



10º Encontro de Ensino Pesquisa e Extensão

Patrocínio, MG, Outubro 2023

APLICAÇÃO DE ALGORITMO DE REGRESSÃO NA PREVISÃO DE ATIVOS DA BOLSA DE VALORES

Carlos Eduardo de Sousa Ávila
Cintia Carvalho Oliveira
Instituto Federal do Triângulo Mineiro
Modalidade: Pesquisa
Formato: Artigo Completo

Resumo

Neste estudo, foi avaliada a eficácia de algoritmos de regressão como ferramenta para previsão de preços no mercado financeiro de ativos. A volatilidade dos preços é uma preocupação constante para investidores, pois pode levar a ganhos substanciais ou a perdas significativas. A previsão de preços desempenha um papel crucial na tomada de decisões de investimento, e esta pesquisa busca contribuir para essa área. Foram coletados dados técnicos históricos de diversos ativos da bolsa de valores brasileira. Esses dados foram utilizados para treinar modelos de regressão polinomial, com diferentes períodos de treinamento. A validação das tendências de preços foi uma métrica essencial para avaliar a precisão dos modelos. Os resultados mostraram que, com um período de treinamento de 60 dias, o modelo de regressão polinomial apresentou maior precisão na previsão de tendências de preços. Foi observado que as previsões mais acuradas ocorreram no 40º dia do período de previsão. Esse estudo contribui para a compreensão da utilidade dos algoritmos de regressão na previsão de valores futuros no mercado financeiro, oferecendo uma abordagem potencialmente mais econômica em comparação com métodos mais complexos, como redes neurais. À medida que as estratégias de investimento são aprimoradas de forma contínua, a previsão de preços permanece como uma ferramenta valiosa para investidores e pesquisadores do mercado financeiro.

Palavras-chave: Algoritmos de regressão; Investimentos; Previsão de preços.

1 Introdução

A volatilidade dos preços no mercado financeiro de ativos é algo que molda a relação de investidores com a aquisições de papéis. Ao mesmo tempo em que é responsável pelos potenciais de ganhos, a flutuação imprevisível dos preços pode resultar em prejuízos significativos.

O comportamento muitas vezes errático do mercado de ações, leva muitos economistas e especialistas a utilizarem, da própria experiência e conhecimento para tomada de decisões, porém esse tipo de método, sofre das fraquezas inerentes ao próprio ser humano, seja de origem emocional, psicológica ou física. (JUNIOR, 2022)

Neste contexto, a previsão de preços é uma ferramenta valiosa que deve ser usada em conjunto de outras métricas para avaliação de aquisições e vendas. Dessa forma, a utilização de algoritmos de regressão surge como um caminho na busca da previsão de valores futuros.

O objetivo deste trabalho é avaliar a eficácia de algoritmos de regressão com o intuito de consolidar uma ferramenta que possa ajudar investidores e pesquisadores do mercado financeiro. Foi feita a catalogação de variáveis técnicas históricas relativas a diversos ativos da bolsa de valores brasileira.

Por fim, o desempenho dos modelos foi avaliado com aplicação de métricas como previsão de movimentações de mercado, verificando se a tendência prevista na regressão foi seguida pelos valores reais de preços e por quanto tempo isso se cumpriu. Além disso, foram testados diferentes tipos de ativos e quantidade de dados de treinamento, buscando o resultado com maior acurácia. Espera-se poder definir a eficácia desse tipo de algoritmo na previsão de valores futuros, podendo ser uma técnica mais barata em termos de desenvolvimento em relação a técnicas mais elaboradas como a criação de redes neurais. Apresentando potencial de previsibilidade de valores, espera-se que esse tipo de algoritmo seja cada vez mais usado e estudado para esse fim.

2 Fundamentação teórica

2.1 Python

A linguagem Python se trata de apenas uma dentre várias linguagens de programação. Assim como idiomas/línguas humanos, existem diversas linguagens computacionais, a maioria é boa em algo em específico, como escrever pequenos programas, acessar banco de dados ou construir páginas web. Python é uma linguagem poderosa, de fácil legibilidade, possui diversas bibliotecas que estendem suas funcionalidades, é grátis e possui diversas aplicações no mundo real. (MENEZES, 2010)

2.2 API

A sigla API em inglês significa interface de aplicação programada. Basicamente se trata de um software responsável por enviar respostas de acordo com a requisição recebida.

