

Previsão de séries temporais do índice negociado austríaco (ATX) usando neural artificial Modelo de rede

Marko MARTINOVIĆ *, Anica HUNJET, Ioan TURCIN

Resumo: Este artigo analisa o Austrian Traded Index (ATX) da Bolsa de Valores de Viena (Wiener Börse) no período de 2009 a 2017, utilizando o método da rede neural artificial (RNA). Os dados de amostragem são retirados da página web do Wiener Börse e filtrados semanalmente para cumprir a sazonalidade semanal no intervalo de oito anos. O objetivo é construir vários modelos de AAN que atendam a determinados critérios e avaliá-los na subamostra de holdout. Além disso, o objetivo é encontrar o melhor modelo que possa prever novos dados futuros ainda não vistos com alta precisão. Uma estrutura de dados para testar o desempenho da previsão é um período de um mês, um quartil, meio ano e um ano para o qual o último ano da amostra de dados é retido (agosto de 2016 a agosto de 2017). Usando vários critérios e parâmetros diferentes, um total de trinta redes foram construídas e testadas e as cinco principais redes foram analisadas com mais detalhes. Os resultados confirmam a alta acurácia do uso do método de redes neurais artificiais, o que é consistente com estudos realizados em casos semelhantes. A correlação das três principais redes selecionadas por subamostra de validação é superior a 0,9. Os erros percentuais absolutos médios (MAPE) para a melhor rede selecionada são 1,76% (mês); 2,11% (quartil); 2,21% (semestral); 2,13% (ano). Mais uma vez, o método ANN provou ser uma poderosa ferramenta de previsão.

Palavras-chave: redes neurais artificiais; ATX; previsão; predição; mercado de ações; análises de séries temporais; Wiener Börse

1. INTRODUÇÃO

A rede neural artificial (RNA) é um conceito não linear popular frequentemente usado na previsão de séries temporais. Sua popularidade ganhou atenção especialmente nas últimas décadas. Existem muitos outros métodos usados para previsão de séries temporais com diferentes abordagens, e o mais popular entre eles é o método linear Autoregressive Integrated Moving Average (ARIMA) [1]. Neste artigo, a abordagem de rede neural artificial é proposta como um método principal. Embora a combinação de diferentes métodos aumente a precisão [2], especialmente a combinação de modelos não lineares e lineares como redes neurais artificiais e modelos lineares como o ARIMA, este artigo é focado apenas nas RNAs e suas combinações, deixando oportunidade de fazer benchmark e comparar as mesmas método com diferentes variações de características, parâmetros e configurações.

A importância de conhecer o comportamento futuro no mundo dos negócios é bem conhecida, especialmente no setor financeiro. A previsão de ações futuras é a base para a tomada de decisões de negócios sólidas, principalmente se forem baseadas em mais de um indicador que possa aumentar a chance de encontrar o curso de ação correto. Os métodos de previsão são frequentemente empregados na resolução de problemas em vários setores, como setor de energia (demanda de energia), ecologia (previsão de poluição do ar), vendas e demanda de produtos, marketing, previsão de falências, falências, crescimento do PIB, mercados de ações, taxas de câmbio, tráfego, ciência etc. Assim, a precisão do modelo de previsão escolhido tem grande impacto na economia de escala e representa um dos desafios importantes na ciência moderna. [3].

Usando conjuntos de dados históricos, vários modelos são propostos para prever o comportamento próximo ao futuro. O desafio é encontrar o melhor modelo que possa se ajustar ao padrão de dados subjacente (mas não excessivamente), fazendo discrepâncias mínimas da série original e cometendo um erro mínimo de validação. A suposição básica é que a escolha do melhor modelo nos dados de amostra representará a melhor solução para o conjunto de dados futuro. Mas isso nem sempre é correto devido a vários fatores subjacentes, como variação na amostra, mudanças na estrutura de dados e incerteza do modelo [4].

Existem duas abordagens importantes na escolha do método de previsão, casual e não casual. Os métodos casuais funcionam em nome de variáveis explicativas de entrada que são consideradas como de impacto importante ou para variáveis previstas, enquanto os métodos de séries temporais que capturam o padrão subjacente usando apenas conjuntos de dados das observações anteriores tornam muito simples a análise e o estudo. Eles não se restringem à necessidade de conhecer a natureza subjacente dos dados gerados nem de fatores importantes que podem influenciar o resultado. Isso pode representar um problema no caso de modelos causais em que a seleção de variáveis de entrada geralmente resulta em suposições e escolhas subjetivas. Este artigo é focado no modelo de séries temporais.

A segunda grande preocupação é assumir o processo de geração de dados subjacente. Geralmente, pode ser linear ou não linear. Apesar do fato de que muitos problemas do mundo real são inerentemente não lineares, métodos baseados em lineares dominaram por décadas. Algumas das razões são sua simplicidade de assumir e implantar modelos. A outra razão pode estar no fato de que novos métodos não lineares emergentes como as RNAs são empregados em uma história relativamente recente, especialmente no final do século 20 fomentado com um rápido desenvolvimento da tecnologia da informação.

Alguns dos representantes dos modelos lineares são o modelo ARIMA, a média móvel (MA), o modelo de suavização exponencial (ES), a regressão linear múltipla e o modelo autorregressivo (AR). Um dos modelos lineares mais populares é o modelo de média móvel integrada autorregressiva (ARIMA), conhecido como Box-Jenkins [1]: combina os modelos autorregressivo (AR) e de média móvel (MA), que são casos especiais do ARIMA.

Além disso, alguns tipos de modelos de suavização exponencial podem ser representados usando o ARIMA.

Por outro lado, os modelos não lineares, como extensão dos esforços anteriores, englobam o modelo autorregressivo de limiar (TAR), o modelo bilinear, o modelo heteroscedástico condicional autorregressivo (ARCH), etc. [5]. Esses modelos são adaptados para padrões de dados não lineares específicos, portanto, têm limitação na generalização da não linearidade. Além disso, muitos modelos não lineares são paramétricos; significa que uma certa suposição deve ser feita antes da implementação do modelo. Aqui, há propensão a certas discrepâncias de

solução ótima se as características reais dos dados não corresponderem às suposições do modelo.

Uma abordagem que pode superar esses problemas é o modelo de rede neural artificial, que é mais flexível em suposições, adaptação e aproximação. As RNAs representam o modelo não paramétrico e não linear generalizado e serão usadas em análises posteriores para combinar e prever o movimento dos preços do índice ATX.

Com relação a uma pesquisa anterior sobre o índice ATX, Haefke C. e Helmenstein C. usaram cointegração e análise de Granger-causalidade para construir modelos lineares e rede neural artificial feedforward para prever ofertas públicas iniciais austríacas (IPOXATX) com base no relacionamento com o índice austríaco do mercado de ações. O modelo construído mostrou melhor resultado do que a abordagem clássica como a média móvel [6].

S. Chittenkopf et al. [7] usaram modelos de RNA como rede de densidade de mistura (MDN) para comparar o desempenho dos modelos de volatilidade aos modelos padrão. Os resultados mostram a importância da memória de longo prazo do modelo para a estratégia de negociação.

