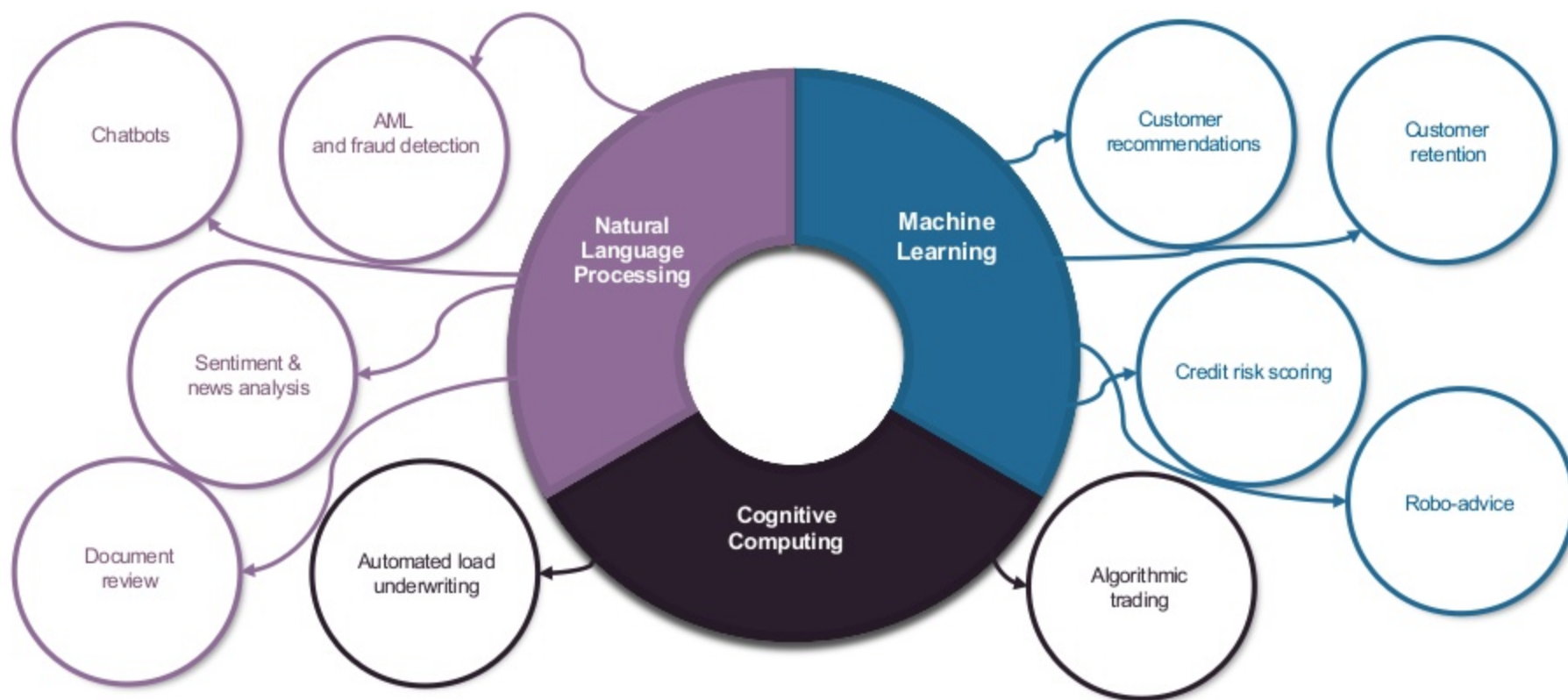


An AI Use Case: Market Event Impact Determination via Sentiment and Emotion Analysis

Lei Gao, IBM
Jin Wang, IBM

#SAISAI6

Opportunities in financial services



What's new in financial services industry?



