



Le Data-mining au service de problématiques économiques

Sandrine LUNVEN
Economiste senior, TAC ECONOMICS

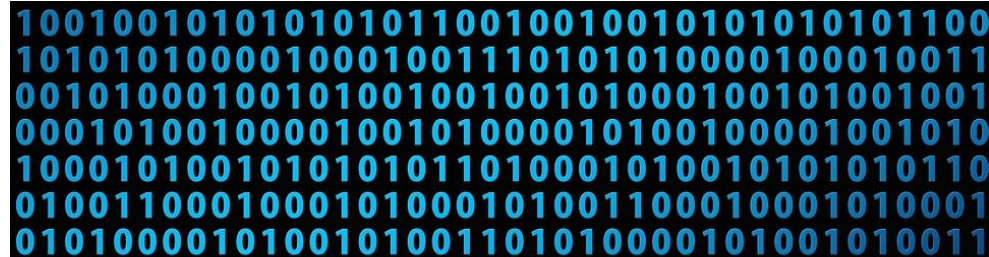
Meetup Breizh Data Club
Avril 2018 - Rennes

TAC ECONOMICS

- Société de recherche appliquée en économie et en finance, qui accompagne les groupes à l'international au travers de notes, de signaux d'alerte et d'échanges.
- Une société qui existe depuis 25 ans. Une équipe d'économistes, « analystes » et « quant ».
- Une clientèle de banques au départ, puis des industriels (gros et moins gros) et des gestionnaires d'actifs.
- Environ 20 salariés, installés au nord de Rennes... et en Inde !
- Quatre grands domaines d'activité: (1) les abonnements aux outils de mesure du risque pays, (2) les études stratégiques, (3) les développements spécifiques, (4) les études pour la Commission Européenne et les Nations Unies... et l'enseignement.

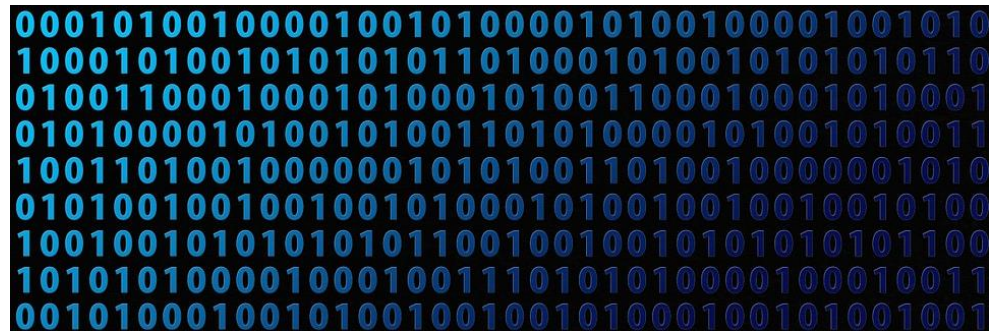
Plan de l'intervention

1. Intérêt du machine learning en macroéconomie
2. Risque pays
3. Marchés financiers
4. Conjoncture : croissance du PIB, pétrole
5. Taux de change EUR/USD
6. Text mining pour mesurer sentiment économique, élections etc.



1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011

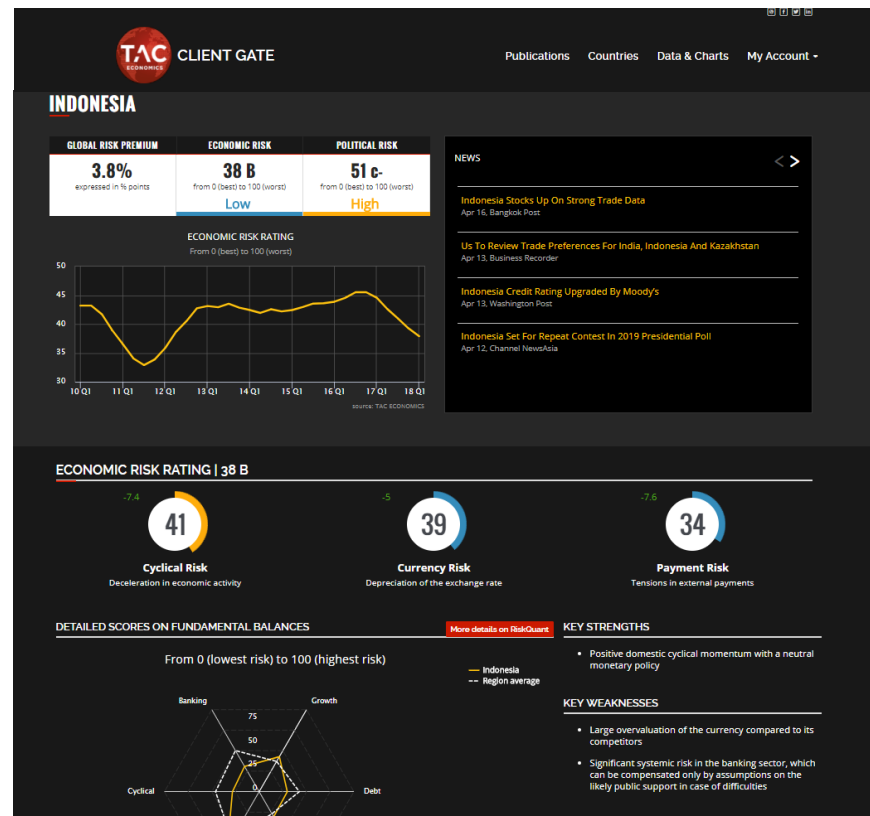
Machine Learning en Macroéconomie



0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001001

Evolution rapide du métier d'économiste « quant »

- **Plus grande accessibilité aux données grâce à l'open Data.** Malgré des historiques parfois courts, de plus en plus d'indicateurs sont disponibles
- Construction d'une base de données TACSTAT mise à jour en temps réel englobant un grand nombre de sources (Eurostat, ECB, IMF, FRED, OECD, EIA, Yahoo etc.) et différents types d'indicateurs (macro, sectoriel, financier) et même des news!
- Utilisation de la base directement sur R et Excel
- **Plus de puissance de calcul, plus de puissance de stockage**



Pourquoi utiliser du machine learning en macroéconomie?

