Mining Social Media to Forecast Stock Market Behavior

Paulo Cortez
ALGORITMI R&D Center
Department of Information Systems
University of Minho
http://www3.dsi.uminho.pt/pcortez
pcortez@dsi.uminho.pt

Work performed in collaboration with **Nuno Oliveira** (former Phd student) and **Nelson Areal** (Dep. Management, U. Minho).

Motivation

Analysis of stock market behavior (e.g., returns, volatility, trading volume) permits more **informed investment decisions**

Investor **sentiment** and **attention** indicators may **affect stock market** behavior

Advantages of social media indicators:

- **Direct** measure
- **Faster** and **cheaper** creation than survey based
- **Diverse frequencies** (e.g., daily, weekly) and **stocks** (e.g., individual stocks, indices)

State of the Art

	Sentiment			Attent.	Financial Analysis				Prediction		
Study	Source[a]	Meth.[b]	Comb.[c]	Source[a]	Per.[d]	Stocks[e]	Meth.[f]	Data[g]	Data[g]	St.[h]	Sur.[i]
(Solt and Statman, 1988)	S				w	Ix	MR	22y			
(Lee et al., 1991)	F				m	Pf	MR	20y			
(Neal and Wheatley, 1998)	F				m,q,a	Pf	MR	60y			
(Fisher and Statman, 2000)	S				m	Ix,Pf	MR	13y			
(Tumarkin and Whitelaw, 2001)	MB			MB	d	I	VAR	11m			
(Lee et al., 2002)	S				w	Ix	GARCH	22y			
(Antweiler and Frank, 2004)	MB	ML		MB	d	I	MR	1y			
(Brown and Cliff, 2004)	F,S		KF,Pca		m,w	Pf	VAR	33y			
(Brown and Cliff, 2005)	S				m	Pf	MR	19y			
(Das et al., 2005)	MB,N	ML		MB,N	d	I	MR	7m			
(Baker and Wurgler, 2006)	F		Pca		m	Pf	MR	38y			
(Qiu and Welch, 2006)	F,S				m,q	Pf	MR	38y			
(Schmeling, 2007)	S				w	Ix	MR	4y			
(Das and Chen, 2007)	MB	ML		MB	d	Ix,I	MR	2m			
(Tetlock, 2007)	N	GL			d	Am,Ix,Pf	VAR	15y			
(Ho and Hung, 2009)	S		Pca		m	I	MR	41y			
(Schmeling, 2009)	S				m	Am,Pf	MR	21y			
(Kurov, 2010)	F,S		Pca		d	Ix,I	MR	14y			
(Yu and Yuan, 2011)	F		Pca		m	Am	MR	42y			
(Bollen et al., 2011)	M	GL			d	Ix	NN	11m	19d		
(Deng et al., 2011)	N	GL		N	d	I	2ML,RW	32m	2y		
(Groß-Klußmann and Hautsch, 2011)	N	P			i	I	VAR	18m			
(Mao et al., 2011)	G,M,N,S	FL,K			d,w	Ix	MR	15m	30d,20w		
(Oh and Sheng, 2011)	M	ML		M	d	I	8ML	4m	10d		
(Sabherwal et al., 2011)	MB	ML		MB	d,i	I	MR	13m			
(Sheu and Wei, 2011)	F				d	Am	MR,TR	4y	59d		
(Zhang et al., 2011b)	M	K			d	Ix	Cor	7m			
(Baker et al., 2012)	F		Pca		m	Am,Pf	MR	25y			
(Schumaker et al., 2012)	N	GL			i	I	SVM	23d	23d		
(Stambaugh et al., 2012)	F,S		Pca		m	Pf	MR	42y			
(Chen and Lazer, 2013)	M	GL			d	Am	MR,TR	97d	25-33d		
(Corredor et al., 2013)	F,S		Pca		m	Pf	MR	18y			
(Garcia, 2013)	N	FL			d	Ix,Pf	MR	100y			
(Hagenau et al., 2013)	N	ML			d	I	TR	14y	12y		
(Smailović et al., 2013)	M	ML			d	I	GC	10m			
(Yu et al., 2013)	B,M,MB,N	ML		B,M,MB,N	d	I	MR	3m			
(Sprenger et al., 2014)	M	ML		M	d	I	MR	6m			
(Al Nasseri et al., 2015)	M	ML			d	Ix	TR	13m	1y	ST	
(Nguyen et al., 2015)	MB	ML,GL		MB	d	I	SVM	13m	78d		

