

Evaluation of FOREX Trading Strategies based in Random Forest and Support Vector Machines

A. D. R. Santuci, E. F. Sbruzzi, L. E. S. Araújo Filho and M. C. R. Leles

Abstract—The Foreign Exchange (Forex) is the largest market in the world and has a daily trading volume of approximately 3.2 trillion dollars. The price movements are influenced by many exogenous factors, thereby it is difficult to predict. As a result, modeling its movement would enable high profitable investment strategies. Aiming at setting up a model that helps the market practitioner make better decisions for trading on Forex, Machine Learning algorithms (Random Forest and SVM) are adopted in this work. Classic technical indicators, like moving averages, are used as features for these algorithms. In order to evaluate these approaches, several simulations were carried out on the pairs Euro/Dollar, Pound/Dollar, Dollar/Swiss Franc and Dollar/Japanese Yen, using three different metrics. For virtually all scenarios investigated, the proposed algorithms outperform the traditional technical indicators.

Index Terms—Financial Machine Learning, FOREX, Random Forest, SVM, Performance of Technical Trading Rules.

I. INTRODUÇÃO

O movimento dos preços dos ativos no mercado financeiro está sujeito a diversos fatores, como: eventos políticos, notícias econômicas, influência internacional, dentre outros. Uma maneira de se atuar nesse mercado é por meio da Análise Fundamentalista, que utiliza fatores econômicos para estimar os valores intrínsecos dos ativos financeiros. Por outro lado, a Análise Técnica (AT) se baseia no estudo da evolução temporal da série histórica de preços. Diversos tipos de estratégias de negociação baseadas na AT foram propostas na literatura [1]. Algumas se baseiam em indicadores “clássicos” como médias móveis, índice de força relativa, entre outros. Em contrapartida, alguns estudos têm combinado essas ferramentas tradicionais com técnicas de inteligência artificial [2–4] e de processamento de sinais [5, 6].

O mercado de câmbio, *Forex*, define os preços das moedas com base na oferta e demanda de uma determinada moeda [7]. Comparado com outros mercados, nele é possível abrir operações 24 horas por dia em 5 dias da semana. Além disso, ele movimentava um volume diário de 3,2 trilhões de Dólares, o que o torna o maior mercado financeiro existente [8].

A análise do mercado *Forex*, portanto, torna-se algo extremamente desafiador, uma vez que o movimento de preços de

um ativo não depende exclusivamente do seu histórico. Desse modo, os métodos estatísticos tradicionais usados apresentam dificuldades para lidar com a descontinuidade, não-linearidade e alta complexidade de séries temporais financeiras [1].

Com a evolução da computação nas últimas décadas, novos caminhos foram emergindo para auxiliar os investidores. Técnicas de *Machine Learning* (ML) tais como Redes Neurais Artificiais, *Support Vector Machines* (SVM) e *Random Forests* se mostram propensas a analisar esses modelos não-lineares e complexos que são dominantes no mercado financeiro [9]. Hu et al. [10] apresentam uma revisão recente de ferramentas de *Deep Learning* aplicadas ao mercado *Forex*, enquanto Fisichella and Garolla [11] avaliam se essas técnicas podem melhorar o desempenho de estratégias baseadas na AT.

Alguns resultados interessantes foram obtidos aplicando técnicas de *Machine Learning* (ML) no problema de modelagem e *trading* no mercado *Forex* [12, 13]. Theofilatos et al. [14] aplicam quatro técnicas de machine learning para esse problema: *K-nearest neighbor classifier* (KNN), *Naïve Bayesian classifier*, *Support Vector Machines* (SVM) e *Random Forests*. Essas técnicas foram comparadas com as técnicas da estratégia de Naïve e da estratégia de convergência e divergência de médias móveis (MACD) a fim de observar se poderiam obter resultados melhores nos dados históricos da taxa do Euro/Dólar (EUR/USD) de 2002 a 2010. Esses autores concluíram que os algoritmos *Random Forest* e SVM obtiveram os melhores resultados.

Com o objetivo de se avaliar ferramentas de auxílio ao investidor no *Forex*, aqui são utilizados dois algoritmos de ML para essa finalidade: *Random Forest* e SVM. Indicadores técnicos clássicos (como médias móveis) entre outros são usados como parâmetros (*features*) para esses algoritmos. Diversas simulações computacionais em quatro pares de câmbio [Euro/Dólar (EUR/USD), Libra Esterlina/Dólar (GBP/USD), Dólar/Franco Suíço (USD/CHF) e Dólar/Yene Japonês (USD/JPY)], são avaliadas a partir de três diferentes métricas. O desempenho desses algoritmos será comparado com alguns dos principais indicadores de AT.