No artigo Previsão de índices e seleção de modelos [8], os autores exploram as diferenças informacionais usando vários princípios de construção de índices do mercado de ações. Eles usaram a previsão do índice um dia à frente para o sinal de compra ou venda. O modelo é testado no Austrian Traded Index usando ações incluídas nele. A previsão foi feita usando métodos econométricos padrão e redes neurais feed forward.

2 METODOLOGIA DE REDE NEURAL ARTIFICIAL

Uma rede neural artificial é um conceito relativamente novo em relação aos métodos estatísticos clássicos [9] e tem sido amplamente utilizado nos negócios e na ciência. As redes neurais artificiais representam uma abordagem não linear, que não requer nenhuma suposição prévia estrita da forma dos dados subjacentes nem do tipo de linearidade da amostra analisada. Independentemente de sua natureza não linear inerente, a RNA também pode lidar com sucesso com problemas lineares. As RNAs têm possibilidades de aproximação muito precisas que podem atender a um amplo conjunto de problemas.

Os modelos de previsão podem ser caracterizados por sua variância e um viés. A variância representa o quanto o modelo é capaz de se adaptar ao conjunto de dados gerado a partir da mesma fonte de dados estática, ou o quanto ele está propenso a mudar devido à mudança de variância. Isso significa, quanto modelo é estável e confiável. Por outro lado, o viés representa o erro do sistema incorporado pela escolha de um determinado modelo pré-especificado. Isso é muito importante no caso de modelos paramétricos, onde suposições sobre o processo de dados subjacentes devem ser feitas antes de implementar o modelo. Outra questão é, o que aconteceu se a estrutura de dados mudou ao longo do tempo. Isso significa, quanto modelo é flexível. Por exemplo, se o processo gerado pelos dados mudar ao longo do tempo, um modelo paramétrico estático pré-especificado terá maior viés. As características preferidas do modelo são baixo viés e baixa variância, mas essas duas são tipicamente exclusivas. Com um modelo mais estável torna-se mais estático, e com um modelo mais flexível, a variação está surgindo.

A RNA representa um modelo flexível e robusto que pode se adaptar facilmente a diversos padrões de dados. Mas o excesso de design (como grande número de nós ocultos e conexões recorrentes) pode impor certas desvantagens. O overfitting é comum em

sistemas complexos, onde muitos parâmetros são incluídos.

Portanto, a variação da amostra de dados é incorporada ao design do modelo, o que não é o preferido. Portanto, a melhor prática é encontrar um equilíbrio entre o viés do modelo e a variação do modelo.

Além disso, uma rede neural é um modelo implícito, o que significa que não pode ser explicado de forma agradável e explícita e se comporta mais como uma "caixa preta" acionada por dados de entrada. Uma vantagem de usar tal abordagem é a possibilidade de capturar não linearidade em padrões de dados sem pressupor a "natureza" dos dados; além disso, possui possibilidades de aproximação funcional universal.

A ideia das RNAs vem da analogia com o cérebro humano, neurônios biológicos e sua fiação.

Embora o cérebro humano seja muito complexo, as redes neurais artificiais são muito mais simples. A unidade-neurônio principal (perceptron) é uma unidade de computação simples que tem a função principal de levar o sinal de entrada para a saída processada. A rede é organizada em camadas interconectadas.

As redes neurais diferem por suas características e design. Existem muitos tipos de RNAs atualmente, mas uma das mais utilizadas para problemas de previsão de séries temporais são as redes feedforward. Neste artigo, redes neurais totalmente conectadas com feedforward de três camadas com um neurônio de saída e algoritmo de aprendizado de retropropagação [10] serão usados.

As redes feedforward (perceptron multicamadas) processam o sinal em apenas uma direção, da entrada à saída, sem qualquer conexão de retorno. Eles são muito simples em design (especialmente no caso de três camadas), mas muito poderosos na resolução de problemas não lineares.

A rede feedforward de três camadas para análises de séries temporais usa observações anteriores como pontos de entrada. Normalmente, um neurônio de saída é usado como previsão um passo à frente. O número de neurônios de saída pode variar para a abordagem multi-step-ahead, onde mais de um neurônio de saída é usado. A primeira camada da rede atua como camada de entrada, a segunda é a camada de nós ocultos (camada oculta) e a terceira camada representa a camada de saída. As camadas geralmente são totalmente conectadas, o que significa que cada neurônio da camada anterior tem uma conexão com todos os nós da camada atual. O Neuron calcula a função trivial como a soma de múltiplas entradas e pesos de conexão e, posteriormente, esse valor é transformado pela função de ativação não linear. Normalmente, a função de ativação para perceptron multicamada é uma função logit ou sigmóide como caso especial Eq. (2).

O perceptron de três camadas pode ser representado com a expressão matemática Eq. (1).

$$y_{ij} = \sum_{j=1}^q w_{ij} y_j + \theta_{ij} \quad (1)$$

Onde:

$$y_j = \frac{1}{1 + e^{-y_j}} \quad (2)$$

y_j - é a função de ativação, p - número de neurônios de entrada, q - número de neurônios ocultos, w - pesos de conexão.

De acordo com o caso específico, esses parâmetros devem ser escolhidos com sabedoria. Enquanto os pesos de conexão são resultados do processo de aprendizado, todos os outros parâmetros são arbitrários. Para o

função de ativação ou transição a função logit é frequentemente usada. O número de neurônios de entrada na camada de entrada representa o número de observações defasadas (observações anteriores) e esse parâmetro desempenha um papel crucial no projeto da rede. A janela de observações anteriores captura a autocorrelação entre pontos de dados adjacentes. Outro parâmetro importante no projeto de rede é o número de nós ocultos. Os nós ocultos são colocados no meio da rede e atuam como um mecanismo de rede. Escolher o número certo de nós ocultos pode melhorar significativamente o desempenho da rede. Se este número for muito pequeno, a rede não tem complexidade para construir um modelo de som. No entanto, se o número for muito grande, a rede enfrenta um problema de overfitting.

Na concepção das RNAs, não existe um procedimento padrão em abordagem sistemática e uniforme, que possa estimar valores exatos dos parâmetros de rede que devem ser utilizados. Para o modelo de previsão de análise de séries temporais, uma saída, várias entradas e vários nós ocultos são usados, geralmente em algumas dezenas ou mais comuns em dez. Muitas vezes, experimentar na amostra de dados é um guia para arbitrar parâmetros, o que muitas vezes pode levar a uma solução abaixo do ideal.

Outros parâmetros que devem ser levados em consideração são o tamanho da amostra e a divisão dos dados. É comumente entendido que a amostra maior tem vantagem na construção de um bom modelo, o que faz sentido, principalmente pela necessidade de mais divisão de dados nos subconjuntos de treinamento, validação e avaliação. Mais observações significam mais chances de uma melhor aproximação da estrutura de dados subjacente. Mas, por outro lado, grandes conjuntos de dados podem sofrer alterações inesperadas no processo de geração de dados subjacentes, o que pode levar a um design de modelo ruim. O tamanho típico da amostra é superior a 100 observações.

