

2018

# Best Practices for Writing Market Intelligence Reports in the Machine Learning Age



NORTHERN LIGHT WHITE PAPER  
C. DAVID SEUSS, CEO

## Traditional Search Processes

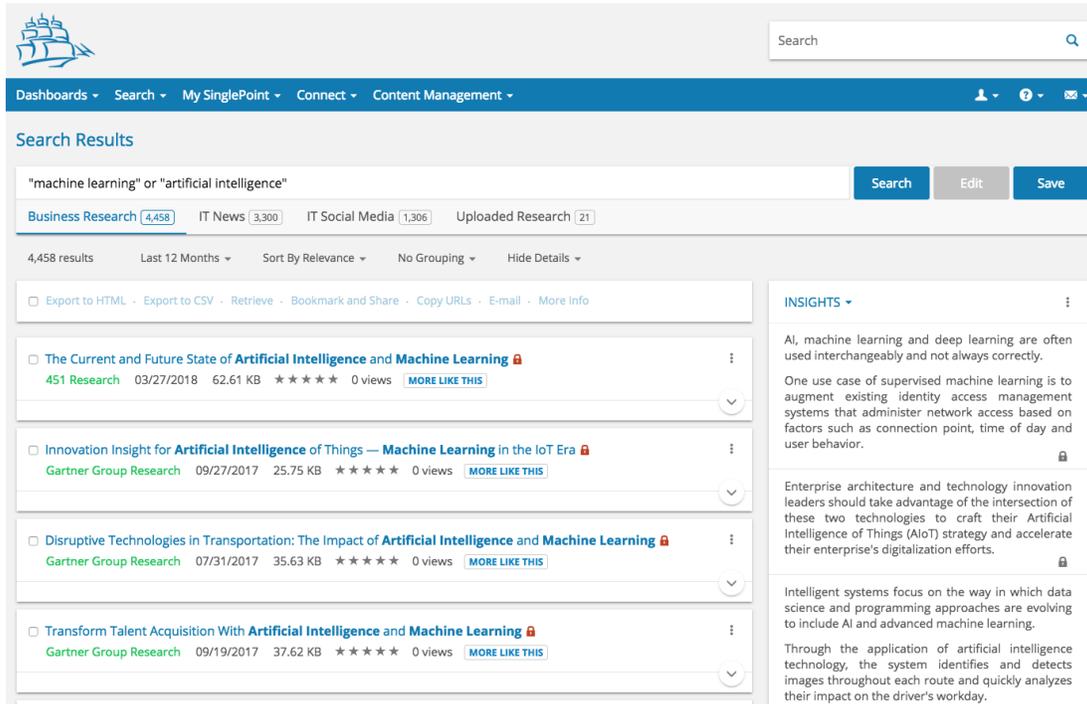
Across our universe of 250,000 market intelligence professionals, using traditional search processes (manually scanning up and down a search result page) users on average download and read only one document per search and therefore fail to acquire the insights from the documents they do not download.

In the machine learning age, users are encouraged to start by reading the Insights Report rather than use the traditional approach of manually examining the search results. By reading the Insights Report, users have access to insights from all the best documents on the search result instead of just one. Northern Light also expects that users will end up downloading and reading more documents when the truly important passages from all of the best documents are presented to them so the Insights Report will become an important vehicle for additional content consumption.

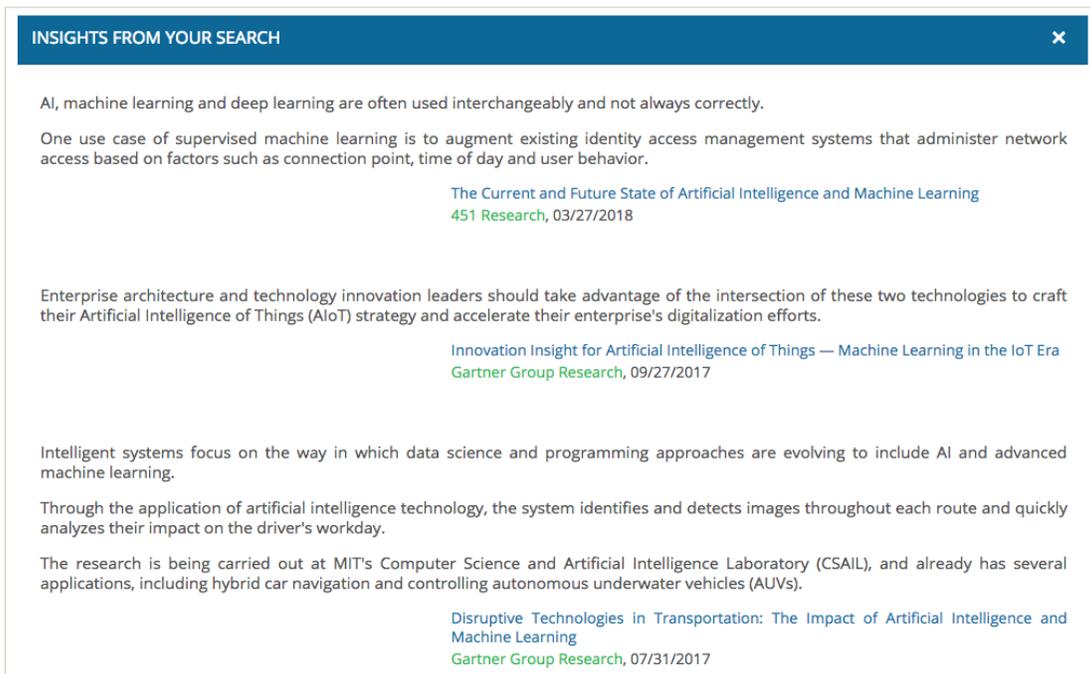
Northern Light's machine learning works as follows to automatically summarize the important ideas in the documents on a search result.

