

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Engenharia
Programa de Pós-Graduação em Engenharia Elétrica

Thiago Henrique Barbosa de Carvalho Tavares

**Modelos Fuzzy com Alterações no Universo de
Discurso para Previsão de Séries Não
Estacionárias com Aplicações em Finanças e
Geotecnia**

Belo Horizonte

2022

Thiago Henrique Barbosa de Carvalho Tavares

**Modelos Fuzzy com Alterações no Universo de Discurso para
Previsão de Séries Não Estacionárias com Aplicações em
Finanças e Geotecnia**

Texto de Tese de Doutorado apresentado ao programa de Pós-graduação em Engenharia Elétrica da UFMG como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica

Orientador: Eduardo Mazonni Andrade
Marçal Mendes

Coorientador: Bruno Pérez Ferreira

Belo Horizonte

2022

T231m

Tavares, Thiago Henrique Barbosa de Carvalho.

Modelos Fuzzy com alterações no universo de discurso para previsão de séries não estacionárias com aplicações em Finanças e Geotecnia [recurso eletrônico] / Thiago Henrique Barbosa de Carvalho Tavares. - 2022.

1 recurso online (114 f. : il., color.) : pdf.

Orientador: Eduardo Mazonni Andrade Marçal Mendes.

Coorientador: Bruno Pérez Ferreira.

Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 107-114.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Predição (Lógica) - Teses. 3. Mercado financeiro - Teses. 4. Geotecnia - Teses. 5. Probabilidades - Teses. I. Mendes, Eduardo Mazzonni Andrade Marçal. II. Ferreira, Bruno Pérez. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 621.3(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

FOLHA DE APROVAÇÃO

**"MODELOS FUZZY COM ALTERAÇÕES NO UNIVERSO DE DISCURSO PARA
PREVISÃO DE SÉRIES NÃO ESTACIONÁRIAS COM APLICAÇÕES EM FINANÇAS
E GEOTECNIA"**

THIAGO HENRIQUE BARBOSA DE CARVALHO TAVARES

Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Engenharia Elétrica. Aprovada em 29 de julho de 2022. Por:

Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes - Orientador
Prof. Dr. Bruno Perez Ferreira
Prof. Dr. Leandro dos Santos Maciel
Prof. Dr. Thiago Henrique Nogueira
Prof. Dr. Adriano Cesar Machado Pereira
Prof. Dr. Víctor Costa da Silva Campos



Documento assinado eletronicamente por **Eduardo Mazoni Andrade Marçal Mendes, Professor do Magistério Superior**, em 29/07/2022, às 16:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Victor Costa da Silva Campos, Professor do Magistério Superior**, em 29/07/2022, às 19:40, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leandro dos Santos Maciel, Usuário Externo**, em 30/07/2022, às 12:35, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Bruno Perez Ferreira, Professor do Magistério Superior**, em 01/08/2022, às 06:33, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Thiago Henrique Nogueira, Usuário Externo**, em 01/08/2022, às 13:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Adriano Cesar Machado Pereira, Coordenador(a) de curso**, em 01/08/2022, às 19:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1636239** e o código CRC **91EFDCE8**.

Esta trabalho é dedicado a todos os amantes de inteligência artificial, tecnologia e finanças e a todos aqueles que acreditam que podem mudar o mundo pois, geralmente, são estes os capazes de fazê-lo.

Agradecimentos

Agradeço aos meus filhos Eduardo, Miguel e Lucas por tornarem minha vida uma grande maravilha na qual todo esforço vale à pena. Agradeço à minha esposa Débora pelo suporte e palavras motivacionais. Agradeço aos meus pais Marinalva e Marcelo pelo apoio durante toda a vida que culminou na realização deste projeto. Agradeço aos meus sogros Fátima e Geraldo por terem me acolhido como um filho. Agradeço à minha irmã Paula pelos momentos felizes. Agradeço à Zelda por ter me feito companhia nas longas noites de trabalho. Agradeço à tia Ione por ser tão presente na minha vida e na vida de meus filhos. Por fim, minha gratidão aos meus orientadores Eduardo e Bruno por terem acreditado em meu potencial e à Capes e UFMG pelos apoios financeiros e estruturais durante o desenvolvimento deste trabalho. Obrigado a todos! A realização deste sonho não teria sido possível sem a ajuda de vocês!

*“O único lugar aonde o sucesso vem antes do trabalho
é no dicionário”
(Albert Einstein)*

Resumo

A modelagem de sistemas a partir de séries temporais visando predição ainda é um campo de pesquisa extremamente explorado nos dias atuais. Os estudos de predição nas áreas relativas ao mercado financeiro possuem relevância ainda mais acentuada, dada à cada vez mais acelerada evolução das tecnologias de transações financeiras com o surgimento de *fintechs* e *criptomoedas*, bem como do estreitamento das relações econômicas entre os países do mundo. Possuindo tal cenário como foco principal, este trabalho visa o desenvolvimento de modelos novos utilizando lógica fuzzy que possuam como principais pontos de inovação a segmentação do universo de discurso nebuloso em clusters utilizando aspectos estatísticos e estruturas computacionais recentes. Em seguida, os modelos são aplicados também a áreas não financeiras (como geotecnia) e em séries não temporais. Dentre os trabalhos gerados, frutos do desenvolvimento da tese, pode-se elencar:

Um primeiro modelo em que é estudada a relação entre a taxa Selic e alguns parâmetros bancários definidos pelos chamados "Acordos de Basileia". A correlação cruzada entre a taxa Selic e os parâmetros é utilizada para explicar como esses parâmetros afetam a taxa Selic e vice-versa, de modo a definir a previsibilidade da taxa Selic utilizando (alguns) desses parâmetros como entradas. Propõe-se então um modelo de previsão da taxa Selic com base em alguns parâmetros específicos usando ideias de lógica fuzzy, que tratam de um particionamento do universo de discurso usando clusters relacionados à distribuição estatística dos dados de saída. O modelo proposto é comparado a outros quatro modelos conhecidos na literatura e mostrou ter um desempenho médio melhor em relação aos mesmos.

Um segundo modelo fuzzy é apresentado onde o universo do discurso é baseado em uma estrutura de dados red-black tree (RBT) para aumentar as possibilidades de obter melhores previsões. O modelo foi aplicado para previsão do índice Ibovespa, no qual os valores de fechamento do índice foram agrupados em clusters após serem organizados utilizando a estrutura RBT. A estrutura de dados RBT é uma estrutura de dados de busca binária que promove um melhor equilíbrio, o que permite uma melhor precisão nos resultados da previsão. O modelo proposto é comparado a modelos fuzzy bastante conhecidos na literatura, apresentando melhores resultados de previsão.

Um terceiro modelo apresenta uma integração entre as escolas de análise técnica e fundamentalista no mercado financeiro. Embora as análises técnica e fundamentalista coexistam em finanças, há dificuldades em integrá-las para auxiliar nas decisões de investimento. Devido a esse fato, muitos pesquisadores as utilizam separadamente, uma seguida da outra. Este modelo pretende integrar ambas as análises em um modelo matemático usando um novo modelo fuzzy de séries temporais. As simulações do modelo pretendem prever os valores futuros do S&P 500 utilizando variáveis fundamentalistas e técnicas. Para verificar a acurácia do modelo, seu desempenho é medido usando a raiz do erro quadrático médio (RMSE) e comparado com outros modelos bem conhecidos na mesma área.

Um quarto modelo apresenta predição para determinação da capacidade de carga em ancoragens. Os muros ancorados no solo no Brasil são muitas vezes dimensionados por metodologias que não se adaptam satisfatoriamente à realidade do solo nacional. Nesse sentido, a pesquisa buscou propor uma metodologia computacional simplificada para previsão da capacidade de carga de âncoras em obras geotécnicas. O modelo proposto foi um modelo fuzzy utilizando características específicas selecionadas dos solos e geometria de sistemas de paredes de amarração, cujas medições são mais fáceis de serem feitas em comparação com a própria capacidade de carga.

Um quinto modelo aplicável a séries não temporais com aplicações na área de geotecnia, no qual é apresentada uma técnica de predição de capacidade de carga geotécnica de tirantes por meio de Lógica Fuzzy . Estes métodos possuem a vantagem de serem tolerantes ao conhecimento impreciso, incompleto e incerto, sendo, portanto, aplicável no encontro de soluções que são aproximações, viáveis e robustas ao mesmo tempo. Este modelo se propõe a ser objeto de consultas para projetistas e engenheiros no que tange ao projeto de tirantes. Para validação e aplicação do modelo, utilizou-se o resultado de 40 ensaios de recebimento (provas de carga) executados na Estação Vila Sônia da Linha 4 – Amarela, São Paulo-SP, dos quais, 32 foram utilizados para treinar o algoritmo e 8 foram utilizados para testar o modelo preditor. Os resultados obtidos foram próximos dos valores extrapolados dos ensaios de desempenho, mostrando-se uma alternativa viável para o dimensionamento de ancoragens.

Palavras-chaves: Predição. Mercado Financeiro. Sistemas Nebulosos. Geotecnia. Árvore rubro-negra.

Abstract

The modeling of systems from time series aiming at prediction is still an extremely explored field of research nowadays. Prediction studies in areas related to the financial market are even more relevant, given the increasingly accelerated evolution of financial transaction technologies with the emergence of *fintechs* and *cryptocurrencies*, as well as the strengthening of economic relationships among the countries of the world. Having this scenario as its main focus, this work aims to develop new models using fuzzy logic that have as main innovation the segmentation aspect of the fuzzy universe of discourse into clusters using statistical aspects and recent computational structures. Then the models are also applied to non-financial areas and in non-time series. Among the works generated, as a result of the development of the thesis, we can list:

A first model in which the relationship between the Selic rate and some banking parameters defined by the so-called Basel Accords is studied. The cross correlation between the Selic rate and the parameters is used to explain how these parameters affect the Selic rate and vice versa, in order to define the predictability of the Selic rate using (some) of these parameters as inputs. A prediction model of the Selic rate is then proposed based on some specific parameters using ideas of fuzzy logic, which deal with a partitioning of the universe of discourse using clusters related to the statistical distribution of the output data. The proposed model is compared to four other known models in the literature and has shown to have a better average performance in comparison to them.

A second fuzzy model is presented where the universe of discourse is based on a red-black tree (RBT) data structure to increase the possibilities of obtaining better predictions. The model was applied to forecast the Ibovespa index where the closing values of the index were grouped into clusters after being organized using the RBT structure. The RBT data structure is a binary search data structure that promotes better balance, which allows for better accuracy in prediction results. The proposed model is compared with well-known fuzzy models in the literature, presenting better prediction results.

A third model presents an integration between the schools of technical and fundamental analysis in the financial market. Although technical and fundamental analysis coexist in finance, there are difficulties in integrating them to assist in investment decisions. Due to this fact, many researchers use them separately, one after the other. This model intends to integrate both analysis into a mathematical model using a new fuzzy time series model. The model simulations aim to predict the future values of the S&P 500 using fundamental and technical variables. To verify the accuracy of the model, its performance is measured using the root mean square error (RMSE) and compared with other well-known models in the same area.

A fourth model presents a prediction for determining the load capacity in anchors. Walls anchored to the ground in Brazil are often dimensioned by methodologies that do not satisfactorily adapt to the reality of the national soil. In this sense, the research sought to propose a simplified computational methodology for predicting the carrying capacity of anchors in geotechnical works. The proposed model was a fuzzy model using selected specific characteristics of the soils and geometry of tie wall systems, whose measurements are easier to be made in comparison with the load capacity itself.

A fifth model applicable to non-time series with applications in the area of geotechnics where a technique for predicting the geotechnical load capacity of ties using Fuzzy

Logic is presented. These methods have the advantage of being tolerant to imprecise, incomplete and uncertain knowledge, being, therefore, applicable in finding solutions that are approximations, viable and robust at the same time. This model is intended to be the object of consultation for designers and engineers regarding the design of tie rods. For validation and application of the model, the results of 40 reception tests (load tests) performed at the Vila Sônia Station on Line 4 – Amarela, São Paulo-SP were used, of which 32 were used to train the algorithm and 8 were used to test the predictor model. The results obtained were close to the values extrapolated from the performance tests, proving to be a viable alternative for the design of anchors.

Key-words: Forecasting. Financial Market. Fuzzy Systems. Geotechnics. Red-black tree.

Lista de ilustrações

Figura 1 – Fluxograma com áreas envolvidas na tese	25
Figura 2 – Exemplo Gráfico de Barras	29
Figura 3 – Exemplo Gráfico de Candle	30
Figura 4 – Exemplo gráfico de candles do Ibovespa	31
Figura 5 – Exemplo de médias móveis	32
Figura 6 – Exemplo de função de pertinência para velocidade alta	37
Figura 7 – Exemplo de função de pertinência de entrada	37
Figura 8 – Exemplo de função de pertinência de saída	38
Figura 9 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais, considerando 4 modelos conhecidos da literatura e o modelo proposto.	54
Figura 10 – Curva de previsão sobre dados de treinamento com gráficos sobrepostos.	54
Figura 11 – Curva de previsão sobre os dados de teste em comparação com os dados experimentais originais, considerando 4 modelos conhecidos da literatura e o modelo proposto.	56
Figura 12 – Curva de previsão sobre dados de teste com gráficos sobrepostos.	56
Figura 13 – Árvore rubro negra após a inserção dos elementos 7, 5 e 10	60
Figura 14 – Árvore rubro negra após a inserção do elemento 2	60
Figura 15 – Árvore rubro negra após a inserção do elemento 6	61
Figura 16 – Árvore rubro negra após a inserção do elemento 1	61
Figura 17 – Árvore rubro negra após a inserção do elemento 3	62
Figura 18 – Árvore rubro negra após a inserção do elemento 4	62
Figura 19 – Estrutura do modelo fuzzy proposto com base em árvores rubro-negras	63
Figura 20 – Exemplo de árvore rubro negra	64
Figura 21 – Desempenho do modelo proposto comparado com o modelo Naive sobre os dados do índice de fechamento Ibovespa.	72
Figura 22 – Zoom sobre a figura do modelo proosto comparado ao modelo naive sobre os dados de fechamento do índice Ibovespa.	73
Figura 23 – Comparação de modelos de previsão	74
Figura 24 – Zoom sobre a Figura Comparação de modelos de previsão	75
Figura 25 – Comparação no desempenho de previsão dos modelos sobre a base de treinamento	80
Figura 26 – Teste de comparação de previsão dos modelos	82
Figura 27 – Comparação de previsão de teste de modelo apenas com Naive	83

Figura 28 – Alternativas de curvas de carga x deslocamento	89
Figura 29 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais.	92
Figura 30 – Curva de previsão sobre os dados de teste em comparação com os dados originais e outros modelos.	94
Figura 31 – Gráfico Q x Q comparando os modelos e os dados experimentais. O gráfico Q x Q é usado para comparar duas distribuições de probabilidade traçando seus quantis um contra o outro. Quanto mais próximos os gráficos estiverem, maior a probabilidade de as distribuições serem semelhantes, o que comprova a acurácia do modelo.	94
Figura 32 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais	99
Figura 33 – Curva de previsão sobre os dados de teste em comparação com os dados originais e outros modelos.	100
Figura 34 – Gráfico Q x Q comparando os modelos e os resultados experimentais.	100

Lista de tabelas

Tabela 1 – Parâmetros bancários e sua correspondente descrição	49
Tabela 2 – Valores p para o teste de hipótese de correlação entra a taxa Selic e certos parâmetros bancários. Um valor- p menor ou igual a 0.05 indica correlação significativa entre a Selic e um determinado parâmetro.	50
Tabela 3 – Descrição das variáveis que serão utilizadas pelo modelo de previsão.	51
Tabela 4 – Dados experimentais usados para treinamento e teste de modelo.	52
Tabela 5 – Previsão de resultados com base nos dados de treinamento.	53
Tabela 6 – Modelos de Comparação de desempenho de modelos sobre dados de treinamento utilizando 55 amostras de treinamento.	55
Tabela 7 – Resultados do teste comparando dados originais com 4 modelos conhecidos e modelo proposto com 55 amostras de treinamento.	55
Tabela 8 – Comparação de desempenho entre o modelo proposto e os modelos existentes no banco de dados de teste. RMSE médio usando amostras de treinamento variando de 48 a 55.	57
Tabela 9 – Resumo da influência do Selic sobre os parâmetros - parte 1	57
Tabela 10 – Resumo da influência do Selic sobre os parâmetros - parte 2	58
Tabela 11 – Exemplo de fuzzyficação. A série temporal fuzzy corresponde à segunda coluna da tabela e é referenciada como $A(t)$	66
Tabela 12 – Regras Fuzzy originadas da Tabela 11	67
Tabela 13 – Regras de Grupo Fuzzy.	67
Tabela 14 – Exemplo de pesos fuzzy.	68
Tabela 15 – Exemplo de Centros de Cluster.	69
Tabela 16 – Exemplo de Conjuntos de previsão, centros de cluster e pesos.	69
Tabela 17 – Dados do índice de fechamento diário Ibovespa usados pasra o teste de treinamento. Os vaores previstos para vários modelos e o modelo proposto são apresentados para comparação.	71
Tabela 18 – Comparação de desempenho entre o modelo proposto e modelos existentes sobre a base de treinamento Ibovespa.	71
Tabela 19 – Resultados de previsão sobre dados do mercado de ações.	76
Tabela 20 – Comparação de desempenho entre o modelo proposto e os modelos existentes na base de testes do Ibovespa.	76
Tabela 21 – Resultados de teste comparados a outros modelos existentes	81
Tabela 22 – Comparação de RMSE entre modelos	81
Tabela 23 – Dados experimentais para treinamento e teste de modelo	90
Tabela 24 – Vetor P para o primeiro treinamento	92

Tabela 25 – Resultados do teste em comparação com dados originais e outros autores. . .	93
Tabela 26 – Comparação RMSE entre o modelo proposto e os modelos existentes	95
Tabela 27 – Dados de entrada do modelo	98
Tabela 28 – Resultados do teste em comparação com dados originais e outros autores (kN).	99
Tabela 29 – Comparação RMSE entre o modelo proposto e os modelos existentes. . . .	101

Sumário

1	Introdução e Cenário atual	19
1.1	Introdução	19
1.2	Cenário atual - Uma perspectiva	23
1.3	Objetivos da tese	25
1.4	Estrutura da Tese	25
2	Materiais e Métodos	27
2.1	Análise Técnica	27
2.1.1	Conceito formal de Análise Técnica	27
2.1.2	A Teoria de Dow	28
2.1.2.1	Representação de papéis na bolsa	28
2.1.3	Métricas de Análise Técnica	29
2.1.3.1	Médias Móveis	30
2.2	Análise Fundamentalista	32
2.2.1	Indicadores Fundamentalistas	33
2.3	Modelagem de Séries Temporais	34
2.3.1	Séries temporais nebulosas (fuzzy)	35
2.3.2	Sistemas Nebulosos	36
3	Desenvolvimento de modelos e resultados - Séries Temporais	43
3.1	Modelo fuzzy baseado na distribuição estatística dos dados de saída	43
3.1.1	Formulação Matemática da lógica fuzzy	43
3.1.2	Modelagem Matemática Fuzzy	43
3.1.2.1	Etapa 1: Coleta de dados	44
3.1.2.2	Etapa 2: Preparação dos dados	44
3.1.2.3	Etapa 3: Fuzzyficação	45
3.1.2.4	Etapa 4: Defuzzyficação	46
3.1.3	Algoritmo do modelo baseado na distribuição estatística dos dados de saída	48
3.1.4	Aplicação para predição de variáveis macroeconômicas	49
3.1.4.1	Dados utilizados	49
3.1.4.2	Estudo de Correlação - Coeficientes de correlação para a hipótese de correlação com significância maior que 95 %	49
3.1.4.3	Simulação	50
3.1.4.4	Base de dados utilizada	51
3.1.4.5	Teste dentro da amostra de treinamento	51

3.1.4.6	Teste fora dos dados de treinamento	55
3.1.5	Discussão	57
3.1.6	Conclusões	58
3.2	Modelo Fuzzy utilizando a árvore rubro negra	59
3.2.1	Árvores rubro-negras (Red-black trees)	59
3.2.2	Construção da árvore rubro negra	60
3.2.3	Desenvolvimento do modelo	62
3.2.4	Algoritmo do modelo baseado na árvore rubro-negra	69
3.2.5	Exemplo de aplicação utilizando a árvore rubro-negra	70
3.2.5.1	Teste com dados de treinamento	70
3.2.5.2	Teste fora dos dados de treinamento	75
3.2.6	Discussão	76
3.3	Modelo Fuzzy integrador	76
3.3.1	Objetivos da simulação	77
3.3.2	Preparação de dados de entradas fundamentalistas	77
3.3.3	Preparação de dados de entradas técnicas	78
3.3.4	Preparação de dados para entrada do modelo fuzzy	78
3.3.5	Resultados Experimentais	79
3.3.5.1	Base de dados utilizada	79
3.3.5.2	Treinando o modelo	79
3.3.5.3	Testando o modelo	80
3.3.6	Discussão	83
4	Desenvolvimento de modelos e resultados - Séries Espaciais	84
4.1	Modelo fuzzy para predição de capacidade de carga em ancoragens	84
4.1.1	Geotecnia - Revisão bibliográfica	84
4.1.2	Conceitos de Geotecnia	87
4.1.3	Resultados experimentais	89
4.1.3.1	Base de dados utilizada	89
4.1.3.2	Treinando o modelo	91
4.1.3.3	Teste fora dos dados de treinamento	93
4.1.4	Discussão	95
4.1.5	Conclusões	95
4.2	Modelo fuzzy para predição de capacidade de carga geotécnica de tirantes	96
4.2.1	Estudo de Caso	97
4.2.2	Algoritmo de predição - Treinamento do modelo	97
4.2.3	Algoritmo de predição - Teste do modelo	97
4.2.4	Estudo da assertividade do modelo	99
4.2.5	Considerações finais	101

5 Conclusão	102
5.1 Conclusões da Tese	102
5.2 Produtos da tese	104
5.2.1 Artigo publicado em periódico	104
5.2.2 Artigo publicado em congresso	104
5.2.3 Artigo publicado em periódico	105
5.2.4 Software publicado	106
5.2.5 Artigo em fase final de revisão em periódico	106
Referências	107

1 Introdução e Cenário atual

Neste capítulo será apresentada a introdução da tese, mostrando o cenário atual envolvendo os conceitos e matérias envolvidas no desenvolvimento do trabalho. Além disso, o capítulo apresentará os principais objetivos da tese bem como a forma como a mesma foi estruturada em termos de capítulos.

1.1 Introdução

Em mercados denominados eficientes, (PINHEIRO, 2013) preconiza que os preços das ações são consequência de toda a informação disponível e que seus valores no futuro estão, de alguma maneira, correlacionados, em parte, com fatos ainda não disponíveis e portanto desconhecidos. Essa afirmação sugere que dados do mercado financeiro possuem uma parcela determinística e outra aleatória. Apesar de não existir modelagem ou predição perfeita em um mercado eficiente, pois esta depende de dados futuros, nota-se que ainda há grande interesse na literatura pela obtenção de melhores modelos que possam, além de permitir predição, acrescentar maior assertividade ao envolver relações entre diferentes mercados, de diferentes países, apontando, por exemplo, o chamado “efeito contágio” conforme salienta (VARTANIAN, 2012). O autor também aponta o crescimento nos estudos de modelos que possam mostrar essas interconexões entre mercados de diferentes segmentos e países.

Em (YING, 2012), é enfatizado que, para facilitar a tomada de decisões em mercados como o da Bolsa de Valores, estudos ainda são conduzidos para viabilizar a implementação de algoritmos computacionais nesses ambientes. Dentre os vários sistemas que buscam modelar o mercado de ações, o autor (IQUIAPAZA, 2010) afirma que “um modelo auto-regressivo (não linear) de transição suave possui capacidade preditiva com retornos estatisticamente positivos”. Esse modelo se baseia em informações de tempo passado dos próprios valores dos ativos financeiros (valores passados dos preços das ações). Conforme salientado pelo autor e por (SANTOS, 2005), a utilização de valores passados de ativos em modelagem não-linear são justificados pela correlação serial já comprovada em séries temporais desses ativos, conforme também afirma (TSAY, 1951). Esse campo do conhecimento que se utiliza basicamente de valores passados de ações para tirar alguma conclusão sobre valores futuros é chamado de Análise Técnica. Pela natureza do problema de predição, a área de pesquisa que trata do mercado de ações investe continuamente em estudos envolvendo probabilidade e estatística, uma vez que esses campos de conhecimento permitem a modelagem adequada da porção determinística dos dados do mercado de ações. Em outras palavras, com a probabilidade e a estatística, pode-se prever, com certo grau de certeza, valores futuros do mercado de ações.

Outro campo bastante estudado na modelagem financeira e predição de oscilação de

ativos envolve estudos psicológicos cujo objetivo é identificar como investidores reagem a fatores externos ao mercado como: notícias veiculadas pela mídia, declarações de líderes de governo, dentre outras. As ações e atitudes desses investidores a partir desses fatores são consideradas, pelos estudiosos do assunto, como as grandes responsáveis pela porção aleatória da flutuação de preços, (MARTINS; JÚNIOR, 2009). Outros estudos, como os apresentados em (CABRERA-PANIAGUA, 2015), tentam modelar as emoções presentes no ser humano, que muitas vezes podem atrapalhar numa correta análise ou decisão dependendo de notícias do mercado financeiro e das muitas formas como as mesmas podem ser apresentadas aos investidores. Nesse contexto, estudos probabilísticos constituem um dos campos de ferramentas mais úteis e promissoras em finanças, uma vez que permitem modelar a incerteza associada a variáveis de risco existentes no mercado financeiro, conforme salienta (BARBACHAN, 2003).

A análise técnica permite a utilização de indicadores visando a predição no mercado financeiro, o que levanta a importância de outra grande área: a de modelagem. Segundo (SANTOS, 2005), a modelagem em finanças e a área de estudo relacionada ao chamado movimento browniano já caminham juntos desde o século passado. Dessa forma, estudos quanto ao movimento browniano são frequentemente aplicados a dados de séries temporais da bolsa de valores. Outros modelos que trabalham com diferentes distribuições de probabilidade e coeficientes aleatórios contínuos ou discretos podem ser observados em (FÁVERO; D., 2010), (AMARAL; EPPRECHT, 2008). Modelos que envolvem distribuições de taxas de retorno se mostram úteis e apresentam bons resultados segundo (CASTRO; SILVEIRA, 2009). A modelagem também se faz presente no mercado de ações por meio, por exemplo, da utilização de funções de correlação e aproximação usando ferramentas conhecidas na literatura como o Filtro de Kalman, conforme bem aponta (QUEIROZ; MEDEIROS; NETO, 2011). Outros trabalhos como (VARTANIAN, 2012) mostram o efeito de correlação direta, que mensura o grau de associação e dependência entre duas variáveis, entre índices como o Dow Jones e Ibovespa, o que sugere uma modelagem mais acurada quando se utiliza um índice como entrada do modelo de previsão do outro índice.

Modelos de redes de nós como os apresentados em (OZSOYLEV, 2014) e (BAI, 2015) buscam relacionar diferentes setores de investimentos (alimentício, vestuário, óleo e gás, educacional, etc) mas deixam lacunas quanto às oportunidades de modelagem que possam interconectar os mesmos setores de forma quantitativa. Estudo similar que trabalha com criação de redes de grafos interconectando empresas e setores do mercado financeiro é apresentado em (CAETANO; YONEYAMA, 2015b). Entretanto nesse trabalho, o foco é o impacto da perturbação em um grafo onde ocorre a retirada de um nó único.

Ainda na questão da modelagem, métodos de regressão aparentemente não relacionados se mostraram bastante úteis no estudo da relação entre retorno de ações e volatilidade de lucro e volume em ativos financeiros, conforme informa (TABAK; GUERRA, 2007). São ferramentas que, conforme bem descreve (QUEIROZ; MEDEIROS; NETO, 2011), se mostraram efetivas em evidenciar a presença de bolhas especulativas no mercado de ações brasileiro. Outras

técnicas estudadas na modelagem de finanças em trabalhos recentes são, por exemplo, as redes neurais (CAETANO; YONEYAMA, 2015a) e (SHEN; CHAO; ZHAO, 2015), e a otimização multiobjetivo para estratégias de investimento (PINTO; NEVES; HORTA, 2015). Apesar das muitas técnicas disponíveis, o desafio de modelagem de séries temporais ainda é presente na área de finanças (CERETTA et al., 2011), principalmente pelo fato de as Séries de Finanças não possuírem um comportamento tão bem definido quanto de outros sistemas físicos, conforme afirma (SANTOS, 2005). Além disso, não existe, ainda, um consenso sobre o melhor modelo a ser aplicado no processo de estimação de volatilidade de preços, retornos e volume de ativos financeiros. Este se apresenta como outro campo que permanece pouco explorado e que pode ser um campo promissor de pesquisa.

Até aqui, foram mostrados que muitos são os estudos que foram conduzidos a fim de se modelar ou prever, de alguma maneira, os valores e flutuações de ativos do mercado financeiro. Entretanto, percebe-se, nesses campos, uma lacuna no sentido de integrar diferentes áreas do conhecimento como usar técnicas de modelagem não apenas utilizando dados numéricos puros, mas também procurando uma maneira de inserir informações a partir de análises históricas das empresas (análises que envolvem histórico da empresa além de dados de balanço patrimonial, lucro trimestral, prejuízo, etc), conforme levanta (XU et al., 2014) e (DIBEH; HARMANANI, 2012). Sendo assim, fica a pergunta: como criar um modelo que inclua, além de variáveis estatísticas relacionadas à série de preços, variáveis contábeis das empresas, diretamente associadas com o histórico das mesmas? Além disso, qual a melhor forma de agrupar tais variáveis de modo a melhorar o desempenho de predição? Para responder a essas questões, além da análise técnica discutida anteriormente, que trabalha com modelagem, também existe a análise fundamentalista, que trabalha com fatores que extrapolam os valores das ações e vão mais a fundo buscando informações sobre a saúde da empresa, bem como o seu histórico de lucros e prejuízos. Dessa maneira, como confirma (OLIVEIRA; FRASCAROLI, 2014), fatos como a relação real existente entre emissões de ofertas públicas de ações com características fundamentalistas, como níveis de produção industrial e retorno de ativos no mercado acionário brasileiro, sugerem que o ramo da análise fundamentalista se mostra promissor e ainda foco de pesquisas em âmbito mundial. Uma opção para tornar essa linha ainda mais efetiva é a agregação de dados fundamentalistas a modelos de análise técnica (conforme citado anteriormente). Alguns trabalhos que mencionam essa questão, embora não façam a integração das duas escolas de análise literalmente são (SANVICENTE; DELGADO, 2010), (PAVLIDIS et al., 2006) e (SHEN; CHAO; ZHAO, 2015). Esses trabalhos utilizam modelos com redes neurais, que também são modelos preditivos com variáveis associadas a análise técnica com rápida taxa de aprendizado, mas que possuem limitações quando procura-se predições de médio prazo. Trabalhos recentes como o de (CAETANO; YONEYAMA, 2012) mostram métodos eficientes para detecção de mudanças abruptas no Mercado Financeiro, também utilizando análise técnica para predição de volatilidade. Entende-se por mudanças abruptas, nesse contexto, alterações drásticas no mercado como quando uma tendência de valorização de um ativo se transforma, de

repente, em uma tendência de desvalorização. Entretanto, uma lacuna neste trabalho é que falta uma análise para mudanças mais suaves no mercado financeiro, que normalmente são focos de estudos em investimentos de médio a longo prazo. Em (XU et al., 2014), vislumbra-se maneiras potenciais de se trabalhar com análise técnica e fundamentalista simultaneamente. Entretanto, estas se limitam a comparar estudos técnicos e fundamentalistas, ao invés de tentar integrá-los. Ainda falta um modelo sistemático para integração entre as duas escolas de análise, seja através da criação de modelos de predição ou através da criação de um índice que auxilie em tomadas de decisão de compra e venda de ações.

Dessa maneira, dado todo o exposto acima, diversos modelos ainda são trabalhados e pesquisados com objetivo de predição e avaliação de risco no mercado de ações. Uma vez que a psicologia no mercado financeiro é responsável pelo fator aleatório na modelagem no mercado de finanças, entendemos que as escolas de análise técnicas e fundamentalista são responsáveis pela porção determinística, e, portanto, modelagem dessas séries temporais (como a de ações). Dessa forma, levanta-se, neste trabalho, a hipótese de que diferentes campos de análise como a análise técnica e fundamentalista possam ser fundidos em um modelo híbrido, aqui denominado técnico-fundamentalista, a fim de se obter melhores resultados no campo da predição de valores de ativos financeiros. Além disso, um modelo com essas características poderia permitir predições não somente de índices financeiros, mas também de sistemas físicos, uma vez que seu processo de predição estaria diretamente relacionado a propriedades estatísticas dos dados. Acredita-se na potencial melhoria de modelos a partir da inserção de dados fundamentalistas como informação a priori em modelos que utilizam análise técnica. Conforme afirma (XU et al., 2014), modelos heterogêneos abordando diferentes campos de análise possuem importante papel no mercado financeiro e vem sendo alvo de estudos cada vez mais frequentes. Na mesma linha de evolução de pesquisa, (DIBEH; HARMANANI, 2012) afirma a tendência de modelos que mesclam mais de um campo de análise, que partem de modelos determinísticos para criar modelos estocásticos que possam capturar flutuações de ruídos presentes em mercados financeiros na dinâmica do mercado de ações, o que corrobora a importância do tema no atual contexto acadêmico.

Além de todo o exposto acima, que basicamente envolve a integração entre as escolas de análise técnica e fundamentalista, pretende-se, nesse trabalho, abordar alguns pontos específicos que caracterizam o ineditismo da tese:

- Criação de um modelo nebuloso que proponha novas metodologias de partição do universo de discurso para as variáveis de entrada..
- Utilização de novas estruturas computacionais para o processo de agrupamento em modelos nebulosos.
- Mesclagem de variáveis técnicas e fundamentalistas em um mesmo modelo matemático para predição de índices no mercado de ações

Uma vez caracterizado o que vem a ser as escolas de análise técnica e fundamentalista, a importância da modelagem no mercado financeiro, a psicologia em finanças e os pontos inéditos pretendidos de estudos para esta tese, mostrar-se-á, na próxima seção, o cenário atual, no que diz respeito ao mercado financeiro, com foco no Brasil e o que os investidores de modo geral pensam a respeito da área de modelagem..

1.2 Cenário atual - Uma perspectiva

É de conhecimento comum que existe uma forte interseção entre as áreas de economia e engenharia e uma dessas áreas envolve a tentativa de criação de modelos, cujo principal objetivo é prever o comportamento do mercado financeiro.

É importante salientar que a previsão supracitada não possui 100% de acuracidade pelo simples fato de séries temporais do mercado financeiro (como séries de valores de ações de empresas, que serão extensamente citadas ao longo desse texto), possuírem um forte fator aleatório. Tal fator aleatório é atribuído ao comportamento do ser humano que, a princípio, é imprevisível e fortemente influenciado por fatores impossíveis de serem controlados como: política, notícias apresentadas pela mídia, recomendações de pessoas, etc.

