Deception And Truth Analysis (D.A.T.A.), Inc.

Data Dictionary

20 November 2023

Below is the Deception And Truth Analysis Data Dictionary. Our API outputs are listed first and in parentheses are the names of these same features on our platform available.

**Company (Company Name)**

This is the formal company name as used by the Securities and Exchange Commission to refer to the entity.

**CIKCode (CIK)**

This is the Securities and Exchange Commission assigned Central Index Key (CIK) that is unique to each filer.

**Ticker (Symbol)**

This is the abbreviation used to uniquely identify publicly traded companies on a particular exchange. Some companies have multiple share classes, but because Deception And Truth Analysis assesses the filings of companies we do not track multiple share classes. Instead, for companies that have multiple share classes we use the main ticker symbol without the designation unique for each share class. For example, both Alphabet and Berkshire Hathaway have multiple share classes, but for Alphabet we use GOOG, and for Berkshire Hathaway we use BRK.

**DocumentType (Type)**

DATAbase features multiple different document types by companies. These include:

* 10-K: Annual 10(k) reports
* 10-Q: Quarterly 10(q) reports
* 8-K: Unscheduled material event 8(k) reports
* MISC: Transcripts, including earnings calls, corporate transaction calls, mergers and acquisitions calls, analyst days, industry presentations, and so on

**FilingDate (Filing Date)**

This is the date in which a filing was first publicly available. The date is in YYYYMMDD format.

**FilingDateUnixTimestamp**

This is the time on the FilingDate that a document was first publicly available.

**ScorePublishedDate**

This is the date on which a DATA Score assessment of a document was first available on the DATAbase platform.

**ScorePublishedDateUnixTimestamp**

This is the time on the ScorePublishedDate for which a DATA Score assessment of a document was first available on the DATAbase platform.

**FiscalPeriod (Fiscal Period)**

If a company’s filing period is on a fiscal period rather than a calendar period this is the last day of its annual reporting period.

**DataScore (D.A.T.A. Score)**

The D.A.T.A. Score ranges from -100% to +100% where negative values suggest the degree of deceptiveness and positive values suggest the degree of truthfulness. Specifically, a highly negative D.A.T.A. Score means that there is a high degree of deceptive language present. Likewise, a high positive D.A.T.A. Score means that there is a high degree of truthful language present.

The D.A.T.A. Score is a scientific evaluation that measures the degree of similarity of a document’s psycholinguistic features to known psycholinguistic features used by deceivers when communicating.

**(DATA Score Y/Y Change)**

This refers to the change in percentage points of a DATA Score for a current document relative to the DATA Score for the same document type a year ago. Note: this score is only available for document types that are released on a predictable schedule, such as 10(k)s, 10(q)s, and earnings call transcripts.

**FragmentsCount**

FragmentsCount is the total number of scorable fragments within a document that contain narrative.

DATA breaks documents down into scorable fragments to provide within document DATA Score assessments. The minimum size of a fragment is 800 words and the size of this fragment is determined by the minimum size necessary to provide a statistically significant assessment that matches our double-blind, scientifically tested accuracy rate of 88.4%.

However, to ensure DATA assessments that make sense, our algorithm will automatically grab text at the beginnings and ends of fragments to ensure that sentences and paragraphs are not cut off in the middle. Additionally, DATA uses a proprietary machine learning algorithm to ensure that the algorithm only scores narratives by excising things such as title pages, tables of content, numerical data, proper nouns, page numbers, URLs, and so on.

**DeceptiveFragmentsCount (# Deceptive Fragments)**

This value shows the number of fragments in a document that have a deceptive D.A.T.A. Score (i.e. a value of -100% to 0%). DATA**RED**line is our premium product which allows you to view each fragment of a document and its level of deceptiveness and truthfulness and how it fluctuates throughout; as well as the underlying language associated with the D.A.T.A. Score.

**DeceptiveFragmentsPercentage (% Deceptive Fragments)**

This is the % of deceptive fragments in a given document. For example, if the number of deceptive fragments (“# Deceptive Fragments”) is 3 and the document has 12 fragments, then the “% Deceptive Fragments” would be: 3 ÷ 12 = 25%.

**DeceptiveFragmentsAverageScore (Average DATA Score Deceptive)**

This shows the average DATA Score for the document fragments that have been assessed as deceptive.

**TruthfulFragmentsCount (# Truthful Fragments)**

This value shows the number of fragments in a document that have a truthful D.A.T.A. Score (i.e. a value of 0% to +100%). DATA**RED**line is our premium product which allows you to view each fragment of a document and its level of deceptiveness and truthfulness and how it fluctuates throughout; as well as the underlying language associated with the D.A.T.A. Score.

**TruthfulFragmentsPercentage (% Truthful Fragments)**

This is the % of truthful fragments in a given document. For example, if the number of truthful fragments (“# Truthful Fragments”) is 88 and the document has 99 fragments, then the “% Truthful Fragments” would be: 88 ÷ 99 = 89%.

