This report explains how the vast amount of content generated across online platforms (big data)\(^1\) is transforming the investment management industry. Data-driven insights have evolved from the consumer products world to the financial world as money managers seek an informational edge to enhance their investment processes. The exponential growth of investment-related content across online platforms provides a vast dataset to analyze in order to extract actionable insights which can potentially identify stocks that will outperform.
Powerful Technologies Can Unlock Investment Insights from Big Data

**BIG DATA, ARTIFICIAL INTELLIGENCE & NATURAL LANGUAGE PROCESSING**

The internet has connected the world as never before, breaking down informational barriers and allowing us to share and collaborate in ways previously unimaginable. Big data has come to represent the collective information that is available from varying sources all over the world. Big data analysis can yield a wealth of knowledge that can be used to track, explain, and potentially predict, the big picture.

**The Digital Universe: 50-fold Growth from the Beginning of 2009 to the End of 2020**

![Image of The Digital Universe: 50-fold Growth from the Beginning of 2009 to the End of 2020]

Source: IDC's Digital Universe Study, sponsored by EMC, December 2012. 2016-2020 growth is estimated. There is no assurance that the projections will be met.

Processing and making sense of big data is beyond what any individual could analyze on their own. Advances in artificial intelligence (AI) techniques used in big data analysis allows for a deeper understanding of vast amounts of data, specifically data from the ever growing social media landscape. While AI may conjure thoughts of cyborgs and robots, at its core, AI is designed to help process data, make sense of it, and act on it.

Natural Language Processing (NLP) is the branch of computer science that deals with the interpretation, meaning, classification, and sentiment analysis of text. Advances in NLP techniques allow for more accurate sentiment analysis than ever before. Moving beyond traditional key word searches, Natural Language Processing today has evolved to provide individuals insights regarding the contextual meaning of datasets.

**GROWTH OF ONLINE PLATFORMS**

The emergence of online platforms has opened up numerous new forums for people to express their opinions and share experiences. In the U.S., 78% of the population has a social network profile. Headlines cross Twitter feeds faster than news networks. The proliferation of online interactions has created a vast set of data that can be analyzed to understand behaviors, trends and sentiments in order to identify actionable insights.
Percentage of U.S. Population with a Social Network Profile from 2008 to 2016

Source: Edison Research; Triton Digital. As of January 31, 2016. 2016 data is an estimate.

BIG DATA ANALYSIS FIRST DEPLOYED FOR BRAND AND CONSUMER ANALYTICS

Consumer product companies were the first to use predictive analysis mined from vast online datasets to help them identify marketing insights and manage their brands. Across various online platforms, including social media, individuals routinely express their opinions of a product, experiences within a store, and general encounters with various brands. Large technology companies such as salesforce.com and IBM offer numerous big data analytical tools for brand and sentiment monitoring in order to help the world’s leading brands capture insights from online and alternative datasets.

ADOPTION OF BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE APPLICATIONS WITHIN THE INVESTMENT COMMUNITY

The investment world – which has always been on the lookout for new ways to gain an informational advantage – has likewise recognized the value of big data. A growing number of asset managers are incorporating big data analytics within their investment processes.

Early attempts at stock sentiment analysis included offline (phone, paper) surveying, in-store questionnaires, and even door-to-door polling. However, it is well documented that people often say one thing when participating in a survey but then act in a different manner. Searching for deeper insights with respect to a company’s investment merits, hedge funds and other asset managers deployed specialty research firms to position canvassers at physical store locations, counting customers and conducting ad-hoc polls in an effort to understand sales patterns and product demand. Not only were these methods time-consuming and resource-intensive, they were also limited in the number of people they could reach.
Today, millions of people voluntarily share their views not only about their own consumer behaviours and brand perception, but also their activities related to their investment ideas and portfolios. Over the past few years, user generated content has accelerated at an exponential rate. The result has been a proliferation in the breadth of discussion, depth of data and diversity of subject matter. Twitter’s adoption of the “Cashtag” methodology ($TICKER$) in the summer of 2012 further enabled the growth of online content by providing a standardized structure to track stock specific discussions. Since then, Twitter has seen growth in finance discussion grow by >4X while dedicated investment community platforms have seen discussion growth exceeding those levels.