As incertezas inerentes a quais decisões referentes a investimentos devem ser tomadas com o objetivo de aumento de patrimônio e maximização de lucro levam à procura por meios de minimizar a probabilidade de erros em investimentos de risco. Um erro em um investimento de risco é aqui considerado como um investimento que leva a algum prejuízo material. A sociedade brasileira, de maneira geral, evita investimentos de risco, buscando alternativas que possuam probabilidade mínima (ou zero) de gerar prejuízos, como investimentos em poupança, CDB e investimentos em tesouro direto nacional.

Os temas de análise fundamentalista e técnica, devido à sua importância, possuirão seções específicas nesta tese para serem tratados. Entretanto, por hora, podemos dizer que a análise fundamentalista trabalha com aspectos financeiros e fundamentos das empresas como lucro, prejuízo, dentre outros, chamados indicadores fundamentalistas. Já a análise técnica trabalha basicamente com análise estatística sobre valores passados de ações e parte do pressuposto de que todo o conhecimento sobre as ações e o mercado financeiro podem ser obtidos diretamente dos valores passados das ações.

Analisando o cenário nacional atual, segundo ([MARTINS; JÚNIOR, 2009](#)), um levantamento sobre a popularidade da poupança mostra que a mesma ainda é o tipo preferido de investimento dos brasileiros.

Conforme ([OZSOYLEV, 2014](#)) ressalta, quando trata de investimentos em bolsa de valores, grande parte da população ainda enxerga este tipo de investimento (e no mercado financeiro de forma geral) como investimentos de altíssimo risco, ou investimentos acessíveis

apenas a pessoas que possuem maior poder aquisitivo. Isso não representa a realidade. Diversos investimentos são acessíveis para pessoas das mais variadas faixas de poder aquisitivo. Todos esses fatos elencados sobre a atual situação e cultura de investimentos brasileiros já justificam o investimento em estudos e análises sobre o mercado financeiro.

Do ponto de vista acadêmico, a justificativa de investimento nesse ramo se dá pela vasta quantidade de trabalhos publicados na área de modelagem financeira, que tenta prever as oscilações no mercado financeiro. Daqui para frente os termos “mercado financeiro” ou “mercado de ações” serão usados de forma intercambiável.

No mercado financeiro, uma ação de uma empresa constitui uma pequena parte da empresa que é vendida a terceiros. As ações podem ser públicas, podendo ser vendidas a qualquer pessoa, ou privadas, caso em que apenas um grupo seleto de pessoas pode comprar ações daquela empresa.

Uma vez que as ações representam uma parte da empresa, possuir uma ação significa ser dono de uma pequena parte da empresa podendo, assim, desfrutar dos lucros da mesma, assim como, obviamente, dos prejuízos.

Os lucros proporcionados por ações podem vir de várias formas distintas. Uma delas, é a própria valorização da ação. Isso pode acontecer porque o mercado (considerado como pessoas de forma geral) considera a empresa um bom investimento e estaria disposto a pagar mais para comprar ações da mesma. Sendo assim, uma forma de obter lucros com ações pode ser comprando as ações quando as mesmas estão baratas e vendendo as mesmas quando estão caras. A grande questão é como saber quando as ações irão subir ou cair, o que é impossível saber com 100% de certeza. Entretanto, existem diversos modelos que tentam prever a tendência de oscilação das ações. Essa previsão possui um certo nível de acurácia/confiabilidade e os modelos estatísticos presentes na literatura tentam possuir o maior nível de confiabilidade possível, em que cada modelo é aplicável a um certo tipo de mercado e investimento financeiro.

Uma das variadas áreas de interseção entre Engenharia e Ciências Econômicas é a área em que procura-se modelar a evolução dos valores de ativos financeiros (como ações da Bolsa de Valores) na tentativa de prevê-los, fazendo trades (investimento envolvendo compra e venda de ações) perfeitos. Como a modelagem não é completamente determinística (não temos conhecimento de todos os fatores que influenciam a variação dos valores das ações), não é possível prever com exatidão os valores futuros das ações. O máximo que se pode fazer é estimar valores com certa margem de assertividade.

Diversas tentativas de modelar evolução temporal de valores de ativos financeiros existem atualmente e cada vez mais acredita-se ser possível estimar a porção determinística desses valores a fim de prevê-los. O mesmo acontece com o mercado de criptomoedas, como, por exemplo, o bitcoin. Conforme pode-se observar em (JANG; LEE, 2018), o treinamento de redes neurais bayesianas pode servir para estimar valores de evolução futura do bitcoin a partir de dados

provenientes da blockchain (espécie de livro contábil virtual que registra, entre outras coisas, transações utilizando moedas virtuais).

1.3 Objetivos da tese

Nesta tese, deseja-se criar um modelo utilizando lógica nebulosa que possua inovações principalmente no que diz respeito à partição do universo de discurso. Deseja-se que este modelo sirva para predição de índices financeiros e possua aplicações a sistemas físicos.

Deseja-se também, criar um modelo de predição híbrido, que use tanto elementos da escola fundamentalista quanto da escola técnica. Elementos inéditos a serem propostos nesta tese envolvem a inclusão de variáveis fundamentalistas específicas no modelo de predição: as métricas corporativas (que também possuem uma seção exclusiva para explicação de seu conceito).

A criação de modelos baseados em regras nebulosas (fuzzy), com regras técnicas e fundamentalistas foram inseridas no modelo através de funções de pertinência.

A Figura 1 representa um fluxograma indicativo que mostra as grandes áreas utilizadas na confecção dos modelos, objetivo desta tese.

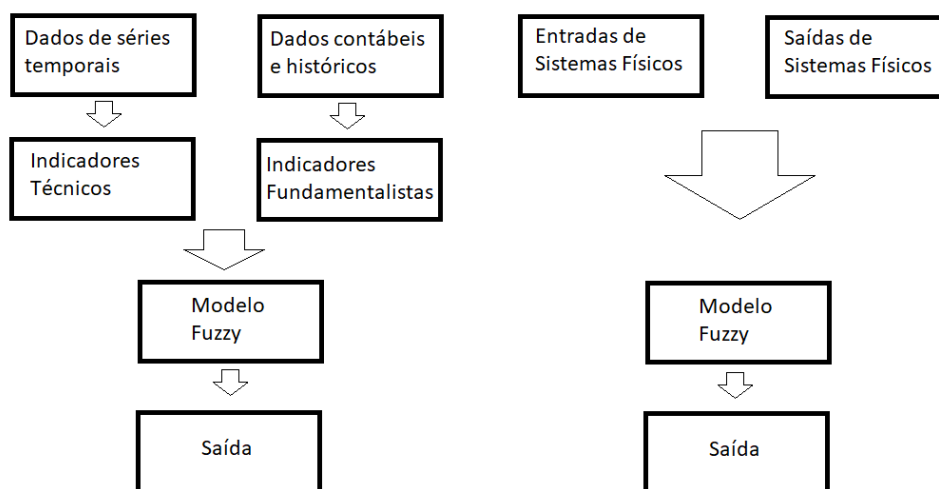


Figura 1 – Fluxograma com áreas envolvidas na tese. Aqui são apresentadas as principais áreas envolvidas do trabalho da tese.

1.4 Estrutura da Tese

Os próximos capítulos deste trabalho estão divididos da seguinte maneira:

- O Capítulo 2 trata do estudo bibliográfico de materiais e métodos. O capítulo, aborda as principais diferenças entre análise técnica e fundamentalista. Além disso, são apresentados alguns conceitos referentes a econometria, que pode ser considerada uma área

da modelagem aplicada a finanças. São apresentadas diversas técnicas de modelagem de séries temporais aplicadas à área de finanças. Aqui serão apresentados trabalhos recentes na área de modelagem que mostram a atual relevância acadêmica deste tema. Além disso, são levantadas algumas lacunas neste trabalho que justificam o desenvolvimento desta tese de doutorado. O capítulo aborda, ainda, conceitos referentes à escola de análise técnica apresentando os indicadores técnicos, por que eles influenciam na decisão de investidores e a forma como são construídos e utilizados em modelos de predição. O capítulo segue tratando conceitos referentes à escola de análise fundamentalista, conceituando-a e mostrando seus diferentes tipo de abordagem. É mencionado como são calculados os indicadores fundamentalistas e como os mesmos influenciam na determinação do valor contábil de uma empresa. São também apresentadas as métricas corporativas, que são variáveis de mensuração de risco normalmente utilizadas pelas empresas a fim de permitir tomadas de decisão com melhor embasamento. Por fim, são apresentados conceitos referentes a séries temporais nebulosas a partir do conceito de lógica nebulosa, também chamada de lógica fuzzy.

- O capítulo 3 apresenta o desenvolvimento dos modelos e os resultados obtidos na tese para séries temporais. O capítulo apresenta de forma detalhada o conceito envolvendo a árvore rubro-negra. A árvore rubro-negra é uma estrutura de dados computacional auto-balanceável que foi utilizada em um dos modelos para a partição do universo de discurso da série nebulosa. Por fim, além de mostrar como os modelos foram construídos, são apresentadas simulações com dados reais comprovando a eficácia dos modelos quando comparados a modelos existentes na literatura. O modelo apresenta discussões detalhadas sobre os resultados de modelagem além de conclusões a respeito de seu desempenho.
- O capítulo 4 apresenta o desenvolvimento dos modelos e os resultados obtidos na tese para séries espaciais. O capítulo trata de conceitos relacionados à área de geotecnia, apresentando conceitos básicos sobre uma área na qual os modelos propostos na tese também se mostraram eficientes. Pode-se dizer que geotecnia consiste na aplicação de métodos científicos e princípios de engenharia com o objetivo de interpretação de dados e utilização do conhecimento adquirido para a solução de problemas envolvendo a crosta terrestre e materiais terrestres. Por fim, são apresentadas construção de modelos, simulações e comparação de resultados de predição com outros modelos existentes em áreas correlatas na literatura.
- O capítulo 5 apresenta a conclusão da tese tecendo comentários sobre os resultados obtidos além de expectativas quanto à produção de trabalhos futuros de pesquisa. O capítulo também apresenta os principais produtos derivados da tese como artigos publicados em congressos e periódicos além de outros em fase de revisão por pares. O capítulo apresenta resumos dos trabalhos realizados.

2 Materiais e Métodos

Neste capítulo são descritos os materiais e métodos que poderão ser usados no desenvolvimento deste trabalho de doutorado.

2.1 Análise Técnica

Uma ação pode ser considerada como sendo uma pequena parte de uma empresa. Muitas empresas e pessoas físicas trabalham diariamente fazendo o que chamamos de "trading", que é basicamente comprar e vender ações visando lucro.

Existem várias estratégias de trading que basicamente norteiam o formato de como as pessoas ou empresas, que serão, a partir daqui, chamados de agentes do mercado financeiro, tomam suas decisões na hora de comprar ou vender ações.

Uma das vertentes de estudo que auxiliam os agentes a tomarem decisões é a chamada escola de Análise Técnica. Essa escola de análise basicamente utiliza fundamentos estatísticos visando identificar momentos ideais para comprar e vender ações de empresas. Alguns dos principais conceitos dessa escola de análise serão explicados a seguir.

Conforme descrito em (LIM, 2016), *a análise técnica é em essência a identificação e a predição do comportamento de mercado potencial baseado largamente nas ações e dinâmicas do mercado em si.*

Análise Técnica é uma escola de análise que trabalha com dados oriundos de papéis, também chamados de ações (ou outros ativos do mercado financeiro) e orienta investidores na tomada de decisões a partir de análises de cunho estatístico sobre esses papéis. A análise técnica parte do pressuposto de que toda a informação necessária para se realizar uma análise reside nos próprios valores dos ativos financeiros.

2.1.1 Conceito formal de Análise Técnica

Conforme bem salienta (LIM, 2016), os seres humanos nascem com certos instintos, moldados pela evolução através do tempo. Todo ser humano busca, de uma forma ou de outra atingir certos objetivos criados a partir desses instintos.

Um desses instintos está diretamente relacionado à melhoria de condições de vida, o que envolve tentativa de maximização de vantagens. No mercado financeiro, essas vantagens estão diretamente relacionadas à obtenção de lucros.

Visando a obtenção de lucros, a análise técnica serve principalmente para dois propósitos:

1. Identificação - Apresenta dados passados e presentes do mercado (como, por exemplo, valores de ativos financeiros). Fornece apenas uma representação descritiva das ações que ocorreram no mercado. Este propósito basicamente permite ao usuário analisar como foi o desempenho do mercado no passado. Este desempenho é normalmente descrito por volatilidade média durante determinado período, tendências de alta ou de baixa ocorridas no passado, volume negociado, impactos de anúncio de mídias e divulgação de decisões políticas.
2. Previsão - Quando uma determinada ação de mercado é identificada, as tendências existentes podem ser utilizadas para se realizar previsão sobre o mercado de ações. Obviamente, para considerar ser possível realizar a predição de valores no mercado de ações, parte-se do pressuposto de que o mercado seja cíclico em algum nível de análise e que, com valores passados, seja possível inferir valores futuros.

2.1.2 A Teoria de Dow

A Teoria de Dow fundou as bases para a análise técnica moderna.

Conforme cita (LIM, 2016), existem 6 princípios básicos da teoria de Dow.

1. As médias descontam tudo
2. O mercado possui 3 tendências específicas
3. Tendências primárias possuem 3 fases
4. Uma tendência persiste até que uma reversão é indicada
5. As médias devem se auto-confirmar
6. Volume deve confirmar a tendência

Além disso, apenas preços de fechamento são reconhecidos pela Teoria de Dow. Para explicar melhor a teoria de Dow, precisamos apresentar como a oscilação de uma ação é representada ao usuário.

2.1.2.1 Representação de papéis na bolsa

Conforme definido em (TAVARES, 2012), denomina-se pregão o processo pelo qual se vende ou compra ações na Bolsa. Durante o dia, os valores das ações passam por grande oscilações. Essas oscilações são medidas através de variações de certos atributos. São eles:

1. Abertura: É valor inicial do papel em um determinado dia.
2. Fechamento: É o valor no qual está o papel quando o dia termina.

3. Máxima: É o valor máximo que o papel atinge ao longo do dia.
4. Mínima: É o valor mínimo que o papel atinge ao longo do dia.

É importante salientar que tais atributos também podem ser apresentados em outra unidade temporal além de diário, podendo ser medidos por semana, mês, ano, etc. Esses atributos são apresentados de formas diferentes. Serão apresentados neste texto duas formas de representação distintas: gráfico de barras e gráfico de candle (ou vela).

A Figura 2 apresenta a estrutura de um gráfico de barras que pode representar a oscilação do papel em qualquer período pré-definido. A partir daqui, quando não for mencionado o período específico de oscilação, será considerado o período de um dia.

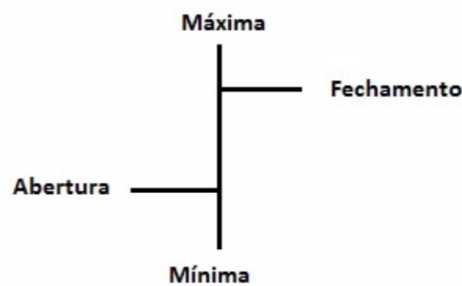


Figura 2 – Exemplo Gráfico de Barras. A linha vertical indica a oscilação diária do papel. O ponto no topo representa o valor máximo que o papel atingiu no dia. O ponto na parte inferior indica o valor mínimo que o papel atingiu. A barra horizontal esquerda indica o valor em que o papel abriu no pregão daquele dia. Já a barra horizontal direita indica o valor em que o papel fechou o pregão no dia. Consira-se um gráfico de alta quando o valor de abertura é menor que o de fechamento e um gráfico de baixa, caso contrário.

O nome candle é atribuído a essa representação pelo fato de que este padrão se assemelha a uma vela, com um corpo principal e um pavio. A estrutura de representação através de candle está apresentada na Figura 3.

2.1.3 Métricas de Análise Técnica

As métricas de análise técnica são normalmente indicadores técnicos utilizados com o objetivo de se analisar o mercado financeiro (ou mercado de ações). Serão apresentados alguns exemplos de métricas para se ter uma maior idéia das vantagens da utilização da análise técnica.

Serão apresentadas três métricas/ferramentas específicas conforme descrito abaixo.

- Médias móveis
- Ondas de Elliot
- Envelopes e Bandas de Bollinger



Figura 3 – Exemplo Gráfico de Candle. Máxima é representada pelo ponto mais alto do candle. Valor mínimo é representado pelo ponto mais baixo. O corpo central do candle possui forma de retângulo. Para candles de alta, como o candle branco da figura, onde a abertura é menor que o fechamento, a abertura é representada pela base do retângulo e o fechamento pelo topo. Para candles de baixa, como o candle preto da figura, onde o valor de fechamento é menor que o de abertura, o fechamento é representado pela base e a abertura, pelo topo do retângulo central. Cabe ressaltar que as cores dos candles são variáveis. Por exemplo existem gráficos onde candles de alta são verdes e candles de baixa, vermelhos.

2.1.3.1 Médias Móveis

A média simples de um conjunto de n dados pode ser dada pela equação 2.1.

$$m_s = \frac{1}{n} \cdot \sum_{k=1}^n x_k \quad (2.1)$$

A média ponderada de um conjunto de dados n ($x_1 \dots x_n$) com pesos ($p_1 \dots p_n$) é dada pela equação 2.2.

$$m_p = \frac{\sum_{k=1}^n x_k \cdot p_k}{\sum_{k=1}^n p_k} \quad (2.2)$$

A média móvel, que pode ser ponderada ou não, é uma função $f(t)$ onde o valor da função em qualquer tempo é a média dos últimos n valores de outra função $g(t)$. A média móvel simples é dada pela equação 2.3.

$$f(t) = \frac{1}{n} \cdot \sum_{k=1}^n g(t - k) \quad (2.3)$$

A média móvel ponderada, onde n amostras ($g(t - 1), g(t - 2), \dots, g(t - k)$) possuem pesos respectivamente iguais a (p_1, p_2, \dots, p_k) pode ser representada pela equação 2.4.

$$f(t) = \frac{\sum_{k=1}^n g(t - k) \cdot p_k}{\sum_{k=1}^n p_k} \quad (2.4)$$

Como exemplo ilustrativo, a Figura 4 mostra o gráfico de candles representando o índice Ibovespa no ano de 2019 de 01/05/2019 até 19/07/2019.

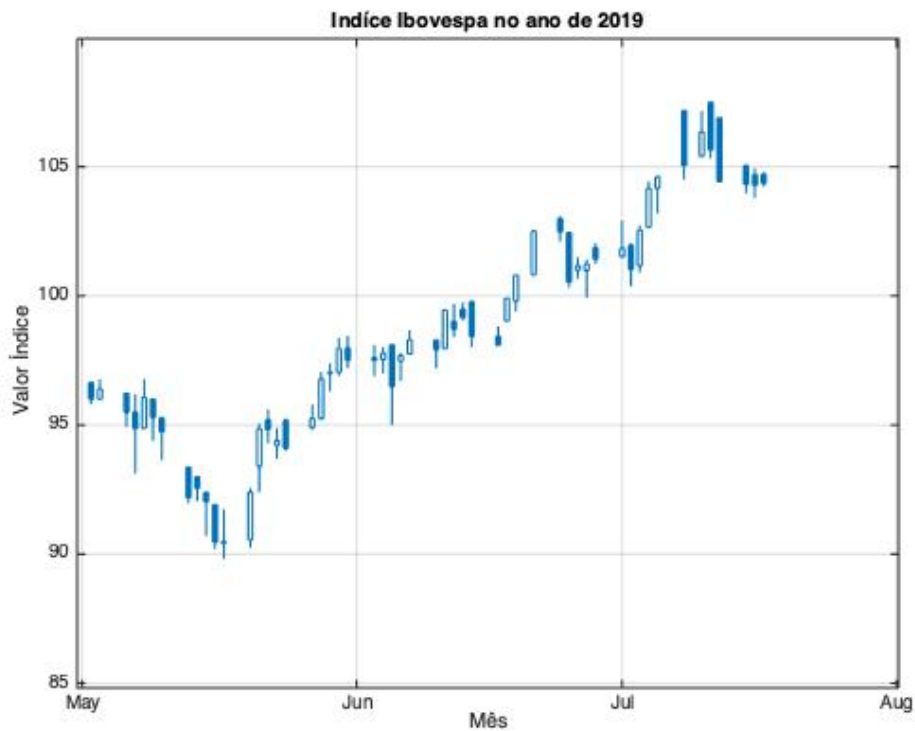


Figura 4 – Exemplo gráfico de candles do Ibovespa. Candles preenchidos são candles de baixa e candles não preenchidos são candles de alta.

Para se apresentar a média móvel, utiliza-se comumente o preço de fechamento no mercado de ações, mas nada impede que valores de abertura, máximo ou mínimo também sejam utilizados. Daqui pra frente, exceto quando mencionado, os valores a serem considerados nas estratégias de investimento citadas serão os valores de fechamento dos papéis.

A Figura 5 mostra o mesmo gráfico da Figura 4, porém com duas curvas de médias móveis. O período representado no gráfico foi reduzido afim de melhorar a visualização das curvas.

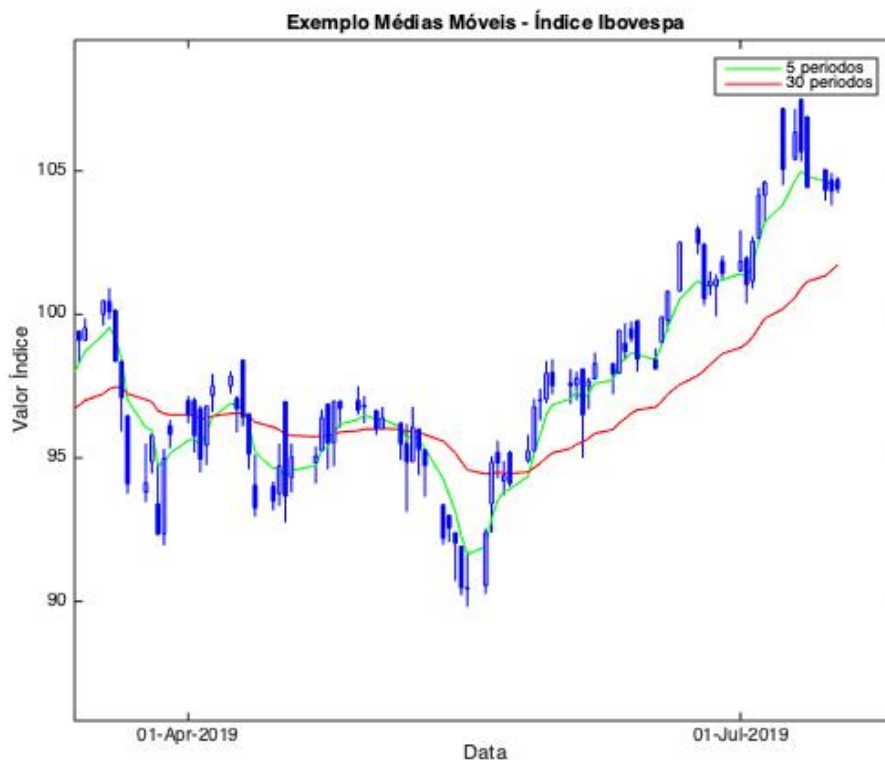


Figura 5 – Exemplo de médias móveis aplicadas sobre o índice Ibovespa entre abril e julho de 2019. Cada média móvel é composta de uma curva representando uma média móvel de n períodos anteriores. No caso da figura, temos duas médias móveis: uma com a média dos 30 períodos anteriores e outra representando uma média móvel de 5 períodos anteriores, conforme indicado na legenda.

Esse formato de representação é bastante usado, pois existem estratégias de investimento em que a compra ou a venda de ações é disparada pelo cruzamento entre duas médias de períodos diferentes. A lógica básica é que quando uma média menor cruza "para cima" outra média maior então pode-se afirmar que existe uma tendência de valorização do papel pois valores recentes estão se tornando maiores que valores antigos, comparando-se um determinado tempo específico. Nesse caso, seria um momento ideal de compra. Já o caso oposto, em que a média menor cruza "para baixo" uma média maior, utilizando a mesma analogia, seria um momento ideal para venda.

2.2 Análise Fundamentalista

A escola de análise fundamentalista é uma escola de análise de valores do mercado mobiliário (assim como a escola de análise técnica) em que são estudados os fundamentos das empresas, onde parâmetros contábeis de balanços são utilizados na tentativa de se identificar o valor real da empresa.

Ainda na área de análise fundamentalista, conforme descrito em (LEE, 1999), métricas corporativas constituem um conjunto de conceitos envolvendo a mensuração do risco de mercado em um ambiente corporativo ou empresarial.

2.2.1 Indicadores Fundamentalistas

No trabalho proposto nesta tese de doutorado existe uma série das chamadas variáveis fundamentalistas que servirão de entrada em um modelo preditivo.

Algumas dessas variáveis serão apresentadas a seguir. Embora nem todas sejam usadas no modelo proposto, entendemos que é importante apresentá-las a fim de se aprofundar na contextualização da análise fundamentalista.

- P/L: Preço por lucro. Indicador que mede a relação entre o preço atual das ações e o lucro por ação acumulado nos últimos 4 trimestres (ou 1 ano).
- P/VP: Preço por valor patrimonial - Preço da ação/Valor Patrimonial.
- P/EBIT: Preço por lucro antes de impostos e despesas. Procura eliminar influência de fatores que não são realmente da empresa.
- PSR: Price Sales Ratio. Valor de mercado da empresa dividido pela receita operacional líquida, ou preço da ação dividido pela receita líquida por ação. Qualquer valor abaixo de um é considerado interessante.
- P/Ativos: Preço da ação dividido pelos ativos totais da empresa.
- P/Capital de Giro: Preço da ação dividido pelo capital de giro por ação. Capital de giro é o ativo circulante menos passivo circulante.
- P/ Ativ Circ Liq: Preço da ação dividido pelo ativo circulante líquido por ação. Ativo circulante líquido é obtido subtraindo o ativo circulante pelo passivo, ou seja, após o pagamento de todas as dívidas, quanto sobraria dos Ativos mais Líquidos da empresa (caixa, créditos, estoques, etc.).
- Div. Yield: É o dividendo pago por ação dividido pelo preço da ação. É o rendimento gerado para o dono da ação pelo pagamento de dividendos.
- EV/EBIT: Valor da empresa dividido pelo EBITDA (quantidade de recursos que a empresa gera apenas em suas atividades principais, sem contar a rentabilidade de investimentos ou descontos de impostos).
- Giro Ativos: Receita líquida dividida pelos ativos totais. Indica a eficiência com a qual a empresa usa seus ativos para gerar vendas.
- Cres. Rec (5a): Crescimento da Receita Líquida nos últimos 5 anos.
- LPA: Lucro por ação. Relação entre o lucro líquido da empresa em um determinado período e o seu número de ações.

- VPA: Valor Patrimonial da ação. $\text{Patrimônio Líquido} / \text{Quantidade de Ações}$
- Margem Bruta: É o Lucro Bruto dividido pela Receita Líquida. Indica a porcentagem de cada R\$1,00 de venda que sobrou após o custo dos produtos/serviços vendidos.
- Margem EBIT: É o EBITDA dividido pela Receita Líquida. Indica a porcentagem de cada R\$1,00 de venda que sobrou após o pagamento dos custos dos produtos/serviços vendidos, das despesas com vendas, despesas gerais e administrativas.
- Margem Líquida: Lucro líquido dividido pela receita líquida.
- EBIT/Ativo: É a relação do EBITDA dividido pelos ativos totais da empresa.
- ROIC: Significa o retorno sobre o capital investido da empresa. Ele mede a capacidade da empresa de obter lucro a partir de seus investimentos. Quanto maior, melhor.
- ROE: Indica o retorno sobre o patrimônio líquido (Equity) da empresa, ou seja, a eficiência da empresa de produzir lucro a partir de seu patrimônio. Quanto maior, melhor. O ideal, também, é comparar o ROIC e o ROE com anos anteriores, a fim de se verificar se o resultado é constante.
- Liquidez Corr: Ativo circulante dividido pelo passivo circulante. Reflete a capacidade de pagamento de dívidas da empresa, no curto prazo.
- Div Br/Patrim: Dívida bruta total (dívida + debêntures) dividido pelo patrimônio líquido.

2.3 Modelagem de Séries Temporais

Nesta seção, os conceitos teóricos necessários para a modelagem de séries temporais são brevemente listados. A estrutura de modelagem será construída através de lógica nebulosa. Serão apresentados trabalhos recentes publicados em áreas cujo conhecimento básico estatístico pode ser encontrado em (MORETTIN; BUSSAB, 2013). Técnicas de processamento digital de sinais bem como conceitos básicos de sinais e sistemas podem ser encontrados em (DINIZ; SILVA; NETTO, 2014) e em (LATHI, 2007). Conceitos de redes neurais podem ser encontrados em (BRAGA; CARVALHO; LUDERMIR, 2007). O processo de modelagem e identificação de sistemas bem como de inferência e simulação de modelos possui conceitos básicos apresentados em (AGUIRRE, 2007) e (LJUNG, 1999).

Na área de modelagem de sistemas financeiros, é de extrema importância o estudo dos perfis dos investidores pois isso se transforma em uma informação a priori a ser incluída em modelos. Um exemplo de trabalho não científico, porém de utilidade clara, é o apresentado em (SEABRA, 2018), em que a autora realiza uma série de entrevistas com grandes nomes de investidores e obtém algumas idéias sobre suas metodologias de tomadas de decisão. A relevância desse trabalho se dá porque grandes investidores com grandes capitais são aqueles

que movimentam a maior parte do mercado em termos de valor. Saber a forma como os mesmos pensam ou mesmo tomam decisões permite a criação de modelos com informações a priori a respeito do formato de tomada de decisões humano, algo que a princípio não seria possível de se modelar.

Ainda na área de modelagem, a identificação de sistemas (AGUIRRE, 2007) e (LJUNG, 1999) é um procedimento no qual dado um conjunto de dados de entrada e saída de um sistema qualquer, é possível a obtenção de um modelo matemático que explique, ainda que de forma aproximada, a relação de causa e efeito presente entre esses dados de entrada e saída. Por exemplo, em massas de dados onde existe ruído presente (que é a grande maioria dos sinais existentes), uma das maneiras de se tirar o ruído puramente aleatório de um sinal é através da realização de vários ensaios e tirando-se a média desses ensaios.

Na área de análise de correlação, o autor (AGUIRRE, 2007) define o que é função de correlação cruzada (FCC) e função de autocorrelação (FAC) bem como função de covariância cruzada e autocovariância. Isso permite verificar a real interdependência entre duas ou mais variáveis antes de se tentar encontrar uma relação de causalidade entre as mesmas.

Um dos ramos de verificação e modelagem de sistemas importante que será agora considerado é a verificação de sistemas nebulosos.

Primeiramente será dada uma pequena introdução de lógica nebulosa e em seguida séries temporais nebulosas serão explicadas. Posteriormente será apresentada uma metodologia de identificação utilizando séries nebulosas.

2.3.1 Séries temporais nebulosas (fuzzy)

O primeiro passo para entender o que é uma série temporal fuzzy é destacando suas diferenças em relação a uma série temporal convencional. Uma série temporal convencional é composta por números reais, enquanto séries temporais fuzzy são compostas por conjuntos fuzzy (SONG; CHISSOM, 1993a). Um conjunto fuzzy é uma classe com limites indefinidos (limites fuzzy).

Vamos definir U o universo de discurso com a seguinte notação apresentada na Equação 2.5.

$$U = \{u_1, u_2, \dots, u_n\} \quad (2.5)$$

Um conjunto fuzzy A de U é definido por:

$$A = \{f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n\} \quad (2.6)$$

onde (CHEN, 1996) f_A é a função de pertinência de A $f_A : U \rightarrow [0, 1]$. $f_A(u_i)$ indica o grau de pertinência de u_i em A, onde $f_A(u_i) \in [0, 1]$ e $1 \leq i \leq n$.

As definições de séries temporais fuzzy são revisadas a seguir.

Definição: Seja $Y(t)(\dots, 0, 1, 2, \dots)$, um subconjunto de \mathbb{R} , o universo de discurso em que conjuntos fuzzy $f_i(t)(i = 1, 2, \dots)$ são definidos e seja $F(t)$ uma coleção de $f_i(t)(i = 1, 2, \dots)$. Então $F(t)$ é chamada de série temporal fuzzy em $Y(t)(\dots, 0, 1, 2, \dots)$.

$F(t)$ pode ser considerada uma variável linguística (ZADEH, 1975) e $f_i(t)(i = 1, 2, \dots)$ pode ser vista como possíveis valores linguísticos de $F(t)$ onde $f_i(t)(i = 1, 2, \dots)$ são representados por conjuntos fuzzy. Note que $F(t)$ é uma função do tempo, uma vez que seus valores podem ser diferentes em momentos diferentes devido ao fato de que as peças do universo de discurso podem ser diferentes em momentos diferentes. De acordo com (SONG; CHISSOM, 1993a), se $F(t)$ é causada por $F(t - 1)$ apenas, então essa relação é representada por $F(t - 1) \rightarrow F(t)$.

Definição: Seja $F(t)$ uma série temporal fuzzy. Se para qualquer tempo t , um deslocamento em t provoca um deslocamento equivalente em $F(t)$, e $F(t)$ somente possui elementos finitos, então $F(t)$ é chamada uma série temporal fuzzy invariante no tempo. Do contrário, a mesma é chamada uma série temporal fuzzy variante no tempo.

A fim de se exemplificar os conceitos de uma série temporal fuzzy suponha que alguém decida classificar a temperatura de um lugar durante o dia entre "muito frio", "frio", "normal", "quente", e "muito quente". Essas classificações são consideradas conjuntos fuzzy e são usadas para representar sentimentos de temperatura.

2.3.2 Sistemas Nebulosos

Conforme apresentado em (WANG, 1997) um sistema nebuloso é um sistema baseado em conhecimentos prévios ou, em outras palavras, baseado em regras. A principal parte de um sistema nebuloso, que pode ser considerada o seu coração, é uma base de conhecimento consistindo de regras do tipo *IF-THEN*.

A Figura 6 apresenta um exemplo de função de pertinência, que também é chamada de função membro. Uma função de pertinência nebulosa pode definir tanto variáveis de entrada de um modelo como variáveis de saída de um modelo. Define-se variáveis linguísticas, aquelas variáveis que possuem algum tipo de significado a princípio não quantificado de maneira exata como: quente, frio, alto, baixo.