Por fim, o código fonte utilizado nesse trabalho pode ser acessado em <https://github.com/armandosantuci/Avaliacao-de-estrategias-de-investimento-em-Forex-utilizando-Random-Forest-e-SVM>.

II. FUNDAMENTAÇÃO TEÓRICA

Nessa seção são apresentados os principais conceitos necessários para a fundamentação teórica do presente trabalho.

A. D. R. Santuci, Universidade Federal de São João del-Rei (UFSJ), Ouro Branco, Minas Gerais, Brasil, (e-mail: armandodiegoad@gmail.com).

E. F. Sbruzzi, professor na Divisão de Ciência da Computação (IEC) do Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, São Paulo, Brasil, (email: elton@ita.br).

L. E. S. Araújo Filho, doutorando no Instituto Tecnológico de Aeronáutica (ITA) área Sistemas e Controle, São José dos Campos, São Paulo, Brasil, (e-mail: lesaf@ita.br).

M. C. R. Leles, professor no Departamento de Tecnologia (DTECH) da Universidade Federal de São João del-Rei (UFSJ), Ouro Branco, Minas Gerais, Brasil, (e-mail: mleles@ufsj.edu.br).

A. Análise do Mercado Forex

Para construir uma estratégia para operar no mercado *Forex* é necessário fazer a análise de eventos e informações disponíveis. As análises podem ser divididas em dois tipos: análise técnica e análise fundamentalista. A análise fundamentalista requer um estudo sobre o cenário político-econômico referente a notícias como indicadores econômicos do país, tais como força da economia subjacente, discursos presidencialistas, política monetária implementada pelo banco central do país em questão, dentre outros. Para o caso do EUR/USD, a política monetária do *FED* (Reserva Federal dos EUA) e do *BCE* (Banco Central Europeu) costumam ser seus maiores influenciadores, bem como o *Nonfarm Payroll* que mede a variação do número de pessoas empregadas durante o último mês de todas as empresas não-agrícolas dos Estados Unidos. Este último indicador pode ser considerado a peça mais importante para oferecer a melhor visão geral da economia.

Em contrapartida, a análise técnica utiliza dos dados passados para supor e tentar prever o movimento futuro do preço. Nesta técnica, são utilizados dados oriundos das séries temporais dos preços dos ativos, permitindo a criação de indicadores técnicos.

B. Séries Temporais

Uma série temporal representa a observação ordenada de uma variável ao longo de uma escala de tempo, por exemplo dados diários, mensais, anuais, entre outros [15]. Os valores de uma série temporal podem ser utilizados para detectar padrões de comportamento, incluindo a movimentação da taxa de câmbio do *Forex*. Esses padrões podem ser detectados a partir de dados obtidos de seu gráfico e analisados com ferramentas estatísticas.

As séries temporais analisadas neste trabalho foram as taxas de câmbio diária dos pares EUR/USD, GBP/USD, USD/CHF e USD/JPY. As escolhas dos ativos em questão é devido aos seus volumes de mercado, pois são os mais negociados no mercado *Forex*. A Tabela (I) faz a relação dos nomes das paridades com as suas respectivas siglas.

TABELA I
PARIDADES ANALISADAS

Nome da Paridade	Sigla
Euro / Dólar	EUR/USD
Libra Esterlina / Dólar	GBP/USD
Dólar / Franco Suíço	USD/CHF
Dólar / Yene Japonês	USD/JPY

C. Indicadores Técnicos

Indicadores técnicos são as ferramentas mais utilizadas pelos investidores que optam por operar com a análise técnica no mercado. Eles ajudam a identificar a impulsão, as tendências, a volatilidade e vários outros aspectos dos ativos [16]. Além disso, são ótimas ferramentas que podem ser empregadas como parâmetros (*features*) em modelos que utilizam técnicas de *machine learning*.

Baseado em suas funções, os indicadores podem ser divididos em três grupos: indicadores de tendência, indicadores de volume (força) e indicadores de momento. Os parâmetros de configurações dos indicadores utilizados são os seguintes:

- *Média Móvel Simples (SMA)*
Função que calcula a média aritmética da série temporal com n dias em determinado período. Quanto maior o seu período, mais suave será sua curva;
- *Stop and Reverse (Parabólica SAR)*
Indica uma tendência predominante em um determinado movimento do preço. Sua direção muda de acordo com a tendência captada e seu comprimento depende da variação dos preços;
- *Average True Range (ATR)*
Indica uma medida de volatilidade com o valor médio do preço no período especificado. Fornece um intervalo de movimentação que o preço pode atingir, podendo assim indicar se o preço está no início ou final de uma tendência;
- *Commodity Channel Index (CCI)*
Mede o nível atual do preço em relação a um nível médio em uma determinada janela de tempo. Ele retorna valores positivos para leituras acima da média e valores negativos para leituras abaixo da média. Desse modo, indica um movimento forte do preço.