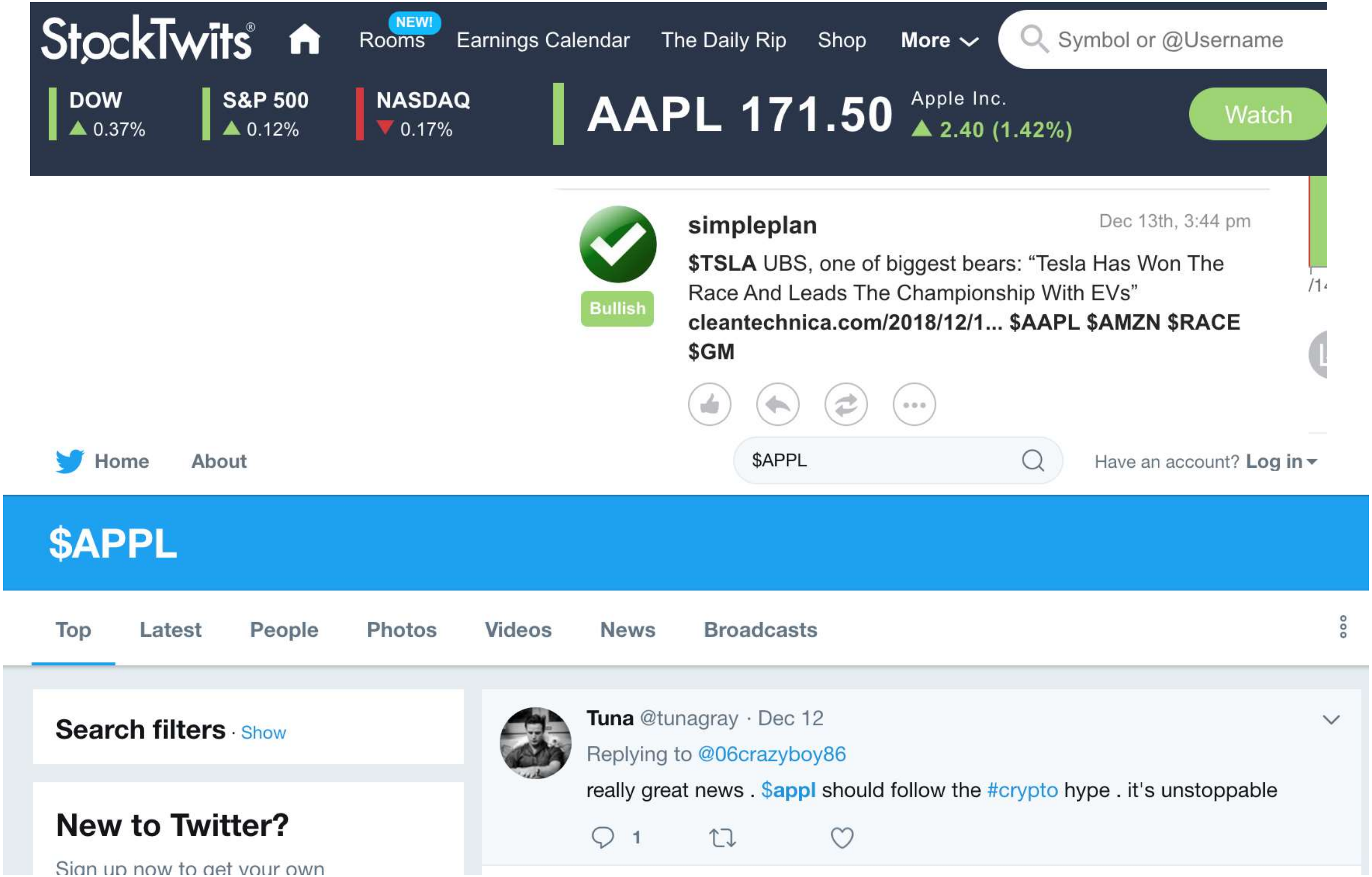
Research Objectives

Two main goals:

- Create a specialized **microblogging stock market lexicon**
- **Rigorous evaluation** of the predictive content of microblogging data for stock market behavior (e.g., diverse ML models, larger test sets, statistical test of predictive accuracy) and existing survey sentiment indices.

Microblogging stock market lexicon

Data: $APPL in StockTwits and Twitter



Methods: fast statistical measures

350,000 labeled messages

Utilization of **three adapted statistical measures**:

- Term Frequency–Inverse Document Frequency (TFIDF)
- Information Gain (IG)
- Pointwise Mutual Information (PMI)

Two **novel complementary statistics**: P_{days} and M_{assoc}

12 lexicon versions for **unique sentiment scores**

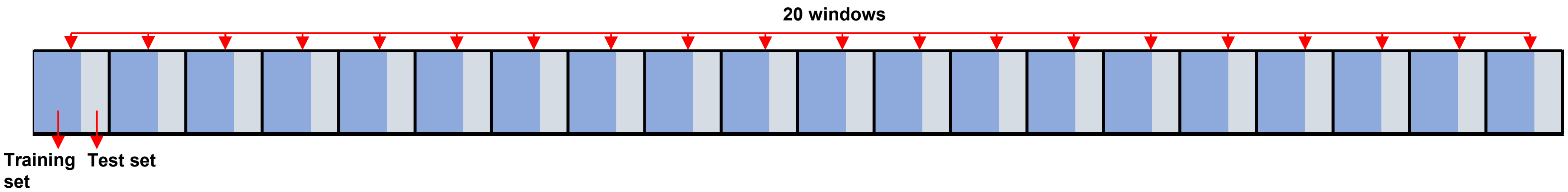
- Four versions for each adapted statistical measure (TFIDF, IG, PMI) (e.g., PMI, PMI x P_{days}, PMI x M_{assoc}, PMI x P_{days} x M_{assoc})

3 versions using sentiment scores for **affirmative** and **negated contexts**

Validation scheme 1: Holdout Split method

- Evaluation metrics such as percentage of correct classifications (CC1) and macro-averaged F-score (F_{Avg})

Validation Scheme 2: Rolling Window method



- **20 parts ordered by time**. Each window has a training set (first 2/3 posts) and a test set (last 1/3 posts)
- **Statistical significance** of CC1 and F_{Avg} improvements for pairs of lexicons: paired Student's t-test and Wilcoxon signed rank test