1. The process extracts all the "summary worthy sentences," which are usually declarative sentences making a statement ending with a period, from all the documents on the search results page.
2. The machine draws a graph of those sentences which measures how related each sentence is to all the others and how important each sentence's relationship partners are. The graph basically looks for the overlap of "features" between the sentences. Features are groups of words that like to hang out together in the group of documents being considered.
3. The graph determines that one sentence is most important. The machine picks it for the report.
4. The graph determines that another sentence is the second most important. The machine picks it for the report.
5. The process continues until the target number of sentences for the report is reached. For Northern Light, this is currently 40 sentences based on our testing of various report lengths.
6. The machine orders the report sentences as they appear in the documents first and also orders the document sentence grouping in the order the documents are on the search result

Below is a sample search result and then the automated Insights Report written by Northern Light machine learning algorithms on that search result.



The screenshot shows a search interface with a search bar at the top right containing the text "Search". Below the search bar is a navigation menu with items: Dashboards, Search, My SinglePoint, Connect, Content Management. The main section is titled "Search Results" and shows a search query: "machine learning" or "artificial intelligence". There are buttons for Search, Edit, and Save. Below the query, there are filters for Business Research (4,458), IT News (3,300), IT Social Media (1,306), and Uploaded Research (21). The results are sorted by relevance and show 4,458 results. A sidebar on the right is titled "INSIGHTS" and contains several paragraphs of text generated by machine learning algorithms, such as "AI, machine learning and deep learning are often used interchangeably and not always correctly." and "Enterprise architecture and technology innovation leaders should take advantage of the intersection of these two technologies to craft their Artificial Intelligence of Things (AIoT) strategy and accelerate their enterprise's digitalization efforts."



The screenshot shows an "INSIGHTS FROM YOUR SEARCH" report. It features a blue header with the title and a close button. The main content area contains several paragraphs of text, each followed by a link to a research report. The first paragraph is: "AI, machine learning and deep learning are often used interchangeably and not always correctly." followed by a link to "The Current and Future State of Artificial Intelligence and Machine Learning" by 451 Research, dated 03/27/2018. The second paragraph is: "Enterprise architecture and technology innovation leaders should take advantage of the intersection of these two technologies to craft their Artificial Intelligence of Things (AIoT) strategy and accelerate their enterprise's digitalization efforts." followed by a link to "Innovation Insight for Artificial Intelligence of Things — Machine Learning in the IoT Era" by Gartner Group Research, dated 09/27/2017. The third paragraph is: "Intelligent systems focus on the way in which data science and programming approaches are evolving to include AI and advanced machine learning." followed by a link to "Disruptive Technologies in Transportation: The Impact of Artificial Intelligence and Machine Learning" by Gartner Group Research, dated 07/31/2017. The fourth paragraph is: "Through the application of artificial intelligence technology, the system identifies and detects images throughout each route and quickly analyzes their impact on the driver's workday." followed by a link to "Disruptive Technologies in Transportation: The Impact of Artificial Intelligence and Machine Learning" by Gartner Group Research, dated 07/31/2017. The fifth paragraph is: "The research is being carried out at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), and already has several applications, including hybrid car navigation and controlling autonomous underwater vehicles (AUVs)." followed by a link to "Disruptive Technologies in Transportation: The Impact of Artificial Intelligence and Machine Learning" by Gartner Group Research, dated 07/31/2017.

There is competition between sentences for a place on the Insights Report based on how important the graph determines that they are.



## Writing for the Machine

Here are Northern Light suggestions for writing market intelligence reports in the Machine Learning age.