SOURCE: World Economic Forum

Machine readable news

- Process news feeds through algorithms in real-time without human interpretation (machine-readable news)
- Discover major events faster than the news through social media / sentiment analysis

Big Data

- Access extensive real-time data sets through specialized databases
- Update and access insights in real-time through cloud-based analytics

AI / Machine Learning

- Ask questions, discover and test hypotheses, and make decisions automatically based on advanced analytics
- Self-correct and continuously improve trading strategies with minimal human interaction

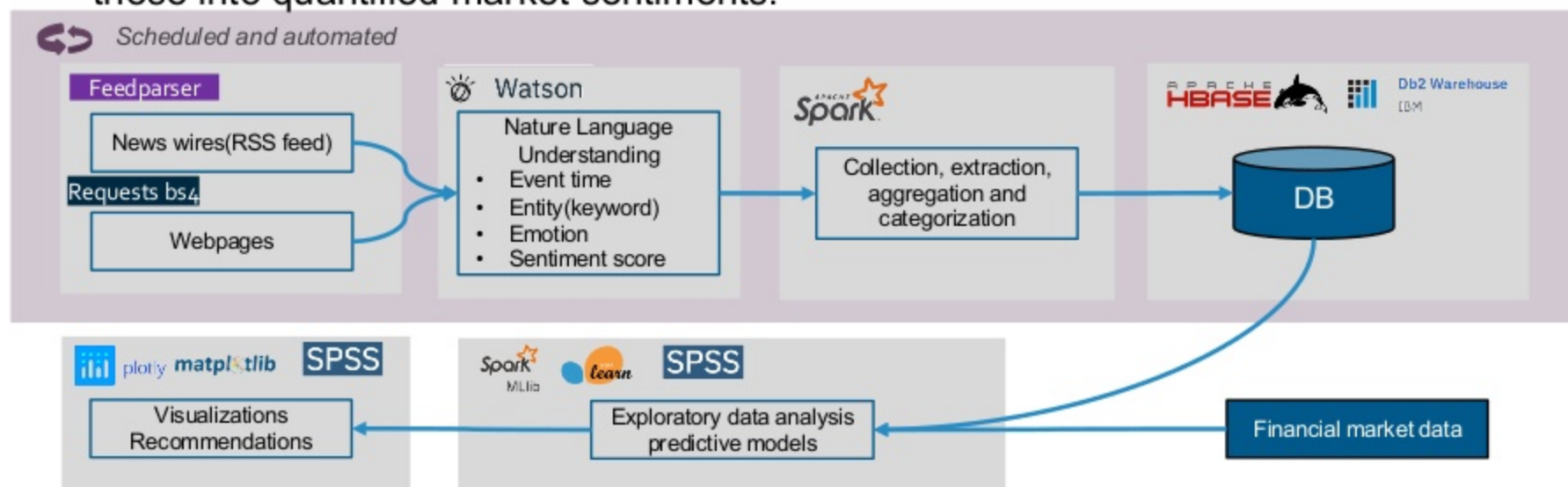
Use Case

Market Event Impact Determination













Natural language processing of news wires to add an additional dimension and timing to market data, then drive machine learning algorithms to determine potential for market impact.

Our Understanding

- Social sentiment analysis uses techniques in the areas of natural language processing, text mining, pattern classification, and econometric modeling.
- The main focus is to automate the process of understanding news presented qualitatively in the form of textual narratives appearing in news wires and social media and turning these into quantified market sentiments.



Watson AI Services

AI Assistant		Discovery		Language			
 Assistant	 Discovery  Natural Language Understanding  Knowledge Studio	 Language Translator  Natural Language Classifier					
Empathy		Visual		Speech		Tools	
 Personality Insights  Tone Analyzer	 Visual Recognition	 Speech to Text  Text to Speech	 Watson Studio				

* Access the services from IBM Cloud: <https://console.bluemix.net/>

Watson AI Services

Catalog

Search

Filter

All Categories

Compute
Containers
Networking
Storage
AI
Analytics
Databases
Developer Tools
Integration
Internet of Things
Security and Identity
Starter Kits
Web and Mobile
Web and Application

AI

Watson Assistant (formerly Conversation)

Lite • IBM

Add a natural language interface to your application to automate interactions with your end users. Common applications

Discovery

Lite • IBM

Add a cognitive search and content analytics engine to applications.

Knowledge Catalog

Lite • IBM

Discover, catalog, and securely share enterprise data.

Knowledge Studio

Lite • IBM

Teach Watson the language of your domain.

Language Translator

Lite • IBM

Translate text, documents, and websites from one language to another. Create industry or region-specific translations via

Machine Learning

Lite • IBM

IBM Watson Machine Learning - make smarter decisions, solve tough problems, and improve user outcomes.