Classification results

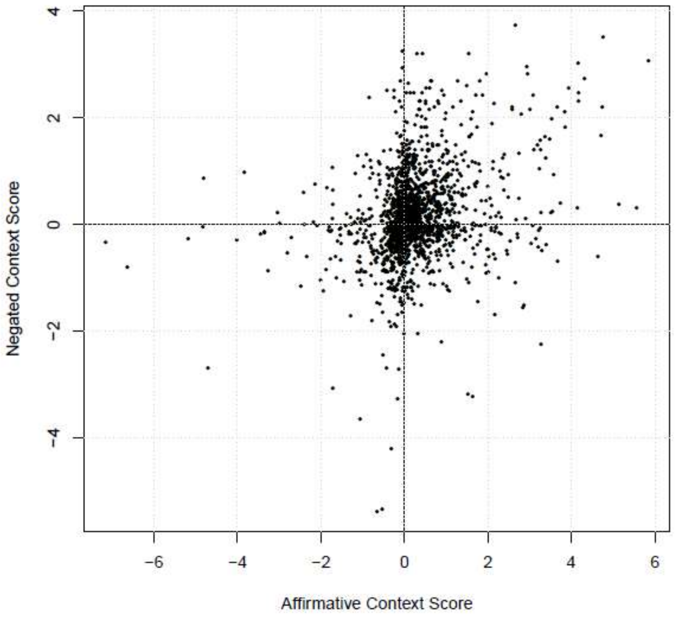
Lexicon	CC1	Unc	CC2	P_{Bull}	R_{Bull}	$F1_{Bull}$	P_{Bear}	R_{Bear}	$F1_{Bear}$	F_{Avg}
Panel A: Evaluation results of holdout split method										
PMI_{Scr}	75.2	0.5	75.5	88.6	76.5	82.1	51.9	71.5	60.1	71.1
PMI_{Assoc}	75.6	0.5	76.0	88.5	77.4	82.6	52.6	70.6	60.3	71.4
PMI_{Days}	78.8	0.5	79.1	86.0	85.4	85.7	59.5	59.6	59.6	72.6
PMI_{All}	78.8	0.5	79.2	86.0	85.5	85.8	59.7	59.5	59.6	72.7
$TFIDF_{Scr}$	74.3	0.5	74.7	88.6	75.1	81.3	50.6	71.9	59.4	70.4
$TFIDF_{Assoc}$	74.8	0.5	75.1	88.4	76.2	81.8	51.3	70.8	59.5	70.7
$TFIDF_{Days}$	78.4	0.5	78.7	85.6	85.4	85.5	58.8	58.3	58.6	72
$TFIDF_{All}$	78.5	0.5	78.8	85.5	85.5	85.5	59.1	58.1	58.6	72.1
IG_{Scr}	70.5	0.5	70.8	89.4	68.5	77.5	46.1	76.3	57.4	67.5
IG_{Assoc}	71.6	0.5	71.9	89.5	70.1	78.6	47.3	75.9	58.3	68.4
IG_{Days}	76.0	0.5	76.4	87.2	79.5	83.2	53.4	65.9	59.0	71.1
IG_{All}	76.4	0.5	76.7	86.9	80.4	83.5	54.1	64.9	59.0	71.3
Panel B: Average evaluation results of rolling window method										
PMI_{Scr}	71.1	0.5	71.4	90.0	69.0	78.0	45.7	77.0	56.9	67.4
PMI_{Assoc}	71.5	0.5	71.9	89.9	69.8	78.4	46.2	76.7	57.3	67.9
PMI_{Days}	77.3	0.5	77.7	87.4	81.5	84.3	54.1	64.4	58.5	71.4
PMI_{All}	77.5	0.5	77.8	87.3	81.7	84.4	54.6	64.2	58.7	71.5
$TFIDF_{Scr}$	71.6	0.5	71.9	89.1	70.8	78.8	45.6	73.4	55.8	67.3
$TFIDF_{Assoc}$	72.1	0.5	72.5	89.0	71.7	79.3	46.3	73.1	56.3	67.8
$TFIDF_{Days}$	77.1	0.5	77.5	86.4	82.4	84.3	54.0	60.5	56.6	70.5
$TFIDF_{All}$	77.3	0.5	77.7	86.5	82.7	84.5	54.5	60.4	56.9	70.7
IG_{Scr}	67.4	0.5	67.7	90.5	63.8	74.2	42.7	78.7	54.3	64.2
IG_{Assoc}	68.0	0.5	68.3	90.5	64.7	74.9	43.2	78.5	54.6	64.8
IG_{Days}	75.2	0.5	75.5	88.4	77.4	82.3	50.6	68.6	57.3	69.8
IG_{All}	75.5	0.5	75.8	88.4	77.8	82.6	51.0	68.5	57.6	70.1

Comparison with six reference lexicons (improved results by a large margin):

Lexicon	CC1	Unc	CC2	P_{Bull}	R_{Bull}	$F1_{Bull}$	P_{Bear}	R_{Bear}	$F1_{Bear}$	F_{Avg}
Panel A: Evaluation results of holdout split method										
PMI_{All}	78.8	0.5	79.2	86.0	85.5	85.8	59.7	59.5	59.6	72.7
FIN	16.8	66.0	49.3	83.5	13.9	23.8	34.3	25.1	29.0	26.4
GI	37.7	25.8	50.8	82.5	36.2	50.3	37.5	42.1	39.7	45.0
MSOL	53.4	1.8	54.3	79.1	58.6	67.3	33.9	38.2	35.9	51.6
MPQA	36.9	37.5	59.0	80.6	40.6	54.0	34.3	26.3	29.8	41.9
OL	31.8	43.0	55.9	82.6	32.7	46.8	37.7	29.4	33.1	39.9
SWN	57.4	5.1	60.5	79.9	59.7	68.3	34.1	50.7	40.7	54.5
Panel B: Average evaluation results of rolling window method										
PMI_{All}	77.5	0.5	77.8	87.3	81.7	84.4	54.6	64.2	58.7	71.5
FIN	17.3	63.7	47.8	84.2	13.9	23.8	32.8	27.4	29.3	26.5
GI	37.4	25.8	50.4	82.6	36.1	50.2	35.8	41.1	37.8	44.0
MSOL	52.3	1.2	53.0	80.2	55.8	65.6	32.5	41.9	36.1	50.8
MPQA	39.1	34.7	59.8	81.3	42.6	55.8	35.2	28.4	31.1	43.5
OL	34.1	39.5	56.3	83.5	34.6	48.9	38.2	32.3	34.7	41.8
SWN	58.9	4.4	61.6	80.6	61.4	69.7	33.8	51.1	40.4	55.0

