

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Lucas de Oliveira Garrigós Leite

**MINERAÇÃO DE DADOS APLICADA À PREVISÃO  
DO PREÇO DE AÇÕES DE CONCESSIONÁRIAS  
DE ENERGIA ELÉTRICA DO ESTADO DE SÃO  
PAULO**

São Carlos  
2016



**Lucas de Oliveira Garrigós Leite**

**MINERAÇÃO DE DADOS APLICADA À PREVISÃO  
DO PREÇO DE AÇÕES DE CONCESSIONÁRIAS  
DE ENERGIA ELÉTRICA DO ESTADO DE SÃO  
PAULO**

Trabalho de Conclusão de Curso apresentado à  
Escola de Engenharia de São Carlos da  
Universidade de São Paulo.

Curso de Engenharia Elétrica – Ênfase em  
Eletrônica

**ORIENTADOR:** Prof. Dr. Ricardo A. S. Fernandes

São Carlos  
2016

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

L933m Leite, Lucas de Oliveira Garrigós  
Mineração de dados aplicada à previsão do preço de  
ações de concessionárias de energia elétrica do Estado  
de São Paulo / Lucas de Oliveira Garrigós Leite;  
orientador Prof. Dr. Ricardo Augusto Souza Fernandes;  
coorientador Prof. Dr. José Carlos de Melo Vieira  
Júnior. São Carlos, 2016.

Monografia (Graduação em Engenharia Elétrica com  
ênfase em Eletrônica) -- Escola de Engenharia de São  
Carlos da Universidade de São Paulo, 2016.

1. Estimação de séries temporais. 2. Mercado de  
ações. 3. Redes Neurais Artificiais. I. Título.

# FOLHA DE APROVAÇÃO

Nome: Lucas de Oliveira Garrigós Leite

Título: "Mineração de dados aplicada à previsão do preço de ações de concessionárias de energia elétrica do Estado de São Paulo"

Trabalho de Conclusão de Curso defendido e aprovado  
em 22/06/2016,

com NOTA 10,0 (Dez, Zero), pela Comissão Julgadora:

*Prof. Dr. Ricardo Augusto Souza Fernandes - (Orientador - DEE/UFSCar)*

*Prof. Associado Rogério Andrade Flauzino - (SEL/EESC/USP)*

*Prof. Dr. Guilherme Guimarães Lage - (DEE/UFSCar)*

Coordenador da CoC-Engenharia Elétrica - EESC/USP:  
Prof. Dr. José Carlos de Melo Vieira Júnior

## **Resumo**

LEITE, L. O. G. (2016). MINERAÇÃO DE DADOS APLICADA À PREVISÃO DO PREÇO DE AÇÕES DE CONCESSIONÁRIAS DE ENERGIA ELÉTRICA DO ESTADO DE SÃO PAULO. Trabalho de Conclusão de Curso – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2016.

A análise deste trabalho tem por intuito prever o fechamento, valor máximo e mínimo de ação de determinadas empresas do ramo de distribuição de energia elétrica, levando em conta somente dados históricos. Têm-se por motivação maior, uma eventual previsão de uma venda ou compra de determinado volume de ações. Desta maneira, foi levado em consideração fatores considerados relevantes no valor final de uma ação das empresas CPFL e Eletropaulo (ambas empresas de capital aberto). Estes fatores foram: Índice IEE, dólar comercial (venda), índice Ibovespa, valor do barril do petróleo, índice Nikkei, temperatura, precipitação e o próprio valor da ação das empresas. Para estes valores foram utilizados uma base de dados de 2006 a 2016, incluindo a média e média móvel de cada data e de cada indicador. Estes fatores foram ranqueados de forma que os primeiros são os que tiveram maior relevância no fechamento das ações.

**Palavras Chave:** Estimação de séries temporais, mercado de ações, redes neurais artificiais.

## **Abstract**

LEITE, L. O. G. (2016). Financial Stock Market Forecast using Data Mining for São Paulo's Power Utilities. Major Thesis – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2016.

The main goal of this major thesis is to predict the closing, maximum and minimum value in two power utilities companies in the state of São Paulo based only on historic datas. The main concept is to estimate that during a certain event, a large number of trades will occur in a day. Due to these facts, many variables were considered: IEE index, dollar, Ibovespa index, Nikkei index, temperature, pluviometric index, oil price and a company's stock.

For these values, data from 2006 to 2016 was used, including the average and rolling average for each value. Ranked on top are the most relevant variables.

**Keywords:** Time series forecasting, stock Exchange, artificial neural networks.

## Lista de Figuras

FIGURA 3.1 – COMPARAÇÃO ENTRE NEURÔNIO BIOLÓGICO E NEURÔNIO ARTIFICIAL .....	15
FIGURA 3.2 – FUNÇÃO DE ATIVAÇÃO DO TIPO DEGRAU BIPOLAR. ....	16
FIGURA 3.3 – FUNÇÃO DE ATIVAÇÃO DO TIPO DEGRAU. ....	17
FIGURA 3.4 – FUNÇÃO DE ATIVAÇÃO DO TIPO LINEAR. ....	17
FIGURA 3.5 – FUNÇÃO DE ATIVAÇÃO DO TIPO TANGENTE HIPERBÓLICA. ....	18
FIGURA 3.6 – FUNÇÃO DE ATIVAÇÃO DO TIPO LOGÍSTICA (SIGMÓIDE). ....	18
FIGURA 3.7 – FUNÇÃO DE ATIVAÇÃO DO TIPO RAMPA. ....	19
FIGURA 3.8 – FUNÇÃO DE ATIVAÇÃO DE BASE RADIAL (GAUSSIANA). ....	19
FIGURA 4.1 – FLUXOGRAMA DA METODOLOGIA PROPOSTA PARA A PREVISÃO DO PREÇO DAS AÇÕES DE CONCESIONÁRIAS DE ENERGIA ELÉTRICA DO ESTADO DE SÃO PAULO. ....	22
FIGURA 4.2 – FLUXOGRAMA DA METODOLOGIA PROPOSTA PARA A PREVISÃO DO PREÇO DAS AÇÕES DE CONCESIONÁRIAS DE ENERGIA ELÉTRICA DO ESTADO DE SÃO PAULO. ....	22



## Lista de Tabelas

TABELA 4.1 – ORGANIZAÇÃO DOS DADOS (ATRIBUTOS CONTIDOS NA BASE DE DADOS).....	25
TABELA 4.2 – RESULTADOS DA SELEÇÃO DE ATRIBUTOS PARA A CPFL.....	26
TABELA 4.3 – RESULTADOS DA SELEÇÃO DE ATRIBUTOS PARA A ELETROPAULO. ....	27
TABELA 4.8 – RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA CPFL: DADOS SEM SELEÇÃO DE ATRIBUTOS E RNA COM 2 CAMADAS NEURAI INTERMEDIÁRIAS. ....	32
TABELA 4.10 – RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA ELETROPAULO: DADOS SEM SELEÇÃO DE ATRIBUTOS E RNA COM 1 CAMADA NEURAL INTERMEDIÁRIA.....	34
TABELA 4.12 – RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA ELETROPAULO: DADOS SEM SELEÇÃO DE ATRIBUTOS E RNA COM 2 CAMADAS NEURAI INTERMEDIÁRIAS. ....	35
TABELA 4.13 – RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA ELETROPAULO: DADOS COM SELEÇÃO DE ATRIBUTOS E RNA COM 2 CAMADAS NEURAI INTERMEDIÁRIAS. ....	36

# Sumário

<b>RESUMO</b>	<b>V</b>
<b>ABSTRACT</b>	<b>VI</b>
<b>LISTA DE FIGURAS</b> .....	<b>VII</b>
<b>LISTA DE TABELAS</b> .....	<b>VIII</b>
<b>CAPÍTULO 1 INTRODUÇÃO AO TRABALHO DE CONCLUSÃO DE CURSO</b> .....	<b>1</b>
1.1 MOTIVAÇÃO E JUSTIFICATIVA AO TEMA DO TRABALHO .....	1
1.2 OBJETIVOS .....	2
1.3 ORGANIZAÇÃO .....	3
<b>CAPÍTULO 2 PREVISÃO DE SÉRIES TEMPORAIS</b> .....	<b>5</b>
2.1 SÉRIES TEMPORAIS .....	5
2.2 MODELOS AUTOREGRESSIVOS E DE MÉDIAS MÓVEIS .....	5
2.3 MODELOS BASEADOS EM SISTEMAS INTELIGENTES .....	7
2.4 BOLSA DE VALORES.....	8
2.4.1 BOVESPA.....	8
2.4.2 NIKKEI.....	9
2.4.3 ÍNDICES DE AÇÕES .....	9
2.5 PRECIPITAÇÃO E TEMPERATURA.....	10
2.6 COTAÇÃO PETRÓLEO.....	10
<b>CAPÍTULO 3 MINERAÇÃO DE DADOS APLICADA À PREVISÃO DE SÉRIES TEMPORAIS</b> .....	<b>12</b>
3.1 FUNDAMENTOS DE MINERAÇÃO DE DADOS .....	12
3.1.1 SELEÇÃO DE ATRIBUTOS .....	12
3.2 REDES NEURAIS ARTIFICIAIS .....	13
3.2.1 ESTRUTURA DO CÉREBRO.....	13
3.2.2 NEURÔNIO ARTIFICIAL .....	14
3.2.3 FUNÇÕES DE ATIVAÇÃO .....	16
3.2.4 ARQUITETURAS DE REDES NEURAIS ARTIFICIAIS .....	19
3.3 REDE NEURAL PROPOSTA PARA PREVISÃO DAS AÇÕES .....	20
<b>CAPÍTULO 4 PREVISÃO DE AÇÕES DE CONCESSIONÁRIAS DE ENERGIA ELÉTRICA DO ESTADO DE SÃO PAULO</b> .....	<b>22</b>
4.1 RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA CPFL.....	30
4.2 RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA ELETROPAULO.....	33
<b>CAPÍTULO 5 CONCLUSÃO</b> .....	<b>38</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>39</b>



# Capítulo 1

## Introdução ao Trabalho de Conclusão de Curso

### 1.1 MOTIVAÇÃO E JUSTIFICATIVA AO TEMA DO TRABALHO

A utilização de mineração de dados e séries temporais vem sendo um tema de pesquisa hoje em dia de grande relevância ao mercado como um todo. Empresas estão fazendo o uso de mineração de dados para rastrear perfil de consumo.

A inteligência competitiva tem o propósito de realizar essa monitoração informacional permitindo às empresas agirem de forma antecipada e reduzir os riscos de operação. Os Sistemas de Apoio à Decisão, dos quais faz parte a Mineração de Dados têm grande potencial para dar suporte a essas atividades, facilitando no processamento, obtenção e disseminação de informações gerenciais, uma vez que a Mineração de Dados tem o intuito de descobrir conhecimento não implícito em bancos de dados (Reis, 2011).

A informação é crucial para que uma organização sobreviva, seja para conquistar novos clientes, seja para se sobressair sobre os concorrentes, ou ainda para melhorar seus processos. Informação passou a ser um recurso da empresa, tal qual os recursos humanos, materiais e financeiros, e juntamente com o conhecimento científico e tecnológico tornou-se o centro das atenções das empresas em novos projetos (Valentim, 2003)

No caso específico de previsão de séries temporais relacionadas ao mercado acionário, é importante prever com maior precisão a tendência do movimento (subida ou descida) da série, visto que tais características podem auxiliar na estratégia de compra e venda das mesmas. Preços de ativos financeiros dependem, em geral, de outros preços do mercado. A volatilidade também pode depender de outros eventos, como anúncios governamentais de taxas de desemprego, de inflação, PIB trimestral, taxa de juros, entre outros. (Amorim, 2008).