Natural Language Classifier

IBM

Natural Language Classifier performs natural language classification on question texts. A user would be able to train their data and the

Natural Language Understanding

Lite • IBM

Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment and more.

Personality Insights

Lite • IBM

The Watson Personality Insights derives insights from transactional and social media data to identify psychological traits

Speech to Text

Lite • IBM

Low-latency, streaming transcription

Text to Speech

Lite • IBM

Synthesizes natural sounding speech from text.

Tone Analyzer

Lite • IBM

Tone Analyzer uses linguistic analysis to detect three types of tones from communications: emotion, social, and

Visual Recognition

Lite • IBM

Find meaning in visual content! Analyze images for scenes, objects, faces, and other content. Choose a default model off the

Watson Studio

Lite • IBM

Embed AI and machine learning into your business. Create custom models using your own data.

FEEDBACK

IBM Cloud: <https://console.bluemix.net/>

Watson Natural Language Understanding(NLU)

With Watson Natural Language Understanding, developers can analyze semantic features of text input, including *categories, concepts, emotion, entities, keywords, metadata, relations, semantic roles, and sentiment* via REST API. The service cleans HTML before analysis by default, which removes most advertisements and other unwanted content.

- **Supported sources:**

- plain text
- HTML
- a public URL

- **Supported SDKs:**

- Android
- Python
- Node.js
- Java
- .NET, Swift, etc.

Watson NLU – Features(1)

Categories

Categorize your content using a five-level classification hierarchy.

Example

- **Input**

url: "www.cnn.com"

- **Response**

/news
/art and entertainment
/movies and tv/television
/news
/international news

Concepts

Identify high-level concepts that aren't necessarily directly referenced in the text.

Example

- **Input**

text: "Natural Language Understanding uses natural language processing to analyze text."

- **Response**

Linguistics
Natural language processing
Natural language understanding

Emotion

Analyze emotion conveyed by specific target phrases or by the document as a whole.

Example

- **Input**

text: "I love apples, but I hate oranges."
targets: "apples", and "oranges"

- **Response**

"apples": joy(score: 0.859042)
"oranges": anger (score: 0.861152)

Watson NLU – Features(2)

Entities

Find people, places, events, and other types of entities mentioned in your content.

Example

- **Input**

text: "IBM is an American multinational technology company headquartered in Armonk, New York, United States, with operations in over 170 countries."

- **Response**

IBM: Company
Armonk: Location
New York: Location
United States: Location

Keywords

Search your content for relevant keywords.

Example

- **Input**

url: <http://www-03.ibm.com/press/us/en/pressrelease/51493.wss>

- **Response**

Linguistics
Natural language processing
Natural language understanding

Metadata

For HTML and URL input, get the author of the webpage, the page title, and the publication date.

Example

- **Input**

url: <https://www.ibm.com/blogs/think/2017/01/cognitive-grid/>

- **Response**

Author: Stephen Callahan
Title: Girding the Grid with Cognitive Computing - THINK Blog
Publication date: January 31, 2017

Watson NLU – Features(3)

Relations

Recognize when two entities are related, and identify the type of relation.

Example

- **Input**

text: "The Nobel Prize in Physics 1921 was awarded to Albert Einstein."

- **Response**

"awardedTo" relation between "Noble Prize in Physics" and "Albert Einstein"
"timeOf" relation between "1921" and "awarded"

Semantic Roles

Parse sentences into subject-action-object form, identify entities/keywords that are subjects/objects of an action.

Example

- **Input**

text: "In 2011, Watson competed on Jeopardy!"

- **Response**

Subject: Watson
Action: competed
Object: on Jeopardy

Sentiment

Analyze the sentiment toward specific target phrases and the sentiment of the document as a whole.

Example

- **Input**

text: "Thank you and have a nice day!"

