

**INVESTIMENTOS EM TEMPOS DE PANDEMIA: UMA ABORDAGEM UTILIZANDO  
MACHINE LEARNING NA CONSTRUÇÃO DE PORTFÓLIOS**

**JOSÉ ERASMO SILVA**

UNIVERSIDADE METODISTA DE PIRACICABA (UNIMEP)

**JORGE LUIZ DOS SANTOS SILVA**

UNIVERSIDADE METODISTA DE PIRACICABA (UNIMEP)

**MARIA IMACULADA DE LIMA MONTEBELLO**

UNIVERSIDADE METODISTA DE PIRACICABA (UNIMEP)

# INVESTIMENTOS EM TEMPOS DE PANDEMIA: UMA ABORDAGEM UTILIZANDO *MACHINE LEARNING* NA CONSTRUÇÃO DE PORTFÓLIOS

## 1 INTRODUÇÃO

O Brasil vive atualmente o menor patamar de taxa de juros da história, fixada em 2,25% em junho de 2020 segundo informações do BACEN (2020). Isso tem feito com que aplicações de renda fixa, famosas por seu alto retorno e baixo risco no mercado brasileiro, se tornem cada vez menos interessantes. Nesse sentido, visando manter os ganhos que antes eram obtidos através da renda fixa, notou-se um incremento substancial no número de investidores na bolsa de valores, saltando de 813.291 no final de 2018 para 2.648.975 em junho de 2020. Tem-se então dois números inéditos para o Brasil, a menor taxa de juros da história e o maior número de investidores registrados na bolsa de valores.

Ao mesmo tempo que se observava a taxa de juros em queda e o número de investidores em alta, o principal índice brasileiro de ações atingiu suas máximas históricas em janeiro de 2020, ultrapassando a marca de 119 mil pontos. Iniciava-se um ano promissor para o mercado brasileiro de ações. Concomitante a isso, surgem na China os primeiros casos de COVID-19, que viria a se tornar uma das piores pandemias já enfrentadas na história. O que a princípio não assustou, começa a surpreender com o número de infectados, número de mortos e a velocidade da sua disseminação.

O primeiro caso de COVID-19 no Brasil foi confirmado em 26/02/2020 fazendo com que a insegurança do que estava por vir fizesse a bolsa despencar 7% nessa data. Depois disso foram seguidas baixas até o índice atingir a mínima do ano aos 63 mil pontos em 23/03/2020, por coincidência, a mesma data em que o governador do estado de São Paulo decretou estado de quarentena. A partir daí o mercado iniciou sua recuperação, apesar de ainda não estar claro como a pandemia seria controlada e como impactaria nas empresas.

Como se pode observar, prever os movimentos do mercado não é uma tarefa fácil e investir no mercado de ações sem uma estratégia definida pode gerar grandes perdas ao investidor. Paiva et al., (2019) descrevem o mercado de ações como sendo dinâmico, complexo, evolucionário, não linear, nebuloso, não paramétrico e caótico por natureza. Além disso, ele é extremamente sensível a fatores políticos, condições microeconômicas e macroeconômicas e expectativas e inseguranças dos investidores.

Wang et al., (2020a) combinaram um algoritmo *Random Forest* em conjunto com um modelo de seleção de ações com múltiplos fatores, tornando a estratégia de seleção de ações mais universal e adaptável. O modelo foi testado no mercado Chinês onde alcançou resultado bastante superior ao principal índice de ações do país, além de ser superior a grandes fundos de investimentos quantitativos. Wang et al., (2020b) propuseram um modelo híbrido. Esses modelos foram trabalhados em duas etapas sendo que, a primeira etapa utilizou o modelo LSTM (*long short-term memory*) e a segunda etapa no processo de formação da carteira utilizando-se o modelo de média variância. Os resultados foram comparados a outros modelos tradicionais, como por exemplo, SVM, RAF, DNN e ARIMA. O retorno acumulado no período de 1994 a 2018 mostrou resultado bastante superior aos modelos utilizados como base na comparação. Joshi e Chauhan (2020) estudaram os determinantes e a acurácia das previsões de múltiplos de preços para empresas do sudeste asiático utilizando dois tipos de abordagens, modelos de regressão *ordinary least square* (OLS) e *machine learning* (Ridge e Lasso). Tanto utilizando modelos convencionais de regressão quanto utilizando *machine learning* os maiores  $R^2$  ajustados e os menores *root mean square error* (RMSE) foram encontrados para a variável P/B. Além disso, observou-se melhora nesses indicadores quando se comparou os modelos convencionais com o modelo baseado em *machine learning*. Lin e Blum, (2020) utilizaram uma abordagem de *Reinforcement Learning*. Este sistema tem a capacidade de aprender estratégias

ótimas a partir de ambientes com informações imperfeitas possibilitando a criação de sistemas robustos de apoio a decisão. Nesse sentido, os autores mostraram que duas técnicas de transformação automática de recursos – *Bayesian recurrent neural network* (BRNN) para modelagem de tendência de preços futuros e *Generative Adversarial* (GANs) para modelagem da variação dos preços no curto prazo – são capazes de melhorar a performance de agentes em problemas relacionados ao gerenciamento de portfólios, principalmente em termos de aumento de lucratividade e redução de riscos. Wu et al., (2020) propuseram duas estratégias de negociação construídas a partir de modelos *reinforcement learning*. As estratégias, chamadas de GDQN (*Gated Deep Q-learning trading strategy*) e GDPG (*Gated Deterministic Policy Gradient trading strategy*), foram testadas em diferentes países e comparadas com outras estratégias tradicionais do mercado. Apesar de utilizarem uma amostra pequena (3 empresas americanas, 3 empresas britânicas e 3 empresas chinesas) as estratégias mostraram resultados melhores que as estratégias tradicionais. De acordo com Yu et al., (2020) muitos fatores que afetam os preços das ações tem características redundantes e não lineares. Sendo assim, modelos clássicos de redução de mapeamento linear, como por exemplo *principal component analysis* (PCA) e *linear discriminant analysis* (LDA) não apresentam bons resultados para problemas não lineares. Nesse sentido, os autores propuseram a utilização em conjunto de dois modelos, o *local linear embedding* (LLE) que é um algoritmo de redução dimensional, com o objetivo de reduzir a dimensão dos fatores que afetam o preço das ações e, posteriormente, os dados gerados neste modelo foram utilizados como *input* em uma rede neural com o algoritmo *back propagation* (BP). Desta forma o modelo foi chamado de LLE-BP e mostrou efetivo e viável na previsão dos preços de ações.

Entende-se que as abordagens baseadas em Inteligência Artificial tem ganhado cada vez mais espaço e tem mostrado cada vez melhores resultados. No entanto, em alguns contextos torna-se quase que impossível prever o futuro dos preços das ações. Nesse sentido outras ferramentas são necessárias para tornar o portfólio seguro e rentável. Segundo Harvey et al (2019) os modelos de alocação *long and short* podem ser utilizados com o objetivo de tornar um portfólio autoimune pois contribuem para que o investidor não seja afetado pela direção do mercado e, sendo assim, são excelentes alternativas de proteção em tempos de crises. Para Jião, Massa e Zhang (2016) esse tipo de modelo contribui ainda por minimizar o custo de investimento pois o lado *short* da carteira serve de recursos para cobrir outras posições. Ainda no contexto da proteção, inserir outros tipos de ativos sem correlação contribui para diversificação e minimização de risco. Dessa forma, um dos ativos muito conhecidos e amplamente estudados é o ouro. Seja físico ou por negociação eletrônica, ele é um dos ativos mais utilizados para proteção de capital. (HARVEY et al., 2019). (NAWAZ; AZAM; ASLAM, 2020). (SAMUEL, 2020). (ALSHAMMARI; ALTARTURI; SAITI, 2020).

