

A Probabilidade de Negociações Informadas (PIN) versus *Disclosure* de Negociações dos *Insiders*: o Caso da JBS S.A. (JBSS3) com Modelos de *High-Frequency Trading*

Probability of Informed Trading (PIN) versus Disclosure of Insider Trading: the Case of JBS S.A. (JBSS3) with High-Frequency Trading Models

Glauco Graco Nóbrega Pordeus

Universidade Federal da Paraíba – UFPB – Brasil
glaucopordeus@outlook.com
ORCID: 0000-0002-2784-8752

Luiz Felipe de Araújo Pontes Girão

TradersClub – TC – Brasil
felipe.pontes@tc.com.br
ORCID: 0000-0002-3639-7340

Filipe Coelho de Lima Duarte

Universidade Federal da Paraíba – UFPB – Brasil
filipe_pb_duarte@hotmail.com
ORCID: 0000-0002-2038-4115

Submetido em 22/12/2020; Aprovado em 15/03/2021.

Resumo

Objetivo: Este trabalho tem como objetivo principal identificar operações de compra ou venda no ativo JBSS3, da empresa JBS S.A., e introduzir uma metodologia de análise de operações com base na *proxy* de informação assimétrica captada pela métrica de Probabilidade de Negociações Informadas (PIN). **Metodologia:** A JBS S.A. é usada como caso de análise nesta pesquisa, pois foi o 1º caso de prisão por *insider trading* no Brasil. Desta maneira, foram usados modelos matemáticos para a estimação da PIN para verificar se era possível se antecipar aos movimentos de uma pessoa com acesso a informações privilegiadas. **Principais resultados:** Os resultados encontrados apontam que a) o relatório obrigatório e mensal exigido pelo órgão regulador do mercado de capitais é pouco útil para proteger o investidor brasileiro; b) que mesmo com um modelo matemático avançado, não foi possível (em partes) obter retorno anormal antecipando as operações dos *insiders*, mas c) o modelo matemático da PIN gerou melhores resultados do que o relatório exigido pelo regulador. **Contribuições acadêmicas:** Foi possível inserir mais uma contribuição com relação ao uso de ferramentas computacionais na literatura sobre o mercado financeiro, como meio de promoção e aprimoramento da eficiência de mercado. **Contribuições práticas:** Investidores poderão se beneficiar com o uso dos resultados desta pesquisa, tanto para se protegerem da seleção adversa e do risco moral e o regulador do mercado de capitais poderá usar técnicas semelhantes para ampliar a fiscalização do mercado, além de ser mais um ponto de partida para a geração de modelos de *high-frequency trading*. **Palavras-chave:** Informação assimétrica. *Insider trading*. Aprendizado de máquina. *High-frequency trading*. PIN.

Abstract

Objective: This work has as main objective to identify the buy or sale procedure in the JBSS3 stock, from JBS S.A. and to introduce a methodology of analysis of operations based on the asymmetric information proxy captured by the Probability of Informed Trading (PIN). **Methodology:** JBS S.A. is used as case in this research as it was the first case of arrest for insider trading in Brazil. In this way, some mathematical models were used to estimate the PIN to check whether it was possible to anticipate the movements of a person with access to private information. **Main results:** The results indicate that a) the mandatory and monthly report required by the capital market regulator is of little use to protect the Brazilian investor, b) even with an advanced mathematical model, it was not possible to obtain abnormal returns anticipating insider operations, but c) the PIN mathematical model generated better results than the report required by the regulator. **Academic contributions:** It was possible to insert

another contribution regarding the use of computational tools in the literature on the financial market, as a means of promoting and improving market efficiency. **Practical contributions:** Investors who benefit from using the results of this research to protect themselves from adverse selection and moral hazard, and the capital market regulator can use similar techniques to expand market surveillance.

Keywords: Asymmetric information. Insider trading. Machine learning. Abnormal return. PIN.

1. Introdução

A homogeneidade informacional é um dos pressupostos básicos para a eficiência do mercado e, para o entendimento desse conceito, Fama (1970) propõe a Hipótese do Mercado Eficiente (HME) em que iniciou seu pensamento de que o mercado pode ser eficiente em três formas (fraca, semiforte e forte), que diferem entre si pelo tempo em que preços absorvem as informações relevantes sobre os ativos.

Ainda sobre o mercado financeiro e sua eficiência, Healy e Palepu (2001) afirmam que podem ocorrer falhas nos momentos em que os negócios forem efetivados pelos agentes compradores e vendedores. Uma grande parte dessas “falhas informacionais” poderia ser resolvida por uma boa *disclosure* corporativa, que é tratada como um meio importante de disseminação de informações ao mercado, sendo constituídas de previsões, relatórios da administração, demonstrações contábeis, negociações dos *insiders* (diretores, controladores etc.) e afins.

Todavia, nem todo mundo tem acesso a essas divulgações ao mesmo tempo, o que faz com que o mercado seja menos eficiente, ainda que Fama (1970) afirme, com a sua forma forte de eficiência, que é possível haver eficiência informacional, mesmo em um mercado em que as pessoas têm acesso assimétrico às informações das empresas investidas.

Um exemplo de falha da eficiência do mercado, na forma forte, ocorreu no dia 17 de maio de 2017, em que após o fechamento do pregão desse dia vieram à tona, por meio do portal da BBC, informações relevantes referentes à empresa brasileira JBS S.A. (seu *ticker* na B3 é JBSS3), informando que os controladores estariam envolvidos em casos de corrupção em conjunto com agentes da União Federal, informação esta que foi corroborada pela delação premiada divulgada pelo Supremo Tribunal Federal, dois dias depois, em 19 de maio de 2017, segundo o site BBC.

Além disso, de acordo com o relatório referente ao art. 11 da Instrução da Comissão de Valores Mobiliários (CVM) de número 358/2002 (ICVM 358), esses controladores venderam parte de suas ações dias antes do vazamento dessas informações, o que poderia, à época, ser um indício de uso de informações privadas (*insider trading* ilegal) para benefício próprio, por parte desses agentes.

Mais à frente, nessa linha do tempo, no dia 14 de setembro de 2017, Wesley Batista, um dos controladores da JBS S.A., foi preso de forma preventiva, sendo esta a 1ª prisão por *insider trading* no Brasil. Por esse motivo, o caso da JBS S.A. foi escolhido para testar as hipóteses que serão apresentadas mais a frente nesta introdução.

Com melhor condição informacional, o mercado como um todo tem a capacidade de se ajustar de forma mais eficiente, visto que o *delay* de informação exigida pela CVM pode chegar a até 40 dias, pois os controladores têm obrigatoriedade de apresentar o relatório de negociações dos *insiders* ao mercado até o décimo dia do mês subsequente ao mês de operação (ICVM 358). Ou seja, todos os acionistas e os brasileiros foram prejudicados nesse evento informacional da JBS S.A. e não havia nada que pudesse ser feito para se proteger do evento, a menos que eles tivessem modelos semelhantes aos que apresentamos nesta pesquisa.

Como exemplo, para que o leitor possa assimilar melhor os prazos e perda informacional, uma operação efetuada no dia 01/02/20XX poderá chegar ao mercado apenas no dia 10/03/20XX. O uso da tecnologia de informação, como enfatizamos no parágrafo anterior, pode ser uma saída para combater estas anomalias e trazer mais eficiência ao mercado.

O desenvolvimento tecnológico aplicado no mercado financeiro dessa maneira já é muito presente. Para ilustrar o que foi dito nos parágrafos anteriores, o uso de *High-Frequency Trading* (HFT) em Wall Street, no ano de 2009, representava em torno de 60% do volume negociado. Da parte dos operadores, o Medallion Fund obteve um rendimento médio de 35%, entre 2000 e 2010 (Aldridge, 2010). Esses modelos não precisam ser usados apenas para HFT, mas podem também ser usados pelo órgão regulador, além de ser possível uma combinação de HFT e *hedge*.

