Corpus / Textual Input

- Corporate Filings (10-K, 10-Q, etc.)
- Analyst / Brokerage Reports
- News / Message Boards
- Social Media
- Meeting / Call Minutes
- (IPO) Prospectuses
- etc.

From →

H o w ↓

Parsing Method

- Labelling Sentences / Paragraphs
- Counting (Key-) Words
- Fog Index
- Term Counts & Term-Frequency-Inverse-Document-Frequency
- Machine Learning / Classifiers: Naïve Bayes, Support Vector Machine / Regression, k nearest neighbors, etc.

Derived Output

- Readability Measures
- Sentiment Scores
- Topic Categories / Scores
- Document Similarities
- Word Lists / Vectors
- Rankings
- (Document) Labels
- etc.

For →

H o w ↓

Model / Fusion Choice

Combining quantitative and qualitative information

- Regressions (linear models, Tobit / Probit / Logit)
- Support Vector Regression
- Time-Series Models
- etc.

Economic Variable

- Stock Returns / CAR
- (Abnormal) Volatility
- (Abnormal) Trading Volume
- Earnings (per share)
- Profitability
- Financial Distress & Bankruptcy
- IPO pricing
- etc.