O mercado de ações desempenha um papel fundamental no sistema econômico de um país e provê oportunidades aos participantes do mercado tal como investidores, *traders*, e o público de maneira geral aumentarem suas riquezas através do investimento em ações. Contudo, medidas preventivas devem ser tomadas para lidar com os riscos de mercado onde o valor do investimento pode aumentar ou diminuir de acordo com as condições do mercado de ações. (ISMAIL et al., 2020).

**A partir de uma abordagem empregando *machine learning* o objetivo deste artigo foi propor um modelo de seleção e alocação de ativos que permita a construção de portfólios rentáveis e seguros, mesmo em tempos de insegurança e pouca previsibilidade.**

Para atingir o objetivo utilizou-se primeiramente o algoritmo *random forest* para seleção mensal das ações. Dentro de um contexto de *long and short* as ações com maiores probabilidades de retornos positivos foram compradas para o lado *long* e as ações com menores probabilidades foram vendidas a descoberto para o lado *short* da carteira. Com esse método projetou-se proteger a carteira das mudanças de tendências no mercado. Por fim, ainda como

uma estratégia de proteção, foi alocado um percentual de ouro na carteira que visa, especialmente em momentos de crise, proteger o capital do investidor. Os resultados se mostraram satisfatórios, especialmente no primeiro semestre de 2020 no qual vivenciou-se as incertezas promovidas pela pandemia de COVID-19.

Na seção 2 apresenta-se uma revisão do referencial teórico e sumariza-se os trabalhos recentes sobre *machine learning* e previsão do preço de ações. Na seção 3 descreve-se a metodologia, amostra, variáveis dependentes, variável dependente e o modelo final. Na seção 4 são apresentados os resultados do trabalho e, por fim, na seção 5 são discutidos os principais achados e também sugestões para trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

A crescente complexidade e propriedade dinâmica dos mercados de ações são os principais desafios do setor financeiro, nos quais estratégias de negociação inflexíveis projetadas por profissionais financeiros experientes não conseguem atingir desempenho satisfatório em todas as condições do mercado. (WU et al., 2020).

No mínimo sob duas perspectivas a maneira de investir mudou nas últimas décadas. Primeiro, as informações financeiras em tempo real são muito mais acessíveis atualmente. Principalmente em função de novas tecnologias que promoveram não só amplo e rápido acesso à internet, mas também permitiram maior eficiência na análise dos dados. Segundo, a inteligência artificial, suportada pelos avanços tecnológicos, tem levado a uma nova maneira de investir, onde as decisões são tomadas com base na análise de um grande volume de informações quantitativas. (CERVELLO-ROYO; GUIJARRO, 2020).

Com o desenvolvimento da inteligência artificial, *big data*, tecnologia de computação em nuvem, as pessoas conhecem gradualmente o investimento quantitativo, que emite instruções de negociação por meio de métodos quantitativos e programação de computadores. (WANG et al., 2020). Nas últimas décadas, investimento quantitativo tornou-se um *hot spot* no desenvolvimento do mercado de capitais na Europa e Estados Unidos. A escala de mercado do investimento quantitativo chegou a 70% do mercado de investimento nos EUA, apresentando desempenho estável, e tornou-se um novo método para investidores. (WANG; XU; LIU, 2020).

Desenvolver um método consistente e acurado de previsão no mercado de ações não é uma tarefa fácil posto que vários fatores, tal como guerras, política etc. afetam os sentimentos dos investidores e, portanto, afetam o comportamento no mercado. (CHEN, 2020). O mercado de ações tem natureza dinâmica, complexa, evolucionária, não linear, nebuloso, não paramétrica e caótica. Além disso, ele é extremamente sensível a fatores políticos, condições microeconômicas e macroeconômicas e expectativas e inseguranças dos investidores. (PAIVA et al., 2019). Para melhor lidar com esse ruidoso ambiente, recentemente, mais e mais pesquisadores estão focados na aplicação e abordagens que utilizam inteligência artificial (AI). (JIANG et al., 2020).

Atendendo a crescente evolução e acessibilidade de novos modelos de investimento, entende-se que antigas estratégias devem ser revistas e evoluídas, bem como novas devem ser criadas. Além de contribuir para maior estabilidade do mercado, a diversidade de estratégias contribui positivamente em momentos de crise e recessão. Quando muitos investidores se agrupam em torno de uma mesma estratégia, os retornos até determinado ponto podem ser conjuntamente positivos, mas em momentos de crise esse comportamento pode ser devastador (AVRAMOV et al., 2017).

Fazer previsões no mercado de ações é um assunto popular e muito desafiador em finanças. Popular porque prever o retorno de ações é muito útil no processo decisório de alocação de ações em carteiras e desafiador por causa das propriedades dinâmicas, caóticas, não lineares e sujeitas a muitos ruídos do mercado. (JIANG et al., 2020). Atualmente,

realização de previsões baseadas em métodos de *machine learning* são consideradas uma das mais efetivas soluções para esse desafio. (NIU et al., 2020). O *machine learning* é um dos ramos da inteligência artificial (IA) e dentro dele existem vários modelos que tem evoluído para atender diferentes demandas.

Um modelo avançado de *machine learning* que vem obtendo bons resultados no mercado financeiro e tem chamado a atenção de investidores e pesquisadores é conhecido como *random forest*. Este modelo é derivado das árvores de decisão criado com a finalidade de melhorar a precisão das mesmas e superar a alta sensibilidade a pequenas alterações nos dados. (FISCHER; KRAUSS, 2018) (RAMASUBRAMANIAN; SINGH, 2016).

O nome *random forest* vem do fato de que os elementos estruturais do algoritmo são árvores de decisão e cada uma dessas árvores é construída com uma parcela de aleatoriedade. Os resultados do processo preditivo deste algoritmo são derivados como votação majoritária das árvores em problemas de classificação e como previsão média em problemas de regressão. (TRATKOWSKI, 2020). O autor destaca ainda que o algoritmo *random forest* pode ser utilizado em complexos ambientes econômicos e também pode ser considerado útil na formulação de estratégias de investimentos.

*Random forest* combina versatilidade e potência em uma única abordagem de *machine learning*. Como o conjunto usa apenas uma parte pequena e aleatória do conjunto completo de recursos, o algoritmo pode lidar com conjuntos de dados extremamente grandes. (LANTZ, 2019).

Wang et al., (2020a) combinaram um algoritmo *random forest* em conjunto com um modelo de seleção de ações com múltiplos fatores, tornando a estratégia de seleção de ações mais universal e adaptável. Foram coletadas informações financeiras de 2012 a 2020 das empresas listadas no CSI 300, que é um dos principais índices de ações da China. A partir do modelo foram rodados testes com 100, 500 e 1000 árvores de decisão, obtendo-se, respectivamente, retornos anualizados de 17,80, 29,40 e 24,50%. Conforme exposto, observa-se que o melhor retorno foi apresentado quando se utilizou 500 árvores de decisão, no entanto, todos eles se mostraram bastante superiores ao principal índice de referência Chinês, o CSI 300, com retorno anualizado de 5,80%. Os resultados foram superiores ainda a alguns fundos quantitativos no mercado.