Sobre *hedge*, por exemplo, se o modelo proposto nesta pesquisa estiver correto e for de fato útil (é possível verificar isso na seção de resultados), a percepção de um volume anormal de vendas pode-

ria fazer o investidor antecipar o evento da delação premiada e entrar na ponta vendida da JBSS3, evitando os prejuízos que foram causados; ou até entrar na ponta comprada do dólar, reduzindo os prejuízos com as outras ações e até ganhando dinheiro com a moeda estrangeira. Sem um modelo semelhante ao usado nesta pesquisa, as pessoas só ficaram sabendo dos dias das negociações dos controladores da JBS S.A. tempos depois, quando não havia mais nada a ser feito.

Com base no que foi exposto até o momento, e com a ciência de que a expansão tecnológica é necessária e contribui para a qualidade das informações no mercado, o presente estudo busca, em seu objetivo geral, identificar se a Probabilidade de Negociação Informada (PIN) tem a capacidade de antecipar as informações apresentadas pelos agentes da JBS S.A. no relatório da do Art. 11 da Instrução CVM 358.

Além disso, este estudo propõe **(i)** verificar se é possível obter retorno acima do mercado por meio das informações obtidas no relatório do Art. 11 da ICVM 358; **(ii)** identificar se é possível obter retorno acima do mercado por meio dos parâmetros obtidos no modelo da PIN adaptado por Lin e Ke (2011); e **(iii)** analisar se é possível obter melhor desempenho financeiro (retorno) operando por meio da PIN quando comparado com as operações efetuadas com base na ICVM 358.

Para isso, foi aplicado um modelo de aprendizado de máquina com o objetivo de identificar pontos de compra e venda no ativo JBSS3, tendo como base a assimetria informacional obtida pela PIN. Essas operações foram comparadas com as operações efetuadas pelos agentes referidos à ICVM 358. Dessa forma, os resultados apontaram que as operações efetuadas com base no modelo de aprendizado de máquina obtiveram desempenho superior às operações efetuadas com base nas negociações dos *insiders* pela ICVM 358.

Esses resultados poderão ser úteis tanto para os investidores no mercado de capitais (que querem se proteger da seleção adversa e do risco moral, conforme informado nos primeiros parágrafos, além de tentarem ganhar dinheiro com modelos HFT) quanto para a CVM (que poderá implementar esse tipo de modelo HFT para identificar transações “anormais”), e também para a teoria financeira, com uma evidência adicional de que as informações privilegiadas, ou privadas, não são captadas normalmente pelos preços dos ativos, mas que, com modelos HFT, é possível captar esse tipo de comportamento – evidência contrária à forma forte da HME.

2. Revisão da literatura e desenvolvimento das hipóteses

A literatura é vasta no que diz respeito à eficiência de mercado. Em Callado (2009) são encontradas diversas referências no que diz respeito ao início da construção do pensamento de eficiência de mercado, no qual são citados os trabalhos de Campbell, Lo e MacKinlay (1997), Leroy (1989) e Ceretta (2001). Todavia, as pesquisas sobre o tema vieram muito antes, destacando-se a importância do estudo de Samuelson (1965), que introduziu a ideia de que os preços dos ativos têm movimentos aleatórios.

Na formulação da HME, Fama (1970) assume que na forma forte da eficiência do mercado, todas as informações, mesmo aquelas que ainda não são de domínio público, já se encontram precificadas; em sua forma semiforte, as informações são precificadas assim que são divulgadas ao público; por fim, em sua forma fraca, o autor coloca que as informações passadas ainda se mantêm absorvidas nos preços presentes e futuros.

Segundo Fama (1970), na HME, a função do mercado de capitais é disponibilizar um meio para que os recursos fluam entre os agentes superavitários e os agentes deficitários da economia. Seu funcionamento é baseado em três premissas-chaves: **(1)** todos os agentes interessados em investir aceitam que os preços dos ativos sejam influenciados pelas *disclosures*; **(2)** as negociações são feitas sem custo de transação; e **(3)** todas as informações disponíveis são gratuitas para todos os agentes do mercado.

Porém, é trivial o entendimento de que, por ser um modelo teórico, este representa uma aproximação da realidade, então a prática de suas três condições em conjunto não é efetivamente contínua. Nesse sentido, Medeiros e Matsumoto (2002) constataram o uso de informações privilegiadas no momento da emissão de novas ações ao mercado, abrindo a possibilidade de obtenção de retornos anormais para os agentes informados.

No mercado de capitais, existem diversas formas de se mensurar os indícios do uso de informações privilegiadas. Algumas dessas formas são: a volatilidade dos retornos dos títulos, volume de negociações, comunicados ao mercado referentes à companhia, lucros extraordinários (Dierkens, 1991);

anúncio de emissão de ações e retornos anormais ao mercado (Garcia, 2002). Outras formas de se identificar indícios de uso de informações não públicas são por meio de modelos de microestrutura de mercado, como já citado neste trabalho.

A PIN é um desses modelos de microestrutura de mercado que foi desenvolvida por Easley et al. (2002) e reflete a assimetria informacional a partir dos dados *intraday* das cotações do ativo analisado. Sua mensuração é captada por meio do desequilíbrio entre as operações de compra e de venda dos agentes presentes no mercado. Esse é um modelo que dispõe aos analistas uma medida sólida do grau de assimetria informacional para o período analisado (Abad & Rubia, 2005). No Brasil, Bosque, Albuquerque, Peng, Silva e Nakano (2020) desenvolveram uma abordagem bayesiana da PIN, com base no modelo aprimorado por Lin e Ke (2011), que possibilita introduzir opiniões dos analistas sobre os parâmetros da PIN.

Ainda no mercado brasileiro, Martins, Paulo e Girão (2016), ao analisarem o *value relevance* no modelo de Ohlson (1995) da assimetria de informação por meio da PIN de Easley et al. (2002), em 198 ações da Bolsa de Valores, Mercadorias e Futuros de São Paulo (atual B3), durante o ano de 2011, encontraram que a PIN agrega informação na avaliação de empresas, mais expressivamente em companhias que não se encontram em segmentos diferenciados de governança corporativa.

Martins, Paulo e Girão (2016) investigaram o possível caso de *insider trading* na OGX e encontraram que o preço da ação da companhia, entre os anos de 2008 e 2014, foi afetado positivamente pelas notícias no site da empresa e comunicados ao mercado; já quando relacionado aos fatos relevantes e formulários da ICVM 358, foi encontrada uma influência negativa. Com relação à PIN, os autores ainda constataram que o mercado precificou o modelo de microestrutura até meados de 2011, o que, para este período, entrou em consonância ao proposto por Demsetz (1986) em relação à possibilidade de os portadores de informações privilegiadas obterem retornos anormais.

Dessa maneira, levando em conta o potencial e a relevância do modelo da PIN proposta por Lin e Ke (2011) – PIN LK, seu desenvolvimento teórico e operacional no Brasil, e sua relevância em estudos internacionais nos quais foi adotado como *benchmark* em alguns dos estudos mais recentes relacionados à probabilidade de negociação com informação privada (Lin & Ke, 2011, Yan & Zhang, 2012, Yan & Zhang, 2014, Gan, Wei & Johnstone, 2015, Ersan & Alici, 2016, Bosque et al, 2020). Isso é o que fundamenta a utilização da PIN LK (2011) como *proxy* para assimetria informacional neste trabalho.

Ainda no que diz respeito à possibilidade de obtenção de retorno anormal por parte dos detentores do capital ao explorarem suas ações, Tonindael e Decourt (2012) encontraram, em empresas brasileiras, que esses agentes conseguem obter desempenho anormal tanto nas operações de compra, quanto nas operações de venda. Os autores constataram que para um mês após as compras dos *insiders*, houve retorno médio de 5,72%, e chegando até 9,87% seis meses após a operação. No que diz respeito às vendas dos *insiders*, houve retorno médio de -3,73%, chegando até -10,66% nos próximos seis meses.

