

UNIVERSIDADE FEDERAL DE ALFENAS
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

João Pedro Molina Peinado

ANÁLISE DE TENDÊNCIAS NO MERCADO FINANCEIRO
ATRAVÉS DE NOTÍCIAS UTILIZANDO MODELOS DE
APRENDIZADO DE MÁQUINA

Alfenas, 15 de Julho de 2019.

UNIVERSIDADE FEDERAL DE ALFENAS
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ANÁLISE DE TENDÊNCIAS NO MERCADO FINANCEIRO
ATRAVÉS DE NOTÍCIAS UTILIZANDO MODELOS DE
APRENDIZADO DE MÁQUINA

João Pedro Molina Peinado

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Alfenas como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Orientador: Prof. Humberto César Brandão de Oliveira

Alfenas, 15 de Julho de 2019.

João Pedro Molina Peinado

**ANÁLISE DE TENDÊNCIAS NO MERCADO FINANCEIRO
ATRAVÉS DE NOTÍCIAS UTILIZANDO MODELOS DE
APRENDIZADO DE MÁQUINA**

A Banca examinadora abaixo-assinada aprova a monografia apresentada como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação pela Universidade Federal de Alfenas.

Prof. Douglas Donizeti de Castilho Braz
Instituto Federal do Sul de Minas

Prof. Rodrigo Martins Pagliares
Universidade Federal de Alfenas

Prof. Humberto César Brandão de Oliveira (Orientador)
Universidade Federal de Alfenas

Alfenas, 15 de Julho de 2019.

Dedico este trabalho aos meus pais, Afonso e Maria, e ao meu irmão Paulo.

AGRADECIMENTO

Primeiramente a meus pais, Afonso Peinado Filho e Maria Cristina Molina Peinado, por todo carinho e amor dados a mim e ao apoio em todas as minhas escolhas.

Ao meu irmão, Paulo Afonso Molina Peinado, por toda a ajuda e todos os momentos de descontração.

À minha namorada, Taila Martins Cardoso, pelo carinho, amor e compreensão em todos os momentos.

Ao meu orientador, Prof. Humberto Brandão, por todos os ensinamentos prestados a mim.

A todas as amizades construídas durante todos esses anos de curso, amizades essas que levarei para a vida toda.

Aos companheiros do Laboratório de Pesquisa e Desenvolvimento (LP&D) pelas discussões e conhecimentos compartilhados.

Aos professores do BCC por todos os conhecimentos passados a mim.

“Só se pode alcançar um grande êxito quando nos mantemos fiéis a nós mesmos.”

Friedrich Nietzsche

RESUMO

Este trabalho propõe uma abordagem, utilizando modelos de aprendizado de máquina, para prever o preço de ativos após tendências causadas por notícias no mercado financeiro. Para realização dos experimentos, foram utilizados os dados da Bolsa de Valores de Londres (LSE) e as notícias das empresas pertencentes ao índice FTSE 100, que é composto pelas 100 empresas mais valiosas listadas na LSE. Este trabalho mostra que, é possível obter resultados positivos ao utilizar técnicas de aprendizado de máquina para prever o preço de ativos com base nos dados de análise de notícias.

Palavras-Chave: notícias, mercado financeiro, aprendizado de máquina.

ABSTRACT

This work proposes an approach using machine learning models to predict asset prices following trends caused by financial market news. In the experiments, we used data from the London Stock Exchange (LSE) and the news of the companies belonging to the FTSE 100 index, which is made up of the 100 most valuable companies listed on the LSE. This work shows that positive results can be obtained by using machine learning techniques to predict asset pricing based on news analytics data.

Keywords: news, financial market, machine learning.

LISTA DE FIGURAS

Figura 1. Impacto de uma notícia positiva no preço do índice FTSE.....	22
Figura 2. Impacto negativo causado por uma notícia positiva no índice FTSE.....	22
Figura 3. Impacto causado por uma notícia de baixa relevância no índice FTSE.....	23
Figura 4. Exemplo de um algoritmo baseado em árvore de decisão.....	30
Figura 5. Arquitetura de uma rede neural artificial simples.....	31
Figura 6. Quantidade de notícias publicadas por hora do ano de 2017.....	33
Figura 7. Densidade, em relação à relevância, das notícias das empresas presentes no FTSE no ano de 2017.....	33
Figura 8. Quantidade de notícias distribuídas entre os tipos de sentimentos no ano de 2017.....	34
Figura 9. Comportamento do valor de fechamento do índice durante o dia 5 de outubro de 2017.....	35

LISTA DE TABELAS

Tabela 1 - Dados gerados pela análise de notícias.....	21
Tabela 2 - Resultados dos modelos para o segundo semestre de 2015.....	37
Tabela 3 - Resultados dos modelos para o segundo semestre de 2016.....	38
Tabela 4 - Resultados dos modelos para o segundo semestre de 2017.....	39

LISTA DE ABREVIACES

GMT	Greenwich Mean Time
LightGBM	Light Gradient Boosting Machine
LSE	London Stock Exchange
MLP	Multilayer Perceptron (Rede Neural Perceptron Multicamada)
RNA	Redes Neurais Artificiais
S&P 500	Standard & Poor 500
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO	15
1.1 JUSTIFICATIVA E MOTIVAÇÃO	16
1.2 PROBLEMATIZAÇÃO.....	16
1.3 OBJETIVOS	17
1.3.1 Gerais	17
1.3.2 Específicos.....	17
1.4 ORGANIZAÇÃO DA MONOGRAFIA	18
2 ANÁLISE DE NOTÍCIAS E O IMPACTO NO MERCADO FINANCEIRO.....	19
2.1 ANÁLISE DE NOTÍCIAS	19
2.2 IMPACTO NO MERCADO FINANCEIRO.....	21
3 TRABALHO RELACIONADOS	25
3.1 NOTÍCIAS E A TENDÊNCIA DO MERCADO FINANCEIRO	25
3.2 PREDIÇÃO DO RETORNO DE AÇÕES	27
4 MATERIAIS E MÉTODOS.....	29
4.1 MÉTODO PROPOSTO.....	29
4.1.1 Aprendizado de Máquina.....	29
4.1.2 Variável Objetivo.....	31
4.2 BASE DE DADOS	31
4.2.1 Base de Dados de Análise de Notícias.....	32
4.2.2 Base de Dados Históricos do FTSE.....	35
4.3 EXPERIMENTAÇÃO	36
5 RESULTADOS	37
5.1 RESULTADOS DO ANO DE 2015	37
5.2 RESULTADOS DO ANO DE 2016	38
5.3 RESULTADOS DO ANO DE 2017	39
5.4 CONSIDERAÇÕES FINAIS.....	39
6 CONSIDERAÇÕES FINAIS.....	41
6.1 CONCLUSÕES	41
6.2 TRABALHOS FUTUROS	42
7 REFERÊNCIAS BIBLIOGRÁFICAS.....	43