Tendo em vista os trabalhos antecédidos, foi notável a preocupação com os dados escolhidos para análise, porém também é importante saber que há interferência externa, ainda mais quando se trata de ações de mercado financeira, onde são muito voláteis à ações do estado, de investimentos de terceiros ou de eventuais problemas, como por exemplo que de ações por envolvimento em corrupção.

Com estas condições, surge uma oportunidade de se pesquisar possíveis variáveis consideradas relevantes ao fechamento do preço de ações de concessionárias de energia elétrica.

## 1.2 OBJETIVOS

Este trabalho tem por objetivo, através da utilização de Redes Neurais, estimar o preço de fechamento, máximo e mínimo das ações de concessionárias de energia elétrica do estado de São Paulo. Para isto, foram considerados diversos fatores, que acreditava ser relevante ao valor das ações. Também é apresentado neste trabalho, dentre os fatores considerados, quais são os mais relevantes para cada empresa. Os fatores elencados, foram:

- Preço máximo
- Preço mínimo
- Preço de fechamento;
- Temperatura;
- Precipitação;
- Índice Nikkei;
- Cotação do Barril de Petróleo;
- Índice de Energia Elétrica (IEE);
- Índice da Bovespa (Ibovespa);
- Cotação do dólar americano.

Para todos estes dados, foram levantados seus valores a partir de 2006 até 2016, tentando dar uma maior fidelidade aos resultados.

Este trabalho teve por finalidade o estudo principalmente dos dados que são relevantes para o valor das ações. Sendo assim, foi considerado somente empresas de capital aberto, de mesmo perfil e do estado de São Paulo. Com isso, foram selecionadas as empresas CPFL e Eletropaulo.

Cabe destacar que foram selecionadas somente concessionárias de distribuição de energia elétrica do estado de São Paulo.

### **1.3 ORGANIZAÇÃO**

O trabalho está organizado da seguinte forma:

No Capítulo 2 são tratados os conceitos e fundamentos referentes às séries temporais, incluindo modelos de previsão. Além disso, é dada uma descrição básica do mercado acionário da bolsa de valores.

No Capítulo 3 os conceitos fundamentais das técnicas para seleção de atributos e das RNAs são apresentados.

O Capítulo 4 é destinado a mostrar os aspectos da metodologia proposta para a devida previsão das ações das concessionárias de energia elétrica paulistas.

No Capítulo 5, discute-se a metodologia empregada e os resultados para a previsão das ações analisadas neste trabalho.

Por fim, no Capítulo 6 são dadas as conclusões gerais e os possíveis trabalhos que podem ser desenvolvidos no futuro.



# Capítulo 2

## Previsão de Séries Temporais

Grande parte dos métodos de previsão tem como base uma série histórica. Esta deve ser composta do maior número possíveis de dados, sendo de extrema importância ter o maior espaço amostral possível, tendo assim uma resposta fiel.

Este capítulo, serão tratados os fundamentos da previsão de séries temporais e apresentado com mais detalhes as variáveis escolhidas para o treinamento desejado.

### 2.1 SÉRIES TEMPORAIS

Uma série temporal, também denominada série histórica, representa uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico. Este conjunto de dados pode ser obtido por meio de observações periódicas do evento de interesse (Latorre e Cardoso, 2001).

Tendo em vista a utilização em diversos meios de produção, a série histórica está se adaptando ao mercado por conta da sua interdisciplinaridade, podendo ser usada desde a previsão de chuva ou não em uma determinada região agrícola, como o impacto que uma fusão de 2 empresas trará na sua ação no dia seguinte. Em Amorim (2008) são apresentados alguns exemplos com os seguintes dados: preços diários de ações, taxa mensal de desemprego, produção industrial, número semanal de casos de uma determinada doença, precipitação pluviométrica, temperatura diária e velocidade do vento.

### 2.2 MODELOS AUTOREGRESSIVOS E DE MÉDIAS MÓVEIS

Através de análise de dados prévios, no caso deste trabalho analisando 15 dias ( D-1, D-2, D-3..., D-15) prévios do fechamento, a previsão de séries temporais tem por objetivo encontrar alguma lógica ou comportamento dos dados analisados e através disso, estimar qual será o valor futuro D+1. Tendo em vista o objetivo do trabalho, utilizaremos a ferramenta de regressão

Atualmente, existem diversos modelos estatísticos que podem ser aplicados à previsão de séries temporais. Dentre estes modelos, pode-se destacar as abordagens baseadas em métodos Autoregressivos e de Médias Móveis (AR, MA e ARMA), bem como o



modelo Autoregressivo Integrado de Médias Móveis (ARIMA). Estes modelos são explanados em maiores detalhes nas subseções que seguem (Laboissiere, 2013).

Cabe dizer que este trabalho faz uso de séries temporais estacionárias, que é quando ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Na prática, a maioria das séries que encontramos apresentam algum tipo de não estacionaridade, por exemplo, tendência.

Dentre os modelos utilizados para a previsão de séries temporais estacionárias, o modelo Autoregressivo (AR) destaca-se por ser extremamente útil para tratar modelos estocásticos. O modelo AR pode ser representado por meio da Equação 2.1:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t \quad (2.1)$$

onde:

- $X_t$  corresponde à observação da série temporal no tempo  $t$ ;
- $\phi_p$  corresponde ao parâmetro real do modelo da série temporal no tempo  $t-p$ ; e
- $e_t$  representa o erro de eventos aleatórios que não podem ser tratados pelo modelo, sendo que  $E(e_t) = 0$  e  $\text{Var}(e_t) = \sigma^2$ .

Já o modelo de Média Móvel (MA) considera um processo linear  $\{X_t, t \in Z\}$  e pode ser representado pela Equação 2.2:

$$X_t = -\theta_1 X_{t-1} - \theta_2 X_{t-2} - \dots - \theta_q X_{t-q} + e_t \quad (2.2)$$

onde:

- $X_t$  corresponde à observação da série temporal no tempo  $t$ ;
- $\theta_q$  corresponde ao parâmetro real do modelo MA no tempo  $t-p$ ; e
- $e_t$  representa o erro de eventos aleatórios que não podem ser explicados pelo modelo, sendo que  $E(e_t) = 0$  e  $\text{Var}(e_t) = \sigma^2$ .

Além dos modelos supracitados, existem os modelos mistos, ou seja, aqueles que fazem uso tanto da autoregressão como da média móvel. Desta forma, foi criado o modelo Autoregressivo de Média Móvel (ARMA), o qual é representado pela Equação 2.3.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \dots - \theta_q X_{t-q} \quad (2.3)$$

Cabe comentar que os parâmetros destes modelos são os mesmos apresentados para as Equações 2.1 e 2.2. Desta forma, torna-se possível observar que os modelos ARMA relacionam os observações passadas com valores futuros, assim como os erros de estimações passadas.

Ressalta-se ainda que os modelos AR, MA e ARMA, são apropriados para descrever séries temporais estacionárias, isto é, séries que se desenvolvem no tempo ao redor de uma média constante.

Entretanto, nem todas as séries podem ser consideradas estacionárias, como é o caso de séries temporais encontradas na indústria, nos negócios e na economia (Box, Jenkins e Rensils, 2008). Com base nesta premissa, surgiu o modelo Autoregressivo Integrado de Média Móvel (ARIMA), o qual também é conhecido por modelo de Box-Jenkins.

O modelo ARIMA pode ser considerado como uma generalização do modelo ARMA, onde a principal consideração é feita ao se definir que o polinômio do modelo AR possui raiz unitária. Por meio desta consideração, foi observado que o modelo ARIMA poderia modelar de forma bastante razoável algumas séries temporais não estacionárias. Com o intuito de melhor representar o modelo ARIMA, o mesmo é descrito pela Equação 2.4.

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 X_{t-1} - \theta_2 X_{t-2} - \dots - \theta_q X_{t-q} \quad (2.4)$$

## 2.3 MODELOS BASEADOS EM SISTEMAS INTELIGENTES

As Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura geral de organismos inteligentes e que adquirem conhecimento através da experiência. No capítulo 3, será apresentado as habilidades das RNAs de mapear padrões, clusterização, identificação, aproximação funcional, predição e estimação.

Algumas possíveis utilizações do uso de redes neurais em mercado de ação é a estimação do fechamento da ação futura para assim poder fazer a compra ou venda. Empresas e corporações também utilizam para fazer uma previsão de seu cenário, como fluxo de caixa, demonstrativo de resultado, valor da ação e cenário de venda por exemplo.

## 2.4 BOLSA DE VALORES

A bolsa de valores é o mercado onde se negociam ações de sociedades de capital aberto e outros valores imobiliários, tais como opções. Pode ser organizado por uma sociedade civil sem fins lucrativos, ou sociedade anônima visando lucro.

Visto que o mercado de capitais é um sistema de distribuição de valores mobiliários que visa proporcionar liquidez aos títulos de emissão de empresas e viabilizar seu processo de capitalização, os principais títulos negociados são os que representam o capital de empresas (as ações) ou empréstimos tomados, via mercado, pelas empresas. Portanto, o mercado de capitais permite a circulação de capital para custear o desenvolvimento econômico de um país, já que pode servir como fonte de renda de finanças pessoais, empresas e governos (Laboissiere, 2013).

Tendo em vista algumas das práticas do mercado de ações, há quem trabalhe exclusivamente com a compra e venda de ações de empresa. Existem dois tipos de investimentos, aqueles com rentabilidade fixa, geralmente praticado por quem não possui uma grande experiência com compra e venda de ações ou quem não quer arriscar no mercado de ações. O outro tipo são as ações que não se tem previsibilidade sobre o retorno, onde é preciso ter uma noção sobre o histórico da empresa, e o macro da economia, podendo porém, ter altos desempenhos no que diz respeito ao investimento inicial.

Conforme Amorim (2008), outro aspecto importante da bolsa de valores é que esta permite o intercâmbio entre os países, o qual está relacionado à troca de moedas, aos investimentos no exterior e aos investimentos em *commodities*. Desta forma, as empresas podem operar por meio da bolsa de valores as transações de suas mercadorias com a finalidade de buscar mecanismos que a protejam contra a oscilação dos preços de seus produtos.

### 2.4.1 BOVESPA

A BM&FBOVESPA (BVMF) foi criada em maio de 2008 com a integração da BM&F (bolsa de derivativos) e da Bovespa (bolsa de ações). Atualmente, é a única bolsa de valores mobiliários do Brasil. É também provedora tanto de serviços de contraparte central, como de custódia de títulos. Além disso, a BM&FBOVESPA ocupa posição destacada entre as maiores bolsas do mundo em valor de mercado e a segunda colocação no ranking das Américas (agosto de 2012).

A BOVESPA Bolsa de Valores, Mercadorias e Futuros está entre as maiores bolsas do mundo em valor de mercado e é líder na América Latina, sendo a principal instituição brasileira de intermediação para operação no mercado de capitais, a companhia desenvolve, implanta e provê sistemas para negociação de ações, derivativos de ações, derivativos financeiros, títulos de renda fixa, títulos públicos federais, moedas a vista e commodities agropecuárias.

