Tapping Unstructured Data for Better Predictions and Decisions
TEXT ANALYTICS is a fast-growing technology that delivers many promises, but almost no application exists that unleashes the predictive power buried in unstructured data to directly improve high-volume predictions and decisions.

FICO has developed a scorecard-based approach that transparently incorporates textual analysis, making it easily usable in automated decision systems. If the promise before was of better decisions through data, we can now make better decisions through all kinds of data.
What Do We Mean By Text Analytics?

These are all interrelated terms that are widely used in a loose way. However, they are not the same.

Text analytics is a technology with many overlapping disciplines. The common characteristic is the need to transform text into a structured numerical format, so that traditional analytical algorithms can be applied.

Text can mean many things: claims, writing on forms, speech, web pages, etc. The value of analyzing this content is enormous.
### How Soon Will Text Analytics Be Widely Adopted?

<table>
<thead>
<tr>
<th>Benefit Level</th>
<th>Less Than 2 Years</th>
<th>2 to 5 Years</th>
<th>5 to 10 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Benefit</td>
<td>Column-Store DBMS</td>
<td>Cloud Computing</td>
<td>Complex Event Processing</td>
</tr>
<tr>
<td></td>
<td>In-Memory Database Management Systems</td>
<td>Content Analytics</td>
<td>Hybrid Cloud Computing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Telematics</td>
</tr>
<tr>
<td>High Benefit</td>
<td>Predictive Analytics</td>
<td>Advanced Fraud Detection</td>
<td>Cloud Parallel Processing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text Analytics</td>
<td>Social Network Analysis</td>
</tr>
<tr>
<td></td>
<td>Predictive Modelling Solutions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social Analytics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cloud-Based Grid Computing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Benefit</td>
<td>Social Media Monitors</td>
<td>Claims Analytics</td>
<td>Cloud Collaboration Services</td>
</tr>
<tr>
<td></td>
<td>Web Analytics</td>
<td></td>
<td>Geographic Visualization and Analytics Systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech Recognition Systems</td>
<td>Video Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MapReduce Alternatives</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Web Experience Analytics</td>
<td></td>
</tr>
</tbody>
</table>

Source: Forbes Estimates to Midstream Adoption, Gartner July 2012
Better decisions through unstructured data mining represents the biggest problem and opportunity in Big Data, yet is poorly understood by many consumers of analytic solutions.

80% of Big Data is unstructured in nature.

Traditional solutions enhanced by textual mining can have a significant bottom-line impact in financial organizations by improving existing business intelligence approaches.

FICO’s approach to text analytics involves:

1. Discovering patterns and hidden value in text-based (unstructured) information by mining data sources using supervised, semi-supervised and unsupervised techniques

2. Consolidating both structured (quantitative) data and unstructured information to gain a more accurate description

3. Producing better predictive models to spot opportunities and recognize trends
How FICO’s Text Analytics Engine Works

Data Analysis and Indexing
- Process documents in multiple formats
- Choose between different text analyzers
- Customizable interface and filtering options
- Create and update indexes both in memory and disk

Analysis and Transformation
- Search through data fast
- Discover hidden information and patterns
- Provide a customizable and expandable analytics on top of search
- Incorporate analytics complexity in an incremental manner

Interface with FICO Tools and Products
- Provides a “ready-to-consume” format for easy integration
- Provides inputs to FICO solutions for specific vertical industries and challenges

Streaming Capabilities and Big Data Integration
- Data-agnostic
- Provide streaming capabilities
- Hadoop compatibility
- Machine Learning Interface (Apache Mahout)
How FICO’s Text Analytics Engine Works

Original Data
- Provider num: NN1
- Date of call: 12-31-13
- Reason code: C5
- Member ID: 12345
- Comment: Paid clm
  2234551 dos 2-21-13. Reports miscalculation. Upset...Request please recoup

Structured Data

Unstructured Data

Linguistic Analysis
- Dependency Analysis
- Named entity extraction
- Intention analysis

Category Dictionary

Synonym Dictionary

Patterns

Index

Metadata
- Category: Miscalculation
- Item: Upset
- Request: Recoup

Mining Engine
- Timeline analysis
- Topic extraction
- 2-D analysis
- Trend analysis
- Distribution analysis

Alerting Engine

Visualization & Mining
How FICO’s Text Analytics Engine Improves Your Analyses

Data Agnostic. Analyzes and indexes structured, semi-structured and unstructured data under a wide diversity of formats. Works with index representations both in memory and disk.