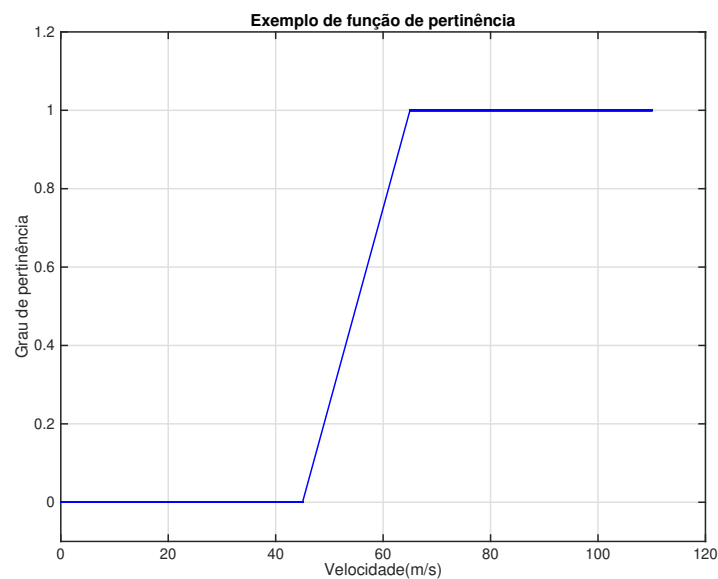


Figura 6 – Exemplo de função de pertinência para velocidade alta. Velocidade iguais ou inferiores a 45 m/s possuem grau de pertinência 0 enquanto velocidades com valores superiores a 65 m/s possuem grau de pertinência 1. Velocidades com valores intermediários a esses limites possuem pertinência entre 0 e 1, que variam linearmente de acordo com a velocidade.

A saída de um modelo de inferência nebuloso (que é um modelo preditivo) é determinada a partir das funções de pertinência de entrada e saída, dos valores de entrada e das regras nebulosas que definem o modelo.

Um exemplo prático de regra de cálculo da saída de modelo (também chamado de defuzzyficação) pode ser encontrado em (CARNEIRO et al., 2013), conforme Figuras 7 e 8.

A Figura 7 apresenta um exemplo de função de pertinência nebulosa de entrada.

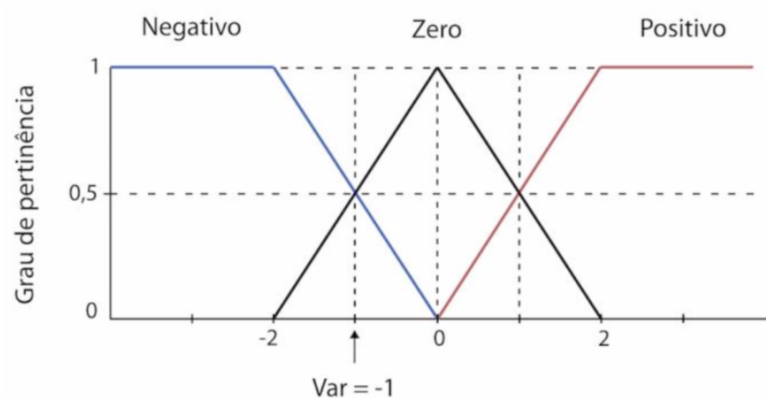


Figura 7 – Exemplo de função de pertinência de entrada. Temos aqui 3 funções nebulosas com nomes: "Negativo", "Zero" e "Positivo". Neste exemplo, nossa variável de entrada chamada de "var" possui, conforme indicado, valor igual a -1". Esse valor de variável faz com que a mesma possua pertinência igual a 0,5 para o conjunto "Negativo" e 0,5 para o conjunto "Zero".

Para maiores detalhes sobre processos de inferência, fuzzyficação e defuzzyficação pode-se consultar livros específicos que tratam o assunto como (WANG, 1997).

Considerando a Figura 7, suponha a seguinte regra nebulosa:

- Se valor da variável de entrada é negativo ou zero, então variável de saída é negativo.

No caso de uma regra do tipo "ou", uma das maneiras de se calcular o valor da pertinência é utilizando o máximo dos valores das pertinências de entrada. Sendo assim, como nossa variável possui valores de pertinência 0,5 para ambos os conjuntos "Negativo" e "Zero", o valor final será o máximo entre 0,5 e 0,5, que, obviamente, é 0,5.

Observando então as curvas de pertinência da saída, observamos a curva "Negativo" e marcamos o valor 0,5. Isso define, conforme mostrado na Figura 8 uma região gráfica abaixo do valor 0,5 na respectiva curva.

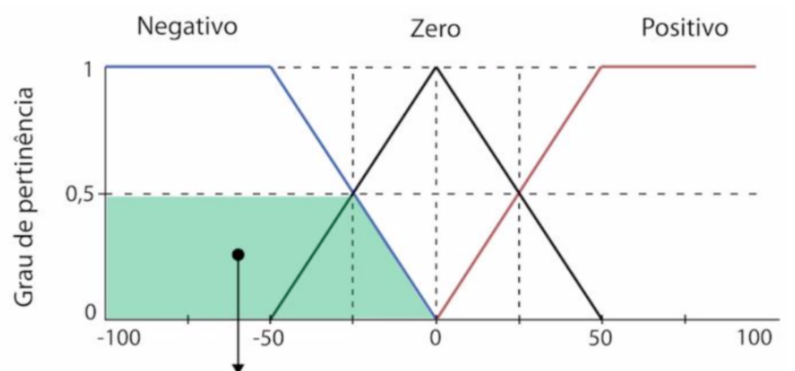


Figura 8 – Exemplo de função de pertinência de saída. O valor 0,5 é obtido a partir da regra nebulosa, que conecta entrada e saída do sistema nebuloso. Esse valor é marcado na função "Negativo" e gera uma figura cuja área está demarcada. O valor de saída representado pela seta é determinado pelo centróide da figura demarcada.

Uma das formas como a saída pode ser calculada é através do centróide da figura marcada, que, no caso da Figura 8 se trata de um trapézio. Para o cálculo do valor final da saída nebulosa, ou processo de defuzzyficação, calculamos o centróide desse trapézio. O valor calculado é aproximadamente -55,95, que seria o valor da saída após o processo de defuzzyficação.

Nas áreas referentes a Sistemas Nebulosos temos diversos trabalhos recentes que mostram o quanto o tema de modelagem financeira ainda é explorado em publicações científicas.

Muitos são os artigos que envolvem a tentativa de utilização de modelos nebulosos para predição no mercado financeiro. Como exemplo, (CHENA; CHENGA; TEOH, 2007) propõe um novo modelo nebuloso que incorpora o conceito da sequência de Fibonacci de forma indireta. Neste trabalho, os autores realizam simulações com dados como séries temporais de ações do TSMC (Taiwan Semiconductor Manufacturing Company) com período de 5 anos e ações TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) com período de 13 anos. No artigo, os autores fazem uma comparação de seu modelo com uma série de outros modelos e concluem que o mesmo possui melhor desempenho que o obtido por modelos de séries temporais nebulosas convencionais.

Ainda no ambiente de modelagem nebulosa, os autores do artigo (BOSE; MALI, 2019) afirmam que a predição de séries temporais por meio de lógica nebulosa se constitui um campo de pesquisa emergente. Vários problemas podem ser resolvidos e modelados através de lógica nebulosa, em especial problemas que tratam de incertezas e imprecisão.

Publicações pioneiras nessa área de modelagem nebulosa podem ser observadas em (SONG; S.CHISSOM, 1993), onde modelos nebulosos de séries temporais são apresentados. O artigo aborda, de maneira introdutória, o processo de modelagem nebulosa, que constitui de três etapas:

- fuzzyficação
- determinação da relação lógica nebulosa
- defuzzyficação

Os autores do artigo também chamam a atenção para possíveis problemas que ocorrem quando se deseja realizar modelagem de séries temporais por meio da utilização da lógica nebulosa como:

- ordem do modelo (valores passados)
- número de intervalos a serem criados
- manuseio de dados online
- manuseio de valores faltantes

Um ponto especial de atenção é a característica sazonal dos dados financeiros. Em outras palavras, é importante levar em conta que dados financeiros não estão disponíveis todos os dias da semana mas apenas em dias úteis.

Um trabalho que atua com modelagem nebulosa porém com modelos com mais de uma variável pode ser visto em (NAN et al., 2012). Os autores alegam existir uma gama grande de trabalhos envolvendo lógica nebulosa para predição no mercado financeiro porém ressaltam uma lacuna em que a maioria dos modelos não considera o fator tendência para realizar a modelagem. O artigo aplica o algoritmo de clusterização (ou agrupamento) c-médias nebuloso para processar séries temporais nebulosas com atributos multi-variados apresentando resultados satisfatórios quando comparados com outros trabalhos existentes na literatura.

Mais um trabalho que apresenta o poder da lógica nebulosa na tarefa de predição de séries temporais financeiras é o apresentado em (SUN et al., 2015). O artigo trata da predição de um índice específico do mercado de ações da China, em especial o referente a preços futuros, utilizando conjuntos nebulosos e métodos de séries de tempo nebulosas multivariadas. O processo de modelagem no artigo é feito através da combinação de modelos de séries temporais nebulosas tradicionais e o método do conjunto aproximado. Durante a criação dos modelos nebulosos, não somente os valores de fechamento das ações, mas também preços spot (meta) e de volume de trade são utilizados na criação dos modelos nebulosos.

O trabalho presente em (TALARPOSHTI et al., 2016) propõe um método que mescla séries temporais nebulosas exponenciais e processo de otimização baseados em aprendizados de autômatos.

Os autores dividem a metodologia em duas fases distintas. Na primeira, intervalos ótimos são obtidos através de séries nebulosas convencionais juntamente com um algoritmo de aprendizado de autômatos para a escolha ideal do comprimento dos intervalos de dados a serem utilizados. Em uma segunda etapa, os intervalos obtidos na primeira fase são utilizados para gerar uma nova série nebulosa. Um processo de otimização é utilizado para estimativa de alguns parâmetros.

O trabalho é utilizado na tentativa de se prever valores em séries temporais de índices de ações. Os resultados apresentados pelos autores se mostram bastante assertivos quando comparados com outras formas de modelagem nebulosa.

(CHENG; YANG, 2018) apresentam uma perspectiva um pouco diferente no que diz respeito a sistemas nebulosos aplicados a modelagem no mercado financeiro. Nada mais lógico é o fato de que predição de valores em mercado de ações se constitui uma das ferramentas mais úteis não somente para o investidor individual, mas também para analistas de Economia e Finanças de forma geral. Os autores desse artigo enfatizam esse fato e salientam que se utilizam de modelos de séries temporais nebulosas que se mostram, na maioria das vezes, mais adequados para previsão de séries temporais como no mercado de ações do que modelos de séries temporais

convencionais.

Os autores inovam um pouco na elaboração do modelo nebuloso e substituem as regras lógicas nebulosas por regras baseadas em algoritmos para realizar a extração de regras de previsão a partir de observações de séries temporais.

A avaliação do método proposto pelos autores se mostrou adequada e assertiva na aplicação de alguns índices principais de bolsas internacionais como TAIEX, Nikkei e HSI por um período de 14 anos. Pelas conclusões dos autores, o modelo proposto supera a performance de modelos tradicionais nebulosos que procuram prever os mesmos tipos de séries temporais financeiras.

O artigo (CHEN; CHEN, 2015) se apresenta como mais um trabalho que levanta em questão o grande esforço de pesquisa implementado para predição de séries temporais no mercado de ações envolvendo lógica nebulosa. As vantagens obtidas a partir de uma predição acurada podem ser percebidas não apenas por pessoas físicas investidoras, mas também por gestores e administradores de empresas que desejam prever riscos assim como analisar tendências econômicas como um todo.

O trabalho citado, como vários outros já descritos nesta tese, também aborda modelagem utilizando modelos de séries temporais nebulosas e aponta um problema que é crucial nesse tipo de modelagem: a definição do comprimento dos intervalos amostrados. Os autores propõem um novo modelo nebuloso para predição das séries temporais baseado em computação granular. Segundo afirmam os autores, o modelo proposto baseia-se na abordagem de computação granular com métodos de partição baseada em binning (separação entre componentes de maior e menor desempenho) e de discretização baseada em entropia. O modelo é testado em séries de dados de índices de ações de diferentes países, dentre eles o Ibovespa e se mostra mais acurado em termos de predição quando comparado a modelos clássicos.

Em (CHU et al., 2009) observamos a abordagem de um tema que há muito já é levantado no mercado de ações: o de que existe uma relação extremamente forte entre o volume transacionado e os retornos proporcionados na compra e venda de ações. Os autores propõem verificar essa relação através de um modelo de séries temporais nebulosas modificado de dois fatores. Os resultados obtidos superam, em termos de performance, alguns modelos listados atualmente na literatura.

Os autores do artigo (ATSALAKIS; DIMITRAKAKIS; ZOPOUNIDIS, 2011) apresentam o chamado sistema WASP (*Wave Analysis Stock Prediction*), que é um sistema que possui uma arquitetura neuro-nebulosa que utiliza aspectos da teoria de ondas de Elliot. A técnica apresentada pelos autores é utilizada para prever a tendência nos preços de ações. Os resultados apresentados no artigo se mostram mais adequados para serem utilizados como estratégia do que a estratégia *buy and hold* que se consiste na compra de um ativo e no manutenção do mesmo por um longo período. Este é mais um trabalho que mostra a utilidade recente de sistemas

neuro-nebulosos para modelagem e predição de dados em séries temporais.

No trabalho de (ATSALAKIS; VALAVANIS, 2009) é apresentado um modelo neural-nebuloso de predição de valores no mercado de ações em que um controlador do tipo ANFIS (*Adaptive Neuro Fuzzy Inference System*) tenta diminuir os erros de predição à medida em que a rede é treinada. O modelo criado pelos autores se mostrou adequado para realizar a predição de um dia à frente no mercado de ações americano. Os resultados apresentados no artigo mostram a enorme utilidade das lógicas de modelagem por redes neurais e sistemas nebulosos na predição de valores de ativos no mercado financeiro. O artigo não contempla taxas de comissão e impostos, algo que não menospreza os resultados da modelagem, uma vez que apenas representam valores fixos ou valores que são funções do volume transacionado.

3 Desenvolvimento de modelos e resultados

- Séries Temporais

O capítulo apresenta o desenvolvimento dos modelos e os resultados obtidos na tese para séries temporais. O conceito e explicação de construção da estrutura de dados árvore rubro-negra são apresentados. É apresentada a forma como os modelos para séries temporais foram construídos, bem como são apresentadas simulações com dados reais comprovando a eficácia dos modelos quando comparados a modelos existentes na literatura. O capítulo finaliza apresentando discussões detalhadas sobre os resultados de modelagem além de conclusões a respeito de seu desempenho.

3.1 Modelo fuzzy baseado na distribuição estatística dos dados de saída

3.1.1 Formulação Matemática da lógica fuzzy

A teoria da Lógica Fuzzy (LF) é utilizada para o tratamento de incertezas, ou seja, informações vagas, ambíguas e qualitativamente incompletas e imprecisas. A LF, também conhecida como Lógica Nebulosa ou Teoria dos Conjuntos Nebulosos possui habilidade em inferir conclusões e gerar respostas baseadas em variáveis linguísticas que podem ser subjetivas. É a teoria clássica para o tratamento da informação imprecisa e vaga. Em outras palavras, ela pode ser entendida tanto como uma maneira de reproduzir o conhecimento e o senso comum funcionando como uma interface entre números e os símbolos, assim como uma ferramenta para construir funções numéricas ou modelos matemáticos. Um tratamento fuzzy possui as seguintes etapas típicas: fuzzyficação, inferência e defuzzyficação.

3.1.2 Modelagem Matemática Fuzzy

Ao invés de usar a Lógica Fuzzy clássica, este modelo usa uma abordagem diferente. Ele começa com os dados de entrada e os mapeia no espaço de saída usando funções lineares de variáveis de entrada. As etapas a seguir mostrarão o modelo proposto para este trabalho.

- Etapa 1 - Coleta de dados
- Etapa 2 - Preparação de dados
- Etapa 3 - Fuzzyficação

- Etapa 4 - Defuzzyficação

Para um estudo mais aprofundado do assunto, recomenda-se os trabalhos de (SONG; CHISSOM, 1993b), (SONG; CHISSOM, 1993a), (CHEN, 1996), (YU, 2005), (TEOH et al., 2008) e (SINGH; DHIMAN; KAUR, 2018).

Cada uma das etapas mencionadas anteriormente serão descritas em detalhes a seguir.

3.1.2.1 Etapa 1: Coleta de dados

Os dados de entrada são organizados na forma de uma matriz de entrada e os dados de saída são organizados na forma de um vetor de saída. A representação da matriz de entrada do modelo é mostrada na Equação 3.1.

$$X = X_{rxN} \quad (3.1)$$

A Equação 3.1 apresenta uma matriz onde cada linha representa uma variável e cada coluna representa uma observação de uma variável. A Equação 3.2 mostra uma forma mais detalhada da Equação 3.1 e é chamada de matriz de entrada.

$$X_{rxN} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2N} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{iN} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{r1} & x_{r2} & \dots & x_{rj} & \dots & x_{rN} \end{pmatrix} \quad (3.2)$$

Seguindo a mesma lógica, o vetor de saída (que representará a variável a ser prevista) é representado nas Equações 3.3 e 3.4.

$$Y = Y_{1xN} \quad (3.3)$$

$$Y_{1xN} = (y_1 \quad y_2 \quad \dots \quad y_j \quad \dots \quad y_N) \quad (3.4)$$

3.1.2.2 Etapa 2: Preparação dos dados

Os dados da matriz de entrada na Equação 3.2 serão agrupados em c clusters (ou grupos) de acordo com os dados de saída Y .

Primeiro, divida os dados de saída Y em c clusters. O valor de c pode ser escolhido com base no desempenho final do algoritmo usando testes. Como exemplo, faça $c = 2$. Isso significa que os dados de saída devem ser divididos em 2 clusters. A divisão deve seguir as etapas:

- Calcule a média de saída y_{media} (ou mediana $y_{mediana}$, dependendo do desempenho do algoritmo)
- Atribua ao cluster 1 todos os valores de saída y_j menores ou iguais a y_{media}
- Atribua ao cluster 2 todos os valores de saída y_j maiores que y_{media}

Observe que o conjunto de dados Y é dividido em 2 clusters. Em seguida, atribua todas as entradas de X , x_{ij} ao mesmo cluster de sua saída correspondente y_j e os limites dos clusters são definidos. A próxima etapa é testar a normalidade sobre os dados de entrada. Como a média populacional e o desvio padrão não estão disponíveis, o teste deve ser feito usando estimativas desses parâmetros com base em amostras.

O teste retornará uma decisão para a hipótese nula de que os dados no vetor X_i (cada linha da matriz X) vêm de uma distribuição normal com uma média e variância estimadas do próprio vetor, usando o teste de ajuste de qui-quadrado.

O objetivo do teste é verificar se um valor observado de um vetor de dados específico está significativamente próximo do valor esperado, supondo que o vetor corresponda a uma distribuição de probabilidade específica. Em outras palavras, ele determina o quão bem a distribuição teórica (como normal, binomial ou Poisson) se ajusta à distribuição empírica. Para o modelo proposto, os dados, conforme declarado acima, são comparados com a distribuição normal.

A Equação 3.5 mostra como o teste de qui-quadrado é calculado.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3.5)$$

onde O_i é o valor observado e E_i é o valor esperado.

Caso o teste rejeite a hipótese nula ou normalidade, os dados devem ser normalizados. A Equação 3.6 mostra uma maneira de normalizar os dados.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.6)$$

onde x é o vetor de dados original e x_{new} é o vetor de dados normalizado usando os valores máximo e mínimo.

3.1.2.3 Etapa 3: Fuzzyficação

Uma vez que os clusters são definidos, a próxima etapa é definir a associação de cada valor x_{ij} para cada cluster. Considerando o cenário de 2 clusters, cada entrada x_{ij} terá 2 valores de pertinência diferentes, cada um correspondendo a cada cluster.

A média e o desvio padrão para cada cluster podem ser expressos pelas Equações 3.7 e 3.8.

$$\mu_i^k = \frac{1}{N^k} \sum_{j=1}^{N^k} x_{ij}^k \quad (3.7)$$

$$\sigma_i^k = \sqrt{\frac{1}{N^k - 1} \sum_{j=1}^{N^k} (x_{ij}^k - \mu_i^k)^2} \quad (3.8)$$

onde μ_i^k é a média aritmética do cluster k, σ_i^k é o desvio padrão do cluster k e N^k é o número de elementos do cluster k.

A função multivariada de densidade de probabilidade normal para uma variável i é usada em cada cluster para todo o conjunto de dados e os parâmetros para o cluster k para definir o valor de associação, como pode ser visto na Equação 3.9:

$$f_{jk}(x_j, \mu^k, \sigma^k) = \frac{1}{\sigma^k \sqrt{(2\pi)}} e^{-\frac{(x_j - \mu^k)^2}{2(\sigma^k)^2}} \quad (3.9)$$

onde f_{jk} é o valor da função, x_j é a j -ésima observação da variável x , μ^k e σ^k são a média e desvio padrão do cluster k.

3.1.2.4 Etapa 4: Defuzzyficação

A previsão da saída y_j , O_j , é uma soma ponderada das funções de saída onde as funções de saída são tantas quanto forem o número de clusters. Os pesos, para cada variável i , podem ser definidos pela Equação 3.10.

$$\tau_{jk} = \frac{f_{jk}}{\sum_{k=1}^c f_{jk}} \quad (3.10)$$

onde τ_{jk} é o peso da j -ésima observação do k -ésimo cluster e c é o número de clusters.

A saída O_j é considerada a soma ponderada das variáveis de entrada como representado na Equação 3.11:

$$\begin{aligned} O_j &= \sum_{k=1}^c \tau_{jk} \vec{a} \vec{x}^T \\ \vec{a} &= \begin{pmatrix} p_{k1} & p_{k2} & \dots & p_{k(r+1)} \end{pmatrix} \\ \vec{x} &= \begin{pmatrix} 1 & x_{1j} & x_{2j} & \dots & x_{rj} \end{pmatrix} \end{aligned} \quad (3.11)$$

onde O_j é a saída prevista, \vec{a} é um vetor de coeficientes desconhecidos, r é o número de variáveis e p_{kq} são os coeficientes desconhecidos. Para o cálculo desses coeficientes desconhecidos, um

problema de otimização pode ser definido com a função objetivo especificada conforme a Equação 3.12:

$$J = \left(\sum_{k=1}^c \tau_{jk} \sum_{j=1}^N \vec{a} \vec{x}^T - y_j \right)^2 \quad (3.12)$$

onde y_j é o j -ésimo valor de saída.

Calculando a derivada de J em relação a p_{kq} e tornando-a igual a zero, o sistema de equações lineares mostrado na Equação 3.13 é obtido.

$$\sum_{k=1}^c \tau_{jk} \sum_{j=1}^N p_{kq} \vec{x}_{qj} - y_j = 0, \forall j \in [1, N] \quad (3.13)$$

onde N é o número de elementos.

A Equação 3.13 pode ser expressa em formato matricial conforme pode ser verificado na Equação 3.14.

$$\underbrace{\begin{pmatrix} \tau_{11} & \dots & \tau_{c1} & \tau_{11}x_{11} & \dots & \tau_{c1}x_{11} & \dots & \tau_{11}x_{r1} & \dots & \tau_{c1}x_{r1} \\ \tau_{12} & \dots & \tau_{c2} & \tau_{12}x_{12} & \dots & \tau_{c2}x_{12} & \dots & \tau_{12}x_{r2} & \dots & \tau_{c2}x_{r2} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots & \dots & \dots & \vdots \\ \tau_{1j} & \dots & \tau_{cj} & \tau_{1j}x_{1j} & \dots & \tau_{cj}x_{1j} & \dots & \tau_{1j}x_{rj} & \dots & \tau_{cj}x_{rj} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots & \dots & \dots & \vdots \\ \tau_{1N} & \dots & \tau_{cN} & \tau_{1N}x_{1N} & \dots & \tau_{cN}x_{1N} & \dots & \tau_{1N}x_{rN} & \dots & \tau_{cN}x_{rN} \end{pmatrix}}_A \times \underbrace{\begin{pmatrix} p_{11} \\ \vdots \\ p_{c1} \\ p_{12} \\ \vdots \\ p_{c2} \\ \vdots \\ p_{1(r+1)} \\ \vdots \\ p_{c(r+1)} \end{pmatrix}}_P = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{pmatrix}}_Y \quad (3.14)$$

Como existem mais parâmetros desconhecidos do que equações, o problema é indeterminado e P pode ser facilmente determinado pelo método dos mínimos quadrados conforme Equação 3.15.

$$P = (A^T A)^{-1} A^T Y \quad (3.15)$$

3.1.3 Algoritmo do modelo baseado na distribuição estatística dos dados de saída

Esta seção apresenta o algoritmo para o desenvolvimento computacional do primeiro modelo apresentado. Posteriormente, o algoritmo foi implementado usando Matlab para o desenvolvimento do trabalho presente nesta tese.

Algoritmo:

- Variáveis de entrada:
 - Faça $x = [\text{Var1}(1:\text{end}-1) \text{Var2}(1:\text{end}-1) \text{Var3}(1:\text{end}-1) \text{Var4}(1:\text{end}-1) \text{Var5}(1:\text{end}-1)]$; até o número de variáveis j
- Variável de saída:
 - $y = \text{Data1}(1:\text{end}-1)$;
- Faça $c =$ número de clusters
- Faça $l = \max(y) - \min(y)$
- Para $i = 1$ até tamanho de y faça
 - Atribua a saída $y(i)$ a um cluster de acordo com o quantil ao qual ela pertence no conjunto y
- Faça $p(i) =$ posições do cluster i na variável de saída y .
- Para $i = 1$ até c faça
 - Para $j = 1$ até número de variáveis de entrada faça
 - * Faça $\mu(i, j)$ a média do cluster i da variável j
 - * Faça $\sigma(i, j)$ o desvio padrão do cluster i da variável j
- Verifique se os dados de cada cluster possuem distribuição próxima da normal. Caso contrário, normalize-os.
- Para $i = 1$ até c faça
 - Para $j = 1$ até número de variáveis de entrada faça
 - * Calcule a pertinência de cada variável j para cada cluster i gerando $e(i, j)$
- Calcule a contribuição de cada cluster na saída fazendo. Para $i = 1$ até c faça
 - Para $j = 1$ até número de variáveis de entrada faça

* Calcule a contribuição do cluster i da variável j na saída gerando $\tau(i, j)$

- Crie a matriz auxiliar conforme Equação 3.14
- Calcule o vetor P conforme Equação 3.15
- Calcule o valor predito $O(j)$ fazendo $O(j) = A * P$ usando:
 - $x = [\text{Var1}(\text{end}) \text{Var2}(\text{end}) \text{Var3}(\text{end}) \text{Var4}(\text{end}) \text{Var5}(\text{end})]$; até Varj
- Repita todo o procedimento para prever cada valor no conjunto y .

3.1.4 Aplicação para predição de variáveis macroeconômicas

3.1.4.1 Dados utilizados

Os dados estudados neste trabalho são o conjunto de séries temporais disponibilizadas publicamente pelo BCB (Bacen - Banco Central do Brasil) em seu site. Os dados podem ser obtidos de páginas como (BRASIL, 2018a), enquanto as informações do Cosif (Plano Contábil das Instituições do Sistema Financeiro Nacional) podem ser obtidas na página Informações para Análise Econômica e Financeira (BRASIL, 2018b), ambas no site do BCB. As mesmas informações também podem ser consultadas, de forma mais consolidada e por meio de recursos gráficos, no site do Banco Data na internet (DATA, 2018).

A Tabela 1 resume como os parâmetros são representados daqui em diante.

Variável	Descrição
AplIfLiq	Interbank investments
CNiv2	Tier 2 Reference Equity
Dispon	Availabilities
RdOC	Loan from Credit Operations
RdS	Rents from Service
RdTVM	Marketable securities
RWACred	Credit risk
TVM_Deriv_A	Derivatives of Securities

Tabela 1 – Parâmetros bancários e sua correspondente descrição

3.1.4.2 Estudo de Correlação - Coeficientes de correlação para a hipótese de correlação com significância maior que 95 %

A Tabela 2 apresenta os valores de p , $p < 0.05$, para os casos em que o coeficiente de correlação entre a taxa de juros Selic e uma determinada variável indicou existência de correlação, ao nível de significância 95%.

Parâmetro	Valor- <i>p</i>
AplIfLiq	0,000001401387586
CNiv2	0,000000000002951
Dispon	0,000177453747735
RdOC	0,000828889429702
RdS	0,001671917767695
RdTVM	0,045510260149990
RWACred	0,000000013197539
TVM_Deriv_A	0,000001871781982

Tabela 2 – Valores *p* para o teste de hipótese de correlação entra a taxa Selic e certos parâmetros bancários. Um valor-*p* menor ou igual a 0.05 indica correlação significativa entre a Selic e um determinado parâmetro.

Na Tabela 2 observa-se que, exceto para RDTVM, todos os parâmetros têm um valor-*p* menor que 0.01. Ou seja, com exceção deste parâmetro, existe uma correlação significativa, superior a 99%, entre a Selic e os parâmetros apresentados na tabela.

3.1.4.3 Simulação

Para validar o modelo proposto e também testar seu desempenho, é considerado um subconjunto de variáveis do mesmo banco de dados experimental utilizado durante a análise de correlação.

O objetivo é criar e testar um modelo de previsão da taxa Selic com base nos valores anteriores da própria taxa Selic juntamente com as variáveis da Tabela 2 que apresentaram correlação significativa com a Selic na análise anterior.

Nesta seção, dois exemplos de teste diferentes são fornecidos para ilustrar o método proposto.

Quatro modelos existentes da literatura (CHEN, 1996), (YU, 2005), (TEOH et al., 2008), (SINGH; DHIMAN; KAUR, 2018) e seus resultados de previsão são comparados aos resultados do modelo proposto sobre o conjunto de dados de treinamento (teste dentro da amostra). Em seguida, os mesmos 5 modelos (os 4 existentes mais o novo modelo proposto) são usados para prever amostras fora do conjunto de dados de treinamento (teste fora da amostra).

A comparação entre os modelos será realizada utilizando *RMSE* e *RMSE comparado ao naive* de acordo com as Equações 3.16 e 3.17 conforme apresentado a seguir.

$$RMSE_{model} = \sqrt{\frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{n}} \quad (3.16)$$

$$RMSE_{relativenaive} = \sqrt{\frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{\sum_{k=1}^n (y(k) - \hat{y}_{naive}(k))^2}} \quad (3.17)$$

Nas Equações 3.16 and 3.17, $y(k)$ é o valor atual do dado, \hat{y} é o dado previsto por um modelo específico e $\hat{y}_{naive}(k)$ é o dado previsto de acordo com o modelo “Naive”.

3.1.4.4 Base de dados utilizada

Os dados a serem usados como entradas durante as fases de treinamento e teste são descritos na Tabela 3.

Variável	Descrição
AplIfLiq	Investimentos Interbancários
CNiv2	Patrimônio de Referência Nível 2
Dispon	Disponibilidades
RdOC	Empréstimo de Operações de Crédito
SelicPrev	Valor anterior (1 passo atrás) da Selic

Tabela 3 – Descrição das variáveis que serão utilizadas pelo modelo de previsão.

A Tabela 4 mostra todos os dados a serem utilizados (um total de 65 amostras), incluindo o valor anterior da Selic, $Selic(t - 1)$, e variáveis com alta correlação com a Selic de acordo com a análise de correlação anterior. Esses dados refletem a Selic trimestral e os parâmetros bancários de dezembro de 2001 a março de 2018.

3.1.4.5 Teste dentro da amostra de treinamento

Para os testes sobre os dados de treinamento, foram utilizadas as primeiras 55 amostras dos valores trimestrais da Selic (taxa de juros principal do Brasil) de dezembro de 2001 a março de 2018.

Para o treinamento do modelo foram utilizadas as primeiras 55 amostras de um total de 65 amostras da Selic e alguns parâmetros bancários específicos.

A Tabela 5 resume os resultados finais do teste em forma numérica, mostrando os dados usados para construir a Figura 9 e a Figura 10. A Tabela 6 resume o teste sobre RMSEs de treinamento. Ela mostra os RMSEs dos quatro modelos existentes e o modelo proposto.

A Figura 9 mostra os resultados do teste durante o treinamento dos modelos em gráficos separados. A Figura 10 mostra os mesmos resultados da Figura 9 com todos os gráficos sobrepostos para análise simultânea.

A partir das Figuras 9 e 10 e Tabelas 5 e 6 pode-se ver facilmente como o modelo proposto supera 3 dos modelos existentes e se aproxima do modelo de Yu. Durante o treinamento, o modelo de Yu seleciona bases específicas para agrupamento em vez de assumir uma distribuição normal. Esse pode ser o motivo pelo qual o modelo Yu foi o melhor durante o treinamento.

Amostra	Original Selic	AplfLiq	CNiv2	Dispon	RdOC	Selicprev
1	1.37	98136953294.7700	12738882946.8000	22600569890.6600	2432397744.96000	1.39
2	1.33	91488965640.0300	13182630825.9200	22776632906.3600	1279213452.41000	1.37
3	1.38	85735258433.1400	19219732948.3300	27722933553.2100	-898444058.710000	1.33
4	1.74	129649019571.840	20563781289.0900	34916309047.2200	-3871007970.23000	1.38
5	1.78	140013532026.820	23916692516.6100	29376322620.9200	-4226890621.20000	1.74
6	1.86	139728229914.750	23450580361.1900	24278086094.5500	-84543371.3100000	1.78
7	1.68	118122758138.230	21944479475.2800	20720205335.1000	-887785201.070000	1.86
8	1.37	147494389808.120	22522185924.8900	22754609121.3800	353783431.530000	1.68
9	1.38	165753514586.430	26054592761.0600	28559772600.9400	620639060.170000	1.37
10	1.23	163817210706.360	26954418135.6400	32873356173.0300	235225172.200000	1.38
11	1.25	161976014791.800	28700272811.1500	33797255585.0100	-485421074.170000	1.23
12	1.48	163734206282.770	29477755179.8800	34659489935.1600	1342901535.07000	1.25
13	1.53	143998238599.030	27866511036.7700	32611200856.6900	3529168640.78000	1.48
14	1.59	144770197521.250	28684822649.6800	31772772653.2300	420378718.050000	1.53
15	1.5	149351381179.610	28581772733.9300	19471637902.5000	2564652411.02000	1.59
16	1.47	187876697772.360	29166912923.5700	23427638014.0900	1936000687.00000	1.5
17	1.42	181852483220.870	34206216235.1400	23414192322.0000	1642377812.81000	1.47
18	1.18	202308364802.490	38024015074.8800	23029886517.4100	2085367788.70000	1.42
19	1.06	216593113651.060	43474968610.1800	22453112025.3800	4161858952.18000	1.18
20	0.99	203163771846.600	50496261859.0700	22653233801.0400	-14322097.8900000	1.06
21	1.05	213484257685.250	53063569835.2600	23696353998.6300	63044801.0100000	0.99
22	0.91	257186175796.000	60434999491.0700	24914635409.6900	382907297.420000	1.05
23	0.8	309764401479.630	72023774319.9400	24948982625.0100	-739437837.030000	0.91
24	0.84	339050472835.330	75011508001.5300	24411741977.2300	1397523808.32000	0.8
25	0.84	341429362746.180	77067116476.6000	29375102265.0500	3805010424.28000	0.84
26	0.96	369838009289.910	78822009180.6300	27970613710.0300	489767769.350000	0.84
27	1.1	422315698466.430	80995754722.0100	29159933329.8200	6739190147.95000	0.96
28	1.12	439912047066.240	79147464550.1500	35697400818.2200	-3748711737.26000	1.1
29	0.97	502218311745.360	84642166973.9600	43687794158.9900	-11257562396.8100	1.12
30	0.76	560931615827.480	86070794312.9700	39128044965.9200	1787406936.74000	0.97
31	0.69	570111083339.090	86141457944.1900	37736263570.9000	7145183649.65000	0.76
32	0.73	623050632857.490	90029598939.1800	37841824350.2100	623648943.290000	0.69
33	0.76	622875934674.180	96026935977.2300	40561063253.5100	1454991075.04000	0.73
34	0.79	591613546505.350	103167994069.190	38762032587.7400	-708059692.840000	0.76
35	0.85	546633827289.460	104665443278.700	40484934819.4600	-365046008.160000	0.79
36	0.93	565254593823.120	109556277945.230	43900284325.1500	-1172030387.76000	0.85
37	0.92	454215512368.050	126131756262.530	50904340454.7200	-1815393234.75000	0.93
38	0.96	559207468534.920	135775997142.260	46983779756.2400	-291951660.400000	0.92
39	0.94	556294750465.990	143675136751.510	57308510631.3400	-2403666813.61000	0.96
40	0.91	611608500513.090	149385635381.210	58716504951.5100	984717830.920000	0.94
41	0.82	592726639345.770	157601405181.480	60924438658.6900	26555873.6600000	0.91
⋮	⋮	⋮	⋮	⋮	⋮	⋮
61	1.05	1432647745784.56	179412671802.430	75805842614.8200	3867613790.26000	1.12
62	0.81	1465533186099.35	173288624404.090	75956481594.3100	17620158.6200000	1.05
63	0.64	1532039630802.90	181274511693.790	82174446187.3300	1897787278.25000	0.81
64	0.54	1516773860894.98	185648691453.900	78342609464.0100	-2051183374.09000	0.64
65	0.53	1407024529850.10	190613090124.350	85932954375.7400	-787842232.370000	0.54

Tabela 4 – Dados experimentais usados para treinamento e teste de modelo.