O Ibovespa é o resultado de uma carteira teórica de ativos, elaborada de acordo com os critérios estabelecidos em sua metodologia e tem por objetivo ser o indicador do desempenho médio das cotações dos ativos de maior negociabilidade e representatividade do mercado de ações brasileiro.

## **2.4.2 NIKKEI**

A Bolsa de Valores de Tóquio, chamada de Tosho ou TSE é assim com a BOVESPA uma casa de compra e venda de ações de empresas de capital aberto. Ela possui 2.292 empresas listadas com um total de capitalização de US\$4,5 trilhões em novembro de 2013.

A TSE possui nove diretores, quatro auditores e oito oficiais executivos. Sua sede é localizada em Tóquio opera em horários diferentes do Brasil, das 8:00 às 11:30 e das 12:30 às 5:00. O Índice Nikkei é o principal rastreador do volume da TSE, de companhias selecionadas pelo maior jornal do Japão, com pontuação na casa das dezenas de milhares.

Um dos objetivos deste trabalho é o de descobrir indicadores que tenham relação com o valor da ação. Dado este objetivo o índice Nikkei foi adicional aos indicadores para se ter um estudo em cima do mesmo e analisar sua real relevância.

## **2.4.3 ÍNDICES DE AÇÕES**

Neste trabalho é feito o uso de dois índices das ações, o índice Ibovespa e o IEE (Índice de Energia Elétrica). Estes representam o comportamento dos preços das ações do segmento que representam no mercado. Estes índices são obtidos com base nos negócios realizados na bolsa de valores, assim cada índice possui uma metodologia de cálculo para quantificar sua variância no período. Analisando de uma ótica de mercado, estes índices servem como termômetro para se ter uma noção se as ações estão valorizando ou desvalorizando.

- Ibovespa - é o mais importante indicador de desempenho médio das cotações do mercado de ações brasileiro, pois, retrata o comportamento dos principais papéis

negociados na bolsa de valores brasileira (exemplo: Vale, Petrobrás, Embraer, etc);

- IEE (Índice de Energia Elétrica) - é um índice setorial que tem o objetivo de medir o desempenho do setor de energia elétrica brasileiro.

Estes dois índices são de extrema importância para o trabalho pois conseguem cobrir uma grande parte do mercado em análise, o Ibovespa já em um cenário macro, analisando o mercado acionário brasileiro, enquanto o IEE tendo uma visão específica do setor energético brasileiro.

## **2.5 PRECIPITAÇÃO E TEMPERATURA**

Este trabalho teve como foco a exploração de potenciais variáveis que possam ser relevantes no fechamento da ação das empresas do estado de São Paulo. Tendo em vista que a matriz é uma usina hidroelétrica, a qual provê a maior parte da energia elétrica para o estado de São Paulo, foi analisado o comportamento da temperatura e chuva na cidade de Santa Terezinha do Itaipú, cidade a qual está acima da usina de Itaipú, esta responsável pelo abastecimento.

Esta temperatura e índice pluviométrico, foi colhido de 3 em 3 horas, durante todo o período da amostra das ações. Tendo estes valores em mãos, foi feito uma média simples para se obter a temperatura e a precipitação diária requerida.

## **2.6 COTAÇÃO PETRÓLEO**

Outra variável importante no cenário global é o valor do barril de petróleo. Este impulsiona diversos segmentos do mercado, sendo utilizado como combustível, na construção civil, lubrificantes, etc.

Esta matéria prima muitas vezes dita o comportamento de ações de determinados segmentos do mercado, dada sua extrema importância na maioria dos segmentos do mercado. Tendo isto em vista, foi utilizado a sua cotação de de máximo, mínimo e fechamento, para se ter uma noção do que se passava no cenário global do petróleo.



# CAPÍTULO 3

## Mineração de Dados Aplicada à Previsão de Séries Temporais

### 3.1 FUNDAMENTOS DE MINERAÇÃO DE DADOS

A mineração de dados segundo a definição de Berestein é um método aplicado na base de dados para a extração de conhecimento novo. Também deve-se ter ciência de que o processo de mineração de dados é apenas uma parte de um processo chamado KDD, que significa Descoberta de Conhecimento em bases de dados ou em inglês Knowledge Discovery Databases.

No caso deste trabalho, para a previsão de fechamento da ação de concessionárias de energia elétrica do estado de São Paulo, o primeiro passo foi listar quaisquer variáveis que fossem consideradas relevantes no fechamento da ação. Após esta listagem, foi analisada uma por uma e assim descartadas as que eram consideradas menos relevantes, tendo no final um número de 8 variáveis.

A mineração de dados entra como um aliado nesta etapa, dado que se torna complexa a análise de grandes quantidades de dados que representam séries temporais. Este trabalho seguirá a definição de Han e Kamber (2001) devido à metodologia proposta de empregar um algoritmo para Seleção de Atributos em sua etapa inicial, onde são descobertas as variáveis com maior relevância, com a utilização do software Weka, à obtenção do padrão comportamental das ações. Ademais, será utilizada uma mineração preditiva, onde são utilizados dados que representam situações passadas do processo/sistema. Assim, por meio destes dados, pretende-se estimar situações atuais e/ou futuras do processo/sistema (Laboissiere, 2013).

#### 3.1.1 SELEÇÃO DE ATRIBUTOS

Da definição de atributos, tem-se que é uma qualidade associada à cada elemento, no caso do estudo, a base de dados. Tendo isto em visto, deve-se utilizar algumas ferramentas na seleção destas características. Para este treinamento, foi utilizado a seleção de atributos Relief-f.

Este treinamento foi proposto por Kira e Rendell em 1992 e tem por objetivo identificar os atributos que exibam valores diferentes para exemplos próximos de classes distintas e valores similares para exemplos de mesma classe. Para sua utilização, deve-se ter a base inteira de dados normalizada, variando seu valor ao mínimo 0 e máximo 1.

No início de cada uma das  $c$  iterações do algoritmo, um exemplo  $E_i$  é selecionado randomicamente do conjunto de dados e se identifica o exemplo mais próximo de classe distinta de  $E_i$ , denotado por  $S$ , e o mais próximo de classe idêntica ao  $E_i$ , denotado por  $A$ . Uma aproximação do valor de importância ( $R$ ) calculado pelo Relief para cada atributo  $X_j$  é dada pela equação:

$$R(X_j) = P(\text{Valor distinto de } X_j|S) - P(\text{valor distinto de } X_j|A)$$

Ao contrário do algoritmo Relief original, o Relief-f é aplicável a problemas monorrótulo com mais de duas classes, isto é, quando  $|L|$  é maior ou igual a 2, e apresenta maior robustez a dados ruidosos e incompletos.

## **3.2 REDES NEURAIS ARTIFICIAIS**

Redes Neurais Artificiais (RNA) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. Uma grande rede neural pode ter centenas ou milhares de unidades de processamento; já o cérebro de um mamífero pode ter muitos bilhões de neurônios.

### **3.2.1 ESTRUTURA DO CÉREBRO**

O cérebro humano é considerado o mais fascinante processador baseado em carbono existente, sendo composto por aproximadamente 10 bilhões neurônios. Todas as funções e movimentos do organismo estão relacionados ao funcionamento destas pequenas células (Tatibana e Kaetsu). Os neurônios estão conectados uns aos outros através de sinapses, e juntos formam uma grande rede, chamada REDE NEURAL. As sinapses transmitem estímulos através de diferentes concentrações de  $\text{Na}^+$  (Sódio) e  $\text{K}^+$  (Potássio), e o resultado disto pode ser estendido por todo o corpo humano. Através desta rede, o corpo apresenta uma capacidade de armazenamento e processamento de informação.

O sistema nervoso é formado por um conjunto extremamente complexo de neurônios. Nos neurônios a comunicação é realizada através de impulsos, quando um impulso é recebido, o neurônio o processa, e passado um limite de ação, dispara um segundo impulso que produz



uma substância neurotransmissora o qual flui do corpo celular para o axônio (que por sua vez pode ou não estar conectado a um dendrito de outra célula). O neurônio que transmite o pulso pode controlar a frequência de pulsos aumentando ou diminuindo a polaridade na membrana pós-sináptica. Eles têm um papel essencial na determinação do funcionamento, comportamento e do raciocínio do ser humano. Ao contrário das redes neurais artificiais, redes neurais naturais não transmitem sinais negativos, sua ativação é medida pela frequência com que emite pulsos, frequência esta de pulsos contínuos e positivos. As redes naturais não são uniformes como as redes artificiais, e apresentam uniformidade apenas em alguns pontos do organismo. Seus pulsos não são síncronos ou assíncronos, devido ao fato de não serem contínuos, o que a difere de redes artificiais.

Os principais componentes dos neurônios são:

- Os dentritos, que tem por função, receber os estímulos transmitidos pelos outros neurônios;
- O corpo de neurônio, também chamado de soma, que é responsável por coletar e combinar informações vindas de outros neurônios;
- E finalmente o axônio, que é constituído de uma fibra tubular que pode alcançar até alguns metros, e é responsável por transmitir os estímulos para outras células.

### 3.2.2 NEURÔNIO ARTIFICIAL

As RNAs são modelos simplificados da arquitetura cerebral, haja visto que estas possuem algumas centenas de neurônios, enquanto o cérebro apresenta em torno de 100 bilhões de neurônios. Outra diferença está no fato de o cérebro utilizar os 5 sentidos como entrada e as RNAs necessitarem, obrigatoriamente, de dados numéricos.

Por meio da Figura 3.1, uma comparação entre o neurônio biológico e o neurônio artificial pode ser visualizada. Nota-se por meio desta figura que o neurônio artificial pode ser dividido nos seguintes elementos básicos: entradas, combinador linear (soma) e função de ativação  $g(\cdot)$ . Na sequência, cada um destes elementos são explanados em maiores detalhes:

1. Sinais de entrada  $\{x_1, x_2, x_3, \dots, x_n\}$  - cada um ponderado pelo seu peso sináptico  $\{w_1, w_2, w_3, \dots, w_n\}$ , os quais são provenientes do meio externo e representam os valores assumidos pelas variáveis de uma aplicação específica. O conjunto de

entradas é usualmente normalizado visando incrementar a eficiência computacional da RNA (Silva, Spatti e Flauzino, 2010);

2. Combinador linear (representado pelo somatório da Figura 3.1) - realiza a soma da multiplicação dos sinais de entrada pelos seus respectivos pesos sinápticos a fim de gerar um valor de potencial de ativação (Ferreira, 2011);
3. Função de ativação  $g(\cdot)$  - objetiva gerar uma saída que esteja dentro de um intervalo de valores válidos. As funções de ativação serão apresentadas na subseção 3.1.3.

