



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO
DEPARTAMENTO DE CIÊNCIAS DE COMPUTAÇÃO

Inteligência Artificial – SCC0230
Prof^a. Dr. Solange Oliveira Rezende
13 de novembro de 2018

Aplicação de Técnicas de Inteligência Artificial na predição de ações da Bolsa de Valores

Autores:

Igor Trevelin Xavier da Silva - 10135354

Maurício Caetano da Silva - 9040996

Vitor Trevelin Xavier da Silva - 9791285

1. Introdução

Apesar do investimento na bolsa de valores ser cada vez mais cogitado, o número total de investidores brasileiros ainda continua sendo muito pequeno em relação à sua grande população [1][2].

Isso se deve, principalmente pelo medo do investidor nacional quanto à grande oscilação do mercado de ações. É sabido que em investimentos de renda variável, como no caso a bolsa, o retorno pode ser muito maior se comparado à investimentos de renda fixa mais conservadores, como imóveis ou a poupança.

Nesse contexto, este trabalho visa a criação de um software preditor que auxilie investidores a entender as tendências do mercado a curto prazo, utilizando para isso técnicas de Inteligência Artificial e a linguagem de programação Python.

2. Identificação do problema

Nesta seção será descrito o problema abordado por este trabalho, juntamente com uma pequena contextualização sobre o tema que o envolve.

2.1 Contextualização

A Bolsa de Valores é um mercado organizado onde são negociadas ações de empresas e outros valores imobiliários. As ações devem ser vistas como “fatias” negociáveis da empresa, ou seja, quando existe compra ou venda de ações o que está sendo negociado são partes de empresas.

Seguindo essa lógica, quando um investidor compra uma ação de uma empresa ele se torna um dos donos desta. Neste momento, esse investidor tem a opção de permanecer com a ação e passar a ganhar uma parte dos lucros da empresa (dividendos), ou então vender essa ação em um momento de valorização.

Assim como em qualquer tipo de mercado, no qual o preço de um produto varia de acordo com a oferta e demanda, o preço da ação de uma empresa na bolsa varia a todo momento de acordo com a quantidade de ofertas de compra e venda de ações, desse modo, o preço de uma ação diminui se a quantidade de ofertas de venda for maior do que as ofertas de compra, e aumenta caso contrário.

Do ponto de vista da empresa, a bolsa de valores é vantajosa pela possibilidade de receber investimentos em grandes quantidades sem a necessidade de recorrer à empréstimos bancários, pagando juros para isso, ou ainda por eliminar a necessidade dos donos injetarem dinheiro próprio na empresa. Para que uma empresa entre no mercado de ações ela deve concluir o processo de abertura de capital realizando uma Oferta Pública Inicial (sigla IPO, do inglês Initial Public

Offering), isso significa que, pela primeira vez, ela estará distribuindo ações em uma bolsa de valores, permitindo a acionistas adquirir partes consideráveis da companhia.

A venda de ações é realizada apenas durante o período de funcionamento da bolsa, esse período é chamado de “pregão”. Hoje em dia, o comércio de ações pode ser realizado tanto em ambiente real como virtual, sendo organizado e mantido pela instituição BM&FBovespa.

Para ser capaz de criar ofertas de compra e venda de ações na Bolsa de Valores pelo ambiente virtual, é necessário primeiramente possuir uma conta numa corretora que será responsável por intermediar toda a atividade do investidor, bem como fornecer os serviços de software para que o mesmo opere na bolsa.

a. Propostas

A proposta inicial deste trabalho seria criar um sistema preditor que trabalhasse em tempo real na bolsa de valores, porém, o principal empecilho para isso foi a obtenção de dados. Na maioria das vezes os sinais em tempo real eram pagos ou, quando gratuitos, tinham um atraso de 15 minutos, em média e isso trouxe uma mudança de estratégia.

Assim, objetivo do trabalho passou a ser a criação de uma ferramenta que, utilizando técnicas de inteligência artificial, fosse capaz de fornecer como resultado uma estimativa para o valor de fechamento do ativo de uma empresa ao fim de um determinado dia de pregão, auxiliando o investidor na tomada de decisão em operações de longo prazo na bolsa.

3. Pré-Processamento

Para a obtenção dos dados utilizamos o software *MetaTrader 5*, obtido por meio do site da corretora de valores *XP investimentos* [3]. É importante destacar que para a utilização desse software é necessária a criação de uma conta em alguma corretora de valores.

Foram utilizados os dados da empresa com o maior volume de ações negociadas no mês de janeiro de 2018 [4], a Petrobras.