Outra área do *machine learning* é conhecida como *Reinforcement Learning*. Esta abordagem tem a capacidade de aprender estratégias ótimas a partir de ambientes com informações imperfeitas possibilitando a criação de robustos sistemas de apoio a decisão. Nesse sentido, Lin e Blum, (2020) mostraram que duas técnicas de transformação automática de recursos – *bayesian recurrent neural network* (BRNN) para modelagem de tendência de preços futuros e *generative adversarial* (GANs) para modelagem da variação dos preços no curto prazo – são capazes de melhorar a performance de agentes em problemas relacionados ao gerenciamento de portfólios, principalmente em termos de aumento de lucratividade e redução de riscos.

Wu et al., (2020) propuseram duas estratégias de negociação construídas a partir de modelos *reinforcement learning*. As estratégias, chamadas de GDQN (*Gated Deep Q-learning trading strategy*) e GDPG (*Gated Deterministic Policy Gradient trading strategy*), foram testadas em diferentes países e comparadas com outras estratégias tradicionais do mercado. Apesar de utilizarem uma amostra pequena (3 empresas americanas, 3 empresas britânicas e 3 empresas chinesas) as estratégias mostraram resultados melhores que as estratégias tradicionais. Os autores deixam como sugestão expandir a técnica utilizada para um número maior de empresas para que a estratégia possa ser utilizada na alocação de portfólios maiores.

Joshi e Chauhan (2020) estudaram os determinantes e a acurácia das previsões de múltiplos de preços para empresas do sudeste asiático utilizando dois tipos de abordagens, a convencional (conforme descrevem os autores) nas quais utilizou-se modelos de regressão

*ordinary least square* (OLS) e as não convencionais, nas quais se utilizou *machine learning* (ridge e lasso). Os múltiplos de preços utilizados na pesquisa foram *price to earnings* (P/E), *price to book* (P/B) e *price to sales* (P/S). Como variáveis explicativas foram utilizadas variáveis comuns e amplamente utilizadas no meio acadêmico para trabalhos dessa área, como por exemplo, taxa de crescimento dos lucros, *payout*, tamanho da firma, volatilidade dos preços, setor, *dividend yield* etc. Tanto utilizando modelos convencionais de regressão quanto utilizando *machine learning* os maiores  $R^2$  ajustados e os menores *root mean square error* (RMSE) foram encontrados para a variável P/B. Além do mais, observou-se melhora nesses indicadores quando se comparou os modelos convencionais com o modelo baseado em *machine learning*. Nesse sentido os autores recomendam o uso do múltiplo P/B para avaliação relativa no contexto das empresas do sudeste asiático.

Evoluções de *machine learning* tem surgido frequentemente no mercado, como é o caso da *deep learning*. Ela é uma técnica avançada de *machine learning* baseada em redes neurais artificiais. Está técnica é considerada um ramo promissor da inteligência artificial e tem atraído muitos estudiosos interessados. (HUANG; CHAI; CHO, 2020). *Deep learning* tem sido amplamente utilizada em visão computacional, processamento de linguagem natural e reconhecimento áudio-visual. O sucesso da *deep learning* como uma técnica de processamento de dados tem despertado o interesse das comunidades acadêmicas. Dada a proliferação das *fintechs* nos últimos anos, o uso da *deep learning* passa a ser predominante em finanças e serviços bancários. (HUANG; CHAI; CHO, 2020). Diferente das redes neurais artificiais do passado, a moderna *deep learning* fornece estabilidade, generalização e escalabilidade de treinamento com *big data*. (CERVELLÓ-ROYO; GUIJARRO, 2020).

Sumarlinda, Supriyanto e Putri, (2020) testaram o poder de previsão da *deep learning* utilizando o método *long short-term memory* (LSTM) para prever o preço de empresas listadas no índice LQ45 (Jakarta Stock Exchange LQ45 Index). Utilizaram como amostra os preços de fechamento diário referentes ao período de 22 de janeiro de 2018 a 31 de outubro de 2019 de duas empresas listadas no índice (ADHI - Adhi Karya (Persero) Tbk e BBRI - Bank Rakyat Indonesia (Persero) Tbk). Os autores concluíram que utilizando o modelo foi possível prever com acurácia o preço das duas empresas, obtendo um *root-mean-square deviation* (RMSE) de 0.023789 para a ADHI e um RMSE de 0.09807 para a BBRI.

Paiva et al., (2019) criaram um modelo de investimento híbrido utilizando *support vector machine* (SVM) e *mean-variance* (MV). O objetivo foi construir um modelo que servisse de ferramenta no processo decisório para *day trading*. Os autores utilizaram como amostra os dados de empresas listadas no índice Bovespa entre junho de 2001 e dezembro de 2016. O modelo apresentou bons resultados até certo nível de custos de transações considerado pelos autores sendo retorno acumulado de 2956,32% quando consideraram 0,05 bps de custos de transação, retorno acumulado de 2289,09% quando foi considerado 0,10 bps. No entanto, quando consideraram 0.50 bps e 1,00 bps o retorno caiu drasticamente, chegando a perda de 61,74% no caso de 1,00 bps.

Pauli, Kleina e Bonat, (2020) verificaram que mesmo em tempos de pandemia de COVID-19 é possível aplicar *machine learning* para prever o retorno das ações. Os autores utilizaram cinco arquiteturas de redes neurais para prever o retorno das seis ações mais negociadas na B3 de março de 2019 a abril de 2020. As redes forneceram ajuste adequado, previsões razoáveis e bons intervalos de confiança. Redes neurais artificiais são efetivas para lidar com dados não lineares. Além disso, quanto maior o volume de dados utilizado, as redes tendem a ser melhores treinadas e apresentam melhores resultados. (PAI; ILANGO, 2020).

De acordo com Yu et al., (2020) muitos fatores que afetam os preços das ações tem características redundantes e não lineares. Sendo assim, modelos clássicos de redução de mapeamento linear, como por exemplo *principal component analysis* (PCA) e *linear discriminant analysis* (LDA) não apresentam bons resultados para problemas não lineares.

Nesse sentido, os autores propuseram a utilização em conjunto de dois modelos, o *local linear embedding* (LLE) que é um algoritmo de redução dimensional, com o objetivo de reduzir a dimensão dos fatores que afetam o preço das ações e, posteriormente, os dados gerados neste modelo foram utilizados como *input* em uma rede neural com o algoritmo *back propagation* (BP). Desta forma o modelo foi chamado de LLE-BP e mostrou efetivo e viável na previsão dos preços de ações.

Conforme observou-se através dos parágrafos anteriores, inúmeras abordagens baseadas em *machine learning* têm sido utilizadas para prever o retorno das ações. No entanto, para alcançar o objetivo do presente artigo pretendeu-se ir além de prever os retornos e utilizar também uma técnica de alocação dos recursos visando rentabilidade e proteção. Esse modelo é conhecido no mercado como um modelo híbrido.