Amostra	Original Selic	Chen, 1996	Yu, 2004	Teoh, 2008	Singh, 2018	Proposed Model
2	1.33	1.49	1.46	1.53	1.51	1.38
3	1.38	1.49	1.34	1.53	1.47	1.47
4	1.74	1.49	1.46	1.53	1.51	1.44
5	1.78	1.69	1.74	1.41	1.51	1.71
6	1.86	1.69	1.82	1.41	1.51	1.69
7	1.68	1.69	1.66	1.41	1.51	1.76
8	1.37	1.49	1.34	1.41	1.45	1.58
9	1.38	1.49	1.46	1.53	1.51	1.38
10	1.23	1.49	1.46	1.53	1.51	1.40
11	1.25	1.21	1.38	1.53	1.47	1.39
12	1.48	1.21	1.38	1.53	1.47	1.36
13	1.53	1.49	1.47	1.41	1.45	1.49
14	1.59	1.49	1.58	1.41	1.53	1.56
15	1.50	1.49	1.46	1.41	1.53	1.50
16	1.47	1.49	1.47	1.41	1.53	1.39
17	1.42	1.49	1.47	1.41	1.45	1.41
18	1.18	1.49	1.13	1.41	1.45	1.28
19	1.06	1.21	1.02	1.09	1.08	1.04
20	0.99	1.01	1.01	0.95	1.08	1.06
21	1.05	1.01	1.02	0.95	1.03	0.99
22	0.91	1.01	1.01	0.95	1.03	1.01
23	0.80	0.82	0.90	0.95	0.92	0.92
24	0.84	0.82	0.82	0.69	0.84	0.80
25	0.84	0.82	0.91	0.95	0.84	0.84
26	0.960	0.82	0.91	0.95	0.84	0.87
27	1.10	1.01	0.98	0.95	0.92	0.99
28	1.12	1.01	1.10	1.09	1.08	1.10
29	0.970	1.21	1.07	1.09	1.08	0.99
30	0.76	1.01	0.98	0.95	0.92	0.94
31	0.69	0.82	0.74	0.69	0.78	0.78
32	0.73	0.73	0.70	0.69	0.75	0.73
33	0.76	0.82	0.77	0.69	0.75	0.71
34	0.79	0.82	0.74	0.69	0.78	0.79
35	0.85	0.82	0.82	0.69	0.78	0.77
36	0.93	0.82	0.91	0.95	0.84	0.81
37	0.92	1.01	0.90	0.95	0.94	0.85
38	0.96	1.01	0.90	0.95	0.94	0.85
39	0.94	1.01	0.98	0.95	0.92	0.91
40	0.91	1.01	0.90	0.95	0.94	0.87
41	0.82	0.82	0.90	0.95	0.92	0.85
42	0.64	0.82	0.82	0.69	0.84	0.81
43	0.54	0.73	0.54	0.69	0.61	0.65
44	0.55	0.73	0.56	0.59	0.61	0.59
45	0.55	0.73	0.56	0.59	0.61	0.61
46	0.61	0.73	0.56	0.59	0.61	0.59
47	0.71	0.73	0.70	0.59	0.61	0.63
48	0.79	0.73	0.77	0.69	0.75	0.76
49	0.77	0.82	0.82	0.69	0.78	0.78
50	0.82	0.82	0.74	0.69	0.78	0.82
51	0.91	0.82	0.82	0.69	0.84	0.87
52	0.96	0.82	0.90	0.95	0.92	0.96
53	1.04	1.01	0.98	0.95	0.92	1.02
54	1.07	1.01	1.01	0.95	1.03	1.10
55	1.11	1.01	1.10	0.950	1.08	1.11

Tabela 5 – Previsão de resultados com base nos dados de treinamento.

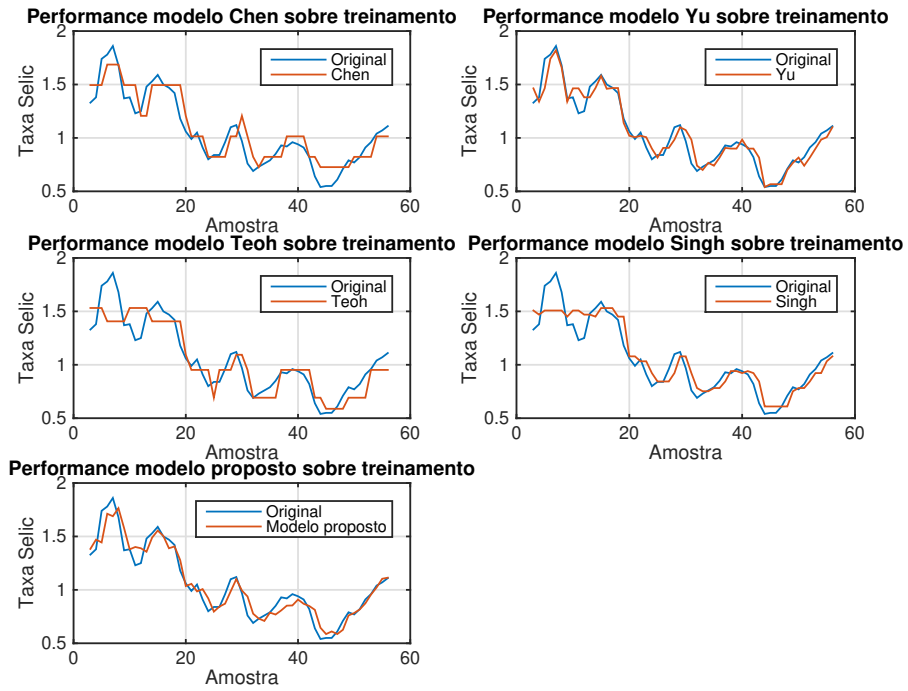


Figura 9 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais, considerando 4 modelos conhecidos da literatura e o modelo proposto.

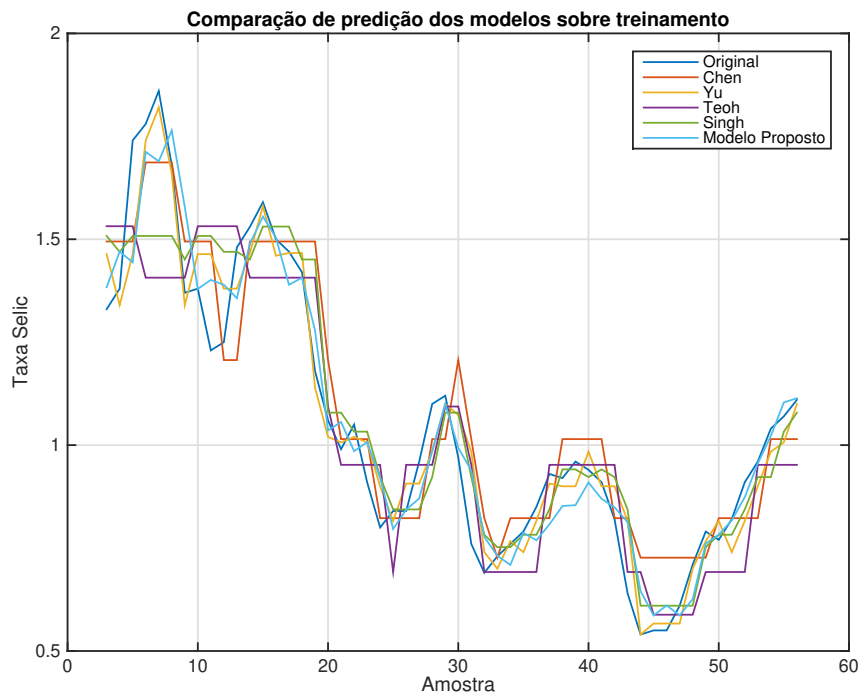


Figura 10 – Curva de previsão sobre dados de treinamento com gráficos sobrepostos.

Modelo	RMSE	RMSE relativo ao Naive
Chen 1996	0.1275	1.1254
Yu 2004	0.0834	0.7364
Teoh 2008	0.1505	1.3287
Singh 2018	0.1217	1.0748
Modelo Proposto	0.0852	0.8177

Tabela 6 – Modelos de Comparação de desempenho de modelos sobre dados de treinamento utilizando 55 amostras de treinamento.

3.1.4.6 Teste fora dos dados de treinamento

Para testar o modelo, as amostras 56 a 65 e uma técnica de janela deslizante foram usadas. Inicialmente, para prever a amostra 56, foi utilizado o modelo treinado com as amostras 1 a 55. A seguir, para prever a amostra 57, foi utilizado o modelo treinado com as amostras 2 a 56. E assim por diante.

A Tabela 7 resume o teste final na forma numérica, mostrando os dados usados para construir a Figura 11 e a Figura 12. A Tabela 8 resume os RMSEs de teste final. Ela mostra o RMSE dos 4 modelos existentes e o modelo proposto considerando o cenário de 55 amostras de treinamento. A tabela também mostra a média relativa do RMSE naive considerando cenários de teste usando 48 a 55 amostras.

A Figura 11 mostra os resultados do teste dos modelos em gráficos separados. A Figura 12 mostra os mesmos resultados da Figura 11 com todos os gráficos sobrepostos para análise simultânea.

A partir das Figuras 11 e 12 e Tabelas 7 e 8 pode-se concluir que o modelo proposto supera os outros modelos com os quais foi comparado. Além disso, foi o único modelo que manteve o RMSE médio em relação ao Naive inferior a 1 no que diz respeito à consistência.

Amostra	Original Selic	Chen, 1996	Yu, 2004	Teoh, 2008	Singh, 2018	Proposed Model
1	1.15	1.01	0.940	1.09	1.07	1.14
2	1.15	1.21	1.02	1.09	1.08	1.15
3	1.15	1.21	1.10	1.08	1.09	1.18
4	1.11	1.21	1.10	1.08	1.09	1.15
5	1.12	1.01	1.07	1.07	1.09	1.19
6	1.05	1.21	1.09	1.05	1.10	1.13
7	0.81	1.03	1.01	1.04	1.02	1.04
8	0.64	0.76	0.82	0.80	0.77	0.74
9	0.54	0.69	0.54	0.69	0.61	0.52
10	0.53	0.69	0.56	0.59	0.61	0.56

Tabela 7 – Resultados do teste comparando dados originais com 4 modelos conhecidos e modelo proposto com 55 amostras de treinamento.

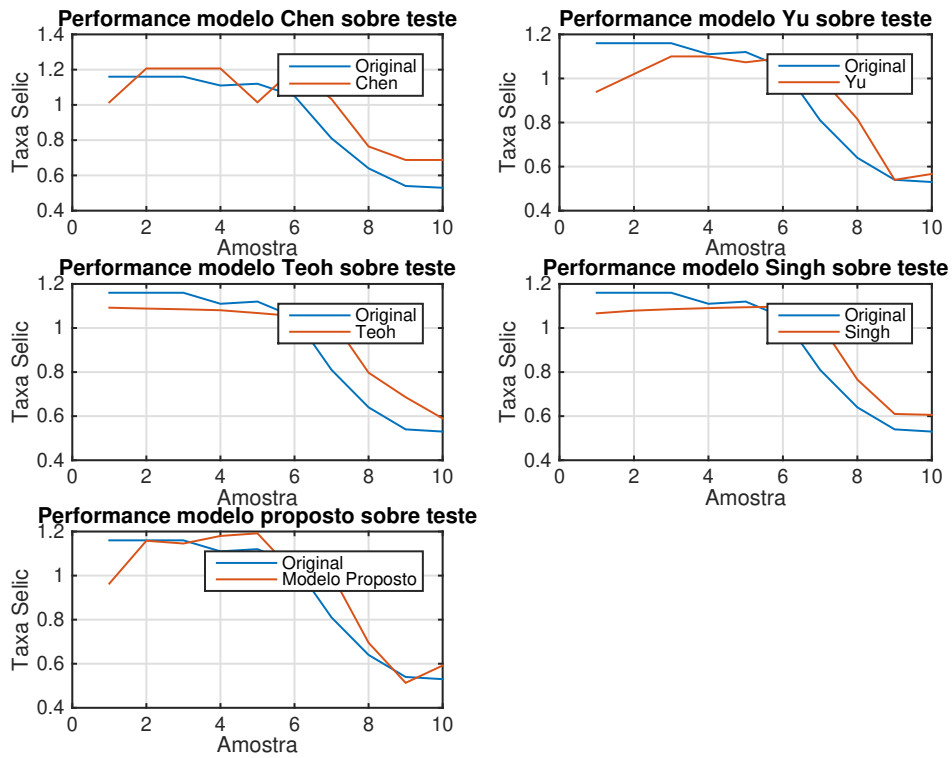


Figura 11 – Curva de previsão sobre os dados de teste em comparação com os dados experimentais originais, considerando 4 modelos conhecidos da literatura e o modelo proposto.

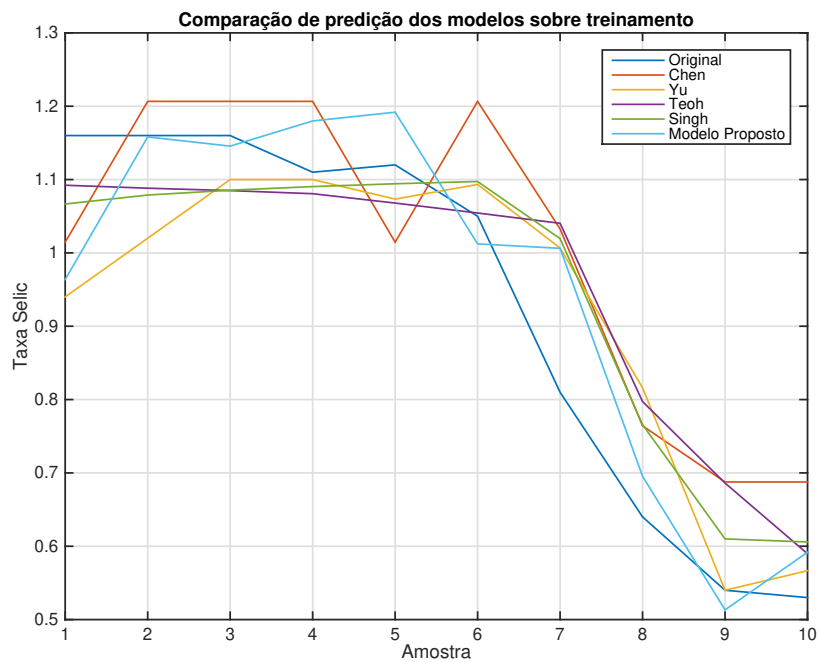


Figura 12 – Curva de previsão sobre dados de teste com gráficos sobrepostos.

Modelo	RMSE 55 Amostras de Treinamento	RMSE relativo ao Naive com 55 amostras de treinamento	RMSE médio relativo ao Naive 48 a 55 amostras de treinamento
Chen 1996	0.1349	1.3070	1.3200
Yu 2004	0.1211	1.1733	1.2138
Teoh 2008	0.1103	1.0684	1.3425
Singh 2018	0.0972	0.9414	1.0988
Modelo Proposto	0.0898	0.8701	0.9688

Tabela 8 – Comparação de desempenho entre o modelo proposto e os modelos existentes no banco de dados de teste. RMSE médio usando amostras de treinamento variando de 48 a 55.

3.1.5 Discussão

A ideia do trabalho foi avaliar o impacto na taxa básica de juros da economia brasileira, a Taxa Selic, em decorrência das variações nos riscos assumidos, na situação de capital das instituições financeiras que atuam no país de dezembro de 2001 a março de 2018, no patrimônio de referência da nível II e algumas contas do Plano de Contabilidade das Instituições do Sistema Financeiro Nacional (Cosif), com o objetivo de tentar explicar como a taxa da economia brasileira se correlaciona com esses parâmetros.

Foi verificada a existência de correlação entre a taxa Selic e alguns desses parâmetros definidos pelos acordos da Basiléia. Com exceção do RdTVM, os dados demonstraram correlação significativa entre a taxa Selic e os parâmetros superior a 99%. Além disso, a análise de correlação mostrou que, na maioria dos casos em que havia correlação significativa, essa correlação com a Selic era negativa. Ou seja, os valores desses parâmetros aumentavam / diminuía à medida que a Selic diminuía / aumentava.

Observou-se que em todos os casos em que houve correlação significativa, essa correlação apresentou o maior valor nos atrasos entre 6 e 12 meses. Isso indica que a maioria desses parâmetros foi influenciada (ou influenciou) pelos valores da taxa Selic de 6 a 12 meses atrás.

Essas relações estão resumidas nas Tabelas 9 e 10 e indicam a confirmação das hipóteses para a maioria dos parâmetros elencados: quando ocorre redução da taxa Selic, as instituições financeiras diversificam suas operações em busca de alternativas de receita. Além disso, a redução da taxa Selic acarreta aumento do patrimônio de referência nível II (CNiv2), reflexo da maior disponibilidade de recursos para financiar as atividades da instituição, conforme esperado.

Selic	RWACred	Cniv2	RdOC	AplifLiq
↑	↓	↓	↓	↓
↓	↑	↑	↑	↑

Tabela 9 – Resumo da influência do Selic sobre os parâmetros - parte 1

Selic	TVM_Deriv_A	Dispon	RdS	RdTVM
↑	↓	↓	↓	↓
↓	↑	↑	↑	↑

Tabela 10 – Resumo da influência do Selic sobre os parâmetros - parte 2

Tais resultados mostram a necessidade de aprofundar o estudo dessa questão, mas, por outro lado, confirmam o conhecimento sobre a relação inversa entre Selic e parâmetros bancários, e acredita-se que contribuam para antecipar as tendências desses parâmetros no futuro próximo, com base em valores atuais da taxa Selic.

3.1.6 Conclusões

A principal contribuição técnica do modelo nesta seção é um algoritmo preciso e eficiente para previsão de séries temporais. O algoritmo agrupa o universo fuzzy de discurso em clusters de acordo com os comandos do usuário. A divisão do universo do discurso em clusters foi feita igualmente, dividindo-se a quantidade total de dados pelo número de clusters. Outra forma de dividir o universo do discurso foi usando quantis, que forneceram resultados ligeiramente diferentes. Em seguida, uma função de densidade de probabilidade normal multivariada foi usada para calcular o valor de pertinência de cada elemento a cada cluster. Para o cálculo da saída (valor previsto) foi utilizado o somatório linear ponderado das variáveis de entrada, onde os pesos foram ajustados por meio do método convencional de mínimos quadrados. O método proposto foi então comparado com quatro modelos existentes na literatura e mostrou ter melhores resultados para as séries temporais investigadas.

A ideia de associar a oscilação da taxa Selic a alguns parâmetros bancários específicos, que podem ser considerados como a contribuição prática do trabalho, foi uma tentativa de responder a questões como: A taxa Selic está correlacionada com parâmetros bancários? É possível prever a Selic com base em alguns parâmetros? As respostas a essas perguntas serviram como uma etapa no método proposto e podem ajudar os gestores governamentais de alto nível a tomar decisões que possam afetar a economia do país como um todo.

Trabalhos futuros nessas áreas podem considerar a inclusão de um maior número de variáveis no modelo fuzzy, bem como a mudança no procedimento de divisão no universo fuzzy de discurso. Outra área emergente que pode fornecer percepções significativas são os conjuntos aproximados. O uso de conjuntos brutos em trabalhos futuros pode fornecer uma alternativa mais eficiente na elaboração de regras de lógica fuzzy (FLRs) no modelo fuzzy.

3.2 Modelo Fuzzy utilizando a árvore rubro negra

3.2.1 Árvores rubro-negras (Red-black trees)

Esta seção cobre a explicação da estrutura de dados em árvore rubro negra, conforme explicado por (GUIBAS; SEDGEWICK, 1978).

Uma árvore rubro negra é uma estrutura de dados em árvore. Uma característica específica da árvore rubro negra é que ela é uma árvore de busca binária com equilíbrio automático (GUIBAS; SEDGEWICK, 1978). O equilíbrio da árvore é mantido pelo uso de operações específicas de reequilíbrio de rotação e / ou recoloração de nós para sempre cumprir certas regras. As principais características / regras da árvore rubro-negra podem ser descritas a seguir.

1. É uma árvore de pesquisa binária com equilíbrio automático
2. Cada nó tem uma cor, vermelho ou preto. Dependendo de sua cor, eles são chamados de nós vermelhos ou nós pretos.
3. O nó raiz é um nó preto.
4. Se um nó é vermelho, seus filhos são pretos
5. Cada caminho de um nó para seu nó nulo (último nó de um ramo) descendente tem o mesmo número de nós pretos.
6. Nenhum nó vermelho tem um filho vermelho. (Caso contrário, há um conflito vermelho-vermelho)
7. Cada nó nulo é um nó preto.

Será usada apenas a função de inserção na árvore rubro negra neste modelo. Nenhuma exclusão será necessária, pois não haverá remoções da árvore após sua construção. As regras para inserção na árvore rubro-negra estão listadas abaixo.

1. Se a árvore estiver vazia, crie um novo nó como um nó raiz com a cor preta.
2. Se a árvore não estiver vazia, crie um novo nó como um nó folha com a cor vermelha.
3. Se o pai do novo nó for preto, termine.
4. Se o pai do novo nó é vermelho, então verifique a cor do irmão do pai do novo nó.
 - Se a cor for preta, faça a rotação adequada e recolora

- Se a cor for vermelha, então recolora ambos o pai e o irmão do pai e também verifique se o pai do pai do novo nó não é um nó raiz. Se não for o nó raiz, recolora-o e verifique novamente se há um conflito vermelho-vermelho considerando este nó como o novo nó.

5. Repita todo o processo até que não haja mais conflitos vermelho-vermelho.

A próxima seção apresenta um exemplo de construção de uma árvore rubro-negra.

3.2.2 Construção da árvore rubro negra

Como um exemplo da árvore rubro negra, vamos criar uma árvore rubro-negra inserindo os seguintes elementos. $X = \{7, 5, 10, 2, 6, 1, 3, 4\}$. Elementos repetidos são ignorados.

Inserindo elementos 7, 5, 10. Elemento 7 será a raiz da árvore e terá cor preta. Elemento 5 é um novo elemento (assim, será vermelho) e uma vez que é menor que 7, será inserido à sua esquerda. Elemento 10 será vermelho, e ficará à direita do elemento 7. A Figura 13 mostra a árvore parcial até o momento.

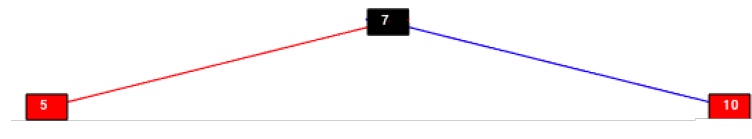


Figura 13 – Árvore rubro negra após a inserção dos elementos 7, 5 e 10

Inserindo elemento 2. Como 2 é menor que 7, ele estará à sua esquerda. Como também é menor que 5, estará à esquerda de 5. Como cada nó, em uma árvore não vazia, sempre é inserido como um nó vermelho, haverá um conflito vermelho-vermelho com os elementos 2 e 5.

O novo nó é 2, o pai do novo nó é 5. Visto que o pai do novo nó é vermelho, deve ser verificada a cor do irmão do pai (ou o “tio” do novo nó se assim quisermos chamá-lo). O novo tio possui valor 10, e é vermelho. Neste caso, uma recoloração deve ser feita (mudar a cor qualquer que seja) do pai e do tio. Neste caso, 5 e 10 serão recoloridos para preto, gerando uma nova árvore rubro negra como mostrado na Figura 14.

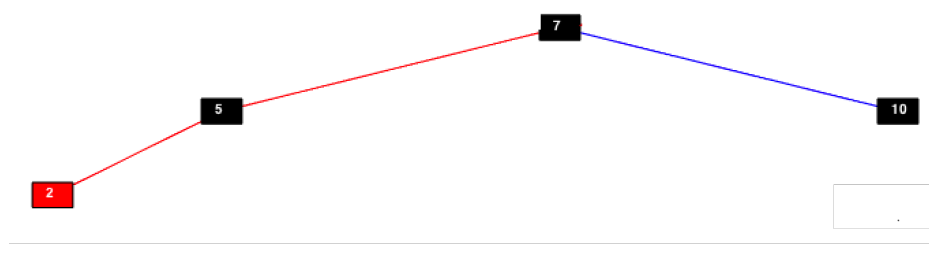


Figura 14 – Árvore rubro negra após a inserção do elemento 2

Inserindo elemento 6. O elemento 6 do novo nó é menor que 7, então ele vai para a esquerda deste. O elemento 6 é maior do que 5, então ele segue à sua direita. Uma vez que o pai do novo nó (5) é preto, simplesmente termine, uma vez que já existe uma nova árvore rubro negra como mostrado na Figura 15.

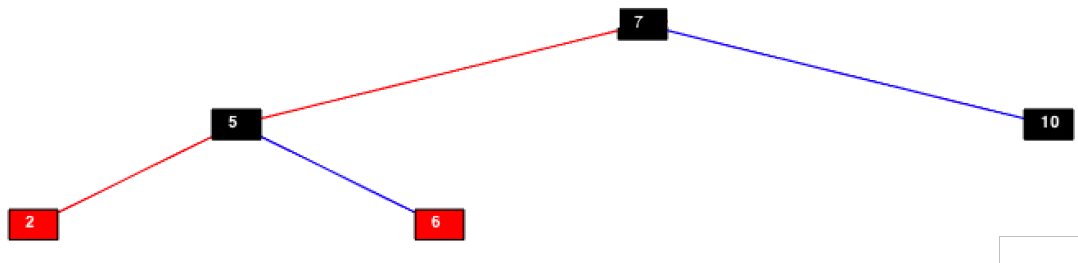


Figura 15 – Árvore rubro negra após a inserção do elemento 6

Inserindo elemento 1. O elemento 1 do novo nó é menor que 7, menor que 5 e menor que 2, então ele vai para a esquerda destes. Como o elemento 2 é vermelho, há um conflito vermelho-vermelho. Verificando o irmão do pai, que é 6, verifica-se que sua cor é vermelha. Em seguida, recolorir os nós 2 e 6 para preto. Agora, verifique o pai do novo nó (1), que é 5. Como não é o nó raiz, ele precisa ser recolorido para vermelho também. Verificar se após a recoloração existe algum conflito vermelho-vermelho. Verifica-se que não existe. A nova árvore rubro negra é mostrada na Figura 16.

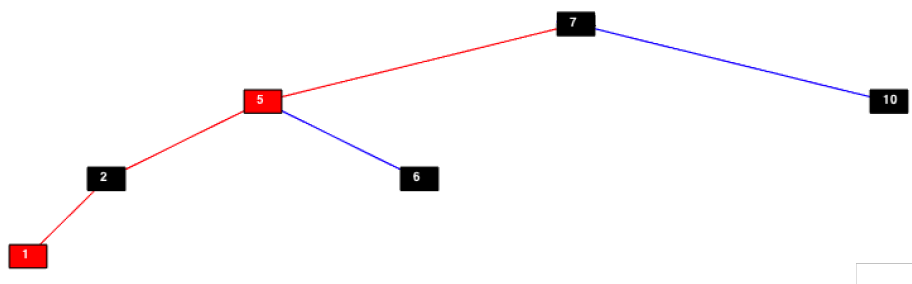


Figura 16 – Árvore rubro negra após a inserção do elemento 1

Inserindo elemento 3. O elemento 3 do novo nó é menor que 7, menor que 5 e maior que 2. Portanto, o nó 3 irá para a direita do elemento 2. Como o pai do novo nó é preto (2 é preto), nenhuma alteração deve ser feita e a nova árvore rubro negra pode ser mostrada na Figura 17.

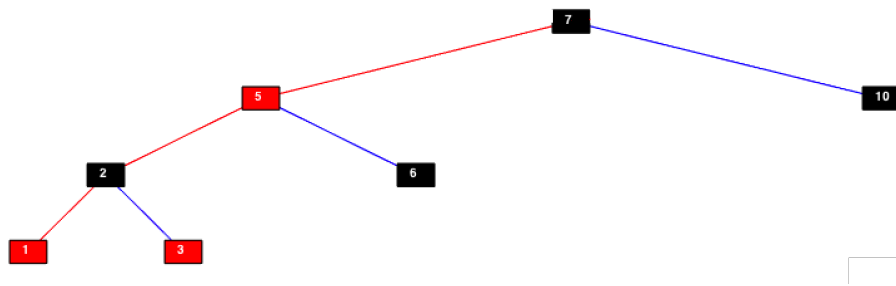


Figura 17 – Árvore rubro negra após a inserção do elemento 3

Inserindo elemento 4. O elemento 4 é um novo nó (portanto, será inserido inicialmente como vermelho) como o filho da direita do elemento 3. Isso causa um conflito vermelho-vermelho porque o novo nó (4) e seu pai (3) são vermelhos. Então, vamos verificar o tio do novo nó (1), que é vermelho. Neste caso, uma recoloração deve ser feita no pai (3) e no tio (1) para preto, sendo 1 = preto, 3 = preto e 4 = vermelho. Agora verifique se o pai do novo nó (2) é o nó raiz. Visto que 2 não é a raiz, devemos recolori-lo para vermelho.

Agora, considerando 2 como o novo nó, um novo conflito vermelho-vermelho é visto envolvendo 2 e seu pai 5. Então, vamos verificar o tio do novo nó, que é 6 e preto. Como o tio é preto, uma rotação e recoloração devem ser feitas.

Rotação significa pegar o trio 2-5-7 e fazer 5 o nó do meio, 2 à esquerda e 7 à direita. Agora recolorir o pai e o pai do pai 5 e 7, respectivamente, para preto e vermelho. O elemento 6 que estava à direita de 5 irá para a esquerda de 7. Agora, verifique se há um conflito vermelho-vermelho, o que não existe. A árvore rubro negra final é mostrada na Figura 18

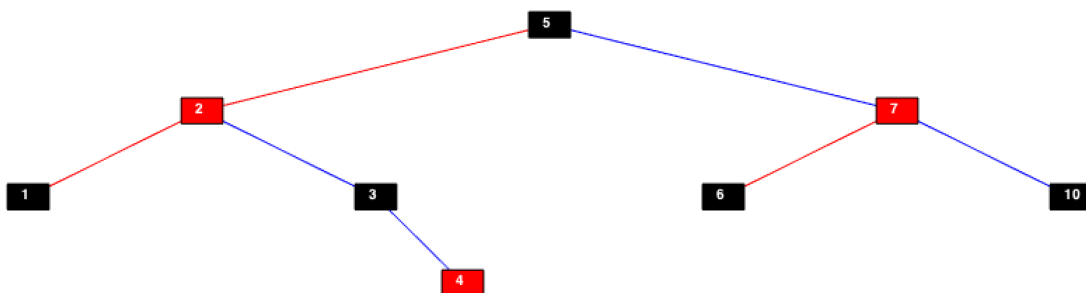


Figura 18 – Árvore rubro negra após a inserção do elemento 4

3.2.3 Desenvolvimento do modelo

Esta seção cobre a estrutura do modelo proposto e os detalhes de implementação. A Figura 19 mostra o diagrama do modelo seguido da descrição de cada fase.

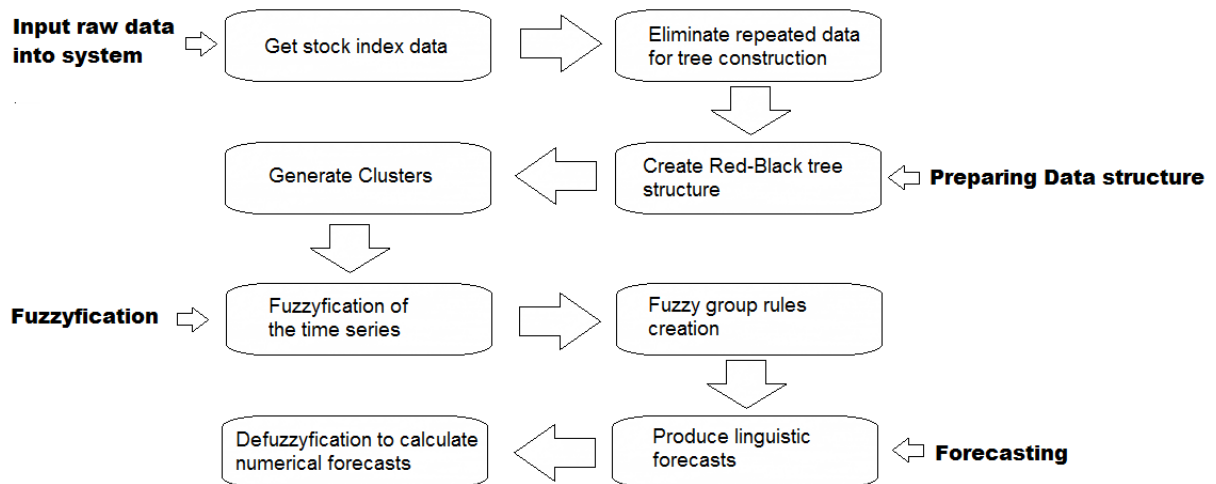


Figura 19 – Estrutura do modelo fuzzy proposto com base em árvores rubro-negras

Passo 1: Obter os dados de índices de ações

Nesta etapa, os conjuntos de dados do índice de preços das ações são reunidos. O conjunto de dados deve ser composto por uma série temporal do preço de fechamento de um índice ou preço de uma ação.