3.1 Base de Dados

A base de dados utilizada representa o histórico de cotações diárias das ações da Petrobras, ativo PETR4, no período de janeiro de 2010 a novembro de 2018. Os dados foram extraídos utilizando o software *Metatrader 5* conectado ao

servidor Demo MQL5 da corretora XP Investimentos, responsável pelo roteamento dos dados fornecidos pela Bovespa.

Os dados estão dispostos em arquivos no formato CSV (Comma-Separated-Values), cada linha apresentando valores referentes à um dia de pregão para o respectivo ativo e as seguintes colunas são definidas:

DATE, OPEN, HIGH, LOW, CLOSE, TICKVOL, VOL e SPREAD

DATE: Data do respectivo pregão no formato YYYY.MM.DD;

OPEN: Preço de abertura do pregão (BRL);

HIGH: Preço máximo do pregão (BRL);

LOW: Preço mínimo do pregão (BRL);

CLOSE: Preço de fechamento do pregão (BRL);

TICKVOL: Quantidade de negociações do ativo do respectivo pregão;

VOL: Quantidade de ações, contratos futuros ou unidades de moeda negociados no ativo no respectivo pregão;

SPREAD: Valor recebido pela corretora por cada transação efetuada;

	DATE	OPEN	HIGH	LOW	CLOSE	TICKVOL	VOL	SPREAD
0	2010-01-04	30.19	30.49	30.08	30.49	13529	13303600	1
1	2010-01-05	30.54	30.58	30.06	30.23	22782	21396400	1
2	2010-01-06	30.06	30.64	30.06	30.64	18647	18720600	1
3	2010-01-07	30.45	30.60	30.29	30.35	12720	10964600	1
4	2010-01-08	30.36	30.55	30.11	30.19	14192	14624200	1
5	2010-01-11	30.40	30.51	29.92	30.09	16438	15317700	1
6	2010-01-12	29.89	29.97	29.44	29.71	18226	14886200	1
7	2010-01-13	29.87	30.02	29.28	29.66	23390	23228200	1
8	2010-01-14	29.64	29.79	29.05	29.14	21130	20075200	1
9	2010-01-15	29.06	29.49	29.01	29.21	19681	21169000	1

Imagem 1: Formatação dos dados de ações da Petrobras obtidos por meio do software MetaTrader 5.

3.2 Sobre a ferramenta

Este projeto foi desenvolvido em Python, uma linguagem aberta, orientada a objetos, de alto nível, multiparadigma e de tipagem dinâmica e forte em conjunto do *Jupyter-Notebook*, um aplicativo da Web de código aberto que permite criar e

compartilhar documentos que contêm código ativo, equações, visualizações e texto narrativo.

Escolhemos essa linguagem de programação pela sua praticidade, gerando alta produtividade, e pela vasta gama de bibliotecas (e.g.: análise de dados, *Machine Learning* e outros).

A seguir listamos as dependências para o correto funcionamento da ferramenta:

- **Python (versão maior ou igual a 3.5.3):** interpretador de comandos da linguagem Python.
- **Pandas:** biblioteca com ferramentas para análise de dados com facilidade de uso e alto desempenho.
- **NumPy:** biblioteca para computação científica, incluindo diversas funções matemáticas e estruturas de dados genéricas.
- **Scikit-learn:** biblioteca de *Machine Learning* com diversas ferramentas para *Data Mining* e *Data Analysis*.
- **Matplotlib:** biblioteca de plotagem 2D que produz figuras com qualidade de publicação em uma variedade de formatos.

Dentre as colunas selecionadas vamos trabalhar apenas com os valores: DATE, OPEN, HIGH, LOW, CLOSE e VOL, sendo as demais colunas descartadas. Essas colunas foram escolhidas pois segundo a análise de correlação, vista na seção “Extração de padrões” abaixo, são essas as que estão mais relacionadas ao nosso objetivo de predição do valor de fechamento.

Vale destacar também que o software *MetaTrader 5* fornece os dados sem valores inválidos nos sinais das ações, como pode ser visto pelo código abaixo.

