

CEFET - Centro Federal de Ensino Tecnológico  
Mestrado em Modelagem Matemática Computacional  
Finanças Computacionais

LUIZ HENRIQUE MAIA CRUZ

**PREVISÃO DE TENDÊNCIAS EM PREÇOS DE AÇÕES UTILIZANDO NOTÍCIAS E  
INDICADORES**

Belo Horizonte – MG  
2016 / 1º semestre

# Sumário

<b>Lista de Figuras .....</b>	<b>iii</b>
<b>Lista de Tabelas .....</b>	<b>iv</b>
<b>Lista de Siglas .....</b>	<b>v</b>
<b>Resumo .....</b>	<b>vi</b>
<b>Abstract.....</b>	<b>vii</b>
<b>1 INTRODUÇÃO .....</b>	<b>8</b>
1.1 Motivação . . . . .	8
1.2 Objetivos . . . . .	8
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>10</b>
2.1 Trabalhos Relacionados . . . . .	10
2.2 Algoritmos Supervisionados . . . . .	11
2.2.1 SVM - Support Vector Machine . . . . .	12
2.2.2 Naive Bayes . . . . .	13
<b>3 METODOLOGIA .....</b>	<b>14</b>
3.1 Tipo de Pesquisa . . . . .	14
3.2 Procedimentos metodológicos . . . . .	14
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>15</b>
<b>5 CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>16</b>

<b>Referências Bibliográficas .....</b>	<b>17</b>
<b>Apêndice A – Linguagem gráfica do WebAPSEE .....</b>	<b>18</b>

## **Lista de Figuras**

## **Lista de Tabelas**

## **Lista de Siglas**

## Resumo

No mercado de ações há duas categorias de métodos de análise que são utilizados como ferramentas pelos investidores para alcançar maiores ganhos, são a análise técnica e a análise fundamentalista. Enquanto na análise fundamentalista destaca-se o estudo de uma empresa e de seu valor de mercado, na análise técnica o que vale é o estudo do movimento dos preços no mercado ao longo do tempo. Na análise técnica os investidores utilizam indicadores, baseados na observação do comportamento do mercado, para prever tendências.

Este trabalho tem por objetivo estabelecer relações entre informações publicadas na mídia e em redes sociais com o desempenho de ações na bolsa de valores. A partir dos dados extraídos e quantificados vamos procurar estabelecer relações com séries históricas de preços de ações. As relações estabelecidas poderão culminar na criação de um indicador a ser utilizado em operações futuras na bolsa. Além das notícias serão realizados experimentos com indicadores conhecidos, em separado e em conjunto com os dados de notícias coletados.

**Palavras-chave:** Finanças, computação, mineração de dados, indicadores, análise técnica, svm, naive bayes.

## Abstract

In the stock market there are two analysis methods categories utilized by the investors to achieve better gain, those are the technical analysis and the fundamentalist analysis. While the fundamentalist analysis involves the study of the price of a company and his market value, for the technical analysis which is important is the study of them prices movement in the market over time. In the technical analysis the investors rely on indicators based on the observation of the market behavior to forecast trends. The objective of this work is to establish relations between published informations on the media and social networks with the performance of assets in the stock Exchange. From the extract and quantified data we will try to establish relations with stock price historic series. The established relations would culminate in one indicator to be utilized in future stock Exchange operations. Beyond the news will be done experiments with known indicators, separate and together with the news collected data.

**Keywords:** Finances, computer, data mining, indicators, technical analisys, svm, naive bayes.



# 1 INTRODUÇÃO

## 1.1 Motivação

Um dos fatores que influenciam a flutuação dos preços no mercado de ações são as informações disponíveis ao público (1). Notícias divulgadas na imprensa e postagens em redes sociais podem provocar movimentos de elevação ou queda no preço de um ativo.

O mercado financeiro é um elemento crucial no desenvolvimento das sociedades modernas. Através do mercado financeiros agentes superavitários, que desejam investir seus recursos podem alcançar agentes deficitários dispostos a empreender. As boas alternativas de investimento, com possibilidade de grandes retornos tem levado um público cada vez maior a buscar o mercado de ações e consequentemente um maior número de pesquisadores tem voltado sua atenção para esse assunto.

No entanto a maior parte das pesquisas se concentra em abordagens da análise técnica. O que se justifica pela anseio dos investidores em ter acesso a métodos quantitativos e a disponibilidade de grandes bases históricas de preços (2). Informações textuais, disponíveis em notícias de jornais, revistas e redes sociais são uma fonte de dados importante para avaliação do desempenho do mercado, entretando devido a natureza não estruturada dessas fontes de dados normalmente são utilizadas apenas em abordagens fundamentalistas. O tratamento dessas informações através de métodos computacionais e seu uso na análise do comportamento de mercado é uma área de pesquisa desafiadora.