- **Response**

Positive sentiment (score: 0.91)

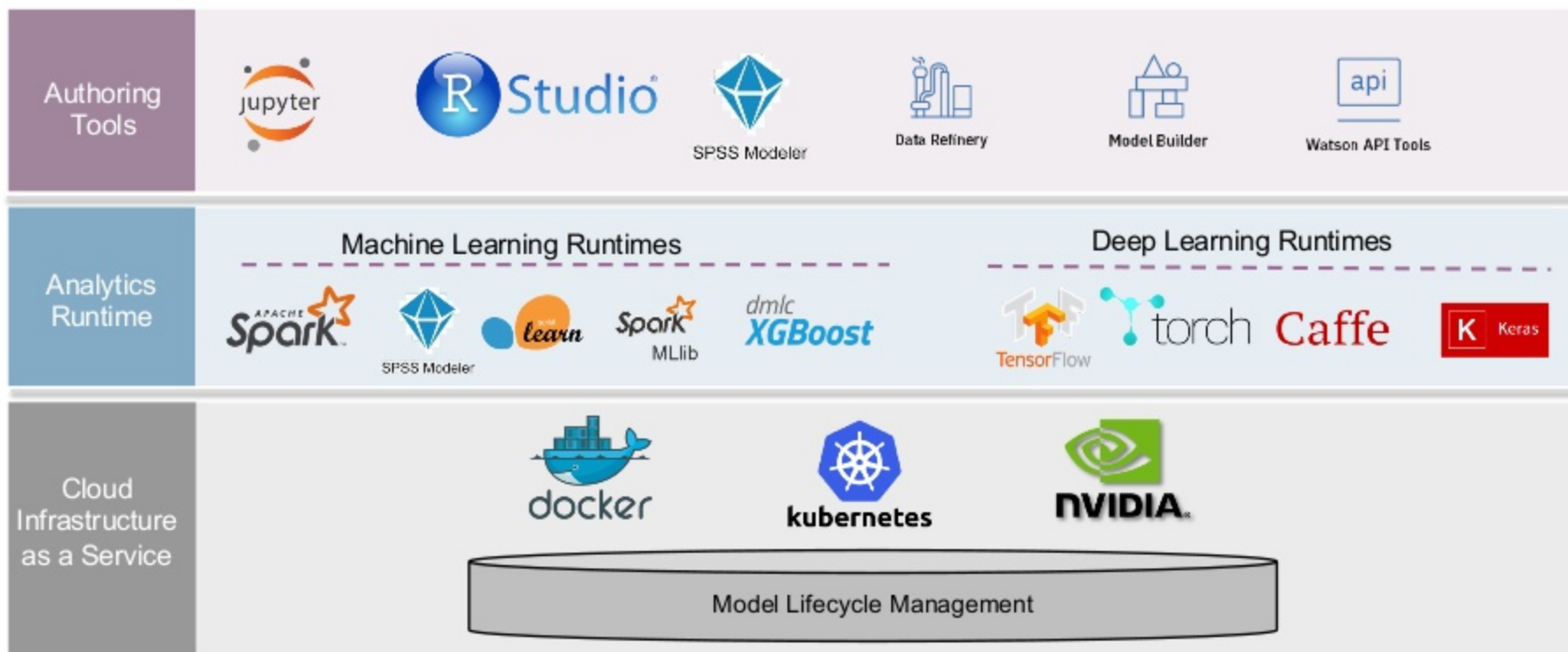
Watson Studio

Watson Studio accelerates the machine and deep learning workflows required to infuse AI into your business to drive innovation. It provides *a suite of tools for data scientists, application developers and subject matter experts to collaboratively and easily work with data* and use that data to build, train and deploy models at scale.

AI Requires Teamwork

- Algorithms
- Data
- People's collaboration

Watson Studio



SPSS

SPSS offers advanced statistical analysis, a vast library of *machine-learning algorithms, text analysis, open-source extensibility, integration with big data and seamless deployment* into applications. Its ease of use, flexibility and scalability make IBM SPSS accessible to users with all skill levels and outfits projects of all sizes and complexity.

Product family

- IBM SPSS Statistics
- IBM SPSS Modeler
- IBM SPSS Collaboration and Deployment Service
- IBM SPSS Analytic Server

SPSS Algorithms on Spark

All IBM Applications built on top of Spark

Spark ML Pipeline & Extended Pipeline

Algorithms in Java

- Scalable
- Designed with Map/Reduce
- Distributed data
- Distributed computation

Descriptives	Linear Engine	Time Casual Model	Scalable Random Trees
Chaid Tree	Generalized Linear	Time Series Explore	Support Vector Machine
Smart Report	Cluster Engine	Point Process Model	Event Based Time Series
Predictor Important	ST Data Preparation	Spatial Tempo Cluster	Smart Data Preprocess
TS Data Preparation	Geo-Spatial AR	Survival Analysis	
Score Engine	Spatial Tempo Prediction	Traditional Time Series	