Nos exemplos cobertos neste trabalho, os valores do conjunto de dados foram retirados de dados históricos do site <http://www.investing.com>. Este é um site conhecido, amplamente utilizado por um grande número de investidores em todo o mundo devido à sua confiabilidade no fornecimento de informações sobre o mercado de ações.

O conjunto de dados de série temporal é alimentado no modelo na forma de matriz orientada usual, conforme mostrado na Equação 3.18

$$x(t) = \{x_1, x_2, \dots, x_n\} \quad (3.18)$$

Na Equação 3.18, t indexa a variável de tempo discreta (x_1 é o primeiro valor da série temporal e x_n é o último; n é o tamanho da série temporal de dados).

Passo 2: Eliminar dados repetidos na construção da árvore

Neste passo outro vetor é criado com base na série temporal. Ele é basicamente a mesma série temporal $x(t)$ com a remoção de elementos repetidos ($x_u(t)$). Este passo é crítico para o passo no qual a estrutura de dados árvore rubro-negra é gerada.

Passo 3: Criação da estrutura de dados árvore rubro-negra

Nesta etapa, a estrutura de dados da árvore rubro-negra é gerada a partir de $x_u(t)$. O modelo proposto requer que a estrutura de dados em árvore não tenha nós repetidos, daí a necessidade de usar $x_u(t)$ em vez de $x(t)$.

Como um exemplo, considere a seguinte série temporal.

$$x(t) = \{6, 7, 15, 7, 10, 2, 8, 20, 5\} \quad (3.19)$$

Após a construção da árvore rubro-negra utilizando os dados da Equação 3.19, nós temos a estrutura de árvore representada na Figura 20.

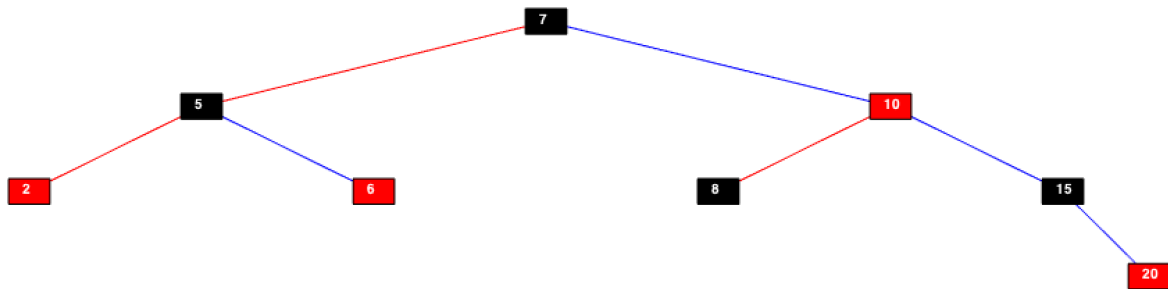


Figura 20 – Exemplo de árvore rubro negra

Passo 4: Geração de clusters

A geração dos clusters definirá os conjuntos fuzzy de forma que cada cluster se conecte a um conjunto fuzzy. Esta etapa recebe como entrada a estrutura de árvore rubro negra (RBT) gerada na etapa anterior. O algoritmo de geração de cluster é definido a seguir.

- Passo 1. Figura 20 mostra uma representação simbólica de uma árvore rubro negra. Inicialmente, é verificado se a árvore possui uma raiz, ou seja, o nó superior da estrutura de dados da árvore (também chamado de Troot). Em seguida, as seguintes etapas são realizadas:
 - Se Troot não possuir filhos, então ele deve pertencer a um novo cluster. Por exemplo, se o último cluster foi C_2 , ele deve pertencer a C_3 , e assim por diante. Se é o primeiro nó, ele deve pertencer a C_1 .
 - Se Troot possui apenas um filho, então faça o filho pertencer ao mesmo cluster que Troot.
 - Se ambos filhos existirem, então compute a diferença entre valores de Troot e seu filho da direita (Troot.RIGHT); e Troot e seu filho da esquerda (Troot.LEFT). Faça o filho com a menor diferença pertencer ao mesmo cluster de Troot. Faça o outro filho pertencer a um novo cluster.
- Passo 2. Para cada nível da árvore e para cada nó do respectivo nível, repita o passo 1, considerando cada nó como uma nova “Troot”.

- Finalize quando todos os nós tiverem sido designados a um cluster.

Se o algoritmo de clusterização for aplicado na Figura 20, os seguintes clusters (ou grupos) serão formados baseados em seus respectivos elementos:

- cluster 1: 7, 5, 6
- cluster 2: 8, 10
- cluster 3: 2
- cluster 4: 15, 20

Passo 5: Fuzzyficação da série temporal

O termo “fuzzyficação” se refere ao processo de transformação da série temporal composta por números reais (x_1, x_2, \dots, x_n) em uma série temporal fuzzy formada por conjuntos fuzzy (A_1, A_2, \dots, A_c) , onde c é o número de diferentes possíveis conjuntos fuzzy. O número de conjuntos fuzzy é o número de clusters na estrutura de árvore rubro negra do passo 4, ou seja, o resultado do processo de clusterização.

No exemplo considerado, $n = 9$ (note que o número de elementos em $x(t)$, e não em $x_u(t)$ deve ser considerado) e $c = 4$.

O processo de fuzzyficação é resumido no seguinte algoritmo.

- Para cada item pertencendo a x_t faça:
 - para cada cluster existente faça:
 - * se os dado da série temporal pertencer ao cluster c faça a série fuzzy igual a A_c
 - fim para
- fim para

Se o algoritmo for aplicado à serie temporal apresentada na Equação 3.19, este resultará na série temporal fuzzy representada na Tabela 11.

Série Temporal	Série Temporal Fuzzy
6	A_1
7	A_1
15	A_4
7	A_1
10	A_2
2	A_3
8	A_2
20	A_4
5	A_1

Tabela 11 – Exemplo de fuzzyficação. A série temporal fuzzy corresponde à segunda coluna da tabela e é referenciada como $A(t)$

Passo 6: Criação de regras de grupo fuzzy

Neste passo, regras de grupo fuzzy são geradas. Num primeiro momento, as regras de grupo fuzzy são identificadas e agrupadas em um momento posterior. No modelo, a seguinte regra de dependência fuzzy foi considerada.

se em uma série temporal fuzzy existe um valor fuzzy $A(t)$ precedendo outro valor $A(t + 1)$, pode-se escrever a Equação 3.20 para representar esta relação.

$$A(t) \rightarrow A(t + 1), \quad (3.20)$$

onde $A(t)$ é chamado de precedente da regra e $A(t + 1)$ é chamado de conseqüente da mesma regra.

Para cada valor de série temporal fuzzy, uma regra será criada definindo o valor da série temporal fuzzy do dia seguinte (ou próxima amostra). Exemplo: o primeiro valor da série temporal fuzzy é A_1 (valor da série temporal original era 6). Dizemos que este valor implicará em A_1 (que é a próxima amostra da série temporal fuzzy, correspondente ao valor 7 na série temporal original). A Tabela 12 mostra as regras fuzzy correspondentes à série temporal fuzzy definida na equação 3.19.

Série Temporal	Série Temporal (t+1)	Série Temporal Fuzzy	Série Temporal Fuzzy(t+1)	Regra Fuzzy
6	7	A_1	A_1	$A_1 \rightarrow A_1$
7	15	A_1	A_4	$A_1 \rightarrow A_4$
15	7	A_4	A_1	$A_4 \rightarrow A_1$
7	10	A_1	A_2	$A_1 \rightarrow A_2$
10	2	A_2	A_3	$A_2 \rightarrow A_3$
2	8	A_3	A_2	$A_3 \rightarrow A_2$
8	20	A_2	A_4	$A_2 \rightarrow A_4$
20	5	A_4	A_1	$A_4 \rightarrow A_1$
5		A_1		

Tabela 12 – Regras Fuzzy originadas da Tabela 11

A seguir, as regras fuzzy (quinta coluna da Tabela 12) são agrupadas. Se existir mais de uma regra com o mesmo precedente, as regras serão agrupadas em uma regra de grupo na qual o precedente é o mesmo, porém com múltiplos consequentes. Como exemplo, supondo que existam as seguintes regras fuzzy:

$$\begin{aligned}
 &A_1 \rightarrow A_1 \\
 &A_1 \rightarrow A_2 \\
 &A_1 \rightarrow A_7
 \end{aligned}
 \tag{3.21}$$

Essas regras podem ser agrupadas em uma única regra como mostrado na Equação 3.22.

$$A_1 \rightarrow A_1, A_2, A_7
 \tag{3.22}$$

Seguindo este exemplo, as seguintes regras de grupo fuzzy (chamadas FGRs - Fuzzy Group Rules) da Tabela 12 podem ser extraídas. A Tabela 13 mostra as FGRs derivadas da Tabela 12.

Regras de Grupo Fuzzy
$A_1 \rightarrow A_1, A_4, A_2$
$A_2 \rightarrow A_3, A_4$
$A_3 \rightarrow A_2$
$A_4 \rightarrow A_1, A_1$

Tabela 13 – Regras de Grupo Fuzzy.

O modelo proposto requer que as regras de grupo fuzzy considerem todos os consequentes, mesmo que repetidos.

Passo 7: Produção de previsões linguísticas

Uma variável linguística é uma variável fuzzy como A_1 . A mesma série temporal $x(t)$ usada no passo anterior será usada para previsão. O processo de previsão será explicado usando o mesmo exemplo de série temporal $x(t)$ representado na Equação 3.19 e que gerou a série temporal fuzzy da Equação 3.23 e as FGRs da Tabela 13. A série temporal é apresentada na Equação 3.23.

$$A(t) = A_1, A_1, A_4, A_1, A_2, A_3, A_2, A_4, A_1 \quad (3.23)$$

Para prever o dado em $t = 2$, considerando a série temporal fuzzy da equação $A(t)$, o valor da série temporal fuzzy no tempo $t = 1$ é considerado. O valor fuzzy em $t = 1$ (ou a primeira amostra) é A_1 .

A partir das FGRs na Tabela 13, é possível se encontrar a regra na qual A_1 é o precedente, ou seja, a primeira regra representada por $A_1 \rightarrow A_1, A_4, A_2$.

Finalmente, a previsão linguística da segunda amostra é o consequente da regra. Em outras palavras, a previsão para a segunda amostra é A_1, A_4, A_2 .

Passo 8: Defuzzyficação para cálculo de previsões numéricas

Começando a partir da previsão linguística do passo 7, o valor previsto fuzzy pode ser defuzzyficado em um valor numérico. Para explicar este procedimento, um exemplo é dado a seguir. Começando a partir da previsão da segunda amostra do passo 7, o valor pode ser representado como na Equação 3.24.

$$previsao_fuzzy(2) = \{A_1, A_4, A_2\} \quad (3.24)$$

Pesos crescentes são designados para cada valor da previsão na Equação 3.24. A Tabela 14 mostra os conjuntos previstos com os correspondentes pesos.

Conjuntos Previstos	Pesos
A_1	1
A_4	2
A_2	3

Tabela 14 – Exemplo de pesos fuzzy.

No próximo passo, para cada conjunto de previsão, é considerado o valor representado pela média simples do conjunto (ou cluster correspondente). Pro exemplo, para o conjunto A_1 , a média do cluster 1 é calculada. Uma vez que o cluster 1 é composto dos elementos $\{7, 5, 6\}$, a média é $(7 + 5 + 6)/3 = 6$. Usando este procedimento para todos os clusters, Centros de Cluster (ou médias) são criados como representado na Tabela 15.

Cluster	Centro de Cluster (Média)
1	6
2	9
3	2
4	17.5

Tabela 15 – Exemplo de Centros de Cluster.

A Tabela 16 resume os conjuntos de previsão, centros de cluster e pesos do exemplo (o objetivo é a previsão da segunda amostra), os quais são necessários para o cálculo da previsão numérica.

Conjuntos de Previsão	Cluster	Centros de Cluster (média)	Peso
A_1	1	6	1
A_4	4	17.5	2
A_2	2	9	3

Tabela 16 – Exemplo de Conjuntos de previsão, centros de cluster e pesos.

Finalmente, a previsão numérica é calculada como a média ponderada dos centros de cluster. Considerando o exemplo, a previsão numérica para a segunda amostra é calculada como:

$$forecast(2) = \frac{6 \times 1 + 17.5 \times 2 + 9 \times 3}{1 + 2 + 3} = 11.333 \quad (3.25)$$

Os passos de 1 a 6 consistem na fase de treinamento do modelo. A previsão então é feita considerando os passos 7 e 8 gerando, ao final, a previsão numérica.

3.2.4 Algoritmo do modelo baseado na árvore rubro-negra

Esta seção apresenta o algoritmo para o desenvolvimento computacional do segundo modelo apresentado. Posteriormente, o algoritmo foi implementado usando Matlab para o desenvolvimento do trabalho presente nesta tese.

Algoritmo:

- Variáveis de entrada:
 - Faça $y = [inputData];(série\ temporal)$
- Faça $uniqueY = y$ retirados elementos repetidos
- Faça $tree =$ árvore rubro-negra construída com $uniqueY$
- Defina os clusters a partir da árvore rubro negra seguindo o passo 4 da criação do modelo.
- Crie a série temporal fuzzy a partir dos clusters gerados.

- Crie um vetor com os centros (médias) dos clusters
- Crie uma matriz que possa encapsular as regras fuzzy a partir do conjunto $uniqueY$
- Faça Para cada $y(i)$ a ser predito:
 - Verifique o valor do conjunto fuzzy (cluster) imediatamente anterior ao $y(i)$ a ser predito e chame de aux
 - Verique a regra fuzzy que possui aux como antecedente
 - Para cada conseqüente da regra anterior armazene em um vetor os centros dos clusters dos conseqüentes
 - Atribua pesos crescentes para cada conseqüente do passo anterior
 - Calcule o valor predito como a média ponderada dos centros de clusters utilizando os pesos atribuídos no passo anterior

3.2.5 Exemplo de aplicação utilizando a árvore rubro-negra

Nesta seção, dois exemplos diferentes são apresentados para ilustrar o método proposto. Quatro modelos conhecidos da literatura existente (CHEN, 1996), (YU, 2005), (TEOH et al., 2008), (SINGH; DHIMAN; KAUR, 2018) usando uma base de dados específica e seus resultados de previsão serão comparados com o modelo proposto. A seguir, os mesmos 5 modelos (4 existente somados ao modelo proposto) serão executados a fim de se prever amostras fora do conjunto de treinamento.

Para fins de comparação, as seguintes medidas de erro serão utilizadas.

- RMSE - Root Mean Square Error
- RMSE comparado ao “modelo naive” (isto é, utilize o último valor conhecido como a previsão atual), também conhecido como RMSE normalizado.

A comparação entre os modelos será realizada utilizando *RMSE* e *RMSE comparado ao naive* de acordo com as Equações 3.16 e 3.17.

3.2.5.1 Teste com dados de treinamento

Para o teste com dados de treinamento, os valores de fechamento do índice Ibovespa (índice de ações brasileiro) de 29/03/2018 a 11/10/2018 foram utilizados. Os dados foram retirados do site *investing.com*. Parte dos dados de resultados de predição dos modelos podem ser encontrados na Tabela 17 (Datas representadas no formato americano).

Date	Ibovespa Dados Reais	Chen 1996 Previsão	Yu 2004 Previsão	Teoh 2008 Previsão	Singh 2018 Previsão	Modelo Proposto Previsão
03/29/2018	85365	-	-	-	-	-
04/02/2018	84666	84790	85148	84721	85141	84820
04/03/2018	84623	83484	84315	84721	84553	84820
04/04/2018	84360	83484	84315	84721	84553	84143
04/05/2018	85210	83484	84315	84721	84553	85167
04/06/2018	84820	84790	85165	84721	85298	84820
04/09/2018	83307	84790	85165	84721	84947	85322
04/10/2018	84510	83484	83481	83302	83247	83076
04/11/2018	85246	83484	84315	84721	84553	84143
04/12/2018	85444	84790	85165	84721	85298	84820
04/13/2018	84334	84790	85148	84721	85141	85214
04/16/2018	82862	83484	84315	84721	84970	85167
04/17/2018	84086	83484	83481	83302	83247	84820
04/18/2018	85776	83484	85648	84721	84970	85560
04/19/2018	85824	84790	85148	84721	85141	85214
04/20/2018	85550	84790	85115	84721	85141	85194
04/23/2018	85603	84790	85148	84721	85097	85214
04/24/2018	85469	84790	85148	84721	85097	85074
⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/05/2018	82322	83484	83481	83302	82998	83076
10/08/2018	86084	83484	83981	83302	84471	82798
10/09/2018	86088	84790	85115	84721	84858	84737
10/10/2018	83679	84790	85115	84721	84858	84737
10/11/2018	82921	83484	84481	83302	83859	83936

Tabela 17 – Dados do índice de fechamento diário Ibovespa usados para o teste de treinamento. Os valores previstos para vários modelos e o modelo proposto são apresentados para comparação.

A avaliação do desempenho do modelo proposto comparado com modelos existentes usando as Equações 3.16 and 3.17 é apresentada na Tabela 18. Percebe-se facilmente que os valores do modelo proposto diferem menos dos dados reais do Ibovespa do que os demais modelos. Isso indica um melhor desempenho de previsão. Os modelos com RMSE menor são os melhores em termos de previsão.

Modelo	RMSE	RMSE relativo Naive
Chen 1996	1235	1.07
Yu 2004	1090	0.946
Teoh 2008	1370	1.189
Singh 2018	1032	0.896
Modelo Proposto	1001	0.869

Tabela 18 – Comparação de desempenho entre o modelo proposto e modelos existentes sobre a base de treinamento Ibovespa.

A Figura 21 mostra o desempenho de previsão do modelo proposto comparado com o modelo naive.



Figura 21 – Desempenho do modelo proposto comparado com o modelo Naive sobre os dados do índice de fechamento Ibovespa.

A Figura 22 mostra um zoom sobre a Figura 21. A linha referente ao modelo proposto está mais próxima, em média, dos dados originais do que o modelo Naive. Isso mostra a melhor acurácia de previsão do modelo proposto em comparação com o Naive.

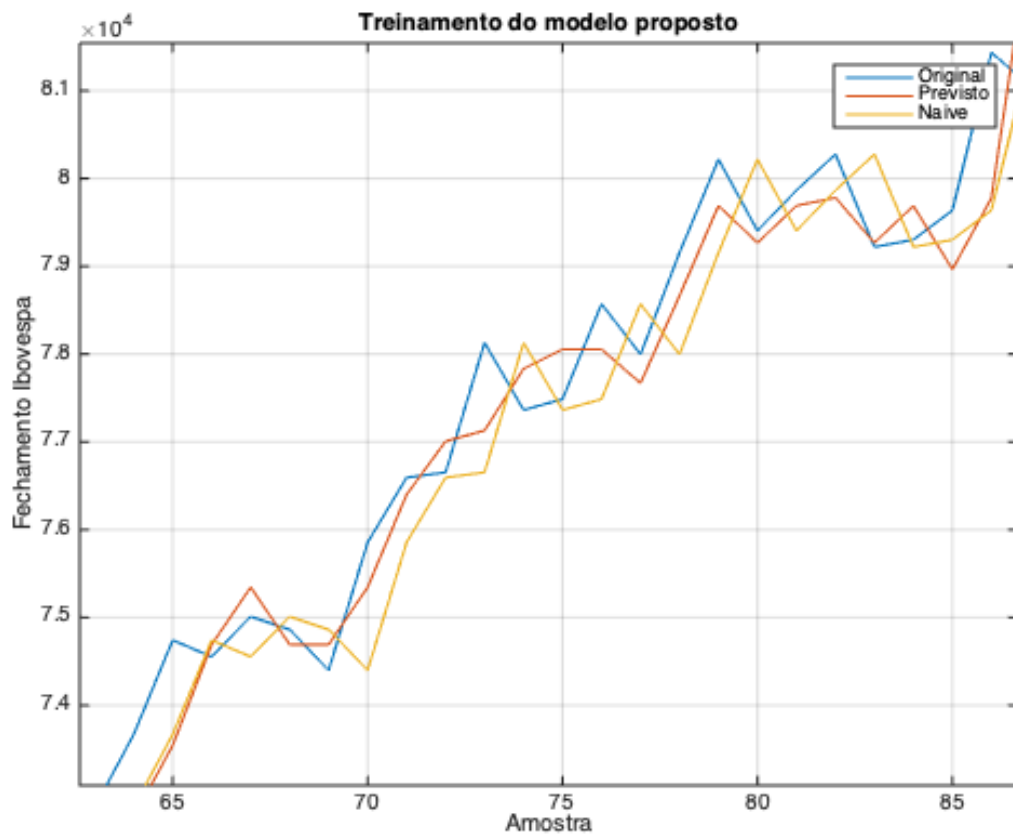


Figura 22 – Zoom sobre a figura do modelo proposto comparado ao modelo naive sobre os dados de fechamento do índice Ibovespa.

A Figura 23 mostra o desempenho do modelo proposto comparado aos 4 modelos existentes (CHEN, 1996), (YU, 2005), (TEOH et al., 2008) e (SINGH; DHIMAN; KAUR, 2018) em termos de previsão. A Figura 24 apresenta um zoom sobre a Figura 23.

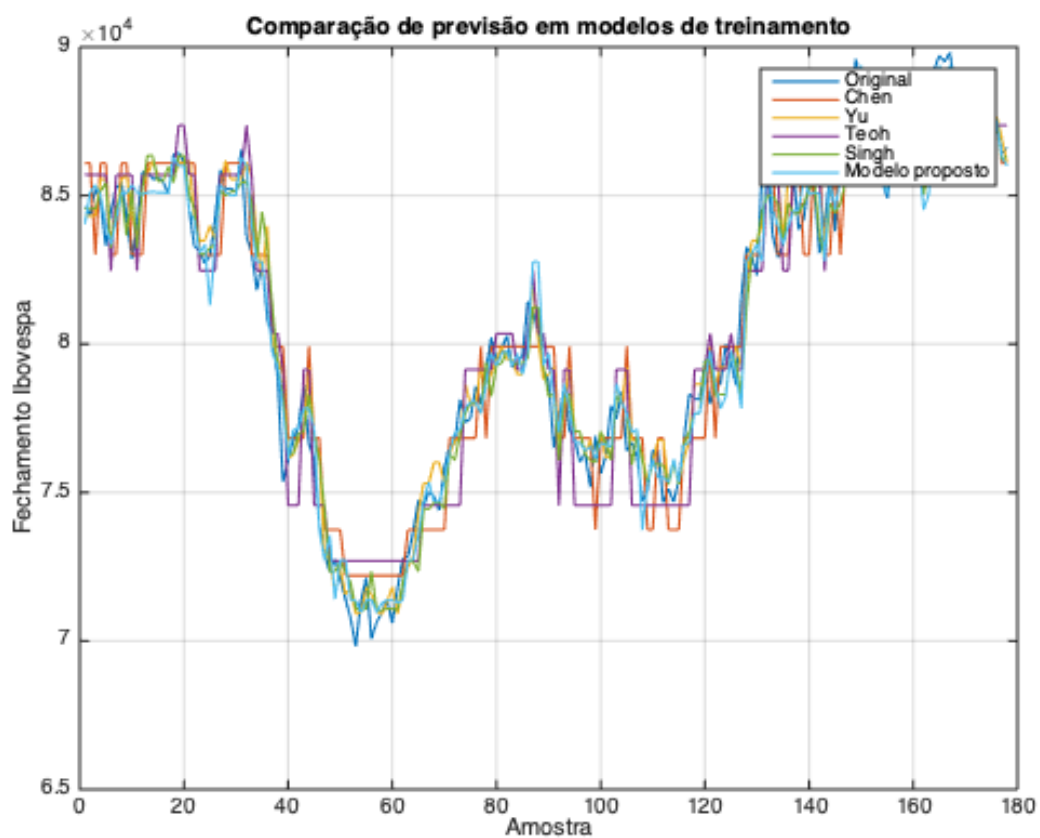


Figura 23 – Comparação de modelos de previsão

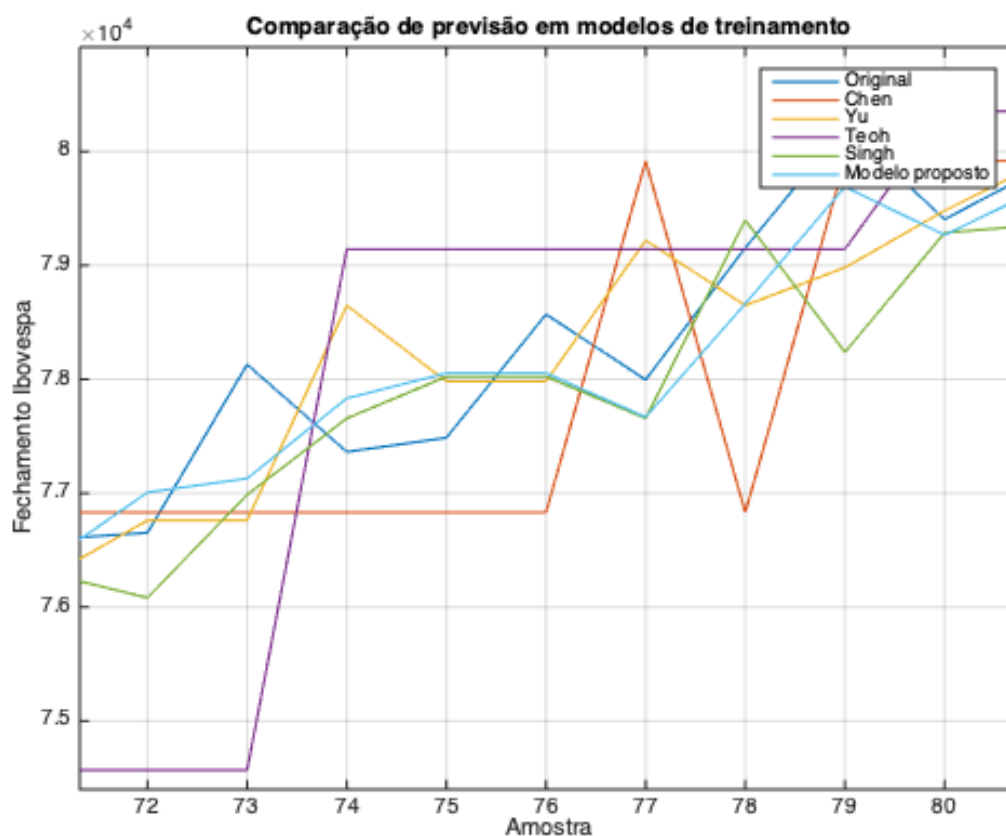


Figura 24 – Zoom sobre a Figura Comparação de modelos de previsão

Pode-se ver facilmente que o modelo proposto supera os modelos existentes em desempenho e acurácia de previsão sobre os dados de treinamento.

Na próxima subseção, a eficácia do método proposto é verificada por meio do teste fora da amostra dos dados de treinamento.

3.2.5.2 Teste fora dos dados de treinamento

O Teste fora dos dados de treinamento foi executado conforme descrito a seguir.

- 1. Execute o algoritmo de treinamento usando a base de dados Ibovespa de 29/03/2018 a 11/10/2018.
- 2. Preveja o preço da próxima amostra do conjunto de dados. Neste caso, preveja a cotação a partir do dia 15/10/2018. (Observe que estão sendo utilizados apenas dias úteis. 12/10/2018 foi feriado no Brasil e a bolsa foi fechada. Também foi uma sexta-feira então não houve negociação nos dias 13/10/2018 e 14/10/2018) .
- 3.Registre essa previsão como forecast(1).

Para calcular a previsão para o dia seguinte, ou seja, forecast(2) a partir do dia 16/10/2018, o conjunto de dados utilizado foi obtido deslocando a amostra de treinamento um passo à frente, utilizando os dados de 02/04/2018 (próximo dia útil) a 15/10/2018 (preço real, não previsto).

Todo o processo foi repetido 4 vezes fornecendo forecast(1) a forecast(4). Os resultados para todos os métodos podem ser vistos na Tabela 19. (Data fornecidas no formato americano)

Date	Ibovespa Dados Reais	Chen 1996 Previsão	Yu 2004 Previsão	Teoh 2008 Previsão	Singh 2018 Previsão	Modelo Proposto Previsão
10/15/2018	83359	83484	83448	83144	82789	82980
10/16/2018	85717	83484	82815	83121	82978	83918
10/17/2018	85763	84790	85315	84559	85389	85389
10/18/2018	83847	84790	85315	84575	85465	85627

Tabela 19 – Resultados de previsão sobre dados do mercado de ações.

A comparação de desempenho baseada entre os métodos usados neste caso é mostrada na Tabela 20.

Modelo	RMSE	RMSE relativo Naive
Chen 1996	1308	0.851
Yu 2004	1642	1.070
Teoh 2008	1480	0.964
Singh 2018	1627	1.060
Proposed model	1293	0.842

Tabela 20 – Comparação de desempenho entre o modelo proposto e os modelos existentes na base de testes do Ibovespa.

3.2.6 Discussão

Tanto o RMSE quanto o RMSE em relação ao Naive mostram que o método proposto supera os outros métodos. O método de Chen foi o mais próximo em termos de desempenho provavelmente pela parcela de áreas de baixa volatilidade no gráfico do Ibovespa.

Comparando os resultados e RMSEs de treinamento e teste, pode-se perceber que o único modelo que foi consistente quanto aos resultados dentro e fora da amostra de treinamento foi o modelo proposto. Além disso, foi também o melhor modelo em termos de desempenho considerando as métricas utilizadas na análise.

3.3 Modelo Fuzzy integrador

Nesta simulação, que trabalha com o mesmo tipo de modelo da Seção 3.1, serão utilizados os dados referentes ao contrato futuro do índice S&P500, um dos principais índices da economia

americana. O Standard & Poor's 500 (S&P 500) é um índice composto por quinhentos ativos, negociados nas bolsas NYSE e NASDAQ. O índice é considerado o principal indicador de valorização de ações no mercado norte-americano. O patrimônio estimado do S&P 500 está avaliado em cerca de \$20 trilhões, com pontuação média de 2.700.

A proporção de ações listadas no índice são avaliadas de acordo com seu valor de mercado. O próprio índice S&P 500 não é negociado no mercado. Em vez disso, seu contrato associado S&P 500 futuro é o instrumento de investimento negociado no mercado de ações. Neste trabalho, as simulações serão feitas usando o contrato futuro do S&P 500 com foco na previsão de seus valores futuros.

3.3.1 Objetivos da simulação

Embora a análise técnica e fundamentalista coexistam em finanças, há dificuldades em integrá-las para auxiliar nas decisões de investimento. Devido a esse fato, muitos pesquisadores as utilizam separadamente, uma seguida do outra. Este trabalho pretende integrar ambas as análises em um modelo matemático usando um novo modelo fuzzy de séries temporais. As simulações do modelo pretendem prever os valores futuros do S&P 500 futuro utilizando variáveis fundamentalistas e técnicas. Para verificar a acurácia do modelo, seu desempenho será medido usando a raiz do erro quadrático médio (RMSE) e comparado com outros modelos bem conhecidos na mesma área.

A seguinte seção cobre a preparação de dados para o modelo fuzzy.

3.3.2 Preparação de dados de entradas fundamentalistas

Esta subseção abordará a preparação das duas entradas fundamentalistas que foram utilizadas neste trabalho (mantendo a nomenclatura americana de origem): P/E (Price to Earnings) e P/B (Price to book). Para o P/E, a estratégia é a seguinte:

1. Recupere os dados publicados de P/E de [28]. Esses dados são trimestrais. Exemplo: recuperar dados de 30/06/2019.
2. Divida o preço de fechamento do S&P 500 do dia 30/06/2019 pelo P/E para obter o lucro desse dia.
3. Replique o valor do lucro de 30/06/2019 até um dia antes do próximo P/E publicado (se o próximo P/E for publicado em 30/09/2019, replicar o lucro até 29/09/2019).
4. Para cada dia de 30/06/2019 a 29/09/2019 calcule o novo P/E (a ser utilizado como entrada) como o preço de abertura do dia dividido pelo ganho daquele dia.

A forma de obter a variável Price to book é a mesma do P/E, exceto pelo fato de que a relação Price to book é publicada mensalmente e os ajustes de replicação precisam ser feitos mês a mês.

Com esta etapa concluída, haverá dois vetores de entrada (um para P/E e outro para P/B) compondo as duas primeiras colunas da matriz de entrada.

3.3.3 Preparação de dados de entradas técnicas

Esta subseção cobrirá a preparação/explicação das quatro entradas técnicas que foram usadas neste trabalho: Banda de Bollinger, retornos, volume e S&P 500 passados. Com esta etapa concluída, haverá mais quatro vetores de entrada (um para cada variável técnica) compondo as colunas 3 a 6 da matriz de entrada. Para as bandas de Bollinger, as construções podem ser feitas a partir da definição direta das bandas de Bollinger. As bandas de Bollinger são 3 médias móveis de 20 períodos separados por 2 desvios padrão que compõem a banda média, banda superior e banda inferior. Neste trabalho com base em um critério de desempenho de execução de programa, foi utilizada a banda superior, compondo a 3ª entrada da matriz de entrada. Para os retornos da variável, sua construção pode ser feita para o retorno no tempo t segundo a Equação 3.26.

$$r(t) = \frac{p(t) - p(t-1)}{p(t-1)} \quad (3.26)$$

onde $p(t)$ é o preço de fechamento dos futuros do S&P 500 no momento (dia) t . Os retornos de entrada compõem a 4ª entrada da matriz de entrada. As variáveis volume e passado S&P500 podem ser recuperadas diretamente dos valores históricos dos valores do contrato futuro do S&P 500. A variável volume reflete o valor negociado do contrato no dia t e compõe a 5ª entrada da matriz de entrada. A variável S&P 500 passado é o valor de S&P no tempo $t-1$ e compõe a 6ª entrada da matriz de entrada.

3.3.4 Preparação de dados para entrada do modelo fuzzy

O modelo fuzzy recebe a matriz de entrada e o vetor de saída. A ideia é criar um modelo fuzzy onde o preço de fechamento de S&P 500 ($S\&P\ 500(t) = p(t)$) seja uma função das variáveis da matriz de entrada de acordo com a Equação 3.27.

$$S\&P500(t) = f(p/e(t), p/b(t), UpperBollingerBand(t-1), r(t-1), volume(t-1), S\&P500(t-1)) \quad (3.27)$$

A forma de divisão dos dados em clusters segue a seguinte lógica:

- Os limites dos clusters são definidos com base nos quantis do vetor de dados de saída. Cada cluster tem um limite inferior e um limite superior. Por exemplo: suponha que o número de clusters seja 3. Defina o limite inferior do cluster 1 como o menor valor dos dados de saída y . O limite superior do cluster 1 será o quantil $1/3$ de y , que também será o limite inferior do cluster 2. O limite superior do cluster 2 será o quantil $2/3$ de y , que também será o limite inferior do cluster 3. O limite superior do cluster 3 será o quantil $3/3$ de y . A Equação 3.28 mostra como calcular os limites de clusters para qualquer número de clusters, onde $ll(c)$ significa limite inferior do cluster " c " e $ul(c)$ significa limite superior do cluster " c ". $c=1\dots n$ onde n é o número de clusters.

$$\begin{aligned} ll(1) &= y(\min) \\ ul(c) &= ll(c + 1) = \text{quantile}(c/n) \text{ of } y \end{aligned} \quad (3.28)$$

- Depois disso, todos os dados em y serão divididos em clusters. Agora, os dados de entrada X , com os valores correspondentes x_{ij} , devem ser atribuídos ao mesmo cluster de sua saída correspondente y_j . Neste momento, tanto a saída quanto as entradas estão divididas em clusters.