1

Introdução

Este capítulo tem por finalidade apresentar as principais motivações e os objetivos deste trabalho. Na Seção 1.1 é apresentada a justificativa e a motivação. Na Seção 1.2 é apresentada a problematização. Os objetivos gerais e específicos são apresentados na Seção 1.3. Por fim, na Seção 1.4 é apresentada a organização da monografia.

Apresenta-se através deste trabalho uma abordagem utilizada para prever o preço de ativos com base nos dados de análise de notícias. Tal abordagem foi desenvolvida com o intuito de auxiliar os investidores em suas tomadas de decisões no mercado financeiro.

A abordagem consiste em utilizar modelos de aprendizado de máquina que, ao serem usados em conjunto aos dados gerados pela análise de notícias, sejam capazes de identificar os padrões responsáveis por causar uma alta ou uma baixa no preço do ativo, assim facilitando sua predição.

Para a realização dos experimentos, utilizou-se os dados históricos da Bolsa de Valores de Londres (LSE) em conjunto aos dados pré processados de análise de notícias. Especificamente para este trabalho, foram utilizadas as notícias das empresas presentes no índice FTSE 100. Segundo Mase (2007) o FTSE 100 é um índice que inclui as 100 empresas mais valiosas listadas na LSE, que representam mais de 80% de sua capitalização.

1.1 Justificativa e Motivação

A chegada de notícias a todo momento muda a percepção ou sentimento em relação a uma determinada empresa ou suas estratégias de negócios adotadas. Atualmente devido à facilidade da Internet, os comerciantes e investidores têm acesso constante às notícias atualizadas, e as notícias constantemente moldam seus sentimentos e afetam suas decisões de investir em uma empresa em particular (Chowdhury *et al.*, 2014).

Neste cenário algumas notícias podem ter uma maior influência na decisão dos investidores e como consequência o mercado pode sofrer um grande impacto. Na literatura são encontradas algumas maneiras de se utilizar notícias para identificar tendências, seja através da relevância ou através do seu sentimento, que é referente a classe de conteúdo, ou seja, positivo ou negativo. No entanto a utilização de somente essas informações não é sempre suficiente para identificar altas e baixas no preço de ativos.

Sendo assim, este trabalho tem como objetivo desenvolver uma abordagem que utiliza técnicas de aprendizado de máquina capaz de prever, com base nas informações geradas pela análise das notícias, o preço de ativos, auxiliando os investidores em suas decisões.

1.2 Problematização

As notícias sempre foram uma fonte importante de informações sobre investimentos. Em mercados cada vez mais competitivos, investidores precisam selecionar e analisar as notícias relevantes, a partir das vastas quantias disponíveis, a fim de tomar decisões boas e oportunas (Mitra & Mitra, 2011).

A identificação de uma tendência no mercado, com base em uma notícia, geralmente é feita através da análise do sentimento e da relevância. Somente aplicar este tipo de análise, não garante que aquela notícia causará uma tendência. A identificação deve ser realizada de maneira adequada, utilizando o maior número possível de informações, para que ela seja realizada de maneira precisa.

Neste contexto, a questão levantada por este trabalho é: “com base nos dados de análise de notícias, é possível utilizar modelos de aprendizado de máquina para prever o preço de ativos?”.

1.3 Objetivos

1.3.1 Gerais

Este trabalho teve como objetivo geral desenvolver uma abordagem capaz de prever o preço de ativos identificando tendências causadas pelas notícias. A abordagem consiste na utilização de modelos de aprendizado de máquina que sejam capazes de identificar padrões nos dados que levam uma notícia a causar uma tendência e assim facilitar a predição do preço do ativo.

1.3.2 Específicos

Para atingir o objetivo geral, os seguintes objetivos específicos foram necessários:

- Encontrar uma abordagem de tratamento dos dados gerados pela análise de notícias. Estes dados devem ser adequados para o modelo utilizado.

- Encontrar um modelo de aprendizado de máquina capaz de identificar os padrões na base de dados das notícias.

1.4 Organização da Monografia

Este trabalho foi dividido da seguinte maneira:

- O Capítulo 2 apresenta os detalhes sobre a análise de notícias e os impactos causados pelas notícias no mercado financeiro;
- O Capítulo 3 contém a revisão bibliográfica, onde se encontram os estudos presentes na literatura sobre os temas abordados neste trabalho;
- O Capítulo 4 contém a metodologia utilizada para a realização do trabalho e as informações sobre a implementação;
- O Capítulo 5 apresenta os resultados obtidos para a base de dados utilizada por este trabalho;
- O Capítulo 6 apresenta as conclusões e os trabalhos futuros sugeridos pelo autor para a continuidade do trabalho.

2

Análise de Notícias e o Impacto no Mercado Financeiro

Neste capítulo são apresentados conceitos que fundamentam a proposta deste trabalho. Na Seção 2.1 é apresentada a análise de notícias e a Seção 2.2, apresenta os impactos causado pelas notícias no mercado financeiro.

2.1 Análise de Notícias

Análise de notícias consiste em transformar as informações presentes nos textos em informações que possam ser interpretadas por um computador. Segundo Mitra & Mitra (2011) um aspecto importante da análise de notícias é descobrir o conteúdo informativo delas. Converter um texto qualitativo em um formato que possa ser interpretado por uma máquina pode ser uma tarefa desafiadora. Pode-se querer determinar se uma notícia possui conteúdo informativo e se esse conteúdo é positivo ou negativo, ou seja, determinar o seu sentimento.