Algorithms in Scala

- Scalable
- DataFrame/DataSet
- With RDD operations
- Distributed data
- Distributed computation

CART Tree	C5 Tree
TS GARCH	TS Modified Croston's
RPI	TS Anomaly Detection

An Example – Oil Price Influenced by News

Targets:

Automatically accumulate and analyze the sentiment/emotions on massive historical or real-time news from RSS feeds, so as to:

- Understand the attitudes from public news regarding oil price.
- Discover the impact of news from internet on oil price.
- Predict the trend of oil price by the recent news sentiment/emotions.

• RSS resources:

- Business News
- Company News
- Earnings News
- General RSS

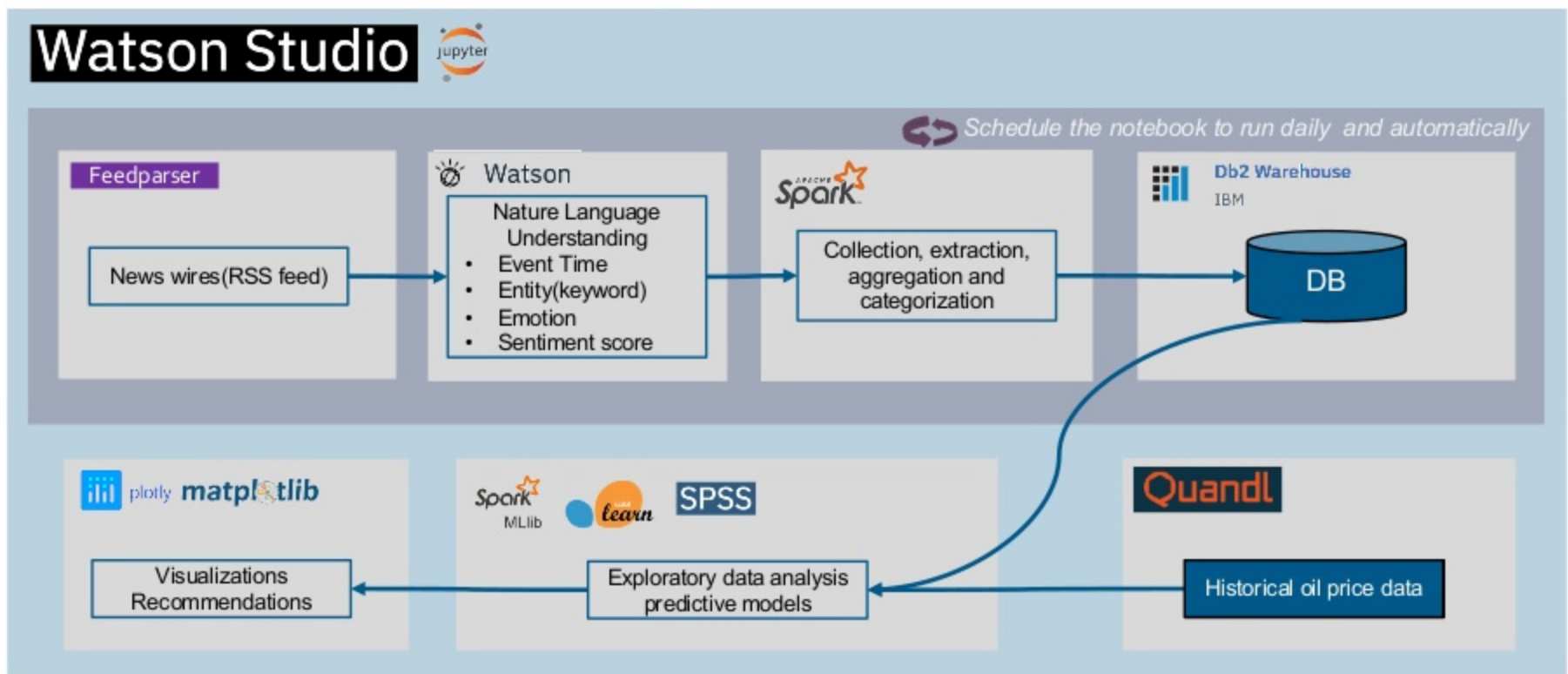
• Keyword categories:

- Types of oil(crude, WTI, etc.)
- Agencies(OPEC, API, etc.)
- Influential countries
- Etc.