Quando relacionado o fator PIN ao retorno das ações, Martins, Paulo e Albuquerque (2013) verificaram que, em 198 empresas listadas na BMF&Bovespa (atual B3), o aumento em 10% da assimetria de informação medida pela PIN ocasiona um aumento no retorno esperado das ações em 8%.

Mais recentemente, Pordeus, Girão e Duarte (2019) demonstraram que é possível observar retornos anormais utilizando a PIN e seus parâmetros para a seleção de carteiras de investimentos em, no mínimo, aproximadamente o retorno observado pelo Ibovespa. Dessa forma, em linha com o objetivo desta pesquisa, mesmo em uma carteira diversificada, é possível também usar a metodologia deste artigo para proteger posições com potencial de informação assimétrica entre investidores e agentes com acesso a informações privilegiadas.

Nesse sentido, com o incentivo de ampliar os estudos na área e apresentar novas evidências no Brasil, são estabelecidas as seguintes hipóteses de pesquisa deste estudo:

- **H1:** Operações baseadas em informações obtidas no relatório exigido pela ICVM 358 superam o retorno de mercado;
- **H2:** Operações baseadas na PIN e seus parâmetros superam o retorno de mercado;
- **H3:** O desempenho do modelo de *trading* baseado na PIN e em seus parâmetros é melhor do que o desempenho do modelo baseado no relatório da ICVM 358.

3. Método

Para identificar se houve retorno acima do mercado nas operações do ativo JBSS3 durante os

anos 2016 e 2017, período suficiente para captar o evento do vazamento da delação premiada dos controladores da JBS S.A., foram simuladas entradas no ativo com ajustes mensais, de acordo com a divulgação dos relatórios da ICVM 358.

Os dados referentes às negociações, utilizados para o cálculo do modelo da PIN e dos retornos, foram extraídos por meio do programa *R*, em que principalmente foi utilizado o pacote *GetHFData* (Perlin & Ramos, 2016) para *download* e tabulação desses dados de negociações diárias do período analisado. Para o cálculo do modelo da PIN, foi utilizado o *script* disponibilizado por Bosque (2016).

Já para a classificação das operações de compra e venda, identificadas pela PIN e seus parâmetros, foi aplicado o algoritmo *XGBoost*. Para tanto, foi utilizado o pacote do *R* denominado de *xgboost*, que é atualizado e mantido pelos próprios autores do algoritmo.

3.1. Estimação da Probabilidade de Negociação Informada

A mensuração da proxy de assimetria de informação nas cotações da JBSS3 foi captada por meio da probabilidade de negociação com informações privilegiadas (PIN) de Easley, Hvidkjaer e O'Hara (2002; 2010) (EHO), adaptada por Lin e Ke (2011), para aumentar a precisão do modelo (Yan & Zhang, 2012; Gan, Wei & Johnston, 2015; Ersan & Alice, 2016). Essa proxy é obtida a partir do número e sentido dos negócios das ações (agressores do *book* de ordens de compra ou venda) em determinado período de tempo.

O modelo da PIN desenvolvido por EHO identifica uma quantidade recorrente de envio de ordens de compra (ε_b) e venda (ε_s) das ações, e toma como base para sinalizar as negociações dos *traders* desinformados. Assim, um nível anormal de compras e vendas é utilizado para classificar as negociações dos agentes informados (μ). A probabilidade de uma negociação de agentes informados (α) é obtida por meio do número de dias em que o nível de negociação foi anormal. A partir disso, a PIN é estimada por meio da união desses fatores simultaneamente, como demonstra a Equação 1.

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \quad (1)$$

Em que:

α é a probabilidade de ocorrer um evento informacional;

μ é taxa de negociações de agentes informados;

ε_b é a taxa de compra de agentes desinformados; e

ε_s é a taxa de venda de agentes desinformados.

Assim, segue o método utilizado para a estimação do modelo LK em que, pós-classificação e apuração das transações de compra e venda, são estimados os parâmetros do modelo apresentados na Equação 1, por meio de um modelo de negociação sequencial. Tal estimativa foi pela maximização de uma função de máxima verossimilhança – conforme a Equação 2:

$$L(B, S) = \sum_{i=1}^n \{ \log \log [\alpha \delta \exp \exp (e_{1i} - e_{maxi}) - \alpha(1 - \delta) \exp(e_{2i} - e_{maxi}) + (1 - \delta) \exp(e_{3i} - e_{maxi})] + B_i \log (\varepsilon_b + \mu) + S_i \log \log (\varepsilon_s + \mu) - (\varepsilon_b - \varepsilon_s) + \varepsilon_{maxi} - \log \log (B_i! S_i!) \},$$

Sendo:

$$\begin{aligned} e_{1i} &= -\mu - B_i \log \log \left(1 + \frac{\mu}{\varepsilon_b} \right) & e_{2i} &= -\mu - S_i \log \log \left(1 + \frac{\mu}{\varepsilon_s} \right) & e_{3i} & \\ &= -\mu - B_i \log \log \left(1 + \frac{\mu}{\varepsilon_b} \right) - \mu - S_i \log \log \left(1 + \frac{\mu}{\varepsilon_s} \right) & e_{maxi} & \\ &= (e_{1i}, e_{1i}, e_{1i}), & i &= 1, 2, \dots, n \end{aligned} \quad (2)$$

Em que:

B e S são os volumes de compras e vendas;

B_i e S_i são as negociações no dia i ;

θ é o vetor de parâmetros (α , μ , δ , ε_b e ε_s);

$(1-\delta)$ é a probabilidade de ser uma “boa notícia”; e

δ é a probabilidade de ser uma “má notícia”.

3.2. Critérios das Operações

As operações foram segmentadas de duas formas, sendo a primeira por meio das informações coletadas no relatório do Art. 11 da ICVM 358 e a segunda foi por meio da PIN e seus parâmetros (α , δ , μ , ε_b e ε_s).

Tendo como base a fundamentação teórica de Grossman e Stiglitz (1980), no Brasil, o trabalho de Girão, Martins e Paulo (2015) ressalta a importância da transparência das informações e reafirma a proposta de Grossman e Stiglitz (1980) no tocante ao acompanhamento das operações dos *insiders*.

Nesse sentido, a identificação das operações por meio da ICVM 358 foi estabelecida se baseando nos critérios adotados por Tonindael e Decourt (2012), que selecionaram as operações com volumes financeiros mais expressivos e relevantes. Seguindo esse pensamento, foram estabelecidas duas formas de operar: **(1)** quando os membros referidos na ICVM 358 negociarem 1% ou mais de suas ações, seguindo o sentido da operação (venda ou compra); **(2)** quando os membros referidos na ICVM 358 negociarem 0,5% ou mais de suas ações, seguindo o sentido da operação (venda ou compra).

Para as operações com base na PIN, foi utilizado o algoritmo *eXtreme Gradient Boosting* (XGBoost). Este algoritmo foi desenvolvido por Chen e Guestrin (2016) e é, segundo os autores, amplamente usado por cientistas de dados para a obtenção de seus resultados em competições de aprendizado de máquina (*machine learning*), e foi vencedor do prêmio John Chambers Award (2016) em estatística computacional.

O XGBoost é um método de classificação com base no conceito de árvores de decisão, porém com diferenças significantes. Esse modelo basicamente treina as novas árvores de decisão com base em suas estruturas fracas anteriores. Inicialmente são modeladas regressões para explicar os dados, de forma randômica, obtendo seus erros, e, em seguida, com base nesses erros, são desenvolvidas novas regressões, extraindo-se um novo modelo com melhor desempenho (*i.e.*, melhor capacidade preditiva). Dessa forma, esse processo segue os seguintes passos: **(1)** aprendizagem do modelo preditivo; **(2)** computação dos resíduos; e **(3)** aprendizagem para predição dos resíduos (Dey *et al.*, 2016).