- **Performance des modèles** avec différents types de cible (cible « équilibrée », signaux « faibles » etc)
- **Performance pour identifier des patterns**, des combinaisons critiques non-linéaires parmi un grand ensemble d'indicateurs
- Beaucoup d'entre eux sont presque entièrement « automatisables » (à la différence des approches économétriques traditionnelles)
- Attention au risque de sur-apprentissage, au « machine learning sauvage », aux risques de bases de données « poubelles » (données erronées, pas de notion d'intégration/stationnarité, liens économiques), phénomènes de « taches solaires », attention au sampling, etc...
- Certains algorithmes sont des « boites noires »... mais pas tous !
- **Meilleure compréhension** des interactions non-linéaires entre les variables économiques, aide à l'analyse!

```
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
```

Machine Learning et Risque Pays

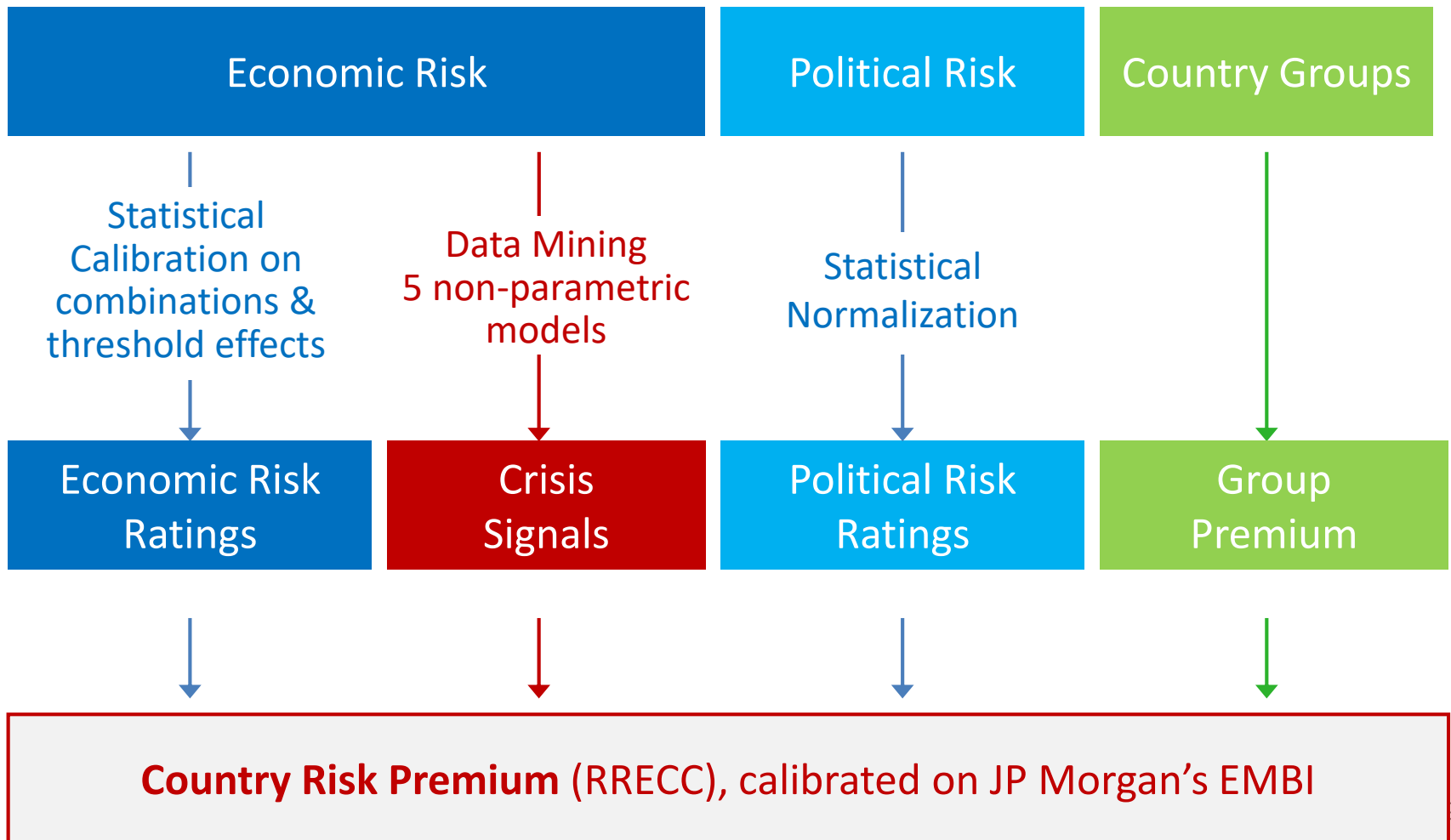
```
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001
```

Prévoir les crises économiques et financières

- De nombreuses études empiriques sur le risque pays et les indicateurs avancés de crises économiques et financières: Krugman (1979), Obstfeld (1994), Cantor and Packer (1996), Eichengreen et al. (1996), Frankel and Rose (1996), Goldstein (1996), Goldstein and Turner (1996), Kaminsky and Reinhart (1999), Komulainen and Lukkarila (2003) , ...
- Mais malgré les classifications existantes, un très grand nombre de crises économiques et financières ont laissé les observateurs perplexes.
- Le manque de relations causales homogènes et des interactions complexes rendent l'identification ex-ante des facteurs de risque extrêmement difficile.

RiskMonitor

- Outil de détection avancée de crises financières dans les économies émergentes et calcul de rating (développé il y a 25 ans!)



Early Warning à différents horizons

- L'outil « Early Warning » fonctionne à l'aide d'une combinaison de méthodes de data-mining à apprentissage supervisé:
 - Réseau de Neurones
 - SVM
 - Random Forest
 - Recursive Partitioning
 - Logit
- Les outils sont calibrés sur des historiques de crises observées sur la totalité de l'échantillon depuis 1970 (bootstrapping/cross validation pour sélection)

Performances de notre outil de « Early Warning »

Signaling power of different models on exchange rate crises

Horizon	Model	Sensitivity (%)	PPV (%)	Accuracy (%)
≤1 year	LDA	0.7	16.7	94.6
1–3 years	LDA	6.1	46.7	90.0
3–5 years	LDA	–	–	–
≤1 year	SVM	19.1	100.0	95.7
1–3 years	SVM	34.2	97.5	93.4
3–5 years	SVM	34.6	100.0	93.8
≤1 year	KNN	90.1	85.8	98.7
1–3 years	KNN	94.3	93.5	98.8
3–5 years	KNN	91.8	91.8	98.4
≤1 year	CDM-Model	97.2	100.0	99.9
1–3 years	CDM-Model	97.8	100.0	99.8
3–5 years	CDM-Model	91.8	99.4	99.2

The signaling power of the different models is calculated using different models of supervised learning over the sample of 50 countries covered by RiskMonitor. This sample contains more than 2000 observations of macroeconomic data and a low percentage of crises (from 5% for the ≤1 year horizon to 10% for the 3 to 5 years horizon).

