RavenPack at a Glance

RavenPack is the leading big data analytics provider for financial services

- Founded: 2003
- Employees: 54
- Location: New York, Marbella
- Languages: Common Lisp, Python, Java

www.ravenpack.com
Our Vision:
Make unstructured content actionable for financial professionals
Outline

- Data for decision-making
- Approaching unstructured content
- Use Cases: Decision-making in Finance
- Key Challenges
Why are we here?

(Not in the existential sense)

Because we believe the web should be codified
Why do we believe the web should be codified?

Because machines need to be able to understand the web.
Why do machines need to understand the web?

Because there is too much content for humans to read.
Why is there so much content to read?

Because we’re constantly generating relevant content.
Why do we want all this content?
Because we need to make better decisions
The Role of Data Analytics in Decision Making

Which of the following inputs did you place the most reliance on for your last big decision?

- 4% Other (e.g. consultants)
- 9% Financial indicators
- 30% Own intuition or experience
- 29% Data & analysis (internal or external)
- 28% Advice or experience of others internally

Only 29% use data

Source: PwC’s Global Data & Analytics Survey 2014, conducted by the Economist Intelligence Unit
The Role of Data Analytics in Decision Making

Key Obstacles:

- Lack of skill/expertise
- Quality/completeness
- Timeliness of data

Source: PwC's Global Data & Analytics Survey 2014, conducted by the Economist Intelligence Unit
Key Decision-making Goals

**ESTIMATE**
Perform timely and better estimates of a company’s revenues and earnings

**MEASURE**
Produce real-time measurements of business or economic indicators like consumer confidence or sentiment

**CONNECT**
Understand contagion effects of information across entire networks of tradable securities (e.g. supply-chain or competitive landscape)

**EXPLAIN**
Come up with new research insights and ways of looking at the market

**PREDICT**
Make more accurate predictions of future asset price moves; forecast demand and supply of resources

**INNOVATE**
Develop differentiated portfolios, improve scalability, and avoid crowded trades
Data for decision-making
Consumer Transactions
Geo-Location
Business Insights
Pricing
Reviews & Ratings
Web Crawled Data
Data Aggregators
App Usage & Web Traffic
Consumer Insights
Public Sector
Online Search
Store Locations
Social Media
Sentiment
Employment
Advertising
Trade
Open Data
Satellite & Weather
B2B Datasets

Source: Eagle Alpha

Possible lack of broader applications holding investors back.
Data Growth in Organizations

Data Growth
Driven by Unstructured Data

125 Exabytes

79.2% Unstructured Data

20.8% Structured Data

Growing 42.5% Per Year

Growing 22.4% Per Year

2013

2017

Source: IDC - 2014, Structured Data vs. Unstructured Data: The Balance of Power Continues to Shift
Surface Web
- Google, Bing, Wikipedia
- CNN, BBC
- Twitter

Publicly accessible

Deep Web
- Internal content, Email, IM
- Government, academic, medical records
- Bank transactions

Private content

Data from the Web
www.ravenpack.com
Big data in organizations

Data in Motion
News, social media, IM, emails, client demand, stock market data

Data at Rest
Purchases, performance & sales, third party data, historical data
Market Size: Data Analytics in Finance

Big Data
$114 Billion
CAGR 30%

Cloud Computing
$129 Billion
CAGR 23%

NLP Technology
$17 Billion
CAGR 16%

Data Analytics for Financial Services
Market Size: $3.1 Billion (CAGR 22%)

High Growth Technology Segments
Serving Financial Markets by 2018

- Banks: $1.6B
- Fund Managers: $1.1B
- Insurance Companies: $325M
- Broker Dealers: $80M
- Data Vendors: $40M
Approaching Unstructured Content

RavenPack
POPULAR FORMATS
RavenPack can process documents containing textual information

- **Emails**: Internal and external including attachments
- **Instant Messages**: IB, Slack, Skype, Symphony, other
- **Text Documents**: Investment notes, business contracts, filings, etc.
- **PDF Documents**: Research reports, newsletters, surveys, etc.

www.ravenpack.com
Adding Structure

**Codify**
Named entity recognition, topic categorization, and temporal classification

**Relevance Tagging**
Quantify importance, rank, and connections

**Novelty Detection**
Identify what is new, original, or unusual

**Sentiment Analysis**
Determine attitude, opinion, and impact
Entities and Taxonomies

Relationship Mapping
- Relationships between entities

Event Taxonomy
- Classification of events and roles
Companies
Over 43,000 companies covering all major stock indexes worldwide

Currencies
161 global currencies including bitcoin

Commodities
All major publicly traded commodities

Organizations
More than 3,000 government and non-governmental entities

www.ravenpack.com
Places
Every economically relevant town, city, region, and country, plus nationalities

Products
Over 1,200 major products and services typically in the media

People
Over 7,000 key business and political figures

Teams
Starting with major leagues such as the NFL, Premier League, and more

www.ravenpack.com
Event Taxonomy

- 5 Topics
- 56 Groups
- 495 Types
- 6895 Roles

Event categories include:
- Earnings
- Acquisitions-Mergers
- Analyst-Ratings
- Assets
- Bankruptcy
- Commodity-Prices
- Credit
- Credit-Ratings
- Dividends
- Equity-Actions
- Exploration
- Indexes
- Business
- Economy
- Environment
- Politics
- Insider-Trading
- Investor-Relations
- Executive-Appointments
- Executive-Compensation
- Executive-Deaths
- Executive-Filings
- Executive-Health
Curated Themes

Global Economic Outlook
Forward looking information, guidance, and forecasts on the world economy.