Wang et al., (2020b) propuseram um modelo híbrido permitindo que eles aproveitassem as vantagens de dois modelos muito conhecidos para criar carteiras de investimentos. Esses modelos foram trabalhados em duas etapas sendo que, a primeira etapa utilizou o modelo *long short-term memory* (LSTM) desenvolvido por Hochreiter e Schmidhuber (1997) para selecionar ações com as melhores previsões de retornos. Após selecionadas as melhores ações implementou-se uma segunda etapa no processo de formação da carteira utilizando-se o modelo de média variância proposto por Markowitz (1952) para obter a proporção de capital alocado para cada uma das ações selecionadas. Os resultados gerados pelo modelo proposto foram comparados a outros tradicionais modelos utilizados para previsão, como por exemplo, SVM, RAF, DNN e ARIMA. O retorno acumulado no período de 1994 a 2018 mostrou resultado bastante superior aos modelos utilizados como base na comparação.

Assim como o artigo de Wang et al., (2020b) o presente estudo também consistiu em duas etapas. Na primeira etapa utilizou-se o algoritmo *random forest* para classificação das ações e, na segunda etapa, as ações foram alocadas nas carteiras *long and short*. Em um estudo realizado por Harvey et al (2019) foram testadas importantes estratégias no mercado americano com o objetivo de verificar qual é mais eficiente para mitigar perdas durante tempos de crise ou de recessão. Além de testar estratégias clássicas de proteção, que são efetivas, mas que correm parte substancial dos ganhos, buscaram testar duas estratégias diferentes visando minimizar os custos de proteção. A primeira, estratégia de *momentum* de séries temporais que apresentou bom desempenho tanto em momentos de crise quanto de recessão. Segundo, estratégia *long and short* em ações, utilizando diferentes métricas de qualidade para classificar as empresas transversalmente. Os melhores resultados foram apresentados por esta última estratégia, que se mostraram satisfatórios ao longo do período estudado (1985-2018), em especial nos momentos de crise e recessão.

As estratégias de investimento *long and short* vêm crescendo e se popularizando, pois permitem trabalhar com um número grande de ativos de forma mais eficiente que os modelos *long-only* (portfólios nos quais são alocadas somente ações compradas). (GRINOLD; KAHN, 2000). Lewin e Campani (2020) utilizaram na alocação de carteiras um modelo introduzido por Hamilton (1989) chamado cadeias de Markov. Esse modelo possibilita que se capture as probabilidades de mudanças entre os estados da economia. O modelo que não permitiu vendas a descoberto (*long only*) gerou um prêmio de risco anualizado de 4,66% com volatilidade de 7,62 e índice de Sharpe de 0,61%. O modelo com vendas a descoberto (*long and short*) apresentou resultado anualizado significativamente maior sendo o prêmio de 36,90%, volatilidade de 33,33 e índice de Sharpe de 1,11%. Apesar do alto prêmio de risco anual da carteira, nota-se também grande volatilidade.

Beaver e McNichols e Price (2016) apontam que apesar da estratégia *long and short* por si só não contribuir com um bom desempenho na carteira, a proteção contra uma direção eventualmente desfavorável no mercado, a minimização do custo de investimento e a

diversificação dessa estratégia fazem com que ela seja muito interessante e adotada por investidores e gestores de fundos de investimento.

Por fim, ainda como medida de proteção para carteira, especialmente tendo em vista as turbulências vividas no mercado em 2020 devido a pandemia do COVID-19, foi alocado um percentual de ouro na carteira. Destaca-se que o ouro é utilizado por muitos investidores como forma de proteção para as carteiras de investimento e que, portanto, justifica-se sua influência nas presentes carteiras. (KOUSAR; BATOOL, 2019). O ouro, seja físico ou por negociação eletrônica, é um dos ativos mais utilizados para proteção de capital. (HARVEY et al., 2019). (NAWAZ; AZAM; ASLAM, 2020). (SAMUEL, 2020). (ALSHAMMARI; ALTARTURI; SAITI, 2020).

A combinação de diferentes ativos em diferentes momentos econômicos proporciona, conforme largamente estudado em finanças, resultados positivos advindos da diversificação. (HARVEY et al., 2019). (NAWAZ; AZAM; ASLAM, 2020). (SAMUEL, 2020). (ALSHAMMARI; ALTARTURI; SAITI, 2020). (KOUSAR; BATOOL, 2019). Apesar de muitos estudos confirmarem o papel do ouro como um ativo de proteção do capital, há ainda pesquisadores que não encontram os mesmos resultados, como por exemplo Maghyereh e Abdoh (2020).

### 3 METODOLOGIA

Para atingir o objetivo do estudo utilizou-se o algoritmo de *machine learning* chamado *random forest*. Ele foi aplicado com a intenção de relacionar as variáveis independentes com a dependente e expor a probabilidade de retornos positivos no mês subsequente aos dados coletados. As ações posteriormente foram alocadas através de um método conhecido como *long and short*. Por fim foi alocado na carteira uma parcela de ouro, que é um ativo de proteção muito utilizado em tempos de crise e incerteza.

No contexto das previsões um dos grandes desafios é reconhecer a relação entre as informações passadas e as informações futuras. Geralmente a previsão do preço de ações é feita utilizando modelos de estatística, matemática, economia e *machine learning* de maneira geral. (JAYAPALAN; SOMASUNDARAM, 2020). Há forte evidência que uma vez que um padrão de comportamento é identificado é possível delinear um modelo de previsão. (PAIVA et al., 2019).

Foram coletadas informações mensais de todas as empresas listadas na B3 referentes ao período de janeiro de 1995 a maio de 2020. Quanto maior o volume de dados coletados, mais acurados se tornam os modelos baseados em *machine learning*. (FISCHER; KRAUSS, 2018) (CERVELLÓ-ROYO; GUIJARRO, 2020). Iniciou-se a criação das carteiras a partir do primeiro mês de 2009 e finalizou-se em junho de 2020. O modelo preditivo foi atualizado mensalmente com o objetivo de integrar os comportamentos mais atuais do mercado, ou seja, para cada um dos 138 meses foi rodado um modelo de treino e teste, conforme apresentado no esquema a seguir.

**Tabela 1 - Períodos de treino e teste do algoritmo**

<b>modelo</b>	<b>treino</b>	<b>teste</b>
1	01/1995 a 12/2008	01/2009
2	01/1995 a 01/2009	02/2009
3	01/1995 a 02/2009	03/2009
4	01/1995 a 03/2009	04/2009
...	...	...
138	01/1995 a 05/2020	06/2020



Servindo-se da Tabela 1 é possível exemplificar o processo de treino e teste. O primeiro modelo apresentado na tabela foi treinado de 01/1995 a 12/2008. Com base nesse treino o modelo preditivo apresenta qual a probabilidade do retorno em 01/2009 ser 1 (positivo) e, a partir daí as ações são alocadas (ou não), conforme estratégia apresentada. Visualiza-se ainda na mesma tabela que o mês previsto é sempre adicionado posteriormente como um mês de treino na previsão dos próximos meses. Conforme exposto anteriormente, esse processo foi repetido 138 vezes formando, dessa forma, 138 modelos preditivos e as carteiras mensais.

A partir das probabilidades são formadas as carteiras *long and short*. O lado *long* da carteira alocou as 15 maiores probabilidades de retornos positivos, desde que essas probabilidades fossem no mínimo 60%. O lado *short* da carteira alocou as 15 menores probabilidades desde que essas probabilidades fossem no máximo 30%. Enfatiza-se que esse procedimento foi repetido em cada um dos 138 meses do período estudado, de maneira automatizada, por meio do software R.

A variável dependente empregada no trabalho é uma *dummy* que recebeu 1 se o retorno posterior a divulgação dos dados foi positivo e recebeu 0 se o retorno no mês posterior a divulgação dos dados foi negativo. O modelo preditivo tem como intuito, tendo em vista o comportamento das variáveis independentes, verificar qual a probabilidade de o retorno ser 1.

