



SUMMIT
ONLINE
JAPAN

AWS Data Exchangeを活用した メディア企業のデータビジネス

株式会社日本経済新聞社

デジタル事業 情報サービスユニット 部長
高松 純子

Agenda

1. 日経のデータビジネスご紹介
2. **AWS Data Exchange** の活用で目指すもの
3. 提供コンテンツ紹介
 - ① 記事データ
 - ② 日経**POS**情報

1. 日経のデータビジネス

日経のデータビジネス

- 1967年 電子計算機部を新設、コンピューターで新聞製作開始
- 1970年 経済データ配信サービス「日経NEEDS」開始
- 1973年 市況情報センター「QUICK」設立
- 1975年 日経ダウ平均（日経平均株価）算出開始
- 1985年 日経POS情報 収集開始
- 1996年 インターネット上で「NikkeiNet」公開
- 1997年 インターネット上で「日経テレコン」サービス開始
- 2010年 「日経電子版」創刊



日経のデータビジネス

- 新聞・電子版 記事データ
- 日経NEEDS
 - 財務データ
 - 株価データ
 - マクロ経済統計
- 業界解説レポート
- 企業活動情報
- 日経POS情報

2018年より、テキストも数字も「人が読む」だけでなく「機械が読む」データとして、購入しやすい価格で販売

⇒ 2020年8月 **AWS Data Exchange** から提供開始

AWS Marketplace now supports Amazon EC2 Spot Instances



LEARN MORE

AWS Marketplace is a digital catalog with thousands of software listings from independent software vendors that make it easy to **find, test, buy, and deploy software** that runs on AWS.

View all products

Find AWS Marketplace products that meet your needs.

Categories

All categories ▾

Vendors

All vendors ▾

Pricing Plans

All pricing plans ▾

Delivery Methods

All delivery methods ▾

Total results: 8810

Clear selection

View results

2. AWS Data Exchange の活用

日経の目指す 法人向けデータ ビジネス

POSEYES

Value
Search

日経
テレコン

日経リスク&
コンプライ
アンス

NOWCAST
FSNなど

アスタ
ミューゼ
レポート

新しいデータ
販売チャンネル

日経API

POS

テレコン(記事)

NEEDS
(数値)

アスタミューゼ
(新技術・
業界動向)

Kepple
(企業情報)

その他

AWS Data Exchangeで提供するメリット

(これまでの課題)

- ✓ 海外チャンネルの拡大、営業力の不足
- ✓ 煩雑な法的手続き、サンプルデータ提供だけでも一苦勞
- ✓ コミュニケーションコストの増加

- AWS Data Exchangeを介して世界中のお客様への販売チャンネルを獲得
- Provider と Buyer の契約手続き、同意取得も画面上で完結
- 販売データの保管とお客様への転送は、Data Exchange が実施
- 請求、料金回収代行、報告もData Exchangeが実施