D. Algoritmos de Machine Learning

Henrique *et al.* [12] apresentam uma revisão recente de aplicações de algoritmos de ML em finanças, enquanto Goodell *et al.* [17] destacam as principais tendências e direções de pesquisa nessa linha. Dentre os diversos algoritmos de ML aplicados em finanças, a escolha dos métodos aqui utilizados se deu, principalmente, em função de sua ampla utilização (SVM) [13] e sua interpretabilidade (Random Forests) [18].

1) *Random Forests*: trata-se de um método de aprendizado supervisionado proposto por Ho [19] que opera construindo múltiplas árvores de decisão na qual cada árvore depende de um vetor com um valor aleatório independente e com a mesma distribuição para todas as árvores na floresta, onde cada árvore vota na classe mais popular referente ao vetor de entrada. É empregado em situações que requeiram reconhecimento de padrões e classificação pela análise de dados, sendo um modelo de aprendizado baseado em árvores de decisões que ajusta vários classificadores destas árvores em várias subamostras do conjunto de dados e usa a média para melhorar a precisão preditiva e o sobreajuste de controle. *Random Forests* também pode ser usado para problemas de regressão no qual são formadas árvores dependendo de um valor aleatório de forma que o preditor da árvore assuma valores numéricos em oposição aos rótulos da classe [20].

2) *Support Vector Machine: SVM*: é um modelo de aprendizagem de máquina utilizado para problemas de classificação e regressão. Foi originalmente definido para a classificação de classes de objetos lineares. Entretanto também pode ser usado para classes que não podem ser separadas por classificadores fixos. Nesse caso, as coordenadas dos objetos são mapeadas em um espaço futuro usando funções não-lineares.

O espaço futuro em que cada objeto é projetado é um espaço multidimensional onde as classes podem ser separadas com o classificador fixo, cuja dimensão depende do número de variáveis do conjunto de dados. Essa separação é feita por um hiperplano que é posicionado de forma a obter a maior distância entre os pontos extremos de cada classe, formando assim a definição de margem máxima. A função *kernel* RBF é utilizado no presente trabalho devido à sua eficiência em prover uma performance superior em problemas de classificação.

E. Modelo de Benchmark: MACD

As estratégias de *machine learning* presentes neste trabalho foram comparadas com a estratégia tradicional de convergência e divergência de médias móveis (MACD) conforme trabalhos correlatos [14][21]. Além disso, o MACD comumente possui resultados melhores quando utilizado para gráficos com tempos maiores como gráficos diários, semanais e mensais. Ele foi introduzido por Gerald Appel em 1970 e está na categoria de indicadores de tendência que mostram a relação entre preços e médias móveis.

Essa estratégia é bem simples: duas médias móveis exponenciais com períodos de 26 e 12 dias, respectivamente, são criadas e a posição é aberta caso haja interseção delas. Se a média móvel com menor período cruza debaixo para cima a média móvel maior abre-se uma posição de compra. Em contrapartida, se a média móvel com menor período cruza a maior vindo de cima para baixo abre-se uma posição de venda.

F. Métricas de Performance

O retorno financeiro gerado não pode ser considerado o principal fator na análise de eficiência de uma estratégia. Existe uma grande diferença entre o resultado obtido pelo sistema operacional quando é aplicado a dados históricos, considerando que seus parâmetros tenham sido otimizados utilizando esses dados; ou seja, o resultado que o mesmo vai produzir quando aplicado a dados reais, ou dados que não foram utilizados durante as simulações. Desse modo, o objetivo das métricas de performance é avaliar e/ou validar o modelo para aplicações em casos reais.

1) *Índice de Sharpe*: Proposto por Sharpe [22], é um indicador financeiro que avalia o desempenho de um investimento através da sua relação risco x retorno em comparação com investimentos de livre risco. Quanto maior o seu valor, melhor é o retorno encontrado pela estratégia analisada. Pode ser calculado por meio da Equação 1.

$$SR = \frac{\mu - r_f}{\sigma} \quad (1)$$

onde μ é a média do retorno esperado, r_f é a taxa de investimento livre de risco no período e σ é o desvio padrão da estratégia analisada, podendo ser entendido como a volatilidade do ativo em questão.