A divisão de dados é o próximo passo para projetar a rede. A amostra pode ser dividida em dois subconjuntos, em dados de amostra e dados fora da amostra. Os dados dentro da amostra são usados para treinamento de rede, enquanto os dados fora da amostra representam dados não vistos usados para avaliação do modelo. A divisão nesses dois conjuntos é comum no projeto de vários modelos lineares e não lineares, mas no caso de RNA, é necessária uma divisão adicional. Os dados na amostra são divididos em um conjunto de treinamento e um conjunto de validação. O conjunto de treinamento é utilizado para a construção de redes, enquanto a subamostra de validação tem a finalidade de selecionar as melhores.

A amostra fora de amostra ou de retenção representa dados reais não vistos que não foram incluídos no processo de construção da rede. O objetivo do out-of-sample é avaliar ou descobrir quanto novo modelo se ajusta aos dados reais, ou quanto processo de aprendizado é realizado com sucesso.

O núcleo do processo de aprendizagem é obtido através da amostra de treinamento; a seleção é feita pela amostra de validação e, finalmente, a avaliação do modelo é feita pelos dados fora da amostra. Mais um parâmetro arbitrário é como dividir esses conjuntos de amostras. Embora não exista um padrão exato para isso, a maior parte da literatura sugere que a maior parte da amostra é dedicada ao aprendizado, cerca de 70% a 90%, e o restante é reservado para a avaliação do modelo.

Determinar uma proporção apropriada é muitas vezes o produto da experimentação na amostra original. Para as RNAs, é necessária uma divisão adicional dos dados na amostra.

Após a determinação dos parâmetros corretos e antes de construir um novo modelo, os dados de entrada devem ser processados adicionalmente. Uma rede neural só é boa como seus dados de entrada.

Uma amostra original deve ser processada ou transformada em uma forma mais adequada, incluindo normalização de dados, dessazonalização e eliminação de tendências [4]. Uma rede neural é especialmente sensível à variação de dados, porque o processo de aprendizado autorregressivo depende totalmente dos dados de entrada, portanto, uma entrada mais uniforme é aconselhável.

Os pesos de conexão de rede são variáveis que devem ser obtidas no processo de treinamento. Existem vários métodos de treinamento, mas os mais utilizados para o perceptron multicamadas são os métodos de aprendizado supervisionado chamados retropropagação. Este algoritmo muda gradualmente (descida) os pesos das conexões tentando encontrar a solução ótima, ou seja, minimizar o erro entre os valores dos dados reais e calculados. Muitas vezes, a soma dos quadrados é usada para a função de erro.

A função de erro pode ser vista como superfície no espaço dimensional $N + 1$, onde N representa o número de pesos. O algoritmo de treinamento está tentando encontrar o ponto mínimo nessa superfície. Os algoritmos de descida de gradiente começam a partir de um ponto aleatório na superfície de erro procurando a direção com a maior inclinação e dão um passo em direção a essa direção. O processo iterativo leva ao ponto mínimo ou submínimo. Assim, o número de processos de treinamento, com diferentes pontos de início aleatórios e parâmetros de rede são necessários para encontrar o mínimo global. Existem outros métodos eficientes de otimização não linear como BFGS-Broyden-Fletcher Goldfarb-Shanno ou algoritmos Scaled Conjugate Gradient que se mostram mais eficientes com menos iteração e mais precisos, mas demandam mais poder computacional para o processo de treinamento.

3 O MODELO E DADOS

Embora se saiba que na maioria dos casos não existe uma solução “melhor para todos” que um método seja superior a todos os cenários possíveis, este trabalho foca-se na RNA como método único e nas suas combinações e variações.

3.1 Os Dados

Estudo de pesquisa traça o índice de negociação austríaca (ATX) da Bolsa de Valores de Viena (Wiener Börse). Os dados são retirados da página oficial da Wiener Börse, de 01.02.1997 a 08.08.2017.

O Austrian Traded Index (ATX) é um índice de preços que contém empresas austríacas de primeira linha e de primeira linha listadas na Wiener Börse. O índice é composto por 20 constituintes com o peso máximo de 20% para uma empresa. O preço é atualizado em tempo real e o índice é revisado duas vezes por ano (março e setembro). Os principais critérios para a seleção de empresas no índice são a capitalização e o volume negociado em bolsa. Por revisão, não mais do que três componentes podem ser alterados. O índice foi lançado em 2 de janeiro de 1991, com valor base de 1000.

Hoje, as 5 principais ações representam mais de 75% do valor do índice, com domínio do setor bancário e da indústria de petróleo/gás.

Os dados originais baixados dos preços do índice ATX são diários, o que significa que, para cada dia de negociação aberta, há um registro com os seguintes atributos: data, preço de abertura, preço alto, preço baixo e variação de preço (em porcentagem). Para os propósitos deste artigo, os dados foram transformados no intervalo de tempo semanal. Na literatura, é mais comum referir-se aos preços semanais [11]. Além disso, os valores semanais são mais adequados

a sazonalidade que é então usada no processamento de dados adicional.

Para a construção de um novo modelo, foram considerados apenas os registros a partir da data de 01.02.2009, embora existam observações mais antigas. O motivo é evitar os efeitos da crise econômica mundial, que altera significativamente a estrutura de dados no período de final de novembro de 2007 ao início de fevereiro de 2009. O tamanho total da amostra de preços semanais conta 450 observações.

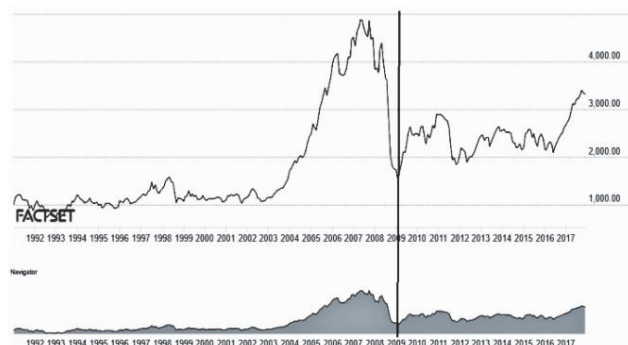


Figura 1 Preço ATX de 1991 a 2017

Os dados à esquerda da linha vertical na Fig. 1 (ponto 2009) foram ignorados. À direita da linha vertical estão os dados de amostra dos modelos. A partir da figura, pode-se ver o aumento e a enorme queda do preço como resultado da crise econômica global.

Tais surtos impulsionados pelas mudanças subjacentes do processo de dados gerados podem levantar um problema de overfitting, especialmente no caso das RNAs.

Embora a RNA seja um modelo robusto que possa se adaptar facilmente a diversas situações sem processamento prévio, uma estrutura mais uniforme é desejável para o treinamento. Isso pode ser alcançado com dados pré-processados, tornando os modelos de previsão mais eficientes e precisos.