Exemple sur la crise de la Thaïlande en 1997

Crisis signals of the CDM-Model on Thailand from 1991 to 1997

Year	≤ 1 year	From 1 to 3 years	From 3 to 5 years
1991	stability	stability	stability
1992	stability	stability	warnings
1993	stability	stability	crisis signal
1994	stability	warnings	crisis signal
1995	stability	crisis signal	warnings
1996	crisis signal	crisis signal	stability
1997	<i>-The exchange rate policy was abandoned on July 2-</i>		

This table presents the crisis signals (within the forecasted horizon) that the CDM-Model rises on Thailand before the financial crisis of 1997, when the Baht depreciated for more than 83% over 6 months.

```
10010010101010101100100100101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
```

Machine Learning et Marchés financiers

```
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001
```

Quantitative Market Alert

Summary of Early Warning Signals

	Market Level (Apr. 03)	04.2018	05.2018	06.2018	3-month outlook	07.2018	08.2018	09.2018	6-month outlook
Equity Indices									
S&P 500	2 614				↗				↗
CAC 40	5 152				↗				↗
DAX 30	12 002				↗				↗
FTSE 100	7 030				↗				↗
Nikkei 225	21 292				↗				↗
10-Year Government Benchmarks									
US Treasury 10 Year Yield	2.78%				↗				↗
French OAT 10 Year Yield	0.74%				↘				↘
German Bund 10 Year Yield	0.50%				↘				↘
UK Gilt 10 Year Yield	1.36%				↘				↘
Japan JGB 10 Year Yield	0.03%				↘				↘
Corporate Spreads									
Euro 5-7 Year A	32 bp				↗				↗
US 5-Year A	94 bp				↗				↘

Source: TAC ECONOMICS

Legend:

Strong improvement	Strong deterioration	Double signal	No sudden change

↗ ↘ Medium term outlook compared to the current market level / - indicates neutral outlook

Quantitative Market Alert

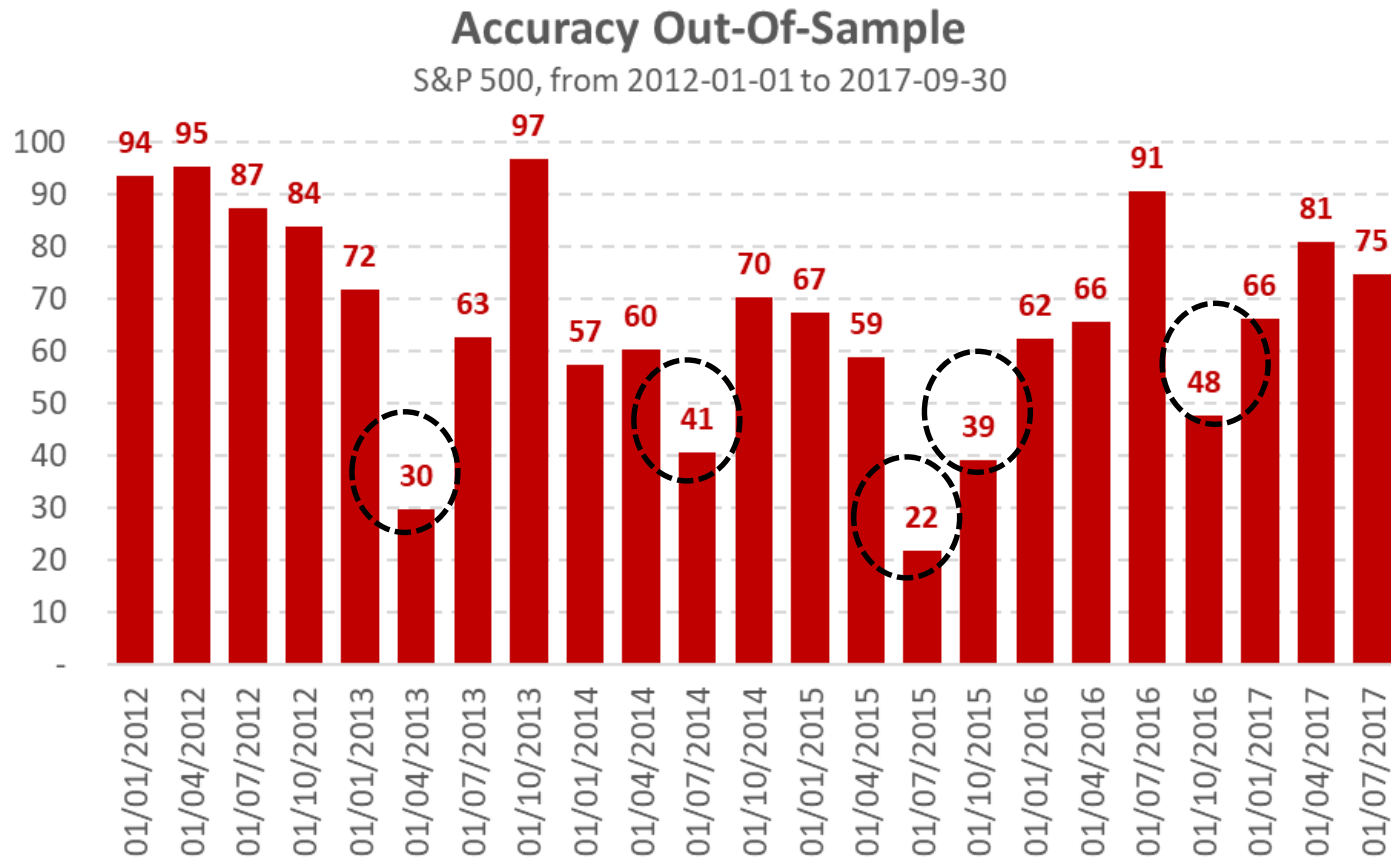
Gini index sur baisse forte du S&P500 à 1 mois

cible	sp500.l.21	
train	01/01/2003	31/12/2014
valid	0.2*Train/50	
test	01/01/2015	31/12/2017

				in	valid
Importance	Category	nb vars	model	gini	gini
All	All	225	logit	86.00	58.40
			cart	90.38	64.73
			rf	100.00	92.35
	Financière	73	logit	95.12	90.22
			cart	91.96	76.55
			rf	100.00	92.12
	Macroéconomique	139	logit	60.28	61.03
			cart	87.77	93.49
			rf	97.71	88.02
	chartisme	6	logit	20.58	46.11
			cart	77.08	43.29
			rf	100.00	41.45
	déclencheur	7	logit	38.73	59.36
			cart	85.99	56.74
			rf	99.99	83.56
1	All	65	logit	89.02	73.45
			cart	94.08	75.32
			rf	100.00	90.69
	Financière	38	logit	81.54	66.62
			cart	93.77	58.59
			rf	100.00	89.38
	Macroéconomique	18	logit	66.79	65.49
			cart	87.85	70.76
			rf	88.38	80.79
	chartisme	6	logit	20.58	46.11
			cart	77.08	43.29
			rf	100.00	41.45
	déclencheur	3	logit	37.01	50.27
			cart	86.61	51.58
			rf	92.68	66.06