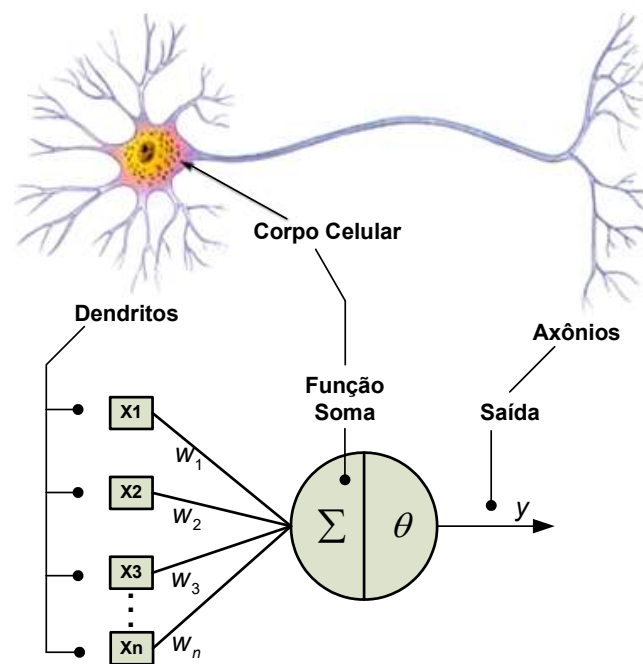


Figura 3.1 – Comparação entre neurônio biológico e neurônio artificial.  
Fonte: (Fernandes, 2009).

Por meio dos três elementos básicos supracitados, a saída de um neurônio artificial pode ser expressa conforme a Equação 3.1:

$$y = g\left(\sum_{i=1}^n w_i x_i - \theta\right) \quad (3.1)$$

onde:

$n$  representa o número de entradas do neurônio;

$w_i$  é o peso sináptico associado à  $i$ -ésima entrada;

$\theta$  é o limiar de ativação do neurônio;

$x_i$  é a  $i$ -ésima entrada do neurônio;

$g(\cdot)$  é a função de ativação do neurônio;

$y$  é a saída do neurônio.

### 3.2.3 FUNÇÕES DE ATIVAÇÃO

Conforme previamente citado, a função de ativação fornece um valor de saída limitado. Segundo Silva, Spatti e Flauzino (2010), as funções de ativação são divididas em dois grupos: parcialmente diferenciáveis e as totalmente diferenciáveis.

Dentre as funções de ativação parcialmente diferenciáveis, destacam-se: degrau, degrau bipolar e rampa simétrica. Já as funções de ativação totalmente diferenciáveis que se destacam em redes neurais artificiais são: logística, tangente hiperbólica, gaussiana e linear.

Algumas das funções de ativação mais utilizadas têm seus cálculos e formatos gráficos mostrados nas Figuras de 3.2 a 3.8:

- Degrau bipolar

$$y(n) = \begin{cases} y = 1, & \text{se } n > 0 \\ y = 0, & \text{se } n = 0 \\ y = -1, & \text{se } n < 0 \end{cases} \quad (3.2)$$

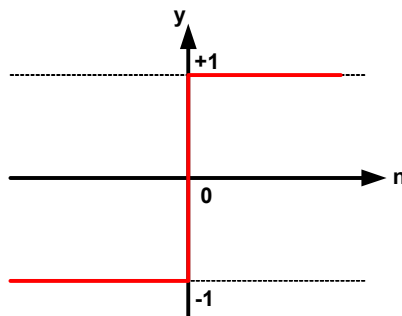


Figura 3.2 – Função de ativação do tipo degrau bipolar.

Fonte: (Fernandes, 2009).

- Degrau

$$y(n) = \begin{cases} y = 1, & \text{se } n \geq 0 \\ y = 0, & \text{se } n < 0 \end{cases} \quad (3.3)$$

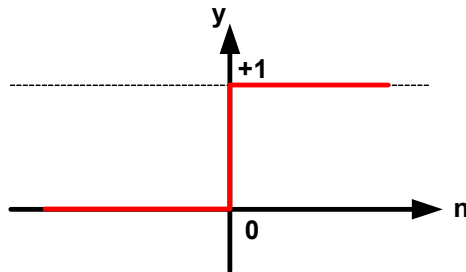


Figura 3.3 – Função de ativação do tipo degrau.

Fonte: (Fernandes, 2009).

- Linear

$$y(n) = n \quad (3.4)$$

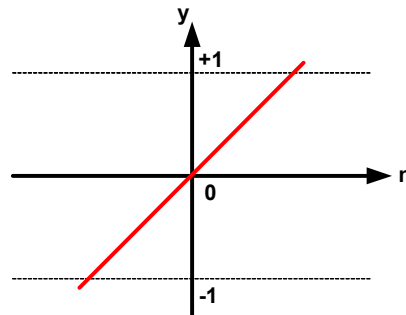


Figura 3.4 – Função de ativação do tipo linear.

Fonte: (Fernandes, 2009).

- Tangente hiperbólica

$$y(n) = \frac{1 - e^{-\beta n}}{1 + e^{\beta n}} \quad (3.5)$$

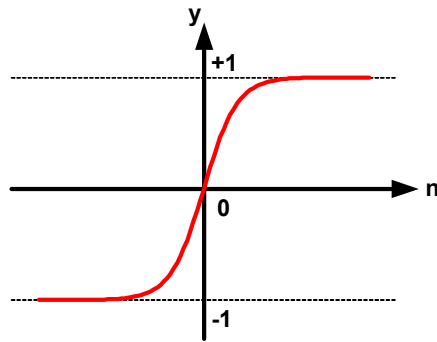


Figura 3.5 – Função de ativação do tipo tangente hiperbólica.

Fonte: (Fernandes, 2009).

onde  $\beta$  é o parâmetro de excentricidade.

- Logística (sigmóide)

$$y(n) = \frac{1}{1 + e^{-\beta n}} \quad (3.6)$$

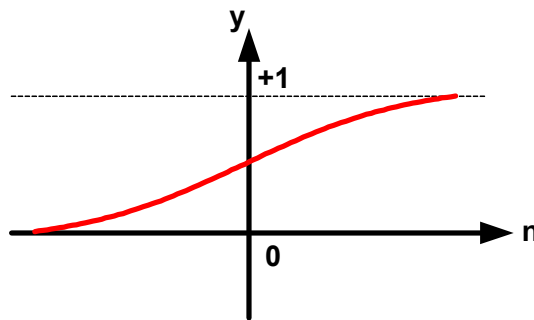


Figura 3.6 – Função de ativação do tipo logística (sigmóide).

Fonte: (Fernandes, 2009).

onde  $\beta$  é o parâmetro de excentricidade.

- Rampa

$$y(n) \begin{cases} y = 1, & \text{se } n > 1 \\ y = n, & \text{se } -1 \leq n \leq 1 \\ y = -1, & \text{se } n < 0 \end{cases} \quad (3.7)$$

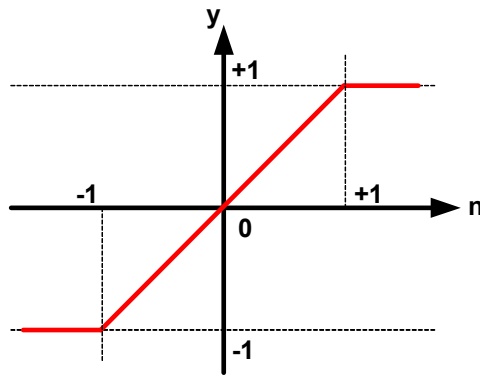


Figura 3.7 – Função de ativação do tipo rampa.

Fonte: (Fernandes, 2009).

- Base radial (gaussiana)

$$y(n) = e^{-\left(\frac{(\sigma n)^2}{2\sigma^2}\right)} \quad (3.8)$$

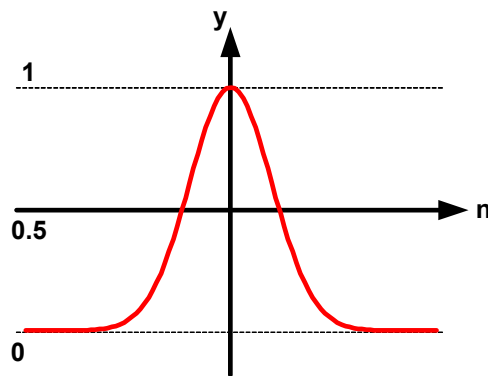


Figura 3.8 – Função de ativação de base radial (gaussiana).

Fonte: (Fernandes, 2009).

### 3.2.4 ARQUITETURAS DE REDES NEURAIS ARTIFICIAIS

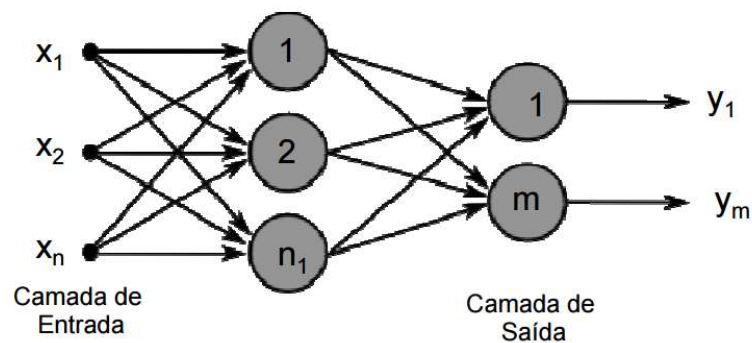
A arquitetura de RNA define a forma com a qual seus diversos neurônios estão dispostos e o direcionamento das conexões sinápticas e sua topologia define composições estruturais, como por exemplo a quantidade de neurônios e quais as funções de ativação.

A rede é basicamente dividida em três partes: Camada de entrada, camadas intermediárias e camada de saída.

- A camada de entrada são todas as entradas; dados, sinais, características ou mediações. São geralmente normalizadas (mantendo seu valor de 0 a 1) em relação a faixa de variação produzida pela função de ativação.

- Já as camadas intermediárias, ou também chamadas de escondidas, ocultas ou invisíveis são realizadas as “extrações” das características da série e inferência da mesma.
- Por último está a camada de saída, a qual é responsável pela produção e apresentação do resultado final.

Quanto ao número de camadas, pode-se categorizá-las em camada única ou múltiplas camadas. Em relação ao tipo de conexão, também são de dois tipos; totalmente conectados ou parcialmente conectados. A topologia pode ser dividida entre Feedforward ou Feedback e neste trabalho usaremos a topologia Feedforward, com múltiplas camadas e neurônios totalmente conectados.



*Figura 3.9 – Feedforward de Camadas Múltiplas*  
 Fonte: *Inteligência Computacional (Aula 2; RNA – Arquiteturas e Treinamento)*

Feedforward de Camadas Múltiplas apresenta uma ou mais camadas escondidas. São empregadas na aproximação de funções, classificação de padrões, otimização e controle de processos.

### 3.3 REDE NEURAL PROPOSTA PARA PREVISÃO DAS AÇÕES

Tendo em visto que o problema proposto é relacionado à aproximação funcional, se paração de mais do que duas classes de dados e estimação/previsão de processos ou sistemas, as redes neurais do tipo Perceptron e Adaline não podem ser empregadas. Assim, neste trabalho são utilizadas redes neurais artificiais de múltiplas camadas do tipo MLP (Multilayer Perceptron).

Para este treinamento, será utilizado o método de Levenberg-Marquardt. Assim como o método de Newton, este é bastante eficiente quando estamos tratando de redes que não possuem mais do que algumas centenas de conexões a serem ajustadas (HAGAN, 1994). Isto deve-se, principalmente, ao fato de que estes algoritmos necessitam armazenar uma matriz quadrada cuja dimensão é da ordem do número de conexões da rede.