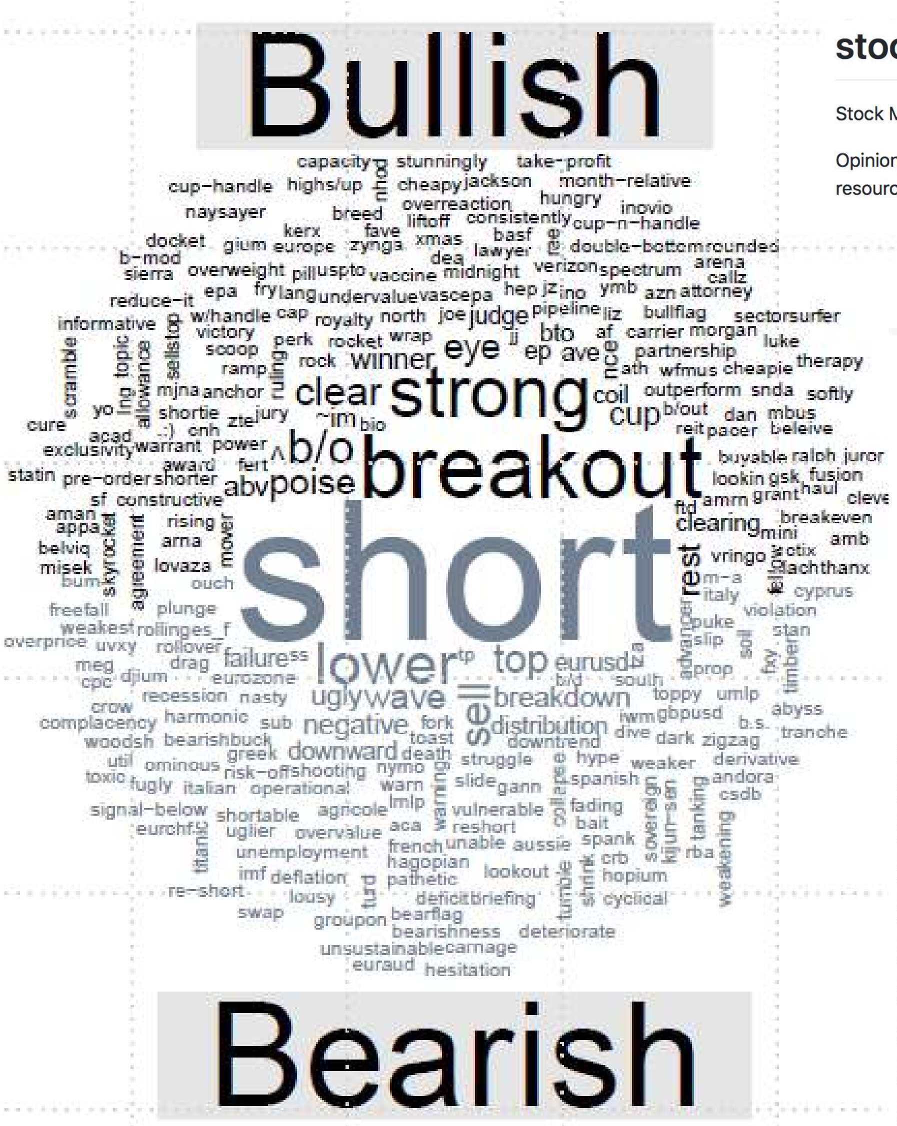
Using separate negative and positive contexts improved the results:

Lexicon	CC1	Unc	CC2	P_{Bull}	R_{Bull}	$F1_{Bull}$	P_{Bear}	R_{Bear}	$F1_{Bear}$	F_{Avg}
Panel A: Evaluation results of holdout split method										
PMI_{All}	78.8	0.5	79.2	86.0	85.5	85.8	59.7	59.5	59.6	72.7
$TFIDF_{All}$	78.5	0.5	78.8	85.5	85.5	85.5	59.1	58.1	58.6	72.1
IG_{All}	76.4	0.5	76.7	86.9	80.4	83.5	54.1	64.9	59.0	71.3
PMI_{BiScr}	79.0	0.5	79.3	86.2	85.4	85.8	59.8	60.3	60.1	73.0
$TFIDF_{BiScr}$	78.5	0.5	78.9	86.0	85.1	85.5	59.0	59.6	59.3	72.4
IG_{BiScr}	76.7	0.5	77.0	87.0	80.8	83.8	54.7	64.8	59.3	71.5
Panel B: Average evaluation results of rolling window method										
PMI_{All}	77.5	0.5	77.8	87.3	81.7	84.4	54.6	64.2	58.7	71.5
$TFIDF_{All}$	77.3	0.5	77.7	86.5	82.7	84.5	54.5	60.4	56.9	70.7
IG_{All}	75.5	0.5	75.8	88.4	77.8	82.6	51.0	68.5	57.6	70.1
PMI_{BiScr}	78.3	0.5	78.7	86.4	84.2	85.2	56.8	60.0	58.1	71.7
$TFIDF_{BiScr}$	77.3	0.5	77.7	86.5	82.6	84.4	54.4	60.7	57.0	70.7
IG_{BiScr}	75.6	0.5	76.0	88.6	77.8	82.7	51.2	69.2	58.0	70.3



Generated lexicon: 1-gram and bi-grams, 20,551 terms https://github.com/nunomroliveira/stock_market_lexicon

Wordcloud:



stock_market_lexicon

Stock Market Lexicon

Opinion Lexicon adapted to stock market conversations in microblogging services (e.g., StockTwits, Twitter). This lexical resource was automatically created using diverse statistical measures and a large set of labeled messages from StockTwits.

```
"Item","POS","Aff_Score","Neg_Score"
"'em","PR",0.379542372881356,0.533419354838710
"'n","CC",1.41317647058824,1.19991666666667
"'n handle","",2.839,2.941
"'s a","",-0.0230769230769231,-0.0136153846153846
"'s abc","",2.26166666666667,2.3
"'s act","",2.403,2.443
"'s again","",-0.156285714285714,-0.0709333333333334
...
"zone num","",-1.19436842105263,-1.17121052631579
"zone on","",-0.8425,-0.822
"zone tkr","",-1.692625,-1.665125
"zone to","",-0.661,-0.641
"zoom","NN",-0.1463,-0.1386
"zoom","VB",-0.652272727272728,-0.649173913043478
"zortrades.com","NN",2.141,2.163
"zte","NN",4.934,5.084
"zuck","NN",-0.237857142857143,-0.185172413793104
"zuckerberg","NN",-0.309666666666666,-0.236923076923077
```