Quantitative Market Alert

Performances DTO 1 mois, 2012-2017 « out of sample »
Taux de « bonnes réponses » S&P500



Quantitative Market Alert

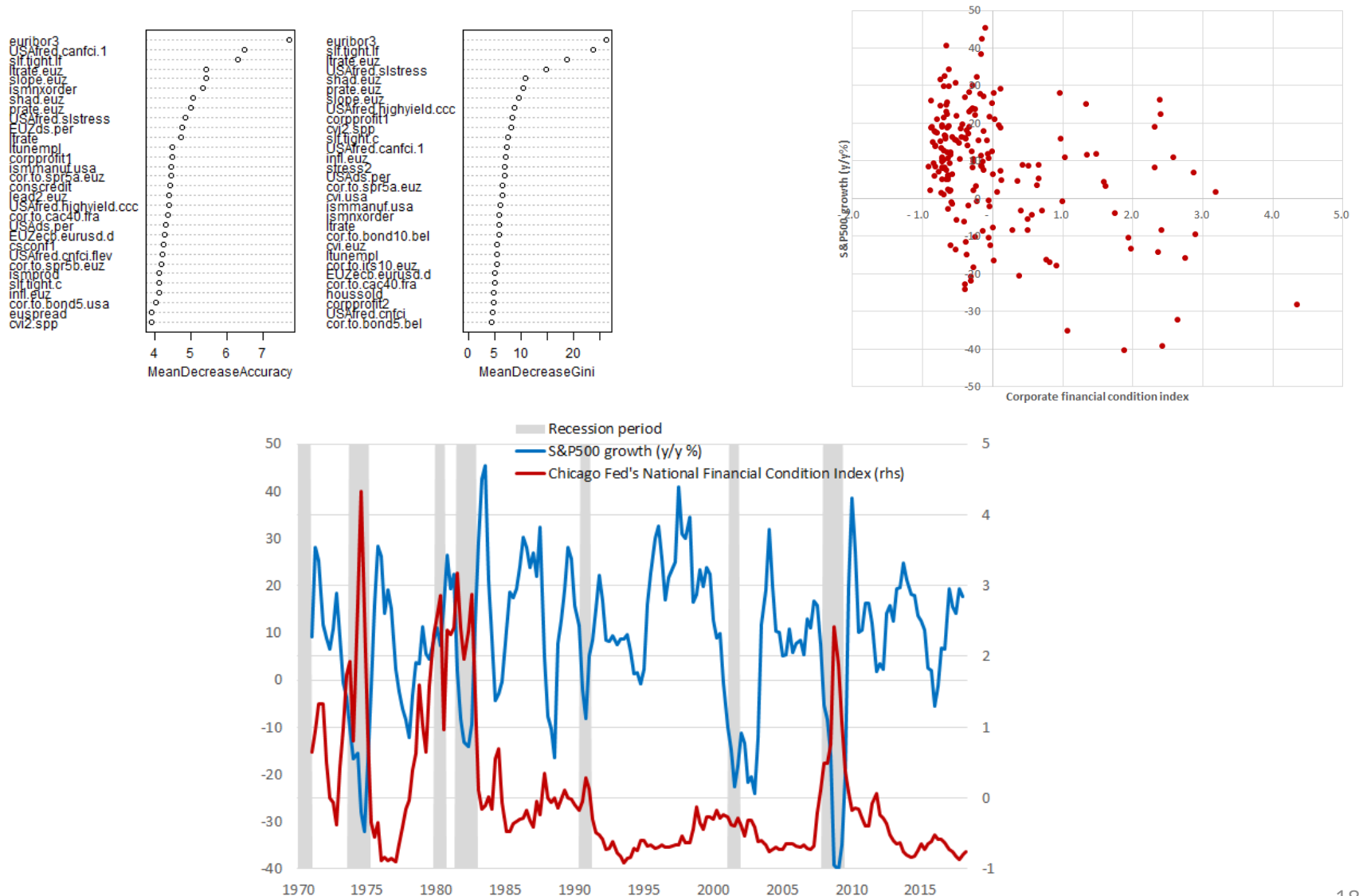
Gini index sur baisse forte du S&P500 à 1 mois

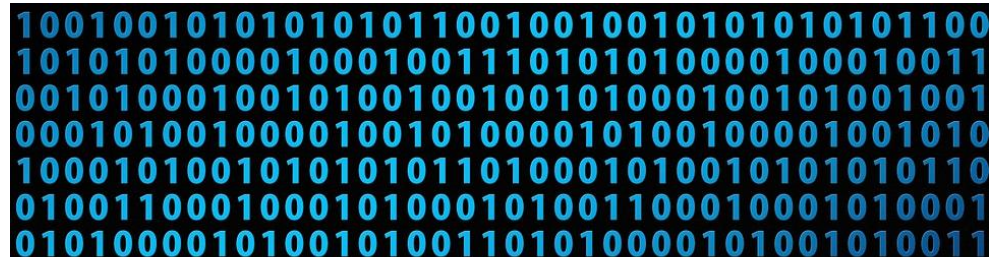
		Observé	
		1	0
Prédit	1	240	55
	0	78	67