Em uma API RESTful, HTTP é o protocolo de comunicação e os serviços disponíveis são definidos como URLs. Geralmente, as requisições são definidas pela construção da URL passando parâmetros definidos pela API [...], a resposta é retornada em uma estrutura já definida. Historicamente, o modelo XML foi utilizado, mas APIs mais recentes costumam utilizar o formato JSON. (SELBY et al., 2019)

2.3 Aprendizado de máquina

Aprendizado de máquina é um ramo em evolução de algoritmos computacionais que têm como objetivo simular a inteligência humana, aprendendo a partir do ambiente no qual está inserido. É considerada uma importante parte na nova era chamada big data. (EL NAQA; MURPHY, 2015)

A capacidade de algoritmos desse tipo de aprenderem a partir de um determinado contexto e conseguirem generalizar para novas tarefas não vistas, é o que torna esse tipo de implementação algo tão poderoso de ser aplicado, não apenas no campo das finanças, mas em áreas como engenharia, entretenimento e medicina.

2.3.1 Aprendizado supervisionado

O aprendizado supervisionado se trata de um ramo do aprendizado indutivo do aprendizado de máquina. Nesse tipo de algoritmo, é apresentado um conjunto de dados que estão relacionados a uma classe ou rótulo específico. É supervisionado já que os exemplos passados para teste têm rótulos conhecidos.

Cada exemplo é descrito por um vetor de valores (atributos) e pelo rótulo da classe associada. O objetivo do algoritmo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados. Para rótulos de classe discretos, esse problema é chamado de classificação e para valores contínuos como regressão. Esse método de aprendizado é o mais utilizado. (LUDERMIR, 2021)

2.3.2 Regressão linear

A regressão linear é uma técnica de aprendizado de máquina que pode ser considerada o primeiro passo em modelos de regressão. O objetivo é modelar uma relação linear entre variáveis para chegar a uma regra que descreva ou estime um valor y com base em um valor x - ou um vetor de valores x . (MAULUD; ABDULAZEEZ, 2020)

Nesse sentido, é feito um treinamento com variáveis correlacionadas, onde seus rótulos são conhecidos, a fim de produzir uma regra matemática que descreva a relação entre os valores. Dessa forma, pode-se aplicar a regra gerada para valores que possuem rótulos desconhecidos e realizar estimativas e previsões.

Em razão de modelos de regressão linear gerarem uma equação de grau um, as projeções geradas estarão alinhadas a uma linha no plano cartesiano. Isso faz com que tais previsões não tenham tanta precisão. Além disso, se os dados utilizados no treinamento do modelo apresentarem desvios muito grandes, a descrição das variáveis estará enviesada, resultando em menor acurácia do modelo.

2.3.3 Regressão polinomial

A regressão polinomial é uma técnica de aprendizado de máquina que possui o mesmo objetivo da regressão linear, além de ser aplicada para projeções após submissão do modelo

ao treinamento com dados com rótulos conhecidos. Em modelos de regressão polinomial, a regra gerada que descreve a relação entre as variáveis não se apresenta de forma linear, podendo ser mais preciso em determinados tipos de base de dados.

Regressão polinomial é um tipo de análise de regressão no (n)ésimo grau polinomial modelo da relação entre variáveis independentes e dependentes. Regressão polinomial é um case especial de regressão linear multivariada na qual a equação polinomial de dados se mistura em uma interação curvilínea entre as variáveis dependentes e independentes. (MAULUD; ABDULAZEEZ, 2020)

2.4 Bolsa de valores

Bolsa de valores se trata de um ambiente no qual é feita a negociação de ativos de renda variável. Ativos de renda variável são os investimentos em que os retornos de capital não podem ser estimados no momento da aplicação. Seus retornos dependem das expectativas do mercado financeiro, podendo variar positivamente ou negativamente. Os investimentos de renda variável mais comuns são as ações, imóveis e commodities. (TAVARES, 2018)

2.4.1 Ações

Ações podem ser definidas como a menor parcela de capital de uma empresa. Quando um investidor adquire ações de uma empresa, ele se torna sócio minoritário da empresa. Os lucros que o investidor obtém estão correlacionados à habilidade da empresa em gerar lucros e à demanda e oferta das ações no mercado financeiro. Algumas empresas possuem parte de seu capital aberto, o que permite que seja negociado na bolsa de valores. Nesse sentido, uma forma comum adotada para obter lucro na negociação de ações é realizar a compra de ativos por um preço e realizar a venda no futuro, quando o preço estiver mais caro. Ações são identificadas na bolsa de valores por meio de *tickers* - código único de cada ativo. (TAVARES, 2018)

2.4.2 Análise técnica

Análise técnica é utilizada para avaliar a compra de um ativo no mercado levando em conta o preço histórico para tentar prever o comportamento futuro do papel.