Quantidade de células inválidas na base de dados

```
1 if ~dataset_clean.isna().sum().all():
2     dataset_clean = dataset_clean.dropna()
3
4 print(dataset_clean.isna().sum())
```

```
DATE      0
OPEN      0
HIGH      0
LOW       0
CLOSE     0
VOL       0
dtype: int64
```

Imagem 2: Número de valores inválidos na tabela da Petrobras.

```

1 # Remove colunas não interessantes da tabela
2 dataset_clean = dataset.drop(['TICKVOL', 'SPREAD'], axis=1)
3
4 # Mostra apenas as 10 primeiras linhas
5 dataset_clean.head(n=10)

```

	DATE	OPEN	HIGH	LOW	CLOSE	VOL
0	2010-01-04	30.19	30.49	30.08	30.49	13303600
1	2010-01-05	30.54	30.58	30.06	30.23	21396400
2	2010-01-06	30.06	30.64	30.06	30.64	18720600
3	2010-01-07	30.45	30.60	30.29	30.35	10964600
4	2010-01-08	30.36	30.55	30.11	30.19	14624200
5	2010-01-11	30.40	30.51	29.92	30.09	15317700
6	2010-01-12	29.89	29.97	29.44	29.71	14886200
7	2010-01-13	29.87	30.02	29.28	29.66	23228200
8	2010-01-14	29.64	29.79	29.05	29.14	20075200
9	2010-01-15	29.06	29.49	29.01	29.21	21169000

Imagem 3: Remoção de colunas não interessantes.

É importante destacar que devido a natureza dos dados utilizados, pouco pré-processamento foi necessário. Além disso, o código disponibilizado no arquivo *Jupyter-Notebook* trabalha com as 5 bases de dados disponíveis, como citado anteriormente na seção 3 (pré-processamento), cabendo ao usuário digitar o nome da ação a ser buscada.

4. Extração de Padrões

Antes de partir para a extração dos padrões é interessante observar como se comportam os dados de um modo mais amplo. Para isso, a imagem abaixo mostra um gráfico do valor de fechamento da ação Petrobras (PETR4) em todo o período analisado da base.

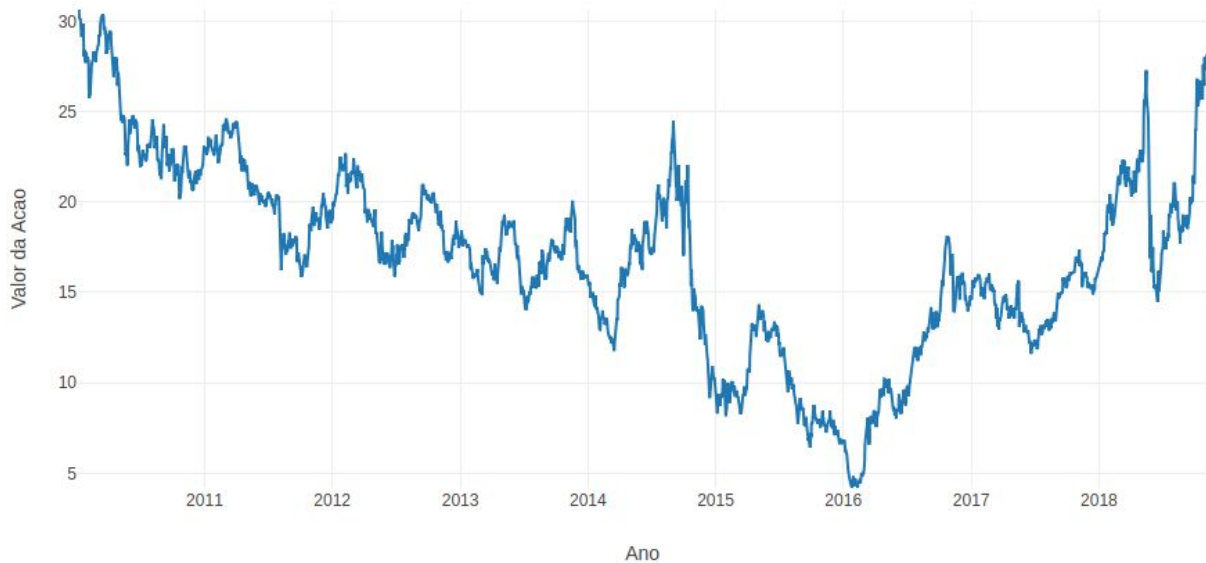


Imagem 4: Gráfico dos valores de fechamento para a ação PETR4 de 2010 a 2018.

É também interessante observar como se comporta a variação dos valores da ação, para isso, podemos observar o gráfico abaixo que mostra todas as variações no período analisado. Vale destacar que variação é calculado pela subtração entre preço de fechamento e preço de abertura.