Sensibilité	75.5%
Spécificité	54.9%
Accuracy	69.8%
PPV	81.4%
NPV	46.2%

Quantitative Market Alert


Indicateurs pertinents





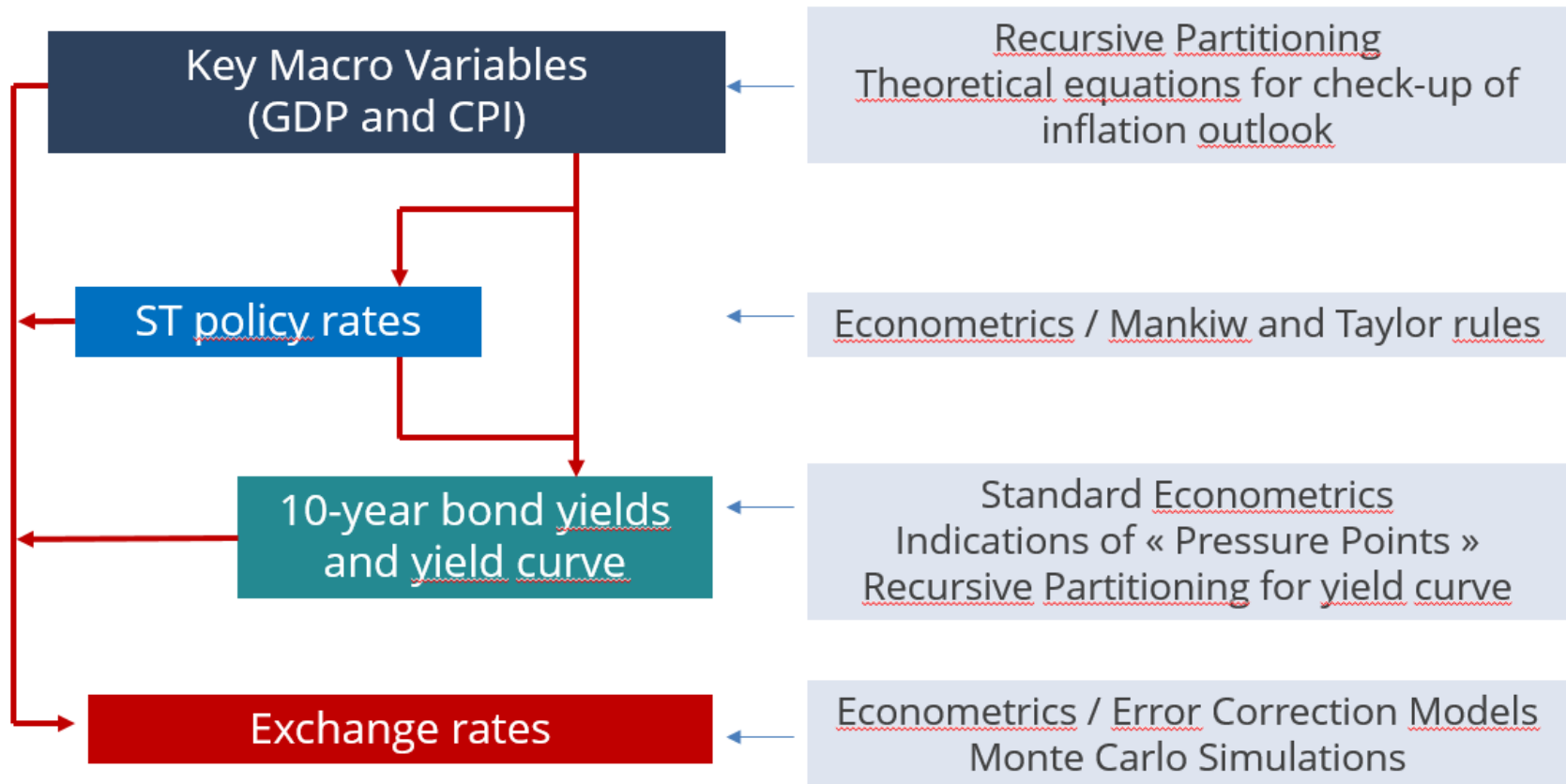
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011

Machine Learning et Conjoncture Economique



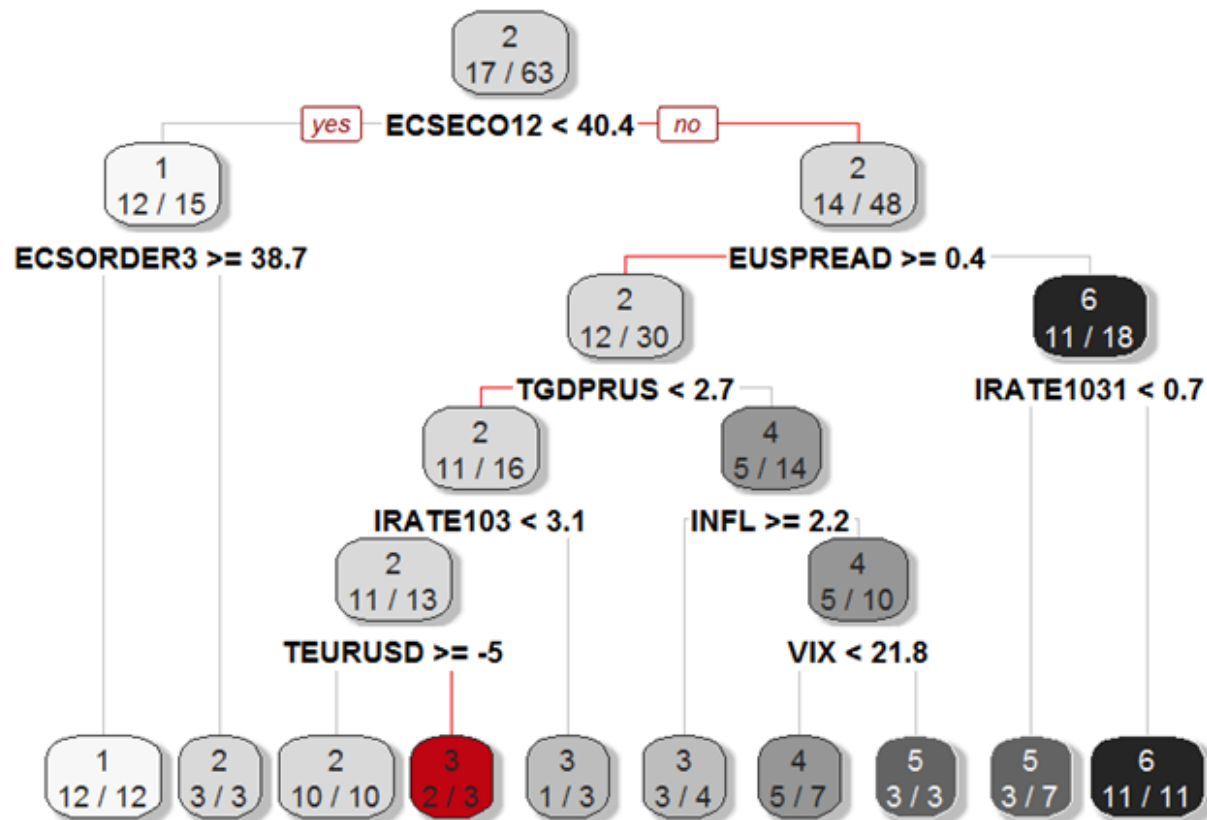
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001001