Para realizar essa previsão, vários dados sobre as ações podem ser considerados, como os preços máximo, médio e mínimo em um determinado período de tempo, o volume, padrões de gráficos [...] e diversos indicadores que são calculados através desses dados. (TAVARES, 2018)

3 Metodologia

O presente trabalho se concentra no estudo de casos múltiplos, nesse caso, relacionado às variáveis técnicas do mercado de renda variável brasileiro. Preços históricos de fechamento e volume de negociação de um determinado ativo são exemplos desse tipo de informação cuja fonte pode ser caracterizada como secundária, já que fora catalogada através de consulta a uma API externa. Os dados foram coletados por meio de levantamento a partir da ferramenta supracitada. Quanto a sua análise, empregou-se a técnica estatística multivariada, realizando recortes temporais a fim de aumentar a diversidade de dados que foram submetidos ao modelo.

Quanto à utilização dos resultados, o presente trabalho se caracteriza como uma pesquisa aplicada, buscando aferir a eficácia dos métodos de regressão na previsão de preços. Ademais, pode ser definido de natureza qualitativa, ao passo que busca abstrair noções de causa e efeito em relação aos dados coletados e projetados.

Por fim, é definido para este trabalho um objetivo exploratório, que busca entender o desempenho desses modelos no contexto proposto. Ao mesmo tempo é empregado a técnica laboratorial experimental, realizando aferições em ambiente controlado e observando os efeitos gerados.

4 Materiais

Foi utilizada a API Alpha Vantage para coleta de dados técnicos históricos, a linguagem de programação Python, que sustentou todo o desenvolvimento desde a aplicação de modelos à visualização e consolidação de resultados. Relativo a isso, boa parte do desenvolvimento se deu por meio de notebooks Python para melhor entendimento do processo. Por fim, foram utilizadas algumas bibliotecas para auxílio de diversas partes do desenvolvimento, dentre elas: SciKit Learn (implementação de modelos de regressão), Matplotlib e Seaborn (visualização e consolidação de resultados).

5 Desenvolvimento

A parte inicial do processo concentrou-se no levantamento de dados para construção dos modelos. Foi criada uma lista de cerca de cem *tickers* negociados na bolsa de valores brasileira. Para cada item da lista, foi feita uma requisição para a API passando o ticker para recuperar seus dados técnicos históricos. Ao fim desse processo, obteve-se um arquivo .csv com tais informações para cada linha da lista de ativos selecionados.

Foi implementado um modelo de regressão polinomial onde foram submetidos os valores de fechamento de cada .csv baixado. Como mostrado nos itens abaixo, cada ativo foi submetido ao modelo utilizando diferentes tamanhos de período de treinamento (direita) e para cada execução foram catalogados determinados dias da previsão (esquerda):

- 365 dias;
- 182 dias;
- 90 dias;
- 60 dias;
- 30 dias;
- 20 dias.
- 1° dia;
- 2° dia;
- 7° dia;
- 15° dia;
- 20° dia;
- 30° dia;
- 40° dia.

Além das variações descritas, pode ser visto na figura 1 os fatiamentos temporais que foram feitos nos preço de fechamento histórico para aumentar a quantidade em que o modelo de regressão seria submetido e construir uma amostragem maior, possibilitando uma aferição mais consistente dos resultados.