Destaca-se que para o treinamento de uma rede MLP é usado o método de treinamento supervisionado, ou seja, são apresentadas à rede a saída desejada correspondente a um conjunto de dados de entrada. Assim, é possível notar que o principal algoritmo de treinamento para redes do tipo MLP é o *backpropagation*. Conforme Araújo e Lotufo (2009), esse método apresenta um elevado esforço computacional, pois, considera as derivadas de primeira ordem nos ajustes dos valores das matrizes de pesos. Em contrapartida, o algoritmo de treinamento de Levenberg-Marquardt emprega derivadas de segunda ordem, assim, o tempo de convergência do mesmo pode ser reduzido e, por consequência, há a redução do esforço computacional (Silva, Spatti e Flauzino, 2010).

Tendo analisado alguns trabalhos realizados com este tipo de treinamento (Levenberg-Marquardt) como base, foi notável o sucesso para com o que foi proposto, o que reforça o uso do mesmo.



## Capítulo 4

# Previsão de Ações de Concessionárias de Energia Elétrica do Estado de São Paulo

Este trabalho foi conduzido, assim como pode-se ver na Figura 4.1, com a ajuda de um fluxograma, onde tem-se uma noção cronológica do trabalho. Nele é possível observar os passos iniciais que começam na coleta de dados históricos, no caso deste trabalho desde 2006 até 2016 e também os passos finais, onde se obtém a previsão dos preços das ações.

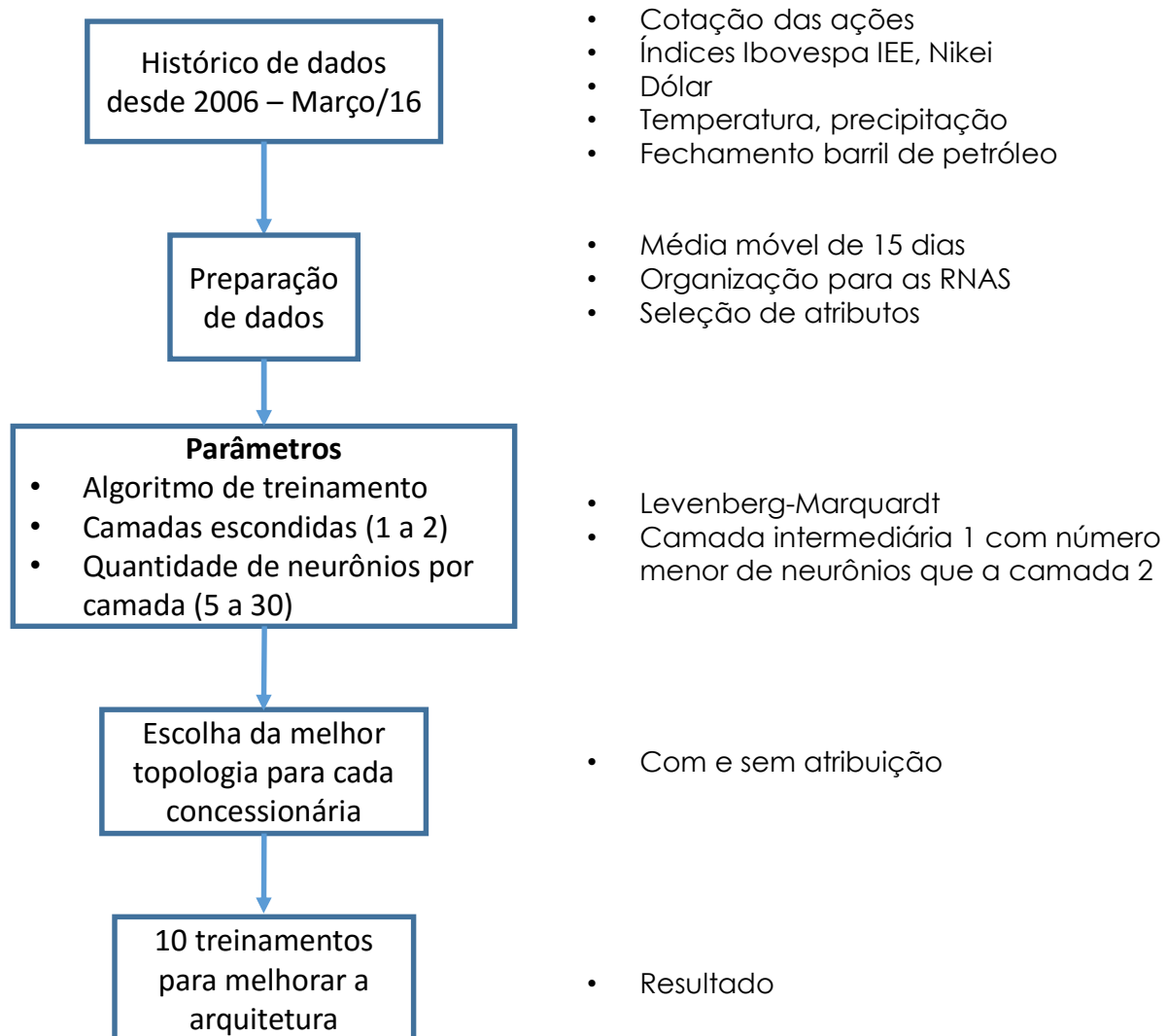


Figura 4.1 – Fluxograma da metodologia proposta para a previsão do preço das ações de concessionárias de energia elétrica do estado de São Paulo.

Partindo do primeiro bloco “Histórico de dados”, pode-se lembrar que foram obtidos e manipulados diversos dados como:

- Cotação das ações: Foram obtidos os dados do valor de fechamento, valor de compra máximo e valor de compra mínimo no período requerido para as empresas CPFL e Eletropaulo. Na Figura 4.2 pode-se notar o comportamento histórico da CPFL (CPFE3.SA) que temos desde 28 de Dezembro de 2006 até 03 de Março de 2016.

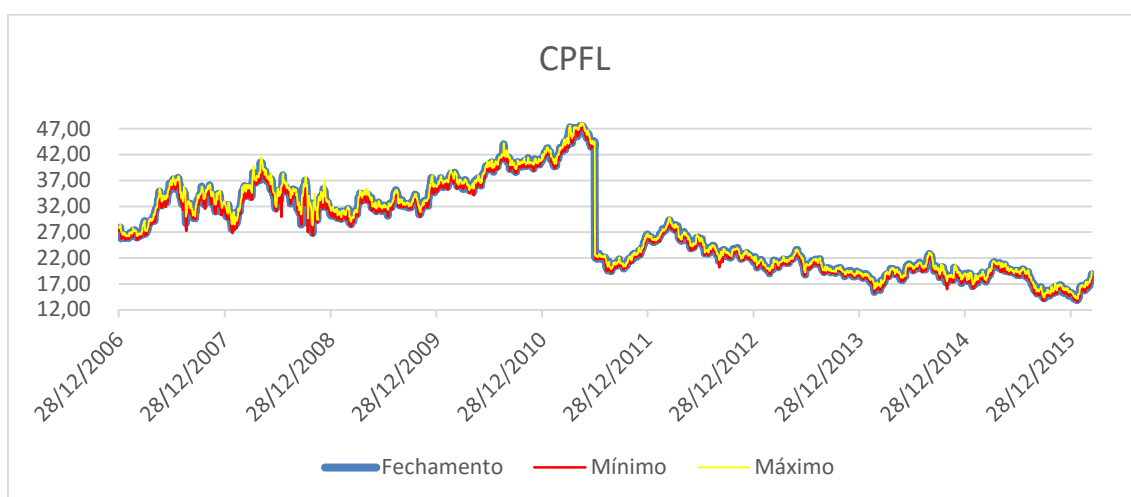


Figura 4.2 – Histórico de cotação da empresa CPFL, fornecido pela UOL Economia

Temos na Figura 4.3 o equivalente, porém para a empresa Eletropaulo (ELPL4.SA). Neste caso foi possível ter o histórico a partir de 26 de Dezembro de 2008, até 10 de março de 2016

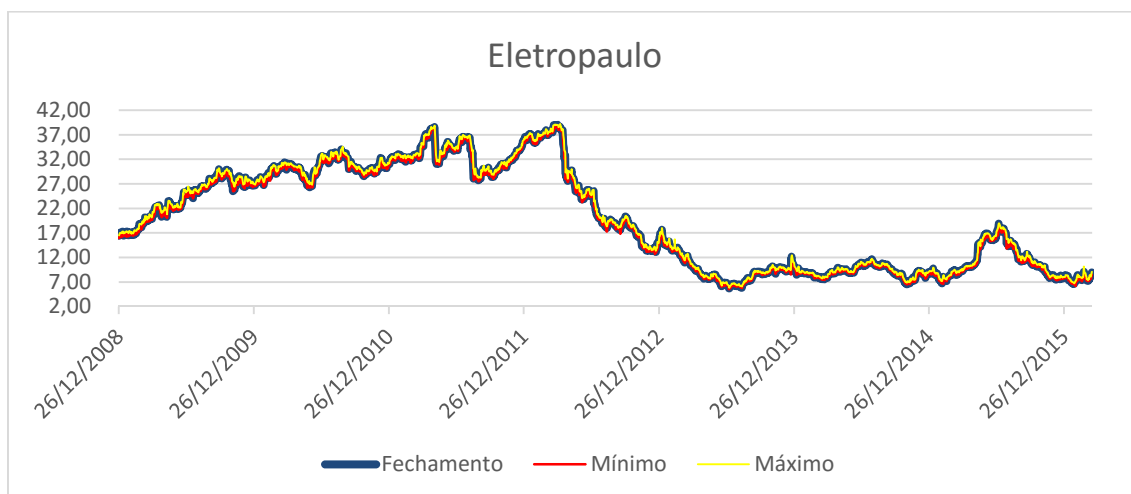


Figura 4.3 – Histórico de cotação da empresa Eletropaulo, fornecido pelo UOL Economia

- Índices IEE, Ibovespa e Nikkei: Os três índices, assim como outros indicadores, foram obtidos através da série histórica disponibilizada pela seção de economia no site da UOL (UOL, 2016).
- Obtenção dos valores históricos de cotação do dólar americano no site do Banco Central do Brasil (Banco Central do Brasil, 2016).
- Temperatura e Precipitação: Estes dados foram obtidos através de uma solicitação ao Instituto Nacional de Pesquisas Espaciais pois a cidade de Santa Terezinha de Itaipu-PR não constava em sites de pesquisa usuais.
- Cotação de Petróleo: Obtido dos valores históricos de cotação do barril do petróleo no site do investing, na seção de commodities, consultado no mês de Março de 2016 (br.investing, 2016).

Todos estes dados foram tratados para que pudessem ser comparáveis, como por exemplo o ajuste da data do índice Nikkei. Por conta de se tratar de um indicador do Japão, o qual possui um calendário diferente do brasileiro, sua data foi alterada com o intuito de poder ser comparáveis.

Na sequência, os dados foram pré-processados, onde foram feitos cálculos de Média Móvel de 15 dias e organização dos dados para que fossem devidamente submetidos ao Seletor de Atributos.

O objetivo deste cálculo é filtrar eventuais flutuações demasiadas e mostrar mais explicitamente a tendência do mercado de ações. Neste trabalho, foi usada a média móvel ponderada para  $n$  termos, cuja notação matemática é expressa pela Equação 4.1:

$$M_{Móvel}(n) = \frac{2}{n(n+1)} \sum_{i=1}^n P[t - (n-i)](i) \quad (4.1)$$

Destaca-se que foi usado  $n = 15$  devido ao número de dias da média móvel. Assim, o objetivo desta média móvel é a de fornecer uma avaliação da série temporal em médio prazo. Cabe ainda comentar que as médias móveis foram calculadas para os índices (Ibovespa, Nikkei e IEE), cotação do dólar americano, para as ações (valores máximos, mínimos e fechamento), para a temperatura, precipitação e para o valor do barril de petróleo.