O pré-processamento de dados para a RNA pode incluir normalização de dados, dessazonalização (ajuste sazonal), eliminação de tendências e transformação. A transformação de dados traduz os valores originais na área traduzida (imagem) por meio da função de transformação. Para fins de previsão, muitas vezes a função logarítmica natural é usada. O ajuste sazonal remove o padrão sazonal da série temporal, tornando-o mais aplicável e conveniente para a construção de um modelo. A redução de tendência remove a tendência da série original, deixando um padrão de dados claro. Nesta pesquisa, foram utilizados o ajuste sazonal e a destendência.

Existem diferentes métodos e soluções estatísticas desenvolvidas por várias instituições como o United States Census Bureau (X-ARIMA), o Banco da Espanha (TRAMO/SEATS) e outros. A série temporal pode ser mostrada como uma soma das componentes Eq. (3).

Neste artigo, o aditivo de série temporal simples decomposição é usado na ferramenta Microsoft Excel [12, 13].

$$Y_t = T_t + S_t + I_t \quad (3)$$

O período escolhido para ajuste sazonal é uma semana do ano, isso porque os dados são apresentados semanalmente. Primeiramente, a razão das médias móveis (RMA) é calculada usando a diferença entre as séries originais e as médias móveis centrais (52 valores). Depois, a sazonalidade

os índices foram calculados como médias normalizadas da RMA para cada número de semana e para cada ano na amostra relevante os índices sazonais ajustados (ASI) estão representados na Fig. 2. A partir da figura (por exemplo) pode-se ver que a diferença entre o semana números 8 e 39 é mais de 10%. Além disso, a primeira metade do ano em média apresenta valores mais elevados do que a outra metade.

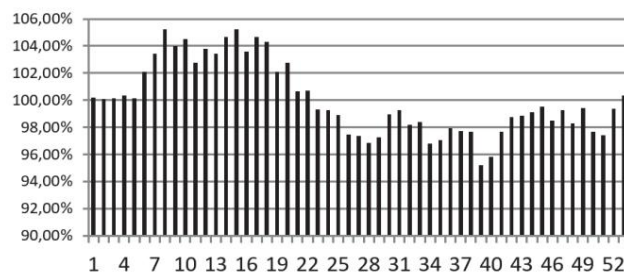


Figura 2 Índices sazonais ajustados semana n. 1-52

Usando os índices sazonais ajustados e a série de dados original, a série ajustada sazonalmente (ASS) pode ser calculada. Depois de determinar o componente S_t da equação Eq. (3), a tendência deve ser removida (Fig. 3).

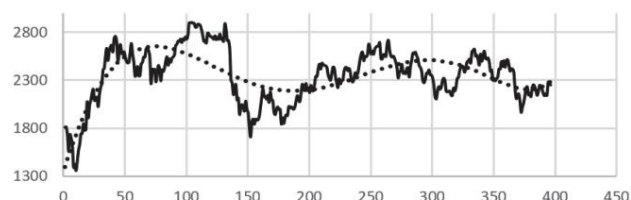


Figura 3 Série sazonal ajustada (ASS) com a linha de tendência

A Fig. 3 inclui valores que estão engajados no processo de aprendizagem, portanto, as últimas 54 observações foram omitidas. Apesar dos ajustes sazonais, ainda existe alguma regularidade. Em vez de usar regressão linear simples para tornar os dados mais uniformes, é usada uma regressão polinomial. No gráfico, a linha pontilhada representa a linha de tendência polinomial de quinta ordem que ajusta os dados com coeficiente de determinação $R^2 = 0,52$. A segunda alternativa foi a linha de tendência linear simples com $R^2 = 0,02$ e inclinação de 0,12, mas não é levada em consideração. Os coeficientes da linha de tendência polinomial são mostrados na Tab. 1.

Tabela 1 Coeficientes da linha de tendência polinomial

x5	x4 x1	x3	x2	b

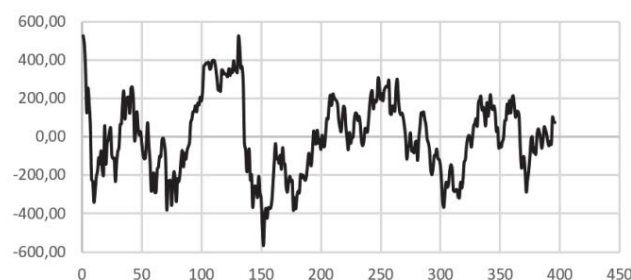


Figura 4 Série original descompensada e com ajuste sazonal (I_t).

Da fórmula Eq. (3), quando os componentes de tendência T e sazonalidade S são removidos, o que resta é It - irregularidade não capturada e função de erro (Fig. 4). Isso representa os dados pré-processados como entrada para a criação dos novos modelos. Na figura, o valor zero (nos eixos x) indica a data de início da amostra (2.1.2009) e 450 indica a data final 8.8.2017. O último ponto do gráfico é 396 (29.07.2016) e é o último ponto da subamostra utilizada para o processo de construção da rede.

3.2 O Modelo

Um modelo de rede neural artificial aplicado nesta pesquisa é o perceptron de três camadas, ou rede neural feedforward de três camadas. A construção e o treinamento foram conduzidos usando o método de mineração de dados StatSoft STATISTICA Automated Neural Network (SANN). O procedimento automatizado combina vários parâmetros de forma a encontrar a melhor solução. As opções que podem ser escolhidas são intervalo de números de nós ocultos, função de ativação (tanto para camada oculta quanto para camada de saída), número de nós de entrada e número de redes para treinar e reter.

As opções definidas para este modelo específico são: número de nós ocultos é [2-12], número de nós de entrada 5, nó de saída 1, funções de ativação são uma do conjunto {identidade, logística, tanh e exponencial}, número de redes treinar e reter é 30; função de erro é a soma dos quadrados.

O número de nós de entrada é um dos parâmetros mais importantes do projeto de rede. Esse número representa quantas observações defasadas o modelo considerará em bloco. Ele captura o padrão de dados movendo a janela de nós de entrada por todas as amostras de treinamento. Foram consideradas duas variantes, a primeira para o período de aproximadamente um mês (5 nós de entrada), e a outra para o período de um ano (52 nós de entrada). Embora ambas as variantes tenham certas vantagens, na literatura mais comum é o uso de um número pequeno, então o número 5 é finalmente aplicado. Um neurônio de saída é definido. Pode ser interpretado como uma previsão antecipada e também é muito comum hoje em dia.

Os dados de séries temporais são divididos em dois subconjuntos principais, nos dados dentro da amostra (representados na Fig. 4) e na amostra fora da amostra ou holdout. A amostra de validação retém observações de preços que não são usadas no processo de construção de rede e representam valores futuros não vistos. O tamanho dos dados da amostra é 396 e o tamanho da amostra de validação é 52.