Méthodes quantitatives d'analyse conjoncturelle



CART pour prévoir la croissance économique en zone Euro

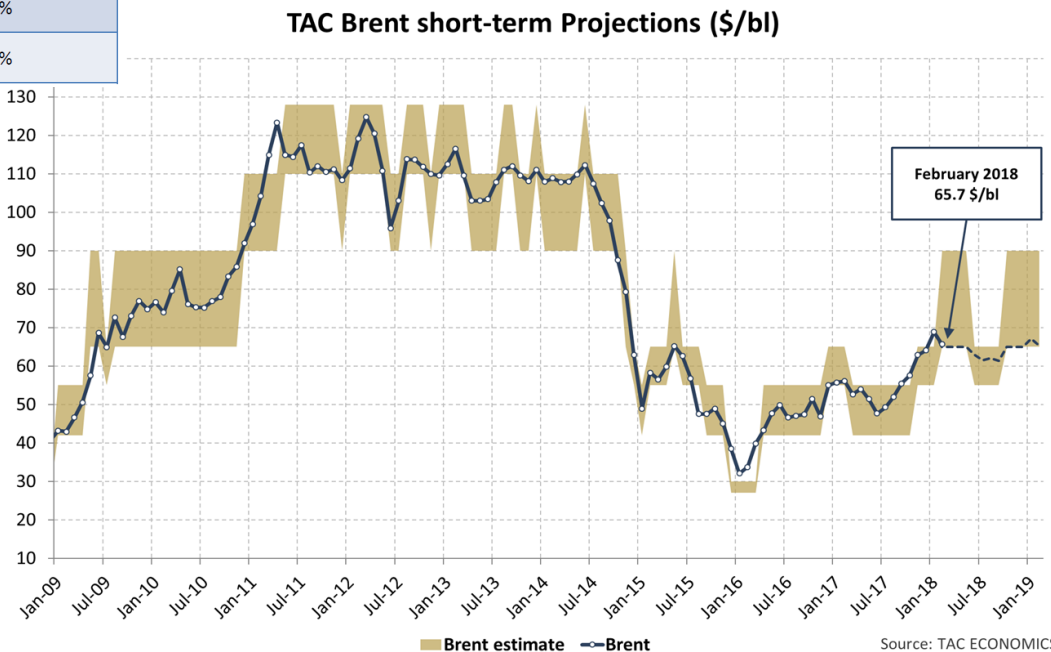
EU GDP growth - Model 2 quarters: 84.1 %



Machine Learning et Prix du pétrole (Brent)

	Average accuracy (training dataset)	Average accuracy (testing, dataset 1)
Naive Bayes	96 %	88 %
Tree Bagging	99.9 %	90 %
Gradient Boosted Machine	100 %	90 %
Supervised SOM	86 %	75 %
Neural Network multilayer perceptron	82 %	74 %
Random Forest	100 %	90 %
Support Vector Machine	96 %	88 %
k-nearest neighbors	84 %	74 %

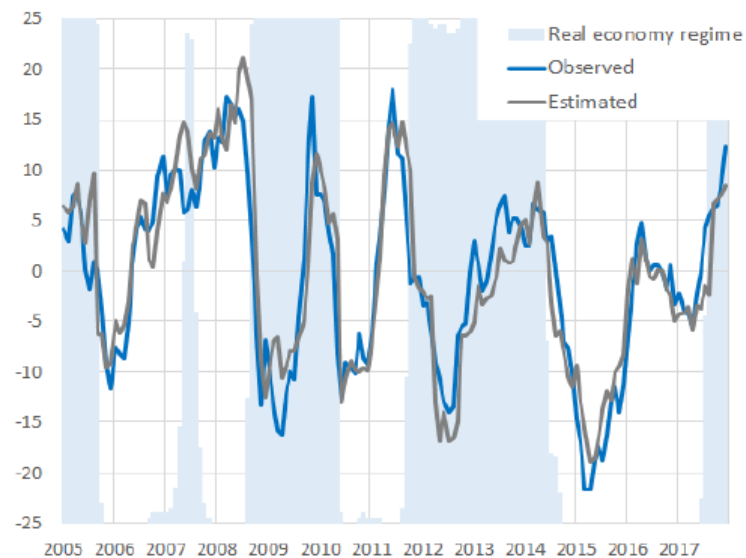
Modèle intégrant des variables macroéconomiques, d'offre et de demande sur le marché du pétrole et des variables financières