War & Conflict
Tracks armed, hostile conflict, terrorism, and hostility between different groups or factions.

Natural Disasters
Assess the impact of natural events such as earthquakes, floods, or hurricanes.

Global Risks
Draws attention to ways that global risks are evolving and interacting in our society.

Commodities
Everything there is to know about commodities including crude oil, gold, silver, wheat, and more.

Elections
A complete picture of the latest campaigns and elections taking place across the globe.

Interest Rate Guidance
Forward looking information, guidance, and forecasts on global interest rates.

Business Confidence
Measures the amount of optimism or pessimism that business managers feel, providing an overview of the economy.
Use Cases: Decision-making in Finance
Equities: corporate sentiment (short-term), economic sentiment (long-term)

Commodities: political/societal/environmental sentiment (short-term), economic sentiment (long-term)

Currencies: economic sentiment

Corporate Credit & Sovereign Debt: corporate & economic sentiment
Quantitative & Algorithmic Trading

Enhanced algorithmic trading execution

Discretionary Trading Applications

Front office decision making tools

Portfolio Selection

Find stocks in the media spotlight

Event based trading

Buy on the rumor, sell on the fact

www.ravenpack.com
Quantitative & Algorithmic Trading
Enhanced algorithmic trading execution

Discretionary Trading Applications
Front office decision making tools

Portfolio Selection
Find stocks in the media spotlight

Event based trading
Buy on the rumor, sell on the fact

www.ravenpack.com
Quantitative & Algorithmic Trading
Enhanced algorithmic trading execution

Discretionary Trading Applications
Front office decision making tools

Portfolio Selection
Find stocks in the media spotlight

Event based trading
Buy on the rumor, sell on the fact

www.ravenpack.com
Quantitative & Algorithmic Trading
Enhanced algorithmic trading execution
Discretionary Trading Applications
Front office decision making tools
Portfolio Selection
Find stocks in the media spotlight
Event based trading
Buy on the rumor, sell on the fact

www.ravenpack.com
Risk Management

Control for entity or event risk

Market Surveillance/Compliance

Investigate insider trading

Competitor Analysis

Stay on top of competitor activity

Supply-chain Analysis

Monitor supply chain
Risk Management
Control for entity or event risk

**Market Surveillance/Compliance**
Investigate insider trading

**Competitor Analysis**
Stay on top of competitor activity

**Supply-chain Analysis**
Monitor supply chain
Risk Management
Control for entity or event risk

Market Surveillance/Compliance
Investigate insider trading

Competitor Analysis
Stay on top of competitor activity

Supply-chain Analysis
Monitor supply chain
Risk Management
Control for entity or event risk

Market Surveillance/Compliance
Investigate insider trading

Competitor Analysis
Stay on top of competitor activity

Supply-chain Analysis
Monitor supply chain
Use Cases: Industry Research
The top decile of our Sentiment Strength factor yields a cumulative return of approximately 600% from December 2000-December 2014 and nearly 400% over the cumulative return of the Russell 2000 index. By focusing on the Active Returns results, we not only see the marked outperformance that can be gained by tilting a strategy toward the top ranked securities as measured by Sentiment Strength, but also a way to filter out underperformers based on the market’s perception. Deciles five through ten all underperform the Russell 2000 to varying extents; being able to monitor your portfolio for signs of poor market perception could inhibit both absolute and active returns in a material way.
In line with intuition, the charts below show that investors react more strongly to unscheduled than scheduled news. For positive news, the unscheduled news has a 3 times bigger impact than scheduled news.
A pairs example: Sport Direct Intl and Dixons Carphone

Pair selection date

Dec-10: worse-than-expected sales performance and questions over staff working conditions

Jan-8: profit warning

Source: Deutsche Bank, FactSet, Axioma, RavenPack
Key Challenges
Alternative data in Finance

Information is at the product or service level and not easily mapped to tradeable securities.

Data sets are so new that there is no or little research on their applications.

The more users, the more likely their basic value will be arbitrated away.

Coverage is limited, providing insights on only a few companies or a specific sector or industry.

Data contains text and requires NLP or images require special processing.

Limited history makes it difficult to determine relevance and perform backtests.

Data may be stored in different formats or contain gaps over time.

www.ravenpack.com
DATA

Content Quality

- Bad encoding
- No paragraphs
- Spelling mistakes
- Bad timestamps
- Bad metadata
- Gaps in history
- Real-time updates
- Source credibility
Point in Time Awareness

Entities change over time, companies are acquired or go out of business, government organizations change, people die.
Interpretation of events can change over time, e.g. Layoffs
Survivorship Bias

We need to ensure that our studies are not just based on entities that exist today.
Security and Privacy Concerns

Data can be sensitive and must be protected. Personal information must be redacted. Need for on-prem solutions and private cloud.
Final Thoughts:
Key Market Trends

Big Data Applications
Growing adoption of larger datasets and the technology to process them

Unstructured Data in the Enterprise
Companies harnessing the value in their internal content

Media Monitoring & Intelligence
Monitoring external data for branding and competitive intelligence

Significance of NLP
Extracting semantic intelligence with text analytics

Unification of Data
Linking of diverse datasets using knowledge representation

Analytics for Decision-making
Descriptive and predictive analytics for evaluating alternatives
Rise to the Challenge!

The RavenPack Team
@ Malaga FC, Spain
Questions?
Bringing Semantic Intelligence to Financial Markets

RavenPack