Efficient, Fast and Customizable Linguistic Processing. The engine dynamically configures pre-defined and customized text analyzers. It integrates five different tokenizers and nine different filters for iterative model design. It also lets you operate with your own sets of stop words, synonyms and customized expressions.

Multi-Mode. Works in both supervised (analyst directed) and unsupervised mode. Lets you describe a target variable based on simple rules.

Metadata. Extracts metadata info if available, and makes it ready for future use.

Multi-Language. Ready for expansion to dozens of languages besides English.

Predictive Modeling Output. Extracts the most predictive terms and association of terms, which can then be used in the FICO Semantic Scorecard.
Now that we’ve extracted insights from the unstructured text—and can represent these insights mathematically—how do we use them to make better predictions of behavior, and ultimately better decisions?
Introducing FICO Semantic Scorecards

FICO combined our text analytics engine with our FICO® Model Builder tool for building predictive models. The result was Model Builder Semantic Scorecards, which leverage the predictive power of unstructured data, adding new predictive features to traditional models, increasing their predictive power.

These scorecards are integrated into the FICO Model Builder tool, giving users the ability to:

- Process unstructured data from various file formats and different sources
- Perform queries, filtering, linguistic pre-processing in a highly efficient way
- Generate and maintain index representations of documents for downstream analysis
- Obtain new features from the unstructured data and combine them with regular structured fields
- Obtain a very fast and scalable solution
Why FICO Model Builder Semantic Scorecards Work So Well

**They’re Transparent.** Semantic scorecards enhance the traditional scorecard technology by adding textual analysis in a clear, natural and logical way.

**They’re Engineerable.** If you don’t like a discovered feature, you can leave it out of your model.

**They’re Improving.** We are rapidly improving the capabilities of our text predictive technology to make it a core component in our decision management technologies.

**They’re Really Fast!** These scorecards can consume tens of thousands of records and produce better predictive models in a matter of seconds.
FICO Semantic Scorecards are the natural extension of traditional scorecards for the unstructured world.

They make possible the integration of structured and unstructured data to produce better predictions that drive better decisions.
Using LDA (Latent Dirichlet Allocation) to Generate Archetypes From Text

FICO analyzes a bank’s call center text to find patterns within the data. The goal is to identify meaningful reasons why customers are calling, and use these insights to better understand and predict attrition risk.

Among the useful discoveries so far: Customers mapping strongly to a frequent traveler archetype are among the least likely to attrite. Moreover, if these customers interact with a call center representative regarding a payment missed due to traveling—resulting in waiver of the late fee—attrition becomes even less likely. This insight could help focus the budget for late fee waivers where it will have a strong impact on improving attrition rates.

**Call Center Text**

_Cy notice. I missed only one payment while I was out of town and now I've been hit by a late fee that I don't feel I should have to pay. I want to make sure my account is not be reported to the credit bureaus as delinquent. I've been a customer for more than five years and I usually pay my bills on time. Why can't the bank know me well enough to understand that I have to travel frequently for my work and therefore can't always meet their imposed deadlines. If you want to keep my business, you need to show more flexibility. Every week I receive plenty of calls. If I receive one more call from your collections department about an unpaid bill I will take them up on their offer._

**“Bag of Words”**

_cancellation, pmt, reason, international due, status, denied, charged, account,Traveling, added, reversal, annual, status, flagging, bad, card, offer, cancelled, monitoring, credit, due, pin, account, verification, charged, cancel, reversal, applied, account, bad, denier, billing, offer, emv, cardholder, denied, flags, statements, offer, account, failed, due, late, int, due, pmt, auth, failed, fraud, annual, advised, auto cardholder, added, reversal, flagging, due, reversed, status, delinquent, int, due, late, pmt, auth, failed, fraud, annual, advised, auto cardholder, added, reversal, flagging, reversed, status, delinquent_

**Generation of call center content archetypes—determining customer topics**

26.2% 18.5% 2.4% 48.7% 4.2%
Learn More

For more information:

North America toll-free
+1 888 342 6336

International
+44 (0) 207 940 8718

email
info@fico.com

web
www.fico.com

FICO, Debt Manager and “Make every decision count” are trademarks or registered trademarks of Fair Isaac Corporation in the United States and in other countries. Other product and company names herein may be trademarks of their respective owners.