• Oil price sources:

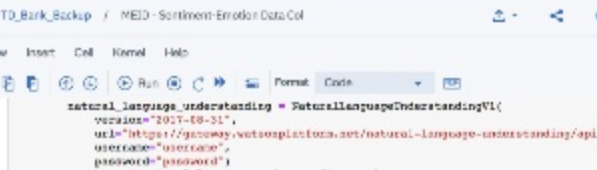
- NYMEX Crude Oil Futures
- E-mini S&P 500 Futures

Oil Price Influenced by News - Solution



Oil Price Influenced by News - Process(1)

- *Generate sentiment/emotion scores by Watson NLU API.*



The screenshot shows the IBM Watson Studio interface. At the top, there are navigation tabs: IBM Watson Studio, Projects, Tools, Community, Services, Manage, Support, and Docs. Below these is a breadcrumb trail: projects / TD_Bank_Backup / MEID - Sentiment-Error Data Col. The main area displays a Jupyter Notebook with the following Python code:

```

natural_language_understanding = NaturalLanguageUnderstandingV1(
    version='2017-08-31',
    url='https://gateway.watsonplatform.net/natural-language-understanding/api',
    username='username',
    password='password')
response = natural_language_understanding.analyze(
    url=url,
    features=features,
    sentiment=sentiment_options(target=keyword_list),
    emotion=emotion_options(target=keyword_list))

print "Sentiment--"
if "sentiment" in response.keys():
    if "targets" in response["sentiment"].keys():
        for x in response["sentiment"]["targets"]:
            for y in x:
                print y + " " + str(x[y])
                record_add_sentiment.update({"target_sentiment_" + y: x[y]})

print "Emotion--"
if "emotion" in response.keys():
    if "targets" in response["emotion"].keys():
        for x in response["emotion"]["targets"]:
            print x["text"]
            record_add_emotion.update({"target_emotion_text": str(x["text"])})
            for y in x["emotion"]:
                print y + " " + str(x["emotion"][y])
                record_add_emotion.update({"target_emotion_" + y: str(x["emotion"][y])})

```

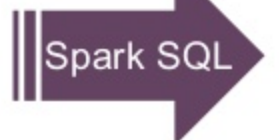
[illegible]

Oil Price Influenced by News - Process(2)

- Combine the sentiment/emotion score data with historical oil price data.

date	open	high	low	volume
2015-10-29	45.63	50.52	45.37	49.45
2015-10-30	45.51	50.13	47.64	47.44
2015-10-31	47.58	49.63	46.46	56.53
2015-11-01	46.59	46.95	45.54	45.25
2015-11-02	46.29	47.02	45.23	55.66
2015-11-03	46.27	47.02	45.23	55.66
2015-11-04	46.27	47.02	45.23	55.66
2015-11-05	46.27	47.02	45.23	55.66
2015-11-06	46.27	47.02	45.23	55.66
2015-11-07	46.27	47.02	45.23	55.66
2015-11-08	46.27	47.02	45.23	55.66
2015-11-09	46.27	47.02	45.23	55.66
2015-11-10	46.27	47.02	45.23	55.66

* Sentiment/emotion scores of RSS feeds



- Aggregation
- Categorization
- Align the data/time

```
df_results = df_collection_via.join(df_oilprice, df_collection_via.RPPRICE_DATE == df_oilprice.date).repartition(1000)
df_results.show()
```

SPARK JOB PROGRESS

JOB	PROGRESS	DURATION	STATUS
4	2 stages	5.57 sec	
5	2 stages	1.85 sec	

date	open	high	low	volume
2015-10-29	45.63	50.52	45.37	49.45
2015-10-30	45.51	50.13	47.64	47.44
2015-10-31	47.58	49.63	46.46	56.53
2015-11-01	46.59	46.95	45.54	45.25
2015-11-02	46.29	47.02	45.23	55.66
2015-11-03	46.27	47.02	45.23	55.66
2015-11-04	46.27	47.02	45.23	55.66
2015-11-05	46.27	47.02	45.23	55.66
2015-11-06	46.27	47.02	45.23	55.66
2015-11-07	46.27	47.02	45.23	55.66
2015-11-08	46.27	47.02	45.23	55.66
2015-11-09	46.27	47.02	45.23	55.66
2015-11-10	46.27	47.02	45.23	55.66

* Historical oil price

Oil Price Influenced by News - Process(3)

- Leverage SPSS Spark Time Series algorithms to build Timeseries model and perform prediction.

