

Using Machine Learning to Prevent Losses in the Brazilian Stock Market During the Covid-19 Pandemic

Arthur Emanuel de Oliveira Carosia

Abstract— The Covid-19 Pandemic caused unprecedented changes in our society, going from human behavior modification areas to stock market crashes around the world. In fact, while the virus spread among several countries, investors watched great stock market losses during the period, especially in Brazil, in which the Ibovespa index presented a fall of almost 50% in its value. Thus, there is a need to investigate strategies to, at least, mitigate the losses during a stock market crisis, and, recently, Machine Learning techniques have played a key role in this process. Therefore, this work aims to propose an investigation of Machine Learning techniques in order to prevent stock market losses in the Brazilian stock market during the Covid-19 pandemic. We study commonly used algorithms in the financial area: Linear Regression, Support Vector Machines, Random Forest, XGBoost, Multilayer Perceptron, and an Ensemble composed by the combination of all of the mentioned models, all of them fed with historical stock prices and technical indicators. Our results show that, when properly tuned, some of the Machine Learning models could even bring a little profit during the Covid-19 pandemic, and, finally, we also present some guidelines for investors' choice when considering investing in a future market crisis.

Index Terms— Stock Market, Machine Learning, Covid-19.

I. INTRODUÇÃO

A pandemia do novo Coronavírus (Covid-19) teve seu começo em 2019 na China e rapidamente se espalhou pelo mundo, causando uma grande crise global na saúde, diretamente, e na economia, indiretamente. Este foi um período sem precedentes na história, levando à mudança de muitos aspectos da vida em sociedade: uso de máscaras, distanciamento social, além de *lockdowns* para manter a população em casa, dentre outros [1, 2].

Desde a chegada do vírus aos demais países do mundo, um dos elementos da sociedade que mais sentiu o efeito de tais medidas foi o mercado financeiro, apresentando quedas bruscas durante o período pandêmico. Por exemplo, no mercado americano¹, o índice Dow Jones apresentou quedas de até 37%, enquanto que o índice S&P 500 apresentou queda de 34%. Em especial, observou-se no mercado de ações brasileiro² queda de até 44% de seu principal índice, o Ibovespa.

Carosia, A. E. O. is with the Federal Institute of Sao Paulo (IFSP), São João da Boa Vista - SP. Email: arthuremanuel.carosia@ifsp.edu.br.

¹ S&P and DOW JONES INDICES - S&P GLOBAL. <https://tinyurl.com/42e6mtm>

² BMFBovespa. Cotações do índice Ibovespa. <https://tinyurl.com/28py92td>

De fato, conforme apontado pela Hipótese de Mercado Adaptativo [3], o mercado financeiro não é composto apenas por investidores racionais, de modo que elementos emocionais - tais como, aversão à perda, risco, ansiedade, dentre outros - podem afetar o modo como investidores alocam o seu patrimônio no mercado. Nesse sentido, em um estudo realizado pela Fundação Oswaldo Cruz³, verificou-se que o período da pandemia de Covid-19 ocasionou um aumento em casos de ansiedade, depressão e tristeza entre os entrevistados, além de mostrar também que pelo menos 50% das famílias tiveram redução da renda. Tendo isso em vista, as crises observadas no mercado de ações, em especial no mercado de ações brasileiro, refletiram, ao menos em parte, o estado emocional da população no período.

Dessa forma, é muito importante para o investidor utilizar-se de recursos que o auxiliem no instante da realização de uma operação frente ao grande número de informações proveniente do mercado financeiro. Dentre estas ferramentas, destacam-se tanto a Análise Técnica quanto a Análise Fundamentalista [4]. Enquanto a Análise Técnica se baseia no comportamento passado das ações esperando que se repitam no futuro, a Análise Fundamentalista é baseada na compreensão de fatores macroeconômicos que demonstrem ao investidor a saúde financeira de uma determinada empresa.

Mais recentemente, a literatura voltada à previsão de mercado tem se baseado em algoritmos de Aprendizado de Máquina, os quais procuram se ajustar a padrões contidos no conjunto de dados utilizados para treiná-los para, posteriormente, realizar previsões a partir de dados desconhecidos pelo modelo. Os algoritmos de Aprendizado de Máquina têm se mostrado mais eficientes do que métodos tradicionais de previsão de mercado [5] e, por isso, são o foco deste trabalho.

Nesse sentido, o desenvolvimento de técnicas que possam, ao menos, mitigar os danos causados pela pandemia e seus efeitos no mercado de ações, é essencial para que investidores possam manter o seu patrimônio em momentos de crise. O objetivo deste trabalho, portanto, é propor a análise de algoritmos de Aprendizado de Máquina comumente utilizados em trabalhos de previsão do mercado de ações, aplicados ao mercado brasileiro no período da pandemia de Covid-19.

³ Convid - Pesquisa de Comportamentos. <https://convid.fiocruz.br/>.

A principal contribuição deste estudo é sugerir ao investidor quais seriam os modelos mais adequados para momentos de grande queda no mercado. Além disso, este trabalho também contribui considerando como estudo de caso o mercado de ações brasileiro, que carece de estudos nesta área [6].