**TruthfulFragmentsAverageScore (Average DATA Score Deceptive)**

This shows the average DATA Score for the document fragments that have been assessed as truthful.

**ScorableWordCount**

This is the total number of scorable words in a document net of the removal by DATA of non-narrative content such as title pages, tables of content, numerical data, proper nouns, page numbers, URLs, and so on.

**PrimaryIndex**

This is the primary stock index for a company. DATA provides stock index coverage for the S&P 500, 400, and 600, as well as the Russell 2000, 1000, and 3000. If a company is a member of one of the S&P indices then its S&P index is considered its primary index, followed by Russell indices.

**DatabaseRank**

This is the ranking of a company’s document at the time of assessment relative to all other documents within DATAbase for that same time period.

**LongTermFocus (Long term focus)**

D.A.T.A. scans all documents for words identified in research done by Loughran and McDonald (2011) as indicative of a company’s ‘long term focus.’ Their work was specifically designed to correct for previous lists of words that had dual meanings. This category of words is one of four meant to provide insight into the tone of a document. The others are short term focus, internal focus, and external focus.

**ShortTermFocus (Short term focus)**

D.A.T.A. scans all documents for words identified in research done by Loughran and McDonald (2011) as indicative of a company’s ‘short term focus.’ Their work was specifically designed to correct for previous lists of words that had dual meanings. This category of words is one of four meant to provide insight into the tone of a document. The others are long term focus, internal focus, and external focus.

**InternalFocus (Internal focus)**

D.A.T.A. scans all documents for words identified in research done by Loughran and McDonald (2011) as indicative of a company’s ‘internal focus.’ Their work was specifically designed to correct for previous lists of words that had dual meanings. This category of words is one of four meant to provide insight into the tone of a document. The others are long term focus, short term focus, and external focus.

**ExternalFocus (External focus)**

D.A.T.A. scans all documents for words identified in research done by Loughran and McDonald (2011) as indicative of a company’s ‘external focus.’ Their work was specifically designed to correct for previous lists of words that had dual meanings. This category of words is one of four meant to provide insight into the tone of a document. The others are long term focus, short term focus, and internal focus.

**FinancialUncertainty (Financial uncertainty)**

D.A.T.A. scans all documents for the Red Flag words identified in research done by Loughran and McDonald (2011) as indicative of a company’s ‘financial uncertainty.’ Their work was specifically designed to correct for previous lists of Red Flag words that had dual meanings. They found that a higher proportion of ‘financial uncertainty’ words was associated with lower excess returns.

**Index Rank**

Shows the company’s D.A.T.A. Score rank within its primary equity index for the previous one-year period. Rank is from most truthful to most deceptive. In the S&P 500, for example, a rank of 13 means the company is the thirteenth most truthful company in the index.

The primary equity index is listed at the top of each Company Page. If a company is in more than one equity index (e.g. DJIA and S&P 500) then its rank within the primary index is shown. Primary index is determined by this default sequence:

1. DJIA
2. S&P indexes included in (i.e. S&P 500, 600, & 400)
3. Russell indexes included in (i.e. Russell 2000, 1000)

Index components are updated at the end of each quarter, not in real time. So if a company has recently been added to, or subtracted from an index there will be a lag in our updating of the index.

**Industry Rank**

Shows the company’s D.A.T.A. Score rank within its industry for the previous one-year fiscal period. Rank is from most truthful to most deceptive. In ‘Pharmaceutical Retailers,’ for example, a rank of 8 means the company is the eighth most truthful company in the industry. D.A.T.A.’s industry categories come from Yahoo! Finance.

**From Average**

Shows the D.A.T.A. Score of a company’s most recently assessed document relative to the entirety of the documents assessed within DATAbase over the preceding rolling one-year period. For example, a figure of -11.09% means that the company’s most recently assessed document is 11.09% below the previous year’s average D.A.T.A. Score for all DATAbase documents. In other words, the company’s assessed truthfulness is trending downward relative to all other companies D.A.T.A. tracks.

**From Index Average**

Shows the D.A.T.A. Score of a company’s most recently assessed document relative to the index’s preceding rolling one-year’s average D.A.T.A. Score for all documents that have been assessed in the index. For example, a figure of +18.98% means that the company’s most recently assessed document is 18.98% above the previous one-year’s average D.A.T.A. Score for all assessed documents in the index. In other words, the company’s assessed truthfulness is trending upward.

**From Industry Average**

Shows the D.A.T.A. Score of a company’s most recently assessed document relative to the industry’s previous rolling one-year’s average D.A.T.A. Score for all documents that have been assessed in the industry. For example, a figure of +25.67% means that the company’s most recently assessed document is 25.67% above the previous one-year’s average D.A.T.A. Score for all assessed documents in the industry. In other words, the company’s assessed truthfulness is trending upward.