As variáveis utilizadas no estudo foram levantadas a partir dos estudos anteriores. Utilizou-se não só variáveis exógenas, mas também endógenas, com o propósito de detectar através das previsões movimentos ocasionados por mudanças nas empresas e também no mercado de maneira geral. A utilização de variáveis macroeconômicas tem ganhado destaque em estudos de alocação de portfólios, bem como os índices representativos de mercado e o preço de commodities, que refletem o comportamento, usualmente, da economia mundial. (SETIADI; MASDUPI, 2019). As variáveis independentes são apresentadas na Tabela 2.

**Tabela 2 - Variáveis independentes**

Variável	Descrição
preço	preço da ação no último pregão do mês
pe	preço no último pregão do mês sobre lucro do último trimestre
ep	lucro do último trimestre sobre preço de fechamento do último pregão do mês
pebitda	preço sobre EBITDA
pfcf	preço sobre fluxo de caixa livre
pbv	preço sobre patrimônio líquido
presence	presença em bolsa
retorno	retorno da ação defasado de 1 a 12 meses
naics	representando o setor no qual a empresa está enquadrada
oz1d	retorno do ouro no mês
ibxx	retorno do IBRX Brasil no mês
lft	LFT acumulada no mês.
dolar	retorno do dólar no mês
wti	retorno do petróleo no mês
djia	retorno do índice Dow Jones
nasdaq	retorno do índice NASDAQ
sp500	retorno do índice S&P 500
rcmt10	Taxa do tesouro americano 10 anos
t-bond10	T-Bond de 10 anos

Fonte: variáveis utilizadas no trabalho observadas em estudos anteriores

Através das variáveis independentes, apresentadas por meio da Tabela 2, procurou-se abranger tanto características isoladas das empresas quanto características do mercado. Isso porque os preços das ações são influenciados por diversos fatores, tanto internos quanto externos.

Um dos modelos de *machine learning* que vem sendo amplamente estudados e tem apresentado resultados positivos tanto para regressões quanto para classificações é chamado de *random forest*, que é considerado um dos modelos mais populares de árvores de decisão. *Random forest* é um classificador que consiste em uma coleção de classificadores estruturados em árvores  $\{h(\mathbf{x}, \theta_k), k = 1, \dots\}$  onde  $\{\theta_k\}$  são vetores aleatórios independentes e identicamente distribuídos e cada árvore emite um voto unitário para a classe mais popular na entrada  $\mathbf{x}$ . (BREIMAN, 2001). A seguir é apresentado um sumário de como o algoritmo *random forest* trabalha:

- 1) Seja  $N$  = número de observações,  $n$  = número de árvores de decisão (entrada do usuário) e  $M$  = número de variáveis no conjunto de dados.
- 2) Escolha  $i$ , subconjunto de  $m$  variáveis de  $M$ , onde  $m \ll M$ , e construa  $n$  árvores de decisão usando um conjunto aleatório de  $m$  variável.
- 3) Cresça cada árvore o máximo possível.
- 4) Use a votação majoritária para decidir a classe da observação.

Um subconjunto escolhido aleatoriamente de  $N$  observações sem substituição (normalmente  $2/3$ ) é usado para construir cada árvore de decisão. (RAMASUBRAMANIAN; SINGH, 2016).

No presente artigo utilizou-se todos os valores *default* do algoritmo *random forest*, com exceção do número de árvores que se aumentou de 500 para 1500 obtendo melhores níveis de acurácia. As carteiras foram avaliadas por meio do retorno médio mensal, retorno acumulado, risco (desvio padrão) e índice de Sharpe. Este índice é calculado dividindo-se a média do retorno excedente pelo desvio padrão da carteira. Dados os valores praticados no mercado brasileiro no que se refere a taxa de juros em alguns períodos, optou-se por utilizar o índice de Sharpe modificado considerando-se somente a média dos retornos dividida pelo desvio padrão. Este índice é um dos mais utilizados na avaliação de risco-retorno de portfólios. (PAI; ILANGO, 2020).

O artigo recorreu a duas técnicas conhecidas para proteção: (1) alocação de um percentual da carteira em ouro, que é largamente reconhecido como um ativo de proteção (HARVEY et al., 2019). (NAWAZ; AZAM; ASLAM, 2020). (SAMUEL, 2020). (ALSHAMMARI; ALTARTURI; SAITI, 2020). (2) alocação das ações no formato *long and short*, que protege a carteira das mudanças de tendências no mercado, em especial em momentos de crise com alta volatilidade. Esse formato minimiza os custos de investimentos promovendo recursos com as ações vendidas a descoberto. (GRINOLD; KAHN, 2000), (LEWIN; CAMPANI, 2020), (BEAVER; MCNICHOLS; PRICE, 2016).

Não pretende-se criar um modelo fixo com capacidade preditiva para o longo prazo, ou seja, com o modelo travado em determinado momento do tempo, mas almeja-se construir um modelo dinâmico que seja atualizado mensalmente com as informações de mercado e que, portanto, pode ser utilizado por longo prazo. Tendo em vista a dinâmica do mercado brasileiro de ações, comportar as mais recentes atualizações faz-se importante para manter a acurácia do modelo.

## 4 RESULTADOS

As tabelas a seguir apresentam as carteiras formadas com base na metodologia proposta. São apresentadas separadamente as carteiras *long* e *short*, posteriormente elas são consolidadas e por fim foi inserido um percentual de ouro nas carteiras.

**Tabela 3 - Carteira *long***

Ano	Desvio	Quant. ações	Retorno	Ibov.	Ret. mensal	Sharpe	Acurácia	Meses positivos
2009	5,1%	9	154%	83%	9%	174,0%	71%	91%
2010	3,1%	9	39%	1%	3%	92,0%	64%	83%
2011	4,8%	8	5%	-18%	1%	10,9%	55%	45%
2012	3,3%	8	61%	7%	4%	125,8%	75%	83%
2013	4,6%	9	17%	-15%	2%	32,9%	53%	55%
2014	5,0%	8	2%	-3%	0%	5,1%	49%	58%
2015	6,9%	7	-20%	-13%	-2%	-25,6%	46%	45%
2016	6,8%	8	55%	39%	5%	76,8%	64%	67%
2017	8,7%	8	142%	27%	9%	99,7%	70%	91%
2018	5,9%	8	21%	15%	2%	38,8%	60%	78%
2019	5,9%	6	101%	32%	7%	114,6%	71%	82%
2020	15,0%	5	-18%	-18%	-4%	-24,8%	54%	50%

Observações: Desvio refere-se ao desvio padrão dos retornos mensais. Quant. ações é a quantidade média de ações alocadas na carteira por mês. Retorno refere-se ao retorno acumulado das carteiras mensais no ano. Ibov refere-se ao retorno do Ibovespa, principal índice de ações brasileiro. Ret mensal refere-se a média mensal dos retornos no ano. Sharpe é o retorno ajustado ao risco. Acurácia é o percentual de acertos apurados através de uma matriz de confusão, em outras palavras, quanto dos resultados previstos como positivos (negativos) foram de fato positivos (negativos). Meses positivos são os meses que fecharam com retorno maiores que zero.