O restante deste trabalho está organizado como descrito a seguir. A Seção II aborda a fundamentação teórica e revisão da literatura, a Seção III apresenta a metodologia do presente estudo, a Seção IV apresenta os resultados obtidos em nossos experimentos e, por fim, a Seção V conclui o presente trabalho.

II. FUNDAMENTAÇÃO TEÓRICA

A previsão do mercado financeiro é uma tarefa muito estudada na literatura acadêmica [6]. Nesse sentido, muitos trabalhos e métodos foram sendo desenvolvidos e, na última década, o principal destaque tem sido o uso de técnicas de Aprendizado de Máquina e da área de Ciência de Dados [5].

Esta Seção apresenta a fundamentação teórica necessária para o desenvolvimento deste trabalho e está organizada como descrito a seguir. A subseção A apresenta os algoritmos de Aprendizado de Máquina selecionados para este trabalho e também uma descrição geral de seu funcionamento. A subseção B apresenta os trabalhos correlatos ao presente trabalho, ou seja, estudos que tratem do tema de previsão do mercado de ações com o uso de algoritmos de Aprendizado de Máquina, bem como também apresenta os diferenciais do presente trabalho perante aos seus correlatos.

A. Algoritmos de Aprendizado de Máquina

Existem vários algoritmos de Aprendizado de Máquina que são utilizados na literatura de previsão de mercado financeiro [7]. É possível perceber que, dentre os mais utilizados, destacam-se: *Linear Regression*, *Support Vector Machines*, *Random Forest*, *XGBoost*, *Multilayer Perceptron*. Além disso, neste trabalho também foi considerada a técnica de *Ensemble*. O funcionamento geral de cada um deles é discutido a seguir.

Linear Regression: técnica baseada na criação de um modelo linear com o objetivo de minimizar a soma do resíduo dos quadrados entre os elementos que são observados no conjunto de dados e os elementos que são previstos pela aproximação linear [8].

Support Vector Machines: técnica cujo funcionamento consiste em maximizar a margem de separação de classes por meio de vetores auxiliares. Essa técnica ainda apresenta a possibilidade do uso de funções de *kernel* para dados não lineares, facilitando encontrar o limite da separação entre classes. Os tipos de *kernel* mais comuns e que serão utilizados neste trabalho são: linear (L), polinomial (P) e radial (RBF). Além disso, vale a pena destacar que a técnica *Support Vector Machines* é utilizada para tarefas de classificação. Neste trabalho, utilizamos a sua adaptação para tarefas de regressão, denominada *Support Vector Regression* [9].

Random Forest: técnica baseada em árvore que funciona com base no mapeamento dos valores de entrada em regras que são encadeadas na forma de ramos (i.e., caminhos). Assim, o algoritmo *Random Forest* é um tipo de algoritmo baseado em árvore com o uso da técnica chamada *bagging*, em

que cada árvore em uma “floresta” de árvores é treinada independentemente [10].

XGBoost: técnica baseada em árvore, que consiste na versão mais recente do algoritmo *Gradient Boosting*, apresentando melhorias para lidar com dados esparsos e recursos de regularização dos dados [11].

Multilayer Perceptron: técnica que consiste em uma Rede Neural Artificial em que cada neurônio, denominado *perceptron*, é seguido por uma função de ativação e cujo agrupamento é realizado em camadas. Cada neurônio recebe impulsos de toda a camada anterior e propaga um novo impulso com base em sua função de ativação durante o treinamento [12].

Ensemble: técnica baseada em combinar os resultados de vários algoritmos de Aprendizado de Máquina com o objetivo de melhorar as previsões. Uma abordagem possível para tarefa de regressão é retornar os valores médios previstos, considerando o resultado de cada modelo. Tal abordagem pode ser útil para equilibrar as fraquezas individuais que eventualmente cada modelo possa apresentar [6].

B. Trabalhos Correlatos

Muitos trabalhos utilizam métodos de Aprendizado de Máquina para prever o mercado de ações. Duas revisões amplas de diversos trabalhos da literatura podem ser encontradas em [6] e em [13], que destacam que os trabalhos mais comuns da área normalmente utilizam *Support Vector Machines* ou redes neurais artificiais, como a rede *Multilayer Perceptron*, ambos utilizados neste trabalho. A seguir, são detalhados os trabalhos mais relevantes ao presente estudo.

O trabalho apresentado em [5] utilizou uma rede neural artificial do tipo *Multilayer Perceptron* para previsão do mercado de ações americano, mostrando eficiência de retorno em até 80% ao ano. O trabalho [5] apresentou uma estratégia de investimento que considera também o seu risco, à qual foi baseada nos experimentos deste trabalho.

Em [7], há um comparativo de diversos algoritmos baseados em Aprendizado de Máquina (como *Support Vector Machines*, *Random Forest*, *XGBoost*, *Multilayer Perceptron* e outros) aplicados ao mercado brasileiro com o uso de notícias para prever quedas do mercado. Seus resultados mostraram que, durante o período de 2017 a 2018, há um relacionamento significativo entre notícias e as mudanças de mercado no Brasil.