Após esta etapa de pré-processamento, os dados foram devidamente organizados e separados em dois conjuntos de amostras (treinamento e validação). A organização destes dados pode ser observada por meio da Tabela 4.1.

Deve-se entender da Tabela 4.1 que  $D$  representa o dia para o qual deseja-se estimar os valores dos preços máximos, mínimos e fechamento, ou seja, a saída desejada da rede neural. Para a metodologia proposta foi considerado apenas o preço de fechamento de  $D$  como possível entrada do método de predição (RNA).

Tabela 4.1 – Organização dos dados (atributos contidos na base de dados).

$D$	Preço de fechamento
$D - i$	Preço de fechamento
	Preço máximo
	Preço de mínimo
Média Móvel de 15 dias	Preço de máximo
	Preço de mínimo
	Preço de fechamento
	Temperatura
	Precipitação
	IEE
	Ibovespa
	Nikkei
	Dólar
	Preço Barril de Petróleo
Saída desejada para $D$	Preço mínimo
	Preço máximo
	Preço Fechamento

Nota-se, por meio da Tabela 4.1, que existem os atributos dos  $D - i$  dias, onde  $i = 1, 2, \dots, 15$ , ou seja, quinze dias anteriores ao dia atual ( $D$ ).

<b>Ibovespa (D-15)</b>				
Fechamento	Mínimo	Máximo	Média	Média Móvel
41.478	41.183	41.698	41.453	40.421,6

Figura 4.4 – Exemplo do esquemático dos dados manipulados

Deste modo, é utilizado as variáveis já citadas acima porém com 15 dias anteriores e 3 variáveis de saída (Preço máximo, Preço mínimo e Preço de fechamento do dia  $D$ ) das ações. Recorda-se que o histórico de dados considerado é de Dezembro de 2006 até Março de 2016.

Tendo em mãos todas as variáveis (atributos) supracitadas, o algoritmo de Seleção de Atributos foi aplicado ao conjunto de dados de treinamento das concessionárias de energia elétrica.

Os resultados das seleções de atributos para cada um dos conjuntos de treinamento podem ser visualizados na sequência por meio das Tabelas 4.2 e 4.3.

Na Tabela 4.2, são apresentados os resultados da seleção de atributos para a CPFL. Para os preços máximo, mínimo e fechamento. Para todos estes valores, foi selecionado as variáveis que apresentavam um valor de atributo inferior ou igual a 0,03.

*Tabela 4.2 – Resultados da seleção de atributos para a CPFL.*

<b>Atributos Selecionados</b>		
<b>Preço Máximo</b>	<b>Preço Mínimo</b>	<b>Preço Fechamento</b>
Temperatura média do dia D-3	CPFL fechamento do dia D-1	Temperatura média do dia D-11
Temperatura média do dia D-5	Temperatura média do dia D-5	Temperatura média do dia D-7
CPFL média móvel do dia D-1	Temperatura média do dia D-8	CPFL fechamento do dia D-1
CPFL fechamento do dia D-1	-	Temperatura média do dia D-8
-	-	Temperatura média do dia D-14
-	-	Temperatura média do dia D-1
-	-	Temperatura média do dia D-2
-	-	Temperatura média do dia D-10
-	-	Temperatura média do dia D-9
-	-	Temperatura média do dia D-3

Na Tabela 4.3, são apresentados os resultados da seleção de atributos para a Eletropaulo. Pode-se notar que para ambas empresas, o valor da temperatura dos dias anteriores foram de alta correlação, tanto para o preço máximo, mínimo e para o preço de fechamento.

Tabela 4.3 – Resultados da seleção de atributos para a Eletropaulo.

<b>Atributos Selecionados</b>		
<b>Preço Máximo</b>	<b>Preço Mínimo</b>	<b>Preço Fechamento</b>
Eletropaulo média móvel do dia D-1	Temperatura média do dia D-13	Temperatura média do dia D-13
Temperatura média do dia D-3	Temperatura média do dia D-1	Temperatura média do dia D-6
Eletropaulo média móvel do dia D-2	Temperatura média do dia D-12	Temperatura média do dia D-3
Temperatura média do dia D-14	Temperatura média do dia D-7	Temperatura média do dia D-8
Eletropaulo média móvel do dia D-3	Temperatura média do dia D-10	Temperatura média do dia D-1
Temperatura média do dia D-10	Temperatura média do dia D-8	Temperatura média do dia D-11
Eletropaulo média móvel do dia D-4	Temperatura média do dia D-6	Temperatura média do dia D-7
Temperatura média do dia D-13	Temperatura média do dia D-3	Temperatura média do dia D-2
Temperatura média do dia D-12	Temperatura média do dia D-4	Temperatura média do dia D-5
Temperatura média do dia D-7	Temperatura média do dia D-2	Temperatura média do dia D-12
Eletropaulo valor máximo do dia D-1	Temperatura média do dia D-5	Temperatura média do dia D-9
Eletropaulo valor máximo do dia D-2	Temperatura média do dia D-11	Temperatura média do dia D-4
Eletropaulo valor máximo do dia D-3	Temperatura média do dia D-9	Temperatura média do dia D-10
Eletropaulo valor médio do dia D-1	Temperatura média do dia D-15	Temperatura média do dia D-14
Eletropaulo valor médio do dia D-2	Precipitação média do dia D-15	Precipitação média do dia D-15
Eletropaulo valor médio do dia D-3	-	-
Eletropaulo valor média móvel do dia D-5	-	-
Temperatura média do dia D-9	-	-
Temperatura média móvel do dia D-11	-	-
Temperatura média do dia D-8	-	-
Temperatura média do dia D-11	-	-
Temperatura média do dia D-6	-	-
Eletropaulo valor média móvel do dia D-6	-	-
Temperatura média do dia D-5	-	-

A partir dos atributos selecionados, as RNAs foram configuradas da seguinte forma:

- Algoritmo de treinamento: Levenberg-Marquardt;
- Erro quadrático médio:  $10^{-12}$ ;
- Número máximo de épocas: 500
- Número de camadas: topologias com 1 e 2 camadas neurais intermediárias;
- Número de neurônios por camada: de acordo com as Tabelas 4.4 e 4.5;

*Tabela 4.4 – Quantidade de neurônios para RNAs com 1 camada neural intermediária.*

<b>Nº de neurônios</b>
5
10
15
20
25

No caso de topologias com 2 camadas neurais intermediárias, foi feita uma combinação entre o número de neurônios da primeira e da segunda camada. Esta combinação foi realizada de acordo com a Tabela 4.5.

*Tabela 4.5 – Quantidade de neurônios para RNAs com 2 camadas neurais intermediárias.*

<b>Número de neurônios da 1ª camada</b>	<b>Número de neurônios da 2ª camada</b>
5	10
	15
	20
	25
	30
10	10
	15
	20
	25
	30
15	10
	15
	20
	25
	30
20	10
	15
	20
	25
	30
25	10
	15
	20
	25
	30

Ressalta-se que as funções de ativação dos neurônios das camadas neurais intermediárias são do tipo tangente hiperbólica e as funções de ativação dos neurônios da camada de saída são do tipo linear.

Realizados todos os treinamentos propostos, foram obtidas 21 topologias de RNAs (com 1 e 2 camadas neurais intermediárias) para cada concessionária. Entretanto, as topologias foram treinadas considerando-se a entrada dos dados originais e a entrada dos dados após a seleção dos atributos, ou seja, ao final de todos os treinamentos obteve-se um total de 42 resultados de previsão para cada concessionária.



Por consequência, houve a necessidade de se estabelecer critérios que pudessem julgar o desempenho das RNAs. Assim, foram realizados cálculos de Erro Absoluto Médio (MAE - *Mean Absolute Error*), Erro Absoluto Percentual Médio (MAPE - *Mean Absolute Percentual Error*) e Erro Médio Quadrático (MSE - *Mean Square Error*) conforme respectivamente apresentados por meio das Equações 4.2 a 4.4:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - \hat{P}_i| \quad (4.2)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \frac{|P_i - \hat{P}_i|}{P_i} \quad (4.3)$$

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \hat{P}_i)^2} \quad (4.4)$$

onde,

$P_i$  representa o preço desejado;

$\hat{P}_i$  representa o preço estimado pela RNA;

$N$  é o número de amostras usadas para validação da RNA.

#### 4.1 RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA CPFL

As Tabelas 4.6 e 4.7 mostram os resultados obtidos para a validação das RNAs treinadas para prever ações da CPFL, onde os resultados da Tabela 4.6 foram obtidos utilizando a base de dados sem seleção de atributos e RNAs com apenas uma camada neural intermediária e os resultados da Tabela 4.7 foram obtidos considerando-se a seleção de atributos e mantendo-se a RNA com apenas uma camada neural intermediária.

*Tabela 4.6 – Resultados obtidos para previsão de ações da CPFL: dados sem seleção de atributos e RNA com 1 camada neural intermediária.*

CPFL - Sem Seleção de Atributos									
Neurônios	Valor Máximo			Valor Mínimo			Fechamento		
	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	0,000248	1,376626	0,000323	0,000259	1,242953	0,000341	0,000496	2,424169	0,000612
10	0,000670	3,699056	0,001066	0,000262	1,363022	0,000337	0,000473	2,269161	0,000651
15	0,001264	5,877369	0,002548	0,000372	1,947537	0,000481	0,000607	3,004742	0,000768
20	0,000555	2,893489	0,000728	0,000516	2,406370	0,000663	0,000529	2,697819	0,000698
25	0,003136	15,077875	0,004006	0,000781	3,597056	0,001069	0,000650	3,050974	0,000850

*Tabela 4.7 – Resultados obtidos para previsão de ações da CPFL: dados com seleção de atributos e RNA com 1 camada neural intermediária.*

CPFL - Com Seleção de Atributos									
Neurônios	Valor Máximo			Valor Mínimo			Fechamento		
	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	<b>0,000199</b>	<b>0,995795</b>	<b>0,000258</b>	<b>0,000247</b>	<b>1,178590</b>	<b>0,000315</b>	<b>0,000273</b>	<b>1,337097</b>	<b>0,000571</b>
10	0,000446	2,100075	0,000613	0,000514	2,154397	0,001552	0,000323	1,527891	0,000715
15	0,000465	2,443960	0,000652	0,000310	1,361211	0,000453	0,000279	1,363630	0,000369
20	0,000553	2,640722	0,000790	0,000363	1,745757	0,000475	0,000261	1,293149	0,000348
25	0,000427	1,784884	0,000757	0,000283	1,290124	0,000430	0,000295	1,419544	0,000402

Por meio das Tabelas 4.8 e 4.9 são mostrados os resultados que foram obtidos utilizando a base de dados sem seleção de atributos e RNAs com duas camadas neurais intermediárias (Tabela 4.8) e os resultados que foram obtidos considerando-se a seleção de atributos e mantendo-se as RNAs com duas camadas neurais intermediárias (Tabela 4.9).

Vale ressaltar que estes valores (MAE, MAPE e MSE) são calculados a fim de julgar o desempenho das RNAs e os resultados que obtiveram melhores performances foram colocados em negrito.