StockTwits conversation volume now exceeds 1.5MM messages per month, covering more than 7000+ $Tickers

"With the introduction of cashtagging, discussions about individual stocks on social media has proliferated."
**FINANCIAL APPLICATIONS OF INVESTMENT INSIGHTS FROM THE SOCIAL MEDIA LANDSCAPE**

Large quantitative hedge funds were early adopters in integrating insights from alternative datasets into their investment processes. This investment approach was further validated when leading financial data providers such as Bloomberg and Thomson Reuters began offering sentiment insights based on online content to their subscribers. More recently, sentiment insights generated by artificial intelligence analytics models have been offered to retail investors of several large brokerage firms. TD Ameritrade and Fidelity now provide sentiment scoring metrics to their clients in addition to traditional fundamental research and technical indicators.

**CAN SENTIMENT PREDICT STOCK PRICES?**

Traditional asset pricing theory, fundamental analysis, relative value and other factor based approaches are typical methods deployed by investors in determining the investment merits of stocks. These methods are proven over a long time horizon; however, they often fail to explain short and medium-term stock price fluctuations. So, what influences asset price movements over these time periods? In our view, the answer is sentiment. Numerous studies have shown that investor sentiment plays an important role in financial markets. Investors have historically relied on rudimentary survey-based measures of sentiment for insights. Investor surveys such as the University of Michigan Consumer Sentiment Index and the AAII investor sentiment survey suffer from limitations associated with traditional survey-based approaches including negative bias potential, low incentive for truth-telling and reporting time-lags. Sentiment indicators derived from insights from online platforms have a distinct advantage relative to traditional survey-based approaches. User comments within the social media landscape are voluntarily generated and posted in real-time. The desire for collaboration and attaining social influence further self-regulates the discussion while fostering an environment with a high incentive for truth-telling.

In fact, there are numerous research studies published by academia relating to the potential power of sentiment insights derived from data from various online platforms. These academic research reports find compelling evidence linking investor sentiment data derived from online sources including social media, news articles and blog posts in predicting future stock performance. Notably, in the report titled “Identification of a Social Media Equity Factor Derived Directly from Tweet Sentiments”, professors Jim Liew and Tamas Budavari found “significant evidence that the characteristics of securities derived from social media information sources have significant power in explaining the time-series of daily returns”, while Patrick Houlihan and Germán Creamer found a similar relationship in their research report titled “Leveraging Social Media to Predict Continuation and Reversal in Asset Prices”. Their research showed “that message volume and sentiment from StockTwit messages contained information about future price changes.”
Summary

**EXponential Growth of Big Data Across Online Platforms**

The exponential growth of discussions about individual stocks from online sources including social media platforms, news articles, and blog posts provides a rich new dataset.

**The Potential to Gain an Informational Edge**

Advances in AI and NLP technologies enable investors to harness this vast dataset to identify actionable investment insights about individual stocks, a task that a few years ago would have been impossible.

**Growing Evidence that Investor Sentiment Can Be Predictive of Stock Returns**

There is a growing body of academic research that links investor sentiment with future stock performance. See Appendix for a sample of related studies.
The BUZZ NextGen AI US Sentiment Leaders Index ("BUZZ Index") employs leading edge analytics to harness the investment insights derived from vast datasets generated across online platforms.

The BUZZ Index identifies 75 U.S. stocks with the most positive investment insights and sentiment measures.

**ONLINE SOURCES**

- Yahoo
- Thomson Reuters
- reddit
- StockTwits
- CNW
- The Wall Street Journal
- Google
- Motley Fool
- MarketWatch
- Twitter
- Financial Times
- CNN Money
- Seeking Alpha
- LinkedIn
- The New York Times
- Investopedia
- Bloomberg
- Twitter
- The Street
- Google
- The New York Times
- Seeking Alpha
- LinkedIn
- CNN Money
- Thomson Reuters
- The Wall Street Journal
- Google
- Motley Fool
- MarketWatch
- Twitter
- Financial Times
- CNN Money
- Seeking Alpha
- LinkedIn
- The New York Times
- Investopedia
- Bloomberg
- Twitter
- The Street
- Google

The data is filtered through an analytics model which utilizes Natural Language Processing Algorithms and Artificial Intelligence applications.