Conclusions

Proposed **automatic procedure to create lexicons**:

Produced lexicons that substantially **outperform six reference lexicons**

Novel complementary metrics proved to be **relevant**

Usage of **affirmative** and **negated** context sentiment values was **useful**

Main Publication

Nuno Oliveira, Paulo Cortez, and Nelson Areal. Stock market sentiment lexicon acquisition using microblogging data and statistical measures. ***Decision Support Systems***, 85:62–73, 2016
(JCR 2015 Q1 in ”Computer Science, Artificial Intelligence” and Q1 in ”Computer Science, Information Systems”; 35 Google Scholar citations)

ELSEVIER

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Stock market sentiment lexicon acquisition using microblogging data and statistical measures

CrossMark

Nuno Oliveira [a,*], Paulo Cortez [a], Nelson Areal [b]

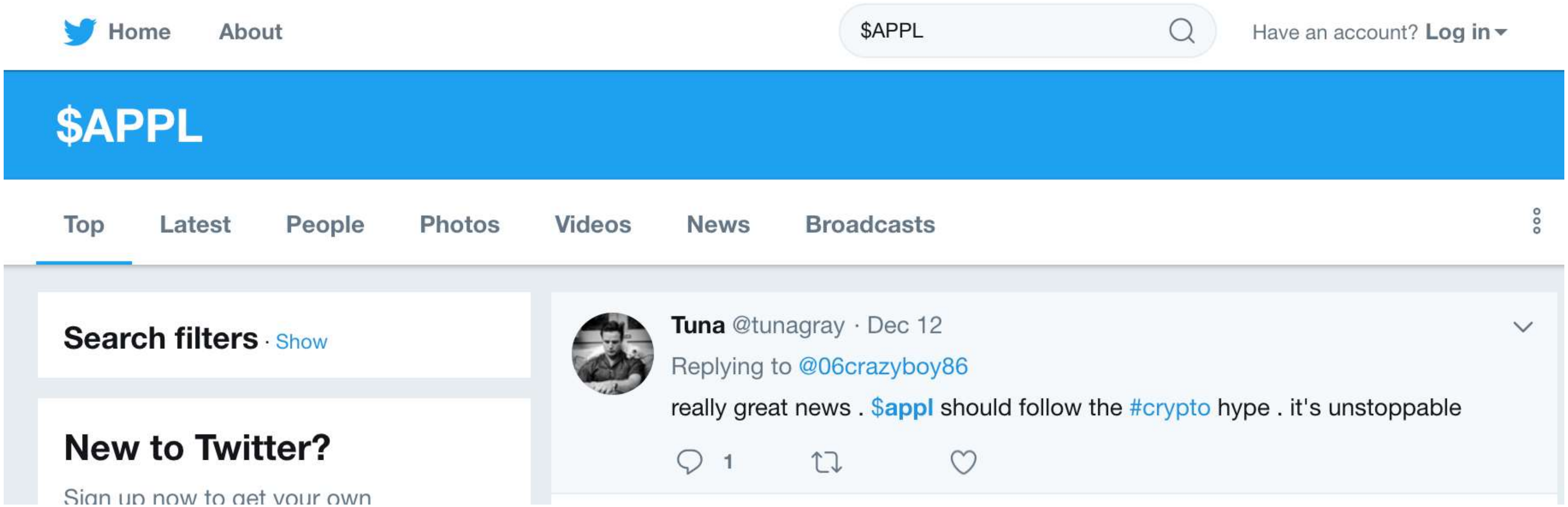
[a] ALGORITMI Centre, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal
[b] School of Economics and Management, Department of Management, University of Minho, 4710-057 Braga, Portugal

Acknowledgements

We would like to thank StockTwits for the provision of their data.

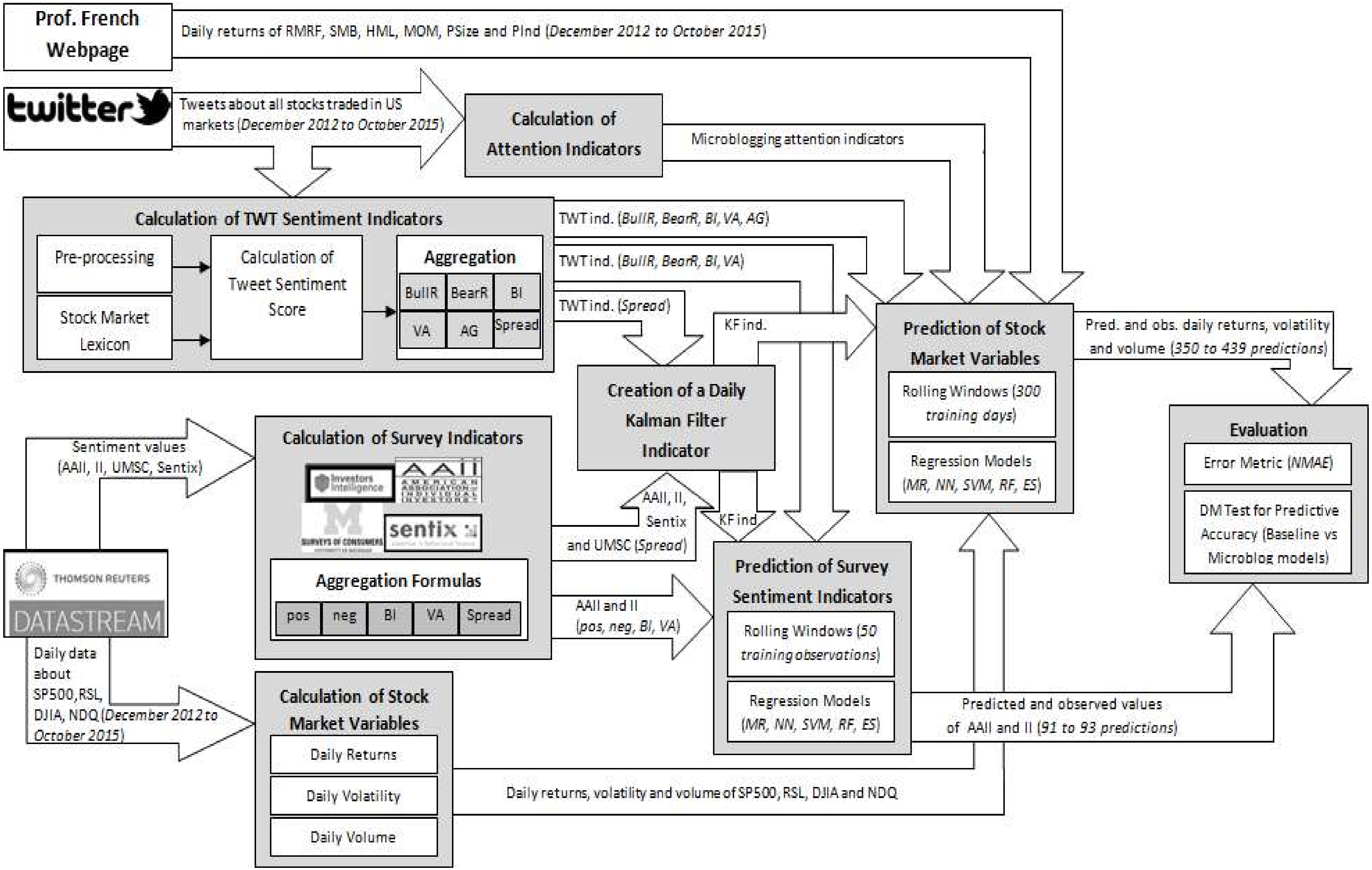
Impact of microblogging data for stock market prediction