2) *Retorno Anualizado*: Consiste na conversão do retorno de qualquer período para o retorno anual. Na Equação (2), k representa o tempo do investimento em anos, $r_{acumulado}$ é o retorno da estratégia em todo o período e $r_{a.a}$ é o retorno anualizado do investimento.

$$r_{a.a} = (1 + r_{acumulado})^{\frac{1}{k}} \quad (2)$$

3) *Máximo Drawdown*: Um *drawdown* é a diferença entre o valor corrente de uma curva de capital e o máximo global dessa curva atingido anteriormente no tempo. É uma métrica que retorna em termos percentuais, a maior perda registrada no período de análise, indicando os riscos que o capital pode estar exposto na estratégia utilizada. Pode ser calculado conforme a Equação (3) onde P representa o máximo anterior do período e L o retorno atual.

$$MDD = \frac{P - L}{P} * 100 \quad (3)$$

4) *Índice de Calmar*: É considerado um indicador de desempenho e foi proposto por Young [23]. É uma função do retorno anual de um investimento pelo máximo *drawdown*. Quanto maior o seu valor, melhor é a relação retorno x risco do investimento. A Equação (4) mostra o cálculo do índice de Calmar, onde: $r_{a.a}$ é a taxa anualizada do retorno do ativo em análise; $r_{f.a.a}$ é o valor anualizado de um investimento livre de risco e MDD é o máximo *drawdown* para o período analisado.

$$Calmar = \frac{r_{a.a} - r_{f.a.a}}{MDD} \quad (4)$$

III. ABORDAGEM PROPOSTA

Nesta seção serão discutidos os métodos utilizados para realização deste trabalho.

A. Análise Exploratória de Dados

Para a aplicação do modelo de ML é necessário fazer uma reorganização dos dados e incluir a criação de parâmetros (*features*) que serão utilizados pelo modelo.

Como parte dos parâmetros utilizados, foi também inserido o cálculo de mudança percentual diária do preço de fechamento. Usou-se também as informações referentes ao corpo do *candlestick* diário, além de seus pavios superiores e inferiores. As configurações dos indicadores seguiram valores que são comumente utilizados em estratégias de negociação e podem ser vistos na Tabela (II).

TABELA II
CONFIGURAÇÃO DOS INDICADORES

Indicadores	Períodos
SMA	50 ; 20
ATR	50
CCI	20
SAR	-

Por ser um modelo para rastrear a tendência do movimento do preço do ativo, é necessário criar uma coluna nos dados de treinamento que represente a correspondência entre as características em cada amostra e prever uma saída que significará o possível movimento de baixa ou alta do preço a partir do preço de abertura do próximo dia. Sendo assim, uma coluna denominada y recebe os sinais de negociação conforme a Equação (4). Um vetor de sinais chamado y_pred

é preenchido com os sinais previstos pelo modelo e segue a mesma interpretação da coluna y .

$$\begin{cases} y = -1 & \text{se } \text{preço}_{t+1} < \text{preço}_t \rightarrow \text{Venda} \\ y = 1 & \text{se } \text{preço}_{t+1} > \text{preço}_t \rightarrow \text{Compra} \end{cases} \quad (5)$$

B. Aplicação do Modelo

Os dados foram divididos entre dados para treinamentos e dados para teste, sendo esse conjunto representado por uma matriz de números reais. Utilizou-se a razão de divisão de 80% dos dados para treinamento e 20% para testes. Para realizar o treinamento de toda uma série histórica foi necessário implementar o método *walk forward* de atualização do modelo.

1) *Método Walk Forward*: Para a implementação do método *walk forward* foi preciso dividir todo o conjunto de dados em subconjuntos, os quais 80% representam os dados de treinamento e os restantes 20% representam os dados de teste. Levando-se em conta os 252 dias úteis no calendário financeiro convencional, foram escolhidas amostras com 300 dias para cada passo de atualização: 240 dias para treinamento e os outros 60 dias para teste. Dessa forma todo o conjunto de dados é retreinado por meio do deslocamento do tamanho do conjunto designado para teste.

2) *Escolha de Parâmetros*: A matriz de parâmetros para o *Random Forests* pode ser visto na Tabela (III) e os parâmetros utilizados para o SVM podem ser visto na Tabela (IV). Para o SVM, os valores de C e γ foram passados com espaçamento exponencial [24]. Destaca-se que não é do conhecimento dos autores do presente trabalho uma metodologia amplamente aceita para definição desses parâmetros. Os valores dos parâmetros dos algoritmos de ML utilizados foram, portanto, escolhidos empiricamente. A cada passo de atualização do *walk forward*, um método de busca extensiva é executado e testado para cada valor de uma matriz de parâmetros, verificando quais combinações de parâmetros geram uma melhor acurácia naquele período de treinamento.