3.3.5 Resultados Experimentais

3.3.5.1 Base de dados utilizada

Para este trabalho foi utilizado o contrato futuro S&P 500 de 07/05/2019 a 29/06/2021.

Ao construir a matriz de entrada, para cada y_j no vetor de saída haverá uma coluna de entrada (que é uma coluna da matriz de entrada). A correspondência entre entrada e saída foi mostrada na Equação 3.27. É importante notar que $y(t) = S\&P500(t)$ depende de algumas variáveis no tempo t e outras no tempo $t - 1$ (para previsão 1 passo à frente). As variáveis de que depende no momento t são aquelas que utiliza o preço de abertura do dia t . Isso é possível porque a intenção é prever o contrato do S&P 500 após o início do dia desde o preço de abertura.

3.3.5.2 Treinando o modelo

A quantidade total de dados é composta por 503 amostras do contrato futuro S&P 500 com datas de 07/05/2019 a 29/06/2021. Para o treinamento do modelo, foram utilizadas 480 amostras cada vez que o modelo previa um passo à frente. O número de clusters (grupos) escolhidos para esta simulação foi 2.

A Figura 25 mostra os resultados da previsão sobre os dados de treinamento comparando-os com modelos já conhecidos na literatura.

A partir da Figura 25 pode-se ver facilmente quão bem o modelo prevê a variável de interesse. Obviamente isso era esperado quando a simulação foi feita sobre os dados de treinamento. Na próxima seção será mostrado o teste fora da amostra de treinamento.

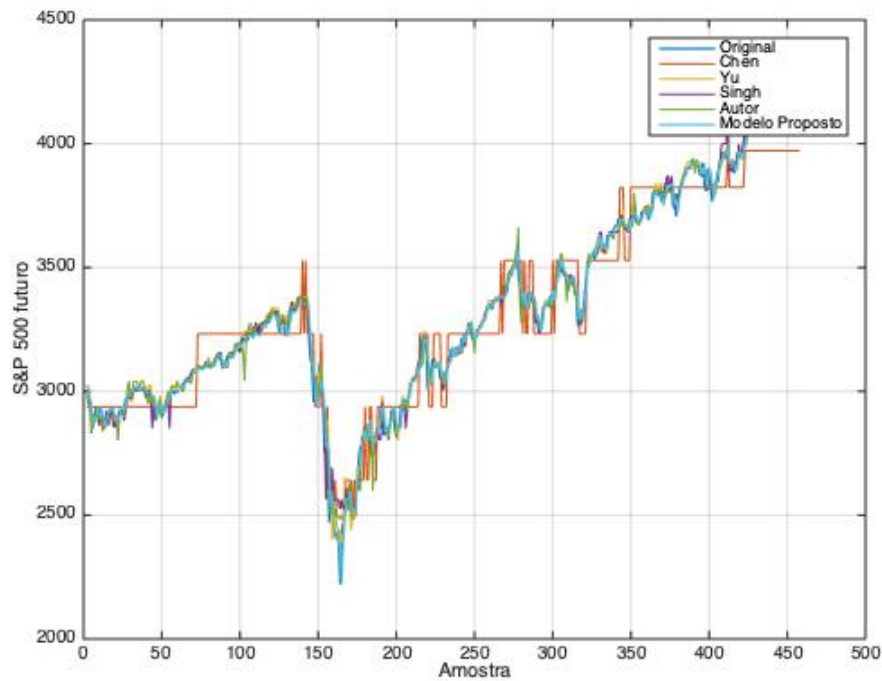


Figura 25 – Comparação no desempenho de previsão dos modelos sobre a base de treinamento

3.3.5.3 Testando o modelo

Para testar o modelo, as amostras testadas foram amostras de índices 481 a 503, totalizando 23 amostras. Nos testes, foi utilizada a técnica de janela rolante. Primeiro, as amostras 1 a 480 foram usadas para prever a amostra 481. Depois, as amostras 2 a 481 foram usadas para prever a amostra 482 e assim por diante. A Tabela 21 mostra os resultados de modelos conhecidos na área fuzzy, bem como o modelo proposto. A Figura 26 mostra os mesmos resultados da Tabela 21 em formato gráfico, incluindo o modelo naive.

Sample	Original S&P 500 futures	Chen, 1996	Yu, 2004	Singh, 2018	Tavares, 2020	Proposed Model
1	4193	3972	4184	4169	4147	4190
2	4199	3972	4184	4197	4185	4197
3	4203	3972	4184	4196	4183	4201
4	4199	3972	4174	4201	4188	4206
5	4206	3972	4184	4200	4208	4204
6	4191	3972	4174	4190	4187	4210
7	4228	3972	4184	4202	4202	4197
8	4226	4123	4174	4191	4188	4230
9	4226	4123	4187	4201	4199	4230
10	4219	4123	4187	4207	4205	4231
11	4238	4123	4187	4193	4192	4222
12	4246	4132	4207	4215	4226	4242
13	4255	4139	4207	4242	4231	4247
14	4247	4147	4241	4246	4236	4256
15	4223	4147	4241	4246	4238	4248
16	4222	4147	4221	4220	4217	4226
17	4187	4147	4221	4219	4217	4222
18	4214	4147	4199	4210	4221	4192
19	4236	4147	4197	4212	4191	4213
20	4232	4147	4221	4249	4226	4239
21	4256	4147	4221	4242	4236	4235
22	4271	4149	4234	4239	4246	4258
23	4281	4163	4247	4245	4253	4274

Tabela 21 – Resultados de teste comparados a outros modelos existentes

Usando os dados da Tabela 21, é possível calcular o RMSE para cada modelo bem como o RMSE normalizado, comparando cada modelo com o modelo naive.

A Tabela 22 apresenta resultados que mostram que, exceto pelo modelo Chen, todos os modelos tiveram bom desempenho na previsão. O modelo proposto superou todos os modelos como pode ser visto na Tabela 22.

Model	Chen, 1996	Yu, 2004	Singh, 2018	Anônimo, 2007	Modelo Proposto
RMSE	176	36	34	33	16
Norm. RMSE	6.4	1.3	1.2	1.2	0.93

Tabela 22 – Comparação de RMSE entre modelos

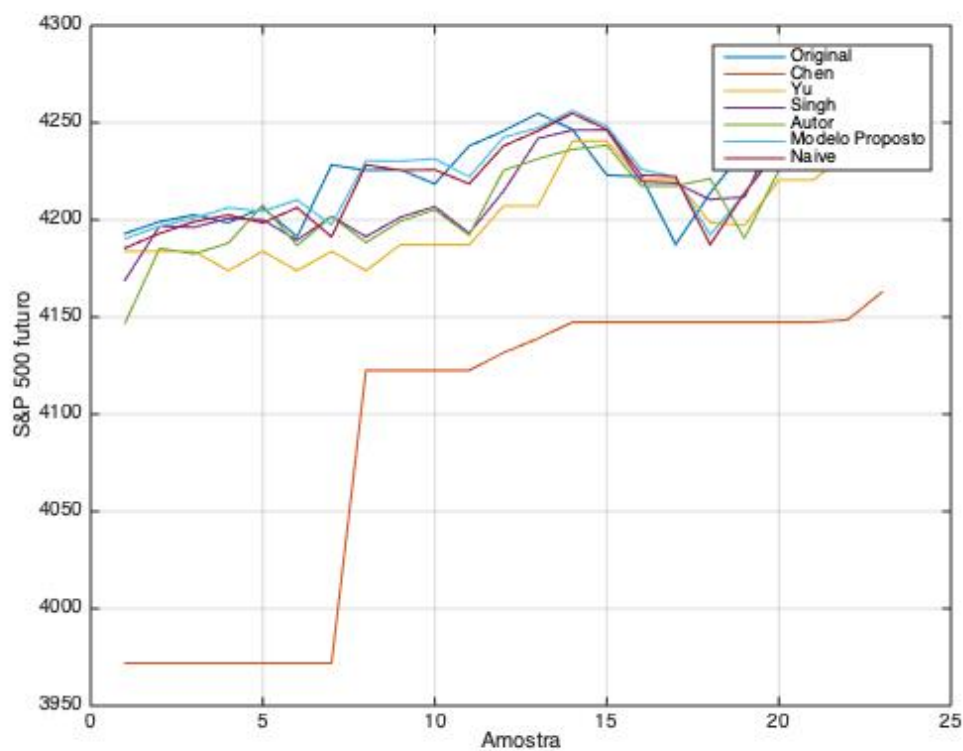


Figura 26 – Teste de comparação de previsão dos modelos

A Figura 27 mostra o desempenho do modelo proposto comparado apenas com o modelo naive.

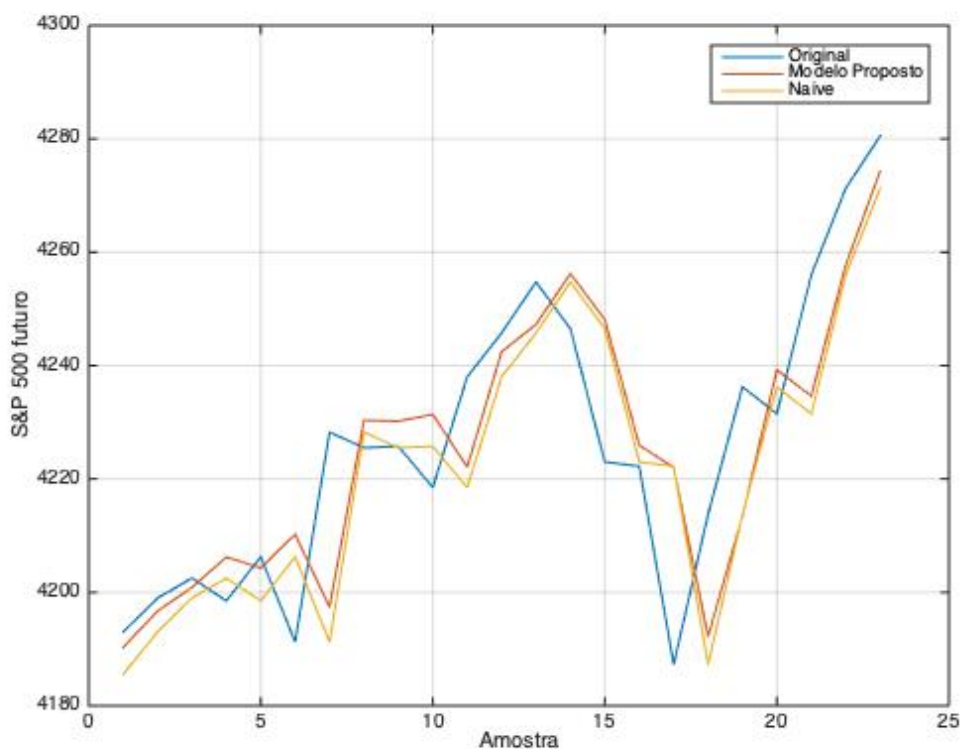


Figura 27 – Comparação de previsão de teste de modelo apenas com Naive

3.3.6 Discussão

A ideia inicial deste modelo foi integrar a análise técnica e fundamentalista do mercado financeiro em um único modelo. Isso foi possível devido ao aspecto subjetivo das séries temporais fuzzy.

Um dos aspectos que mais influenciou o desempenho do modelo foi a forma como o universo do discurso foi particionado. Foram utilizados quantis para dividir os dados ao invés de valores fixos de comprimento dos clusters. Isso permitiu que o modelo capturasse variáveis estatísticas dentro de cada cluster, tornando-as mais separáveis e, portanto, mais capazes de fazer previsões com maior acurácia.

O modelo foi comparado com outros modelos e superou todos eles. A comparação foi feita utilizando o RMSE e o RMSE normalizado (comparando os modelos com o naive).

Trabalhos futuros envolvem um estudo mais aprofundado sobre o universo da divisão do discurso usando diferentes estruturas como árvores binárias e florestas aleatórias.

A forma como as variáveis fundamentalistas são criadas também impõem grande importância no desempenho do modelo.

4 Desenvolvimento de modelos e resultados - Séries Espaciais

O capítulo apresenta o desenvolvimento dos modelos e os resultados obtidos para séries espaciais. Inicialmente, são apresentados conceitos relacionados à área de geotecnia. Em seguida, são apresentadas as etapas para a construção de modelos aplicáveis a séries espaciais. Por fim, são apresentadas simulações e comparações de resultados de predição com outros modelos existentes em áreas correlatas na literatura. Os modelos apresentados neste capítulo seguem a mesma lógica de criação da Seção 3.1.

4.1 Modelo fuzzy para predição de capacidade de carga em ancoragens

A seção apresentará conceitos básicos sobre geotecnia e resultados experimentais de um modelo aplicado à predição de capacidade de carga em ancoragens.

4.1.1 Geotecnia - Revisão bibliográfica

O uso de alguma técnica de contenção para garantir a estabilização de aterros e escavações apresenta-se como uma alternativa técnico-econômica viável e em expansão em todo o mundo, especialmente nos grandes centros urbanos, onde a possibilidade de retirada é quase impossível (VASCONCELOS, 2016), (CARVALHO; GOMES; PORTO, 2016), (PORTO; GOMES; MARTINI B.D.AND ARAUJO, 2016). Entre os tipos de contenções existentes, a solução de cortina de concreto armado aliada ao uso de âncoras rejeitos e pré-estressadas é a solução mais adequada, se não a única (grandes alturas e áreas restritas (PORTO; FERREIRA; SILVA, 2017), (PORTO; TORRES; GOMES, 2017), (VASCONCELOS; PEREIRA; PORTO, 2018a), (VASCONCELOS; PEREIRA; PORTO, 2018b)).

O interesse dos profissionais geotécnicos na técnica de paredes em queda se justifica por sua praticidade, velocidade de execução, versatilidade e, sobretudo, segurança (garante pequenas deformações), quando comparadas a outras técnicas utilizadas para estabilizar cortes em encostas naturais. Apesar do amplo uso dessa técnica no Brasil (desde a década de 1950), pouca pesquisa tem sido feita sobre o método de execução (SANTOS et al., 2019), (OLIVEIRA et al., 2019), (DIAS, 2019), (GONTIJO, 2020), (PEREIRA; PORTO, 2020), (DIAS; GOMES; PORTO, 2021). Nem mesmo uma quantificação dos fatores que afetam o comportamento das âncoras rejeitáveis, tais como: tipo de solo, metodologia de perfuração executiva, diâmetro perfurado, procedimento de injeção, proteção e fluência foram realizados com profundidade

acadêmica/conceitual substancial. Assim, um estudo científico dessa técnica, que ajuda sua compreensão e desenvolvimento, é oportuno e necessário (HUANG; ZHOU; YUE, 2012), (SAHAJDA, 2014), (ZHANG et al., 2015), (CARVALHO et al., 2018), (AZIZ et al., 2018), (JADID; SHAHRIAR; RAHMAN, 2019), (LANDE et al., 2020), (CAPUTO et al., 2021).

Ao analisar a literatura luso-brasileira, como publicações feitas por (FERRARI, 1980), (SOUZA, 2001), (CARVALHO, 2009) e (DIAS; GOMES; PORTO, 2021), nota-se que o conhecimento e o aprimoramento da técnica de cortinas jogadas no solo com o uso de âncoras rejeitadas e as pré-estressadas vêm principalmente da execução e monitoramento das obras realizadas, ou seja, da experiência dos contratantes. Nesse contexto, alguns conceitos e metodologias de execução estão sendo aprimorados, mas sem uma consistência teórica que permita avanços técnicos de forma racional. (CHEN; SAYDAM; HAGAN, 2018) e (FABRIS et al., 2021).

A estabilidade de uma contenção lançada é estudada em seu estado limite final (ELU). Assim, um parâmetro muito importante responsável pelo mecanismo de transferência de carga e restrição de circulação da massa do solo, durante e após sua escavação, é, segundo (SOUZA, 2001), *a capacidade de carga específica (t_u) que é dada em termos de resistência por unidade de comprimento. Dividindo-o pelo perímetro médio do bulbo tem-se a resistência ao cisalhamento (q_s) desenvolvida na interface entre a seção ancorada do tirante e o solo adjacente. Em termos práticos, quanto maior este parâmetro, melhor será o desempenho da ancoragem para ajudar a estabilizar o maciço de solo.*

A quantificação dos parâmetros t_u e q_s pode ser obtida, aproximadamente, a partir de formulações teóricas e empíricas, ou mesmo de forma real, a partir de testes de recepção, qualificação e/ou elementos básicos estabelecidos por (ABNT, 2018), devidamente extrapolado. Como comentado de forma simplificada anteriormente, os fatores que podem influenciar o desempenho de ancoragem são: tipo de solo, diâmetro do orifício, fluido de perfuração, características do rejunte de cimento, uso de aditivos no rejunte de cimento, fator água/cimento, tempo de execução, metodologia de injeção e volume injetado.

Esta pesquisa abrange um amplo programa experimental que vem sendo desenvolvido no campo desde o final da década de 1970 em São Paulo, na construção do Metrô. Esses testes foram realizados na tentativa de contribuir para o estabelecimento da melhor forma de realização dos vínculos, associados às condições geotécnicas locais. O banco de dados disponibilizado nesta pesquisa vem da realização de testes básicos em diversos trabalhos localizados no estado de São Paulo, com diferentes metodologias executivas.

Por essas razões, esta pesquisa constitui uma excelente oportunidade para catalogar a experiência disponível e aumentar o nível de conhecimento sobre a técnica de paredes caindo com âncoras rejeitáveis e pré-estressadas, permitindo assim disseminá-la de forma mais robusta no âmbito da engenharia geotécnica nacional. Os resultados desta pesquisa encontrarão aplicação prática imediata, pois tem relação direta com o desempenho e o custo da técnica de fixação de parede de concreto no Brasil.

Por todas as razões expostas, é óbvia a importância de tentar criar modelos para prever a capacidade de carga da âncora. Existem diversas áreas de conhecimento com foco na previsão de valores em uma determinada função ((CHENG et al., 2008), (ADOKO; WU, 2011), (SAMUI, 2020), (MOAYEDI; MOSALLANEZHAD; RASHID, 2020) e (EBID, 2021)). Atualmente, existem diversas metodologias e ferramentas para auxiliar na previsão da capacidade de carga da âncora, notadamente: Regressão logística, razão de frequência, pesos de evidência, rede neural artificial, processo de hierarquia analítica, índice estatístico, valor da informação, índice de entropia, Regressão vetorial de suporte, fator de certeza, Dempster-Shafer, mapeamento do índice de estabilidade, árvore de decisão, floresta aleatória, função de crença probatória, avaliação de múltiplos critérios espaciais, combinação linear ponderada, análise discriminante, modelo aditivo generalizado, modelo de splines de regressão adaptativa multivariada, avaliação de vários critérios, conjuntos ásperos, Análise de cluster, Naïve Bayes, Análise estatística bivariada, Classificação e árvore de regressão, detecção automática de interação qui-quadrado, análise discriminante linear de Fisher, análises de decisão multi-critérios, Apoiar a regressão vetorial, processo de rede analítica, associação de mineração de regras.

A maioria desses modelos são Estatísticos ou Heurísticos. Alguns são deterministas, porém seu uso é restrito. A área de conhecimento difusa (fuzzy) tem sido utilizada em diversas áreas de campo, como economia, finanças e engenharia, devido à sua instalação, que descrevem variáveis subjetivas, como sentimentos e intenções humanas (EBID, 2021).

(CHEN; TANUWIJAYA, 2011) mostra um método de previsão fuzzy multivariada baseado em uma série de tempo fuzzy e técnicas de clustering. Sua principal contribuição reside no desenvolvimento de um novo método de agrupamento dos dados, a fim de gerar diferentes intervalos de comprimento para os conjuntos fuzzy, em vez de colocá-los em intervalos de comprimento estático. Em (CHEN; CHANG, 2010) é apresentado um método para previsão fuzzy multivariável com base em clusters e técnicas de interpolação de regras fuzzy. As taxas de variação nos conjuntos de dados de treinamento são usadas para criar amostras de treinamento que são usadas para criar regras fuzzy (cada uma correspondente a um cluster) fazendo uso do algoritmo de clusterização C-means (CHENG et al., 2008).

(JR.; JR., 2016) apresenta um trabalho cuja principal intenção é propor uma nova metodologia para previsão difusa em séries temporais. O método usa a clássica função de associação triangular junto com o α -cut e aplica a função de autocorrelação estendida.

Nas áreas como economia e finanças, onde modelos não estacionários tendem a aparecer, os modelos GARCH (heterocedasticidade condicional autoregressiva generalizada) apresentam-se como modelos mais adequados para a previsão. Uma fusão significativa entre modelos Garch e técnicas difusas tem aparecido nos últimos anos. (TAKIYAR; UPADHYAY; SINGH, 2016) apresenta uma série de tempo de carga modelada usando modelagem fuzzy e transformação de wavelet. Em (HANAPI et al., 2020) uma técnica de regressão linear difusa foi combinada com um modelo GARCH para estimar os parâmetros do modelo e uma variância de janela deslizante

difusa foi desenvolvida para fazer a estimativa do peso da previsão. (ALAMEER et al., 2019) propõe um modelo que misture um sistema de inferência neuro-fuzzy adaptável (ANFIS) e um algoritmo genético (GA) para verificar sua previsão de melhorias nas redes em relação a modelos conhecidos como ARIMA (média móvel integrada autoregressiva) e GARCH.

Considerando todo o trabalho que foi publicado recentemente na academia, esta pesquisa propõe um novo modelo, incluindo a lógica fuzzy com a intenção de prever a capacidade de carga. A principal contribuição aqui é fornecer um novo método para clusterizar os dados com a intenção de ter clusters melhores distribuídos. Essa distribuição permite uma melhora no erro de previsão em relação aos modelos clássicos existentes na literatura. A principal motivação na redação deste trabalho foi alcançar melhores resultados em relação aos estudos anteriores, como deve ser visto.

4.1.2 Conceitos de Geotecnia

Para definir a capacidade de carga geotécnica das ancoragens, duas metodologias podem ser utilizadas:

- Formulações Semi-Empíricas, tais como (NUNES, 1987), (BUSTAMANTE; DOIX, 1985), (ABNT, 2006), (FALCONI, 2005), (JR; MALLMANN; IORIO, 2004) e (PORTO, 2015).
- Formulações matemáticas, tais como, por exemplo, (MASSAD, 1986) and (VEEN, 1953).

Para a definição da carga final de teste, pode-se considerar a proposta (VEEN, 1953) de Van der Veen. Este autor associa a curva carga x deslocamento usando a equação 4.1:

$$F = F_R(1 - e^{-ad}), \quad (4.1)$$

onde:

- F = carga aplicada no topo do tirante (kN);
- F_R = última carga correspondente à assíntota da curva vertical (kN);
- e = base dos logaritmos naturais;
- a = coeficiente que define o formato da curva (mm⁻¹);
- d = deslocamento correspondente à carga F (mm).

Como a função anterior tem duas incógnitas (F_R and a), sua solução é obtida por meio de tentativas. Assim, a expressão de Van der Veen pode ser reescrita pela Equação 4.2.

$$1 - \frac{F}{F_R} = e^{-ad} \quad (4.2)$$

ou, de outra forma, como na Equação 4.3.

$$ad = -\ln\left(1 - \frac{F}{F_R}\right) \quad (4.3)$$

O gráfico que mais se aproxima de uma linha indicará o valor F_R (desconhecido do problema). A inclinação desta linha é a própria variável a (a incógnita do problema).

A equação foi reescrita por (AOKI, 1976), resultando na Equação 4.4:

$$F = F_R(1 - e^{-ad+b}), \quad (4.4)$$

onde:

- b = ponto de intersecção com o eixo das ordenadas da reta obtida na escala semilogarítmica.

Com este artifício, (AOKI, 1976) pretendeu obter um valor de r^2 mais próximo de 1, possibilitando assim um melhor ajuste da curva carga x deslocamento, com os pontos intermediário e final de carregamento.

De acordo com (CINTRA et al., 2013), o método Van der Veen foi originalmente usado manualmente. Com o avanço dos programas de computador, foi possível utilizar o método dos mínimos quadrados para regressão linear em cada tentativa de valor de RF.

Embora o método seja muito prático, ele deve ser usado com cautela. A interrupção prematura do ensaio pode resultar em carga limitada x curvas de deslocamento para interpretação. (MAGALHÃES, 2005) apresenta três alternativas de curva carga x deslocamento, nas quais:

- Figure 28 (a): Teste interrompido no alongamento elástico: muito difícil determinar a carga final;
- Figure 28 (b): Teste interrompido no início da plastificação do sistema solo-bulbo: é possível prever a carga final por meio de extrapolação matemática;
- Figure 28 (c): Teste com grandes deformações para pequenos incrementos de carga: a carga final é evidente, não exigindo métodos de extrapolação.

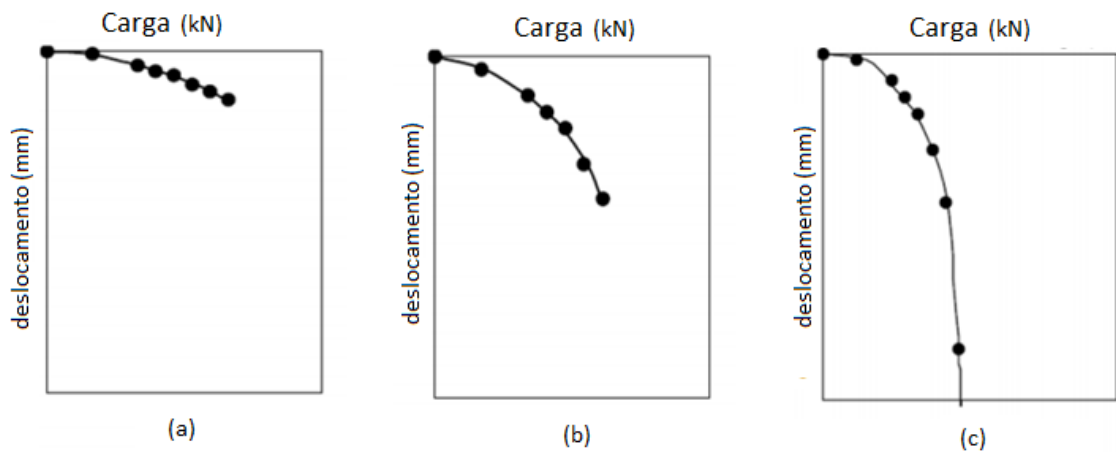


Figura 28 – Alternativas de curvas de carga x deslocamento

4.1.3 Resultados experimentais

Para validar o modelo proposto neste trabalho e também testar seu desempenho será utilizado um banco de dados experimental para diferentes tipos de solos com algumas variáveis específicas de cada amostra juntamente com a respectiva capacidade de carga. A intenção é criar e testar um modelo para prever a capacidade de carga de um solo com base em alguns dados específicos facilmente medidos.

As previsões serão comparadas às capacidades de carga experimentais medidas em campo para testar a acurácia do modelo.

4.1.3.1 Base de dados utilizada

A Tabela 23 mostra os dados de diferentes amostras de solos com algumas entradas medidas e a saída medida (*capacidade de carga*).

Índice da Amostra	Entrada 1 Tipo de Solo	Entrada 2 SPT	Entrada 3 Diâmetro Perfurado	Entrada 4 Lb	Entrada 5 k	Entrada 6 β	Saída Capacidade de Carga
1	areia siltosa	5	0,114	10	2,29880091	2,2	900
2	areia siltosa	5	0,114	11	2,271542402	2,2	1100
3	areia siltosa	5	0,114	12	1,982299336	2,2	1020
4	areia siltosa	5	0,114	7	1,950411207	2,2	525
5	areia siltosa	5	0,114	7	2,684314107	2,2	840
6	argila siltosa	11	0,114	9	1,18867702	2,1	630
7	argila siltosa	11	0,114	5	1,220587813	2,1	450
8	argila siltosa	11	0,114	5	1,328286737	2,1	500
9	argila arenosa	38	0,114	8	1,191790268	2,1	960
10	argila arenosa	38	0,114	8	1,484630163	2,1	1000
11	argila arenosa	38	0,114	7	1,260865459	2,1	840
12	areia argilosa	10	0,114	5	2,349543673	2,2	550
13	areia argilosa	10	0,114	5	3,508323851	2,2	800
14	areia argilosa	7	0,117	8	3,180597362	2,2	840
15	areia argilosa	7	0,117	8	2,422941285	2,2	840
16	areia argilosa	7	0,117	7	3,123001032	2,2	840
17	areia argilosa	7	0,117	8	3,200074382	2,2	920
18	areia argilosa	7	0,117	5	1,925887735	2,2	475
19	areia argilosa	7	0,117	5	2,923111158	2,2	575
20	areia argilosa	7	0,117	9	2,688694402	2,2	990
21	areia argilosa	7	0,117	9	2,977819943	2,2	1035
22	areia argilosa	7	0,117	10	2,502407527	2,2	1000
23	areia argilosa	7	0,117	7	1,558161598	2,2	560
24	areia argilosa	7	0,117	7	1,796337728	2,2	665
25	areia argilosa	7	0,117	9	2,94319413	2,2	990
26	areia argilosa	7	0,117	9	3,355241308	2,2	1035
27	areia argilosa	7	0,117	5,5	3,572439591	2,2	632,5
28	areia argilosa	13	0,114	7	1,239353534	2,2	665
29	areia argilosa	13	0,114	6,5	1,376205227	2,2	747,5
30	areia argilosa	13	0,114	6,5	1,319311827	2,2	617,5
31	areia argilosa	13	0,114	5	1,559186704	2,2	450
32	areia argilosa	13	0,114	5	1,04945259	2,2	450
33	silte arenoso	24	0,1	10	1,651817498	2,25	1900
34	silte arenoso	24	0,1	9,5	1,813809445	2,25	2090
35	silte arenoso	24	0,1	10	1,762070704	2,25	2300
36	silte arenoso	24	0,1	6	2,753029163	2,25	1380
37	silte argiloso	11	0,114	9	2,49454926	1,97	1530
38	silte argiloso	11	0,114	6,5	3,100742175	1,97	1007,5
39	silte argiloso	11	0,114	9,5	2,134988033	1,97	1235
40	silte argiloso	11	0,114	9	2,537069986	1,97	1530

Tabela 23 – Dados experimentais para treinamento e teste de modelo

Uma breve explicação sobre as entradas:

- Tipo de solo: características geológicas e geotécnicas do solo;
- SPT: Número de golpes necessários para que o amostrador padrão do "teste de penetração padrão" penetre 30 cm de um total de 45 cm registrado no solo. Esta variável representa a resistência do solo. Quanto maior o SPT, maior será a rigidez do solo;

- Diâmetro perfurado (m): Diâmetro perfurado no solo para inserir o sistema de parede de amarração;
- Lb (m): Comprimento de amarração ancorada;
- k (kN/m²): coeficiente de ancoragem ao solo;
- β : coeficiente de aumento do diâmetro do bulbo devido à injeção de cimento.

A capacidade de suporte de saída (kN) define a carga máxima que uma amostra de solo específica manipula dependendo de suas características (conforme definido pelas entradas). Para as entradas do modelo, consideraremos as entradas 2 a 6, uma vez que a entrada 1 não é numérica e o modelo proposto trata apenas de variáveis quantitativas.

Os dados possuem 40 amostras e foram divididos aleatoriamente entre conjuntos de treinamento e teste. A divisão foi feita de forma aleatória para que o modelo não fosse enviesado. Por exemplo, se fosse usado durante o treinamento apenas amostras de um tipo específico de solo, o modelo não seria capaz de prever as cargas com acurácia quando uma amostra de um tipo de solo diferente fosse fornecida. 80% (32 amostras) dos dados foram usadas para treinamento e 20% foram usadas para teste.

A proporção 80% - 20% para treinamento e teste é uma proporção comumente usada ao modelar séries. Foi feito um teste nessa proporção e os resultados foram satisfatórios. O algoritmo selecionou amostras para treinamento e teste aleatoriamente, de forma que a base do banco de dados pudesse ter amostras de todos os tipos de solos e o modelo não fosse tendencioso.

4.1.3.2 Treinando o modelo

Os dados apresentados na Tabela 23 tiveram suas linhas randomizadas e 32 amostras do total de 40 amostras (80%) foram usadas para treinar o modelo conforme mostrado na Figura 25.

Após múltiplos testes randomizando as linhas, o melhor resultado foi alcançado quando as linhas na Tabela 23 foram organizadas na seguinte ordem de amostras

$$S = \{7, 38, 29, 10, 18, 26, 14, 16, 13, 33, 4, 19, 20, 15, 31, 12, 40, 22, 37, 30, 39, 9, 3, 6, 34, 36, 35, 24, 32, 17, 2, 8, 1, 23, 25, 28, 21, 11, 5, 27\}$$

O primeiro treinamento foi feito com as primeiras 32 amostras de S, nas quais temos o primeiro conjunto de treinamento:

$$S_{Train} = \{7, 38, 29, 10, 18, 26, 14, 16, 13, 33, 4, 19, 20, 15, 31, 12, 40, 22, 37, 30, 39, 9, 3, 6, 34, 36, 35, 24, 32, 17, 2, 8\}$$

Os resultados deste primeiro treinamento podem ser resumidos como segue. O vetor P alcançado é mostrado na Tabela 24.

\vec{P}
4.23×10^6
8.02×10^3
-4.46×10^6
7.02×10^1
1.94×10^2
-1.69×10^6
-9.07×10^6
1.41×10^4
1.66×10^7
1.96×10^2
-8.16×10^1
3.12×10^6

Tabela 24 – Vetor P para o primeiro treinamento

A Figura 29 mostra os resultados da previsão sobre os próprios dados de treinamento.