Cada observação representa um preço semanal considerado como último preço de fechamento no último dia de negociação da semana. Normalmente é sexta-feira, mas pode cair na quinta ou até na quarta em casos especiais (feriados). A ideia é manter observações inéditas por um período de um ano, que engloba os últimos 52 preços semanais (5.8.2016 - 4.8.2017) do conjunto.

A partir dessas observações, será feita a avaliação do modelo selecionado. Além disso, os dados na amostra são divididos em subconjuntos de treinamento e validação. Usando a rede de subamostra de treinamento é construída e usando a subamostra de validação, as melhores redes são selecionadas. A subamostra de validação engloba as últimas 52 observações dos dados da amostra (7.8.2015 - 29.7.2016). A subamostra de treinamento contém todos os outros valores. A distribuição das subamostras é: {344/52/52} do total de 448 unidades, ou em porcentagem {77%/11,5%/11,5%}, que é uma relação típica encontrada na literatura.

O BFGS-Broyden-Fletcher-Goldfarb-Shanno é usado como algoritmo de treinamento - um método iterativo de resolver problemas

problemas de otimização linear. Também é chamado de Quasi-Newton e é um algoritmo de treinamento de segunda ordem com convergência rápida. Ele exige mais memória e poder do processador do que o algoritmo básico de treinamento de descida de gradiente, como o método de retropropagação. Isso se deve ao cálculo e armazenamento da Matriz Hessiana.

No processo de construção, um total de trinta redes foram construídas e as cinco principais redes foram analisadas com mais detalhes. Os resultados (gráficos, desempenho e erros) são apresentados nos prazos de um mês, trimestre, semestre e ano.

4 RESULTADOS

Após a execução do procedimento de rede neural automatizada, novas redes foram construídas com as seguintes características médias (Tab. 2 e Tab. 3).

Tabela 2 Desempenho médio de 30, 5 e 1 das principais redes selecionadas

	Treinamento de avaliação de desempenho	Desempenho erro de validação	Desempenho erro de avaliação
Top 1	0,8420	0,9156	0,9405
AVR5	0,8620	0,8593	0,8990
AVR30	0,8545		

Tabela 3 Erro SOS médio de 30, 5 e 1 das principais redes selecionadas

	Erro de treinamento	Erro de validação	Erro de avaliação
Top 1	2370,71	2664,35	1707,10
AVR5	2419,60	1724,37	4155,06
AVR30	2556,50		

O desempenho na ferramenta SANN representa a correlação entre os dados de destino reais (entrada do modelo) e as previsões feitas pelo modelo (saída). Em teoria, o coeficiente de correlação é o valor entre \hat{y}_1 e \hat{y}_1 , onde 1 representa a correspondência perfeita. Para desempenho no conjunto de treinamento, correlação próxima ou igual a 1 nem sempre é um bom sinal. Isso pode indicar o problema de overfitting ou incorporação de variância e ruído aleatório no modelo, o que leva a capacidades de generalização mais fracas.

Aba. 3 mostra o desempenho agregado de todas as redes treinadas, as 5 melhores e a melhor. A primeira linha se refere à melhor rede, a segunda linha representa as médias das 5 principais redes e a última linha calcula as médias de todas as 30 redes construídas. A melhor rede selecionada não é a melhor rede (entre essas 30), mas é a melhor rede selecionada por desempenho de validação.

Pode-se observar que o desempenho na subamostra de treinamento possui valores um pouco maiores do que no desempenho de validação ou avaliação. O desempenho médio na subamostra de validação mostra o valor mais baixo, mas os resultados da avaliação são novamente mais altos. Isso significa que os modelos construídos em geral têm melhores pontuações de desempenho nos dados não vistos do que nos dados da amostra.

Horizontalmente (na terceira linha Tab. 3), a diferença nas médias de desempenho de todas as redes treinadas entre os grupos {treinamento/validação, treinamento/avaliação} é {8,9%, 4,2%}. No caso da rede melhor selecionada (primeira linha Tab. 3) esses números são {8,5%; 2,9%}. A diferença de 2,9% indica que a eficiência da amostra de validação para a melhor rede selecionada é pouco menos de 3% menor que a eficiência demonstrada no conjunto de treinamento. Além disso, não há grande desvio entre os valores no mesmo grupo.

Por exemplo, na coluna de validação, a diferença entre a melhor rede selecionada e as médias de todas

redes treinadas é de 0,87%. Para o caso do conjunto de avaliação essa diferença é de 1,8% e para o conjunto de treinamento de 0,47%.

A Soma dos Quadrados (SOS) é utilizada como função de erro e os resultados na Tab. 3 indicam o menor erro médio da amostra de validação, enquanto a amostra de avaliação possui o maior valor de erro. Novamente, as diferenças das médias de todas as redes (30), das 5 principais redes e da rede de melhor desempenho dentro dos mesmos grupos (treinamento, validação, avaliação) são de 7,3%, 4,1% e 59,3%. A maior diferença no valor do erro está no grupo de avaliação.

Indica que a distribuição de erros do grupo de avaliação é amplamente dispersa, apesar de resultados de desempenho muito estreitos (correlação). O erro SOS com desempenho semelhante (correlação) pode ser imaginado como uma distância entre linhas paralelas com padrões semelhantes, deslocadas umas das outras. Quanto maior a distância o erro é maior. A diferença do erro médio entre as subamostras pode ser devido à diferença das características dos dados incluídos nas amostras.

Tabela 4 Estatísticas de dados da amostra

Conjuntos de dados	ATX' (alvo)
Mínimo (Trem)	$\hat{y}565.564$
Máximo (Trem)	526.259
Média (Trem)	$\hat{y}1.622$
Desvio padrão (Trem)	212.288
Mínimo (Validação)	$\hat{y}286.967$
Máximo (Validação)	214.162
Média (Validação)	10.733
Desvio padrão (validação)	112.595
Mínimo (Avaliação)	96.570
Máximo (Avaliação)	755.578
Média (Avaliação)	434.396
Desvio padrão (Avaliação)	459.628

Isso pode ser observado na Tab. 4, e na Fig. 5, onde o intervalo entre os valores máximo e mínimo no conjunto de validação é o menor, ainda menor do que na amostra de treinamento, apesar do desempenho na amostra de treinamento apresentar melhores resultados. Além disso, a partir da Fig. 5 pode-se observar que o conjunto de treinamento tem a forma mais próxima da distribuição normal. É lógico, porque a amostra do trem tem mais de 7 vezes mais observações do que os dois conjuntos restantes.

O maior desvio está na amostra de avaliação. Em valores absolutos, o desvio mínimo está no grupo de validação, mas em percentuais (desvio em relação ao intervalo), o mínimo está no conjunto de treinamento. Os desvios padrão (em percentuais) dos grupos de treinamento, validação e avaliação em relação ao intervalo (max-min) são: {19,4%; 22,46%; 53,9%}. Essa distribuição desigual pode impor certos problemas, portanto, estatísticas de dados mais uniformes entre os grupos são preferíveis. O forte desvio no grupo de validação é devido ao aumento inesperado do preço ATX no período do último ano observado. Apesar disso, os modelos de RNA propostos se ajustam muito bem ao problema.