Esse modelo foi usado em aplicações no mercado financeiro para previsão das tendências das ações, no qual teve um acerto de 87% para os períodos de 60 e 90 dias (Dey *et al.*, 2016). Yu (2017) utilizou essa ferramenta para prever o *Chicago Board Options Exchange* (CBOE) *Volatility Index* – VIX, do primeiro até o sexto pregão subsequentes da análise, obtendo uma taxa de acerto entre 55% e 65%, dependendo da janela analisada, sendo a melhor acurácia obtida aplicando o modelo para 5 e 6 dias.

Sob outra perspectiva, Carmona, Climent e Momparler (2018) utilizaram o algoritmo XGBoost com o objetivo de prever a falência de bancos no mercado dos Estados Unidos entre os anos de 2001 e 2015, com uma série de 30 indicadores financeiros de 156 bancos comerciais, obtendo um acurácia de 94,74%. Ou seja, o melhor desempenho quando comparado aos demais modelos testados em seu trabalho, a Regressão Logística e *Random Forest* que obtiveram acertos de 84,21% e 92,11% respectivamente.

Os *inputs* do modelo preditivo (XGBoost) serão os *outputs* do modelo LK, sendo eles a própria PIN, α , δ , Negociações Informadas e Negociações Desinformadas (ε_b e ε_s); tendo como variável dependente os retornos observados acima do mercado, sendo ajustados para saídas binárias (0 ou 1), computando 1 quando os retornos forem maiores do que o mercado, e 0 para retornos abaixo do mercado, o que leva, portanto, à formação de um modelo de classificação.

Ressalta-se que é importante considerar não só a PIN, mas também os seus parâmetros de modo a explorar todos os componentes de um “dia normal” de negociação de ações (Easley, Kiefer & O’Hara, 1997A, B), além do paradoxo de Grossman e Stiglitz (1980).

Para a *proxy* de retorno do mercado foi estabelecido o Índice Bovespa (Ibovespa), que é o índice que concentra as maiores e mais líquidas empresas do mercado brasileiro. Os retornos ajustados ao mercado são expressos da seguinte maneira, como demonstra a Equação 3:

$$Ra = \left(\frac{Pa_{tf}}{Pa_{t0}} - 1 \right) - \left(\frac{Pm_{tf}}{Pm_{t0}} - 1 \right) \quad (3)$$

Em que:

Ra é o retorno do ativo ajustado ao mercado;

Pa_{tf} é o preço do ativo na data futura;

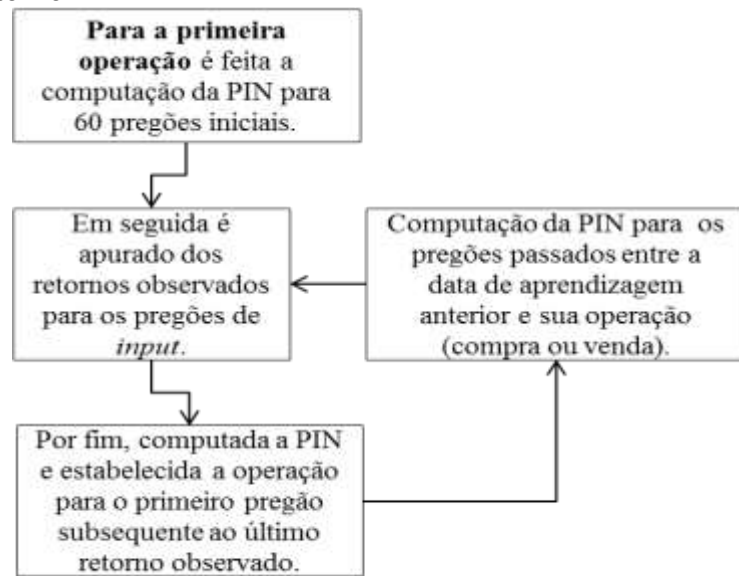
Pa_{t0} é o preço do ativo na data de estimação da PIN;

Pm_{tf} é o preço do Ibov na data futura; e

Pm_{t0} é o preço do Ibov na data estimação da PIN.

Mais especificamente, o retorno é calculado de ponta-a-ponta, no caso: pregão 1 é formulada a carteira que ficará em "custódia" até o 20º pregão e assim sucessivamente. Esses valores pontuais de "20 em 20 pregões" são utilizados para calcular o retorno acumulado total da carteira ao término do experimento que buscamos testar.

Após a computação dos valores da PIN, seus parâmetros e do retorno ajustado ao mercado para os dias da amostra, que consiste nos pregões entre os dias 13/01/2016 e 08/12/2017 (totalizando 475 pregões), é realizada a primeira aprendizagem (estimação) do modelo para classificação da primeira operação (compra ou venda), aplicando o XGBoost, tendo como entrada para aprendizagem os dados dos 60 primeiros pregões da amostra. A Figura 1 demonstra o fluxo do processo e aprendizagem e observação do retorno.



Fonte: Elaborado pelo autor

Figura 1: Rotina de aprendizagem e previsão do XGBoost

Neste estudo, se optou por testar duas janelas de entrada para a variável dependente (retorno ajustado ao mercado). Portanto, foi desenvolvida a análise para dois modelos distintos: no **primeiro modelo** são observados os retornos ajustados ao mercado no 20º pregão subsequente à apuração dos parâmetros da PIN; no **segundo modelo**, ampliamos a sua temporalidade para o 60º pregão posterior, com o objetivo de captar o *delay* em relação à reação do mercado frente ao cenário na data de apuração da PIN, visto que Tonidael e Decourt (2012) encontram que, testando os mesmos espaços temporais, quanto maior o período de observação, maior é o retorno anormal.

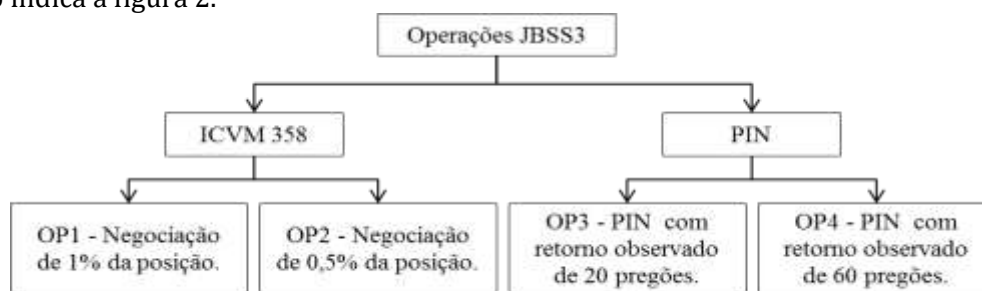
Com relação aos parâmetros adotados no XGBoost, o modelo é utilizado por meio do comando *xgboost*, do pacote do R com o mesmo nome, sendo o *output* por meio de regressão logística binária, com valores probabilísticos entre 0 e 1. Dessa forma, para os *outputs* com valores acima de 0,5 foi adotada a decisão de se realizar operações de compra; para os *outputs* com valores menores que 0,5, foi tomada a decisão de se realizar operações de venda.

As saídas das regressões logísticas, ou probabilísticas, estimam a probabilidade de 0-100% de um determinado evento ocorrer. Nesse caso, se assumiu que uma saída probabilística acima de 0,5 (ou 50%) significa que existe mais chance de o retorno ajustado ser positivo, sendo classificada como uma operação de compra, assim como o inverso é verdadeiro para o experimento testado nesta pesquisa.

