

Análise de Sentimentos: Abordagem Lexical de classificação de opinião no contexto mercado financeiro brasileiro



Vitor Miguel Xavier Peres
Mestrando - PPGCC
PUCRS - Escola Politécnica
Porto Alegre, Brasil

Dr. Renata Vieira
PPGCC
PUCRS - Escola Politécnica
Porto Alegre, Brasil

Dr. Rafael Bordini
PPGCC
PUCRS - Escola Politécnica
Porto Alegre, Brasil



Agenda

- Análise de Sentimentos
 - Estado da Arte
 - Contexto e Dados
 - Modelo de Análise de Sentimentos
 - Construção de Léxico
 - Experimentos e Resultados
- Considerações Finais
- Continuidade da Pesquisa
- Referências

Análise de Sentimentos



Business Plan

EXPERIENCE

POSITION TITLE for company tld
Present
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2013 - 2016
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2012 - 2013
Short description of the position and the responsibilities you had in this position.
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Inusami qui spe volur new.

POSITION TITLE for company tld
2003 - 2010
Short description of the position and the responsibilities you had in this position.

ADDRESS
125 Name Street,
Town / City,
State / Country,
Postal / ZIP code

PHONE
0028 01234 5678

EMAIL
info@sambblaq.com

WEBSITE
www.mypage.com

SKYPE
skype: sambqak

HOBBIES
creating websites
swimming
photography
body building

PROFESSIONAL STATEMENT
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse suscipit efficitur lectur. Fusce iaculis, leo nec vulputate efficitur. Lorem interdum elit, ut vestibulum nid met

REFERENCES

ELIOT BROWN
0028 01234 5678
eliot@mypage.com



Estado da Arte

<p>Por definição de Liu [4], a análise de sentimentos dedica-se ao tratamento de opiniões ou sentimentos que foram expressados em textos, e que através de um léxico de sentimento, que corresponde à uma orientação semântica ou a polaridade de palavras, assinalam quais sentidos textos se definem.</p>	<p>Mizumoto [5] propõe um estudo para determinar a polaridade de sentimentos em textos de notícias sobre mercado financeiro, utilizando dicionários léxicos. Além da construção de um dicionário léxico, fazendo o uso de aprendizado semi-supervisionado, anotando manualmente a polaridade de um determinado número de termos/palavras dos textos destas notícias.</p>	<p>San [6] também utiliza a abordagem léxica para a análise de sentimentos em notícias do mercado financeiro, sendo o mesmo feito sobre dois conjuntos de experimentos, que utilizam, ou não, o processo de stemming.</p>	<p>Palanisamy [7] através de léxicos, procuram-se identificar qual tipo de sentimento de tweets que contenham variações de palavras, emoticons e hashtags. Em [8], o autor propõe o emprego de coesão léxica baseada em textos, na detecção de sentimentos e polaridade de textos.</p>

Contexto e Dados

01



TradingView

Ferramenta gráfica com o objetivo de compartilhar e visualizar possíveis operações financeiras.

02

InfoMoney

InfoMoney

É o maior site especializado em investimentos pessoais e educação financeira do Brasil.

03



Investing.com

Portal Financeiro global de dados, que fornecem notícias, analyses, cotações, gráficos das principais ações.

ANÁLISE TÉCNICA

Fechamento do Mercado – 29/06/2016

As bolsas ao redor do mundo sobem pelo segundo pregão consecutivo após um forte movimento negativo nos últimos pregões, o Ibovespa fechou com alta de 1,99% cotado a 51.002.

Análise técnica na prática



Índice Bovespa Futuro – Tendência Lateral 83.300 / 90.375

Por Alexandre Ragazzino - 13 horas atrás

Ainda acompanhamos no gráfico intradiário de 01 hora o fato de os preços estarem dentro da Tendência Lateral entre 83.300 e 90.375. A dúvida é se os...