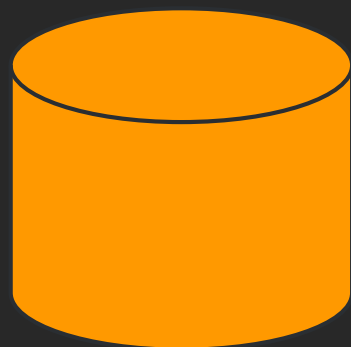
3 . AWS Data Exchangeでの 提供コンテンツ

① 新聞記事データ

大量のテキストデータとしてAI学習用教師データや、テキストマイニングなどの分析用に活用する需要が2017年ごろから発生

⇒ 分析用に販売、**Keyword**や公開日時についての構造化データ

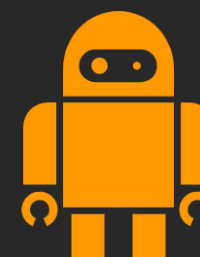
⇒ 大量のテキストデータを適切な価格で



Nikkei Article DB



Article Data Files



AI, Text Mining

項目名	備考	実例
記事ID(DBユニークキー)	"NIRKDB20150715NKM0351" など	NIRKDB20200628NKM0048
掲載日(datetime型)	"2016-09-25T23:33:26Z" など ※新聞各紙の時分は00:00:00固定	2020/6/28 9:00
媒体略号	媒体のユニークコード	NKM
媒体名称	"日本経済新聞 朝刊"、"日経速報ニュースアーカイブ"など	日本経済新聞 朝刊
絵・写真・表の有無	"" or "有"	
記事本文段落数	段落数	6
記事本文文字数	文字数	721
記事見出し	見出し	P C R 需要、企業で拡大、大手の2割が実施検討、海外渡航に必要な例も。
記事本文	本文	<p>新型コロナウイルス感染を判断するP C R検査の需要が大手企業の間で広がっている。日本経済新聞が主要85社に聞き取り調査したところ、三菱商事が既に検査を実施したと回答した。全体の約2割の16社が「実施を検討中」と答えた。海外事業の比重が高い機械や自動車関連の企業で海外渡航時にP C R検査を求められるケースが出てきている。（関連記事7面に）</p> <p>P C R検査はウイルス感染の判定などに使う遺伝子検査の一つ。感染拡大当初は発熱が一定期間続くなど実施には条件があったが、最近は感染の疑いがない無症状者であっても受け付ける民間の診療所が増えている。</p> <p>アンケートでは「希望する従業員にP C R検査を受けさせる計画はあるか」と各企業に聞いた。三菱商事は海外駐在を予定している社員など対象を限定。6月中旬から実施している。検査には社内診療所に加え外部の検査機関を活用している。</p> <p>「検討中」はブリヂストンやDMG森精機、日本航空やテルモなどの16社。東京エレクトロンは海外派遣する技術者に検査を受けさせることを検討している。「顧客から（P C R検査の）陰性証明書（3面きょうのこと</p>
キーワード	記事の文中から主題語として切り出したワード（またはその正式名称）	三菱商事,テルモ,ブリヂストン,大丸松坂屋百貨店,日本航空,DMG森精機,西武ホールディングス,東京エレクトロン,検査,P C R検査,P C R需要,感染,新型,日経調査,ウイルス,コロナウイルス,大手企業,実施検討
分類語	記事内容のテーマコード（#W～）・業界コード（#B～）、証券コード等の会社コード（T～、N～、PD～）、紙面名等の記事分類キーワード（\$～）、コラム名（「～」）	T 8 0 5 8,PD 4 3 1,T 4 5 4 3,PD 3 1 3,T 5 1 0 8,PD 1 3 1,T 9 2 0 1,PD 6 1 1,T 6 1 4 1,PD 2 1 1,T 9 0 2 4,PD 5 5 1,T 8 0 3 5,PD 2 3 9,\$ 一面,# K 7,# B 0 3 2 0,# B 0 4 7 0,N 0 0 0 1 5 9 2,N 0 0 1 3 6 1 2,N 0 0 0 0 6 9 7,N 0 0 0 1 6 5 5,N 0 0 0 1 9 4 1,N 0 0 0 1 1 5 4,N 0 0 3 8 6
株式コード	東証コード	8058,4543,5108,,9201,6141,9024,8035

① 新聞記事データ

(事例①)

保険会社様から災害に関する記事データ

(事例②)

IoT社会の実現に関するレポート作成のため社会面記事データ

(事例③)

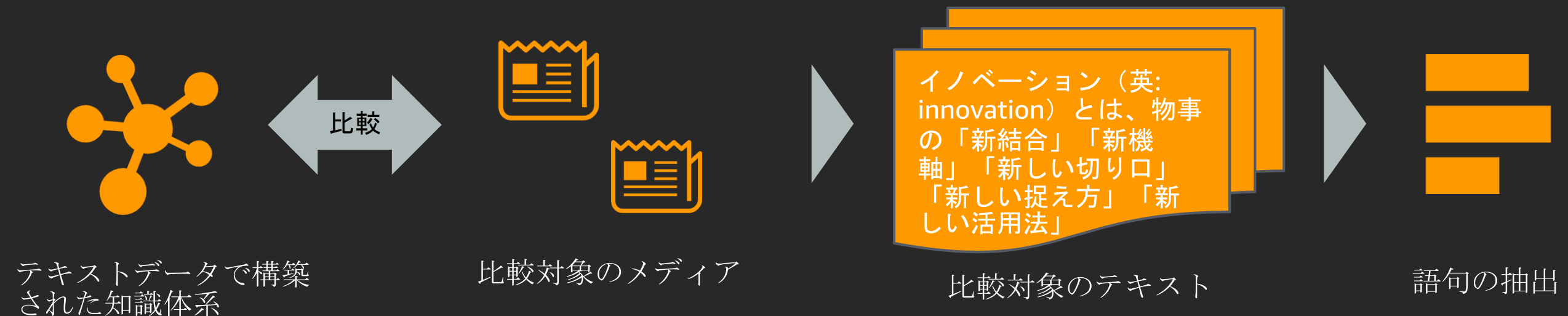
株価予測モデル構築のため記事データ10年分

事例) 日経の報道テーマの検証・評価

課題：他メディアでは多く取り上げられているが日経では低頻度？

手法：

- Neo4jでテキストデータから知識構造を構築し、他メディアと比較する
- 記事がもつ階層構造と自然言語処理の結果を蓄積
- 特定のテーマへの偏重可視化や、取り上げるべき課題の提案に寄与