Além disso, o trabalho [14] utilizou a rede neural *Multilayer Perceptron* para previsão do mercado de criptomoedas, em especial o *Bitcoin*, durante o período de 2017 a 2021. Seus resultados mostraram que a redução da porcentagem de erro médio foi significativa, ainda que em momentos críticos (isto é, de grande queda) para a moeda.

O trabalho [15], assim como este estudo, apresentou um comparativo de seis algoritmos de Aprendizado de Máquina para previsão do índice S&P 500. Os resultados mostram que não existe uma técnica de Aprendizado de Máquina que possa ser adotada para prever as tendências de um determinado preço de ação em todas as situações consideradas. No entanto, o trabalho destacou que, dentre as técnicas testadas, aquelas que foram baseadas em redes neurais artificiais (como a *Multilayer Perceptron*), são as técnicas mais promissoras até então.

Por fim, os trabalhos [16] e [17] utilizaram *Support Vector Machines* para previsão dos mercados da Indonésia e dos Estados Unidos, respectivamente. Ambos trabalhos mostraram que a acurácia do modelo foi capaz de superar os modelos e estratégias *baselines*.

Em relação aos trabalhos considerados, o presente estudo apresenta os seguintes diferenciais:

- Comparar diferentes algoritmos de Aprendizado de Máquina, para dar um panorama geral ao investidor sobre qual modelo seria o mais adequado para momentos de crise do mercado financeiro global.
- Considerar o mercado de ações brasileiro, que possui poucas publicações na área, que são mais voltadas ao mercado asiático e americano [6].

III. METODOLOGIA

Este trabalho tem como objetivo prever, por meio de uma de algoritmos de Aprendizado de Máquina, o mercado de ações brasileiro durante o período da pandemia de Covid-19. A metodologia proposta para este fim é apresentada na Fig. 1. Primeiramente, os valores de ações, juntamente com indicadores técnicos (i.e., cálculos matemáticos realizados a partir dos valores das ações com o objetivo de identificar tendências no mercado) são utilizados como entrada para o treinamento de modelos de Aprendizado de Máquina comumente utilizados na área de previsão de mercado de ações. Em seguida, há a seleção dos melhores modelos e seus respectivos parâmetros com base em medidas de erro de treinamento. Por fim, a partir de cada modelo selecionado na etapa anterior, é realizada uma simulação de investimento para analisar a qualidade das previsões na prática. As subseções a seguir aprofundam o detalhamento de cada uma das etapas da metodologia deste trabalho.

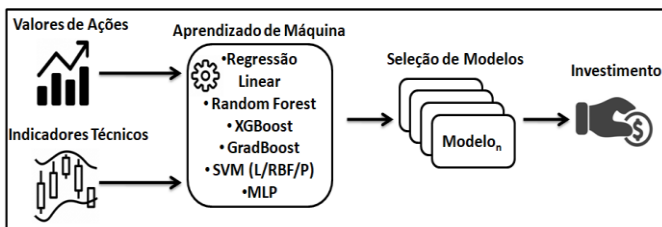


Fig. 1. Metodologia de previsão do mercado de ações.

A. Conjunto de Dados

O conjunto de dados utilizado nos experimentos consiste em cotações diárias do índice Bovespa (Ibovespa), que representa os papéis mais negociados na B3 (Brasil Bolsa Balcão), a bolsa de valores do Brasil. Os dados utilizados compreendem o período de 01-01-2015 até 23-03-2020, dia que corresponde ao menor valor de mercado do Ibovespa durante a pandemia de Covid-19. A Fig. 2 ilustra o comportamento do mercado de ações brasileiro no período correspondente à crise de mercado do ano de 2020. A distribuição dos dados utilizados nos experimentos é apresentada na Tabela 1. Os dados de cotações são compostos pelas seguintes características: valor de abertura, valor máximo, valor mínimo, valor de fechamento, além do volume de transações.

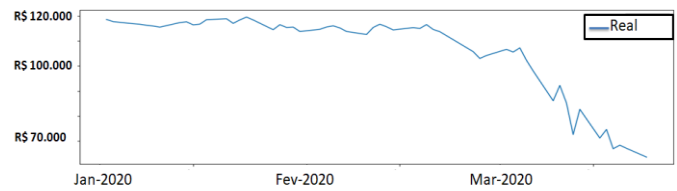


Fig. 2. Período de queda durante a pandemia de Covid-19 do índice Ibovespa, de 01-01-2020 até o dia 23-03-2020.

