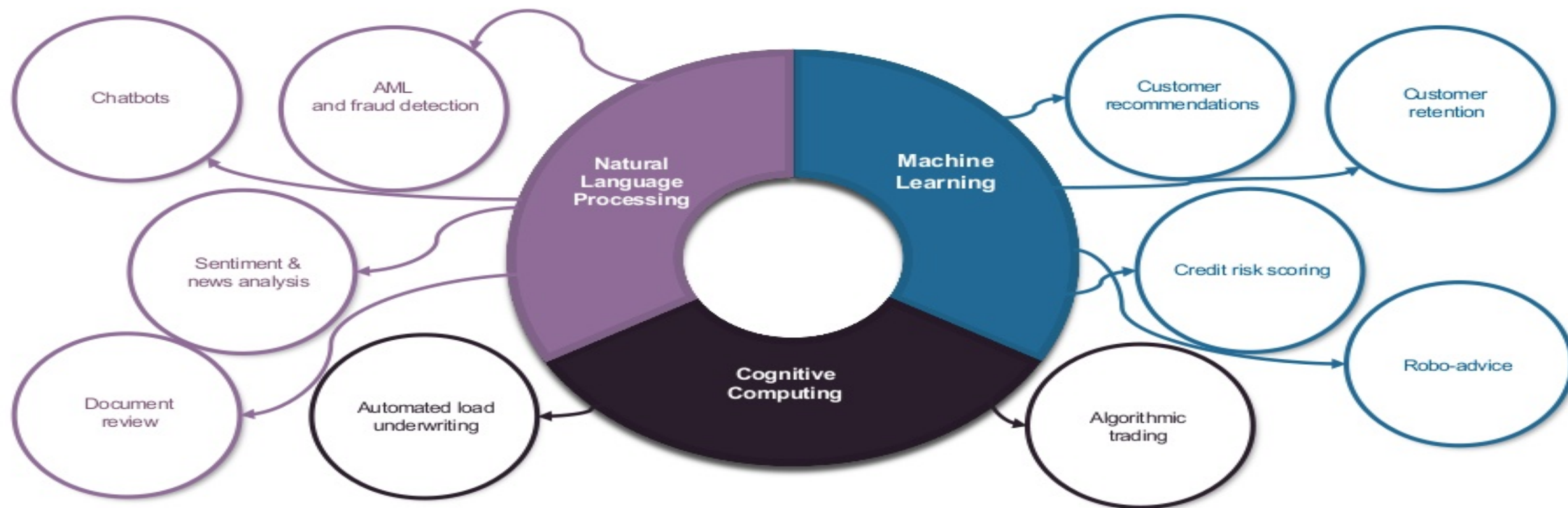


# **An AI Use Case: Market Event Impact Determination via Sentiment and Emotion Analysis**

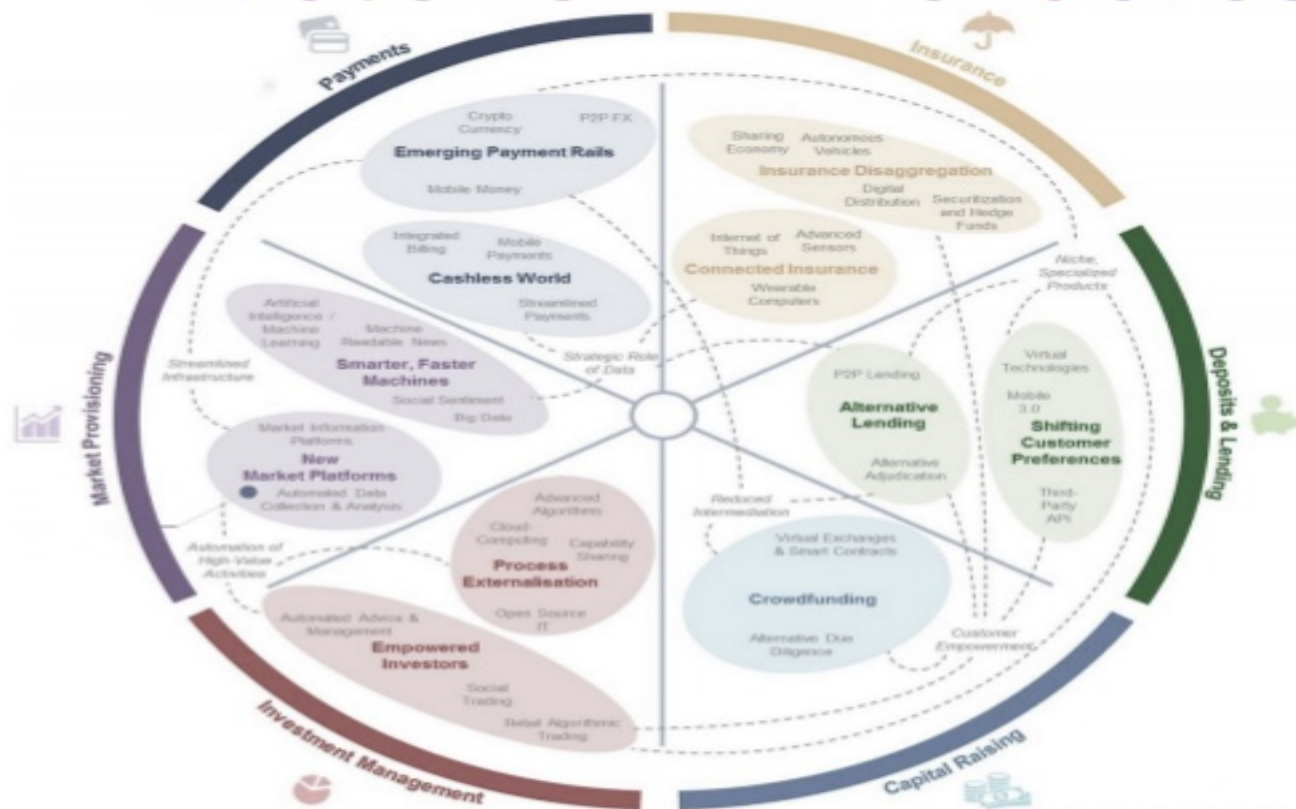
Lei Gao, IBM  
Jin Wang, IBM

**#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

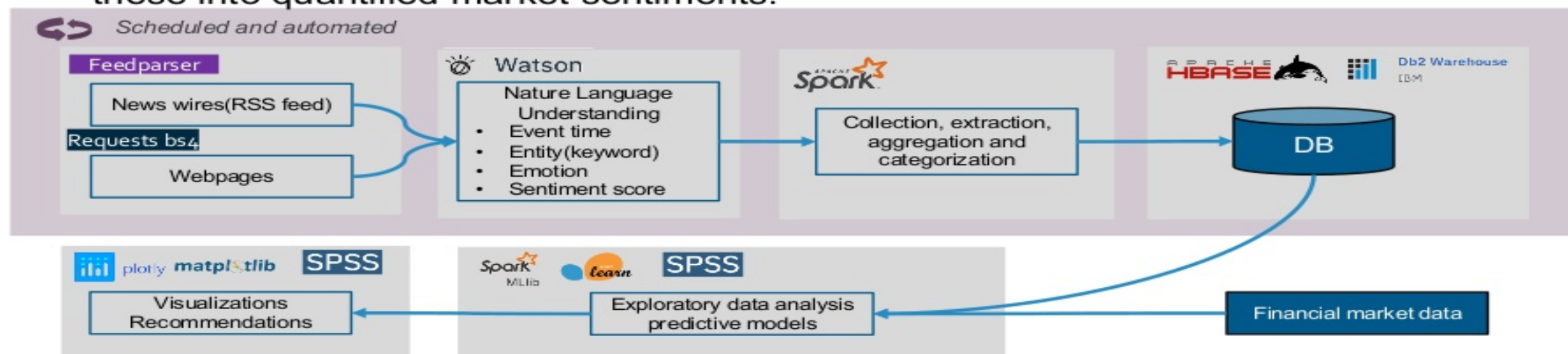
### ***Market Event Impact Determination***

*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.*





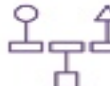









## Our Understanding

- 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		 Language Translator	
		 Natural Language Understanding		 Natural Language Classifier	
		 Knowledge Studio			
Empathy		Visual		Speech	
 Personality Insights		 Visual Recognition		 Speech to Text	 Watson Studio
	 Tone Analyzer			 Text to Speech	