- ❖ Write in simple declarative sentences that have noun subjects and predicate verbs that express whole ideas within each sentence. Declarative sentences always end with a period as the terminator, never a question mark or exclamation mark.

**Don't say:** The study showed how fast [CompanyName] [ProductName] is. The result: blazing speed!

- ✓ The machine uses sentences as a fundamental unit of idea organization. It will not automatically associate a word in one sentence with an object in another sentence, even in an adjacent sentence. There is no overlap in the words in the two sentences above and the machine will not relate them directly to one another in the graphs. The second sentence is not a complete sentence in that it has no verb as a predicate and it will be ignored by our application as a fragment.

**Do say:** The study showed that [CompanyName] [ProductName] is really fast and has blazing speed.

- ❖ Questions are to be avoided whenever possible since usually the answer follows the question in another sentence or in tables or lists. We don't usually care about the questions, it is the answers we want. Put the answers into context with a simple declarative sentence ending in a period.

**Don't say:** How often do users update their antivirus program? 23% said daily, 67% said weekly, and 10% said "been meaning to do that but never get around to it."

- ✓ The machine won't be able to associate the answers to the questions because there is no overlap of word (the graph won't connect them) and because the answers have such generic words and figures the machine will consider the answers as less important.

**Do say:** Most users update their antivirus programs weekly (67%), some daily (23%), and some never at all (10%).



- ❖ Avoid putting the important ideas in lists, or at least only in lists.

**Don't say:** The key factors in enterprise cloud vendor selection are:

1. Implementation assistance
2. Near instant response to increased server needs
3. Uptime history
4. Cost

- ✓ The machine will not be able to associate the list to the key sentence above the list that provides context because the words do not overlap.

**Do say:** The key factors in enterprise cloud vendor selection are implementation assistance, near instant response to increased server needs, uptime history, and cost.

- ❖ If you need a list or just want one for readability issues, repeat the findings in a sentence in notes, a sidebar, appendix, etc. using the principles of simple declarative sentences ending with periods.

- ❖ Avoid pronouns (he, she, they, it) and pronoun-like words (company, firm) whenever possible.

**Don't say:** We interviewed key executives in the field such as Frank Smith, SVP of Cloud Computing. He said cloud computing rocks.

- ✓ The machine may not be able to find the antecedent of "He" in all cases and if the wrong antecedent is found and substituted, you have just generated fake news. If we stick to the words as they are written for the automated summary you will get: "He said cloud computing rocks," leaving the reader of the automated summary to wonder who "He" is.

**Do say:** We conducted interviews of executives in the field and Frank Smith SVP of Cloud Computing said cloud computing rocks.

**Don't say:** [CompanyName] had a good quarter market share wise. The company said cloud computing is winning one big new client after another.

- ✓ While not formally a pronoun, company has pronoun-like qualities in that we don't know who the company is if we only have the second sentence above to work with.



**Do say:** [CompanyName] increased its market share during the quarter as cloud computing won one big account after another.

❖ **Restrain the urge to write with creative expression.**

**Don't say:** The study shows that compared to XYZ's servers, [CompanyName] [ProductName] are quicker than a greased pig.

- ✓ Other writers won't use the greased pig expression and the sentence will be considered less important by the machine.

**Do say:** The study shows [CompanyName] [ProductName] were faster than XYZ's servers by a large margin.

❖ **Use standard vocabulary words that other writers are likely to use.**

**Don't say:** Geospatial information systems are an essential element of driverless car technology.

- ✓ Other writers might call it GPS and if so the above sentence will be considered less important.

**Do say:** GPS is an essential element of driverless car technology.

❖ **Don't write with social media conventions like hashtags, especially hash tags as subjects or objects in noun phrases.**

**Don't say:** “#IBM bring the technology and #Accenture bring the industry expertise” - sums up why #partnering is so important i... <https://t.co/fCGDJrXXM9>

- ✓ By the way, the above is a recent tweet by an IDC analyst. In our implementation, which is not tuned for social media, we look for actual nouns and proper nouns as



the subjects of sentences and we would conclude this sentence does not have a real subject.

**Do say:** IBM bringing the technology and Accenture bringing the industry expertise sums up why partnering is so important.

- ❖ If you use bullet points make them complete sentences with noun subjects a period as the terminator.
  - ✓ As previously mentioned, the machine looks for sentences as a fundamental unit of idea organization. It won't interpret incomplete phrases that are bullet points as meaningful.
  
- ❖ When using PowerPoint, put any essential oral commentary and the interpretation of graphical elements into the notes field on each slide using the above principles of simple declarative sentences ending with periods. Or append slides that use text sentences that recap the findings in simple declarative sentences ending in periods.
  
- ❖ If you put a conclusion or comment in a text box on a slide with a graphical element, use a complete sentence ending with a period in the text box.