TABELA I
DISTRIBUIÇÃO DOS DADOS UTILIZADOS NOS EXPERIMENTOS

	Abertura	Máxima	Mínima	Fecha- mento	Volume
Itens	1483	1483	1483	1483	1483
Média	75311,83	76077,37	74583,02	75357,40	4755610, 20
Dev.					
Pad.	21410,56	21560,69	21261,03	21430,00	2645519, 50
Min.	37501	38031	37046	37497	0
25%	56303	56903	55742	56428	3197100
50%	74543	75260	73851	74553	3888400
75%	95376	96327	94273,5	95367	5195600
Máx.	119528	120150	118919	119528	19300000

Por fim, os dados de cotações obtidos são utilizados para o cálculo de Indicadores Técnicos, que são valores que apresentam tendências de mercado ao investidor. Nesse sentido, tanto as cotações como os indicadores são utilizados como entrada para os algoritmos de Aprendizado de Máquina, apresentados na subseção a seguir. Os seguintes indicadores foram utilizados [4, 7]:

- **Média Móvel** (do inglês, *Moving Average* - MA): consiste no valor médio de um ativo em um dado período, com parâmetros de 10, 20 e 30 dias.
- **Média móvel de convergência e divergência** (do inglês, *Moving Average Convergence Divergence* - MACD): consiste em um indicador de tendência, com parâmetros 12, 26 e 9 dias em cada média móvel.
- **Índice de Força Relativa** (do inglês, *Relative Strength Index* - RSI), com os parâmetros 6, 12 e 24 dias.
- **Bandas Bollinger** (do inglês, *Bollinger Bands* - BB): consiste em um indicador de volatilidade do mercado. Não possui parâmetros adicionais.
- **Índice de Fluxo do Dinheiro** (do inglês, *Money Flow Index* - MFI): consiste em um indicador da intensidade do fluxo de dinheiro para dentro ou fora de um ativo, com parâmetro de 14 dias.
- **Balanco de Volume** (do inglês, *On Balance Volume* - OBV): consiste em um indicador de volume de transações. Não possui parâmetros adicionais.
-

B. Algoritmos de Aprendizado de Máquina

Neste trabalho, foram selecionados algoritmos de Aprendizado de Máquina que estão dentre os algoritmos mais comumente utilizados em trabalhos da literatura sobre

previsão do mercado de ações [7]. A listagem dos algoritmos escolhidos é a seguinte:

- Linear Regression (Lin. Reg.).
- Support Vector Regression (SVR), considerando três núcleos diferentes: RBF, polinomial (P) e linear (L).
- XGBoost.
- Random Forest (Rand. Forest).
- Multilayer Perceptron (MLP).

Além dos algoritmos considerados, também utilizamos a técnica de *Ensemble*, que consiste em combinar as previsões de diversos algoritmos de Aprendizado de Máquina, por meio do cálculo da média de seus resultados. Neste trabalho, portanto, o *Ensemble* foi composto da combinação dos cinco algoritmos apresentados anteriormente, com o intuito de explorar a sua capacidade preditiva em conjunto.

As entradas de cada um dos algoritmos são compostas das seguintes características: cotação histórica (que inclui os valores de abertura, máxima, mínima, fechamento, fechamento ajustado e volume) e indicadores técnicos (MA, MACD, RSI, BB, MFI e OBV). Antes de serem introduzidas nos algoritmos de Aprendizado de Máquina, as entradas passam por um processo de normalização *Min-Max* [5].

Nesse sentido, a partir destas entradas normalizadas, o objetivo é prever o valor de fechamento da ação no dia seguinte, ou seja, a partir de dados fornecidos sobre o dia d , queremos prever o valor de fechamento da ação no dia $d+1$.

Para isso, o treinamento dos algoritmos de Aprendizado de Máquina escolhidos é feito do seguinte modo. Como cada algoritmo possui uma série de parâmetros para serem investigados, utilizamos uma busca em grade (do inglês, *grid search*), para escolher os parâmetros mais adequados, conforme apresentado na TABELA . Para o desenvolvimento dos algoritmos de Aprendizado de Máquina utilizados neste trabalho, foi utilizada a linguagem de programação *Python* em conjunto com a biblioteca *Scikit-Learn* (disponível em <https://scikit-learn.org/>).

Além disso, como o conjunto de dados utilizado para treinamento apresenta dependência temporal, utilizamos o método de validação cruzada com janela incremental (do inglês, *increasing window cross validation*) com $k=3$ passos, conforme usada por [5, 8]. Neste método, uma janela é utilizada para, a cada passo k , aumentar o tamanho do conjunto de dados que é utilizado no treinamento do algoritmo escolhido. Além disso, o método reserva um subconjunto de dados, posterior ao conjunto usado para treinamento, para ser utilizado como conjunto de validação, conforme apresentado na Fig. 3.

Para o desenvolvimento dos algoritmos de Aprendizado de Máquina utilizados neste trabalho, foi utilizada a linguagem de programação *Python* em conjunto com a biblioteca *Scikit-Learn* (disponível em <https://scikit-learn.org/>).

C. Avaliação

Para avaliar e selecionar o modelo com parâmetros mais adequados para a tarefa de previsão do mercado brasileiro utilizou-se a métrica erro quadrático médio (do inglês, *Mean Squared Error* - MSE), muito utilizado em trabalhos da área [5, 8].