Microblog Data: Twitter



We collected around 31 million tweets from 22nd of December 2012 to 29th of October 2015 holding cashtags of all stocks traded in US markets (nearly 3800 stocks).

Methodology



Objectives:

• Application and comparison of **diverse Machine Learning methods:** Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF), Ensemble Averaging (EA)

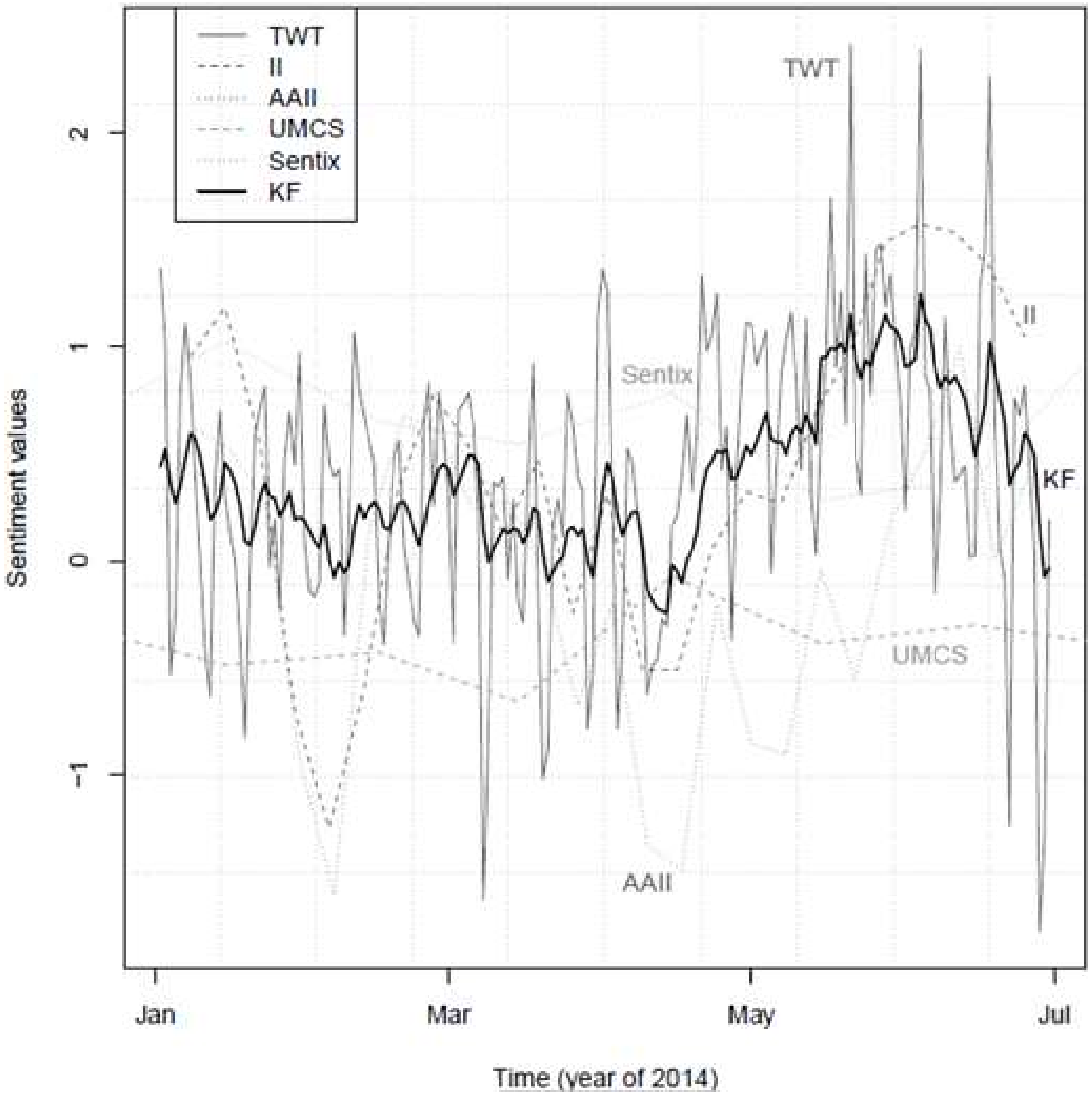
• Application of a **Kalman Filter** (KF) procedure to **aggregate diverse sentiment indicators**