記事の階層構造をそのままにDB化、グラフで表現

● 階層的な構造をグラフで表現する

- Package: ある日の朝刊
- Page: 面
- Article: 朝刊に掲載された記事
- Headline: 記事の見出し
- Sentence: 記事に含まれる文章
- Word: 文章に含まれる語句



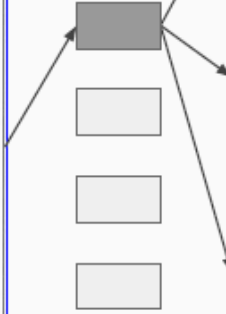
Package



Page



Article



Paragraph

固有表現抽出の結果は **NER: エッジ** で表現

学生データ活用 悩む企業 リクナビ問題で基準示されず 判例は採用に幅広い裁量

Headline

就職活動をする学生に関するデータの活用はどこまで認められるのか、企業の困惑が続いている。

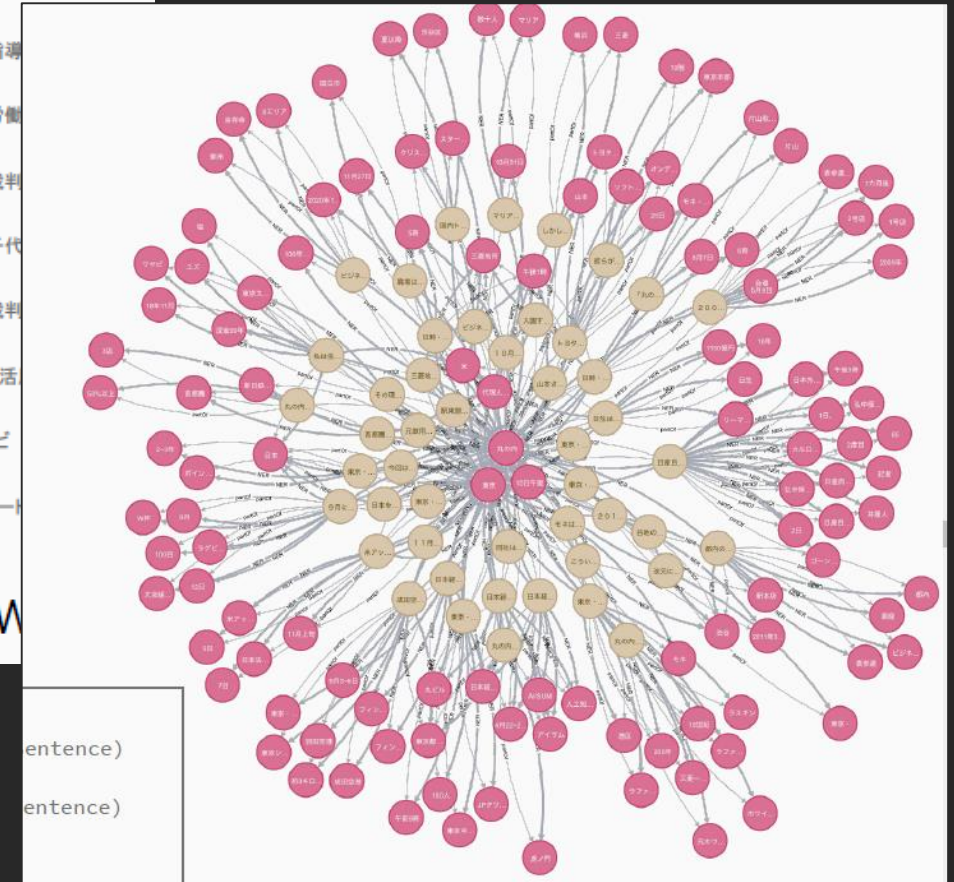
就職情報サイト「リクナビ」を運営するリクルートキャリア(東京・千代田)の「内定辞退率」販売問題で厚生労働省は利用企業を行政指導した。

一方で最高裁判例は企業の採用に幅広い裁量を認めており、適正なデータ活用の基準は不明確なままとなっている。

NER: LOCATION

NER: COMPANY

- 就職活動
- 内定辞退率販売問題
- 就職情報サイト
- 行政指導
- 厚生労働
- 最高裁判
- 東京千代
- 最高裁判
- データ活
- リクナビ
- リクルー



entence)
entence)