Foi aplicado o comando *early stopping* para parar as simulações automaticamente quando encontrado um valor mínimo para o erro. É importante utilizar a ferramenta, dado que ela ajuda a minimizar o *overfitting* do modelo, visto que ele para o processo de treinamento quando não é identificado um ganho na função custo. Para identificar a perda de ganho "manualmente" seria necessário fazer testes de "força bruta".

Como forma de ilustração, a Figura 2 representa o fluxograma de extração das operações por meio dos quatro modelos de decisão aplicados neste estudo. Na descrição dos resultados deste trabalho, serão apresentados os frutos dos modelos, denominados de operações 1, 2, 3 e 4 (OP1, OP2, OP3 e

OP4), como indica a figura 2.



Fonte: Elaborado pelo autor

Figura 2 – Fluxo do Processo Decisório das Operações

A data das operações é referente à data de divulgação dos relatórios da ICVM 358. Portanto, a cada divulgação, é analisado se será efetuada compra, venda ou não operar, para ocasiões que não houver informações relevantes para a fundamentação de uma posição.

4. Resultados

4.1 Estatísticas descritivas

A Tabela 1 demonstra a distribuição da amostra em seus valores mínimos, máximos, médios, e em seu primeiro, segundo e terceiro quartil – compondo as estatísticas descritivas.

Tabela 1 – Estatística descritiva da amostra para 2016 e 2017

Retornos	Volume de negócios		PIN e seus parâmetros de estimação						
	Re 20p	Re 60p	ICVM	PIN	α	δ	ε_b	ε_s	μ
Mínimo	-48,49%	-58,52%	0%	8,54%	13,33%	0%	6694	6188	4276
1º quartil	-12,28%	-18,84%	0%	12,26%	26,53%	26,21%	7645	7733	7481
Mediana	-2,95%	-7,50%	0%	15,10%	33,33%	40,01%	8524	8521	9843
Média	-5,07%	-8,99%	0,25%	15,13%	32,96%	44,90%	9243	8715	10471
3º quartil	5,97%	7,58%	0,03%	17,66%	38,33%	57,63%	9634	9479	12238
Máximo	27,06%	26,63%	2,70%	22,54%	58,81%	100%	15876	11631	21662

Fonte: Elaborado pelos autores.

Nota: Re 20p se refere aos retornos no 20º pregão subsequente ao da apuração da PIN; Re 60p se refere aos retornos no 60º pregão subsequente ao da apuração da PIN; ICVM se refere ao volume (%) de negociação da posição dos *insiders*; PIN se refere à probabilidade de negociação com informações privadas; α se refere à probabilidade de ocorrência de um evento informacional; δ se refere à probabilidade de o evento ser uma má notícia; ε_b se refere à taxa de compra dos agentes desinformados; ε_s se refere à taxa de venda dos agentes desinformados; μ se refere à taxa de negociação dos agentes informados.

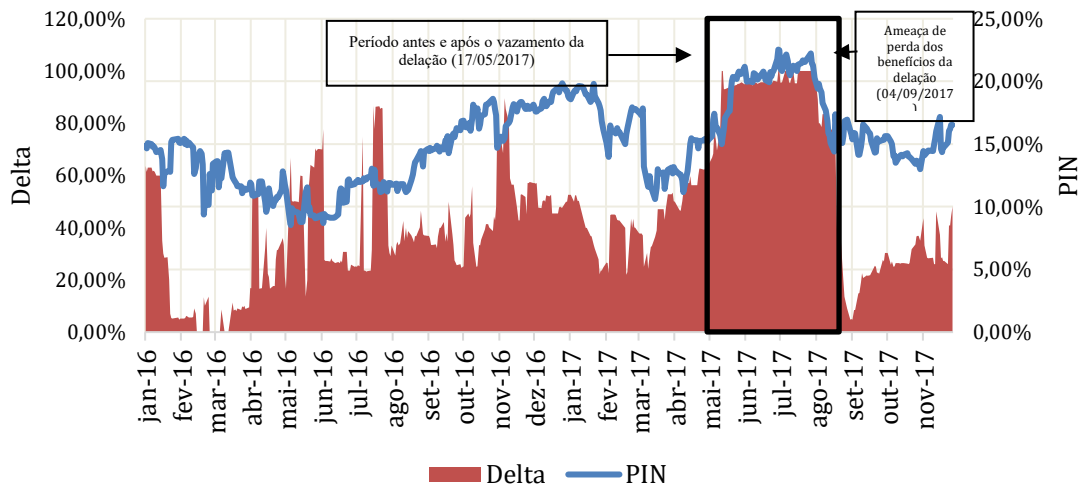
A partir da Tabela 1, pode-se observar que, em média, os retornos ajustados ao mercado são negativos, em -5,07% e -8,99%, para os intervalos de 20º e 60º pregão posterior, respectivamente, e que, com exceção do terceiro quartil, o retorno ajustado ao mercado em todos os intervalos da amostra para o 60º pregão posterior são inferiores ao retorno ajustado da amostra do 20º pregão, o que indica uma tendência de baixa para o ativo da JBSS3, em um horizonte de tempo maior.

Com relação às informações obtidas por meio da ICVM 358, observa-se uma baixa movimentação dos *insiders*, visto que em 75% dos meses os *insiders* negociaram menos que 0,03% de sua posição total (este ponto será abordado posteriormente na apresentação da Tabela 4 deste trabalho).

De acordo com o modelo LK, a probabilidade diária de haver uma negociação informada, para os anos de 2016 e 2017 no ativo JBSS3, foi em média 15,13%, valor consistente com o observado em pesquisas anteriores sobre o assunto, como observado por Martins, Paulo e Albuquerque (2013). Com relação ao Delta, em média, a chegada de notícias é positiva, o que é representado por um delta abaixo de 50%, informação corroborada pela sua média (44,90%) e pelo número de negociações efetuadas pelos agentes compradores, em relação aos agentes vendedores, tendo em média um número de negociações 8,54% superior, quando observado os intervalos da amostra apresentada na Tabela 1.

Com intuito de promover uma melhor visualização da relação entre a probabilidade de negociação informada (PIN) e a probabilidade desta informação ser negativa (Delta ou δ), apresenta-se o Gráfico 1:

Gráfico 1: Relação intraday entre PIN e o Delta entre janeiro de 2016 e dezembro de 2017



Fonte: Elaborado pelos autores.

Nota: PIN se refere à probabilidade de negociação com informações privadas; Delta (δ) se refere à probabilidade de o evento informacional ser uma má notícia.

A média da PIN para o período analisado foi 15,13%, como visto na Tabela 1, e, com base nisso, é possível notar que nos meses anteriores à delação, 01 de novembro de 2016 e 17 maio de 2017 (data do vazamento da delação), o ativo JBSS3 teve uma PIN média de 17,79%, valor 17,55% superior à média total do período. Nesse mesmo período em que a PIN estava acima da média, o sentido do delta se manteve praticamente neutro, com uma média de 46,40%, o que é 3,35% superior à média total da amostra e simboliza incerteza com relação ao sentido da negociação informada.

Outro período de destaque está entre o começo de julho e o final de agosto de 2017, quando a PIN média foi de 20% e o Delta médio de 90,49%, ficando num patamar superior à média total da amostra em 32,20% e 101,54%, respectivamente. Tais fatores demonstram uma alta probabilidade de negociação com informação negativa, privada ao mercado, fato que pode ser relacionado ao pronunciamento feito pelo então Procurador-Geral da República (PGR), Rodrigo Janot, no dia 4 de setembro de 2017, direcionando a possibilidade do cancelamento dos benefícios da delação premiada feita pelos controladores da companhia – isso ainda é discutido hoje, em 2020, pelo atual PGR.