• **Prediction** of daily variables (returns, volatility and trading volume) of **diverse indices and portfolios**

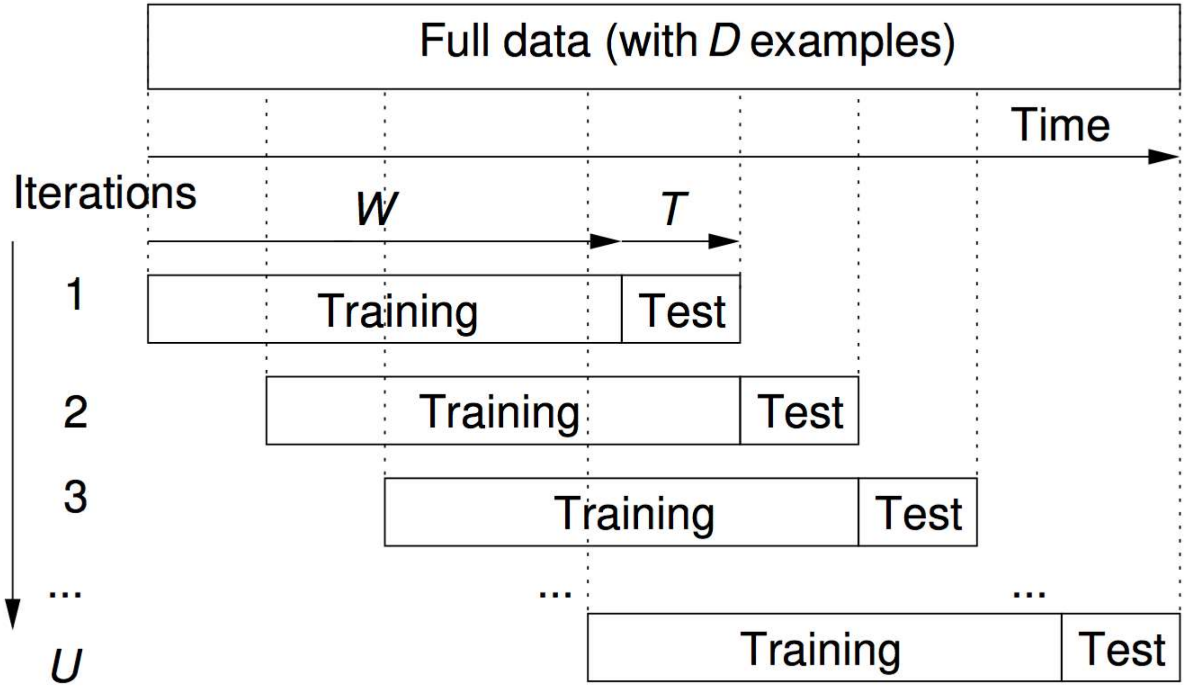
• **Forecasting of survey** sentiment: American Association of Individual Investors (AAII) and Investors Intelligence (II)

• Utilization of **microblogging stock market lexicon**

Daily sentiment indicator created by Kalman Filter procedure



Evaluation: rolling window and NMAE metric



$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

$$NMAE = \frac{MAE}{y_H - y_L}$$

Results: Prediction of **returns** of **indices**

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n predictions: 414; returns range: 7.53)	MR: 8.12	SVM MRet2 (KF): 7.98*	SVM MRet3 (BullR): 8.01* SVM MRet2 (KF): 7.98*
HML (n predictions: 392; returns range: 3.36)	SVM: 10.29	SVM MRet3 (BullR): 10.24	
MOM (n predictions: 392; returns range: 4.63)	SVM: 10.78	SVM MRet2 (KF): 10.69*	SVM MRet2 (KF): 10.69*
NDQ (n predictions: 439; returns range: 9.35)	SVM: 7.61	SVM MRet7: 7.58	
RMRF (n predictions: 392; returns range: 7.58)	SVM: 8.27	SVM MRet3 (KF): 8.19*	SVM MRet3 (KF): 8.19*
RSL (n predictions: 439; returns range: 7.02)	EA: 11.02	EA MRet1: 11.02	
SMB (n predictions: 392; returns range: 3.36)	MR: 12.44 SVM: 12.44	SVM MRet4 (BullR): 12.27*	SVM MRet4 (BullR): 12.27*
SP500 (n predictions: 439; returns range: 7.85)	SVM: 7.87	SVM MRet3 (KF): 7.79**	SVM MRet7: 7.80** SVM MRet6: 7.81* SVM MRet4 (VA): 7.81* SVM MRet5 (VA): 7.81* SVM MRet3 (KF): 7.79**

Summary of the results of the prediction of returns

- **Microblogging sentiment and attention** indicators had **predictive value** for **SP500, portfolios of smaller market capitalization** (Lo20 and Lo30) and some sectors (**High Technology, Energy and Telecommunications**)

- **KF indicators** were important for **SP500, Lo20, High Technology** and **Energy** industries