Recorde do Alibaba no Dia dos Solteiros Pode Não Reverter a Queda das suas Ações

Por Haris Anwar/Investing.com - 16 horas atrás

O Alibaba (NYSE:BABA), gigante chinês do comércio eletrônico, mais uma vez deu um grande espetáculo no seu tão aguardado festival de compras do "Dia dos...



Com Puxada no Finalzinho, IBOV Respeita Topo Anterior

Por Thiago Bisi - 09.11.2018

O dia foi predominante de queda após um repique logo após a abertura. Mas uma puxada forte nos momentos finais salvou o índice de fechar negativo. Alta de 0,02%. Temos um candle...

PETR4 confirma movimento baixista e perde os 26 reais

PETR4, 1D VIÉS DE BAIXA



PCTEDESCO TOP

153 visualizações · Nov 8

PETR4 confirma seu movimento de correção baixista e rompe o marco dos 26 reais. Já opera abaixo da MME20, as médias móveis (MME5 e MME20) tem viés baixista, bem como o HiloActivator. Testa agora região de suporte, um fundo anterior na faixa d...



1

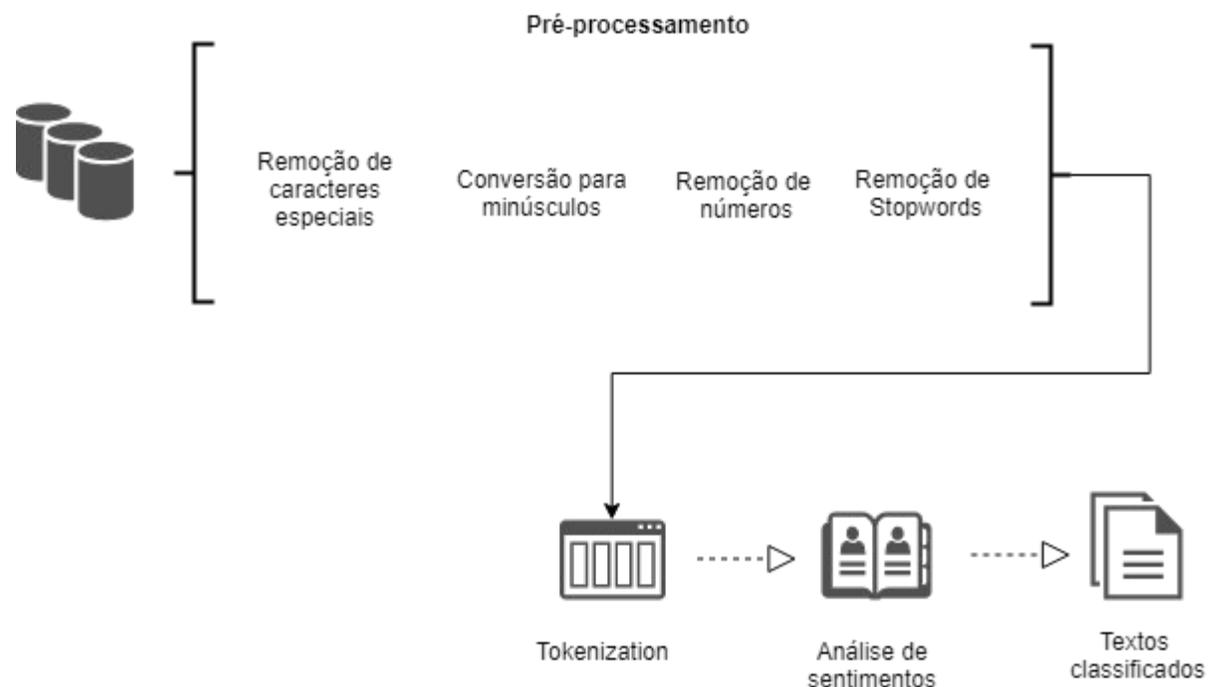


0

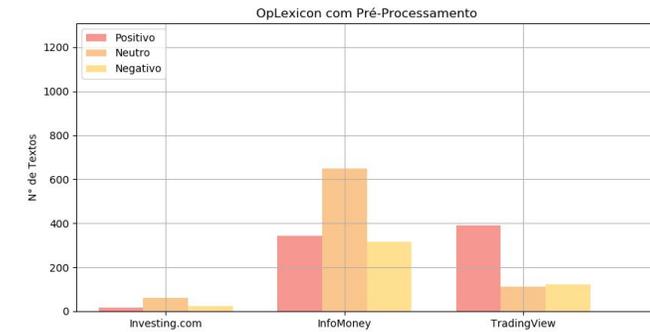
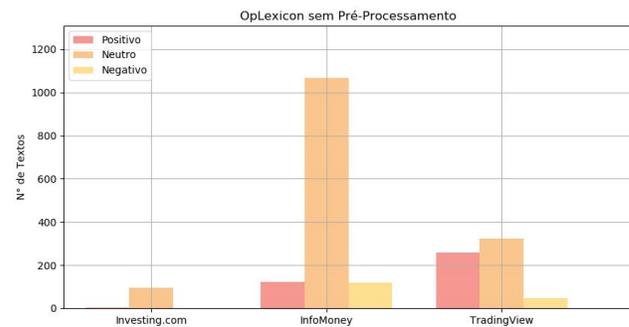


Modelo de Análise

Para uma melhor performance do algoritmo criado, houve a necessidade de realizar a limpeza dos dados, que possibilita através de abordagem de dicionário léxico percorrer as sentenças de maneira simples. Esse processo tem sido aplicado a textos que não possuem a estrutura formal, no caso textos de Twitter, e contém uma série significativa de ruídos [1].



Pré-processamento: Sim ou Não?



Resultados da Análise

	SentiLex			OpLexicon			UniLex		
	POS	NEU	NEG	POS	NEU	NEG	POS	NEU	NEG
InfoMoney	174	999	135	343	649	316	531	247	527
Investing.com	5	86	9	17	60	23	40	27	32
TradingView	379	157	90	390	112	124	388	49	189

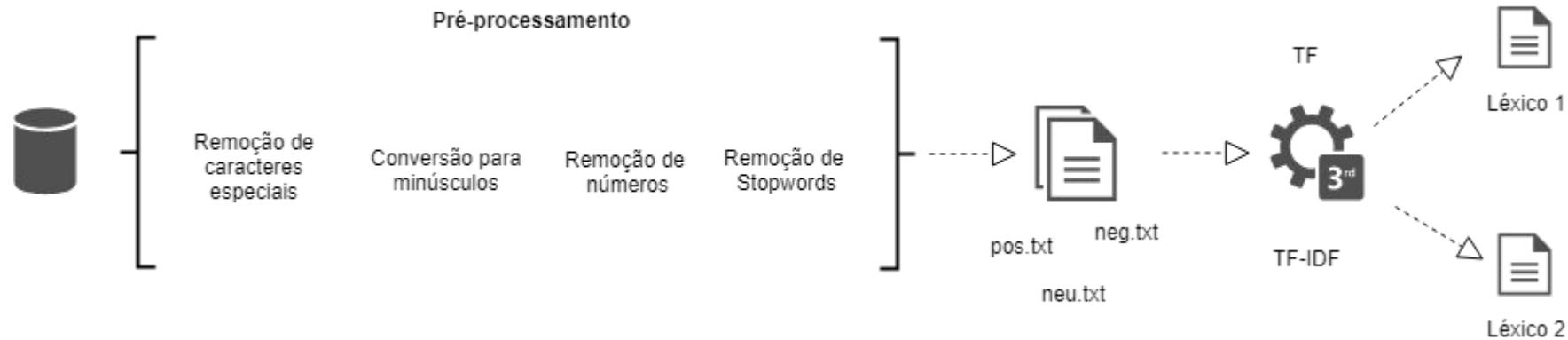
1308 (InfoMoney), 100 (Investing.com), 626 (TradingView)