O processamento dos textos de notícias pode ser realizado utilizando os sistemas automáticos de classificação de textos. O sistema identifica os principais conceitos com base em uma análise de probabilidade estatística da frequência e das relações de termos em um texto. Se o sistema for usado para analisar a sinopse textual de um filme, o conceito-chave seria filme, e o filme poderia até mesmo ser classificado em uma categoria pré-definida, como comédia, romance, ação, aventura ou ficção científica (Brown *et al.* 2002).

Atualmente existem várias empresas que oferecem *software* que capturam características particulares de notícias em tempo real. Essas ferramentas analisam eletronicamente as informações textuais disponíveis usando algoritmos de reconhecimento de padrões linguísticos. Palavras, padrões de palavras, a novidade de uma notícia, seu tipo e outras características são traduzidos em indicadores de relevância, novidade e o tom do item (Groß-Klußmann & Hautsch, 2010).

Ainda segundo os autores, a chegada de notícias é gravada com registros de horário Greenwich Mean Time (GMT) com precisão de até um milissegundo. Cada mensagem de notícia fornece um indicador de sentimento, novidade e relevância, como pode ser observado na Tabela 1. Os indicadores são produzidos com base em uma análise linguística automatizada do texto das notícias.

Tabela 1 – Dados gerados pela análise de notícias.

Data	Empresa	Relevância	Sentimento	Sentimento	Sentimento	Tipo Sentimento
			Negativo	Neutro	Positivo	
2014/09/15 13:00:07	3i Group	1.0	0.090223	0.152625	0.757151	1
2014/09/15 14:00:07	Experian	1.0	0.044799	0.130197	0.825004	1
2014/01/24 08:11:33	Pearson	0.040161	0.109096	0.158286	0.732645	1
2014/10/23 12:38:02	Tesco	0.353553	0.818456	0.125193	0.056351	-1

2.2 Impacto no Mercado Financeiro

Qualquer notícia pode causar impacto no preço das ações das empresas, podendo esse impacto ser positivo ou negativo. Na Figura 1 é possível observar o comportamento, anterior e posterior a uma notícia de um dos ativos presentes no índice FTSE, dos valores do índice. A notícia observada possui alta relevância e um elevado sentimento positivo o que acarreta uma valorização do índice.

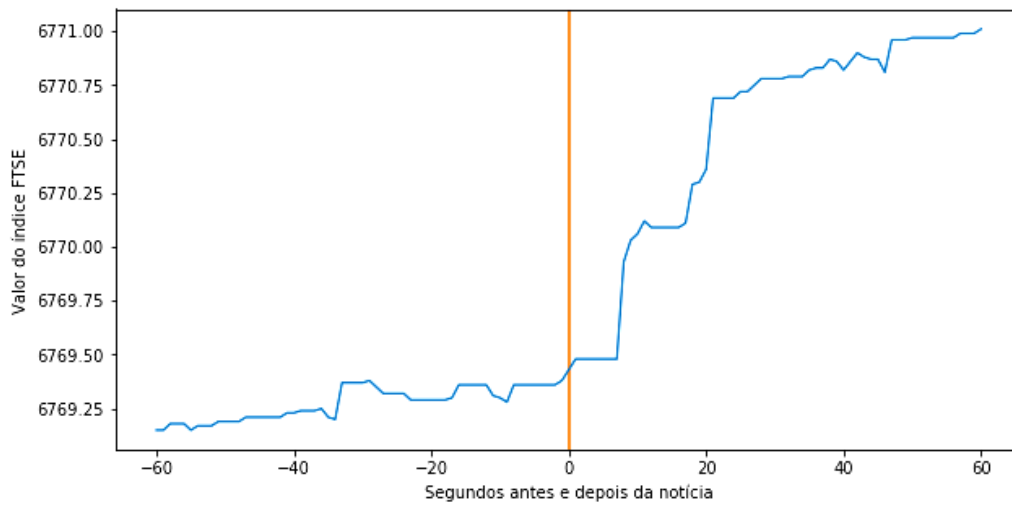


Figura 1. Impacto de uma notícia positiva no preço do índice FTSE.

Porém as informações de sentimento e relevância não são totalmente confiáveis pois uma notícia pode ter impacto contrário ao seu sentimento, como pode ser observado na Figura 2, no qual uma notícia de sentimento positivo acaba causando uma desvalorização do índice.

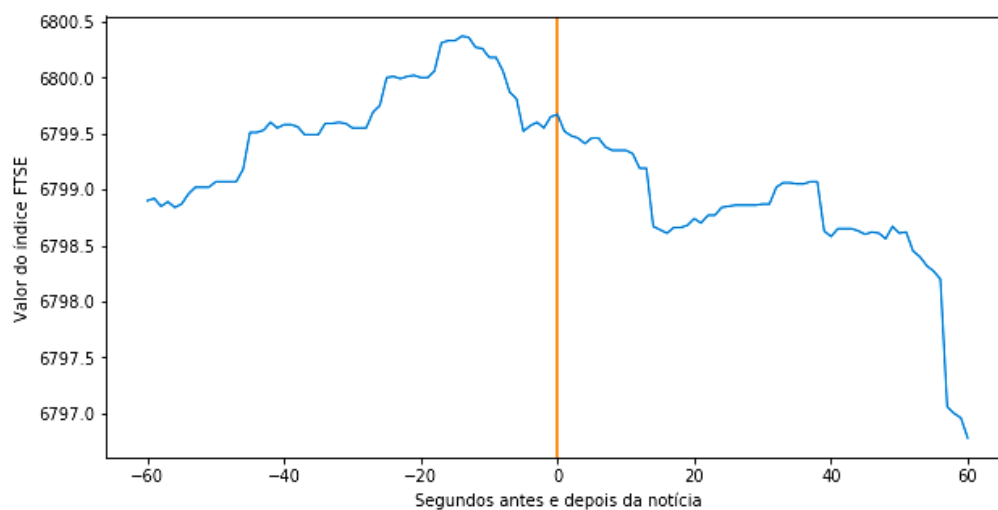


Figura 2. Impacto negativo causado por uma notícia positiva no índice FTSE.