Os trinta modelos treinados e as séries de dados originais são plotados na Fig. 6. A seta sólida aponta para o grupo de redes (cluster) onde as séries originais estão localizadas.

Basicamente, pode-se notar que existem dois grandes clusters de rede e uma rede que se destaca das demais. A partir da Fig. 6 grupos de redes podem ser facilmente observados; o cluster inferior está marcado com seta pontilhada, o cluster superior (com linha contínua) e a rede de anomalia, acima de todos eles.

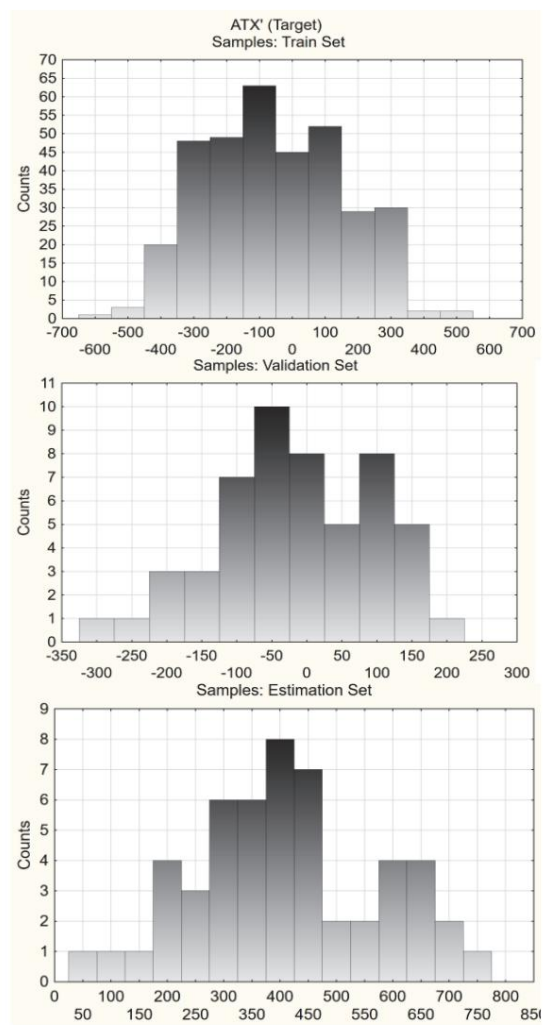


Figura 5 Histograma de treinamento, validação e as amostras de avaliação

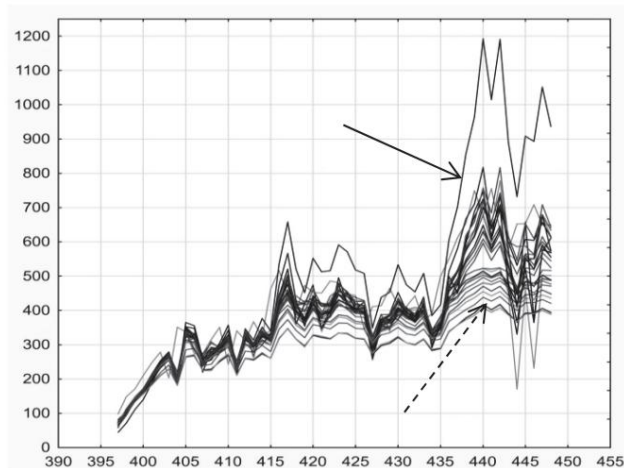


Figura 6 Previsões de todas as 30 redes treinadas

As primeiras 15 a 20 observações da amostra de avaliação estão praticamente alinhadas para todas as redes. Isso pode ser interpretado como um alto desempenho de previsão para o período de aproximadamente quatro primeiros meses do período observado, antes da dispersão em clusters. Independentemente da dispersão, padrões muito semelhantes (correlação) podem ser observados em todos os casos. Além disso, pode-se notar que o cluster superior, no qual a série temporal original está localizada (seta sólida) possui linhas muito mais densas.

Um método de referência de adequação da rede de classificação é o desempenho feito na subamostra de avaliação.

A SANN usa um procedimento de validação cruzada combinando vários números de parâmetros e sugerindo novas redes. Nesta primeira etapa, as configurações de rede, como pesos de conexão, são definidas. Junto com isso, o treinamento foi guiado usando a subamostra de validação para ajuste fino dos parâmetros definidos anteriormente. A amostra de validação é utilizada como seletor, para propor a(s) melhor(is) rede(s).

No final, a amostra de avaliação informa como um bom trabalho foi feito.

Após gerar trinta novos modelos, as 5 principais redes neurais foram listadas na Tab. 5. Todas as redes foram classificadas por coluna de desempenho de validação; assim, dados não vistos não estavam influenciando a seleção final do modelo.

Tabela 5 Desempenho e taxas de erro dos 5 principais modelos treinados.

Tabela 5 Desempenho e taxas de erro dos 5 principais modelos treinados.										
Configuração de ID	Desempenho do treinamento	Desempenho de validação	Avaliação de desempenho	Treinamento erro	Validação erro	Avaliação erro	Algoritmo de aprendizagem	Função de erro	Função de ativação (oculta)	Função de ativação (saída)
33 MLP 5-3-1	0,942748	0,861987	0,915584	2370,709	1707,102	2664,35	BFGS 127 SOS		Tanh	Exponencial
8 MLP 5-5-1	0,942014	0,859308	0,909845	2400,361	1716,701	2935,67	BFGS 61 SOS Exponencial	Exponencial		
11 MLP 5-3-1	0,940490	0,858781	0,918686	2462,696	1729,497	3627,76	BFGS 49 SOS Exponencial	Exponencial		
15 MLP 5-12-1	0,941676	0,858355	0,897422	2413,984	1709,037	5086,06	BFGS 29 SOS Exponencial			Logística
10 MLP 5-4-1	0,940783	0,857913	0,895508	2450,240	1759,531	6461,17	BFGS 36 SOS Exponencial			Tanh
28 MLP 5-4-1	0,938771	0,852333	0,928949	2531,055	1763,481	2487,30	BFGS 20 SOS 0,928949	2370,71	1707,10	Identidade
- Máx. / Mín.	0,942748	0,861987				2320,81	-	-	-	-

O desempenho da validação segue de perto os conjuntos de treinamento e avaliação. A última linha da Tab. 5 contém os melhores resultados gerais de todas as redes para cada categoria. O melhor valor do desempenho de validação é, obviamente, na primeira linha, porque toda a tabela é classificada por essa coluna. Um bom indicador é que a rede melhor validada é também a rede mais bem treinada, mas é mais provável que seja esperado, pois ambos os conjuntos estão incluídos no processo de construção da rede. Para o grupo de avaliação isso não é verdade. Os melhores modelos por desempenho de avaliação (0,9289) nem aparecem na lista das 5 principais redes. No conjunto das 25 redes restantes, o número daquelas com valor de desempenho de avaliação superior a 0,9 é 15, e três redes apresentam erro SOS de avaliação inferior a 2450. Esse fato indica um problema de seleção das melhores redes. A melhor rede selecionada não precisa ser a melhor rede do conjunto de treinamento. A rede mais precisa é marcada com ID28 (segunda última linha) e está em 25º lugar de 30 (classificada por desempenho de validação). Porém, o interessante é que a rede 28 não apresenta o menor erro de avaliação apesar da melhor nota de desempenho.