TABELA III
MATRIZ DE PARÂMETROS *Random Forests*

Parâmetro	Valores
bootstrap	[True]
max_depth	[50, 60, 70, 80, 90, 100, 110]
max_features	[2, 3]
min_samples_leaf	[3, 4, 5]
min_samples_split	[5, 6, 7, 8, 9, 10, 11]
n_estimators	[200, 300, 400, 500, 600]
random_state	[0, 42]

TABELA IV
MATRIZ DE PARÂMETROS SVM

Parâmetro	Valores
C	$[2^{-15}, \dots, 2^5]$
γ	$[2^{-15}, \dots, 2^3]$
kernel	[RBF]

3) *Execução e Retorno da Estratégia*: As previsões geradas pelo modelo foram alocadas no vetor y_{pred} e comparada com o valor real da coluna y . Foi simulado um saldo de mil dólares e caso a previsão estivesse certa somava-se os pontos referente ao corpo do *candlestick*, caso a previsão fosse errada, esses pontos eram subtraídos.

Por se tratar de uma estratégia fictícia, não foram considerados os valores de corretagens e demais taxas de negociações. Tais valores variam de corretora para corretora e podem mudar com o passar do tempo ou sofrer isenções de acordo com os valores e os volumes de *trades* envolvidos.

IV. RESULTADOS E DISCUSSÃO

A Tabela (V) mostra a divisão de tempo utilizada no modelo, esta corresponde ao período entre 01/01/2019 a 30/11/2020. As cotações históricas dos ativos foram obtidos *feed* histórico do site Dukas Copy (<https://www.dukascopy.com/swiss/english/marketwatch/>).

A amostra total dispõe de 600 dias, com subamostras de 300 dias nos quais 240 são utilizados para treino e os outros 60 dias para teste, que representa o intervalo de tempo em que as estratégias realizam as operações. Para fins de comparação, a estratégia MACD foi colocada para operar nos mesmos dias que as estratégias com *machine learning*.

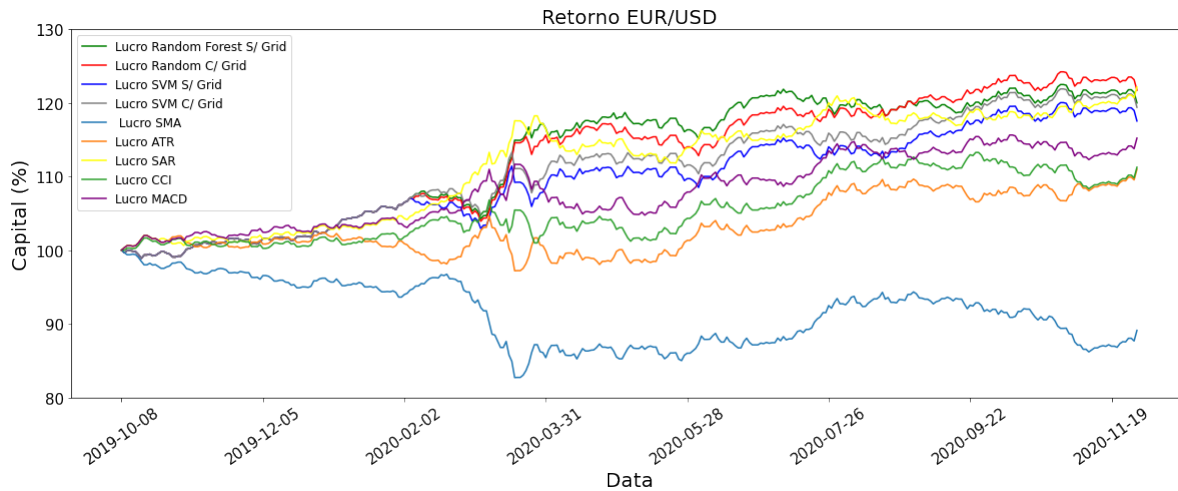
TABELA V
DIVISÃO DA AMOSTRA

Divisão da Amostra	Dias
Tamanho total da Amostra	600
Janela de Treinamento	240
Dias de Teste e validação	60
Total de Dias Posicionados	360

A Fig. 1 apresenta a evolução das curvas de capital e das respectivas ordens de mercado emitidas pelas estratégias avaliadas no par EUR/USD. A Fig. (1b) mostra a sinalização das ordens de compra e venda no par EUR/USD pelos algoritmos de *machine learning* utilizados, as ordens de compra e de venda são mostradas como valores positivos e negativos, respectivamente.