Além do campo de sentimento, é possível observar que notícias de baixa relevância podem ser relevantes para o mercado, assim como mostrado na Figura 3.

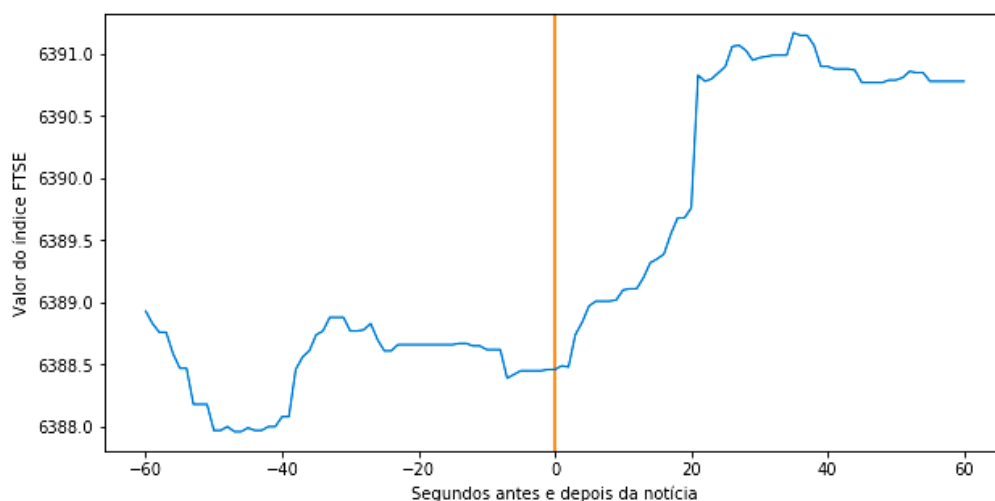


Figura 3. Impacto causado por uma notícia de baixa relevância no índice FTSE.

Assim como ilustrado, tomar decisões somente através dos campos de relevância e sentimento podem causar resultados contrários aos desejados. O foco da abordagem deste trabalho é justamente minimizar estes casos através da identificação de variações de padrões entre os dados das notícias.

3

Trabalho Relacionados

Este capítulo tem como objetivo apresentar alguns trabalhos que abordam os principais conceitos tratados nesta monografia. Na Seção 3.1 estão presentes os trabalhos referentes às tendências do mercado e na Seção 3.2 os trabalhos referentes à predição do retorno de ações.

Este trabalho propõe uma abordagem para prever o preço de ações com base em dados de notícias. Nesta seção se encontram os principais trabalhos que serviram de base para a pesquisa. Esta seção pode ser dividida em duas partes:

- A primeira parte contém os trabalhos que abordam os impactos e as tendências que são causados pelas notícias no mercado financeiro;
- A segunda apresenta trabalhos que abordam diferentes técnicas para predição do retorno das ações utilizando dados de notícias.

3.1 Notícias e a tendência do Mercado Financeiro

Wüthrich *et al.* (1998) analisam como o mercado financeiro de Hong Kong se comportava diante da publicação de notícias. Para o estudo foram coletadas notícias dos 5 mais populares sites financeiros do mundo, disponíveis antes da abertura do mercado. O estudo utilizou dados diários e mostrou que, para as notícias analisadas, o mercado não sofre influência.

Groß-Klußmann & Hautsch (2010) analisam o impacto na volatilidade, liquidez e no retorno de ações britânicas. Eles utilizaram 39 ações negociadas na Bolsa de Valores de Londres e os dados de análise de notícias pré-processados pela empresa Thomson Reuters. Os autores identificaram que somente o volume acumulado de negociações e a volatilidade sofrem influências das notícias e que os indicadores de sentimento são capazes de prever tendências futuras de preços.

Leinweber & Sisk (2012) analisam o impacto de notícias no mercado financeiro norte-americano. No estudo, eles utilizam dois intervalos de tempo diferentes para realizar a simulação, os intervalos são de 2003 até 2008 e de 2006 a 2009. A simulação demonstra que é possível capitalizar nos dias em que muitas notícias novas acontecem. Além disso, as simulações demonstram que é possível criar estratégias melhores do que utilizar a estratégia clássica de vender quando há notícias negativas e comprar quando há notícias positivas.

Chowdhury *et al.* (2014) discutem como as notícias de domínio público influenciam o mercado no dia a dia, com intuito de comprar, vender ou deter uma ação específica. Os autores utilizam um período total de 4 semanas e mostram que existe uma forte correlação entre o sentimento das notícias e a curva de preço das ações.

Heston & Sinha (2016) utilizam dados de notícias norte-americanas, durante os anos de 2003 a 2010, para prever o retorno de ações no mercado. O estudo demonstra que notícias com sentimento positivo impactam o mercado em até uma semana e que o impacto de notícias pode ter um atraso perto de datas onde ocorrem os anúncios de ganhos das empresas.

Kelly & Ahmad (2018) investigam o impacto de notícias com diferentes níveis de sentimento, eles propõem um método de quantificação do sentimento e utilizam métodos estatísticos para estimar o impacto causado pelos sentimentos dessas notícias no mercado. Os resultados demonstram que o sentimento das notícias é capaz de identificar tendências do mercado, e como consequência, aumentar o ganho de estratégias de negociação.

3.2 Predição do Retorno de Ações

A utilização de notícias para prever o retorno de ações tem sido abordado em diferentes estudos realizados pelo mundo todo. Nesta seção serão apresentados alguns trabalhos que utilizam diferentes técnicas com objetivo de realizar tal predição.

Mittermayer (2004) apresenta um sistema para prever tendências no mercado financeiro imediatamente após a publicação das notícias. O sistema é composto por três componentes, o primeiro é utilizado para identificar informações relevantes nas notícias através de técnicas de pré-processamento de texto, o segundo componente é responsável por classificar as notícias em categorias pré-definidas e, por fim, o último componente define as estratégias de negociação que serão adotadas. A utilização do sistema demonstra que, em estratégias específicas, a categorização de notícias é capaz de melhorar a predição do retorno das ações.