Mixte Econométrie / CART pour prévoir EUR/USD

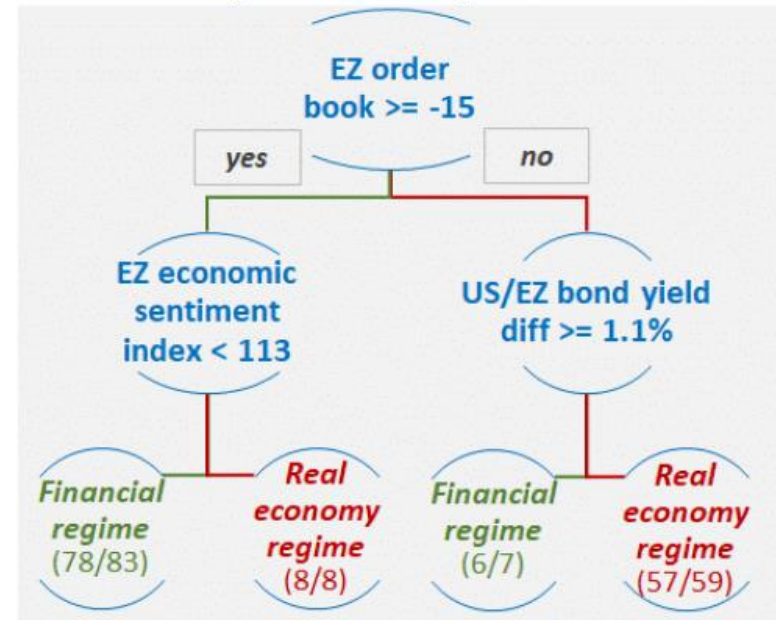
EUR/USD growth - Markov Switching model

Y/Y, in %

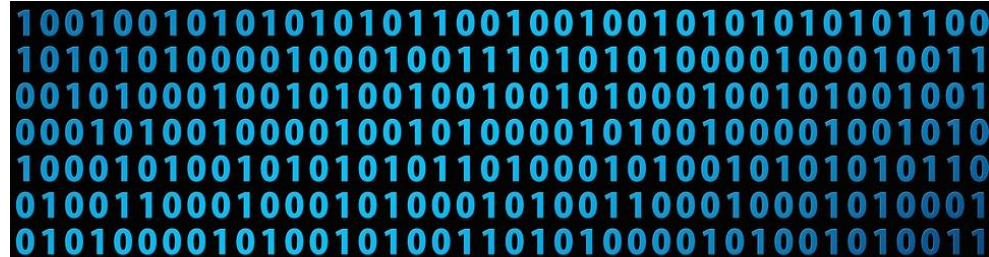


Source: TAC ECONOMICS

Factors for regime-switching- Coincident Quarter

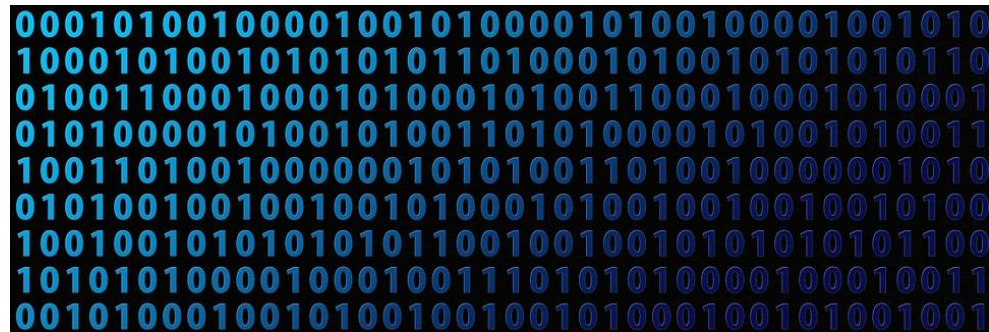


Source: TAC ECONOMICS



1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011

Text Mining sur la Presse Internationale



0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001001

Text mining

The database of news & tweets

- A list of 1000 websites, newspapers, information providers and blogs, in different languages.
- Every 3 hours our web-crawlers store the information published on these websites on internal databases (approx. 20,000 per day).
- A different set of crawlers extract and store tweets on a pre-defined list of keywords or corporates.

Sentiment Analysis by Country & Corporate

- Every day, internal algorithms estimate the sentiment using different techniques, on all “articles” and tweets.
- With the number of news and tweets, we also calculate a “buzz index” for each country, corporate or theme analysed.

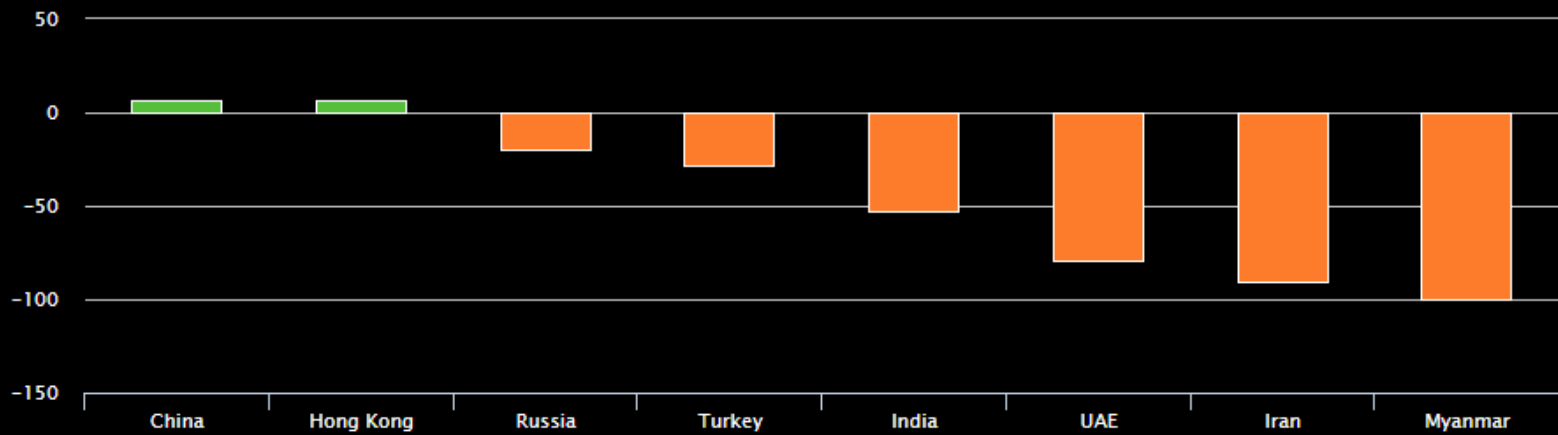
Extraction of Hot Topics in the News

- An internal algorithm to extract most important articles and topics, on a daily, weekly and monthly basis.
- The list of key topics is available on a large sample of countries and some multinational companies.
- The list can be modified to include any theme or keyword.
- Topics automatically published on our portal in real-time.

TAC Mood Index by Country

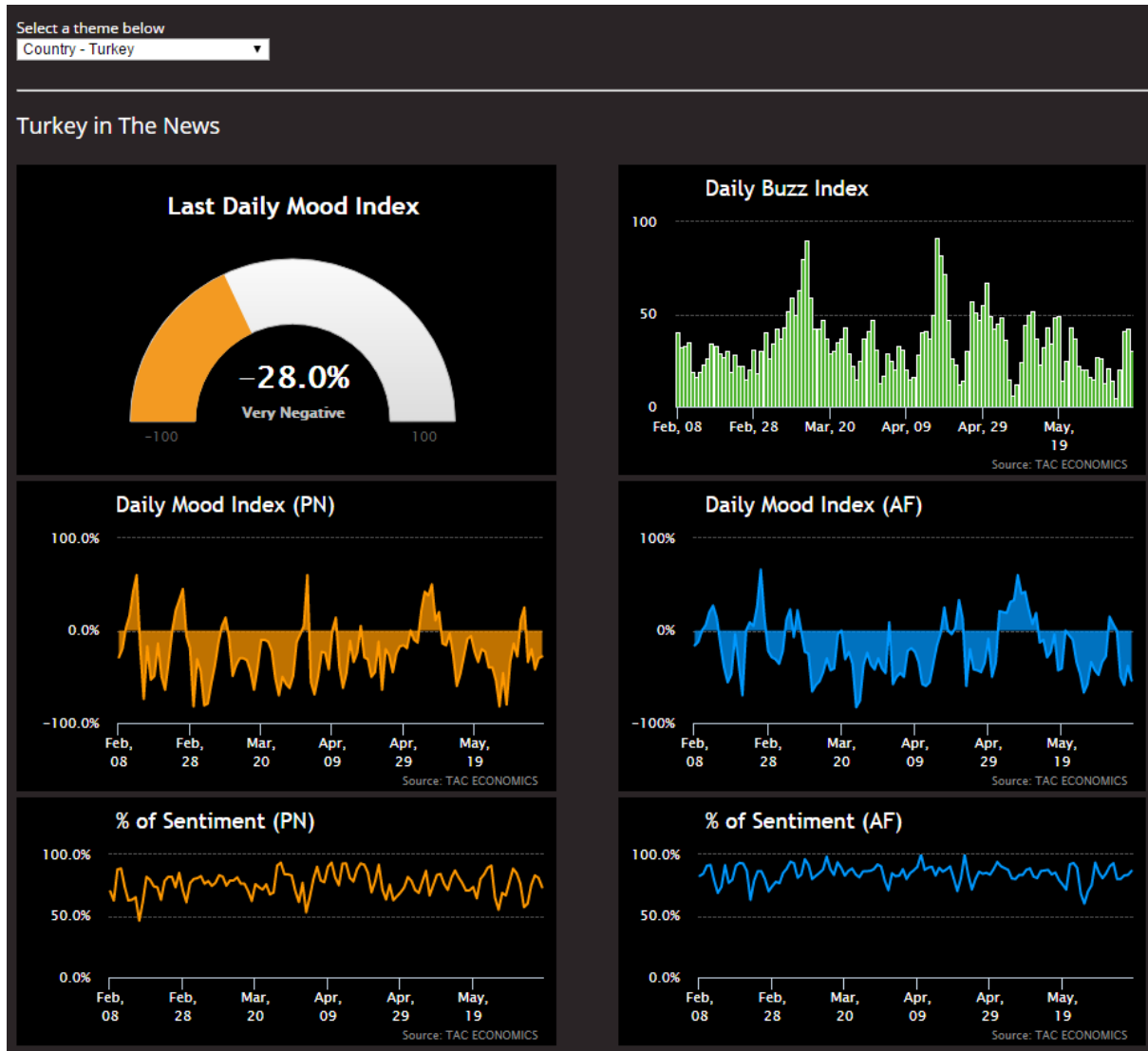
The Mood Index by country on June 08, 2017