Na segunda linha da Tab. 5 são as configurações de rede. A estrutura de configuração tem três números conectados, que indicam três camadas na arquitetura perceptron multicamadas. O primeiro número representa o número de nós de entrada, o segundo o número de nós ocultos e, finalmente, o número 1 na última camada representa um neurônio de saída para uma previsão antecipada. No procedimento SANN de validação cruzada, o intervalo de nós ocultos é definido entre 2 e 12, mas quase todas as principais redes mostram que esse número é inferior ou igual a 5. Dos 5 principais resultados apenas uma rede possui a configuração com número maior de neurônios ocultos, neste caso 12. De acordo com o princípio da parcimônia [2], se as redes tiverem desempenho semelhante, aquela com design mais simples tem vantagem sobre a complexa. Além disso, o menor número de nós ocultos confirma que uma rede simples pode ser muito poderosa e superar as contrapartes mais complexas. Para as principais redes selecionadas, a função de ativação exponencial domina, mas o melhor modelo geral tem a função de ativação de identidade.

No próximo gráfico Fig. 7, as 5 principais redes são mostradas juntamente com a série original.

A linha ATX' da série temporal de entrada original é marcada com pontos pretos e a melhor rede selecionada é marcada com um

quadrado. O ATX' representa o índice de negociação austríaco após o processamento e preparação para o treinamento do modelo. Para calcular erros de modelos, devem ser feitas transições de volta para a série original (ATX). As discrepâncias no desempenho das 5 principais redes devem-se à classificação em diferentes clusters, mencionadas anteriormente. Das 5 principais redes selecionadas, três delas podem ser colocadas no cluster superior (ID 33; 8; 11) e duas no inferior (ID 15, 10).

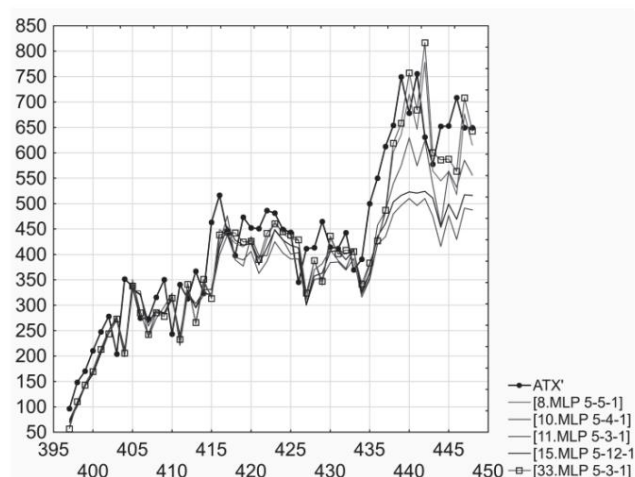


Figura 7 Previsões para as 5 principais ANNs

De acordo com a Figura 6, pode-se observar que aproximadamente metade dos modelos gerados são classificados no cluster superior e a outra metade no cluster inferior. Uma proporção semelhante é encontrada nas 5 melhores redes selecionadas (3/2). Embora a série ATX original domine no topo dos 5 modelos, existem várias redes em sua classe cujos valores estão em média acima da série original.

No entanto, não é incomum que a série original e a previsão do modelo sigam uma à outra (abaixo ou acima, dependendo se a tendência é alta ou baixa), o que é parcialmente causado pelo atraso. Isso pode ser visto em vários exemplos e pesquisas anteriores [5, 14, 15, 16]. Uma das razões é que o modelo que usa entrada de um passo para prever um passo em avançado e o processo de aprendizado da RNA pode ser explicado como equação diferencial de primeira ordem

$$y(t) = y(0) + \int_0^t \frac{dy}{dt} dt \quad (4)$$

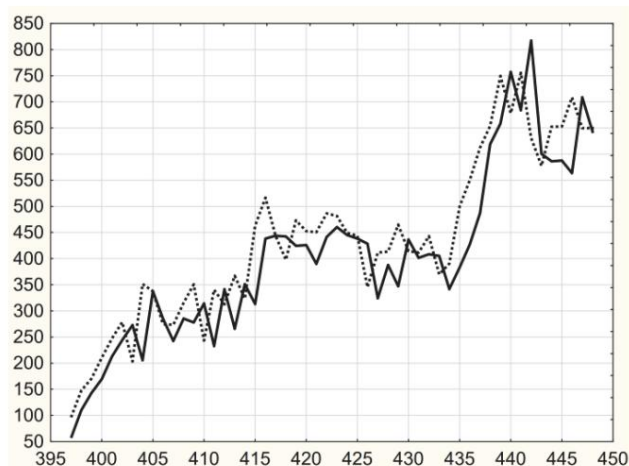


Figura 8 Resultados da previsão da melhor rede selecionada (por validação)

Finalmente, na Fig. 8, a comparação entre a série original e a melhor série de modelo selecionada (ID 33) é plotada. A série original ATX' foi desenhada com linha pontilhada e a série modelo com linha sólida. A linha da série de previsão segue de perto a linha de dados de destino. A Fig. 8 prevê o movimento do preço ATX nas próximas 52 semanas ou no período de um ano, com base na melhor rede selecionada (ID 33) via amostra de validação.

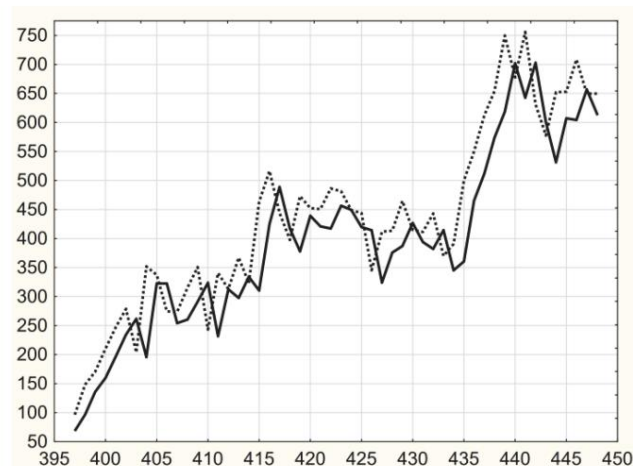


Figura 9 Resultados da previsão da melhor rede (por avaliação)

Por outro lado, a Fig. 9 compara as séries originais (linha pontilhada) com as previsões do melhor modelo (ID 28, linha completa) encontrado no conjunto de treinamento. Este modelo não seria levado em consideração, pois teve um desempenho de validação muito ruim. A partir da Fig. 9 pode ser visto que ambos os padrões de linhas são praticamente iguais, assim a correlação resulta no valor mais alto no conjunto (0,92849).