Zhai *et al.* (2007) apresentam um experimento utilizando *Support Vector Machines* (SVM) para prever o comportamento das ações de uma empresa na Bolsa de Valores Australiana. O experimento é feito utilizando notícias publicadas no maior jornal relacionado a negócios e finanças da Austrália. O resultado do experimento mostra que, sem a utilização das notícias, o modelo atinge uma acurácia na previsão de cerca de 58% e com a sua utilização foi capaz de aumentá-la para 70%.

Larkin & Ryan (2008) utilizam programação genética juntamente com dados de notícias, gerados em tempo real, para prever variação do preço de ações a curto prazo. Utilizando somente os dados das notícias, o estudo é capaz de identificar grande saltos de preços, do índice Standard & Poor 500 (S&P 500), até uma hora antes dos saltos de fato ocorrerem.

Ho & Wang (2016) desenvolvem uma abordagem utilizando redes neurais para prever o movimento no preço de ações durante o dia. No trabalho, são utilizados os dados de notícias pré processadas da bolsa norte-americana e os experimentos são feitos utilizando os preços das ações da Google Inc. O experimento mostra que é potencialmente lucrativo criar uma estratégia de compra e venda baseado nas predições da rede neural.

4

Materiais e Métodos

Esta Seção tem por objetivo descrever os materiais e métodos deste trabalho. A Seção 4.1 apresenta os detalhes sobre o método proposto, na Seção 4.2 estão descritos os detalhes da base dados utilizada neste trabalho e na Seção 4.3 é descrita a experimentação.

4.1 Método Proposto

4.1.1 Aprendizado de Máquina

Os algoritmos de aprendizado de máquina escolhidos para este trabalho foram um algoritmo baseado em uma floresta de árvores de decisão e um algoritmo de redes neurais artificiais.

Uma árvore de decisão utiliza os padrões de uma entrada para tomar decisões, sendo assim, ela é capaz encontrar a melhor sequência de decisões em cada padrão que possa alcançar a melhor resposta para aquela entrada. Um exemplo de algoritmo de árvore de decisão pode ser observado na Figura 4. Neste exemplo, ao analisar os dados de uma pessoa referente ao seu atraso, distância até o seu destino e quantidade de trânsito, é possível classificar se a pessoa é capaz de chegar ao seu destino percorrendo o caminho criado pelo algoritmo.

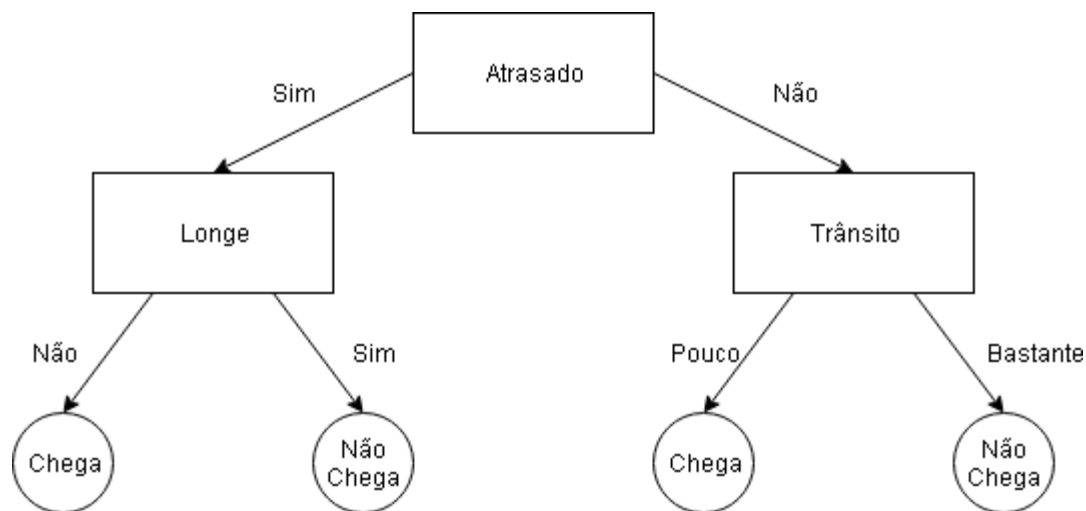


Figura 4. Exemplo de um algoritmo baseado em árvore de decisão.

O algoritmo baseado em árvores utilizado neste trabalho é o *Light Gradient Boosting Machine* (LightGBM) que é um algoritmo de aprimoramento gradiente baseado em uma floresta de árvores de decisão.

Os modelos de redes de neurais artificiais são algoritmos baseados em camadas que são compostas por unidades chamadas de neurônios e cada um dos neurônios é conectado por um vínculo que possui um valor numérico denominado peso. Os pesos são ajustados a cada vez que o conjunto de dados do treinamento é apresentado à rede neural. Por fim é utilizado os pesos finais para classificar uma entrada. Na Figura 5 é possível observar a arquitetura de uma rede neural artificial (RNA).

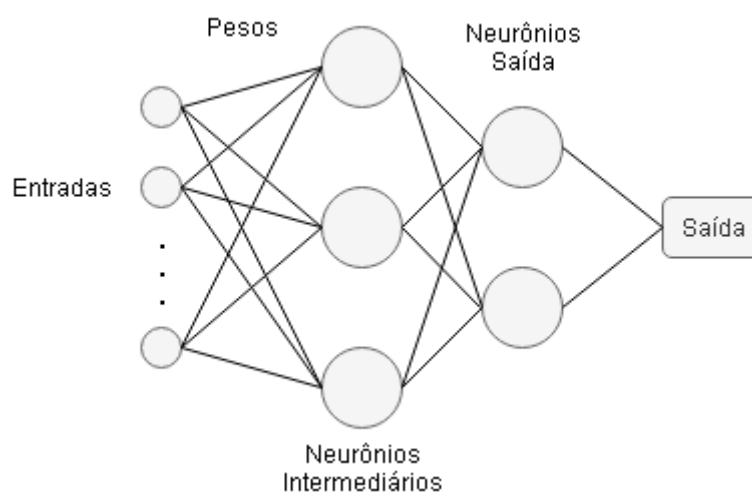


Figura 5. Arquitetura de uma rede neural artificial simples.