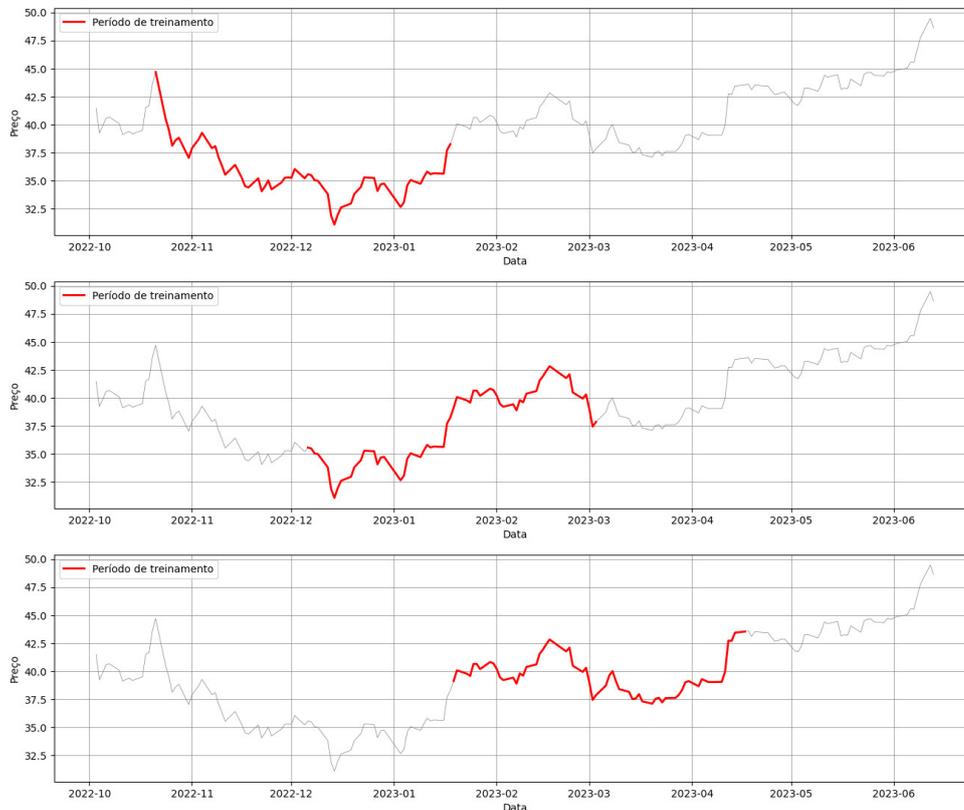


Figura 1: Fatiamento temporal do preço de fechamento histórico.

Para a consolidação dos resultados, foi construída uma planilha que em cada linha trazia dados a respeito de:

- O ticker do ativo relativo ao valor previsto
- A tamanho do período de treinamento utilizado (visto na tabela 1)
- Ponto na base histórica onde foi definido o intervalo de treinamento
- Ordem do dia relativo ao valor previsto (visto na tabela 1)
- Valor real daquele dia
- Valor previsto daquele dia

- Acurácia com base no valor real e previsto
- Validação de tendência, uma variável que é verdadeira se o algoritmo previu corretamente a tendência dos preços ou não independente de valores nominais

6 Resultados

Com base na planilha de resultados foram feitas diversas plotagens para tentar identificar onde o algoritmo de regressão polinomial conseguiu produzir previsões úteis. Nesse sentido, a validação de tendência foi a variável mais importante para identificar eventual justificativa para seu uso. Em uma plotagem inicial, como é mostrado na figura 2, fora levantada a quantidade de previsões certas e erradas em todos os mais de 45000 resultados.

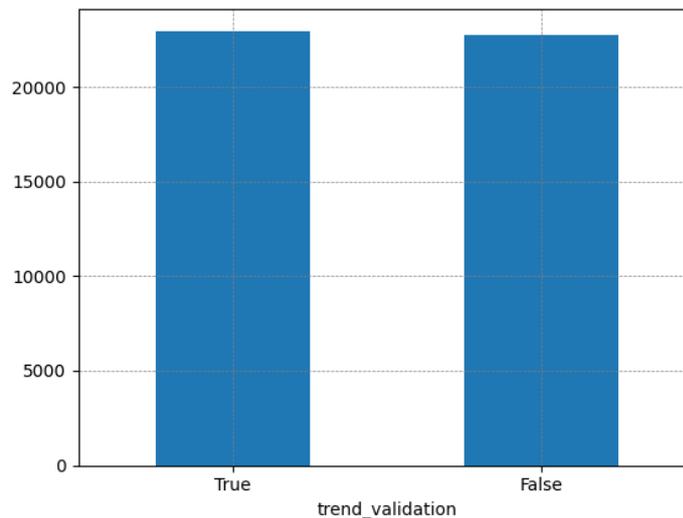


Figura 2: Validação de tendência.