[illegible]

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with tabs: Home, Projects, Tools, Community, Services, Manage, Support, and Docs. Below this is a breadcrumb trail: My Projects > 10. IBM Studio > Consideration of RMI/EC2 instances and... The main workspace is titled 'Tutorial: Python 2 with Spark 2.1'. It contains a JupyterLab environment with a table of data. The table has 6 columns: 'datestamp', 'price', 'open', 'open_predicted', 'open_act', and 'open_act'. The data is organized into 10 groups, each with a 'datestamp' and a 'price' column, followed by a table of 'open' and 'open_predicted' values. The 'open_act' column is also present. The table is displayed in a grid format with a light blue header and alternating row colors.

datestamp	price	open	open_predicted	open_act	open_act
2015-10-17	0	45.81	null	null	null
2015-10-18	0	47.46	null	null	null
2015-10-19	0	46.59	null	null	null
2015-10-20	0	46.38	null	null	null
2015-10-21	0	46.87	null	null	null
2015-10-22	0	47.37	47.8830927318868	48.033889945121	48.4588888888889
2015-10-23	0	48.25	47.3641983355937	44.1675942033087	55.8284881272784
2015-10-24	0	45.88	46.1824243127142	44.1271657555478	49.4327286470612
2015-10-25	0	46.37	46.1766595833274	43.6525545181245	49.8321621283216
2015-10-26	0	46.44	46.4388708672861	43.7418884324982	49.8501768226589
2015-10-27	0	46.14	46.1753983018298	43.81508138614128	48.13158833780866
2015-10-28	0	45.18	45.6381287898545	42.959483674181285	48.17641793183234
2015-10-29	0	44.44	44.6827342881842	42.2246341222755	47.5393263933936
2015-10-30	0	46.56	46.1473229281361	44.2584685448468	47.2716336846462
2015-10-31	0	46.18	46.2796884436253	45.4627043522515	47.8384251812226
2015-11-01	0	46.43	46.8279441418136	45.87181818178388	48.1842188381838

Oil Price Influenced by News – Insights with Plots

- From Fig. 1, the timeseries fit values meet the open price well and can reflect the trend on oil price.
- From Fig. 2, the last 100 selected timeseries points has an error between fit and target value, but it is acceptable because the predict value is in the range of estimation interval.
- From both Fig. 1 and Fig. 2, We can reasonably infer that, the predict value in coming steps can be a reference for future oil price trend.



Fig. 1

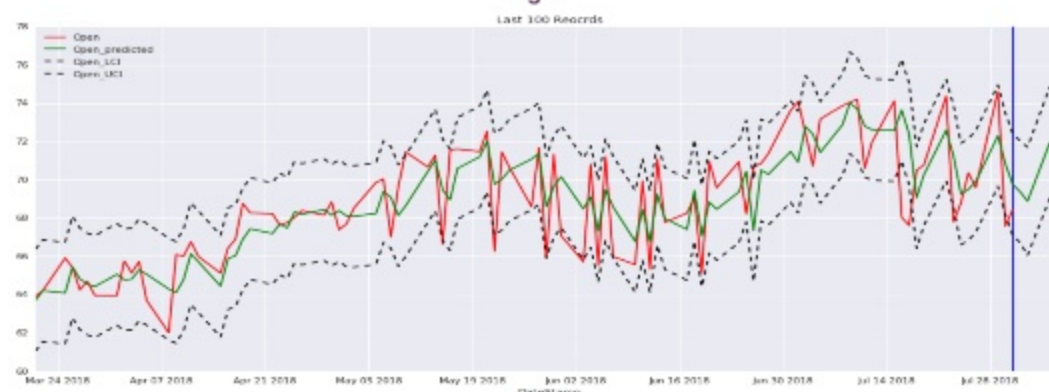


Fig. 2

Thank You!