\* 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

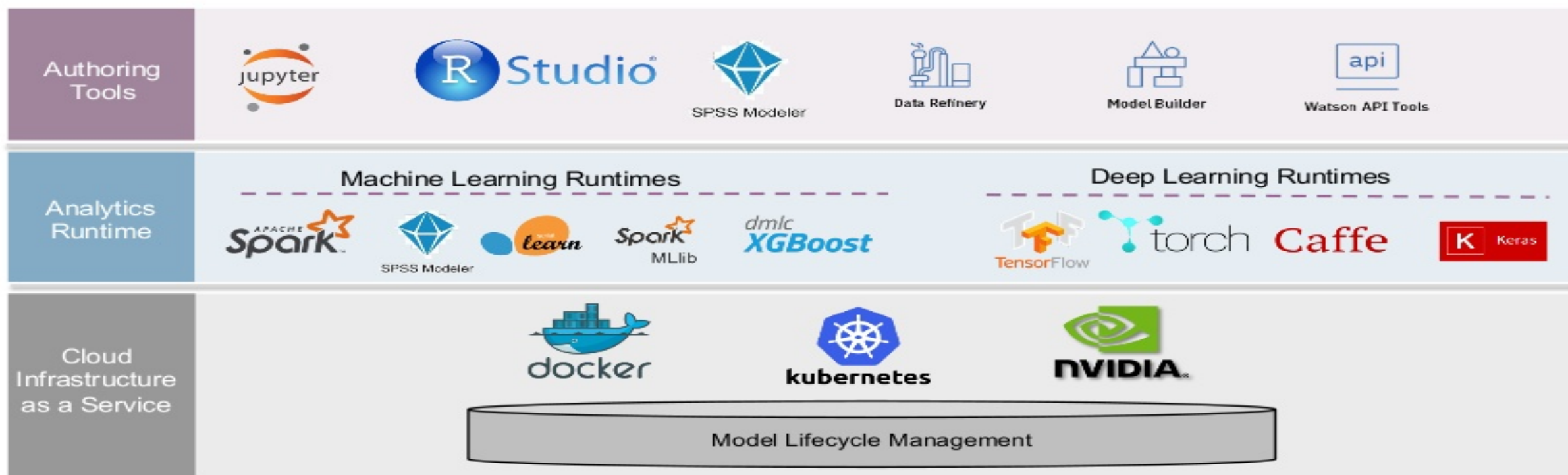
Watson Studio accelerates the machine and deep learning workflows required to infuse AI into your business to drive innovation. It provides *a suite of tools for data scientists, application developers and subject matter experts to collaboratively and easily work with data* and use that data to build,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

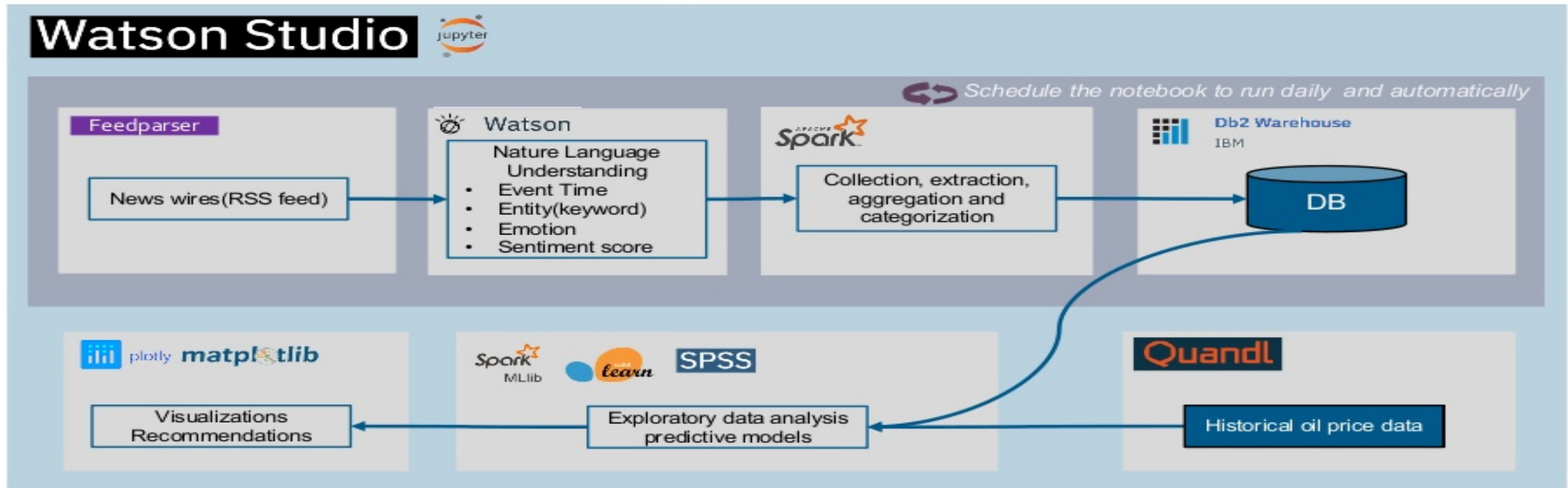
- Types of oil(crude, WTI, etc.)
- Agencies(OPEC, API, etc.)
- Influential countries
- Etc.