## 1.2 Objetivos

Neste trabalho tentaremos quantificar a relevância das informações publicadas e tentar estabelecer uma relação entre uma notícia veiculada e a variação no preço de uma ação.

Para alcançar os objetivos propostos utilizaremos técnicas de mineração de dados. Trabalharemos com algoritmos como SVM e Naive Bayes, conhecidos como algoritmos supervisi-

onados. Eles são chamados dessa forma por se caracterizarem por trabalhar em duas etapas: a etapa de treinamento, no qual aprendem um padrão treinando sobre uma base de dados conhecida e classificadas e outra etapa de teste no qual tentamos fazer previsões sobre outra base.

Nossas bases de dados são constituídas por séries históricas de preços de ações, obtidas a partir do serviço Bloomberg Professional, contratado pelo CEFET/MG e dados históricos de notícias conseguidos através do Observatório da Web, da UFMG.

## 2 REFERENCIAL TEÓRICO

### 2.1 Trabalhos Relacionados

Nesta seção analisamos alguns trabalhos relevantes em mineração de dados voltados para a aplicação no mercado financeiro.

Em (3) é descrito o NewsCATS, um sistema que automaticamente analisa e classifica em categorias notícias da imprensa e partir dessa classificação provê recomendações para negociações na bolsa de valores.

Tendo objetivo selecionar uma estratégia de investimento que possa alcançar uma rentabilidade acima da média de mercado, o NewsCATS tenta prever tendências nos preços das ações no momento imediatamente após a publicação de notícias na imprensa. Ele trabalha com tres componentes, o primeiro utiliza tecnicas de pré-processamento para recuperar informações relevantes, o segundo classifica as nformações em categorias e o terceiro deriva, a partir das informações fornecidas pelos outros dois componentes, estratégias para investimento.

O sistema foi implementado em JAVA, são utilizadas técnicas de pré-processamento como remoção de stop-words e steeming e a represenção das características dos textos é feita utilizando modelos como TF (term-frequency), IDF (Inverse Document Frequency) e TF x IDF. O algoritmo utilizado na classificação das notícias é o SVM.

Colocar resultados do NewsCATS

Um exemplo de trabalho utilizando redes sociais está presente em (4). Nesse artigo os pesquisadores confrontam dados coletados no Twitter com retorno diário das ações de empresas que fazem parte do índice S&P 500. Esse trabalho estabelece uma medida das emoções expressas nos tweets em relação a uma determinada empresa com o retorno das ações dessa empresa. O artigo trabalha com o conceito de valência emocional, podendo a valência ser positiva ou negativa.

O artigo investiga o período pelo qual as informações publicadas no Twitter permanecem influenciando um ativo e a relevância do número de seguidores de um usuário na influência

que ele exerce.

A partir das análises os autores chegaram a conclusão que os tweets influenciam as ações dentro do mesmo dia e em um intervalo de 10 dias. Além disso usuários com um número de seguidores acima da média são capazes de exercer uma influência mais imediata.

O trabalho de (5) aborda a utilização de uma síntese de técnicas de linguística, finanças e estatística para criar o Arizona Financial Text System (AZFinText), um sistema para previsão de preços discretos de ações.

O AZFinText utiliza uma combinação entre os preços das ações e o texto de notícias financeiras para fazer previsões a respeito do preço discreto de uma ação em um período curto de tempo. O sistema confia na hipótese de que há uma pequena janela de tempo que o mercado leva para encontrar o equilíbrio após uma nova informação ser noticiada, dessa forma uma abordagem automatizada é capaz de explorar essa janela e conseguir lucros acima da média do mercado.

Os autores utilizaram o algoritmo SVR, uma versão do SVM que utiliza regressão. As entradas para o algoritmo são as notícias e o preço atual dos ativos e o alvo é o preço a ser previsto. Para avaliar o retorno o AZFinText foi executado por um ano e obteve um retorno de 8,5

## 2.2 Algoritmos Supervisionados

Os algoritmos supervisionados caracterizam-se por utilizar uma base de dados de treino, na qual para cada elemento da base o valor alvo é conhecido, ou seja, dada uma questão a ser respondida pelo algoritmo, a base de dados de treino consiste em um conjunto de pares de entradas e saídas.

O algoritmo supervisionado analisa essa base de treino e a partir dela busca inferir uma função a partir da qual possam ser estabelecidas as saídas para novas entradas.

Podemos enumerar algumas etapas básicas que devem ser realizadas na aplicação de algoritmos supervisionados:

1. Determinar um conjunto de exemplos de treinamento para ser utilizado. Este conjunto deve ser representativo para modelar o problema a ser resolvido. Deve ser conseguido um conjunto de entradas e determinadas suas respectivas saídas, em alguns casos será necessário fazer um trabalho manual para conseguir um conjunto classificado.