Optou-se no presente artigo por mostrar as tabelas *long* e *short* de forma separada, isto é, sem consolidar os retornos para que seja possível visualizar quais são os benefícios alcançados com essa estratégia no período. Observa-se por meio da Tabela 3 que o melhor resultado foi alcançado no ano de 2009, quando o Ibovespa também alcançou seu melhor resultado no período. Esse resultado se deu principalmente pela recuperação que houve no mercado após a crise do *subprime*, ocorrida entre 2007 e 2008. A carteira *long* apresentou resultado positivo em 10 dos 12 anos levantados, o que corresponde a 83% do período. Apresentou ainda resultado maior que o Ibovespa em todos os anos, exceto 2015. Ressalta-se que o resultado de 2020 foi apurado até 06/2020 e apresenta prejuízo de 18%, coincidentemente o mesmo do Ibovespa.

**Tabela 4 - Carteira *short***

Ano	Desvio	Quant. Ações	Retorno	Ibov.	Ret. Mensal	Sharpe	Acurácia	Meses positivos
2009	5,7%	10	-20%	83%	-4%	-70,8%	40%	20%
2010	0,0%	5	1%	1%	1%		60%	100%
2011	8,5%	6	7%	-18%	2%	21,1%	51%	60%
2012	9,7%	3	28%	7%	5%	56,5%	72%	60%
2013	21,3%	7	-54%	-15%	-3%	-15,4%	60%	64%
2014	9,4%	7	13%	-3%	1%	15,5%	72%	58%
2015	11,6%	10	60%	-13%	5%	47,2%	74%	60%
2016	20,9%	6	49%	39%	6%	28,5%	65%	70%
2017	10,2%	7	32%	27%	3%	28,2%	76%	67%

2018	11,6%	7	2%	15%	1%	7,5%	58%	55%
2019	25,2%	5	-73%	32%	-8%	-33,5%	61%	56%
2020	26,6%	8	36%	-18%	20%	73,5%	67%	50%

Observações: Desvio refere-se ao desvio padrão dos retornos mensais. Quant. ações é a quantidade média de ações alocadas na carteira por mês. Retorno refere-se ao retorno acumulado das carteiras mensais no ano. Ibov refere-se ao retorno do Ibovespa, principal índice de ações brasileiro. Ret mensal refere-se a média mensal dos retornos no ano. *Sharpe* é o retorno ajustado ao risco. Acurácia é o percentual de acertos apurados através de uma matriz de confusão, em outras palavras, quanto dos resultados previstos como positivos (negativos) foram de fato positivos (negativos). Meses positivos são os meses que fecharam com retorno maiores que zero.

A Tabela 4 apresenta o lado vendido da carteira, em outros termos, o lado *short*. É possível observar que o desvio padrão do lado *short* é bastante superior ao que pode ser visto no lado *long* da carteira. Observa-se ainda que em alguns anos houve perda substancial do patrimônio, como por exemplo os anos 2013 e 2019. Uma sugestão nesse ponto seria inserir uma estratégia de *stop loss* para que as perdas fossem limitadas a determinado patamar. Em função da carteira *short* se beneficiar com a queda das ações, observa-se para 2020 resultado positivo de 36%.

**Tabela 5 - Carteira long and short consolidada**

Ano	Desvio	Quant. Ações	Retorno	Ibov.	Ret. Mensal	<i>Sharpe</i>	Acurácia	Meses positivos
2009	5,8%	19	100%	83%	6%	106%	55%	83%
2010	3,1%	14	38%	1%	3%	89%	62%	83%
2011	4,7%	14	8%	-18%	1%	15%	53%	50%
2012	4,0%	11	65%	7%	4%	108%	73%	83%
2013	9,5%	15	-16%	-15%	-1%	-10%	56%	67%
2014	4,6%	15	9%	-3%	1%	19%	61%	58%
2015	4,9%	17	22%	-13%	2%	37%	60%	58%
2016	9,4%	14	66%	39%	5%	54%	64%	73%
2017	5,3%	15	89%	27%	6%	106%	73%	83%
2018	8,4%	15	3%	15%	1%	8%	59%	67%
2019	10,2%	11	-6%	32%	0%	1%	66%	67%
2020	8,7%	13	13%	-18%	3%	32%	60%	60%

Observações: Desvio refere-se ao desvio padrão dos retornos mensais. Quant. ações é a quantidade média de ações alocadas na carteira por mês. Retorno refere-se ao retorno acumulado das carteiras mensais no ano. Ibov refere-se ao retorno do Ibovespa, principal índice de ações brasileiro. Ret mensal refere-se a média mensal dos retornos no ano. *Sharpe* é o retorno ajustado ao risco. Acurácia é o percentual de acertos apurados através de uma matriz de confusão, em outras palavras, quanto dos resultados previstos como positivos (negativos) foram de fato positivos (negativos). Meses positivos são os meses que fecharam com retorno maiores que zero.

Através da Tabela 5 observa-se a consolidação dos dois lados. De maneira geral esse tipo de alocação tem os lados em equilíbrio, isto significa, a quantidade total vendida a descoberto é utilizada na compra. No entanto, considerando que para entrar nas carteiras era necessário atender aos critérios de probabilidade definidos, nem sempre foi possível manter um equilíbrio entre os lados. Observa-se, em especial no ano de 2020 onde o mercado sofreu com alta volatilidade em função da pandemia de COVID-19, a alocação por meio do *long and short* foi efetiva em reduzir o desvio padrão. Conforme apresentado, o desvio padrão de 2020 foi de 15% para a carteira *long*, 26,6% para a carteira *short* e 8,7% para a carteira *long and short*. Todavia é importante destacar que esse comportamento de redução do desvio padrão não foi encontrado em outros anos, como por exemplo o período de 2016 a 2019. Percebe-se, portanto, que a efetividade da minimização de riscos propostas por esse tipo de carteira pode não ocorrer em momentos com tendência de alta, conforme ocorreu nos anos exemplificados.

**Tabela 6 - Carteira *long and short* consolidada com alocação de 1/3 de ouro**

Ano	Desvio	Quant. Ações	Retorno	Ibov.	Ret. Mensal	Sharpe	Meses positivos
2009	3,6%	19	42%	83%	3%	85,7%	83%
2010	2,3%	14	35%	1%	3%	109,8%	92%
2011	3,8%	14	13%	-18%	1%	28,5%	58%
2012	2,1%	11	44%	7%	3%	147,1%	92%
2013	6,4%	15	-13%	-15%	-1%	-14,1%	42%
2014	2,9%	15	11%	-3%	1%	30,7%	67%
2015	3,9%	17	26%	-13%	2%	51,3%	50%
2016	6,6%	14	32%	39%	3%	38,6%	67%
2017	3,5%	15	61%	27%	4%	116,4%	83%
2018	5,5%	15	11%	15%	1%	19,2%	75%
2019	7,2%	11	6%	32%	1%	10,2%	75%
2020	5,0%	13	31%	-18%	5%	92,9%	67%

Observações: Desvio refere-se ao desvio padrão dos retornos mensais. Quant. ações é a quantidade média de ações alocadas na carteira por mês. Retorno refere-se ao retorno acumulado das carteiras mensais no ano. Ibov refere-se ao retorno do Ibovespa, principal índice de ações brasileiro. Ret mensal refere-se a média mensal dos retornos no ano. Sharpe é o retorno ajustado ao risco. Meses positivos são os meses que fecharam com retorno maiores que zero.