Tabela 4.84 – Resultados obtidos para previsão de ações da CPFL: dados sem seleção de atributos e RNA com 2 camadas neurais intermediárias.

CPFL - Sem Seleção de Atributos										
Neurônios camada 1	Neurônios camada 2	Valor Máximo			Valor Mínimo			Fechamento		
		MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	10	0,000892	4,213340	0,001050	0,000547	2,539470	0,000785	0,000689	3,241653	0,000897
	15	0,000462	2,335874	0,000632	0,000363	1,860592	0,000460	0,000738	3,177996	0,001188
	20	0,001037	5,123908	0,001349	0,000458	2,258735	0,000583	0,000615	3,044691	0,000748
	25	0,000654	2,868437	0,001012	0,000592	2,653331	0,000788	0,000639	3,554281	0,000868
	30	0,000916	4,226820	0,001205	0,000379	1,757001	0,000476	0,000605	2,847178	0,000789
10	10	0,001061	4,449598	0,001723	0,000530	2,424208	0,000676	0,000841	3,631399	0,001106
	15	0,001642	7,007260	0,002115	0,000401	2,072805	0,000529	0,000786	3,990592	0,001021
	20	0,000804	4,112829	0,001004	0,000327	1,633034	0,000427	0,000601	2,837138	0,000789
	25	<b>0,000401</b>	<b>2,166803</b>	<b>0,000515</b>	<b>0,001080</b>	<b>5,159111</b>	<b>0,001397</b>	<b>0,000692</b>	<b>3,362121</b>	<b>0,000864</b>
	30	0,000791	4,366207	0,001110	0,000439	2,380333	0,000528	0,000534	2,664470	0,000663
15	10	0,000695	3,193682	0,000892	0,000710	3,623714	0,000893	0,000870	3,886586	0,001191
	15	0,001473	7,723144	0,001986	0,000737	3,571984	0,001019	0,000563	2,804261	0,000716
	20	0,000755	4,117173	0,000993	0,000490	2,576021	0,000619	0,001331	5,769319	0,001907
	25	0,002110	8,792735	0,003138	0,000463	2,512858	0,000641	0,000550	2,664746	0,000718
	30	0,001178	6,666803	0,001806	0,000647	2,783147	0,000911	0,001019	5,277388	0,001240
20	10	0,000652	3,505424	0,000914	0,000810	3,453275	0,001057	0,000868	3,800080	0,001139
	15	0,002468	14,982413	0,003274	0,000444	2,064672	0,000614	0,000497	2,578764	0,000639
	20	0,002531	12,805066	0,003121	0,000707	3,234712	0,000959	0,001232	5,287719	0,001661
	25	0,000868	4,559431	0,001129	0,001306	6,036528	0,001873	0,000927	4,391355	0,001135
	30	0,000642	3,476252	0,000814	0,000407	2,305096	0,000589	0,000962	4,277023	0,001326
25	10	0,001856	9,564070	0,002285	0,000773	3,641744	0,001036	0,000557	3,114416	0,000715
	15	0,000622	3,869384	0,001055	0,000445	2,300187	0,000559	0,001025	4,821997	0,001366
	20	0,000499	2,439426	0,000652	0,000471	2,301524	0,000643	0,001280	6,177683	0,001638
	25	0,000953	4,802414	0,001232	0,001494	6,349504	0,002085	0,000433	2,217753	0,000544
	30	0,002788	12,973994	0,003764	0,001590	7,121558	0,002081	0,002527	9,933801	0,003585

Tabela 4.9 – Resultados obtidos para previsão de ações da CPFL: dados sem seleção de atributos e RNA com 2 camadas neurais intermediárias.

CPFL - Com Seleção de Atributos										
Neurônios camada 1	Neurônios camada 2	Valor Máximo			Valor Mínimo			Fechamento		
		MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	10	0,000482	2,192464	0,000651	0,000245	1,134423	0,000324	0,000287	1,380216	0,000589
	15	0,000221	1,052884	0,000298	0,000303	1,359015	0,000850	0,000280	1,360697	0,000372
	20	0,000228	1,103096	0,000297	0,000258	1,200493	0,000371	0,000257	1,278792	0,000343
	25	<b>0,000187</b>	<b>0,94995</b>	<b>0,000241</b>	<b>0,000241</b>	<b>1,177764</b>	<b>0,000319</b>	<b>0,000263</b>	<b>1,298553</b>	<b>0,000348</b>
	30	0,000512	2,606368	0,000612	0,000288	1,306316	0,000866	0,000291	1,400609	0,000386
10	10	0,000323	1,447057	0,000534	0,000264	1,190506	0,000353	0,000285	1,384216	0,000379
	15	0,000280	1,329192	0,000383	0,000490	2,022970	0,002139	0,000300	1,434900	0,000499
	20	0,000371	1,615492	0,000705	0,000259	1,187224	0,000379	0,000288	1,408751	0,000388
	25	0,000500	2,283596	0,000828	0,000270	1,250487	0,000371	0,000304	1,449550	0,000493
	30	0,000486	2,108544	0,001489	0,000330	1,462859	0,000483	0,000298	1,463744	0,000566
15	10	0,000434	1,980845	0,004421	0,000327	1,519601	0,000432	0,000305	1,457658	0,000409
	15	0,000805	3,209650	0,002553	0,000233	1,110715	0,000321	0,000316	1,537827	0,000564
	20	0,000241	1,128128	0,000347	0,000271	1,240591	0,000389	0,000341	1,602769	0,000485
	25	0,001025	4,576183	0,001597	0,000286	1,277647	0,000429	0,000355	1,643641	0,000573
	30	0,001143	5,005770	0,001813	0,000454	2,081597	0,000594	0,000268	1,319095	0,000355
20	10	0,000232	1,130962	0,000318	0,000296	1,335200	0,000428	0,000315	1,513303	0,000429
	15	0,000622	2,829918	0,001028	0,000272	1,253912	0,000371	0,000279	1,360994	0,000372
	20	0,000246	1,169233	0,000370	0,000297	1,415312	0,000456	0,000334	1,565662	0,000578
	25	0,000417	1,813968	0,001221	0,000281	1,348394	0,000410	0,000278	1,355905	0,000372
	30	0,000741	3,313465	0,002054	0,000336	1,507584	0,000499	0,000330	1,545548	0,000490
25	10	0,000384	1,688144	0,000579	0,000235	1,097710	0,000335	0,000300	1,450150	0,000393
	15	0,000480	2,109485	0,002356	0,000254	1,180285	0,000351	0,000296	1,433382	0,000395
	20	0,000752	3,067233	0,002524	0,000295	1,338792	0,000420	0,000332	1,554812	0,000540
	25	0,000940	4,110907	0,001508	0,000300	1,365442	0,000432	0,000338	1,603523	0,000453
	30	0,000250	1,227,080	0,000337	0,000355	1,559546	0,000537	0,000336	1,579818	0,000449

## 4.2 RESULTADOS OBTIDOS PARA PREVISÃO DE AÇÕES DA ELETROPAULO

Por meio das Tabelas de 4.10 a 4.14 é possível visualizar os resultados obtidos para a validação das RNAs que foram treinadas visando prever as ações da Eletropaulo. O resultado da previsão realizada sem a etapa de seleção de atributos encontra-se nas Tabelas 4.10 e na tabela 4.11 é possível visualizar o resultado da previsão com a etapa de seleção, frisando que ambas possuem 1 camada neural intermediária.

*Tabela 4.105 – Resultados obtidos para previsão de ações da ELETROPAULO: dados sem seleção de atributos e RNA com 1 camada neural intermediária.*

Eletropaulo - Sem Seleção de Atributos									
Neurônios	Valor Máximo			Valor Mínimo			Fechamento		
	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	0,000447	4,349320	0,000881	0,000489	2,749475	0,000703	0,000640	3,400051	0,000821
10	<b>0,000346</b>	<b>3,159537</b>	<b>0,000445</b>	<b>0,000473</b>	<b>2,554772</b>	<b>0,000626</b>	<b>0,000894</b>	<b>4,649180</b>	<b>0,001136</b>
15	0,025727	210,037779	0,035249	0,000799	4,363274	0,001127	0,000630	3,409597	0,000816
20	0,017293	140,357439	0,024032	0,000906	5,119615	0,001192	0,001179	7,165820	0,001607
25	0,001401	12,639042	0,001971	0,000914	5,111145	0,001266	0,000986	5,718066	0,001268

*Tabela 4.11 – Resultados obtidos para previsão de ações da ELETROPAULO: dados com seleção de atributos e RNA com 1 camada neural intermediária.*

Eletropaulo - Com Seleção de Atributos									
Neurônios	Valor Máximo			Valor Mínimo			Fechamento		
	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	0,000355	2,444140	0,003113	0,005309	23,811320	0,007116	0,005096	26,795891	0,006382
10	0,004411	36,430265	0,005367	0,004857	22,664739	0,006437	0,004984	23,528420	0,006396
15	0,006396	1,824178	0,000415	0,004651	22,077672	0,005946	0,004872	23,364322	0,006288
20	0,006288	3,143665	0,004742	0,004632	22,748435	0,006065	0,003951	21,917623	0,004961
25	<b>0,000246</b>	<b>1,972664</b>	<b>0,000412</b>	<b>0,004911</b>	<b>23,403270</b>	<b>0,006512</b>	<b>0,004259</b>	<b>21,145719</b>	<b>0,005382</b>

Já as Tabelas 4.12 e 4.13 apresentam 2 camadas neurais intermediárias, sendo a primeira tabela sem seleção de atributos e a segunda tabela com seleção de atributos respectivamente.

Ressalta-se que os resultados, novamente, foram obtidos para estimações de preços máximos, mínimos e de fechamento das ações.

Tabela 4.62 – Resultados obtidos para previsão de ações da ELETROPAULO: dados sem seleção de atributos e RNA com 2 camadas neurais intermediárias.