Uma vez selecionado o modelo com parâmetros mais adequados, foi realizada uma simulação de investimento considerando o valor inicial de R\$ 200.000,00. A simulação de investimento é realizada da seguinte forma: (1) realiza-se uma previsão p do valor de fechamento da ação no dia $d+1$; (2) se p maior ou igual a 0, utiliza-se todo o orçamento disponível para comprar a ação ou, caso já possua a ação, o investidor a mantém em sua posse; e (3) se p menor que 0, vende-se todas as ações que o investidor possua ou, caso não possua a ação, o investidor não realiza nenhuma ação.



Fig. 3. Método de validação cruzada com janela incremental - *Increasing Window Cross Validation*

TABELA II
LISTA DE PARÂMETROS

Modelo	Parâmetros
Linear Regression (Lin. Reg.)	-
Support Vector Regression (SVR)	C : {0.001, 0.01, 0.1, 1} Kernel: {Linear, Poly, RBF}
XG Boost	booster: {gbtree, gblinear, dart}, max_delta_step: {0,1,5}, lambda: {1,3,5,10,50,100}
Random Forest (Rand. Forest)	criterion: {squared_error,poisson}, n_estimators: {150}, max_leaf_nodes: {5,10,35,None}, min_samples_leaf: {1,3,5}
Multilayer Perceptron (MLP)	max_iter: {250,500,1000}, learning_rate_init: {0.01,0.001}, hidden_layer_sizes: {(10,10,10),(25,25,25),(50,50,50),(10,10,10,10),(25,25,25,25,25),(50,50,50,50,50)}, early_stopping = true

Assim, os resultados da simulação de investimento são avaliados por meio das seguintes métricas: (1) valor final de investimento, que representa o valor resultante ao final do período considerado; (2) *Sharpe Ratio* (SR), que representa o risco do investimento, sendo que valores mais elevados de SR representam menores riscos enquanto que valores baixos representam grandes riscos; e (3) *Maximum Drawdown*

(MDD), que representa a maior queda de um ativo desde o seu valor mais elevado, sendo que, valores menores de MDD indicam uma pequena queda da ação em relação ao seu máximo, enquanto que valores elevados de MDD representam uma queda significativa.

A Seção a seguir apresenta os resultados obtidos nos experimentos deste trabalho.

IV. RESULTADOS

Nesta seção, são apresentados os resultados dos experimentos desenvolvidos neste trabalho. Desse modo, esta seção está dividida da seguinte forma: a subseção A apresenta os resultados do treinamento dos diversos algoritmos de Aprendizado de Máquina utilizados neste trabalho, bem como também os parâmetros mais adequados a cada um deles; a subseção B apresenta os resultados simulação de investimento utilizando os modelos selecionados; a subseção C discute os resultados obtidos.

A. Treinamento de Algoritmos de Aprendizado de Máquina

Para o treinamento de cada um dos algoritmos de Aprendizado de Máquina selecionados neste trabalho, consideramos o conjunto de dados de entrada como sendo as cotações do índice Bovespa. Para isso, o conjunto de dados foi dividido da seguinte forma para realização das etapas de treinamento e teste:

- Conjunto de dados de treinamento: 01-01-2015 a 31-12-2019.
- Conjunto de dados de teste: 01-01-2020 a 23-03-2020.

A TABELA apresenta os resultados contendo os parâmetros mais adequados a cada um dos algoritmos considerados, ou seja, os algoritmos cujos parâmetros apresentaram menor medida de erro durante a etapa de treinamento. Dessa forma, dentre todas as combinações de parâmetros testadas com o método de busca em grade, conforme apresentado na Seção III, os modelos cujos parâmetros estão mostrados na tabela serão utilizados nos experimentos de simulação de investimento da subseção subsequente. Vale a pena destacar que o *Ensemble* utiliza os modelos cuja combinação de parâmetros está também nesta tabela. Por fim, a TABELA apresenta os resultados das medidas de erro de treinamento, considerando o conjunto de dados de validação e a métrica MSE, para os algoritmos cujos parâmetros estão apresentados na TABELA, além da técnica de *Ensemble*.

B. Simulação de Investimento

A partir dos modelos selecionados no experimento anterior, foi realizada uma simulação de investimento considerando o patrimônio inicial de R\$ 200.000,00 durante a pandemia de Covid-19, ou seja, considerando o conjunto de testes que abrange o período de 01-01-2020 a 23-03-2020.

Como *baseline* para este experimento, consideramos a estratégia de investimento *Buy&Hold*, que consiste em o investidor comprar um ativo (no caso, uma fração do índice Bovespa) no início do período de investimento e mantê-lo sob sua posse durante todo o período considerado. O resultado do investimento com essa estratégia é apresentado na TABELA.