Por fim, apresentou-se a Tabela 6 na qual considera-se a alocação de 1/3 de ouro na carteira, por meio do contrato OZ1D. Constata-se que, especialmente em 2020, foi possível reduzir ainda mais o desvio padrão e incrementar o resultado. Em outros anos também é possível observar diminuição no desvio padrão, confirmando o papel do ouro como um ativo de proteção. Ao considerar os valores anuais, somente o ano de 2013 apresentou resultado negativo representando 8% do período, melhor dizendo, em 92% dos retornos anuais o resultado foi positivo. A seguir, a Tabela 7 apresenta os resultados consolidados do período.

**Tabela 7 - Resultado consolidado de 2009 a 2020**

	<i>long</i>	<i>short</i>	<i>long and short</i>	<i>long and short com ouro</i>	Ibov.
Desvio padrão	7%	16%	7%	5%	6%
Retorno acumulado	4073%	-34%	1795%	1136%	153%
Média	3%	1%	2%	2%	1%
Sharpe	47%	8%	34%	40%	14%
Acurácia	61%	65%	63%		
Meses positivos	70%	59%	70%	71%	57%

Conforme observa-se por intermédio da Tabela 7, onde apresenta-se os resultados consolidados, quando se considera o período total o comportamento muda levemente. Acredita-se que, nesse sentido, o modelo deve ser aperfeiçoado considerando diferentes pesos aos investimentos em diferentes momentos da economia. Conforme exposto na metodologia, considerou-se o peso igual para todas as ações distribuídos com os mesmos pesos para as diferentes carteiras, inclusive para o ouro. A primeira coluna da tabela apresenta o consolidado geral do período considerando somente as ações compradas, isto é, o lado *long*. Verifica-se que a carteira de forma isolada gerou o melhor dos retornos, inclusive com o melhor retorno ajustado ao risco que foi de 47%. A carteira *short* gerou prejuízo de 34%, no entanto com *sharpe* positivo em 8% em função das médias. Ressalta-se que um dos objetivos da alocação *long and short* é minimizar o risco, objetivo este que não foi alcançado, levando em consideração o período todo, tendo em vista que o desvio padrão permaneceu em 7% após a consolidação das

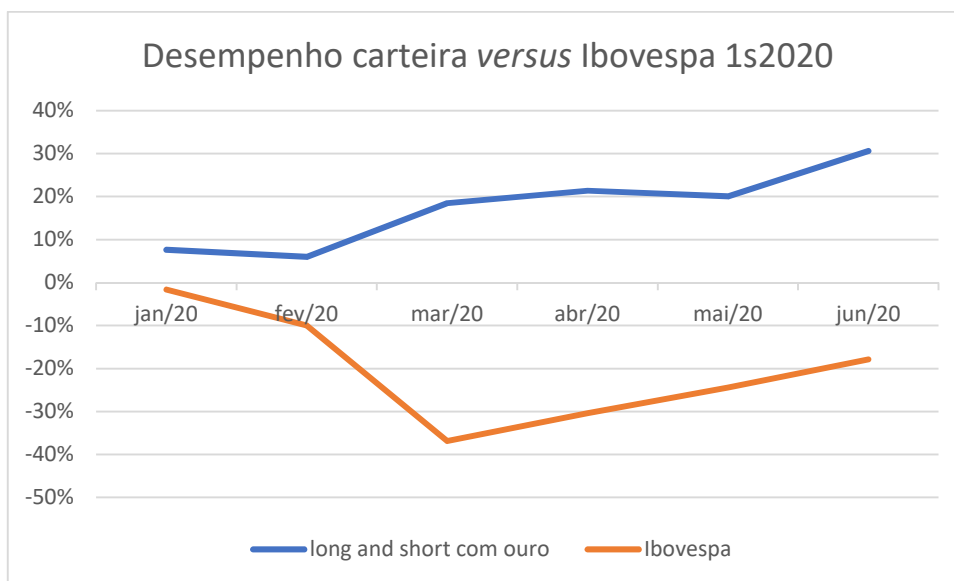
carteiras, ou seja, o mesmo desvio do lado *long* isolado. Apesar disso vale destacar que outro objetivo do lado *short* é diminuir os custos de investimento, promovendo através da venda a descoberto recursos para outras alocações. Na penúltima coluna inseriu-se 1/3 de ouro nas carteiras, ficando então 1/3 *long*, 1/3 *short* e 1/3 ouro. Nesse caso é possível observar que o risco da carteira diminuiu de 7% para 5%. A última coluna apresenta os resultados do Ibovespa para o período. Em todos os quesitos avaliados o índice Bovespa foi inferior a carteira *long and short* com alocação de ouro.

Acredita-se que que o poder do ouro e da alocação *long and short* mostra melhores resultados principalmente em momentos de grande incerteza, conforme exposto por Harvey et al (2019). Deste modo consolidou-se na Tabela 8 os dados referentes ao primeiro semestre de 2020.

**Tabela 8 - Resultado consolidado referente ao primeiro semestre de 2020**

	<i>long</i>	<i>short</i>	<i>long and short</i>	<i>long and short com ouro</i>	Ibov.
Desvio padrão	15%	27%	9%	5%	14%
Retorno acumulado	-18%	36%	13%	31%	-18%
Média	-4%	20%	3%	5%	-2%
<i>Sharpe</i>	-25%	74%	32%	93%	-15%
Acurácia	54%	67%	60%		
Meses positivos	50%	50%	60%	67%	50%

Conforme visualiza-se por meio da Tabela 8 o desvio das carteiras *long* e *short* isoladas aumentou bastante, sendo 15% para a *long* e 27% para *short*, confirmando a alta volatilidade do mercado no período. Contudo, quando se consolida ambas o desvio diminuiu para 9%, mostrando nesse caso o poder das carteiras *long and short* na redução de risco em momentos de incerteza. Destaca-se que, conforme apresentado na terceira coluna por meio dos lados integrados, mesmo em um momento de maior volatilidade e incerteza foi possível obter um retorno de 13% utilizando esse modelo de alocação. Quando se inseriu o ouro o resultado surpreendeu ainda mais, diminuindo o desvio padrão de 9% para 5% e aumentando o retorno de 13% para 31% no primeiro semestre de 2020. Salienta-se que o resultado ajustado ao risco também foi superior com a alocação do ouro neste período.



**Figura 1 - Retornos acumulados mensais referentes a 1s2020**

Por fim, através da Figura 1 demonstrou-se o retorno acumulado mês a mês das carteiras *versus* Ibovespa para o primeiro semestre de 2020. Nela pode-se observar como o índice sofreu com a insegurança do mercado marcando seus valores mínimos em março e posteriormente recuperando. Quanto a carteira, a mesma se mostrou bastante estável, mesmo diante das incertezas geradas pelo COVID-19.

No cenário de alocação das ações para as carteiras, sugere-se para estudos futuros calcular os custos de transações bem como utilizar algum critério de *stop loss*. Supõe-se que um critério de *stop loss* deve incrementar ainda mais os resultados, principalmente nas perdas ocasionadas no lado *short* da carteira entre 2009 e 2019. Ainda no cenário de alocação, julga-se interessante promovê-la com diferentes pesos, de acordo com os momentos do mercado e da economia, conforme metodologia aplicada por Lewin e Campani (2020). Na contextura de *machine learning* sugere-se testar outros algoritmos, como por exemplo os baseados em *Reinforcement Learning*.