Dados Classificados TradingView:

- Positivos: 286
- Neutros: 19
- Negativos: 130
- Total: 435

Métrica	OpLexicon	SentiLex	UniLex
Precision	0,693	0,721	0,612
Recall	0,763	0,822	0,631
F1-score	0,725	0,76	0,621
Acuracy	0,6	0,673	0,502

Construção de Léxico



Bag of Words

A essência desta técnica é converter documentos de textos em vetores, de modo que cada documento seja convertido em um vetor que representa a frequência de todas as palavras distintas presentes em um espaço vetorial de um documento.



TF-IDF

Indica a ponderação de termos que ocorrem em termos proporcionalmente à sua frequência [3]. Esta técnica foi originalmente desenvolvida como uma métrica para funções de classificação, na busca de resultados em mecanismos de busca, baseadas em consultas de usuários.

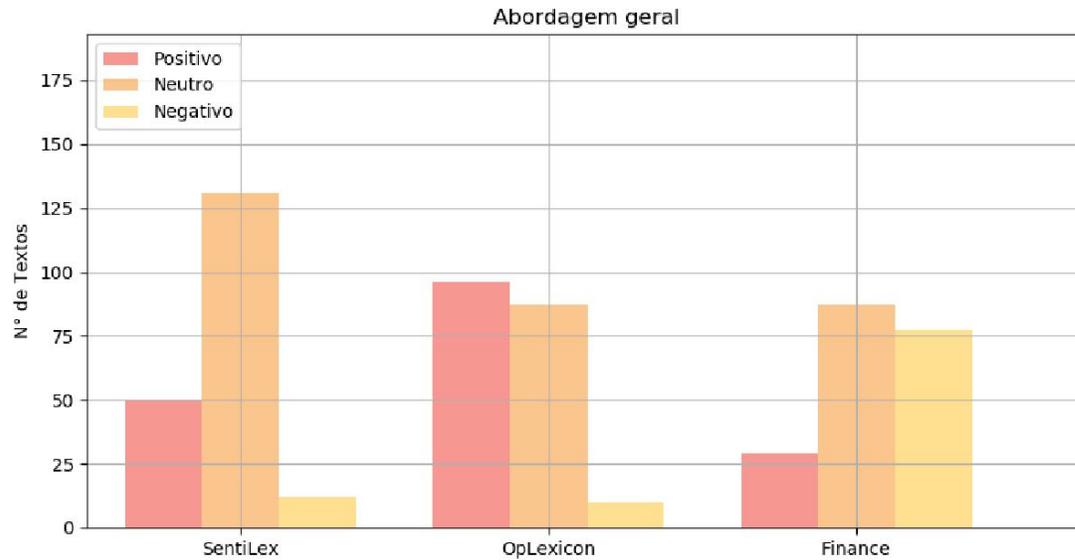
Geração de Léxicos

Bag of Words	
Termos	Polaridade
Vermelho	-1
Resistencia	-1
Negativo	-1
Alta	1
Queda	1
Oportunidade	1
Rompendo	1
Força	0
Retração	0
Romper	0
Risco	0
Queda	0

TF-IDF	
Termos	Polaridade
Revertendo	-1
Falha	-1
Difícil	-1
Acumular	1
Reagiu	1
Caiu	1
Disparando	1
Ideal	0
Sustentado	0
Margem	0
Correção	0
Gerou	0

Léxico		
Termos	Pol	Pos
Pior	-1	Adj
Rompido	-1	Adj
Perder	-1	Vb
Segurar	1	Vb
Atentar	1	Vb
Impulsionar	1	Vb
Sólido	1	Adj
Corrigir	0	Vb
Sinalizar	0	Vb
Negociar	0	Vb
Atingido	0	Adj
Permanecer	0	Vb

Aplicando o léxico...