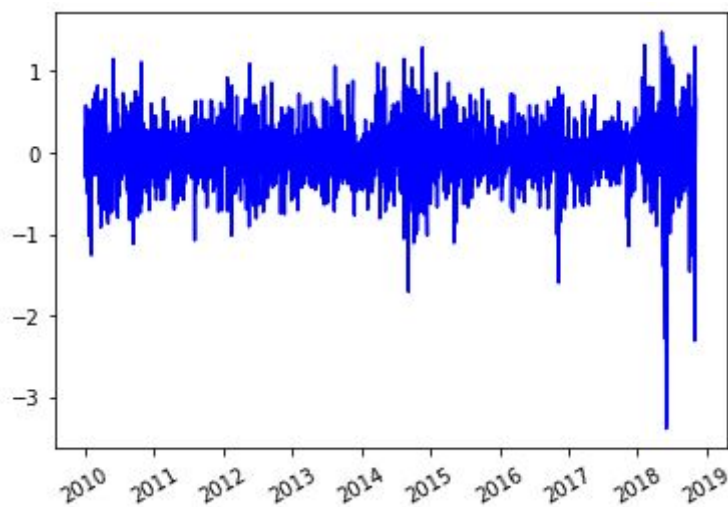


Imagem 5: Gráfico das variações para a ação PETR4 de 2010 a 2018.

Neste trabalho, os valores de predição serão feitos com base no conceito de regressão linear. A regressão linear é um modelo estatístico que examina a relação linear entre duas variáveis (Regressão Linear Simples) ou mais (Regressão Linear Múltipla) - uma variável dependente (Y) e variáveis independentes (X). Relacionamento linear basicamente significa que quando uma (ou mais) variáveis independentes aumenta (ou diminui), a variável dependente aumenta (ou diminui).

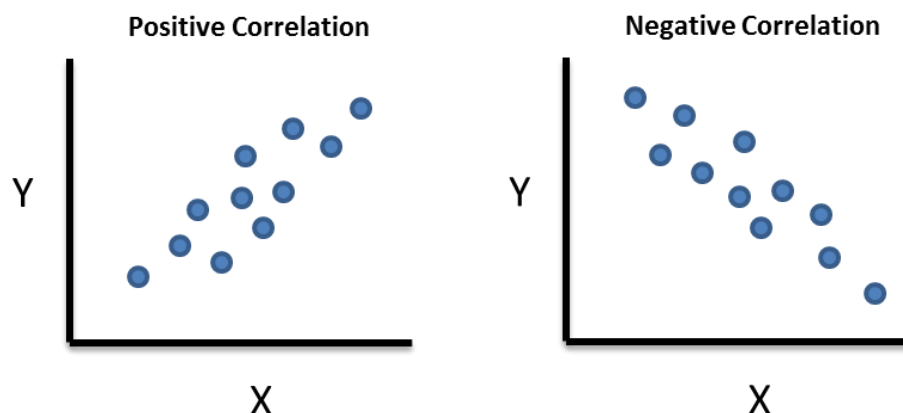


Imagem 6: Exemplos de correlação linear negativa e positiva.

Uma relação linear entre as variáveis Y e X é representada por esta equação:

$$Y = mX + b$$

Nesta equação, Y é a variável dependente - ou a variável que busca-se prever ou estimar; X é a variável independente - a variável que utilizada para fazer previsões; m é o declive da linha de regressão - que representam o efeito X tem em Y . Em outras palavras, se X aumenta em 1 unidade, Y aumentará exatamente em m unidades. Essas ressalvas nos levam a uma Regressão Linear Simples

Em um modelo de regressão, busca-se minimizar esses erros encontrando a “linha de melhor ajuste” - a linha de regressão dos erros seria mínima e isso está relacionado com (ou equivalente a) minimizar o erro quadrático médio.

No caso deste trabalho, foi utilizado um modelo de Regressão Linear Múltipla. A equação de regressão é praticamente a mesma que a equação de regressão simples, apenas com mais variáveis:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

em que cada valor β_i são os coeficientes associados a cada variável independente X_i .

O Python possui algumas bibliotecas que implementam a regressão linear e optou-se por utilizar a biblioteca *Scikit-Learn*.

Antes de se aplicar a regressão linear devemos observar se existe correlação entre as variáveis independentes selecionadas (valores de abertura, máxima, mínima e volume) com a variável dependente (fechamento).

4.1 Correlação entre variáveis

A seguir estão alguns gráficos que mostram a correlação entre os dados analisados e o valor de fechamento para ações da Petrobras.