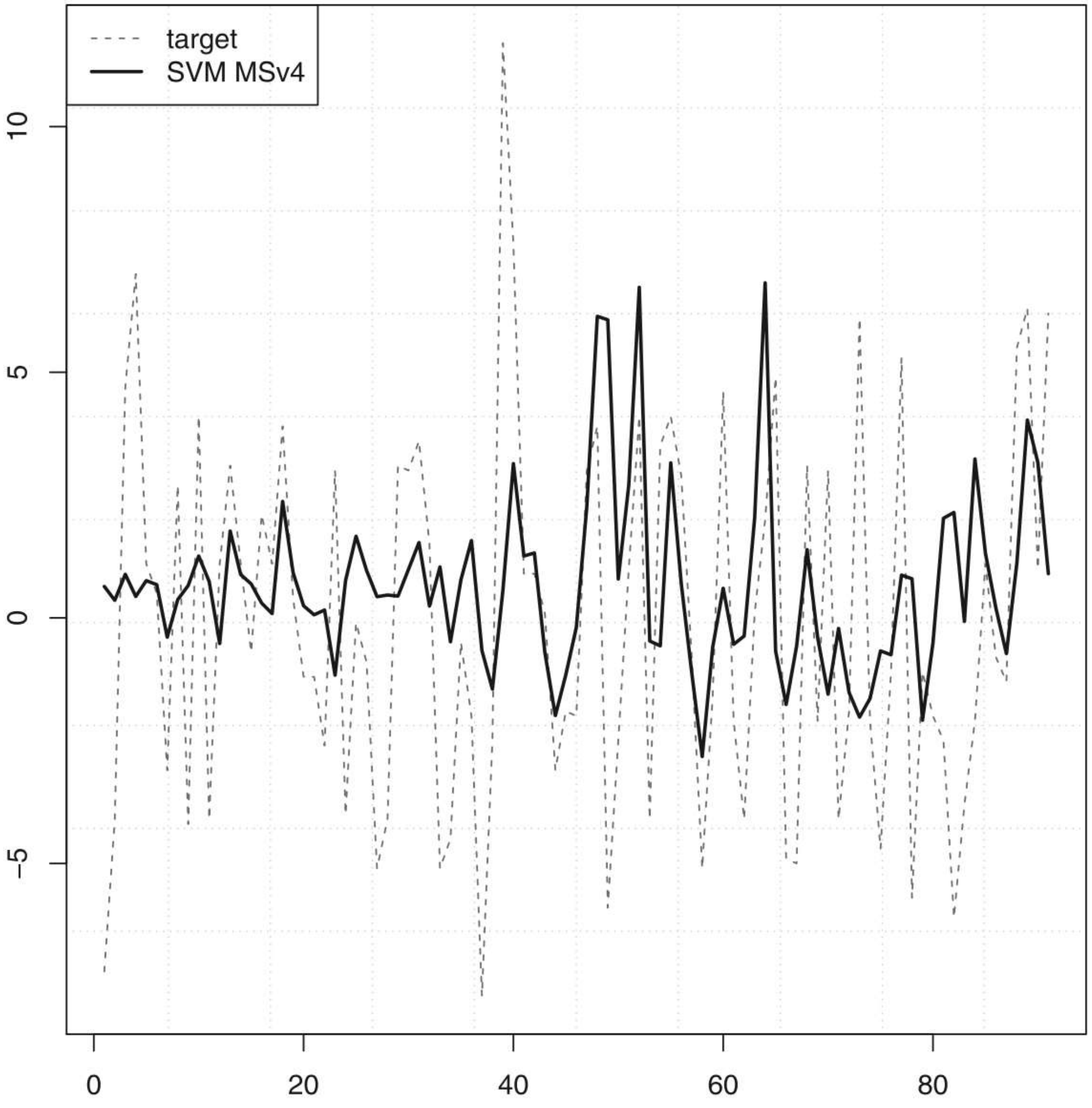
Prediction of volatility

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n predictions: 413; realized volatility range: 92.41)	MR: 2.91	SVM MVIt3 (AG): 2.79	MR MVIt3 (KF): 2.85*
NDQ (n predictions: 413; realized volatility range: 57.18)	MR: 3.89	SVM MVIt3 (VA): 3.87	
RSL (n predictions: 412; realized volatility range: 39.56)	MR: 5.71	MR MVIt1: 5.71	
SP500 (n predictions: 413; realized volatility range: 67.90)	EA: 3.34	EA MVIt1: 3.34 EA MVIt3 (AG): 3.34	
VIX (n predictions: 413; VIX range: 30.42)	EA: 3.26	SVM MVIt3 (BullR): 3.25	

Prediction of trading volume

Index	Baseline	Lowest NMAE	Statistical significant results
DJIA (n predictions: 414; volume range: 310804)	SVM: 6.00	SVM MVIt5 (BullR): 5.84*	SVM MVIt5 (BullR): 5.84* SVM MVIt5 (BI): 5.85*
SP500 (n predictions: 413; volume range: 1636036)	SVM: 4.98	SVM MVIt1: 4.98	

Prediction of II index using KF



Conclusions

Proposed **automatic procedure to create lexicons**:

- Produced lexicons that substantially **outperform six reference lexicons**
- **Novel complementary metrics** proved to be **relevant**
- Usage of **affirmative** and **negated** context sentiment values was **useful**

Robust evaluation of the predictive value of Twitter data (e.g., larger test sets, statistical test for predictive accuracy, application and comparison of diverse ML methods)

Microblogging indicators were particularly **informative** for:

- Prediction of **returns** of **SP500, portfolios of lower dimension** and some sectors (**High Technology, Energy and Telecommunications**)

Conclusions

KF indicators were informative for the prediction of **returns of some portfolios and indices**

Prediction of survey values:

- **Twitter sentiment indicators** were informative to **predict negative values of AAII**
- **KF indicators** were informative for the **prediction of II computed by VA formula**

Advantages of microblogging sentiment indicators:

- **Direct** sentiment measure
- **Faster** and **cheaper** creation than survey based
- **Personalized frequency** (e.g., daily) and **targets** (e.g., stocks, indices)

Main Publication

Nuno Oliveira, Paulo Cortez, and Nelson Areal. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. ***Expert Systems with Applications***, 73:125–144, 2017
(JCR 2015 Q1 in ”Computer Science, Artificial Intelligence”; 40 Google Scholar citations)

ELSEVIER

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices

CrossMark

Nuno Oliveira[a,*], Paulo Cortez[a], Nelson Areal[b]

[a] *ALGORITMI Centre, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal*
[b] *School of Economics and Management, Department of Management, University of Minho, 4710-057 Braga, Portugal*

Contacts

Nuno Oliveira: nunomroliveira@gmail.com

Paulo Cortez: pcortez@dsi.uminho.pt

Nelson Areal: nareal@eeg.uminho.pt