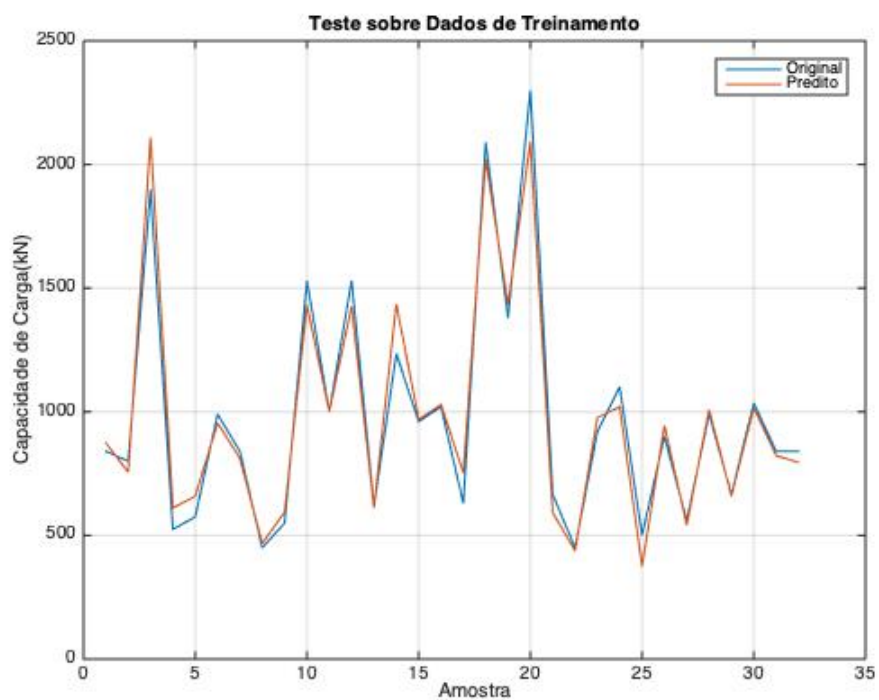


Figura 29 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais.

A partir da Figura 29 pode-se ver facilmente o quão bem o modelo prevê a capacidade de carga nos dados de treinamento. Na próxima seção, o modelo será testado para dados fora do conjunto de treinamento.

4.1.3.3 Teste fora dos dados de treinamento

Para este teste, usaremos uma técnica de janela deslizante. Primeiro, os dados a serem testados serão as amostras no seguinte conjunto (índices da Tabela 23):

$$S_{Test} = \{1, 23, 25, 28, 21, 11, 5, 27\}$$

O teste, usando a técnica de janela deslizante, será realizado da seguinte forma:

- Para prever a primeira amostra de S_{Test} , que é a amostra de número 33 de S (índice 1 da Tabela 23), o modelo será treinado usando amostras de 1 a 32 de S .
- Para prever a segunda amostra de S_{Test} , que é a amostra de número 34 de S (índice 23 da Tabela 23), o modelo será treinado usando amostras de 2 a 33 de S .
- Para prever a n -ésima amostra de S_{Test} , que é a amostra de número $n + 32$ de S , o modelo será treinado usando amostras de n a $n + 31$ de S .

A Tabela 25 resume o teste final. Ela mostra as amostras testadas (com índices da Tabela 23), os valores previstos como em (COULOMB, 1776), (OSTERMAYER, 1974), (BUSTAMANTE; DOIX, 1985), (PORTO, 2015) e as previsões pelo modelo proposto neste trabalho.

Índice de Dados	Valor Experimental (kN)	Coulomb, 1776 (kN)	Ostermayer, 1974 (kN)	Bustamante, 1985 (kN)	Porto, 2015 (kN)	Modelo Proposto, 2021 (kN)
1	900	563	1339	1181	966	944
23	560	279	962	1075	588	628
25	990	406	1237	1237	1428	954
28	665	582	938	1378	729	684
21	1035	406	1237	1528	1444	1001
11	840	530	895	1453	1814	744
5	840	364	938	965	789	728
27	632	248	756	889	1059	839

Tabela 25 – Resultados do teste em comparação com dados originais e outros autores.

A Figura 30 mostra a curva com os resultados da Tabela 25. Outra visão sobre a validade dos resultados de cada modelo em comparação com os dados experimentais originais pode ser vista na Figura 31 que mostra a distribuição de 15 quantis de cada modelo, uniformemente espaçados.

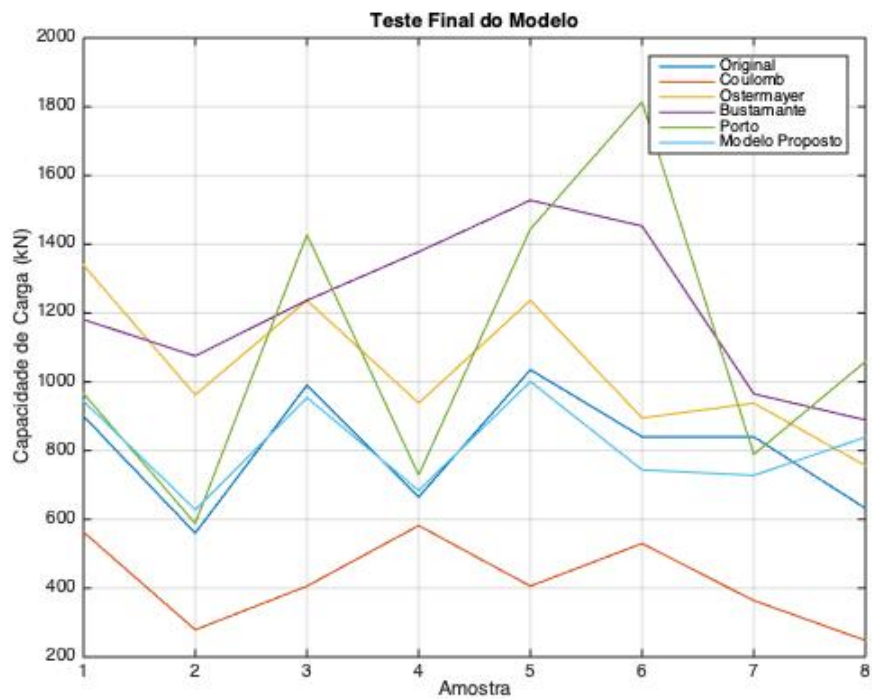


Figura 30 – Curva de previsão sobre os dados de teste em comparação com os dados originais e outros modelos.

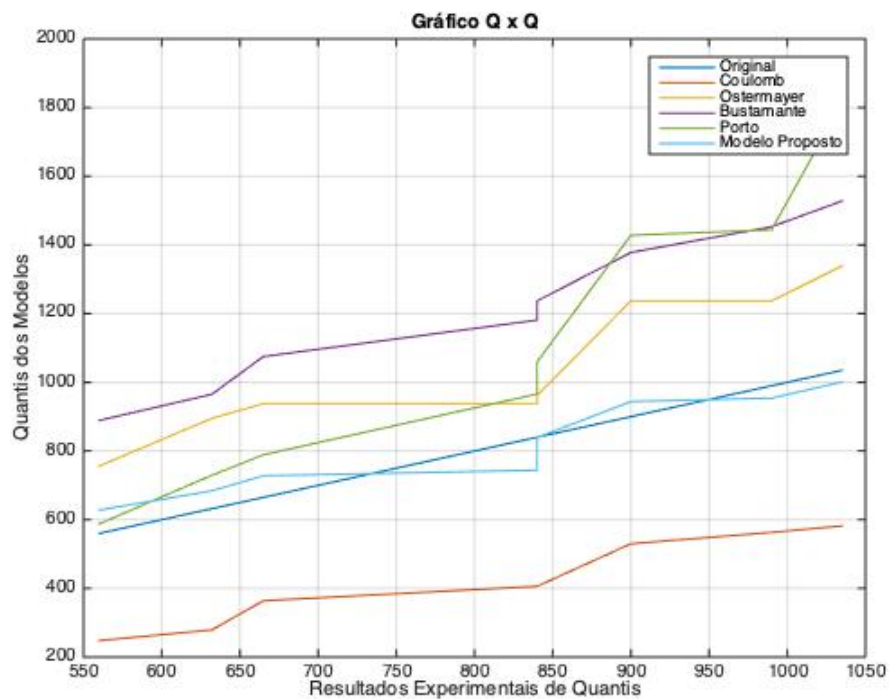


Figura 31 – Gráfico Q x Q comparando os modelos e os dados experimentais. O gráfico Q x Q é usado para comparar duas distribuições de probabilidade traçando seus quantis um contra o outro. Quanto mais próximos os gráficos estiverem, maior a probabilidade de as distribuições serem semelhantes, o que comprova a acurácia do modelo.

Podemos usar os dados da Tabela 25 para calcular o RMSE do modelo proposto e dos demais modelos existentes a fim de se verificar a melhoria proporcionada pelo modelo.

O RMSE (root mean square error) é uma medida do erro de modelo comparando-o com os dados esperados a serem previstos. O RMSE pode está rerepresentado na Equação 4.5.

$$RMSE_{modelo} = \sqrt{\frac{\sum_{k=1}^n (y(k) - \hat{y}(k))^2}{n}} \quad (4.5)$$

A Tabela 26 mostra os valores de RMSE de cada modelo analisado e o modelo proposto.

RMSE Coulomb, 1776	RMSE Ostermayer, 1974	RMSE Bustamante, 1985	RMSE Porto, 2015	RMSE Modelo Proposto
418.8	264.8	449.6	433.5	96.1

Tabela 26 – Comparação RMSE entre o modelo proposto e os modelos existentes

A partir da Figura 30 e Tabela 26 pode-se ver a melhora significativa que o modelo proposto tem sobre os outros modelos.

4.1.4 Discussão

O modelo proposto demonstrou ter melhores resultados que 4 modelos conhecidos na literatura. A partir da Tabela 26, pode-se ver facilmente que o modelo proposto forneceu o menor RMSE (erro médio) sendo aproximadamente 66% melhor que o modelo Ostermayer e 75% melhor que os demais modelos. A Figura 30 mostra todos os gráficos de previsão sobrepostos para facilitar a análise. A figura mostra que a curva do modelo proposto está muito mais próxima dos dados originais do que as demais curvas, o que comprova visualmente seu melhor ajuste.

4.1.5 Conclusões

Na Engenharia Geotécnica, é extremamente comum tentar encontrar a carga máxima de ancoragem permitida para um tipo de solo. A melhor maneira de encontrar pode ser muito difícil, especialmente se vários experimentos em campo tiverem que ser realizados. Este trabalho forneceu um novo modelo para encontrar a carga máxima sem ter que fazer experimentos em campo. Trata-se de um novo modelo para ser utilizado em Engenharia Civil (subárea Geotécnica) misturando aspectos de inteligência computacional em que um modelo fuzzy é utilizado para fins de classificação tendo o universo particionado em clusters com base na distribuição dos dados estatísticos de saída. A previsibilidade do modelo foi comparada com trabalhos existentes na área para verificar sua acurácia. Os resultados deste estudo indicaram melhor desempenho do modelo proposto em comparação a esses.

Foi criado um modelo fuzzy baseado em algumas características específicas de cada tipo de solo. O modelo mostrou-se adequado na previsão da carga máxima em cada tipo de solo. Diferentemente do algoritmo fuzzy C-means e algoritmos de agrupamento subtrativo, o modelo proposto gera os agrupamentos com base na distribuição estatística dos dados de saída, ao invés de utilizar variáveis como similaridade e distância.

Este estudo apresentou um novo método de incorporação de incerteza no processo de previsão da capacidade de carga de ancoragem. A incerteza em duas fontes é abordada: incerteza nos parâmetros de entrada do solo para modelos de suporte de carga de ancoragem e incerteza nos próprios modelos de suporte de carga de tirantes. O novo método usa conjuntos fuzzy para caracterizar a incerteza. O modelo proposto é utilizado para processar os dados fuzzy em modelos determinísticos de capacidade de carga. O novo método produz melhores resultados. Com base nos resultados de teste de carga disponíveis, o modelo proposto previu consistentemente as capacidades de carga das âncoras com um erro aceitável em termos de RMSE.

Aumentar a variabilidade das propriedades do solo aumenta a incerteza na capacidade prevista das âncoras, como esperado. A capacidade do modelo proposto em considerar a incerteza nos parâmetros do solo permite avaliar a variação da capacidade prevista da ancoragem. Isso é importante, pois o engenheiro pode superestimar (ou subestimar) a capacidade de carga da âncora se os parâmetros "reais" do solo se desviarem dos valores estimados usados na análise. Com o modelo proposto nesta pesquisa, o projetista pode agora incorporar sistematicamente a incerteza percebida ou estimada na análise da capacidade de carga das âncoras e obter uma estimativa com mais confiança.

4.2 Modelo fuzzy para predição de capacidade de carga geotécnica de tirantes

Para o desenvolvimento do modelo utilizou-se um banco de dados de 40 (quarenta) tirantes, dos quais, 32 (trinta e dois) foram utilizados para treinar o algoritmo e 8 (oito) foram utilizados para testar o modelo preditor. As 40 (quarenta) amostras foram divididas aleatoriamente entre o conjunto de treinamento (80%) e de teste (20%). O algoritmo selecionou amostras para treinamento e teste aleatoriamente, de forma que a base do banco de dados de teste e treinamento pudessem ter amostras representativas quanto as variáveis de entrada do problema. Assim, o modelo não seria tendencioso. A capacidade de carga dos tirantes é a variável de saída do modelo e a unidade é kN. Para o treinamento do algoritmo foi utilizado os valores de capacidade de carga extrapolados da curva carga x deslocamento dos boletins de protensão através da função de transferência de carga elastoplástica proposta por (SOUZA, 2001). Por fim, as previsões serão comparadas às capacidades de carga inferidas na extrapolação (dados experimentais) para testar a acurácia do modelo. A mensuração da acurácia do modelo será realizada através do erro quadrático médio (RMSE). Desse modo, pretendeu-se contribuir com

o atual estado da arte do projeto de tirantes no Brasil.

4.2.1 Estudo de Caso

O estudo de caso contemplado nesta seção refere-se a uma estrutura de contenção empregada na execução da Estação Vila Sônia do Metrô (Linha 4 – Lote 2) Amarela, ao longo da Avenida Francisco Morato, no bairro do Butantã em São Paulo-SP. A Tabela 27 mostra os dados de 40 tirantes ensaiados na obra, com suas respectivas informações geométricas e de caracterização geotécnica.

4.2.2 Algoritmo de predição - Treinamento do modelo

Apresenta-se nos itens que se seguem os resultados obtidos utilizando-se a Lógica Fuzzy, bem como discussão dos resultados. A fim de se avaliar o desempenho do modelo preditor proposto, uma comparação com métodos clássicos (Mohr-Coulomb, Ostermayer, Bustamante e Porto) foi realizada, mostrando-se que os resultados encontrados foram satisfatórios.

Os dados apresentados na Tabela 27 tiveram suas linhas tomadas como aleatórias, das quais 32 amostras do total de 40 amostras (80%) foram “sorteadas” e utilizadas para treinar o modelo.

Após múltiplos testes de randomização das linhas, o melhor resultado foi alcançado quando as linhas da Tabela 27 foram organizadas na seguinte ordem de número de amostras: $S = \{12, 8, 20, 35, 24, 21, 33, 27, 2, 29, 1, 31, 34, 17, 25, 5, 10, 14, 39, 15, 13, 19, 23, 38, 36, 9, 26, 40, 3, 7, 37, 22, 28, 32, 16, 6, 18, 4, 11, 30\}$. O primeiro treinamento foi feito com as primeiras 32 amostras de S , nas quais tem-se o primeiro conjunto de treinamento: $STrei = \{12, 8, 20, 35, 24, 21, 33, 27, 2, 29, 1, 31, 34, 17, 25, 5, 10, 14, 39, 15, 13, 19, 23, 38, 36, 9, 26, 40, 3, 7, 37, 22\}$. Os resultados deste primeiro treinamento podem ser observados na Figura 32.

A partir da Figura 32, pode-se ver facilmente o quão bem o modelo prevê a capacidade de carga nos dados de treinamento. Na próxima seção, o modelo será testado para dados fora do conjunto de treinamento.

4.2.3 Algoritmo de predição - Teste do modelo

Os dados apresentados na Tabela 27 tiveram suas linhas tomadas como aleatórias, das quais 8 amostras do total de 40 amostras (20%) foram “sorteadas” e utilizadas para testar o modelo.

Os dados a serem testados serão as amostras no seguinte conjunto (Linhas da Tabela 27): $STeste = 28, 32, 16, 6, 18, 4, 11, 30$.

O teste usando a técnica da janela deslizante será realizado da seguinte forma:

Amostra	Tirantes	Solo	SPT	Diâmetro do furo (m)	Lb (m)	K (kN/m ²)	β	Capacidade de Carga (kN)
1	TR-A-119	Silte Argiloso	22,00	0,127	9	2,839712482	1,97	1674,00
2	TR-A-47	Silte Argiloso	14,67	0,127	8	3,888833283	1,97	1440,00
3	TR-A-20	Silte Argiloso	13,00	0,127	8	4,246209861	1,97	1424,00
4	TR-A-19	Silte Argiloso	13,00	0,127	8	4,651746757	1,97	1560,00
5	TR-A-128	Silte Argiloso	22,00	0,127	9	3,282463353	1,97	1935,00
6	TR-A-124	Silte Argiloso	22,00	0,127	9	2,748108854	1,97	1620,00
7	TR-A-127	Silte Argiloso	22,00	0,127	9	2,748108854	1,97	1620,00
8	TR-A-126	Silte Argiloso	22,00	0,127	9	3,053454282	1,97	1800,00
9	TR-A-112	Silte Argiloso	22,00	0,127	9	2,671772496	1,97	1575,00
10	TR-A-44	Silte Argiloso	14,67	0,127	8	4,212902724	1,97	1560,00
11	TR-A-40	Silte Argiloso	14,67	0,127	8	4,104879577	1,97	1520,00
12	TR-A-15	Silte Argiloso	13,00	0,127	8	4,246209861	1,97	1424,00
13	TR-A-14	Silte Argiloso	13,00	0,127	8	4,246209861	1,97	1424,00
14	TR-A-13	Silte Argiloso	13,00	0,127	8	3,93609341	1,97	1320,00
15	TR-A-12	Silte Argiloso	13,00	0,127	8	3,816817852	1,97	1280,00
16	TR-A-06	Silte Argiloso	13,00	0,127	9	3,816817852	1,97	1440,00
17	TR-A-07	Silte Argiloso	13,00	0,127	9	4,198499637	1,97	1584,00
18	TR-A-50	Silte Argiloso	14,67	0,127	8	3,888833283	1,97	1440,00
19	TR-A-54	Silte Argiloso	14,67	0,127	8	4,32092587	1,97	1600,00
20	TR-A-53	Silte Argiloso	14,67	0,127	8	4,32092587	1,97	1600,00
21	TR-A-52	Silte Argiloso	14,67	0,127	8	4,32092587	1,97	1600,00
22	TR-A-51	Silte Argiloso	14,67	0,127	8	3,67278699	1,97	1360,00
23	TR-A-49	Silte Argiloso	14,67	0,127	8	3,456740696	1,97	1280,00
24	TR-A-48	Silte Argiloso	14,67	0,127	8	3,780810137	1,97	1400,00
25	TR-A-46	Silte Argiloso	14,67	0,127	8	3,888833283	1,97	1440,00
26	TR-A-28	Silte Argiloso	13,00	0,127	8	4,174644526	1,97	1400,00
27	TR-A-27	Silte Argiloso	13,00	0,127	8	3,864528075	1,97	1296,00
28	TR-A-26	Silte Argiloso	13,00	0,127	8	3,93609341	1,97	1320,00
29	TR-A-25	Silte Argiloso	13,00	0,127	8	4,055368968	1,97	1360,00
30	TR-A-24	Silte Argiloso	13,00	0,127	8	4,055368968	1,97	1360,00
31	TR-A-23	Silte Argiloso	13,00	0,127	8	4,246209861	1,97	1424,00
32	TR-A-22	Silte Argiloso	13,00	0,127	8	4,055368968	1,97	1360,00
33	TR-A-21	Silte Argiloso	13,00	0,127	8	4,174644526	1,97	1400,00
34	TR-A-125	Silte Argiloso	22,00	0,127	9	3,511472424	1,97	2070,00
35	TR-A-123	Silte Argiloso	22,00	0,127	9	3,206126996	1,97	1890,00
36	TR-A-122	Silte Argiloso	22,00	0,127	9	3,328265167	1,97	1962,00
37	TR-A-121	Silte Argiloso	22,00	0,127	9	3,206126996	1,97	1890,00
38	TR-A-120	Silte Arenoso	29,33	0,127	12,5	1,770688694	1,97	1875,00
39	TR-A-09	Silte Argiloso	10,67	0,127	8	4,496397616	1,97	1288,00
40	TR-A-08	Silte Argiloso	10,67	0,127	8	5,027028391	1,97	1440,00

Tabela 27 – Dados de entrada do modelo

1. Para prever a 1ª amostra de STeste, que é a 33ª amostra de S, o modelo será treinado usando as amostras 1 a 32 de S.
2. Para prever a 2ª amostra do conjunto de STeste, que é a 34ª amostra de S, o modelo será treinado usando as amostras 2 a 33 de S.
3. Para prever a enésima (n^a) amostra de STeste, que é a $(n+32)^a$ amostra de S, o modelo será treinado usando as amostras n até $(n + 31)$ de S.

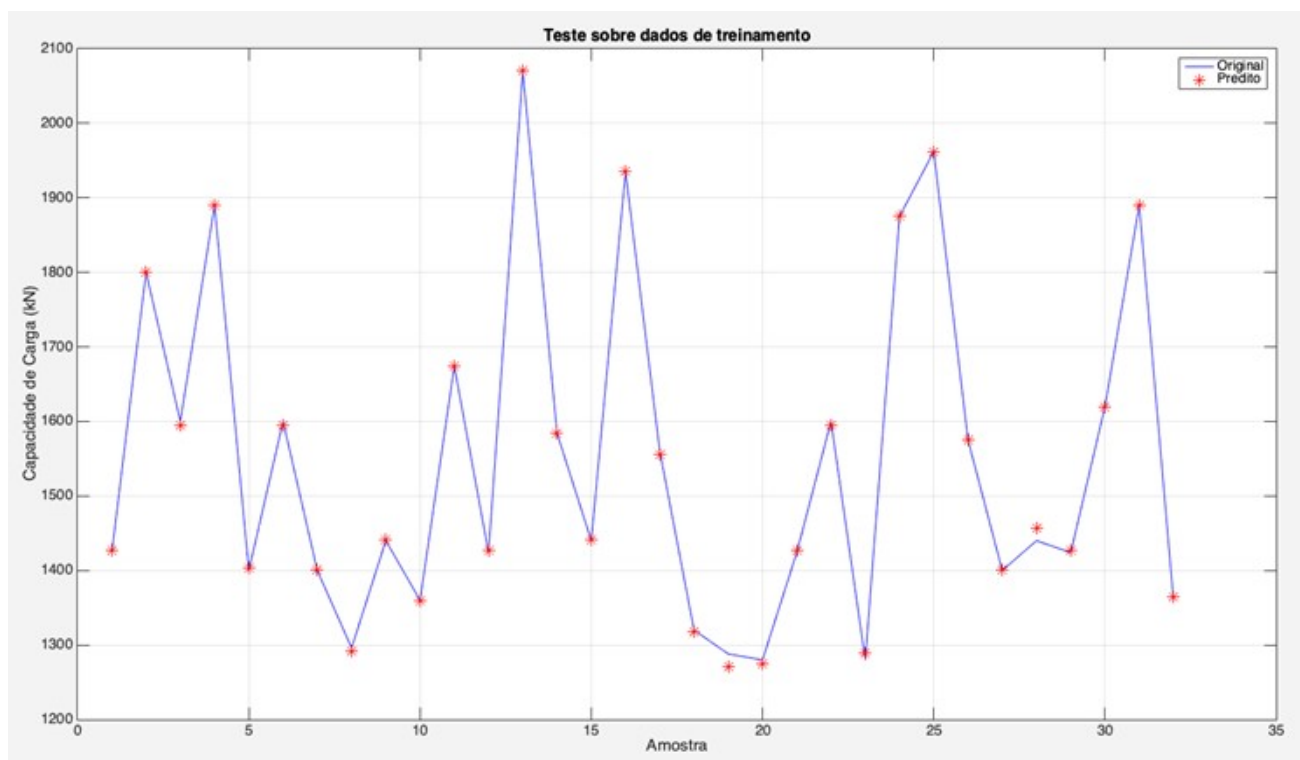


Figura 32 – Curva de previsão sobre os dados de treinamento em comparação com os dados experimentais originais

A Tabela 28 resume o teste final e mostra as amostras testadas, os valores previstos como em ((COULOMB, 1776)), ((OSTERMAYER, 1974)), ((BUSTAMANTE; DOIX, 1985)), ((PORTO, 2015)) e as previsões geradas pelo modelo proposto.

Índice dos dados amostrais	Valores Experimentais	Mohr-Coulomb (1776)	Ostermayer (1974)	Bustamante (1985)	Porto (2015)	Modelo Proposto
28	1320	756	1070	2439	2639	1317
32	1360	756	1070	2339	2719	1359
16	1440	906	1203	2518	2879	1450
6	1620	917	1204	2914	3240	1619
18	1440	729	1070	2439	2879	1440
4	1560	756	1070	2640	3119	1568
11	1520	729	1070	2439	3039	1515
30	1360	756	1070	2339	2719	1359

Tabela 28 – Resultados do teste em comparação com dados originais e outros autores (kN).

A Figura 33 mostra o gráfico com os resultados apresentados na Tabela 28.

4.2.4 Estudo da assertividade do modelo

Outra visão sobre a acurácia dos resultados de cada modelo em comparação com os dados experimentais originais pode ser vista na Figura 34, que mostra a distribuição de 15 quantis de resultados de cada modelo igualmente espaçados.

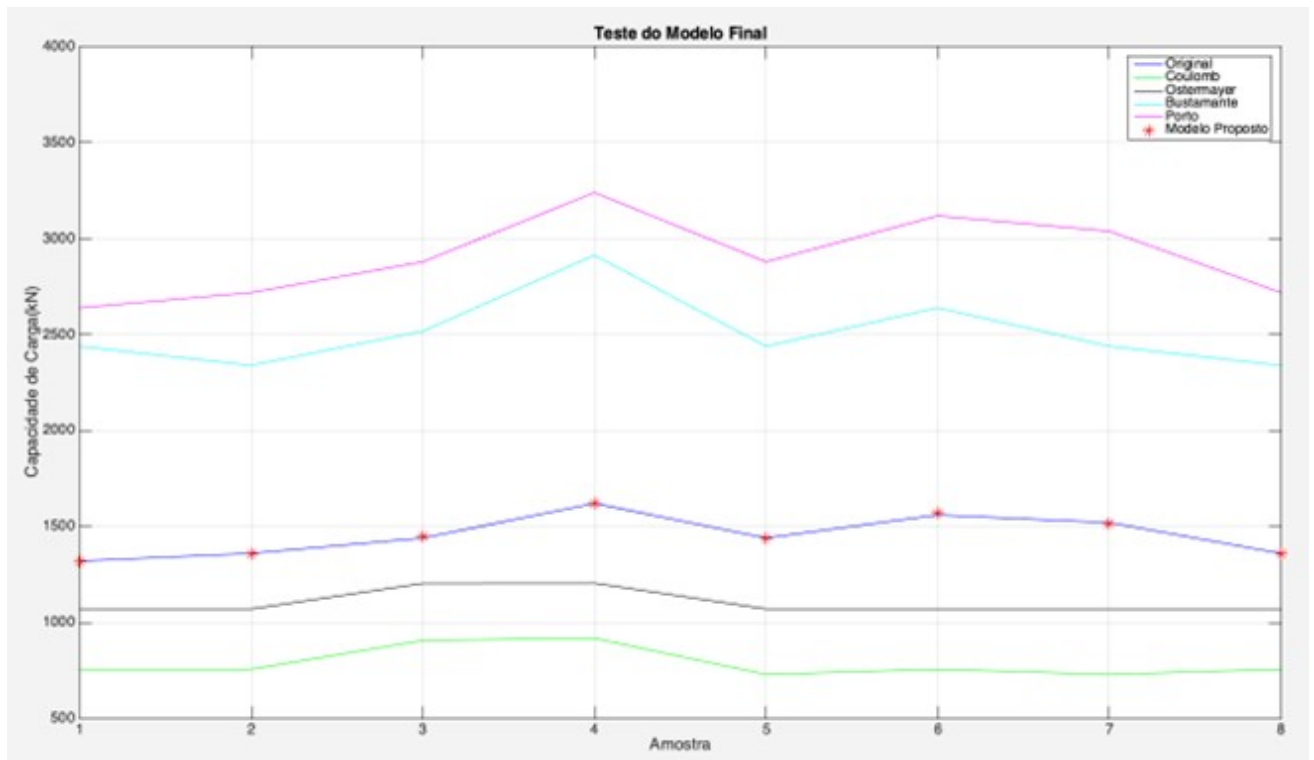


Figura 33 – Curva de previsão sobre os dados de teste em comparação com os dados originais e outros modelos.

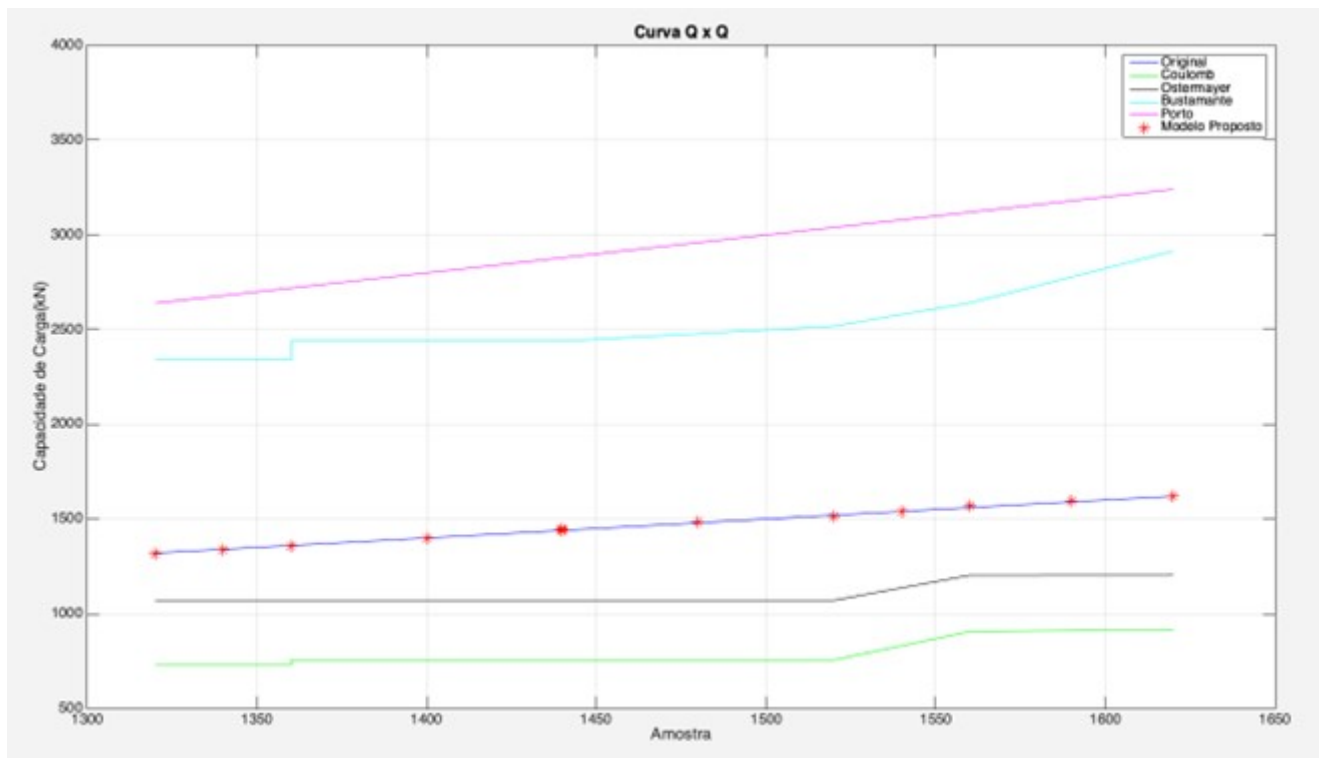


Figura 34 – Gráfico Q x Q comparando os modelos e os resultados experimentais.

Pode-se usar os dados da Tabela 28 para calcular o RMSE do modelo proposto e dos outros modelos existentes para verificar a melhoria proporcionada pelo modelo. O RMSE (erro quadrático médio) é uma medida de um erro de modelo comparando-o com os dados esperados a serem previstos. A Tabela 29 mostra os valores de RMSE de cada modelo analisado e o modelo proposto nesta simulação.

RMSE				
Mohr-Coulomb (1776) 671	Ostermayer (1974) 360	Bustamante (1985) 1062	Porto (2015) 1456	Modelo Proposto 4,9

Tabela 29 – Comparação RMSE entre o modelo proposto e os modelos existentes.

4.2.5 Considerações finais

Na Engenharia Geotécnica é extremamente comum tentar encontrar a carga máxima que a ancoragem pode suportar considerando formulações analíticas que possuem correlações com SPT. Nas últimas décadas vários pesquisadores se dedicaram a esse desafio, no entanto, nenhum obteve 100% de êxito. Este trabalho forneceu um novo modelo para encontrar a carga máxima das ancoragens considerando uma grande variabilidade de fatores que influenciam o comportamento dos tirantes. Os resultados obtidos foram próximos dos valores extrapolados dos ensaios de desempenho, mostrando-se uma alternativa viável para o dimensionamento de ancoragens.

Foi criado um modelo fuzzy baseado em algumas características específicas da ancoragem e seu meio. O modelo mostrou-se adequado na previsão da carga máxima de tirantes protendidos e reinjetáveis.

É importante salientar que como qualquer modelo, o modelo proposto precisou de dados (amostras de treinamento). No caso de uma obra nova em que não existem ensaios antes dos projetos, os dados de treinamento podem ser criados usando informações de obras semelhantes feitas com o mesmo material (ou material semelhante) da obra em estudo. Portanto, para validar o modelo, é importante uma calibração prévia considerando um banco de dados de obras similares existentes. Para trabalhos futuros, os autores pretendem aprimorar o modelo acrescentando características adicionais, como as qualitativas (metodologia executiva, experiência do construtor, etc). Outro ponto de melhoria é mesclar o modelo com redes neurais artificiais (RNA) para aprimorar seus recursos de previsão.

5 Conclusão

O capítulo apresenta a conclusão da tese bem como os produtos desenvolvidos durante a realização da mesma.

5.1 Conclusões da Tese

No trabalho desenvolvido nesta tese, os fundamentos teóricos e as bases de modelagem de séries temporais foram revisadas para fomentar a criação de modelos de predição de ativos e índices financeiros. Além disso, o modelo pôde ser aplicado a uma base de dados da área de geotecnia, onde o foco foi prever a máxima carga suportada por tirantes e ancoragens.

Conforme apresentado no texto, os modelos utilizaram séries temporais e espaciais fuzzy para a realização das predições.

O primeiro modelo foi criado utilizando lógica fuzzy no qual o processo de fuzzificação se baseou na distribuição estatística dos dados de saída do modelo. Basicamente, o modelo utiliza os dados de entrada e os mapeia no espaço de saída utilizando funções lineares das variáveis de entrada. Os dados, então, foram agrupados em clusters que guardam similaridades estatísticas entre si, através da utilização da média (ou mediana) dos dados de saída. O processo de fuzzyficação consiste em aproximar a distribuição dos dados de saída do modelo por uma distribuição normal e, com isso, agrupar os dados de entrada em clusters específicos, baseados na distribuição dos dados de saída. Por fim, o processo de predição se tornou uma função da soma ponderada dos dados de entrada, no qual os pesos são associados ao grau de pertinência dos dados de entrada à função de distribuição normal estimada sobre a saída. Para efeitos de simulação, o modelo foi testado para a previsão da taxa básica de juros da economia brasileira, a Selic, usando como dados de entrada séries temporais disponibilizadas pelo próprio Banco Central do Brasil que podem possuir algum tipo de relação de influência sobre a taxa Selic. A simulação mostrou que o modelo se apresentou adequado para predição da taxa Selic quando comparado a outros modelos existentes na literatura.

