



Universidade Federal
do Rio de Janeiro

Escola Politécnica

ANÁLISE DE NOTÍCIAS DO MERCADO FINANCEIRO UTILIZANDO
PROCESSAMENTO DE LINGUAGEM NATURAL E APRENDIZADO DE
MÁQUINA PARA DECISÕES DE SWING TRADE

Lucas Gama Canto

Projeto de Graduação apresentado ao Curso de Engenharia de Controle e Automação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientador: Heraldo Luís Silveira de Almeida

Rio de Janeiro
Março de 2020

ANÁLISE DE NOTÍCIAS DO MERCADO FINANCEIRO UTILIZANDO
PROCESSAMENTO DE LINGUAGEM NATURAL E APRENDIZADO DE
MÁQUINA PARA DECISÕES DE SWING TRADE

Lucas Gama Canto

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO
CURSO DE ENGENHARIA DE CONTROLE E AUTOMAÇÃO DA ESCOLA
POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE ENGENHEIRO DE AUTOMAÇÃO.

Examinado por:

Prof. Nome do Primeiro Examinador Sobrenome, D.Sc.

Prof. Nome do Segundo Examinador Sobrenome, Ph.D.

Prof. Nome do Terceiro Examinador Sobrenome, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2020

Gama Canto, Lucas

Análise de Notícias do Mercado Financeiro Utilizando Processamento de Linguagem Natural e Aprendizado de Máquina Para Decisões de Swing Trade/Lucas Gama Canto. – Rio de Janeiro: UFRJ/ Escola Politécnica, 2020.

X, 8 p.: il.; 29,7cm.

Orientador: Heraldo Luís Silveira de Almeida

Projeto de Graduação – UFRJ/ Escola Politécnica/ Curso de Engenharia de Controle e Automação, 2020.

Referências Bibliográficas: p. ?? – ??.

1. Primeira palavra-chave.
 2. Segunda palavra-chave.
 3. Terceira palavra-chave.
- I. Silveira de Almeida, Heraldo Luís. II. Universidade Federal do Rio de Janeiro, Escola Politécnica, Curso de Engenharia de Controle e Automação. III. Título.

*Ao povo brasileiro, pela total
contribuição em minha
graduação.*

Agradecimentos

Gostaria de agradecer a todas as pessoas e situações que tornaram este momento possível. Em especial, meus pais Benedita e Manoel, pelo suporte e esforço incondicional em apoiar minha decisão de vir estudar engenharia no Rio de Janeiro, aos professores da graduação, que me fizeram evoluir no âmbito acadêmico, profissional e pessoal e aos amigos que me apoiaram e participaram do meu processo de graduação.

Resumo do Projeto de Graduação apresentado à Escola Politécnica/ UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Automação.

ANÁLISE DE NOTÍCIAS DO MERCADO FINANCEIRO UTILIZANDO
PROCESSAMENTO DE LINGUAGEM NATURAL E APRENDIZADO DE
MÁQUINA PARA DECISÕES DE SWING TRADE

Lucas Gama Canto

Março/2020

Orientador: Heraldo Luís Silveira de Almeida

Curso: Engenharia de Controle e Automação

Com o objetivo de automatizar análises fundamentalistas de mercado, o uso de tecnologia para processamento de texto vem sendo utilizado constantemente no meio acadêmico e profissional. De forma a contribuir para este campo em crescimento, este trabalho discorre um estudo acerca da criação de modelos preditivos sobre a valorização ou desvalorização de ações na bolsa de valores do Brasil (B3, antiga Bovespa) a partir de notícias sobre o mercado financeiro brasileiro de forma a auxiliar decisões de Swing Trade, ou seja, compra e venda de ações dentro de uma janela de tempo maior que um dia.

Para isto, o presente projeto utiliza o framework PyText, que se baseia em conceitos de Aprendizado de Máquina, Redes Neurais e Processamento de Linguagem Natural de forma a desenvolver modelos preditivos com a tarefa de classificação textual.

Abstract of Undergraduate Project presented to POLI/UFRJ as a partial fulfillment of the requirements for the degree of Engineer.

FINANCIAL MARKET NEWS ANALYSIS USING NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING FOR SWING TRADE DECISIONS

Lucas Gama Canto

March/2020

Advisor: Heraldo Luís Silveira de Almeida

Course: Automation and Control Engineering

In order to automate fundamental market analysis, the use of text processing technology has been constantly used in academic and professional means. To contribute to this growing field, this paper discusses a study about the creation of predictive models regarding the valuation or devaluation of shares on the Brazilian stock exchange (B3, former Bovespa) based on news about the Brazilian financial market in order to assist Swing Trade decisions, that is, buying and selling stocks within a time window longer than one day.

To this end, the present project uses the PyText framework, which is based on Machine Learning, Neural Networks and Natural Language Processing concepts in order to develop predictive models with the task of textual classification.