A relação entre previsões de tendências certas e erradas é um bom indicador para aferir a confiabilidade do modelo para aplicação em cenários reais. Nesse sentido, quanto maior a margem de previsões corretas sobre erradas, mais válida a aplicação do modelo, pois significa que ele consegue descrever a tendências de preços de forma predominantemente correta.

Na busca por encontrar um recorte onde o modelo fora mais preciso, foi feita uma plotagem de validação de tendência segundo a quantidade de dias de treinamento, visto na figura 3. Como base nessa visualização, foi percebido que, de seis diferentes tamanhos de período de treinamento, apenas dois apresentaram validação correta em maior número que

validações equivocadas, sendo o período de 60 dias o mais promissor.

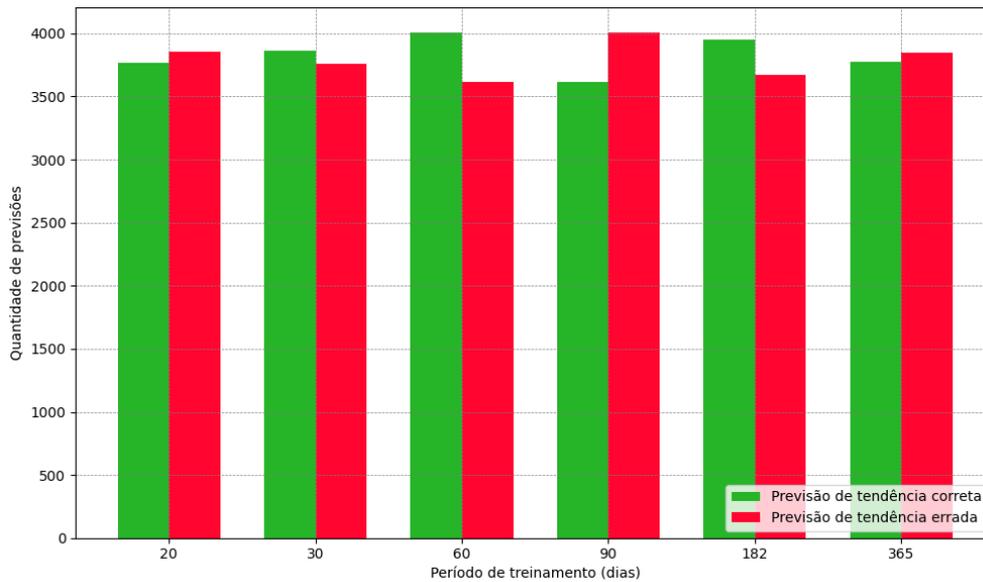


Figura 3: Validação de tendência em diferentes períodos de treinamento.

Buscando identificar de forma mais restrita onde a previsão foi mais de acordo, foram plotadas as validações de tendência segundo a ordem do dia de previsão para casos cujo tamanho do período de treinamento fora 60 dias, como pode ser visto na figura 4. Nessa visualização, foi possível identificar que em todas as ordens de dias previstos, a quantidade de tendência prevista de forma correta foi maior que as previsões erradas, sendo o 40º dia a melhor margem.