### • 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.

```
IBM Watson Studio  Projects  Tools  Community  Services  Manage  Support  Docs

projects / TD_Bank_Backup / MEID - Sentiment-Emotion Data Col

Edit  View  Insert  Cell  Kernel  Help

natural_language_understanding = NaturalLanguageUnderstandingV1(
    version="2017-08-01",
    url="https://gateway.watsonplatform.net/natural-language-understanding/api",
    username="username",
    password="password")
response = natural_language_understanding.analyze(
    url=url,
    features=[
        sentiment=SentimentOptions(targets=keywordlist),
        emotion=EmotionOptions(targets=keywordlist)
    ])

print "\n---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 "\n---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": x["text"]})
            for y in x["emotion"]:
                print y + " : " + str(x["emotion"][y])
                record_add_emotion.update({"target_emotion_" + y: str(x["emotion"][y])})
```



DATE	ITEM	URL	TARGET	SENTIMENT	SCORE	TARGET	EMOTION	SCORE	TARGET	EMOTION	SCORE
2015-12-11 14:14:00	http://oilprice.com	neutral	0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-10 17:44:00	http://oilprice.com	neutral	0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-10 00:40:00	http://oilprice.com	negative	-0.570451	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-09 18:16:00	http://oilprice.com	negative	-0.401004	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-08 18:08:00	http://oilprice.com	negative	-0.405414	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-05 10:07:00	http://oilprice.com	negative	-0.440136	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-05 17:09:00	http://oilprice.com	negative	-0.369667	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-04 17:06:00	http://oilprice.com	negative	-0.427960	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-03 17:19:00	http://oilprice.com	negative	-0.427960	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-03 16:46:00	http://oilprice.com	negative	-0.405704	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-02 10:10:00	http://oilprice.com	negative	-0.405704	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-01 10:06:00	http://oilprice.com	neutral	0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-28 14:58:00	http://oilprice.com	neutral	0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2015-12-27 10:57:00	http://oilprice.com	neutral	0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

# Oil Price Influenced by News - Process(2)

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

Spark SQL interface showing sentiment/emotion scores of RSS feeds. The table has columns: DATE, OPEN, HIGH, LOW, LAST, CHANGE, GAIN, VOLUME, PREVIOUS DAY OPEN, INTEREST.

DATE	OPEN	HIGH	LOW	LAST	CHANGE	GAIN	VOLUME	PREVIOUS DAY OPEN	INTEREST
2015-10-28	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-29	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-30	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-31	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-01	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-02	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-03	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-04	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-05	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-06	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-07	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-08	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-09	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-10	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-11	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-12	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-13	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-14	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-15	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-16	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-17	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-18	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-19	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-20	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-21	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-22	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-23	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-24	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-25	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-26	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-27	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-28	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-29	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-30	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0

\* Sentiment/emotion scores of RSS feeds



- Aggregation
- Categorization
- Align the data/time

Spark SQL interface showing the correlation of NYMEX futures and oil prices. The table has columns: DATE, OPEN, HIGH, LOW, LAST, CHANGE, GAIN, VOLUME, PREVIOUS DAY OPEN, INTEREST.