O segundo modelo apresentado utilizou uma estrutura de dados denominada "árvore rubro negra" para a realização da divisão do universo de discurso em clusters (ou agrupamentos). A árvore rubro negra consiste em uma árvore de busca binária com equilíbrio automático. Após a construção da árvore rubro negra com os dados de entrada do modelo, a mesma foi utilizada para a divisão dos dados em clusters culminando na fuzzificação dos referidos dados. Regras fuzzy foram então criadas para a construção do modelo de predição. Foram utilizadas métricas de avaliação do modelo como o RMSE (root mean square error) para comparação da performance do mesmo com outros modelos existentes na literatura. Para simulação e teste, foram utilizados

dados do índice Ibovespa. O modelo se mostrou adequado quando comparado a outros modelos existentes no que tange à capacidade de predição do índice considerado.

O terceiro modelo se apresentou como um modelo híbrido mesclando aspectos técnicos e fundamentalistas de séries temporais do mercado de ações. O modelo pode ser considerado um modelo integrador, onde variáveis técnicas e fundamentalistas servem simultaneamente como variáveis de entrada do modelo. Neste modelo, uma preparação dos dados mais detalhada se fez necessária. As entradas fundamentalistas utilizadas foram P/E (Price to Earnings) e P/B (Price to Book). As entradas técnicas utilizadas foram Banda de Bollinger, retornos, volume e S&P 500 passados. Os limites dos clusters são definidos com base nos quantis do vetor de dados de saída. Cada cluster tem um limite inferior e um limite superior. Para as simulações realizadas, foi utilizado o contrato futuro S&P 500 de 07/05/2019 a 29/06/2021. Nos resultados das simulações, o aspecto que mais influenciou o desempenho do modelo foi a forma como o universo de discurso foi particionado, através da utilização de quantis ao invés de valores fixos do comprimento dos clusters. O modelo foi comparado a outros modelos existentes e superou os mesmos no que tange ao desempenho na realização da predição.

O quarto modelo forneceu uma maneira de encontrar a carga máxima de ancoragem para um tipo permitido de solo sem ter que fazer experimentos em campo. Trata-se de um novo modelo para ser utilizado em Engenharia Civil (subárea Geotécnica) misturando aspectos de inteligência computacional em que um modelo fuzzy é utilizado para fins de classificação tendo o universo particionado em clusters com base na distribuição dos dados estatísticos de saída. A previsibilidade do modelo foi comparada com trabalhos existentes na área para verificar sua precisão. Os resultados deste estudo indicaram melhor desempenho em comparação a esses. Foi criado um modelo fuzzy baseado em algumas características específicas de cada tipo de solo. O modelo mostrou-se adequado na previsão da carga máxima em cada tipo de solo. Diferentemente das médias Fuzzy C e algoritmos de agrupamento subtrativo, o modelo proposto gera os agrupamentos com base na distribuição estatística dos dados de saída, ao invés de aspectos como similaridade e distância.

O quinto modelo apresentou um sistema de predição de cargas de tirantes utilizando lógica fuzzy. Neste caso, utilizou-se uma série de dados espaciais onde para cada saída (neste caso a carga do tirante a ser predito) tem-se uma série de entradas como tipo de solo, comprimento de perfuração ancorada, diâmetro perfurado, entre outros. O processo de fuzzyficação seguiu a mesma lógica do processo utilizado no modelo 1. As simulações demonstraram, utilizando critérios como o RMSE, que o modelo obteve um desempenho de predição superior ao de outros modelos de predição de carga já consagrados na literatura.

A pesquisa do doutorado permitiu um aprofundamento em modelos fuzzy com inovações na partição do universo de discurso e na definição das regras fuzzy de predição. Os modelos apresentaram bons desempenhos na predição de séries temporais não estacionárias, como séries financeiras e na predição da capacidade de carga de ancoragens e tirantes a partir de variáveis de

entrada específicas.

Como trabalhos futuros, pretende-se trabalhar na inovação de modelos fuzzy, com foco na partição do universo de discurso e na geração das regras fuzzy a fim de permitir predições mais acuradas e com aplicações mais abrangentes.

5.2 Produtos da tese

Nesta seção serão apresentados os principais produtos de pesquisa da tese, incluindo artigos publicados, com aceite e em etapa de revisão. Os títulos e resumos dos trabalhos gerados são apresentados na mesma língua em que os mesmos foram submetidos/publicados.

5.2.1 Artigo publicado em periódico

Periódico: Journal of Intelligent & Fuzzy Systems, vol. 42, no. 6, pp. 5681-5694, 2022
DOI:10.3233/JIFS-212128

Título: Relationship between Selic rate and Basel III parameters - A statistics approach and a fuzzy forecasting model

Resumo do artigo: In this work the relationship between the Selic rate and some bank parameters defined by the so-called Basel Accords is studied. The cross-correlation between the Selic rate and the parameters is used to explain how these parameters affect the Selic rate and vice-versa so as to define the predictability of the Selic rate using (some of) these parameters as inputs. A model is then proposed for predicting the Selic rate based on some specific parameters using fuzzy logic ideas, which dealt with a partitioning of the universe of discourse using clusters related to the output data distribution. The proposed model is compared to four other known models in the literature and showed to have better performance in average compared to all other models.

5.2.2 Artigo publicado em congresso

Congresso: 5º Congresso Brasileiro de túneis e estruturas subterrâneas - Seminário Internacional "Latina America Tunnelling Seminar - Lat 2021"

Título: Proposta de modelo preditor de capacidade de carga de tirantes usando lógica fuzzy

Resumo do artigo: Pretendeu-se nesta pesquisa desenvolver um modelo de predição de capacidade de carga geotécnica de tirantes por meio de Lógica Fuzzy (LF), também conhecida como Lógica Difusa ou Nebulosa. Trata-se de um ramo da Inteligência Computacional (IC) que se transforma em realidade na Geotecnia. A IC consiste de paradigmas, conceitos, algoritmos e implementações de sistemas numéricos que se adequam a ambientes complexos. Estes métodos

possuem a vantagem de que são tolerantes ao conhecimento impreciso, incompleto e incerto, sendo, portanto, aplicável no encontro de soluções que são aproximações, viáveis e robustas ao mesmo tempo. As bases teóricas da IC são transdisciplinares e trazem conceitos da ciência da computação. Neste contexto, a LF se adequa bem aos problemas de Engenharia Geotécnica, relacionados ao dimensionamento de ancoragens reinjetáveis e protendidas. As variáveis do problema, muitas das vezes são incertas, ficando a critério do projetista firmar premissas simplificadoras para que uma solução seja possível. As versões de 1996 e 2006 da ABNT NBR 5629 apresentavam uma formulação semiempírica para o cálculo da capacidade de carga geotécnica de tirantes. Uma vez que a ordem de grandeza do resultado obtido é função da natureza e qualidade dos dados de entrada, a comissão ABNT CE-02:152.10 responsável pela atualização da NBR 5629 optou por excluir da publicação atual de 2018 o item que auxiliava no dimensionamento do trecho ancorado do tirante e criou o item 4.5.3.1, esclarecendo que o dimensionamento deve ser feito por métodos constantes em publicações técnicas que refletem o estado da arte na mecânica dos solos. Este trabalho se propõe a ser objeto de consultas para projetistas e Engenheiros no que tange o projeto de tirantes. Para validação e aplicação do modelo, utilizou-se o resultado de 40 ensaios de recebimento (provas de carga) executados na Estação Vila Sônia da Linha 4 – Amarela, São Paulo-SP, dos quais, 32 foram utilizados para treinar o algoritmo e 8 foram utilizados para testar o modelo preditor. Os resultados obtidos foram próximos dos valores extrapolados dos ensaios de desempenho, mostrando-se uma alternativa viável para o dimensionamento de ancoragens.

5.2.3 Artigo publicado em periódico

Periódico: Applied Soft Computing, vol. 127, 2022

DOI:<https://doi.org/10.1016/j.asoc.2022.109323>

Título: Fuzzy Time Series Model based on Red-Black Trees for Stock Index Forecasting

Resumo do artigo: Forecasting data, specially in stock markets, is still an extensively investigated area of research. The subjectivity of the elements that influence the market oscillation is the main challenge that any forecasting model faces. In this context, existing fuzzy models have tried to increase forecasting accuracy in financial markets over the years. Fuzzy returns of the phenomenon under investigation helps to mitigate the subjective part of the financial market, specially regarding the human feeling influence over it. Although there are several data structures that can help to define the proper clusters from the universe of discourse of a fuzzy model, this paper proposes a novel fuzzy model from which the universe of discourse is based on a red-black tree (RBT) data structure so as to increase the possibilities of obtaining better predictions. The RBT data structure is a binary search tree data structure that promotes a better balance, which allows a better accuracy in the forecasting results. The proposed model is compared to well known fuzzy models in the literature showing better forecasting results.

5.2.4 Software publicado

Plataforma de publicação: Code Ocean

Link/DOI: <https://doi.org/10.24433/CO.5424678.v1>

Título: Fuzzy Red-BlackTree model

Resumo: O software publicado refere-se ao código criado para gerar os resultados do artigo publicado "Fuzzy Time Series Model based on Red-Black Trees for Stock Index Forecasting" no periódico "Applied Soft Computing".

5.2.5 Artigo em fase final de revisão em periódico

Este artigo encontra-se em fase de revisão apenas por editor.

Periódico: International Journal of Geomechanics

Título: Clustering based fuzzy model for predicting anchor bearing capacity

Resumo do artigo: Despite the widespread use of the anchoring technique in geotechnical works in Brazil, mainly in underground works and retaining wall, limited research has been carried out on the determination of load capacity in anchors. Ground anchored walls in Brazil are often dimensioned by methodologies that are not satisfactorily adapted to the reality of the national soil. In this sense, this research sought to propose a simplified computational methodology for forecasting the load capacity of anchors in geotechnical works. The proposed model was a fuzzy model using selected specific characteristics of the soils and geometry of tieback wall systems, whose measurements are easier to be made compared to the load capacity itself. For validation, calibration and application of the model, the results of 40 performance tests (load tests) performed by the Institute of Technological Research and presented by (PORTO, 2015) were used, of which 32 were used to train the algorithm and 8 were used to test the predictor model. The results obtained were close to the extrapolated values (VEEN, 1953) of the performance tests, proving to be a viable alternative for the design of anchors. Finally, the proposal presented in this work is an unprecedented approach on the subject, which should certainly be improved over time, with the inclusion of a new test database of others places of the world.

Referências

- ABNT. *NBR 5629 - Execução de Tirantes Acorados no Terreno. Associação Brasileira de Normas Técnicas*. Rio de Janeiro, RJ, 2006. Citado na página 87.
- ABNT. *NBR 5629 - Tirantes Acorados no Terreno - Projeto e Execução. Associação Brasileira de Normas Técnicas*. Rio de Janeiro, RJ, 2018. Citado na página 85.
- ADOKO, A. C.; WU, L. Fuzzy inference systems-based approaches in geotechnical engineering- a review. *EJGE*, v. 16, p. 1543–1558, 2011. Citado na página 86.
- AGUIRRE, L. A. *Introdução à Identificação de Sistemas: Técnicas Lineares e Não-Lineares aplicadas a Sistemas Reais*. [S.l.]: Editora UFMG, 2007. Citado 2 vezes nas páginas 34 e 35.
- ALAMEER, Z. et al. Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. *Natural Resources Research*, v. 28, p. 1385–1401, 2019. Citado na página 87.
- AMARAL, R. M.; EPPRECHT, E. K. Compra e venda de ações baseada em gráfico de controle e modelo ar. *Encontro Nacional de Engenharia de Produção*, 2008. Citado na página 20.
- AOKI, N. Considerações sobre a capacidade de carga de estacas isoladas. In: *Curso de Extensão Universitária em Engenharia de Fundações*. Rio de Janeiro, RJ: [s.n.], 1976. Citado na página 88.
- ATSALAKIS, G. S.; DIMITRAKAKIS, E. M.; ZOPOUNIDIS, C. D. Elliott wave theory and neuro-fuzzy systems, in stock market prediction: The wasp system. *Expert Systems With Applications*, v. 38, n. 8, p. 9196–9206, 2011. Citado na página 41.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications*, v. 36, p. 10696–10707, 2009. Citado na página 42.
- AZIZ, N. et al. Soil shear strength as a function of nspt via van der veen extrapolation. *Journal of Rock Mechanics and Rock Engineering*, v. 51, p. 2207–2221, 2018. Citado na página 85.
- BAI, L. e. a. Market turning points forecasting using wavelet analysis. *Physica A: Statistical Mechanics and its Applications*, v. 437, p. 184–197, 2015. Citado na página 20.
- BARBACHAN, J. F. *Probabilidade para Finanças*. [S.l.]: IBMEC, 2003. Citado na página 20.
- BOSE, M.; MALI, K. Designing fuzzy time series forecasting models: A survey. *International Journal of Approximate Reasoning*, v. 111, p. 78–99, 2019. Citado na página 39.
- BRAGA, A.; CARVALHO, A. de; LUDERMIR, T. *Redes Neurais artificiais: Teoria e Aplicações*. [S.l.]: LTC, 2007. Citado na página 34.
- BRASIL, B. C. do. *The Basel Committee on Banking Supervision - <https://www3.bcb.gov.br/ifdata/>*. 2018. Disponível em: <<https://www3.bcb.gov.br/ifdata/>>. Citado na página 49.

BRASIL, B. C. do. *Informações para Análise Econômico Financeira* - <http://www.bcb.gov.br/pt-br>. 2018. Disponível em: <<http://www.bcb.gov.br/pt-br>>. Citado na página 49.

BUSTAMANTE, M.; DOIX, B. Une méthode pour le calcul des tirants et des micropieux injectés. *Bulletin de liaison des laboratoires des ponts et chaussées*, v. 140, p. 75–92, 1985. Citado 3 vezes nas páginas 87, 93 e 99.

CABRERA-PANIAGUA, D. Decision-making system for stock exchange market using artificial emotions. *Expert Systems with Applications*, v. 42, n. 20, p. 7070–7083, 2015. Citado na página 20.

CAETANO, M. A. L.; YONEYAMA, T. A method for detection of abrupt changes in the financial market combining wavelet decomposition and correlation graphs. *Physica A: Statistical Mechanics and its Applications*, v. 391, n. 20, p. 4877–4882, 2012. Citado na página 21.

CAETANO, M. A. L.; YONEYAMA, T. An autocatalytic network model for stock markets. *Physica A: Statistical Mechanics and its Applications*, v. 419, p. 122–127, 2015. Citado na página 21.

CAETANO, M. A. L.; YONEYAMA, T. The effects of node exclusion on the centrality measures in graph models of interacting economic agents. *Physica A: Statistical Mechanics and its Applications*, v. 430, p. 216–223, 2015. Citado na página 20.

CAPUTO, G. V. et al. Improved method for the seismic design of anchored steel sheet pile walls. *ASCE. J. Geotech. Geoenviron. Eng.*, v. 147, p. 943–5606, 2021. Citado na página 85.

CARNEIRO, R. E. et al. Avaliação de habilidade cognitiva em sistemas adaptativos utilizando lógica fuzzy e teoria da resposta ao item. In: BRAGA, A. d. P.; Bastos Filho, C. J. A. (Ed.). *Anais do 11 Congresso Brasileiro de Inteligência Computacional*. Porto de Galinhas, PE: SBIC, 2013. p. 1–6. Citado na página 37.

CARVALHO, E. P. et al. Comparison of test methodologies to evaluate steel-concrete bond strength of thin reinforcing bar. *Construction and Building Materials*, v. 147, p. 243–252, 2018. Citado na página 85.

CARVALHO, L. A.; GOMES, R. C.; PORTO, T. B. Ancoragens em solo: estudo de caso, proposta e comparação entre métodos semi-empíricos e extrapolação de van der veen. In: *Congresso Brasileiro de Mecânica dos Solos e Engenharia Geotécnica*. UFMG, Belo Horizonte, Brasil: [s.n.], 2016. Citado na página 84.

CARVALHO, M. A. R. *Ancoragens pré-esforçadas em obras geotécnicas. Construção, ensaios e análise comportamental*. Tese (Doutorado) — Faculdade de Engenharia da universidade do Porto, Porto, Portugal, 2009. Citado na página 85.

CASTRO, F. H. F.; SILVEIRA, H. P. Modelagem das distribuições das taxas de retorno dos índices ibovespa e s&p500. *Revista de Administração Mackenzie*, v. 10, n. 1, p. 114–133, 2009. Citado na página 20.

CERETTA, P. S. et al. Intraday volatility forecasting: analysis of alternative distributions. *Revista Brasileira de Finanças*, v. 9, n. 2, p. 227–256, 2011. Citado na página 21.

- CHEN, J.; SAYDAM, S.; HAGAN, P. C. Numerical simulation of pull-out behaviour of fully grouted cable bolts. *Construction and Building Materials*, p. 1148–1158, 2018. Citado na página 85.
- CHEN, M.-Y.; CHEN, B.-T. A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Information Sciences*, v. 294, p. 227–241, 2015. Citado na página 41.
- CHEN, S.-M. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*, v. 81, p. 311–319, 1996. Citado 5 vezes nas páginas 36, 44, 50, 70 e 73.
- CHEN, S.-M.; CHANG, Y.-C. Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. *Information Sciences*, v. 180, p. 4772–4783, 2010. Citado na página 86.
- CHEN, S.-M.; TANUWIJAYA, K. Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques. *Expert Systems with Applications*, v. 38, p. 10594–10605, 2011. Citado na página 86.
- CHENA, T.-L.; CHENGA, C.-H.; TEOH, H. J. Fuzzy time-series based on fibonacci sequence for stock price forecasting. *Physica A*, v. 380, p. 377–390, 2007. Citado na página 39.
- CHENG, C.-H.; YANG, J.-H. Fuzzy time-series model based on rough set rule induction for forecasting stock price. *Neurocomputing*, v. 302, p. 33–45, 2018. Citado na página 40.
- CHENG, M. Y. et al. Evolutionary fuzzy neural inference system for decision making in geotechnical engineering. *Journal of computing in civil engineering*, v. 22, p. 272–280, 2008. Citado na página 86.
- CHU, H.-H. et al. Fuzzy dual-factor time-series for stock index forecasting. *Expert Systems with Applications*, v. 39, n. 1, p. 165–171, 2009. Citado na página 41.
- CINTRA, C. A. et al. *Fundações, ensaios estáticos e dinâmicos*. [S.l.]: Oficina de Textos, 2013. Citado na página 88.
- COULOMB, C. A. *Essai sur une application des regles des maximis et minimis a quelques problemes de statique relatifs, a la architecture*. Tese (Doutorado) — Acad. Roy. Div. Sav., France, 1776. Citado 2 vezes nas páginas 93 e 99.
- DATA, B. *BANCO DATA* - <https://bancodata.com.br>. 2018. Disponível em: <<https://bancodata.com.br>>. Citado na página 49.
- DIAS, M. V. L. *Proposta de Dimensionamento Geotécnico de Tirantes por Meio de Ensaio de Recebimento*. Dissertação (M.S. thesis) — Programa de Pós-Graduação em Geotecnia da UFOP, Ouro Preto, 2019. Citado na página 84.
- DIAS, M. V. L.; GOMES, R. C.; PORTO, T. B. Estudo do dimensionamento do bulbo ancorado de tirantes através da análise de ensaios de recebimento. *Revista Geotecnia*, v. 151, p. 89–104, 2021. Citado 2 vezes nas páginas 84 e 85.
- DIBEH, G.; HARMANANI, H. A stochastic chartist–fundamentalist model with time delays. *Computational Economics*, v. 40, n. 2, p. 105–113, 2012. Citado 2 vezes nas páginas 21 e 22.
- DINIZ, P. S. R.; SILVA, E. A. B. da; NETTO, S. L. *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. [S.l.]: Bookman, 2014. Citado na página 34.

- EBID, A. 35 years of (ai) in geotechnical engineering: State of the art. *Geotech Geol Eng*, v. 39, p. 637–690, 2021. Citado na página 86.
- FABRIS, C. et al. Numerical simulation of a ground anchor pullout test monitored with fiber optic sensors. *J. Geotech. Geoenviron*, v. 147, 2021. Citado na página 85.
- FALCONI, F. Capacidade geotécnica de ancoragens reinjetáveis. In: *IV COBRAE - Conferência Brasileira sobre Estabilidade de Encostas*. Salvador-BA: [s.n.], 2005. Citado na página 87.
- FERRARI, A. O. *Um estudo sobre os resultados dos testes fundamentais para o atirantamento provisório no solo de São Paulo*. Dissertação (M.S. thesis) — Escola Politécnica da Universidade de São Paulo, São Paulo, SP, 1980. Citado na página 85.
- FáVERO, L. P. L.; D., C. Modelos multinível de coeficientes aleatórios e os efeitos firma, setor e tempo no mercado acionário brasileiro. *Pesquisa Operacional*, v. 30, n. 3, p. 703–727, 2010. Citado na página 20.
- GONTIJO, G. M. *Reavaliação geotécnica das ancoragens de uma estrutura de contenção com patologias diversas*. Dissertação (M.S. thesis) — Núcleo de Geotecnia, Universidade Federal de Ouro Preto, Ouro Preto, 2020. Citado na página 84.
- GUIBAS, L. J.; SEDGEWICK, R. A dichromatic framework for balanced trees. *19th Annual Symposium on Foundations of Computer Science (sfcs 1978)*, p. 8–21, 1978. Citado na página 59.
- HANAPI, A. L. M. et al. A novel fuzzy linear regression sliding window garch model for time-series forecasting. *Applied Sciences*, v. 10, p. 1949, 2020. Citado na página 86.
- HUANG, M.; ZHOU, Z.; YUE, S. A novel durable intelligent fiber reinforced polymer anchor with embedded optical fiber bragg grating sensors. *ci. China Technol. Sci.*, v. 55, p. 1455–1462, 2012. Citado na página 85.
- IQUIAPAZA, R. A. e. a. Previsão não-linear de retornos na bovespa: Volume negociado em um modelo auto-regressivo de transição suave. *Revista de Administração Contemporânea*, v. 14, n. 1, p. 149–171, 2010. Citado na página 19.
- JADID, R.; SHAHRIAR, A.; RAHMAN, M. Evaluation of theoretical models to predict the pullout capacity of a vertical anchor embedded in cohesionless soil. *Geotech Geol*, v. 37, p. 3567–3586, 2019. Citado na página 85.
- JANG, H.; LEE, J. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, v. 6, p. 5427–5437, 2018. Citado na página 24.
- JR, I. O. J.; MALLMANN, W.; IORIO, W. R. Método de cálculo para estimativa da carga de ruptura de tirantes auto-perfurantes tipo tubular. In: *Seminário de Engenharia de Fundações Especiais e Geotecnia – SEFE V*. São Paulo, SP: [s.n.], 2004. Citado na página 87.
- JR., J. C.; JR., C. C. Identification method for fuzzy forecasting models of time series. *Applied Soft Computing*, 2016. Citado na página 86.
- LANDE, E. L. et al. Effects of drilling for tieback anchors on surrounding ground: Results from field tests. *Geotech. Geoenviron. Eng.*, v. 145, 2020. Citado na página 85.

- LATHI, B. P. *Sinais e Sistemas Lineares*. [S.l.]: Bookman, 2007. Citado na página 34.
- LEE, A. Y. *Corporate Metrics*. USA: Risk Metrics Group, 1999. Citado na página 32.
- LIM, M. A. *The Handbook of Technical Analysis*. [S.l.]: Wiley, 2016. Citado 2 vezes nas páginas 27 e 28.
- LJUNG, L. *System Identification: Theory for the User*. [S.l.]: Prentice Hall, 1999. Citado 2 vezes nas páginas 34 e 35.
- MAGALHÃES, P. H. L. *Avaliação dos Métodos de Capacidade de Carga e Recalque de Estacas Hélice Contínua via Provas de Carga*. Dissertação (M.S. thesis) — Programa de Pós Graduação em Engenharia Civil), UNB, Brasília, 2005. Citado na página 88.
- MARTINS, L. A.; JÚNIOR, E. Aversão a perdas: Comparação de decisões de investimento entre investidores individuais e fundos de pensão no Brasil. *Brazilian Review of Finance*, v. 7, n. 4, p. 429–457, 2009. Citado 2 vezes nas páginas 20 e 23.
- MASSAD, F. D. Notes on the interpretation of failure load from routine pile load tests. *Solos e Rochas*, v. 9, p. 33–36, 1986. Citado na página 87.
- MOAYEDI, H.; MOSALLANEZHAD, M.; RASHID, A. A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications. *Neural Comput & Applic*, v. 32, p. 495–518, 2020. Citado na página 86.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. [S.l.]: Editora Saraiva, 2013. Citado na página 34.
- NAN, G. et al. Heuristic bivariate forecasting model of multi-attribute fuzzy time series based on fuzzy clustering. *International Journal of Information Technology & Decision Making*, v. 11, n. 1, p. 167–195, 2012. Citado na página 40.
- NUNES, A. I. C. Ground prestressing – first Casagrande lecture. In: *VIII CPAMSEF*. Cartagena, Colombia: [s.n.], 1987. Citado na página 87.
- OLIVEIRA, J. D. C. T. de; FRASCAROLI, B. F. Impacto dos fatores macroeconômicos na emissão de ações na bolsa de valores. *Revista de Gestão, Finanças e Contabilidade*, v. 4, n. 1, p. 30–51, 2014. Citado na página 21.
- OLIVEIRA, P. V. G. et al. Utilização do método de porto (2015) no projeto de ancoragens de reforço de cortina atirantada em Belo Horizonte - MG. *Revista Multidisciplinar do Nordeste Mineiro*, v. 1, p. 1–23, 2019. Citado na página 84.
- OSTERMAYER, H. Construction, carrying behavior and creep characteristics of ground anchors. In: *Proceedings of the Conference on Diaphragm Walls and Anchorages*. London: [s.n.], 1974. p. 141–151. Citado 2 vezes nas páginas 93 e 99.
- OZSOYLEV, H. N. e. a. Investor networks in the stock market. *Review of Financial Studies*, v. 27, n. 5, p. 1323–1366, 2014. Citado 2 vezes nas páginas 20 e 23.
- PAVLIDIS, N. et al. Financial forecasting through unsupervised clustering and neural networks. *Operational Research*, v. 6, n. 2, p. 103–127, 2006. Citado na página 21.

- PEREIRA, A. B.; PORTO, T. B. Soil-structure interaction applied to anchored retaining wall-geotechnical and structural analysis. *International Journal of Geoscience*, v. 2, p. 1–13, 2020. Citado na página 84.
- PINHEIRO, C. A. O. Abnormal gains among historical series of trade scrips on corporative business practices at brazil. *Revista De Gestao, Financas E Contabilidade*, v. 3, n. 2, p. 121–134, 2013. Citado na página 19.
- PINTO, J. M.; NEVES, R. F.; HORTA, N. Boosting trading strategies performance using vix indicator together with a dual-objective evolutionary computation optimizer. *Expert Systems with Applications*, v. 42, n. 19, p. 6699–6716, 2015. Citado na página 21.
- PORTO, T. B. *Comportamento geotécnico e metodologia via web para previsão e controle*. Tese (Doutorado) — Universidade Federal de Ouro Preto, Ouro Preto, MG, Brasil, 2015. Citado 4 vezes nas páginas 87, 93, 99 e 106.
- PORTO, T. B.; FERREIRA, D. C.; SILVA, C. R. Estruturas ancoradas e suas aplicações na geotecnia. *REVISTA FUNDACOES & OBRAS GEOTECNICAS*, v. 1, p. 42–53, 2017. Citado na página 84.
- PORTO, T. B.; GOMES, R. C.; MARTINI B.D.AND ARAUJO, D. A. Proposta de metodologia via web para previsão e controle de ancoragens em solos. In: *VII Simpósio Brasileiro de Engenheiros Geotécnicos Jovens*. Belo Horizonte, Brasil: [s.n.], 2016. Citado na página 84.
- PORTO, T. B.; TORRES, A. C. A.; GOMES, R. C. Behavior of reinjectable and prestressed anchors in soil masses: Construction case study in congonghas - brazil. *SOILS & ROCKS*, v. 40, p. 177–186, 2017. Citado na página 84.
- QUEIROZ, T.; MEDEIROS, O.; NETO, J. O. Evidências de bolhas especulativas na bovespa: uma aplicacao do filtro de kalman. *Revista Brasileira de Financas*, v. 9, n. 2, p. 257–265, 2011. Citado na página 20.
- SAHAJDA, K. Ground anchor loads measured on an excavation sheet pile wall. *Tunneling and Underground Construction*, v. 145, p. 974–983, 2014. Citado na página 85.
- SAMUI, P. Application of artificial intelligence in geo-engineering. *Springer Series in Geomechanics and Geoengineering*, 2020. Citado na página 86.
- SANTOS, D. E. et al. Numerical study of slope stability in an anchored wall in belo horizonte, brazil. *International Journal of Science and Engineering Investigations*, v. 8, 2019. Citado na página 84.
- SANTOS, E. B. *Modelagem de Séries Temporais Financeiras Multidimensionais via Processos Estocásticos e cópulas de Lévi*. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo, 2005. Citado 3 vezes nas páginas 19, 20 e 21.
- SANVICENTE, A. Z.; DELGADO, R. T. Learning theory and equity valuation: an empirical analysis. *Revista Brasileira de Financas*, v. 8, n. 2, p. 113–129, 2010. Citado na página 21.
- SEABRA, L. *Conversas com Gestores de Ações Brasileiros: A fórmula dos grandes investidores para ganhar dinheiro na bolsa*. [S.l.]: Editora Schwarcz S. A., 2018. Citado na página 34.
- SHEN, F.; CHAO, J.; ZHAO, J. Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing*, v. 167, p. 243–253, 2015. Citado na página 21.

- SHEN, F.; CHAO, J.; ZHAO, J. Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing*, v. 167, p. 243–253, 2015. Citado na página 21.
- SINGH, P.; DHIMAN, G.; KAUR, A. A quantum approach for time series data based on graph and schrodinger equations methods. *Modern Physics Letters A*, v. 33, n. 35, p. 1–23, 2018. Citado 4 vezes nas páginas 44, 50, 70 e 73.
- SONG, Q.; CHISSOM, B. S. Forecasting enrollments with fuzzy time series. *Fuzzy Sets and Systems*, v. 54, p. 1–9, 1993. Citado 3 vezes nas páginas 35, 36 e 44.
- SONG, Q.; CHISSOM, B. S. Fuzzy time series and its models. *Fuzzy Sets and Systems*, v. 54, p. 269–277, 1993. Citado na página 44.
- SONG, Q.; S.CHISSOM, B. Fuzzy time series and its models. *Fuzzy Sets and Systems*, v. 54, n. 3, p. 269–277, 1993. Citado na página 39.
- SOUZA, R. N. *Ancoragens reinjetáveis e protendidas em solo: previsão de comportamento e controle de execução*. Tese (Doutorado) — Escola Politécnica da Universidade de São Paulo, São Paulo, SP, 2001. Citado 2 vezes nas páginas 85 e 96.
- SUN, B. et al. Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing*, v. 151, p. 1528–1536, 2015. Citado na página 40.
- TABAK, B. M.; GUERRA, S. M. Stock returns and volatility: the brazilian case. *Economia Aplicada*, v. 11, n. 3, p. 329–346, 2007. Citado na página 20.
- TAKIYAR, S.; UPADHYAY, K. G.; SINGH, V. Fuzzy artmap and garch-based hybrid model aided with wavelet transform for short-term electricity load forecasting. *Energy Science & Engineering*, v. 4, p. 14–22, 2016. Citado na página 86.
- TALARPOSHTI, F. M. et al. Stock market forecasting by using a hybrid model of exponential fuzzy time series. *International Journal of Approximate Reasoning*, v. 70, p. 79–98, 2016. Citado na página 40.
- TAVARES, T. H. B. de C. *Simulação de Estratégias de Investimento de Médio Prazo*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Minas Gerais, 2012. Citado na página 28.
- TEOH, H. J. et al. Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. *Data and Knowledge Engineering*, v. 67, p. 103–117, 2008. Citado 4 vezes nas páginas 44, 50, 70 e 73.
- TSAY, R. *Analysis of Financial Time Series*. [S.l.]: John Wiley and Sons, 1951. Citado na página 19.
- VARTANIAN, P. R. Impactos do índice dow jones, commodities e câmbio sobre o ibovespa: uma análise do efeito contágio. *Revista de Administração Contemporânea*, v. 16, n. 4, p. 608–627, 2012. Citado 2 vezes nas páginas 19 e 20.
- VASCONCELOS, L. A. C. *Estudo da capacidade de carga de ancoragens protendidas e reinjetáveis em maciços geotécnicos*. Dissertação (M.S. thesis) — Núcleo de Geotecnia, Universidade Federal de Ouro Preto, Ouro Preto, 2016. Citado na página 84.

- VASCONCELOS, L. A. C.; PEREIRA, J. A.; PORTO, T. B. Anchored structures: State of art in brazil and in the world. *International Journal of Science and Engineering Investigations*, v. 7, p. 61–70, 2018. Citado na página 84.
- VASCONCELOS, L. A. C.; PEREIRA, J. A.; PORTO, T. B. Soil shear strength as a function of nspt via van der veen extrapolation. *REM - International Engineering Journal*, v. 71, p. 513–519, 2018. Citado na página 84.
- VEEN, V. D. The bearing capacity of a pile. In: *Proceedings of the Third Int. Conference on Soil Mechanics and Foundation Engineering*. Zurique, Suíça: [s.n.], 1953. p. 84–90. Citado 2 vezes nas páginas 87 e 106.
- WANG, L.-X. *A Course in Fuzzy Systems and Control*. USA: Prentice-Hall International, Inc., 1997. Citado 2 vezes nas páginas 36 e 38.
- XU, H.-C. et al. Wealth share analysis with “fundamentalist/chartist” heterogeneous agents. *Abstract and Applied Analysis*, v. 2014, p. 1–11, 2014. Citado 2 vezes nas páginas 21 e 22.
- YING, W. e. a. Design and development of financial applications using ontology-based multi-agent systems. *Computing and Informatics*, v. 28, n. 5, p. 635–654, 2012. Citado na página 19.
- YU, H.-K. Weighted fuzzy time series models for taiex forecasting. *Physica A*, v. 349, p. 609–624, 2005. Citado 4 vezes nas páginas 44, 50, 70 e 73.
- ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, v. 8, p. 199–249, 1975. Citado na página 36.
- ZHANG, K. et al. Experimental study of a large-scale ground anchor system with frp tendon and rpc grout medium. *Journal of Composites for Construction*, v. 19, 2015. Citado na página 85.