Métrica	OpLexicon	SentiLex	UniLex	Finance
Precision	0,693	0,721	0,612	0,78
Recall	0,763	0,822	0,631	0,64
F1-score	0,725	0,76	0,621	0,69
Acuracy	0,6	0,673	0,502	0,62



Considerações Finais

EXPERIENCE

POSITION TITLE for company tld
Present
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2013 - 2016
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2012 - 2013
Short description of the position and the responsibilities you had in this position.
Lorem ipsum dolor sit amet, consectetur adipiscing elit. In ut enim euismod, nisi ut elit.

POSITION TITLE for company tld
2003 - 2010
Short description of the position and the responsibilities you had in this position.

REFERENCES

ELIOT BROWN
0028 01234 5678
eliot@mypage.com

Business Plan

ADDRESS
125 Name Street,
Town / City,
State / Country,
Postal / ZIP code

HOBBIES
creating websites
swimming
photography
body building

PHONE
0028 01234 5678

EMAIL
info@sambalq.com

WEBSITE
www.mypage.com

SKYPE
skype: sambalq

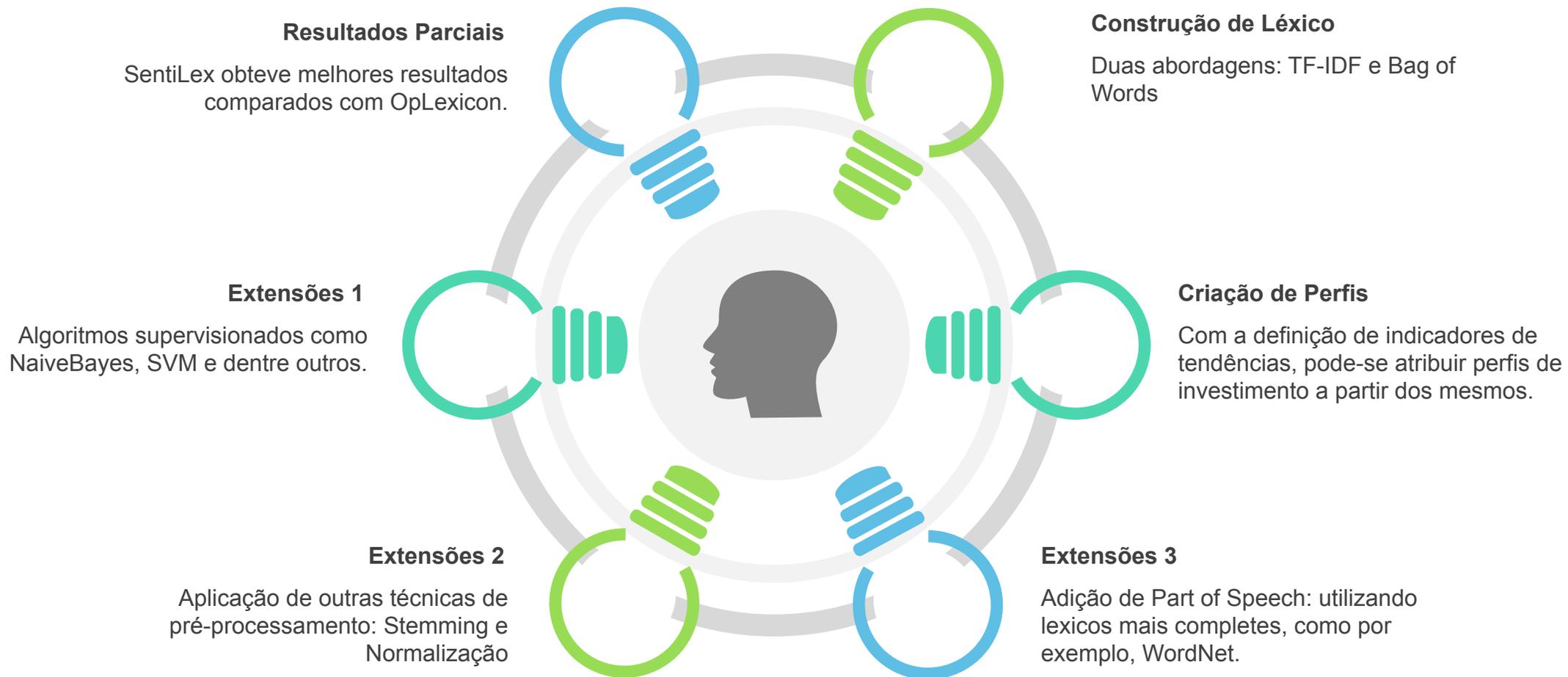


PROFESSIONAL STATEMENT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse suscipit efficitur lectus. Fusce iaculis, leo nec vulputate efficitur. In ut enim euismod, nisi ut elit.



ESCOLA
POLITÉCNICA



Referências

EXPERIENCE

POSITION TITLE for company tld
Present
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2013 - 2016
Short description of the position and the responsibilities you had in this position.

POSITION TITLE for company tld
2012 - 2013
Short description of the position and the responsibilities you had in this position.
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Inusami qui spe volur new.

POSITION TITLE for company tld
2003 - 2010
Short description of the position and the responsibilities you had in this position.

REFERENCES

ELIOT BROWN
0028 01234 5678
eliot@mypage.com

Business Plan

ADDRESS
125 Name Street,
Town / City,
State / Country,
Postal / ZIP code

HOBBIES