TABELA III
SELEÇÃO DE PARÂMETROS PARA CADA ALGORITMO DE APRENDIZADO DE MÁQUINA

Modelo	Parâmetros Adequados
Lin. Reg.	-
SVR (Linear)	C: 1
SVR (Poly)	C: 0.001
SVR (RBF)	C: 1
XG Boost	booster: gbtree, lambda: 100, max_delta_step: 0
Rand. Forest	criterion: squared_error, max_leaf_nodes: None, min_samples_leaf: 5, n_estimators: 150
MLP	hidden_layer_sizes: (10, 10, 10), learning_rate_init: 0.01, max_iter: 500

TABELA IV
MEDIDA DE ERRO (MSE) CONSIDERANDO OS ALGORITMOS DE APRENDIZADO DE MÁQUINA COM PARÂMETROS SELECIONADOS NO EXPERIMENTO ANTERIOR.

Modelo	MSE
SVR (RBF)	0,0639
SVR (Poly)	0,0246
SVR (Linear)	0,0210
RandForest	0,0189
XGBoost	0,0161
Ensemble	0,0005
MLP	0,0004
Lin. Reg.	0,0002

TABELA V
RESULTADO DE INVESTIMENTO COM A ESTRATÉGIA BUY&HOLD.

Estratégia	Valor Final	MDD	SR
Buy & Hold	R\$ 148.505,00	26,19	-0,224

A TABELA apresenta o resultado de investimento utilizando os algoritmos de Aprendizado de Máquina e a técnica de Ensemble, conforme os parâmetros selecionados na subseção anterior. Esta tabela apresenta tanto o número de acertos no movimento do mercado (i.e., os movimentos de alta e baixa) como também o valor final de investimento e as medidas de risco MDD e SR. Finalmente, a Figura 4 apresenta a comparação entre os movimentos reais que ocorreram com o índice Bovespa no período e o movimento previsto por cada um dos modelos analisados neste trabalho.

TABELA VI
RESULTADO DE INVESTIMENTO COM ALGORITMOS DE
APRENDIZADO DE MÁQUINA.

Modelo	Movi- mento	Valor Final	MDD	SR
Ensemble	36	R\$ 219.483,00	1,3635	0,204
SVR (L)	35	R\$ 219.473,00	0,052	0,213
MLP	31	R\$ 205.930,00	1,801	0,093
SVR (RBF)	32	R\$ 200.186,00	0,627	-0,017
XG Boost	34	R\$ 199.553,00	5,537	-0,002
Rand. Forest	32	R\$ 181.202,00	14,257	-0,100
Lin. Reg.	25	R\$ 160.798,00	20,919	-0,184
SVR (P)	24	R\$ 148.505,00	26,190	-0,224

A subseção a seguir analisa e discute os resultados obtidos nestes experimentos.

C. Discussão

Em se tratando dos resultados obtidos neste trabalho, primeiramente, vale a pena destacar o experimento voltado à seleção de parâmetros para cada um dos modelos a serem selecionados. Uma observação é a de que, conforme apresentada em outros trabalhos da área de previsão de mercado com algoritmos de Aprendizado de Máquina [5, 7], é de extrema importância a necessidade de analisar o funcionamento dos modelos na prática, isto é, em simulação de investimentos. Por exemplo, o modelo *Linear Regression* apresentou a menor medida de erro para o conjunto de dados de validação. No entanto, este modelo não apresentou resultados satisfatórios na simulação de investimento. Por outro lado, o modelo MLP e a técnica de *Ensemble* apresentaram, respectivamente, o segundo e o terceiro menor valor dentre as medidas de erro e, na simulação de investimento, apresentaram resultados significativos quando comparado ao modelo *Linear Regression*. Tal resultado pode ter sido obtido, provavelmente, porque o modelo *Linear Regression* apresenta características lineares, o que não acontece com modelos baseados em redes neurais, como a MLP ou a combinação de modelos promovida pelo *Ensemble*. De fato, o comportamento do mercado financeiro também é de característica não-linear [19]. Portanto, esses resultados mostram que nem todos os modelos que apresentam menor valor de erro durante o treinamento também apresentam melhores resultados quando aplicados ao contexto de investimentos.

Além disso, os resultados obtidos em nossos experimentos mostraram que a proposta de Ensemble e os modelos SVR com kernel linear, MLP e SVR com kernel RBF foram capazes, inclusive, de trazer lucro durante o período considerado, além de reduzirem riscos (MDD e SR) associados ao período de investimento e, conforme apresentado também na Figura 4, eles foram particularmente úteis na previsão da queda do mercado, acompanhando significativamente o valor real do índice Bovespa no período. Desse modo, os algoritmos MLP e SVR são altamente recomendados aos investidores para períodos de crise e devem ser considerados em estratégias de investimento em tais momentos. Eventualmente, a combinação de tais modelos em um único sistema de previsão pode ser uma estratégia interessante, o que de fato foi observado pelos resultados da proposta de Ensemble.