Most Optimistic and Most Pessimistic



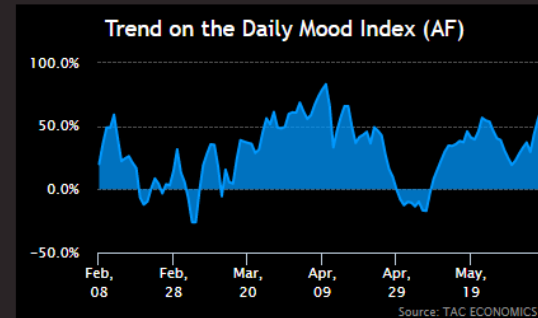
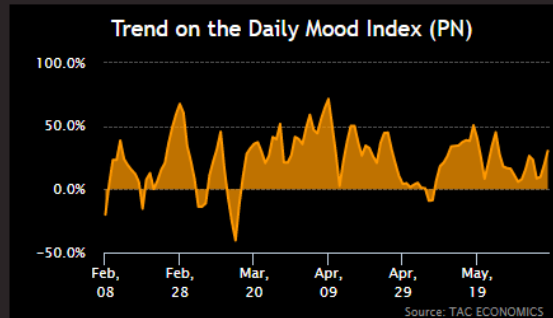
Source: TAC ECONOMICS

Sentiment Analysis: Turkey



Sentiment Analysis: Macron

Trends on Macron in The News



Monthly Hot Topics on Macron

French President Macron's Party To Win Clear Majority In Parliament: Poll

Jun, 07, REUTERS

PARIS (Reuters) - French President Emmanuel Macron's party is on course to win a commanding majority in this month's parliamentary election, an IFOP poll showed on Wednesday, reinforcing the trend seen in other surveys also pointing to a victory for Macron's camp...

The Latest: Putin, Macron Recall Czar's 1717 Paris Visit - Washington Post

May, 29, Google News

Washington PostThe Latest: Putin, Macron recall czar's 1717 Paris visitWashington PostPARIS □ The Latest on Russian President Vladimir Putin's visit to France (all times local): 7:30 p.m.: French President Emmanuel Macron and Russian President Vladimir Putin have visited an exhibit marking the ...

'We Have Disagreements' - Newly Elected French President Macron Pulls No Punches In Meeting With Putin

May, 29, World News network

French President Emmanuel Macron hosted Russia's Vladimir Putin at the sumptuous royal palace in Versailles today to talk about cooperating on fighting terror. French President Emmanuel Macron, right, and Russian President Vladimir Putin pose in the garden of the Versailles Palace following their ...

Macron And Putin Clash Over Russia's Alleged Election Meddling

May, 29, CBC

French President Emmanuel Macron welcomed Russian President Vladimir Putin at the Palace of Versailles on Monday. Past suspicions of Russian meddling in the French election resurfaced, with Macron denouncing Russian media and Putin denying hacking allegations...

Macron Explains His Tense Handshake With Trump - Fox40

May, 29, Google News

FOX40Macron Explains His Tense Handshake with TrumpFOX40WASHINGTON □ French President Emmanuel Macron confirmed that there was indeed a deeper significance to the prolonged handshake he shared with U.S. President Donald Trump in Brussels. □My handshake with him, it's not innocent,□ Macron told ...

Trump, Macron Handshake Turns Into Showdown - Cbs News

May, 25, Google News

CBS NewsTrump, Macron handshake turns into showdownCBS NewsBRUSSELS □ U.S. President Donald Trump met his match in a handshake showdown with France's new president, Emmanuel Macron. At their first meeting, ahead of a NATO summit in Brussels on Thursday, the two men locked hands for so long that

```
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100101000100101001001
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
```

Conclusion

```
0001010010000100101000010100100001001010
1000101001010101011010001010010101010110
0100110001000101000101001100010001010001
0101000010100101001101010000101001010011
1001101001000000101010011010010000001010
0101001001001001010001010010010010010100
1001001010101010110010010010101010101100
1010101000010001001110101010000100010011
0010100010010100100100100101000100101001001
```


Des mutations profondes

- Plus de données, de meilleurs modèles, plus de puissance de calcul, plus de puissance de stockage.
- L'économétrie a aussi évolué depuis 30 ans, mais ces nouvelles méthodes, outils et données changent notre métier plus profondément encore.
- Les systèmes vont continuer à s'automatiser de plus en plus dans les années qui viennent.
- Les « acteurs » de l'analyse et de la prévision ne sont plus les mêmes qu'il y a 10 ans et le « data scientist » devient roi.

Une opportunité qu'il faut savoir saisir

- Le « data scientist » informaticien, statisticien, économiste existe-t-il vraiment ?
- Plutôt que des menaces, ces évolutions sont aussi de formidables opportunités.
- Il faudra toujours des analystes pour concevoir, estimer et contrôler les systèmes... et éviter qu'un Random Forest ne prévoit une ère glaciaire car vous avez placé la sonde de température dans un frigo par erreur.
- La connaissance « métier » va (re-)devenir fondamentale. L'analyste qui sait marier sa compréhension des données, ses connaissances théoriques, ses connaissances du métier à une meilleure connaissance des principes généraux des méthodes du machine learning « moderne », aura un avantage comparatif considérable...
- ... mais il ne faut pas rater le train !

Merci de votre Attention !



Nous Contacter

Sandrine Lunven

sandrine.lunven@taceconomics.com

Commercial et Communication

taceconomics@taceconomics.com

Tel: +33 (0)2 99 39 31 40 – www.taceconomics.com