Eletropaulo - Sem Seleção de Atributos										
Neurônios camada 1	Neurônios camada 2	Valor Máximo			Valor Mínimo			Fechamento		
		MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	10	0,000892	4,213340	0,001050	0,000547	2,539470	0,000785	0,000689	3,241653	0,000897
	15	0,000462	2,335874	0,000632	0,000363	1,860592	0,000460	0,000738	3,177996	0,001188
	20	0,001037	5,123908	5,123,908	0,000458	2,258735	0,000583	0,000615	3,044691	0,000748
	25	0,000654	2,868437	0,001012	0,000592	2,653331	0,000788	0,000639	3,554281	0,000868
	30	0,000916	4,226820	0,001205	0,000379	1,757001	0,000476	0,000605	2,847178	0,000789
10	10	0,001061	4,449598	0,001723	0,000530	2,424208	0,000676	0,000841	3,631399	0,001106
	15	0,001642	7,007260	0,002115	0,000401	2,072805	0,000529	0,000786	3,990592	0,001021
	20	0,000804	4,112829	0,001004	0,000327	1,633034	0,000427	0,000601	2,837138	0,000789
	25	<b>0,000401</b>	<b>2,166803</b>	<b>0,000515</b>	<b>0,001080</b>	<b>5,159111</b>	<b>0,001397</b>	<b>0,000692</b>	<b>3,362121</b>	<b>0,000864</b>
	30	0,000791	4,366207	0,001110	0,000439	2,380333	0,000528	0,000534	2,664470	0,000663
15	10	0,000695	3,193682	0,000892	0,000710	3,623714	0,000893	0,000870	3,886586	0,001191
	15	0,001473	7,723144	0,001986	0,000737	3,571984	0,001019	0,000563	2,804261	0,000716
	20	0,000755	4,117173	0,000993	0,000490	2,576021	0,000619	0,001331	5,769319	0,001907
	25	0,002110	8,792735	0,003138	0,000463	2,512858	0,000641	0,000550	2,664746	0,000718
	30	0,001178	6,666803	0,001806	0,000647	2,783147	0,000911	0,001019	5,277388	0,001240
20	10	0,000652	3,505424	0,000914	0,000810	3,453275	0,001057	0,000868	3,800080	0,001139
	15	0,002468	14,982413	0,003274	0,000444	2,064672	0,000614	0,000497	2,578764	0,000639
	20	0,002531	12,805066	0,003121	0,000707	3,234712	0,000959	0,001232	5,287719	0,001661
	25	0,000868	4,559431	0,001129	0,001306	6,036528	0,001873	0,000927	4,391355	0,001135
	30	0,000642	3,476252	0,000814	0,000407	2,305096	0,000589	0,000962	4,277023	0,001326
25	10	0,001856	9,564070	0,002285	0,000773	3,641744	0,001036	0,000557	3,114416	0,000715
	15	0,000622	3,869384	0,001055	0,000445	2,300187	0,000559	0,001025	4,821997	0,001366
	20	0,000499	2,439426	0,000652	0,000471	2,301524	0,000643	0,001280	6,177683	0,001638
	25	0,000953	4,802414	0,001232	0,001494	6,349504	0,002085	0,000433	2,217753	0,000544
	30	0,002788	12,973994	0,003764	0,001590	7,121558	0,002081	0,002527	9,933801	0,003585

Novamente foi feito destacado em negrito os valores mais relevantes, que no caso da Tabela 4.10 foi obtido no treinamento com 10 neurônios, na Tabela 4.11 no treinamento com 25 neurônios, na Tabela 4.12 10 neurônios na camada 1 e 25 neurônios na camada 2 e na Tabela 4.13 houve 10 neurônios na camada 1 e 15 neurônios na camada 2.

Tabela 4.73 – Resultados obtidos para previsão de ações da ELETROPAULO: dados com seleção de atributos e RNA com 2 camadas neurais intermediárias.

Eletropaulo - Com Seleção de Atributos										
Neurônios camada 1	Neurônios camada 2	Valor Máximo			Valor Mínimo			Fechamento		
		MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE	MAE	MAPE (%)	MSE
5	10	0,001122	9,934910	0,001404	0,004138	20,187978	0,005848	0,005464	28,115408	0,007125
	15	0,000304	2,396331	0,001013	0,004767	22,673785	0,006483	0,005019	22,966840	0,006762
	20	0,000227	1,875231	0,000548	0,004713	21,363335	0,006649	0,004319	20,232617	0,005989
	25	0,000193	1,608987	0,000319	0,005283	28,232917	0,006793	0,004706	23,209835	0,006125
	30	0,000200	1,722612	0,000434	0,004511	21,240051	0,006368	0,004089	19,989041	0,005425
10	10	0,000235	1,837519	0,000514	0,004230	20,030602	0,005469	0,003779	18,438279	0,005162
	15	<b>0,000192</b>	<b>1,613873</b>	<b>0,000385</b>	<b>0,004098</b>	<b>20,801515</b>	<b>0,005429</b>	<b>0,004366</b>	<b>20,892255</b>	<b>0,005757</b>
	20	0,000240	1,916096	0,000625	0,004319	20,726558	0,005796	0,004749	22,746637	0,006217
	25	0,000470	3,101093	0,004312	0,004848	23,015547	0,006265	0,004965	24,969512	0,006376
	30	0,004256	27,823641	0,007301	0,004175	21,330939	0,005561	0,004895	23,882461	0,006205
15	10	0,001082	11,296826	0,001236	0,004650	21,341282	0,006293	0,004645	21,813361	0,006112
	15	0,011949	100,204485	0,016171	0,004336	24,117661	0,005594	0,004786	21,693240	0,006408
	20	0,000590	5,546675	0,001788	0,005099	25,124119	0,006705	0,004284	20,388061	0,005691
	25	0,000388	2,798894	0,000660	0,005047	23,673486	0,006675	0,004090	19,766413	0,005519
	30	0,000346	2,430939	0,002233	0,004775	22,833195	0,006266	0,004675	22,713909	0,006110
20	10	0,003903	26,822308	0,006569	0,004508	21,360963	0,005927	0,004377	23,228579	0,005629
	15	0,000287	2,117018	0,001193	0,004556	21,476369	0,006034	0,004039	19,882405	0,005364
	20	0,000736	4,353121	0,006848	0,004422	21,290178	0,005800	0,004799	23,051892	0,006165
	25	0,005586	35,462685	0,008532	0,004466	21,075563	0,006095	0,004409	23,076567	0,005658
	30	0,000203	1,699075	0,000431	0,004644	21,139868	0,006277	0,005366	26,409305	0,006753
25	10	0,004034	32,779591	0,006183	0,004678	21,329477	0,006200	0,004016	18,754293	0,005401
	15	0,000446	3,298396	0,000667	0,004264	20,181102	0,005745	0,004133	19,410347	0,005520
	20	0,000518	3,654908	0,004077	0,004538	21,460063	0,005983	0,004406	23,936445	0,005490
	25	0,000546	3,847252	0,002755	0,005288	27,841869	0,006536	0,004226	20,251739	0,005503
	30	0,000442	2,975324	0,000791	0,004761	23,457557	0,005999	0,004170	22,153503	0,005364





## Capítulo 5

# Conclusão

Um dos intuitos deste trabalho também foi o de explorar possíveis variáveis que poderiam influenciar na saída do treinamento da rede neural. Por este motivo foi analisado índices como temperatura e chuva, os quais foram bem relevantes para os atributos com seleção, em especial para a Eletropaulo. Ao mesmo tempo tem-se por conhecimento que fatores como valor do dólar e índice IEE fornecem um indicativo do comportamento geral tanto da bolsa de valores como do setor elétrico brasileiro.

Portanto, ao se trabalhar com séries temporais, é importante questionar quais são os tipos de variáveis que devem formar a base de dados, pois pode-se ter valores de alta correlação como a temperatura, que mostra-se importante na seleção porém não se sabe ao certo se influi ou não na previsão da série temporal. É importante também destacar que a base de dados deve ter uma organização cronológica pois como se trata do fator tempo, esta organização é crucial para a previsibilidade da série.

Cabe comentar ainda que o emprego da seleção de atributos possibilitou a redução de dimensionalidade dos dados de entrada que, por consequência, contribuiu para a minimização do esforço computacional. Além disso, a seleção de atributos faz-se importante devido à eliminação de variáveis que pudessem contribuir para o mal desempenho das Redes Neurais Artificiais durante o processo de previsão.

Por fim, a análise de diversas configurações das Redes Neurais (variação de camadas e número de neurônios) junto ao emprego do algoritmo de Levenberg-Marquardt possibilitou a obtenção de bons resultados para as concessionárias de energia elétrica selecionadas para este trabalho. Portanto, pode-se dizer que a Mineração de Dados baseada em Seleção de Atributos e Redes Neurais Artificiais é eficiente para a previsão das ações destas concessionárias.

# Referências Bibliográficas

REIS, E. J. R. *Benefícios do uso de mineração de dados para a inteligência competitiva das organizações*. Monografia de Final de Curso. Instituto de ciências exatas, Universidade Federal de Minas Gerais, 2011.

AMORIM, M. C. *Previsão de séries temporais usando séries exógenas e combinação de redes neurais aplicada ao mercado financeiro*. Dissertação de Mestrado. Centro de Informática, Universidade Federal de Pernambuco, 2008.

VALENTIM, Marta Lígia Pomim; et. al. *O Processo de Inteligência Competitiva em Organizações*. 2003.

LABOISSIERE, L. A. Estimação do preço de ações de concessionárias de energia elétrica brasileiras por meio de redes neurais artificiais. Escola de Engenharia de São Carlos, Universidade de São Paulo, 2013.

TATIBANA, C.Y. e KAETSU, D.Y. Uma introdução às Redes Neurais. <http://www.din.uem.br/ia/neurais/#neural>

ARAÚJO, K. A. M. e LOTUFO, A. D. P. Previsão de cargas elétricas através de uma rede neural feedforward com treinamento por Levenberg-Marquardt utilizando as toolboxes do MATLAB. *XXI Congresso de Iniciação Científica da UNESP*, 2009.

BANCO CENTRAL DO BRASIL. "Cotação de Moeda Estrangeira: Dólar Americano", Disponível em: <http://www4.bcb.gov.br/pec/taxas/port/ptaxnpesq.asp?id=txcotacao>, Acessado em 04 de Outubro de 2013.

BM&FBOVESPA. "Introdução ao Mercado de Capitais", Disponível em: <http://www.bmfbovespa.com.br/pt-br/a-bmfbovespa/download/merccap.pdf>, Acessado em 02 de Setembro de 2013a.

BM&FBOVESPA. "Histórico de Ações das Concessionárias de Energia Elétrica", Disponível em: <http://www.bmfbovespa.com.br>, Acessado em 04 de Outubro de 2013b.

BOX, G. E. P.; JENKINS, G. M. e REINSEL, G. C. *Time Series Analysis, Forecasting and Control*, Wiley Series in Probability and Statistics, 2008.

CERETTA, P. S.; BRUTTI, M. R. e SCHLENDER, S. G. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. *Informações Econômicas*, v.40, pp. 15-17, 2010.

COAKLEY, J. R. e BROWN, C. E. Artificial Neural Networks in Accountig and Finance: Modeling Issues. *International Journal of Intelligent Systems in Accounting, Finance & Management*, v. 9, pp. 119-144, 2000.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. e UTHURUSAMY, R. *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996.

FERNANDES, R. A. S. *Identificação de fontes de correntes harmônicas por redes neurais artificiais*. Dissertação de Mestrado. Escola de Engenharia de São Carlos, Universidade de São Paulo, 2009.

GATELY, E. *Neural Networks for Financial Forecasting*, John Wiley & Sons, 1995.

HALL, M. A. *Correlation-based Feature Selection for Machine Learning*, Ph.D. Thesis, The University of Waikato, 1999.

HALL, M. A. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. *Proc. of the 17<sup>th</sup> International Conference on Machine Learning (ICML)*, pp. 359-366, 2000.

HAN, J. e KAMBER, M. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2001.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999.

HU, J. W.; HU, Y. C. e LIN, R. R. Applying Neural Networks to Prices Prediction of Crude Oil Futures. *Mathematical Problems in Engineering*, v. 2012, 2012

KAASTRA, I. e BOYD M. Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, v.10, pp. 215-236, 1996.

KIMOTO, T.; ASAKAWA, K.; YODA, M. e TAKEOKA, M. Stock market prediction system with modular neural networks, *IJCNN International Joint Conference on Neural Networks*, pp.1-6, 1990.

LATORRE, M. R. D. O. e CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*, v. 4, n. 4, pp.145-152, 2001.

BERESTEIN, M. *Uso de Mineração de Dados na Bolsa de Valores – Trabalho de Conclusão de Curso pela Universidade do Vale do Itajaí, 2010*