Tabela 6 Medidas de erro

Período de tempo	LOUCO	LOUCO	MSE	MSE	MAPA	MAPA
	ATX(28)	ATX(33)	ATX(28)	ATX(33)	ATX(28)	ATX(33)
Mês (4)	40,22	35,72	1714,36	1302,54	1.764%	1.571%
Trimestre (13)	50,11	42,98	3500,89	2948,09	2.110%	1.812%
Metade (26)	54,92	51,45	4453,87	3993,32	2.209%	2.070%
Ano (52)	58,38	58,62	4952,48	5308,66	2.128%	2.123%

Um bom indicador da adequação do modelo é o erro calculado em dados reais não processados, que mede o quanto as previsões estão próximas das séries de dados originais reais. A série ATX, juntamente com os resultados da previsão, devem ser traduzidos de volta ao espaço original, adicionando a tendência e a sazonalidade.

As medidas de erro que são tomadas em consideração são MAD (Mean Absolute Desvio), MSE (Mean Squared Error) e MAPE (Mean Absolute Percentage Error). Na Tab. 6 valores de erro são mostrados para o melhor modelo selecionado (ID 33 - desempenho de validação mais alto) e o modelo com o melhor desempenho de avaliação (ID 28). Os erros foram calculados para o período de um mês (4 observações), quartil (13), semestre (26) e ano (53 observações). Todos os tipos de erros aumentam seus valores com horizonte de tempo progressivo, o que é esperado. Embora o modelo 33 tenha desempenho inferior no conjunto de avaliação do que o modelo 28, quase todas as medidas de erro registram melhores resultados para o modelo 33, exceto para o caso do horizonte de um ano, onde os erros são ligeiramente a favor do modelo 33. o erro percentual é de 1,57% para o período de um mês do modelo 33.

5. CONCLUSÕES

Apesar da notável diferença de estrutura de dados entre a subamostra de validação e os dados dentro da amostra, os resultados gerais confirmam a alta precisão do uso do método de redes neurais artificiais. Este jornal mostra vários

combinações de redes neurais construídas e seus desempenhos.

Futuras pesquisas podem ser direcionadas na combinação da RNA com outro método, especialmente o linear. A modificação do número de nós de entrada e saída também pode ajudar a encontrar uma solução mais eficiente.

6 REFERÊNCIAS

- [1] Box, GE & Jenkins, GM (1976). Análise de séries temporais. Previsão e controle. *Na série Holden-Day no tempo*. *Análise de série, ed. revisada, San Francisco: Holden-Day.*
- [2] Clemen, RT (1989). Combinando previsões: uma revisão e bibliografia comentada. *Jornal Internacional de Previsão*, 5(4), 559-583. [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5)
- [3] Zhang, GP (Ed.). (2004). *Redes neurais nos negócios previsão*. IGI global. <https://doi.org/10.4018/978-1-59140-176-6>
- [4] Zhang, GP (2004). Previsão de negócios com artificial redes neurais: uma visão geral. Em *redes neurais em previsão de negócios*, 1-22. <https://doi.org/10.4018/978-1-59140-176-6.ch001>
- [5] Zhang, GP (2003). Previsão de séries temporais usando um híbrido ARIMA e modelo de rede neural. *Neurocomputação*, 50, 159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- [6] Haefke, C. & Helmenstein, C. (1996). Redes neurais no mercado de capitais: uma aplicação para previsão de índices. *Economia Computacional*, 9(1), 37-50. <https://doi.org/10.1007/BF00115690>
- [7] Schittenkopf, C., Dorffner, G., & Dockner, EJ (1998). Previsão de volatilidade com redes de densidade de mistura. *Dentro*

- Conferência Internacional sobre Redes Neurais Artificiais*, 929-934).
https://doi.org/10.1007/978-1-4471-1599-1_145
- [8] Haefke, C. & Helmenstein, C. (2002). Previsão de índice e seleção do modelo. *Sistemas Inteligentes em Contabilidade, Finanças & Management*, 11(2), 119-135.
<https://doi.org/10.1002/isaf.214>
- [9] Zhang, GP (2004). Uma combinação ARIMA e neural abordagem de rede para previsão de séries temporais. Em *Neural Redes em 213-225*. *Previsão de Negócios*, <https://doi.org/10.4018/978-1-59140-176-6.ch011>
- [10] Werbos, PJ (1990). Retropropagação ao longo do tempo: o que é faz e como fazer. *Anais do IEEE*, 78(10), 1550-1560. <https://doi.org/10.1109/5.58337>
- [11] Martinovic, M., Stoic, M., Duspara, M., Samardzic, I., & Stoic, A. (2016). Conversão algorítmica dos dados exibidos semanalmente para o nível mensal usando a planilha. *Procedia Engineering*, 149, 288-296.
<https://doi.org/10.1016/j.proeng.2016.06.669>
- [12] Martinović, M., Samardžić, I., & Stoić, A. (2016). Análise da curva de demanda do laptop por meio da pesquisa na internet. *Inovações, Tecnologia, Educação e Gestão / Anais da 4ª Conferência Internacional, ITEM*, 2016.
- [13] Martinović, M., Požega, Ž., & Crnković, B. (2017). Análise de Séries Temporais para o Par de Moedas Kuna Croata / Euro. *Ekonomskivjesnik/Econviews-Review of Contemporary Negócios, Empreendedorismo e Questões Econômicas*, 30(1), 37-49.
- [14] Li, LK, Pang, WK, Yu, WT e Troutt, MD (2004). Previsão de taxas de câmbio de curto prazo: uma abordagem neural recorrente abordagem de rede. Em *Redes Neurais nos Negócios Previsão*, 195-212.
<https://doi.org/10.4018/978-1-59140-176-6.ch010>
- [15] Brownlee, J. (2019). Previsão de séries temporais com LSTM Redes neurais recorrentes em Python com Keras. Recuperado de <https://machinelearningmastery.com/time-series-predict-lstm-recurrent-neural-networks-python-keras/>
- [16] Nayak, RK, Pavitra, SYH, Tripathy, R., & Prathyusha, Vol. K. (2019). Previsão de Preço de Câmbio de Moeda Estrangeira usando memória de longo prazo com K-Nearest Neighbor Método. *Revista Internacional de Engenharia e Tecnologia Avançada (IJEAT)*, 9 (2).
<https://doi.org/10.35940/ijeat.B3551.129219>

Informações de Contato:

Marko MARTINOVIĆ, mag.ing.rađ, v.pred.

(Autor correspondente)

Universidade de Slavonski Brod,

Trg Stjepana Miletija 12, 35000 Slavonski Brod, Croácia E-

mail: marko.martinovic@vusb.hr

Anica HUNJET, PhD, Professora Associada

University North,

Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croácia

E-mail: anica.hunjet@unin.hr

Ioan TURCIN, Assistente em pesquisa e ensino

CAMPUS 02 Universidade de Ciências Aplicadas,

Körblergasse 126, 8010 Graz, Áustria

E-mail: ioan.turcin@campus02.at