Stratifyd

A.I.-Powered Data Analytics Platform

Simplifying the Science of Analytics

CONTENTS

Introduction

Traditional data analytics requires the user to know what they are looking for at the beginning of a study. Analysis is often performed when there is good prior knowledge about the characteristics of the data set and the problem being addressed. This allows questions to be developed in advance of the analysis.

A challenge is presented when customers, products, and environments change, and the traditional analytics tools of taxonomies and ontological maps must evolve. Another challenge is specialized skills of data scientists are required to update the tools. These updates need to be specifically identified and implemented, resulting in time delays in responding to changes in the data. Fortunately, this complex process of analytics can now be simplified using Machine Learning.

Machine Learning

Machine Learning algorithms learn by example

They do so automatically, and will generalize data based on what is learned (i.e. inductive learning). Generalization is a critical component is determining how well Machine Learning performs, and includes:

- Finding natural groupings within a data set
- Classifying categories in which data points belong
- Forecasting and Regression analysis
- Determining the most important variables that affect outcomes

Machine Learning is not a new concept, however cloud computing is making it more accessible and affordable than ever before. Compute power and storage is plentiful in the cloud, and it includes the ability to scale capacity in minutes to handle large data jobs. You no longer need to maintain your own processing and storage infrastructure.

Inductive learning lies at the heart of Machine Learning

Inductive learning lets the data tell its own story

There is no need to have prior knowledge of the data. There is no need to make assumptions in the data set, or work with small samples. This method of unsupervised learning removes the potential for bias to be inserted into the analysis, and makes identification of unknown issues to surface. Inductive learning lets the data tell its own story. This approach is an even greater asset when there is a large amount of data, and there is little advance knowledge of the story the data tells, or the questions to ask.

One of the most difficult and untapped problems in analytics has been unstructured textual data. This problem is an excellent use case for Machine Learning and statistical Natural Language Processing (NLP). Eliminating taxonomical and ontological analytics methods simplifies the process, improves the time to glean insights from data, reduces the operational cost of analytics, and removes inherent biases in a predetermined interpretation framework.

The best results are achieved when the Human Factor is added to Machine Learning

Automation is susceptible to a loss of insights from an experienced human analyst. Harmonizing human decisions with machine learning is key to unlocking insights through analytics. An Augmented Intelligence approach can leverage both human and machine learning, and analytics efforts are best served when harmonizing the two. The famous chess battles with IBM's Deep Blue chess playing machine are examples.

Chess master Gary Kasparov designed a tournament to determine which grouping would garner superior results – solely humans, solely computers, or a combination of humans and computers. Teams comprised of humans leveraging computer assistance consistently defeated all challengers including the strongest "computer only" and "human only" teams. A surprising finding, however, unfolded and was reported by Kasparov. "The winner was revealed to be not a grandmaster with a state-of-the-art computer, but a pair of amateur American chess players using three computers at the same time."
Combining the speed and processing power of a computer with a human's interpretive capabilities leads to superior results. This is Augmented Intelligence in its basic form - computational algorithms helping to put the human end user in the best position to make an informed decision. Algorithms read and analyze data to show the "who", "what", "when" and "where". A human then uses this intelligence to determine the "why" and decides on the appropriate course of action.

Algorithms read and analyze data to show the "who", "what", "when" and "where"

Why Data Visualization

Let the data tell the story through visualization

People generally don't want to consume content in the form of long spreadsheets of numbers or text. Insights are simplified when the results are easy to understand, leaping off the page in the form of graphics and charts, color coded, and easy to digest. Additional layers of content can be made available by clicking, or hovering over specific visuals to see additional information. Comparison of data sets side by side, comparisons over time periods, and comparison ranking simplify the story. The key is to help business teams quickly spot key trends, patterns, and themes so they can better anticipate and refine strategies and best practices in real-time.

An interactive UI allows virtually anyone to interact with data

An interactive UI allows virtually anyone to interact with data, empowering many in an enterprise to consume analytics. Interactive visualizations assist with one of the most fundamental concepts in investigative analysis, the four W's (who, what, when, where). Visualizations depict each of the four W's, while the human analysts build their own conclusions on the "why". Most importantly, by augmenting intelligence through deep computing processes, the "why" can be turned into specific action.