Tabela 2 - Cronologia dos fatos relacionados à delação da JBS, de janeiro até dezembro de 2017

Fato	Data
Operação Cui Bonno	13 de janeiro de 2017
Marcelo Miller pede exoneração	23 de fevereiro de 2017
Temer recebe Joesley	07 de março de 2017
Operação carne fraca e diálogo com Ricardo Saud	17 de março de 2017
Joesley grava Aécio Neves	24 de março de 2017
Negociação com a Procuradoria-Geral	27 de março de 2017
Marcelo Miller deixa o Ministério Público	05 de abril de 2017
Ações controladas e depoimentos	Em abril de 2017
Assinatura da delação	03 de maio de 2017
Homologação da delação	10 de maio de 2017
Vazamento da delação	17 de maio de 2017
Aécio Neves é denunciado	02 de junho de 2017
Michael Temer é denunciado	27 de junho de 2017
Câmara barra denúncia de Aécio Neves	02 de agosto de 2017
Acordo de delação pode ser anulado	04 de setembro de 2017

Fonte: adaptado de Nexo

Já no que diz respeito à negociação dos *insiders*, representada pelas informações obtidas no relatório da ICVM 358, a atuação mais relevante desses agentes é efetivada durante os meses de maio, junho e julho de 2017, nos quais eles negociaram, ao todo, 85,6% das ações do total analisado, entre 2016 e 2017. A Tabela 2 apresenta a cronologia dos fatos ocorridos no período antes e depois da delação premiada dos agentes da JBS S.A.

Sabendo da cronologia dos fatos, pode-se correlacionar, na Tabela 3, os meses em que houve movimentação nas posições dos Irmãos Batista, a data de apresentação do relatório da ICVM 358 ao mercado, sentido da negociação (compra ou venda), posição negociada em relação ao total e o quanto a negociação representa em relação ao total do período analisado.

Tabela 3 – Negociação dos *insiders* apresentados por meio da ICVM 358, entre janeiro e dezembro de 2017

Data de entrega do relatório ao mercado	Compra	Venda	Quantidade de ações negociada	Proporção das negociações em relação ao período	Posição negociada (%)
10/02/2016	191mil	197,2mil	388,2mil	0,56%	0%
10/03/2016	1.169mil	1.227,7mil	2.396,7mil	3,43%	0%
08/04/2016	-	63,1mil	63,1mil	0,09%	0%
10/05/2016	-	10mil	10mil	0,01%	0%
09/01/2017	-	40mil	40mil	0,06%	0%
09/01/2017	-	10mil	10mil	0,01%	0%
08/03/2017	-	254,5mil	254,5mil	0,36%	0%
10/05/2017	-	31.777,5mil	31.777,5mil	45,46%	2,63%*
10/06/2017	-	18.635,7mil	18.635,7mil	26,66%	1,58%*
10/07/2017	-	9.413,3mil	9.413,3mil	13,47%	0,81%*
10/08/2017	-	385,5mil	385,5mil	0,55%	0,03%
08/09/2017	-	6.520mil	6.520mil	9,33%	0,57%*
10/11/2017	-	1mil	1mil	0%	0%

Fonte: Elaborado pelos autores.

Nota: * Meses em que a negociação foi fora da normalidade, caracterizando operação de acordo com o modelo 1 e/ou 2 referentes a ICVM 358.

Uma curiosidade adicional para os leitores da pesquisa, visto que estamos em um universo de *small data*, é que o tempo despendido no processo de treinamento e predição do modelo é irrisório, de poucos minutos. Porém, o tempo gasto para o cálculo da PIN para um pregão é de cerca de 15 segundos, no geral – dependendo da máquina que está efetuando o procedimento. Nesta pesquisa, gastaram-se, no total, aproximadamente 2 horas para computar a PIN e seus parâmetros para o ativo JBSS3.

4.2 Análise das hipóteses com base nos relatórios da ICVM 358 e com os modelos baseados na PIN

De acordo com as estatísticas descritivas apresentadas anteriormente, as operações com a ICVM 358 foram simuladas para maio, junho, julho e setembro de 2017, seguindo o sentido das operações dos *insiders*. Para a PIN, foram simuladas entradas em 11 meses com o modelo com retorno observado no 20º pregão subsequente e em 9 ocasiões para o modelo com retorno observado no 60º pregão subsequente. Já para operações de venda foram 9 para o modelo com retorno observado no 20º pregão subsequente e 8 para o modelo com retorno observado no 60º pregão subsequente.

A Tabela 4 compara o resultado das simulações, incluindo o Ibovespa – referência utilizada neste estudo. As operações, quando efetuadas com base na ICVM 358, obtiveram retorno positivo, sendo que não é possível assumir a significância dessas operações, visto que, dos 24 períodos observados, foram identificadas apenas 2 operações referentes ao primeiro modelo de seleção, quando os agentes previstos na ICVM 358 negociarem mais que 1% de suas ações, que obtiveram 100% de acerto.

Vale ressaltar que o período em que esse modelo sinalizou para operar foi nos meses de maio e abril, meses subsequentes ao vazamento da delação premiada dos seus controladores. Já para o segundo modelo houve mais pontos de operações, quatro ao todo, sendo que destas, 50% foram operações falhas, ou seja, com perdas.

No que diz respeito à rentabilidade desses modelos baseados na ICVM 358, o primeiro obteve um retorno total de 49,49%; o segundo, 10,8%, sendo que para o mesmo período o Ibovespa obteve uma rentabilidade de 60,62%. Isso significa que ambos os modelos obtiveram um desempenho inferior ao mercado em -11,13% e -49,82, respectivamente. Dessa maneira, a **H1**, que buscou verificar se seria possível obter retorno acima do mercado por meio de operações efetuadas com base nas informações obtidas no relatório referente ao Art. 11 da ICVM 358, é rejeitada – tanto pelo retorno total, quanto pela relação risco-retorno.

Tabela 4 – Comparativo das operações efetuadas com informações de *insiders* entre janeiro de 2016 e dezembro de 2017

	Ibovespa	ICVM 1%	ICVM2 0,5%	PIN 20p	PIN 60p
Meses Analisados	-	24	24	24	24
Sem Operações	-	22*	20*	4**	7**
Compras	-	0	0	11	9
Vendas	-	2	4	9	8
Taxa de Acerto***	-	100,0%	50,0%	60%	64,71%
Retorno Total	91,79%	49,49%	10,8%	78,13%	127,84%
Retorno Médio	3,82%	2,06%	0,45%	3,26%	5,33%
Desvio-padrão	6,15%	8,03%	9,56%	14,42%	12,92%
Retorno/Risco****	0,62	0,26	0,05	0,23	0,41

Fonte: Elaborado pelos autores.

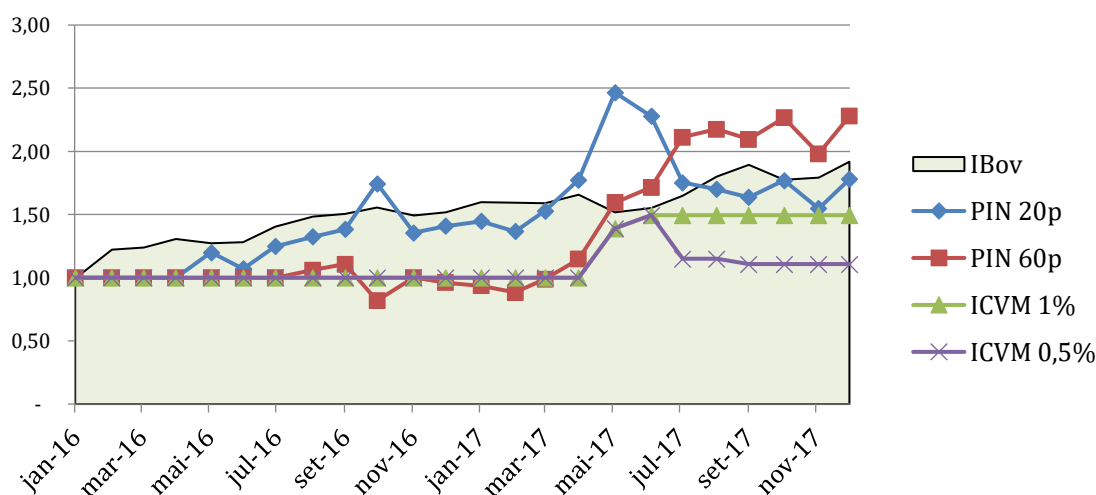
Notas: Ibovespa representa a referência para o mercado brasileiro de ações; ICVM 1% representa a estratégia baseada na simulação dos *trades* dos *insiders* quando eles negociam mais de 1% das suas ações; ICVM2 0,5% representa a estratégia baseada na simulação dos *trades* dos *insiders* quando eles negociam mais de 0,5% das suas ações; PIN 20p representa o retorno apurado 20 pregões depois do cálculo da PIN; PIN 60p representa o retorno apurado 60 pregões depois do cálculo da PIN.