DATE	OPEN	HIGH	LOW	LAST	CHANGE	GAIN	VOLUME	PREVIOUS DAY OPEN	INTEREST
2015-10-28	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-29	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-30	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-10-31	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-01	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-02	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-03	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-04	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-05	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-06	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-07	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-08	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-09	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-10	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-11	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-12	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-13	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-14	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-15	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-16	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-17	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-18	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-19	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-20	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-21	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-22	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-23	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-24	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-25	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-26	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-27	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-28	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-29	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0
2015-11-30	45.51	46.02	45.37	45.49	0.2	0.4	400040.0	255295.0	0.0

\* Historical oil price



## Oil Price Influenced by News - Process(3)

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

[illegible]

IBM Watson | Projects | Tools | Community | Services | Manage | Support | Docs

My Projects > 10. Bank Backup > Overview of RNNBD Patterns and...

File | Edit | View | Insert | Cell | Kernel | Help

Run | Format | Curl

Date/Time	Status	Open	Open Predicted	Open Act	Open Act
2015-10-15	O	10 steps	0.48 sec	0.48 sec	0.48 sec
2015-10-12	O	10 steps	1.08 sec	1.08 sec	1.08 sec
2015-10-12	O	1 step	0.12 sec	0.12 sec	0.12 sec
2015-10-14	O	10 steps	0.08 sec	0.08 sec	0.08 sec
2015-10-14	O	1 step	0.08 sec	0.08 sec	0.08 sec
2015-10-14	O	2 steps	0.08 sec	0.08 sec	0.08 sec
2015-10-15	P	10 steps	0.48 sec	0.48 sec	0.48 sec
2015-10-15	P	10 steps	1.08 sec	1.08 sec	1.08 sec
2015-10-15	P	1 step	0.12 sec	0.12 sec	0.12 sec
2015-10-15	P	10 steps	0.08 sec	0.08 sec	0.08 sec
2015-10-15	P	1 step	0.08 sec	0.08 sec	0.08 sec
2015-10-15	P	2 steps	0.08 sec	0.08 sec	0.08 sec



# 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.

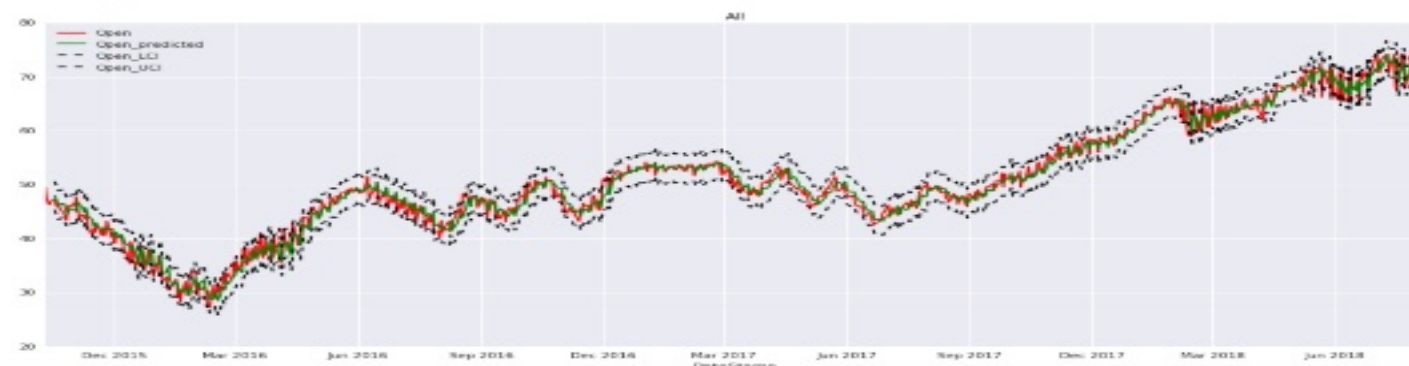


Fig. 1



Fig. 2

# Thank You!