4.1.2 Variável Objetivo

Uma variável objetivo é aquela que o modelo de aprendizado utilizará para o seu aprendizado durante o treinamento e é a variável que o modelo tentará prever. No exemplo de árvore da Seção 4.1.1 a variável objetivo é se uma pessoa é capaz de chegar até o seu destino a tempo.

A variável objetivo deste problema é definida como sendo o valor do índice FTSE para três tempos diferentes, (5, 10 e 30) segundos após a publicação de uma notícia.

4.2 Base de Dados

Neste trabalho foram utilizadas duas bases de dados. A primeira base contém os dados de análise de notícias das empresas que estavam presentes no índice FTSE 100 do ano de 2015 ao ano de 2017. A segunda base contém os dados históricos do índice FTSE 100.

4.2.1 Base de Dados de Análise de Notícias

A primeira base de dados é a base de análise das notícias, nela estão presentes as seguintes informações:

- *Date-Time*: campo que indica em qual dia e hora foi feita a publicação.
- *Asset Name*: nome da empresa abordada.
- *Relevance*: campo que contém a relevância.
- *Sentiment Class*: campo que indica a classe de sentimento.
- *Sentiment Positive*: nível de sentimento positivo.
- *Sentiment Negative*: nível de sentimento negativo.
- *Sentiment Neutral*: nível de sentimento neutro.
- *Novelty*: campo que indica a novidade do assunto tratado.

Uma das informações mais importantes é a data, pois é através desta informação que é possível comparar o impacto da notícia no mercado. Além disso, é importante analisar somente as notícias que são publicadas durante o horário de funcionamento do mercado, pois a abordagem tenta identificar impactos em segundos após a publicação de uma notícia. Na Figura 6 é possível observar a quantidade de notícias publicadas por hora de funcionamento do mercado no ano de 2017.

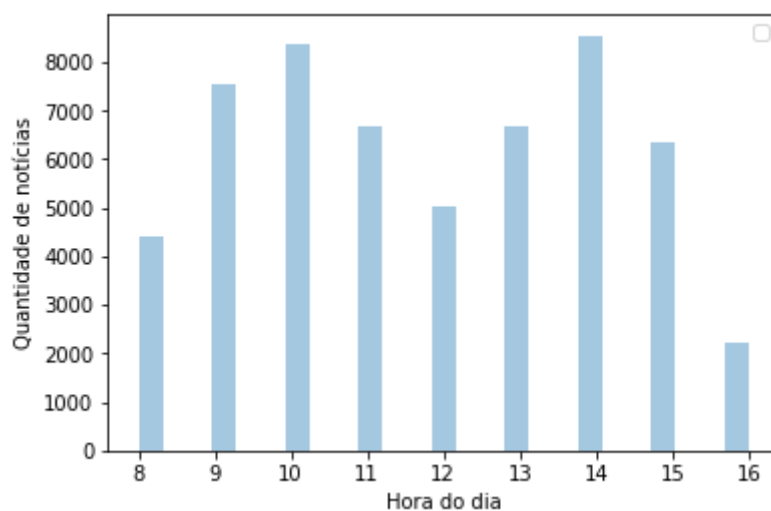


Figura 6. Quantidade de notícias publicadas por hora do ano de 2017.

O campo de relevância possui valores distribuídos entre 0 e 1, onde valores mais próximos a 1 indicam uma maior relevância enquanto valores próximos de 0 indicam notícias menos relevantes. Na Figura 7 é possível observar que notícias mais relevantes são mais comuns dentre as empresas presentes no índice FTSE.

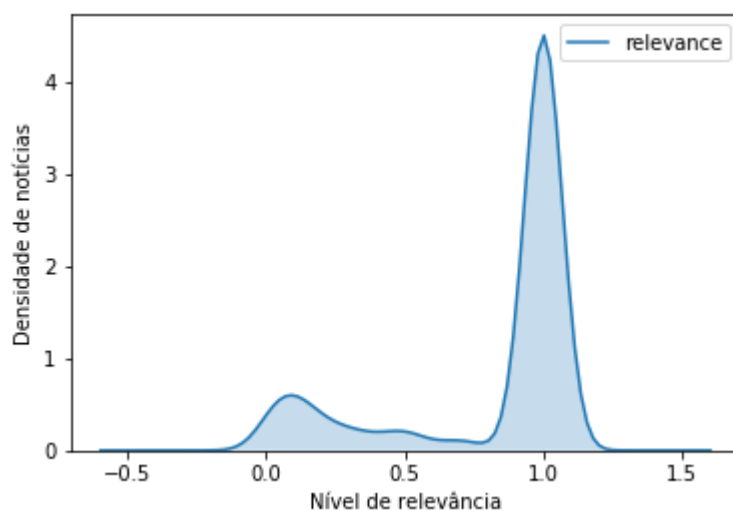


Figura 7. Densidade, em relação à relevância, das notícias das empresas presentes no FTSE no ano de 2017.

O sentimento das notícias é classificado em três diferentes tipos: positivo, neutro e negativo. O campo de classe de sentimento possui valor 1 para sentimento positivo, valor 0 para sentimento neutro e valor -1 para sentimento negativo. Além disso existem classificadores quantitativos para cada um dos sentimentos. É possível observar na Figura 8 que notícias positivas estão presentes em maior quantidade em relação aos outros dois tipos.

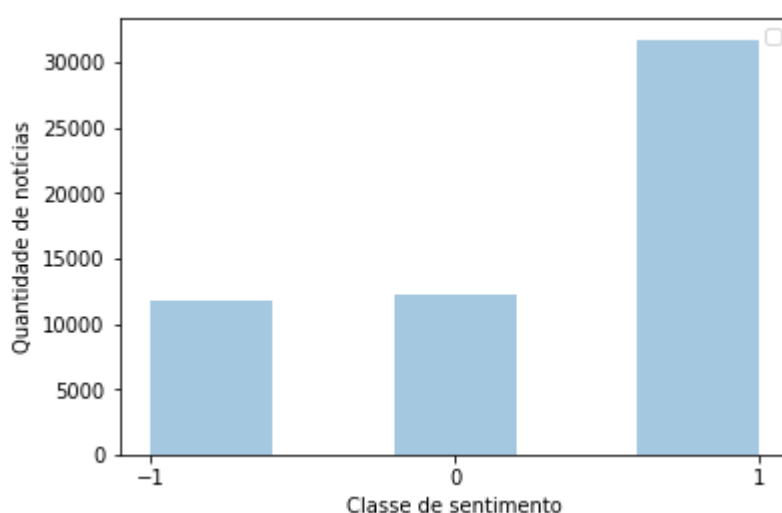


Figura 8. Quantidade de notícias distribuídas entre os tipos de sentimentos no ano de 2017.