## 5 CONSIDERAÇÕES FINAIS

O trabalho alcançou o objetivo e mostrou ser possível construir carteiras de investimentos rentáveis e seguras mesmo em tempos de maior incerteza e volatilidade, como no caso de 2020 onde presenciou-se a pandemia de COVID-19. Notou-se que o modelo é efetivo em momentos de crise, mas deve ser ajustado em momentos de maior previsibilidade, como o período de 2016 a 2019 no qual a bolsa esteve em tendência de alta.

No que se refere ao desempenho das carteiras no período total do trabalho, o melhor resultado considerando retorno acumulado, média mensal e *sharpe* foi encontrado na carteira *long only*. O menor desvio padrão e, portanto, e menor risco sob esta perspectiva foi encontrado na carteira *long and short* com alocação de ouro, que também apresentou resultado ligeiramente melhor em termos de meses positivos, mostrando maior estabilidade.

Ao isolar o primeiro semestre de 2020 o resultado sofre alterações. Notou-se que nesse período a carteira *long and short* com alocação de ouro obteve melhores resultados em praticamente todos os quesitos, como por exemplo o desvio padrão menor, retorno acumulado maior, *sharpe* maior e percentual de meses positivos também maior.

## REFERÊNCIAS

ALSHAMMARI, A. A.; ALTARTURI, B.; SAITI, B. The impact of exchange rate , oil price and gold price on the Kuwaiti stock market : a wavelet analysis. **The European Journal of Comparative Economics**, v. 17, n. 1, p. 31–54, 2020.

BEAVER, W.; MCNICHOLS, M.; PRICE, R. The costs and benefits of long-short investing: A perspective on the market efficiency literature. **Journal of Accounting Literature**, v. 37, p. 1–18, 2016.

BREIMAN, L. Random forests. **Machine Learning**, 2001.

CERVELLÓ-ROYO, R.; GUIJARRO, F. Forecasting stock market trend: a comparison of machine learning algorithms. **Finance, Markets and Valuation**, v. 2020, n. 1, p. 37–49, 2020.

CHEN, L. Using Machine Learning Algorithms on Prediction of Stock Price. **Journal of Modeling and Optimization**, v. 12, n. 2, p. 84–99, 2020.

FISCHER, T.; KRAUSS, C. Deep learning with long short-term memory networks for financial market predictions. **European Journal of Operational Research**, v. 270, n. 2, p.

654–669, 2018.

GRINOLD, R. C.; KAHN, R. N. The Efficiency Gains of Long-Short Investing. **Financial Analysts Journal**, v. 56, n. 6, p. 40–53, 2000.

HARVEY, C. R. et al. The Best of Strategies for the Worst of Times: Can Portfolios be Crisis Proofed? **The Journal of Portfolio Management**, 2019.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997.

HUANG, J.; CHAI, J.; CHO, S. Deep learning in finance and banking: A literature review and classification. **Frontiers of Business Research in China**, v. 14, n. 1, 2020.

ISMAIL, M. S. et al. Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange. **Applied Soft Computing Journal**, v. 93, p. 106422, 2020.

JAYAPALAN, V.; SOMASUNDARAM, K. Machine Learning based comparison of financial forecasting methods. **International Journal of Advanced Science and Technology Vol.**, v. 29, n. June, p. 8902–8907, 2020.

JIANG, M. et al. The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. **Annals of Operations Research**, 2020.

JIAO, Y.; MASSA, M.; ZHANG, H. Short selling meets hedge fund 13F: An anatomy of informed demand. **Journal of Financial Economics**, v. 122, n. 3, p. 544–567, 2016.

JOSHI, H.; CHAUHAN, R. Determinants and Prediction Accuracy of Price Multiples for South East Asia : Conventional and Machine Learning Analysis. **Indonesian Capital Market Review**, v. 12, p. 42–54, 2020.

KOUSAR, A. S. AND S.; BATOOL, S. A. Impact of gold and oil prices on the stock market in Pakistan. **Journal of Economics, Finance and Administrative Science**, p. 47–63, 2019.

LANTZ, B. **Machine Learning with R**. Third edit ed. Birmingham B3 2PB, UK: Packt Publishing Ltd, 2019.

LEWIN, M.; CAMPANI, C. H. Gestão de Carteiras sob Múltiplos Regimes: Estratégias que Performam Acima do Mercado. **Revista de Administração Contemporânea**, v. 24, n. 4, p. 300–316, 2020.

LIN, Y.-C.; BLUM, J. FEATURE TRANSFORMATION AND SIMULATION OF SHORT TERM PRICE VARIABILITY IN REINFORCEMENT LEARNING FOR PORTFOLIO MANAGEMENT. **Society for Modeling & Simulation International (SCS)**, 2020.

MAGHYEREH, A.; ABDOH, H. Tail dependence between gold and Islamic securities. **Finance Research Letters**, 2020.

MARKOWITZ, H. Portfolio Selection. **The Journal of Finance**, v. 7, n. 1, p. 77–91, 1952.

NAWAZ, M. S.; AZAM, M.; ASLAM, M. Probable daily return on investments in gold. p. 47–54, 2020.

NIU, T. et al. Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. **Expert Systems with Applications**, v. 148,



2020.

PAI, N.; ILANGO, D. V. **Neural Network Model for Efficient portfolio Management and Time Series Forecasting**. Proceedings of the International Conference on Intelligent Computing and Control Systems. **Anais...**2020

PAIVA, F. D. et al. Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. **Expert Systems with Applications**, v. 115, p. 635–655, 2019.

PAULI, S. T. Z. DE; KLEINA, M.; BONAT, W. H. Comparing Artificial Neural Network Architectures for Brazilian Stock Market Prediction. **Annals of Data Science**, 2020.

RAMASUBRAMANIAN, K.; SINGH, A. **Machine learning using R**. [s.l: s.n.].

SAMUEL, S. S. A STUDY ON GROWTH OF GOLD ETFS : AS AN EFFECTIVE INVESTMENT TOOL OVER PHYSICAL GOLD. **UGC Care Journal**, v. 40, n. 40, p. 2328–2331, 2020.

SETIADI, J.; MASDUPI, E. **The Effect of Macroeconomic Variables on Banking Stock Price Index in Indonesia Stock Exchange**. 4th Padang International Conference on Education, Economics, Business and Accounting (PICEEBA-2 2019). **Anais...**2019

SUMARLINDA, S.; SUPRIYANTO, N.; PUTRI, D. K. LQ45 Stock Price Predictions Using The Deep Learning Method. **International Journal of Advanced Research and Publications**, v. 4, n. 4, p. 20–23, 2020.

TRATKOWSKI, G. **Construction of Investment Strategies for WIG20, DAX and Stoxx600 with Random Forest Algorithm**. Contemporary Trends and Challenges in Finance, Springer Proceedings in Business and Economics. **Anais...**Springer Nature Switzerland, 2020

WANG, S. et al. Stock selection strategy of A-share market based on rotation effect and random forest. **AIMS Mathematics**, v. 5, n. April, p. 4563–4580, 2020a.

WANG, W. et al. Portfolio formation with preselection using deep learning from long-term financial data. **Expert Systems with Applications**, v. 143, p. 113042, 2020b.

WU, X. et al. Adaptive stock trading strategies with deep reinforcement learning methods. **Information Sciences**, v. 538, p. 142–158, 2020.

YU, Z. et al. Stock price forecasting based on LLE-BP neural network model. **Physica A: Statistical Mechanics and its Applications**, v. 553, p. 124197, 2020.