Para os pares EUR/USD e USD/JPY a busca extensiva na parametrização a cada passo de deslocamento da janela temporal fez a diferença. Em contrapartida, com o par GBP/USD, o *Random Forests* isso não ocorreu. No par USD/CHF o algoritmo com SVM conseguiu se sobressair. É interessante observar também que os algoritmos com *Random Forests* obtiveram os piores resultados quando comparados até mesmo com a estratégia do MACD.

As Tabelas (VI) e (VII) mostram os índices de de Sharpe e de Calmar, respectivamente. Esses índices refletem o quão viável é o investimento ante o risco que ele apresenta ao investidor. Os pares EUR/USD e GBP/USD tiveram os melhores índices com as estratégias que utilizavam SAR, seguidas pelas estratégias com *Random Forests*, SVM e MACD, respectivamente. Para o par USD/CHF, os melhores índices foram obtidos com as estratégias que utilizam SVM, seguidas pela estratégia do MACD, do SAR e as estratégias com *Random*



(a) Evolução das estratégias: Par EUR/USD.



(b) Ordens de mercado: Par EUR/USD. Os pontos foram deslocados no eixo vertical para facilitar a visualização.

Fig. 1. Evolução das curvas de capital e das respectivas ordens de mercado emitidas pelas estratégias avaliadas.

Forests. No caso do USD/JPY, os melhores resultados foram com *Random Forests*, seguida pela estratégia com SVM, SAR e MACD. Nesse caso, pode-se avaliar o máximo *drawdown* de cada modelo durante o período de análise e negociação. Nos pares EUR/USD e USD/JPY, a estratégia escolhida como *benchmark* (MACD) teve a maior perda quando comparada aos modelos propostos e aos modelos que utilizaram SAR. No par GBP/USD, as perdas foram menores quando utilizados os algoritmos com SAR, *Random Forests* e MACD. Com exceção do par GBP/USD que apresentou a menor perda com o algoritmo que utiliza o indicador SAR, as perdas foram menores quando utilizado pelo menos uma das estratégias com *machine learning* nos outros três pares analisados.

A Tabela (VIII) mostra os resultados do retorno anualizado de cada estratégia em cada ativo analisado.

Os melhores resultados obtidos por Theofilatos et al. [14] ao modelar o par EUR/USD ocorreram através de algoritmos que utilizam SVM e *Random Forests*. No presente trabalho, o maior retorno anualizado foi obtido pela estratégia utilizando o SAR com 9,67%, seguidos pelos *Random Forests* com 9,51%,

SVM com 8,63% e MACD com 5,96%. Com relações às maiores perdas, a ordem foi SMA com -17,30%, ATR com -7,14%, MACD com -6,58%, -4,84% com SVM e, finalmente, *Random Forests* com -3,71%.

De uma forma geral, os resultados do presente trabalho reforçam que os algoritmos apresentados têm um desempenho superior à estratégia escolhida como *benchmark*.

Finalmente, é necessário levar em conta que o modelo aqui apresentado não considera qualquer custo envolvido nas negociações (taxas, impostos, spreads e corretagens), então os lucros apresentados são fictícios. Segundo Olivares-Nadal and DeMiguel [25], pode-se considerar cerca de 0,1% a 0,5% de taxa de negociação por valor transacionado. Desse modo, uma estratégia que tem mais gatilhos e gera um maior número de operações pode facilmente se tornar menos rentável do que outra que dê menos gatilhos de entrada.

V. CONCLUSÃO E TRABALHOS FUTUROS

A aplicação ferramenta para auxílio nas negociações no mercado de câmbio é algo desafiador. No presente trabalho

TABELA VI
ÍNDICE DE SHARPE

Estratégia	EUR/USD	GBP/USD	USD/CHF	USD/JPY
<i>Random Forests</i> com busca extensiva	2,25	2,57	1,23	3,24
<i>Random Forests</i> sem busca extensiva	2,10	2,71	0,90	1,79
SVM com busca extensiva	2,00	2,27	2,57	2,59
SVM sem busca extensiva	1,80	2,35	1,89	2,54
MACD	1,25	1,92	1,15	0,75
SMA	-0,96	0,97	-0,79	-0,28
ATR	1,08	-0,64	-1,15	0,26
SAR	2,28	2,69	1,75	2,43
CCI	2,43	0,65	1,05	-0,27