O campo de novidade indica o quão novo é o assunto presente em uma notícia, ou seja, notícias que demoram mais tempo para serem publicadas terão valores de novidades maiores do que as notícias que são publicadas mais rapidamente. Na base de dados existem indicadores de novidade para as últimas 12 e 24 horas e para os últimos 3, 5 e 7 dias.

4.2.2 Base de Dados Históricos do FTSE

A segunda base de dados contém as informações do índice FTSE no mercado financeiro. Nela estão presentes dados em relação ao seu preço, sendo eles: preço de abertura, fechamento, alta e baixa. Além destas informações, a base também contém o volume negociado do índice. Todas essas informações foram gravadas segundo a segundo do ano de 2015 ao ano de 2017.

O comportamento do valor de fechamento do índice pode ser observado, segundo a segundo, na Figura 9. Como pode ser observado o índice possui uma variação alta durante o dia, possuindo momentos de baixa e momentos de alta.

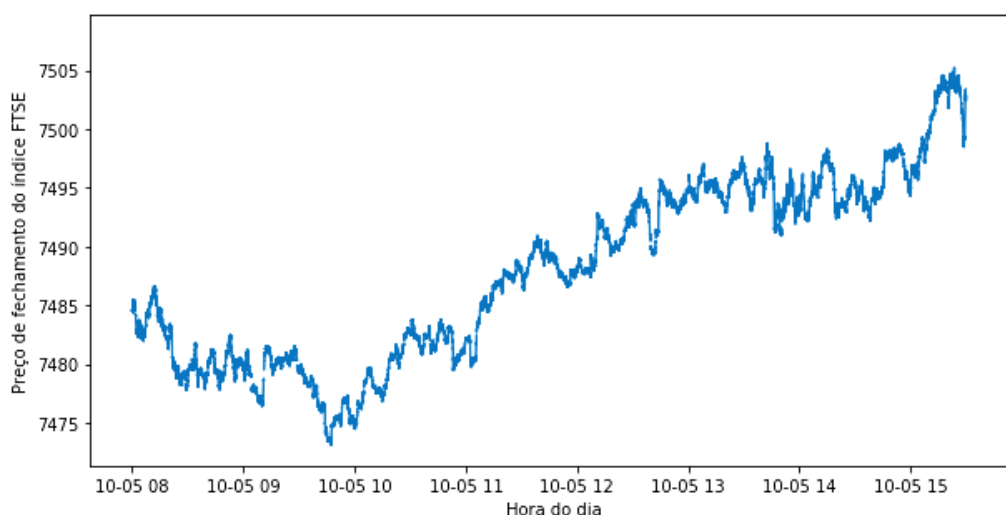


Figura 9. Comportamento do valor de fechamento do índice durante o dia 5 de outubro de 2017.

4.3 Experimentação

Para avaliar a abordagem desenvolvida neste trabalho, foi necessária uma divisão dos dados para cada um dos anos avaliados. Para cada ano, foram utilizadas as notícias dos seis primeiros meses como treinamento dos modelos e os dados dos seis últimos meses como base de avaliação.

Todos os modelos de aprendizado de máquina utilizados neste trabalho seguiram esta divisão na base de dados. Para cada um dos modelos foram realizados testes prevendo a variável objetivo descrita na Seção 4.1.2.

Para a comparação de resultados foi feita a análise do ganho obtido em pontos base (bps), que corresponde a 0,01 pontos percentuais, ou seja, 1% é equivalente a 100 bps. O ganho é calculado seguindo as seguintes estratégias:

- Se o modelo prever uma alta, o ganho é calculado com o valor de venda no momento de alta menos o valor de compra no momento da notícia;
- Se o modelo prever uma queda, o ganho é calculado com o valor de venda no momento da notícia menos o valor de compra no momento da queda.

Os experimentos são realizados todos de maneira automatizada, após o tratamento das bases de dados, é feita a modelagem de ambos os modelos utilizados para prever o preço do índice. Os resultados são calculados a partir das previsões de ambos os modelos.

5

Resultados

Este capítulo é responsável por apresentar os resultados do método proposto. Divide-se da seguinte forma: a seção 5.1 apresenta os resultados para o ano de 2015, na seção 5.2 os resultados para o ano de 2016, na seção 5.3 os resultados para o ano de 2017 e por fim, na seção 5.4 as considerações finais.

A apresentação dos resultados é feita separadamente para cada um dos anos e, para cada ano, são apresentados os ganhos totais, calculados seguindo as estratégias citadas na Seção 4.4.

5.1 Resultados do ano de 2015

Para os dados de 2015, os modelos são treinados com aproximadamente 30 mil notícias e são avaliadas aproximadamente 34 mil. A Tabela 2 apresenta os resultados obtidos em pontos bases para cada um dos tempos avaliados.

Tabela 2 - Resultados dos modelos para o segundo semestre de 2015.

Modelo	5 segundos	10 segundos	30 segundos
LightGBM	115,6	76,9	62,6
Rede Neural (MLP)	112,6	51,8	42,9

É interessante analisar que, para as notícias do segundo semestre de 2015, ambos os modelos conseguem um resultado positivo em todos os cenários. Além disso, o intervalo de 5 segundos se mostrou melhor do que os demais intervalos, utilizando o LightGBM o resultado foi aproximadamente 50% melhor do que utilizando o intervalo de 10 segundos, já utilizando a rede neural o resultado foi 119% melhor analisando os mesmos intervalos.

5.2 Resultados do ano de 2016

Utilizando os dados de 2016, os modelos são treinados com cerca de 36 mil notícias e são avaliadas cerca de 24 mil, diferente do ano de 2015 que possuía mais notícias nos dados de treinamento do que nos dados de avaliação. A Tabela 3 apresenta os resultados obtidos.