Sumário

Lista de Figuras	ix
Lista de Tabelas	x
1 Introdução	1
1.1 Tema	2
1.2 Delimitação	2
1.3 Justificativa	2
1.4 Objetivos	2
1.5 Metodologia	2
1.6 Descrição	2
2 Fundamentação Teórica	3
2.1 Economia Financeira	3
2.2 Aprendizado de Máquina	3
2.3 Processamento de Linguagem Natural	3
2.4 Redes Neurais	3
3 Pré-processamento de Dados	4
4 Treinamento	5
5 Conclusões	6
6 Revisão Bibliográfica	7
A Algumas Demonstrações	8

Lista de Figuras

1.1	Logotipo da POLI-UFRJ.	2
1.2	Logotipo da COPPE.	2

Lista de Tabelas

1.1	Siglas dos cursos de engenharia da Escola Politécnica da UFRJ. . . .	1
1.2	Siglas dos programas de pós graduação da COPPE.	2
6.1	Exemplos de citações utilizando o comando padrão <code>\cite</code> do <code>L^AT_EX</code> e o comando <code>\citet</code> , fornecido pelo pacote <code>natbib</code>	7

Capítulo 1

Introdução

Segundo a norma de formatação de teses e dissertações do Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia (COPPE), toda abreviatura deve ser definida antes de utilizada.

Do mesmo modo, é imprescindível definir os símbolos, tal como o conjunto dos números reais \mathbb{R} e o conjunto vazio \emptyset .

Você deve selecionar seu curso de engenharia usando o comando `\department{Sigla}` e no lugar de Sigla inserir a sigla referente ao seu curso de engenharia. A tabela 1.1 relaciona as siglas dos cursos de engenharia da Escola Politécnica da Universidade Federal do Rio de Janeiro (POLI-UFRJ), enquanto que a tabela 1.2 relaciona as siglas dos programas de pós graduação da COPPE.

Tabela 1.1: Siglas dos cursos de engenharia da Escola Politécnica da UFRJ.

Sigla	Curso
EA	Engenharia Ambiental
ECV	Engenharia Civil
ECI	Engenharia de Computação e Informação
ECA	Engenharia de Controle e Automação
EMAT	Engenharia de Materiais
EPT	Engenharia de Petróleo
EPR	Engenharia de Produção
EEC	Engenharia Eletrônica e de Computação
EET	Engenharia Elétrica
EMC	Engenharia Mecânica
EMET	Engenharia Metalúrgica
ENO	Engenharia Naval e Oceânica
ENU	Engenharia Nuclear

Note também que todas as figuras ou tabelas devem ser citadas no texto. Como ocorre com as tabelas 1.1 e 1.2. Para ilustrar o uso de figuras em \LaTeX , considere as figuras 1.1 e 1.2.

Tabela 1.2: Siglas dos programas de pós graduação da COPPE.

Sigla	Curso
PEB	Engenharia Biomédica
PEC	Engenharia Civil
PEE	Engenharia Elétrica
PEM	Engenharia Mecânica
PEMM	Engenharia Metalúrgica e de Materiais
PEN	Engenharia Nuclear
PENO	Engenharia Oceânica
PPE	Planejamento Energético
PEP	Engenharia de Produção
PEQ	Engenharia Química
PESC	Engenharia de Sistemas e Computação
PET	Engenharia de Transportes



Figura 1.1: Logotipo da POLI-UFRJ.

1.1 Tema

1.2 Delimitação

1.3 Justificativa

1.4 Objetivos

1.5 Metodologia

1.6 Descrição



Figura 1.2: Logotipo da COPPE.

Capítulo 2

Fundamentação Teórica

2.1 Economia Financeira

2.2 Aprendizado de Máquina

2.3 Processamento de Linguagem Natural

2.4 Redes Neurais

Capítulo 3

Pré-processamento de Dados

Capítulo 4

Treinamento

Capítulo 5

Conclusões

Capítulo 6

Revisão Bibliográfica

Para ilustrar a completa adesão ao estilo de citações e listagem de referências bibliográficas, a Tabela 6.1 apresenta citações de alguns dos trabalhos contidos na norma fornecida pela CPGP da COPPE, utilizando o estilo numérico.

Tabela 6.1: Exemplos de citações utilizando o comando padrão `\cite` do \LaTeX e o comando `\citet`, fornecido pelo pacote `natbib`.

Tipo da Publicação	<code>\cite</code>	<code>\citet</code>
Livro	[?]	?]
Artigo	[?]	?]
Relatório	[?]	?]
Relatório	[?]	?]
Anais de Congresso	[?]	?]
Séries	[?]	?]
Em Livro	[?]	?]
Dissertação de mestrado	[?]	?]
Tese de doutorado	[?]	?]

É importante notar que, segundo a Norma para a Elaboração Gráfica do Projeto de Graduação da Escola Politécnica da UFRJ para trabalhos de conclusão de curso de engenharia de julho de 2012, as referências bibliográficas podem ser apresentadas de duas formas: (i) Referências numeradas e (ii) Referências em ordem alfabética. Para exibição numerada, em que a exibição das referências bibliográficas segue a ordem de citação usada no texto, use o comando `\bibliographystyle{coppe-unsrt}`. Para exibição de referências bibliográficas em ordem alfabética, basta usar o comando `\bibliographystyle{coppe-plain}` ao final do documento.

Apêndice A

Algumas Demonstrações