No entanto, este estudo possui algumas limitações que serão consideradas em trabalhos futuros. A primeira delas é que, na área de mercado financeiro, ganhos passados não significam ganhos futuros. Portanto, em eventos de crises futuras, o investidor deve agir com cautela, ainda que escolha os modelos mais adequados estudados em crises passadas. Outra limitação deste trabalho é não considerar o uso de notícias, que apresentam grande impacto e podem mudar o fluxo do mercado diretamente, principalmente em instabilidades políticas. Vale a pena abordar também que existem diferentes indicadores de Análise Técnica, e a sua escolha é altamente dependente do investidor. A escolha de diferentes indicadores, portanto, poderia alterar os resultados obtidos neste trabalho e, assim, cabe ao investidor selecionar e utilizar os indicadores mais adequados aos seus objetivos.

V. CONCLUSÃO

Este trabalho apresentou o uso de algoritmos de Aprendizado de Máquina voltados para a previsão do mercado de ações brasileiro com o objetivo de reduzir as perdas enfrentadas no mercado durante a pandemia de Covid-19. Os resultados mostraram que a técnica de *Ensemble*, combinando diversos algoritmos de Aprendizado de Máquina, apresentou os melhores resultados dentre os modelos analisados, reduzindo o risco de investimento e inclusive conseguindo trazer algum lucro ao investidor durante o período.

As principais contribuições deste trabalho são: (1) o comparativo de diversos algoritmos de Aprendizado de Máquina, com o intuito de apresentar ao investidor qual modelo pode ser o mais adequado durante o período de crises; e (2) o uso de estudo de caso voltado ao mercado de ações brasileiro, que carece de estudos na área.

Como trabalho futuros, pretendemos ampliar os experimentos, considerando: (1) outros períodos de crise, como a crise do ano de 2008; (2) outros mercados mundiais, como o Americano e o Asiático; e (3) utilizar modelos de Aprendizado Profundo (do inglês, *Deep Learning*) que sejam adaptados para dados com características temporais.

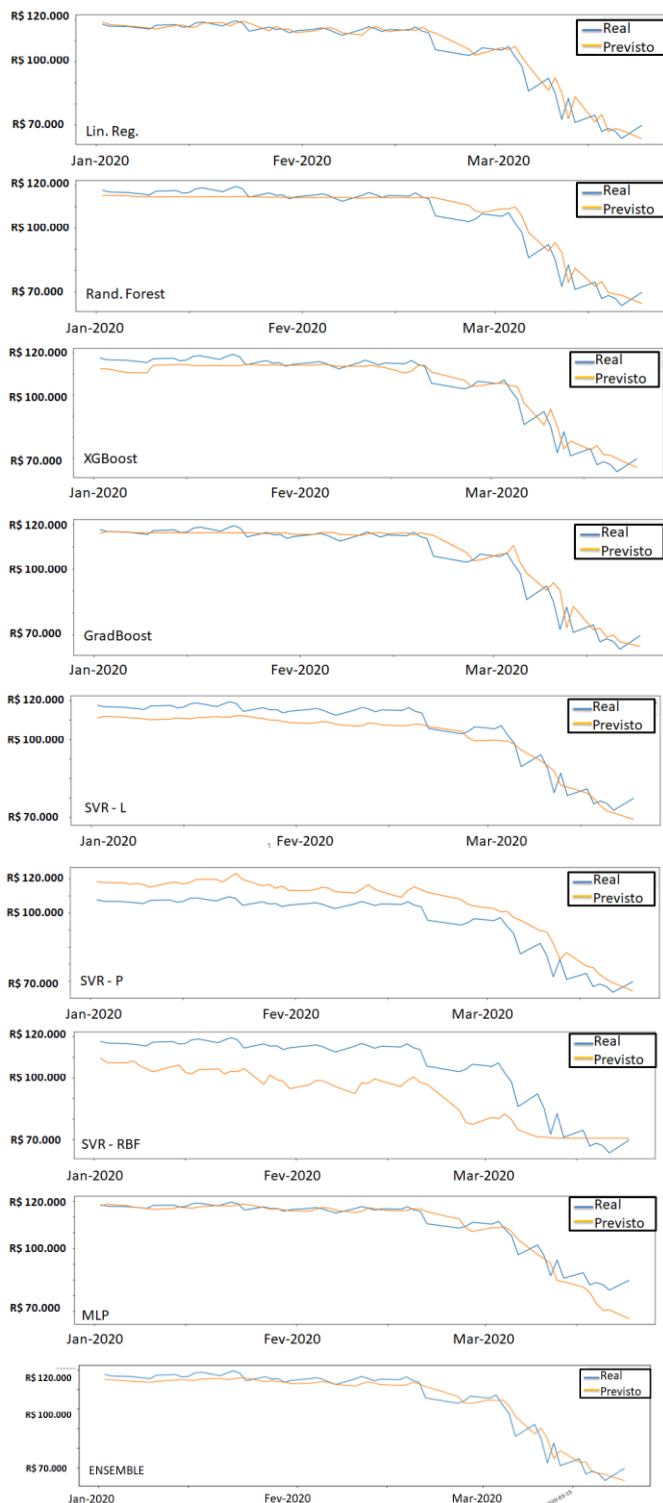


Fig. 4. Gráfico comparando as previsões com os valores reais das ações de todos os algoritmos de Aprendizado de Máquina considerados.