* não houve operação significativa dos *insiders*; ** refere-se aos meses em que foram computadas as entradas para o modelo XGBoost; *** refere-se à relação entre o que foi proposto pelos modelos sobre o observado no mercado; **** Retorno Médio dividido pelo Desvio-padrão, quanto maior esse número, melhor é o desempenho da estratégia analisada.

Já com relação às operações indicadas pelo XGBoost, tendo como base a PIN, foram efetuadas 20 operações para o modelo com retorno no 20º pregão subsequente (PIN 20p), e 17 operações para o modelo com retorno no 60º pregão subsequente (PIN 60p), acertando, respectivamente, 12 das 20 operações (60%) e 11 das 20 operações simuladas (64,71%).

No tocante à rentabilidade desses modelos, o modelo com análise do 20º pregão obteve um retorno total de 78,13%; o segundo modelo, por sua vez, obteve um retorno observado de 127,84%. Assim, apenas um dos modelos com base na PIN obteve retorno acima do Ibovespa, modelo com maior janela de retorno, conforme esperado (Tonidael & Decourt, 2012). Portanto, rejeita-se a H2 para o modelo de 20 períodos, mas não rejeitamos a H2 para o modelo de 60 períodos – considerando o retorno total. Se considerarmos a relação risco-retorno, nenhum deles superou o Ibovespa, porém o Ibovespa foi composto por uma carteira *buy and hold*, enquanto os modelos de PIN foram compostos por *trades* eventuais – ficando menos expostos às oscilações gerais do mercado.

Gráfico 2 - Capitalização dos modelos e comparativo com o Ibovespa, entre 2016 e 2017



Fonte: Elaborado pelos autores.

Notas: Ibovespa representa a referência para o mercado brasileiro de ações; ICVM 1% representa a estratégia baseada na simulação dos *trades* dos *insiders* quando eles negociam mais de 1% das suas ações; ICVM2 0,5% representa a estratégia baseada na simulação dos *trades* dos *insiders* quando eles negociam mais de 0,5% das suas ações; PIN 20p representa o retorno apurado 20 pregões depois do cálculo da PIN; PIN 60p representa o retorno apurado 60 pregões depois do cálculo da PIN.

Com relação à H3 desta pesquisa, encontramos evidências que a suportam, visto que ambos os

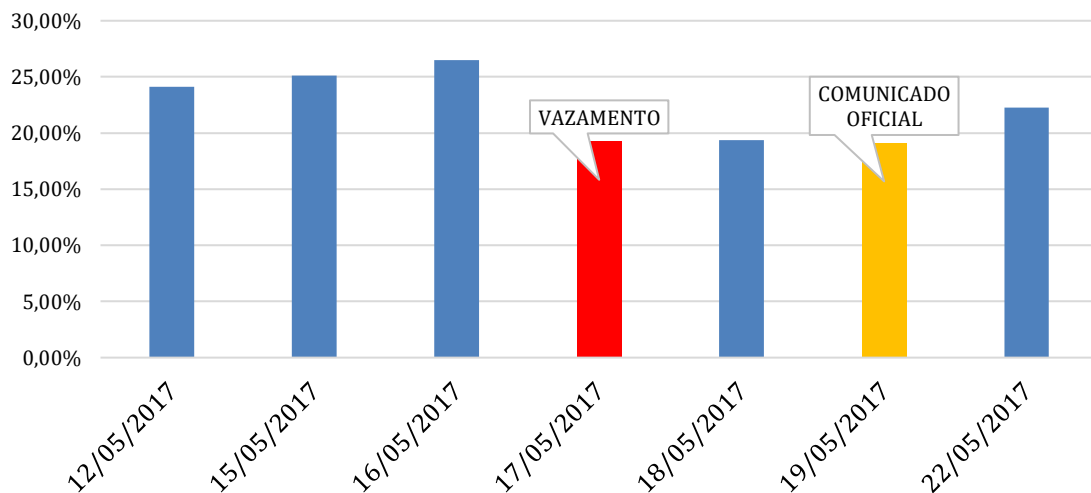
modelos com base na PIN, nesta amostra estudada, obtiveram melhores rendimentos quando operando por meio da PIN do que por meio das informações presentes no relatório da ICVM 358.

Por fim, com o objetivo de visualizar melhor o desenvolvimento dos modelos durante o período, está disposto o Gráfico 2, no qual se comparam todos os modelos abordados neste estudo e o desempenho do Ibovespa, adotando como o início da série o valor de uma unidade monetária, para efeito de capitalização. É relevante ressaltar que ambos os modelos da PIN, a partir de fevereiro de 2017, meses antes do vazamento da delação, apresentaram alta expressiva.

4.3 Teste de robustez: quem perdeu e quem ganhou com as operações dos insiders?

Para testar a robustez dos dados e uma linha do tempo dos eventos, com o modelo EHO original, apresentamos, no Gráfico 3, a evolução da PIN antes e após o vazamento da delação dos Irmãos Batista. Conforme o relatório da ICVM 358 de maio de 2017, os controladores da JBS venderam 984.900 ações no dia 16/05/2017, ao preço médio de R\$ 10,11, e venderam mais 3.635.000 ações no dia 17/05/2017 (o vazamento da delação ocorreu após o fechamento do mercado), ao preço médio de R\$ 9,66.

Gráfico 3 - Evolução da PIN da JBSS3 antes, durante e depois do vazamento da delação dos Irmãos Batista



Fonte: Elaborado pelos autores.

Notas: As cores das barras servem para diferenciar as PINs nas datas indicadas pelos dias sem evento específico e dias com o vazamento e o comunicado oficial da delação premiada.

Ressalta-se que as negociações com a PGR começaram em 27/03/2017 e, durante abril, foram coletados os depoimentos que seriam usados como parte da delação premiada. Ainda, a última vez que os acionistas controladores haviam vendido ações da companhia foi em fevereiro de 2016, antes de venderem um pouco em abril (quando já estavam em negociações com a PGR e gravando depoimentos) e maio (1 dia antes e no dia do vazamento da delação) de 2017.

Enquanto os controladores vendiam suas ações, no mesmo dia, 17/05/2017, a tesouraria da JBS S.A. comprou 3.166.100 ações pelo Goldman Sachs, ao preço médio de R\$ 9,66; 222.400 ações pela XP Investimentos, ao preço médio de R\$ 9,62; e 140.000 ações pela Ativa, ao preço médio de R\$ 9,65.

