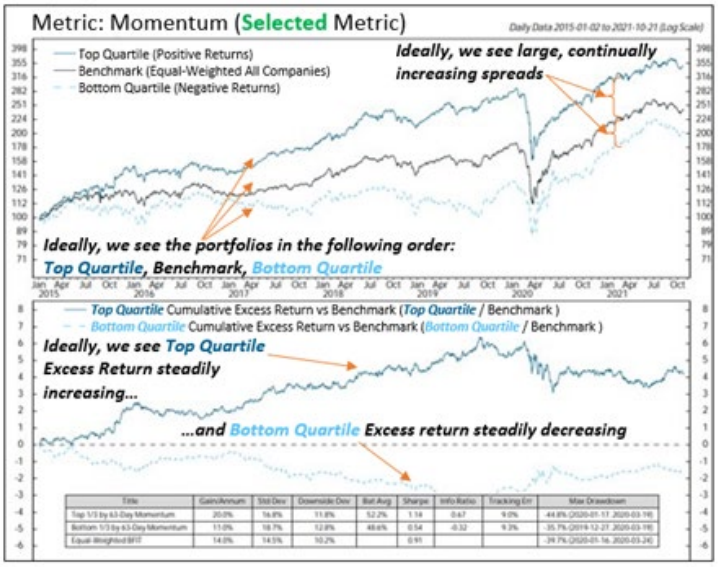
Metric: Momentum Distribution of Stocks in Universe

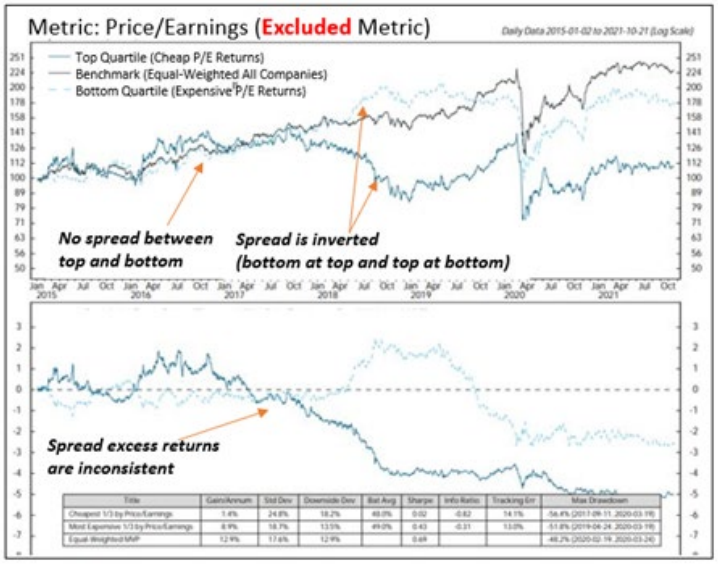


If the metric is predictive, the Top Quintile portfolio should consistently outperform and the Bottom Quintile portfolio should consistently underperform the benchmark. This relationship is best observed by examining the bottom panel of the charts on the right, which show the cumulative excess return of each portfolio relative to the benchmark. The chart at top right (Selected metric) shows a consistent, steady, positive slope of the Top Quintile performance relative to the benchmark in the bottom panel (solid grey line), indicating that the Momentum metric is consistently adding value. The consistent, steady negative slope of the Bottom Quintile shows that companies with relatively low Momentum consistently underperform. The trends observed in both the Top and Bottom Quintiles indicate that Momentum is correctly differentiating between high- and low-performing companies and should be included in the Portfolio Selection Tool.

Metrics that are not predictive of quality exhibit inconsistent spreads between the Top and Bottom Quintiles or counter-intuitive relationships. These trends suggest that the metric is not informative and should be excluded from the tool (see bottom chart).

Example of Metric Evaluation Charts





References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[4] Andrew Trask, Phil Michalak, and John Liu. sense2vec - a fast and accurate method for word sense disambiguation in neural word embeddings, 2015.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[6] Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston. Starspace: Embed all the things!, 2017.

[7] Stefan Jansen. Hands-on Machine Learning for Algorithmic Trading, Packt Publishing, 2018.

[8] Wikipedia, "Metaverse," 9 February 2022. [Online]. Available: https://en.wikipedia.org/wiki/Metaverse.

[9] Global X, "Beyond Baby Boomers," 2017. [Online]. Available: https://www.globalxetfs.com/content/files/Beyond-Baby-Boomers-2017.pdf

Disclaimer

Ned Davis Research, Inc. (NDR), any NDR affiliates or employees, or any third-party data provider, shall not have any liability for any loss sustained by anyone who has relied on the information contained in any NDR publication. In no event shall NDR, any NDR affiliates or employees, or any third-party data provider, be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs) in connection with any use of the information contained in this document even if advised of the possibility of such damages.

The data and analysis contained in NDR's publications are provided "as is" and without warranty of any kind, either expressed or implied. The information is based on data believed to be reliable, but it is not guaranteed. **NDR DISCLAIMS ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY, SUITABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE.**

Full terms of service, including copyrights, terms of use, and additional disclaimers are available at https://www.ndr.com/web/ndr/terms-of-service.