Collaborative analytics

Usability is further simplified and enhanced when visualizations are customized, annotated, and shared with colleagues. Collaboration on analytics benefits users in many ways:

- Speeds the time from data to revenue, using social interaction to speed decision-making.
- Opens communication, identification of insights, and creation of action plans helps align business strategies. This keeps everyone on the same page.
- Enables companies to extend the analysis to a greater portion of the enterprise so timely and useful insights can be added by other team members.
- Increases information visibility across the company.
- Improves organizational cohesion.

Simply put, business teams can make better decisions by working collaboratively, rather than working independently and linearly.

Signals' analytics engine dynamically creates semantic associations across all textual elements

How Signals™ Helps

Deep Textual Learning

Patented Signals™ analysis technology reads, assembles and processes word grams of varying length within every textual element to "learn" statistically relevant topics within the text. Built on a mixture of Recurrent Neural Networks (RNN), Semantic Parsing, Multi-Model Data Connection, Graph Models, and Long Short-Term Memory architecture (LSTM), Signals' analytics engine dynamically creates semantic associations across all textual elements, grouping related elements together into categories, top trends, and geo clusters.

Autonomously learning the content of any textual data provides a fast, unbiased method for processing customer comments, product reviews, chat sessions, voice to transcript interactions, survey data and other unstructured content.

Smart Structure Data Discovery

Signals™ solves the challenge of textual analytics and enables in-line structured data exploration, automated predictive results, and effortless collaboration within your analytics context. Specifically, the Signals™ intelligence core engine will automatically categorize your structured data based on known Numerical, Categorical, Ordinal, Temporal, User and Geographical data types.

With automatic spatial-temporal algorithms, such as KDE, Graph-Model, Convolutional Neural Network (CNN) and Page-Rank weighting, Signals™ performs multiple complex processing elements to identify key elements within any data set. From ad hoc jobs to enterprise data sets, the ability to select and dimension data, including sentiment, geography, category and topics associated with unstructured textual data across these structured dimensions for discrete analysis is fast and easy.

Advanced Natural Language Processing

Natural Language Processing is a fundamental processing core within Signals™. Supporting and performing NLP natively in 24 distinct languages, Signals™ eliminates core language translation and human tagging, automatically detecting originating language identifiers without user interaction.

Signals™ automatically performs complex parts of speech parsing, tagging, named entity extraction, lemmatization (cleaning), tokenization, knowledge-graph association, emoji identification, and other language nuances to navigate through the ambiguity and imprecision of written and spoken language. The end result is the ability to discern relevant content, emotion and even motion within unstructured textual elements.

Multi-Lingual Sentiment Analysis

Signals™ uses proprietary technology, including Hidden Markov Model and Deep Learning (RNN), opinion analysis, and statistical analysis to assess sentiment across all 24 supported languages. Sentiment is calculated on multiple levels, leveraging Part-Of-Speech identification to build dynamic sentiment scores.

Signals™ provides an intuitive visual interface within the UI to set or tune sentiments specific to product, industry or specialty.

Dynamic Topic Summarization

Signals™ identifies clusters of semantically similar textual elements and groups them into unique topic groups for each processed data set. Centered around these opinion-critical, unique categorical n-grams, the engine further identifies specific words, phrases, and language constructs that are associated with positive and negative meanings in the topical categories. In addition to words and phrases, a contextual analysis around the data-driven category is utilized to improve the accuracy of the sentiment and properly reflect its semantic differences within each category.

Global Geographical Analysis

Signals™ provides street-level geo encoding with a 95%+ global accuracy rate. The robust Geo-Analytics processing engine automatically performs space-time-kernel density estimation to generate a relative Heat Index for different sentiment categories across geo dimensions.

Temporal Analysis

Temporal information accompanying structured and unstructured data is paramount in understanding quantitative events and their potential underlying relationships across disparate data sets. Signals™ utilizes Time-Series Predictive analysis, Deep Learning, and Event-Analysis to identify unexpected trends and patterns across structured and unstructured data.

Conclusion

Advanced data analytics tools have traditionally been costly, time-intensive, complicated, and often reserved for the large enterprise. They often require complex IT integration and are difficult to use. We believe in making data discovery simple and approachable for everyone - no data scientist required. We make advanced data analytics more accessible and approachable, while allowing our customers to discover actionable insights in minutes and hours, rather than weeks or months, and at a fraction of the cost. We help business teams quickly visualize key patterns and behaviors, so they can better anticipate and refine strategies and best practices in real-time.