creating websites
swimming
photography
body building

PHONE
0028 01234 5678

EMAIL
info@sambalq.com

WEBSITE
www.mypage.com

SKYPE
skype: sambqak



PROFESSIONAL STATEMENT
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse suscipit efficitur lectur. Fusce laculis, leo nec vulputate efficitur. Lorem interdum elit, ut vestibulum nid met



ESCOLA
POLITÉCNICA

- [1] Symeonidis, S.; Effrosynidis, D.; Arampatzis, A. “A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis”, *Expert Systems with Applications*, vol. 110, 2018, pp. 298 – 310.
- [2] Bisi, T. “Análise técnica”. In: *L&S Educação - Traders ensinando traders*, Educação, L. (Editor), 2017.
- [3] Sarkar, D. “Text Analysis with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data”. Apress, 2016.
- [4] Zhao, J.; Liu, K.; Xu, L. “Sentiment analysis: Mining opinions, sentiments, and emotions bing liu (university of illinois at chicago) cambridge university press, 2015 isbn 9781107017894”, *Computational Linguistics*, vol. 42, 06 2016, pp. 1–4.
- [5] Mizumoto, K.; Yanagimoto, H.; Yoshioka, M. “Sentiment analysis of stock market news with semi-supervised learning”. In: 2012 IEEE/ACIS 11th International Conference on Computer and Information Science, 2012, pp. 325–328.
- [6] Im, T. L.; San, P. W.; On, C. K.; Alfred, R.; Anthony, P. “Analysing market sentiment in financial news using lexical approach”. In: 2013 IEEE Conference on Open Systems (ICOS), 2013, pp. 145–149.
- [7] palanisamy, P.; Yadav, V.; Elchuri, H. “Serendio: Simple and practical lexicon based approach to sentiment analysis”. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 2013, pp. 543–548.
- [8] Devitt, A.; Ahmad, K. “Sentiment polarity identification in financial news: A cohesion-based approach”. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007, pp. 984–991.

Obrigado!

Vitor Miguel Xavier Peres
vitor.peres@edu.pucrs.br
Mestrando - PPGCC
PUCRS - Escola Politécnica
Porto Alegre, Brasil