Dessa forma, os controladores da empresa evitaram uma perda de aproximadamente R\$ 48 milhões, às custas da empresa e, em consequência, dos seus acionistas. Adicionalmente, a Caixa Econômica Federal e o BNDES Participações tinham posição na empresa, detendo, respectivamente, 4,92% e 21,32% do capital da JBS S.A. Ou seja, os contribuintes brasileiros também perderam dinheiro, mesmo sem ter ações da JBS S.A. No final das contas, os controladores venderam 4.619.900 ações no entorno da delação e a tesouraria da companhia comprou 3.528.500.

5. Implicações teóricas e práticas

Esta pesquisa pode contribuir tanto com a teoria, quanto com a prática de investidores e também para o órgão regulador do mercado de capitais brasileiro da seguinte maneira:

a) **Teoria Financeira:** os resultados evidenciaram que de fato os preços dos ativos não se ajustam

automaticamente à chegada de ordens de compra ou venda com base no uso de informações privilegiadas e que, em um caso que foi o 1º que gerou prisão por *insider trading* no Brasil, foi possível testar a adequação do modelo, indicando, mais uma vez, que o mercado não é eficiente na forma forte;

- b) **Prática de investidores:** é possível que investidores usem algoritmos de alta frequência (HFT), tanto para se proteger da seleção adversa e do risco moral, conforme explicado na introdução, quanto para gerar operações de compra e venda, com a detecção de chegada de “negociadores bem informados” no livro de ofertas de um determinado ativo, como no exemplo da JBSS3; e, por fim
- c) **Regulação do mercado de capitais:** os resultados também apontam que o órgão regulador poderá utilizar os resultados desta pesquisa para elaborar modelos de HFT que detectem antecipadamente algum negociador informado em busca de expropriar os negociadores menos informados, ou comparar os resultados ora apresentados nesta pesquisa com os modelos que eles eventualmente utilizem no dia a dia. A BSM, vinculada à B3, também monitora os dados de negociação, porém não podemos afirmar que usam um modelo semelhante; logo, eles também poderiam se beneficiar desses resultados para comparar com as ferramentas usadas para detectar o uso de informações privilegiadas.

6. Considerações finais

Este estudo teve como objetivo identificar operações com base no potencial uso de informações privilegiadas. Para tanto, aplicou-se o modelo de aprendizado de máquina aos *outputs* da PIN, obtendo um retorno acima do observado para as operações efetuadas pelos agentes internos à companhia analisada: JBS S.A.

Os resultados encontrados para o período analisado constataram que as operações efetuadas por meio do modelo XGBoost obtiveram um resultado superior às operações efetuadas pelos *insiders*, tendo em vista que, de acordo com a ICVM 358, essas operações são influenciadas pelo *delay* entre as reais operações dos agentes internos à companhia e a disseminação dessa informação ao mercado, visto que os *insiders* têm até o décimo dia útil do mês subsequente para informar as operações efetuadas, o que pode comprometer o desempenho daqueles que operam seguindo suas operações (o que demonstra as simulações com base nessa informação), ou que querem se proteger da expropriação gerada pelo uso de informações privadas.

Sobre algumas escolhas metodológicas, destaca-se que os 20 pregões foram escolhidos por caracterizar, de forma genérica, o período de um mês, fazendo com que os testes com a PIN fiquem mais em linha com as informações disponibilizadas pela ICVM, ganhando um maior grau de comparabilidade. Já a utilização de 60 pregões tem sua origem na ideia de *trades* com prazo maior, dado que os *insiders* podem tentar se antecipar às divulgações de eventos e “treidar” antes para não chamar tanta atenção. Para esse caso específico, é possível verificar algumas evidências disso com base nas tabelas 2 e 3 desta pesquisa. Em pesquisas futuras, seria possível avaliar outras metodologias de escolha com prazos maiores ou menores.

Vale destacar que este estudo está limitado ao ativo JBSS3 para o período de janeiro de 2016 até dezembro de 2017, visto que esse foi o primeiro caso de potencial *insider trading* com prisão no Brasil e a fonte de dados da B3 não está mais disponível para consulta.

Ademais, é importante ressaltar a relevância do avanço da tecnologia de inteligência de máquina e seu acesso mais intuitivo. A união desse nicho de conhecimento com as demais áreas da ciência, inclusive com o mercado de capitais, é fundamental, aliado aos *insights* gerados pela experiência do usuário, para aperfeiçoamento do processo de tomada de decisão e manuseio de novas fontes de informações.

Referências

- Abad, D & Rubia, A. (2005). Modelos de estimación de la probabilidad de negociación informada: una comparación metodológica en el mercado Español. *Revista de Economía Financeira*, (7), 1-37.
- Aldridge, I. (2013). *High-frequency trading: a practical guide to algorithmic strategies and trading systems* (2nd ed). John Wiley & Sons.
- Bosque, L., Albuquerque, P., Peng, Y., Da-Silva, C., & Nakano, E. (2020). Probability of informed trading: a Bayesian approach. *International Journal of Applied Decision Sciences*, 13(2), 183-214.

- Bosque, L. (2016). *Estimação da probabilidade de negociação privilegiada por meio de inferência bayesiana*. Dissertação de mestrado, Universidade de Brasília, Brasília, DF, Brasil.
- Callado, A. A. C. (2009). *Eficiência do mercado acionário brasileiro: retorno das ações negociadas na Bovespa, variáveis macroeconômicas, causalidade e fatores condicionantes*. Tese de doutorado, Universidade Federal de Pernambuco, Recife, PE, Brasil.
- Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, (61), 304-323.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, CA, USA, 22.
- Easley, D., Hvidkjaer, S., & O'hara, M. (2002). Is information risk a determinant of asset returns? *The journal of finance*, 57(5), 2185-2221.
- Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). *Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting*. PESIT, Bengaluru, India, Working Paper.
- Easley, D., Kiefer, N. M., O'hara, M., & Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *The Journal of Finance*, 51(4), 1405-1436.
- Easley, D., Kiefer, N. M., & O'Hara, M. (1997A). The information content of the trading process. *Journal of Empirical Finance*, 4(2-3), 159-186.
- Easley, D., Kiefer, N. M., & O'Hara, M. (1997B). One day in the life of a very common stock. *The Review of Financial Studies*, 10(3), 805-835.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393-408.
- Girão, L. F. D. A. P., Martins, O. S., & Paulo, E. (2015). Insider Trading B-side: relevance, timeliness and position influence. *Brazilian Review of Business Management*, São Paulo, 17(58), 1341-1356.
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of accounting and economics*, 31(1-3), 405-440.
- Martins, O. S., & Paulo, E. (2013). A Probabilidade De Negociação Com Informação Privilegiada No Mercado Acionário Brasileiro (The Probability of Informed Trading in the Brazilian Stock Market). *Revista Brasileira de Finanças*, 11(2), 249-280.
- Martins, O. S., Paulo, E., & Albuquerque, P. (2013). Negociação com informação privilegiada na BM&FBovespa e seu reflexo no retorno das ações (the probability of informed trading in BM&FBovespa and its reflection on stock returns). *Revista de Administração Empresas (RAE)*, 53(4), 350-362.
- Martins, O. S., Paulo, E., & Girão, L. F. D. A. P. (2016). Preço da ação, disclosure e assimetria de informação: o caso OGX. *Revista Universo Contábil*, 12(1), 06-24.
- Perlin, M., & Ramos, H. (2016). GetHFData: A R package for downloading and aggregating high frequency trading data from Bovespa. *Brazilian Review of Finance*, 14(3), 443-478.
- Pordeus, G. G. N., Girão, L. F. A. P. & Duarte, F. C. L. (2019, Julho). Seleção de Carteiras por Meio do Aprendizado de Máquina com Informação Assimétrica: uma abordagem com o XGBoost. *XIX Congresso USP de Iniciação Científica em Contabilidade*, São Paulo, SP, Brasil, 19.
- Yu, M. Y. (2017). *Predicting the Volatility Index Returns using Machine Learning*. Tese de doutorado, University of Toronto, Toronto, ON, Canadá.