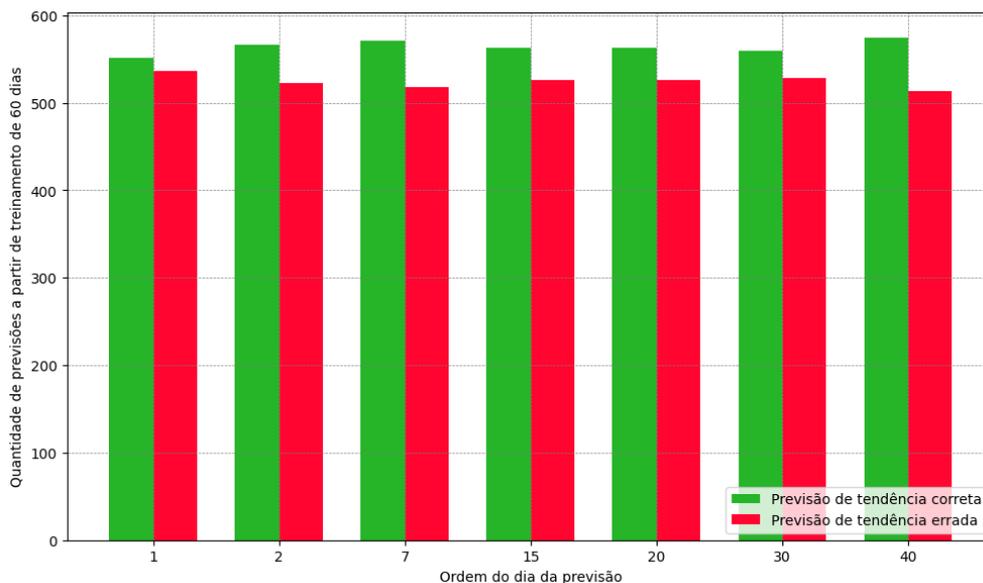


Figura 4: Validação de tendência para diferentes dias previstos com treinamento de 60 dias.

7 Considerações finais

Entende-se os resultados apresentados aqui são de característica preliminar, não sendo indicada a aplicação do modelo no estado atual em aplicações reais. Apesar de ser possível identificar recortes em que o modelo produziu projeções majoritariamente corretas, as margens apresentadas não são grandes o bastante para conseguir lucros de forma consistente no médio e longo prazo.

De forma a estender o que foi produzido, cabem a classificação e aferição de desempenho em diferentes tipos de ativos, aplicações de outros tipos de algoritmos de regressão, utilização de diferentes variáveis técnicas de forma conjunta e/ou inclusão de indicadores para treinamento do modelo.

Pode-se entender que a estratégia de aplicação de modelos de regressão para previsão de preços de ativos da bolsa de valores, no estado apresentado, não apresenta margens que forneçam segurança. São necessárias análises mais profundas e uma quantidade maior de testes para validar a eficácia do método, avaliando diferentes tipos de algoritmos de regressão e consolidando resultados a partir de métricas mais diversas, como aferição de desempenho segundo as características de cada ativo.

REFERÊNCIAS BIBLIOGRÁFICAS

EL NAQA, I.; MURPHY, M. J. O que é aprendizado de máquina? In: **Aprendizado de Máquina em Oncologia Radioterápica: Teoria e Aplicações**. Edição: Issam El Naqa, Ruijiang Li e Martin J. Murphy. Cham: Springer, 2015. cap. 1, p. 3–11. ISBN 978-3-319-18305-3. DOI: 10.1007/978-3-319-18305-3_1. Disponível em: <https://doi.org/10.1007/978-3-319-18305-3_1>.

JUNIOR, L. F. F. Previsão de preços de ações da bolsa de valores por redes neurais artificiais, 2022. Disponível em: <<http://hdl.handle.net/11449/238784>>.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, 2021. DOI: 10.1590/s0103-4014.2021.3510.0010.

MAULUD, D. H.; ABDULAZEEZ, A. M. A Review on Linear Regression Comprehensive in Machine Learning. **Computational Intelligence Magazine, IEEE**, IEEE, v. 01, n. 4, p. 140–147, 2020.

MENEZES, N. N. C. **Introdução à Programação com Python**. Primeira edição. [S.l.]: Novatec, 2010. ISBN 978-85-7522-250-8. Disponível em: <<https://s3.novatec.com.br/capitulos/capitulo-9788575222508.pdf>>.

SELBY, P. et al. BrAPI—an application programming interface for plant breeding applications. **Bioinformatics**, v. 35, n. 20, p. 4147–4155, mar. 2019. ISSN 1367-4803. DOI: 10.1093/bioinformatics/btz190. eprint: https://academic.oup.com/bioinformatics/article-pdf/35/20/4147/48976014/bioinformatics_35_20_4147.pdf. Disponível em: <<https://doi.org/10.1093/bioinformatics/btz190>>.

TAVARES, J. T. S. Sistema Automático de Negociação para a Bolsa de Valores Utilizando Redes Neurais Multilayer Perceptron e Regressão Linear. Universidade Estadual de Feira de Santana, Feira de Santana, 2018. Disponível em: <<http://tede2.uefs.br:8080/handle/tede/747>>.