② 日経POS情報

日経が独自に収集しているPOS（販売情報管理）データ

国内約1500店舗のデータを直接買い付け、商業統計に合わせて店舗を選択して提供

「どのメーカーの商品がどれだけ売れたか」実数値を提供、2日ディレイで日本の消費動向を把握できる



全国1,500店舗のスーパーマーケット、コンビニエンスストア、ドラッグストア



NIKKEI



小売



金融



メーカー



官公庁

項目名	内容	実例
地域	47都道府県別に集計可能	全国
業態		日経収集店舗・全スーパー
年月日	日次・週次・月次・年次	2020年7月21日
対象店舗数	集計日に該当商品の売上があった集計店舗数	458
来店客数	集計日に該当商品の売上があった集計店舗に来店した人数	1397537
営業日数	のべ営業日数	458
商品名		グリコ 朝食りんごヨーグルト 140G
商品コード	JANコード	45184703
登録日	POSデータに記録の残る最古の売上日付	2018年7月12日
最新売上日	POSデータに記録の残る最新の売上日付	2020年7月23日
メーカー名		江崎グリコ
メーカーコード	GS1事業者コード	4901005
東証コード	東証コード	2206
大分類名		ヨーグルト
大分類コード		45
小分類名		フルーツ入りヨーグルト
小分類コード		45002
千人当り金額	対象店舗・期間での販売金額合計÷対象店舗・期間の来客数合計×1000	29
金額シェア	対象店舗・期間の全対象商品の販売金額の合計を100とした時の個々の割合	0.09
販売金額	対象店舗・期間での販売金額合計	39923
特売販売金額	販売金額実数値のうち特売で販売された販売金額	27011
最低価格	対象店舗・期間・期種で最も安い価格	50
最高価格	対象店舗・期間・期種で最も高い価格	128
カバー率		29.7
加重カバー率		40.5
出現店来店客数		565913
販売店舗数		136

② 日経POS情報

(事例①)

自社スーパーと全国平均とのギャップ分析、売上テコ入れ

(事例②)

食品メーカーの商品開発、長期サイクルでのブーム予測

(事例③)

店頭売り上げの増減から、食品メーカーの業績予測モデル

事例) 日経CPINOW/ Firm Sales Now



NOWCAST社と日経の共同事業

(日経CPINOW)

日本初の「日次」消費者物価指数、日経と東大が**2013年**に開発

原データとして、主に日経**POS**データを活用、販売価格と販売数量のデータから速報性・精度共に優れた消費者物価を日々計測

(Firm Sales Now)

スーパー店頭での売上から、食品メーカー・日雑品メーカーの業績を予測

速報性が高く、正確な業績予測が可能

日経CPINOW



DASH BOARD



NEWS



GRAPH



DATA



REPORT



SEMINAR

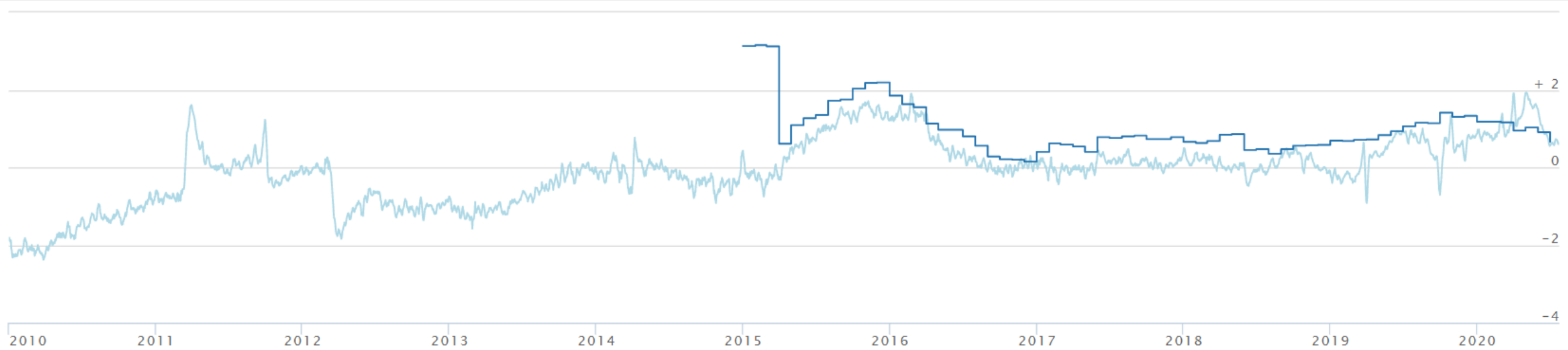


SCHEDULE



FAQ

日経CPINow (日経ナウキャスト日次物価指数)