TABELA VII
ÍNDICE DE CALMAR

Estratégia	EUR/USD	GBP/USD	USD/CHF	USD/JPY
<i>Random Forests</i> com busca extensiva	3,79	4,67	1,30	7,90
<i>Random Forests</i> sem busca extensiva	3,75	4,70	0,73	3,87
SVM com busca extensiva	2,63	3,53	5,57	5,13
SVM sem busca extensiva	2,36	3,65	7,91	5,05
MACD	1,32	3,21	2,17	0,86
SMA	-0,45	0,70	-0,40	-0,20
ATR	1,06	-0,45	-0,70	0,26
SAR	2,64	5,78	1,98	6,25
CCI	1,42	0,70	1,08	-0,27

foram utilizados dois algoritmos de *machine learning* para essa finalidade. Foram realizadas diversas simulações computacionais em quatro pares de câmbio (EUR/USD, GBP/USD, USD/CHF e USD/JPY), usando três diferentes métricas de desempenho. Na maioria dos cenários investigados, esses modelos obtiveram resultados superiores quando comparados a indicadores técnicos tradicionais, se mostrando, portanto, ferramentas que podem auxiliar o investidor em suas operações nesse mercado.

Para trabalhos futuros, pode ser considerado a adaptação do modelo para escalas de tempos menores, *intraday*. Isso possibilitaria o estudo da aplicação de técnicas de alta frequência e a implementação em tempo real. A adição dos custos de

TABELA VIII
RETORNO ANUALIZADO

Estratégia	EUR/USD	GBP/USD	USD/CHF	USD/JPY
<i>Random Forests</i> com busca extensiva	9,51%	17,00%	4,78%	13,84%
<i>Random Forests</i> sem busca extensiva	8,86%	17,72%	3,52%	8,58%
SVM com busca extensiva	8,63%	15,14%	3,52%	12,12%
SVM sem busca extensiva	7,91%	15,63%	9,34%	11,89%
MACD	5,96%	13,68%	7,09%	3,88%
SMA	-5,70%	7,20%	-3,77%	-1,61%
ATR	5,24%	-6,00%	-5,35%	1,45%
SAR	9,67%	17,74%	6,63%	11,09%
CCI	5,35%	5,41%	4,21%	-1,49%

corretagem nas simulações seria outro caminho de pesquisa. Outra linha que vem tendo bastante atenção utiliza ferramentas de *Deep Learning*. Finalmente, a inclusão de fatores externos como inflação e taxas de juros poderia enriquecer a abordagem aqui proposta.

AGRADECIMENTOS

Esse trabalho foi parcialmente financiado pela Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG, Processo APQ-00913-19).

REFERÊNCIAS

- [1] R. W. Colby, *The Encyclopedia of Technical Market Indicators*. McGraw-Hill Professional, 2012.
- [2] M. C. R. Leles, M. V. L. Pereira, R. A. Iquiapaza, E. F. Sbruzzi, and C. L. N. Júnior, "Evaluation of technical analysis trading rules in a artificial stock market environment," *IEEE Latin America Transactions*, vol. 18, no. 10, pp. 1707–1714, 2020.
- [3] M. C. R. Leles, L. A. Mozelli, E. F. Sbruzzi, C. L. Nascimento Júnior, and H. N. Guimarães, "A multicriteria trading system based on singular spectrum analysis trading rules," *IEEE Systems Journal*, 2019.
- [4] M. C. R. Leles, E. F. Sbruzzi, C. L. Nascimento Júnior, and J. M. P. Oliveira, "Trading Switching Setup Based on Reinforcement Learning Applied to a Multiagent System Simulation of Financial Markets," in *2019 Annual IEEE International Systems Conference*. IEEE, 2019, pp. 1–8.
- [5] M. C. R. Leles, L. A. Mozelli, C. L. Nascimento Júnior, E. F. Sbruzzi, and H. N. Guimarães, "Study on Singular Spectrum Analysis as a new technical oscillator for trading rules design," *Fluctuations and Noise Letters*, vol. 17, no. 4, p. 1850034, 2018.
- [6] M. C. R. Leles, L. A. Mozelli, and H. N. Guimarães, "New trend-following indicator: Using SSA to design