REFERÊNCIAS

- [1] S. Pilz, V. Theiler-Schwetz, C. Trummer, R. Krause, e J. P. A. Ioannidis, “SARS-CoV-2 reinfections: Overview of efficacy and duration of natural and hybrid immunity”, *Environ. Res.*, vol. 209, no. January, pp. 112911, 2022, doi: 10.1016/j.envres.2022.112911.
- [2] J. Herby, L. Jonung, e S. H. Hanke, “A literature review and meta-analysis of the effects of lockdowns on COVID-19 mortality”, *Stud. Appl. Econ.*, no. 200, 2022.

- [3] A. Lo, “The adaptive market hypothesis: market efficiency from an evolutionary perspective”, *J. Portf. Manag.*, vol. 30, no. 5, pp. 15–29, 2004.
- [4] S. B. Achelis, “Technical Analysis from A to Z”, *Search*, vol. 77, pp. 33–4, 2000, [Online]. Available at: <http://books.google.com/books?hl=en&lr=&id=ZproZYDvKq sC&oi=fnd&pg=PR13&dq=Technical+Analysis+from+A +to+Z&ots=AI16VIRoS2&sig=cujwHZYaCu3dAI-urVEj9ykUEm8>
- [5] A. Picasso, S. Merello, Y. Ma, L. Oneto, e E. Cambria, “Technical analysis and sentiment embeddings for market trend prediction”, *Expert Syst. Appl.*, vol. 135, pp. 60–70, nov. 2019, doi: 10.1016/j.eswa.2019.06.014.
- [6] M. M. Kumbure, C. Lohrmann, P. Luukka, e J. Porras, “Machine learning techniques and data for stock market forecasting: A literature review”, *Expert Syst. Appl.*, vol. 197, pp. 116659, jul. 2022, doi: 10.1016/j.eswa.2022.116659.
- [7] J. J. Duarte, S. Montenegro González, e J. C. Cruz, “Predicting Stock Price Falls Using News Data: Evidence from the Brazilian Market”, *Comput. Econ.*, no. 0123456789, 2020, doi: 10.1007/s10614-020-10060-y.
- [8] X. Su, X. Yan, e C.-L. Tsai, “Linear regression”, *WIREs Comput. Stat.*, vol. 4, pp. 275–294, 2012, doi: 10.1002/wics.1198.
- [9] W. S. Noble, “What is a support vector machine?”, *Nat. Biotechnol.* 2006 2412, vol. 24, no. 12, pp. 1565–1567, dez. 2006, doi: 10.1038/nbt1206-1565.
- [10] L. Breiman, “Random Forests”, *Mach. Learn.* 2001 451, vol. 45, no. 1, pp. 5–32, out. 2001, doi: 10.1023/A:1010933404324.
- [11] T. Chen e C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, doi: 10.1145/2939672.
- [12] M. W. Gardner e S. R. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences”, *Atmos. Environ.*, vol. 32, no. 14–15, p. 2627–2636, ago. 1998, doi: 10.1016/S1352-2310(97)00447-0.
- [13] B. M. Henrique, V. A. Sobreiro, e H. Kimura, “Literature review: Machine learning techniques applied to financial market prediction R”, *Expert Syst. Appl.*, vol. 124, p. 226–251, 2019, doi: 10.1016/j.eswa.2019.01.012.
- [14] S. Rajabi, P. Roozkhosh, e N. M. Farimani, “MLP-based Learnable Window Size for Bitcoin price prediction”, *Appl. Soft Comput.*, vol. 129, p. 109584, nov. 2022, doi: 10.1016/J.ASOC.2022.109584.
- [15] I. Medarhri, M. Hosni, N. Nouisser, F. Chakroun, e K. Najib, “Predicting Stock Market Price Movement using Machine Learning Techniques”, *2022 8th Int. Conf. Optim. Appl.*, pp. 1–5, out. 2022, doi: 10.1109/ICOA55659.2022.9934252.
- [16] Z. Rustam e P. Kintandani, “Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation”, *Model. Simul. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/8962717.
- [17] M. Sedighi, H. Jahangirnia, M. Gharakhani, e S. F. Fard, “A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine”, *Data* 2019, vol. 4, pp. 75, vol. 4, no. 2, p. 75, maio 2019, doi: 10.3390/DATA4020075.
- [18] X. Li, P. Wu, e W. Wang, “Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong”, *Inf. Process. Manag.*, p. 102212, fev. 2020, doi: 10.1016/j.ipm.2020.102212.
- [19] G. S. Atsalakis e K. P. Valavanis, “Surveying stock market forecasting techniques – Part II: Soft computing methods”, *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5932–5941, abr. 2009, doi: 10.1016/J.ESWA.2008.07.006.



Arthur Emanuel de Oliveira Carosia has a Bachelor’s degree and a Master’s degree in Computer Science from the University of São Paulo – USP, and a Ph.D. in Technology from the University of Campinas - UNICAMP. His research interests are Data Science and Computational Intelligence, with recent work on machine learning and stock market prediction.