T指数

2020/07/23

0.63%



S指数

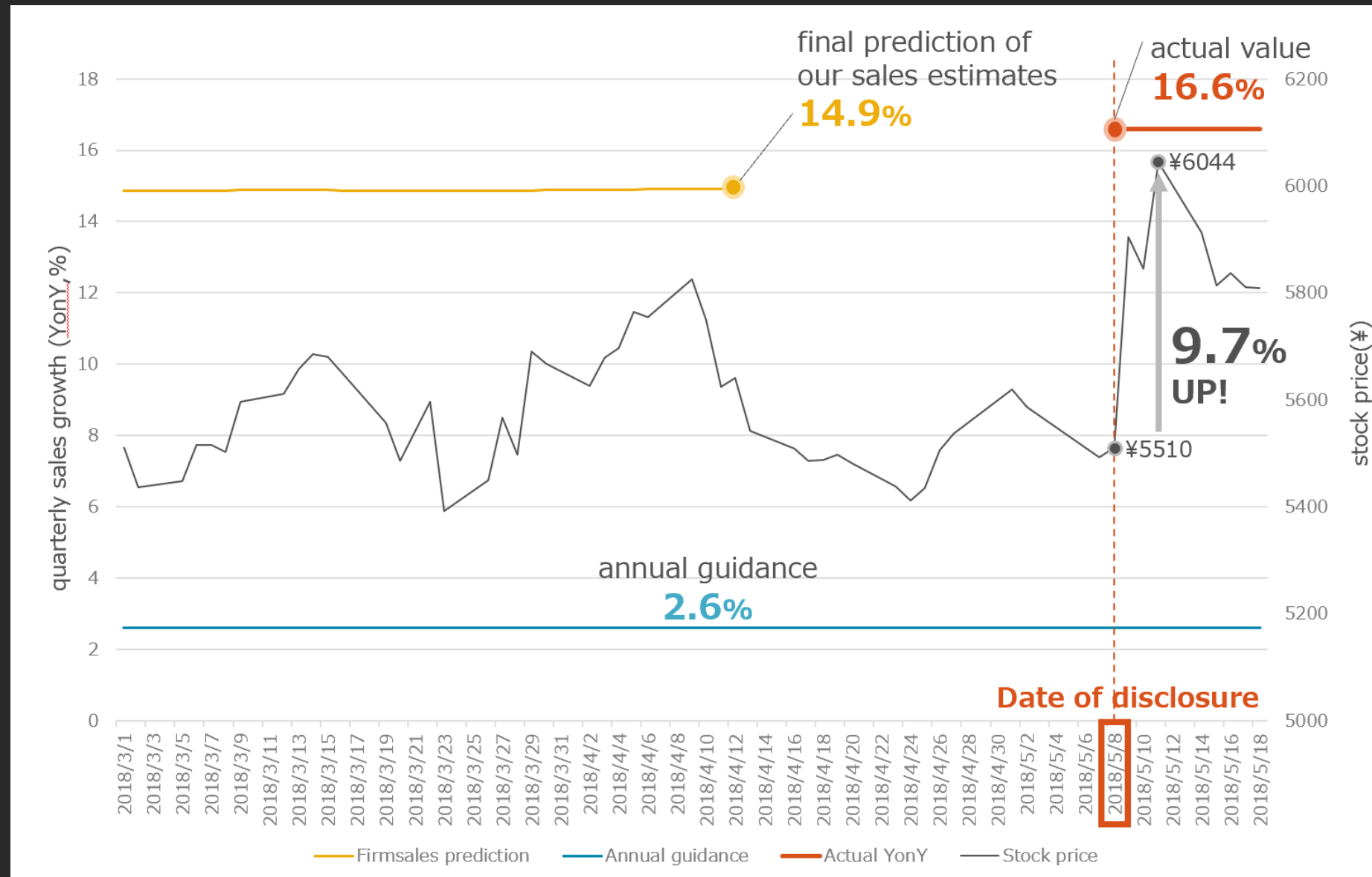
2020/07

0.66%



ニュース一覧

日経POS情報による企業業績の予測



(2018年、国内食品会社の事例)

市場の売上予測（コンセンサス）は、前年比で+ 2.6%だった。

FirmSalesNowの売上予測では、大幅な成長が見込まれ、前年比+ 14.9%と予測。

決算発表での実際の結果は+ 16.6%。

ラインナップ

日経オルタナティブデータHP :

<https://nkbb.nikkei.co.jp/alternative/>

日経POS情報HP :

<https://nkpos.nikkei.co.jp/>

お問い合わせ :

nikkei-data@nex.nikkei.co.jp

Thank you!