- trading rules,” *Fluctuation and Noise Letters*, vol. 16, no. 2, p. 1750016, 2017.
- [7] M. Amiri, M. Zandieh, B. Vahdani, R. Soltani, and V. Roshanaei, “An integrated eigenvector–dea–topsis methodology for portfolio risk evaluation in the forex spot market,” *Expert systems with applications*, vol. 37, no. 1, pp. 509–516, 2010.
- [8] M. Cerrato, N. Sarantis, and A. Saunders, “An investigation of customer order flow in the foreign exchange market,” *Journal of Banking & Finance*, vol. 35, no. 8, pp. 1892–1906, 2011.
- [9] B. Kaczorowski, M. Kleina, M. M. Marques, and W. Silva, “Artificial intelligence and the multivariate approach in predictive analysis of the small cap index of the brazilian stock exchange,” *IEEE Latin America Transactions*, vol. 19, no. 11, pp. 1924–1932, 2021.
- [10] Z. Hu, Y. Zhao, and M. Khushi, “A survey of forex and stock price prediction using deep learning,” *Applied System Innovation*, vol. 4, no. 1, p. 9, 2021.
- [11] M. Fisichella and F. Garolla, “Can deep learning improve technical analysis of forex data to predict future price movements?” *IEEE Access*, vol. 9, pp. 153 083–153 101, 2021.
- [12] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Literature review: Machine learning techniques applied to financial market prediction,” *Expert Systems with Applications*, vol. 124, pp. 226–251, 2019.
- [13] M. Islam, E. Hossain, A. Rahman, M. S. Hossain, K. Andersson *et al.*, “A review on recent advancements in forex currency prediction,” *Algorithms*, vol. 13, no. 8, p. 186, 2020.
- [14] K. Theofilatos, S. Likothanassis, and A. Karathanasopoulos, “Modeling and trading the eur/usd exchange rate using machine learning techniques,” *Engineering, Technology & Applied Science Research*, vol. 2, no. 5, pp. 269–272, 2012.
- [15] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [16] F. Lemos, *Análise Técnica dos mercados financeiros: um guia completo e definitivo dos métodos de negociação de ativos*. São Paulo, Saraiva Educação, 2015., 2015.
- [17] J. W. Goodell, S. Kumar, W. M. Lim, and D. Pattnaik, “Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis,” *Journal of Behavioral and Experimental Finance*, vol. 32, p. 100577, 2021.
- [18] X. Man and E. P. Chan, “The best way to select features? comparing mda, lime, and shap,” *The Journal of Financial Data Science*, vol. 3, no. 1, pp. 127–139, 2021.
- [19] T. K. Ho, “Random decision forests,” in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.
- [20] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [21] C. L. Dunis, J. Laws, and G. Sermpinis, “Modelling and trading the eur/usd exchange rate at the ecb fixing,” *The European Journal of Finance*, vol. 16, no. 6, pp. 541–560, 2010.
- [22] W. F. Sharpe, “Mutual fund performance,” *The Journal of business*, vol. 39, no. 1, pp. 119–138, 1966.
- [23] T. W. Young, “Calmar ratio: A smoother tool,” *Futures*, vol. 20, no. 1, p. 40, 1991.
- [24] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, “A practical guide to support vector classification,” 2003.
- [25] A. V. Olivares-Nadal and V. DeMiguel, “Technical note—a robust perspective on transaction costs in portfolio optimization,” *Operations Research*, vol. 66, no. 3, pp. 733–739, 2018.



Armando D. R. Santuci Bachelor's Degree in Mechatronics Engineering from Universidade Federal de São João del Rei (UFSJ) and Technische Universität Ilmenau (TU Ilmenau). While at undergraduate level he worked at a project to build and parametrize a 3D print under the guidance of Dr. Bruno Nazário Coelho. He works with research in Data Science area applying machine learning and quantitative analysis at Forex market.



Elton F. Sbruzzi Sbruzzi PhD em Finanças Computacionais pela *University of Essex*, UK. Mestre em Economia pela Universidade Federal do Rio Grande do Sul. Graduado em Economia pela Universidade Estadual de Campinas. Professor na Divisão de Ciência da Computação do Instituto Tecnológico de Aeronáutica. Desenvolve pesquisa interdisciplinar abrangendo a aplicação de Ciência de Dados e Inteligência Artificial em Finanças e Investimentos.



Luiz E. S. Araújo Filho Master of Science from the Instituto Tecnológico de Aeronáutica (ITA) in Electronics and Computer Engineering, Systems and Control field. Bachelor's Degree in Electrical Engineering from the Universidade Federal do Maranhão (UFMA), while at the undergraduate level he coordinated the Mobile Robotics and Wireless Communication Laboratory (LRC) at UFMA under the guidance of Prof. Dr. Luciano Buonocore. He currently participates in the Laboratory of Intelligent Machines at ITA as a doctoral student under the supervision of Prof. Dr. Cairo L. Nascimento Júnior. He works with research and development in the areas of mobile robotics, multi-agent systems, modeling and control of dynamic systems, computer vision and artificial intelligence.



Michel C. R. Leles received the Ph.D. degree in signals and systems from the electrical engineering graduate program, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. He is an Associate Professor with the Department of Technology at Federal University of São João del-Rei, Brazil, since 2010. His main research interests include data science in general and, particularly, its main applications to digital signal processing, timeseries analysis, and computational finance.