2. Definir uma representação das características para as entradas de dados da função de aprendizagem. O mais comum é representar uma entrada como um vetor de características.
3. Escolher o algoritmo de aprendizagem mais adequado.
4. Dependendo da escolha do algoritmo, algumas parâmetros são necessários para o processamento. Ajustar esses parâmetros é fundamental para um bom resultado. Uma técnica utilizada para chegar a valores de parâmetros satisfatórios é a validação cruzada.
5. Avaliar a acurácia da função de aprendizagem em um conjunto de testes.

Nos experimentos realizados neste trabalho foram utilizados dois algoritmos supervisionados, o SVM e o Naive Bayes.

### 2.2.1 SVM - Support Vector Machine

A técnica de aprendizado de máquina SVM - Support Vector Machine - constitui um campo de pesquisa importante entre os algoritmos supervisionados. O SVM apresenta resultados superiores a outros algoritmos supervisionados, como as redes neurais, quando aplicado a campos como mineração de textos e reconhecimento de imagens.

Dado um conjunto de vetores de treinamento, onde os valores alvo dentro de cada vetor pertencem a duas classes, sendo essas classes linearmente separáveis, o SVM é capaz de encontrar um hiperplano que melhor separa os conjuntos de vetores nas suas devidas classes.

Mais formalmente podemos dizer que dado  $X_i$  para  $i$  em  $1, 2, \dots, N$  os vetores do conjunto de treinamento. Esses vetores pertencem às classes  $W_1$  e  $W_2$ , que assumimos como classes linearmente separáveis. O objetivo do SVM é encontrar um hiperplano  $g(x) = W^T x + w_0 = 0$  que separa corretamente os vetores.

Há situações em que os vetores não podem ser separados linearmente, nesses casos é aplicada uma transformação. Essa transformação consiste em uma função não-linear, conhecida como função kernel, que mapeia as entradas do conjunto de treinamento em uma dimensão na qual é possível encontrar um hiperplano que separe corretamente os vetores para suas respectivas classes.

As funções kernel mais utilizadas são:

O tipo de kernel e os parâmetros a serem utilizados variam de acordo com a natureza do problema no qual estão sendo aplicados.

### 2.2.2 Naive Bayes

Os algoritmos Naive Bayes constituem uma família de algoritmos baseadas na aplicação do Teorema de Bayes. O nome naive se deve ao fato do algoritmo assumir a hipótese de independência entre as características do problema a ser tratado.

A aplicação do naive bayes na categorização de textos é bastante comum. O seu uso se justifica pela simplicidade do algoritmo e os bons resultados apresentados, em muitos casos alcançando um resultado melhor que abordagens mais complexas.

As implementações do naive bayes variam e sua utilização depende da distribuição probabilística do conjunto de dados no qual será aplicado. Algumas implementações são naive Bayes multinomial, gaussiano e bernoulli.

### **3 METODOLOGIA**

#### **3.1 Tipo de Pesquisa**

#### **3.2 Procedimentos metodológicos**

## **4 RESULTADOS E DISCUSSÃO**



## **5 CONCLUSÕES E TRABALHOS FUTUROS**

## Referências Bibliográficas

- 1 CHAN, Y.-c.; CHUI, A. C. W.; KWOK, C. C. Y. The impact of salient political and economic news on the trading activity. *Pacific-Basin Finance Journal*, v. 9, n. 3, p. 195–217, 2001. Disponível em: <<http://EconPapers.repec.org/RePEc:eee:pacfin:v:9:y:2001:i:3:p:195-217>>.
- 2 NASSIRTOUSSI, A. K. et al. Review: Text mining for market prediction: A systematic review. *Expert Syst. Appl.*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 41, n. 16, p. 7653–7670, nov. 2014. ISSN 0957-4174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2014.06.009>>.
- 3 MITTERMAYER, M.-A. Forecasting intraday stock price trends with text mining techniques. In: *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 3 - Volume 3*. Washington, DC, USA: IEEE Computer Society, 2004. (HICSS '04), p. 30064.2–. ISBN 0-7695-2056-1. Disponível em: <<http://dl.acm.org/citation.cfm?id=962751.962898>>.
- 4 SUL, H.; DENNIS, A. R.; YUAN, L. I. Trading on twitter: The financial information content of emotion in social media. In: *2014 47th Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2014. p. 806–815. ISSN 1530-1605.
- 5 SCHUMAKER, R. P.; CHEN, H. A discrete stock price prediction engine based on financial news. *Computer*, IEEE Computer Society, Los Alamitos, CA, USA, v. 43, n. 1, p. 51–56, 2010. ISSN 0018-9162.

## **APÊNDICE A – Linguagem gráfica do WebAPSEE**

WebAPSEE-PML (*Process Modeling Language*) é a linguagem gráfica usada para modelar processos no ambiente Open-WebAPSEE. Nesta linguagem, um modelo de processo pode ser construído a partir de símbolos gráficos conectados e o detalhamento do relacionamento com os outros componentes do modelo é feito através de formulários específicos que apóiam essa tarefa.