Tabela 3 - Resultados dos modelos para o segundo semestre de 2016.

Modelo	5 segundos	10 segundos	30 segundos
LightGBM	111,9	93,6	68,9
Rede Neural (MLP)	42,8	81,7	29,2

Similar aos resultados obtidos no ano anterior, os resultados deste ano demonstram que, para as notícias de 2016, também pode-se observar um resultado positivo nos casos analisados. Porém, é possível observar que a rede neural obtém um melhor resultado quando utilizando o intervalo de 10 segundos após a publicação da notícia, diferente do LightGBM que obtém o melhor resultado no intervalo de 5 segundos.

5.3 Resultados do ano de 2017

No ano de 2017, são avaliadas cerca de 30 mil notícias ao mesmo tempo que são utilizadas 25 mil notícias para o treinamento. Na Tabela 4 é possível observar os resultados dos modelos.

Tabela 4 - Resultados dos modelos para o segundo semestre de 2017.

Modelo	5 segundos	10 segundos	30 segundos
LightGBM	150,5	193,4	197,5
Rede Neural (MLP)	95,5	184,4	160,4

Diferente dos anos anteriores, é possível observar nos resultados de 2017 que os modelos obtiveram um resultado maior para o intervalo de 30 segundos após a publicação de notícias, sendo o resultado do LightGBM 32% superior no intervalo de 30 segundos se comparado ao intervalo de 5 segundos e o resultado 68% melhor utilizando a rede neural.

5.4 Considerações Finais

A partir dos experimentos realizados, é possível observar que ambos os modelos utilizados conseguem um resultado positivo em todos os cenários analisados. Além disso, também é possível observar que o LightGBM obteve resultados melhores que o modelo de redes neurais.

6

Considerações Finais

Este capítulo apresenta as conclusões sobre o trabalho desenvolvido e traz algumas sugestões de trabalhos futuros. É organizado da seguinte forma: Na seção 6.1, são apresentadas as conclusões em relação aos objetivos do trabalho, e na Seção 6.2 são sugeridas propostas para trabalhos futuros.

6.1 Conclusões

Através dos resultados é possível concluir que a meta definida pelo objetivo geral do trabalho foi atingida. Dessa forma, foi possível, com as bases de dados utilizadas neste trabalho, desenvolver uma abordagem com modelos de aprendizado máquina, capaz de atingir resultados positivos ao prever o preço de ativos com base nas tendências causadas por notícias. Em relação aos objetivos específicos:

- Foi possível realizar o tratamento dos dados de análise de notícias para que eles fossem utilizados de maneira adequada por cada um dos modelos utilizados neste trabalho;
- Foi possível também utilizar modelos de aprendizado de máquina que fossem capazes de atingir resultados positivos ao prever o preço de ações, assim os modelos identificam os padrões nos dados que levam a tendências no mercado.

Com base nos resultados, também é possível concluir que os impactos das notícias no mercado não possuem um padrão de tempo para ocorrerem, sendo necessário um estudo mais aprofundado para definir qual o melhor intervalo de tempo a ser adotado para o treinamento dos modelos.

6.2 Trabalhos Futuros

Algumas sugestões para a continuidade do trabalho:

- Realizar experimentos, usando as mesmas bases, com o objetivo de identificar impactos na volatilidade e no volume de negociações;
- Realizar o experimento de maneira iterativa, realizando a predição de uma notícia de cada vez, sempre adicionando as notícias já avaliadas na base de treinamento;
- Avaliar a abordagem desenvolvida neste trabalho em outros intervalos de tempos;
- Aplicar a abordagem desenvolvida neste trabalho em ações de outros mercados financeiros.

7

Referências Bibliográficas

Mase B. (2007). The impact of changes in the FTSE 100 index, *Financial Review*, 42:461–84.

Chowdhury S. G., Routh S., Chakrabarti S. (2014). News Analytics and Sentiment Analysis to Predict Stock Price Trends, *International Journal of Computer Science and Information Technologies*, Vol. 5 (3), 3595–3604.

Mitra L. and Mitra G. (2011). Applications of news analytics in finance: A review. *The Handbook of News Analytics in Finance*, 596:1.

Groß-Klußmann A., Hautsch N. (2010). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, *Journal of Empirical Finance*, vol. 18 (pg. 321-40).

Brown D., Janes B., Steele M., Cooper R. (2002). *United States Patent No. US 20020099730A1*.

Disponível em: <https://patents.google.com/patent/US20020099730A1> Acesso em: 01 de Julho de 2019.

Wüthrich B., Cho V., Leung S., Permuntilleke D., Sankaran K., Zhang J., and Lam W. (1998). "Daily Stock Market Forecast from Textual Web Data", *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, IEEE Computer Society Press, Los Alamitos, CA, pp. 2720-2725.

Kelly, S. and Ahmad, K. (2018). Estimating the impact of domain-specific news sentiment on financial assets. *Knowl. - Based Syst.* 150, 116-126.

Mittermayer, M.-A. (2004). Forecasting Intraday Stock Price Trends with Text Mining Techniques. Proceedings of the 37th Hawaii International Conference on Social Systems. Hawaii.

Zhai Y., Hsu A., Halgamuge, S.K. (2007). Combining news and technical indicators in daily stock price trends prediction. Proceedings of the 4th international symposium on neural networks: advances in neural networks, Part III, Springer-Verlag, Nanjing, China, pp. 1087-1096.

Larkin F., Ryan C. (2008) Good News: Using News Feeds with Genetic Programming to Predict Stock Prices. In: O'Neill M. et al. (eds) Genetic Programming. EuroGP 2008. Lecture Notes in Computer Science, vol 4971. Springer, Berlin, Heidelberg.

Leinweber D., Sisk J. (2012). Relating News Analytics to Stock Returns. The Handbook of News Analytics in Finance.

Heston S. L., Sinha N. R. (2016). News versus sentiment: predicting stock returns from news stories. Technical Report 2016-048: Board of Governors of the Federal Reserve System, Washington.

Ho K.-Y., (Walter) Wang W. (2016). Predicting Stock Price Movements with News Sentiment: An Artificial Neural Network Approach. Springer International Publishing Switzerland.