**THE BUZZ PROCESS**

<table>
<thead>
<tr>
<th>NATURAL LANGUAGE PROCESSING SOFTWARE</th>
<th>ARTIFICIAL INTELLIGENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. First, from a defined investment universe of U.S. stocks, the BUZZ Index targets the most mentioned stocks across online sources including social media platforms, news articles and blog posts.</td>
<td>2. The BUZZ Index looks at what the online community is saying about these most-mentioned stocks and whether the sentiment is positive, negative or neutral on a collective basis.</td>
</tr>
<tr>
<td>3. The BUZZ Index then measures the influence within the online community of those individuals who are discussing each stock and their historical accuracy.</td>
<td></td>
</tr>
</tbody>
</table>

**INSIGHTS SCORING**

1. Select the 75 stocks with the most bullish scores
2. Reconstitute the Index each month to capture the latest Big Data driven insights, positioning the Index to capture the highest return potential.
Appendix

The following is a sample of available research within the field:

BUZZ NEXTGEN AI SERIES INDICES: US SENTIMENT LEADERS INDEX
A Primer for Investors

IMPORTANT DISCLOSURES & DEFINITIONS

BUZZ NextGen AI US Sentiment Leaders Index and any other owned by BUZZ (the “BUZZ Indexes”) are a product of BUZZ Indexes Inc. (“BUZZ”). “BUZZ Indexes” and “BUZZ Social Media Insights” are trademarks of BUZZ. All materials published on and services offered through the BUZZindexes.com website (the “Content”) are protected by copyright and owned or controlled by BUZZ or the party credited as the provider of such content. Redistribution or reproduction of any BUZZ trademark or Content in whole or in part is prohibited without written permission.

The Content does not constitute an offer of services in jurisdictions where BUZZ or its affiliates do not have the necessary licences. All information provided by BUZZ or its affiliates is impersonal and not tailored to the needs of any person, entity or group of persons. BUZZ receives compensation in connection with licensing the BUZZ Indexes to third parties. Past performance of the BUZZ Indexes is not a guarantee of future results.

It is not possible to invest directly in any BUZZ Index. BUZZ does not sponsor, endorse, sell, promote or manage any investment fund or other investment vehicle that is offered by third parties and that seeks to provide an investment return based on the performance of a BUZZ Index. BUZZ does not make any assurance that investment products based on a BUZZ Index will accurately track index performance or provide positive investment returns. BUZZ is not an investment advisor and it makes no representation regarding the advisability of investing in any such investment fund or other investment vehicle. A decision to invest in any such investment fund or other investment vehicle should not be made in reliance on any of the statements in the Content. Prospective investors are advised to make an investment in any investment fund or other vehicle only after carefully considering the risks associated with investing in such funds or vehicles, as detailed in an offering memorandum or similar document that is prepared by or on behalf of the issuer of the investment fund or other vehicle. Inclusion of a security within a BUZZ Index is not a recommendation by BUZZ to buy, sell, or hold such security, nor is it considered to be investment advice. Any data on past performance, modelling or back-testing contained in the Content is no indication as to future performance. No representation is made as to the reasonableness of the assumptions made within or the accuracy or completeness of any modelling or back-testing. BUZZ Index returns represent past performance and are not indicative of any specific investment. The BUZZ Indexes and Content (including any of the output derived from any analytic tools or models) is not intended to predict actual results, which may differ substantially from those reflected. Past performance is not necessarily indicative of future results.

The Content, including any data, quantitative models, analytic tools and other information and services made available in the Content, are believed to be reliable, but BUZZ does not represent that the Content (including information obtained from third party sources) is accurate, complete or error free, and it should not be relied upon as such. The Content is provided for informational purposes only and is made available “as is”. Because of the possibility of human and mechanical errors as well as other factors, there can be no assurance of timeliness, reliability, performance or continued availability of the Content. BUZZ accepts no responsibility for any errors, omissions, delays, interruptions, completeness, continued availability or currency of or in the Content (including but not limited to the calculation or performance of any index and/or the output of any quantitative model or analytic tool). In no event shall BUZZ or its affiliates 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 Content even if advised of the possibility of such damages. BUZZ disclaims any and all express or implied warranties, including, but not limited to, any warranties of merchantability or fitness for a particular purpose or use, freedom from bugs, software errors or defects, that the Content availability will be uninterrupted or that the Content will operate with any software or hardware configuration.