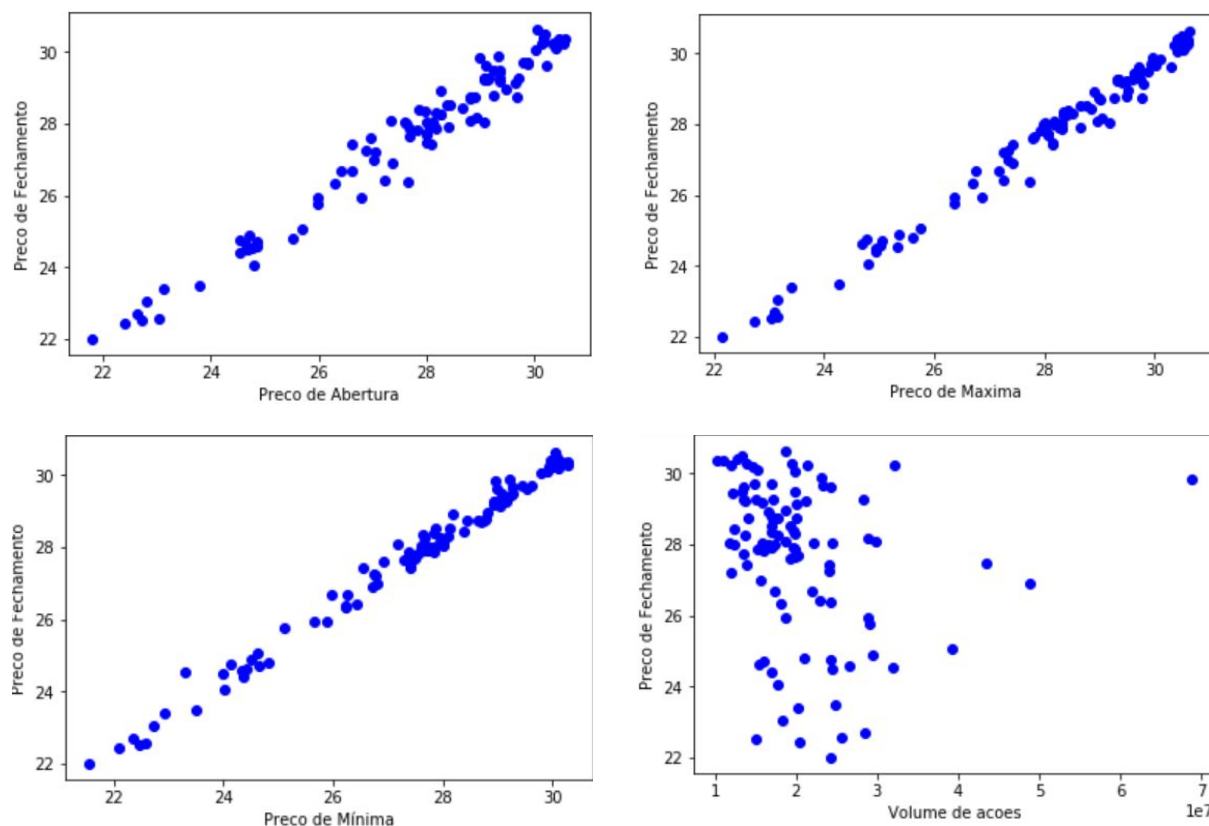


Imagem 7: Gráficos de correlação das variáveis selecionadas em relação ao preço de fechamento da ação da Petrobras.

Por meio da observação da imagem acima, pode-se afirmar que existe forte correlação entre os valores de mínima, máxima e abertura com relação ao preço de fechamento. Já para o caso do volume de ações podemos afirmar que existe uma fraca correlação, mas mesmo assim essa informação é interessante.

4.2 Aplicação da Regressão Linear

Ao se aplicar o algoritmo de Regressão Linear pelo Método dos Mínimos Quadrados nesse conjunto de dados, obtemos os seguintes resultados apresentados nas imagens a seguir:

DATE	OPEN	HIGH	LOW	VOL	REAL	PREDICTED	ERROR
2017-11-01	16.93	17.04	16.82	38875500	16.84	16.92	-0.08
2017-11-03	16.90	16.95	16.62	32605400	16.88	16.70	0.18
2017-11-06	16.99	17.37	16.92	46046000	17.36	17.23	0.13
2017-11-07	17.24	17.28	16.44	61098400	16.44	16.60	-0.16
2017-11-08	16.63	16.89	16.45	41179600	16.89	16.68	0.21
2017-11-09	16.83	16.88	16.66	29387900	16.66	16.72	-0.06
2017-11-10	16.65	16.75	16.45	35959400	16.66	16.56	0.10
2017-11-13	16.63	16.71	16.33	28207200	16.58	16.44	0.14
2017-11-14	16.44	16.46	15.22	88430700	15.29	15.44	-0.15
2017-11-16	15.56	15.98	15.42	42703800	15.75	15.78	-0.03

Imagem 8: Tabela com o valor predito e o respectivo erro para as dez primeiras tuplas do conjunto de teste.

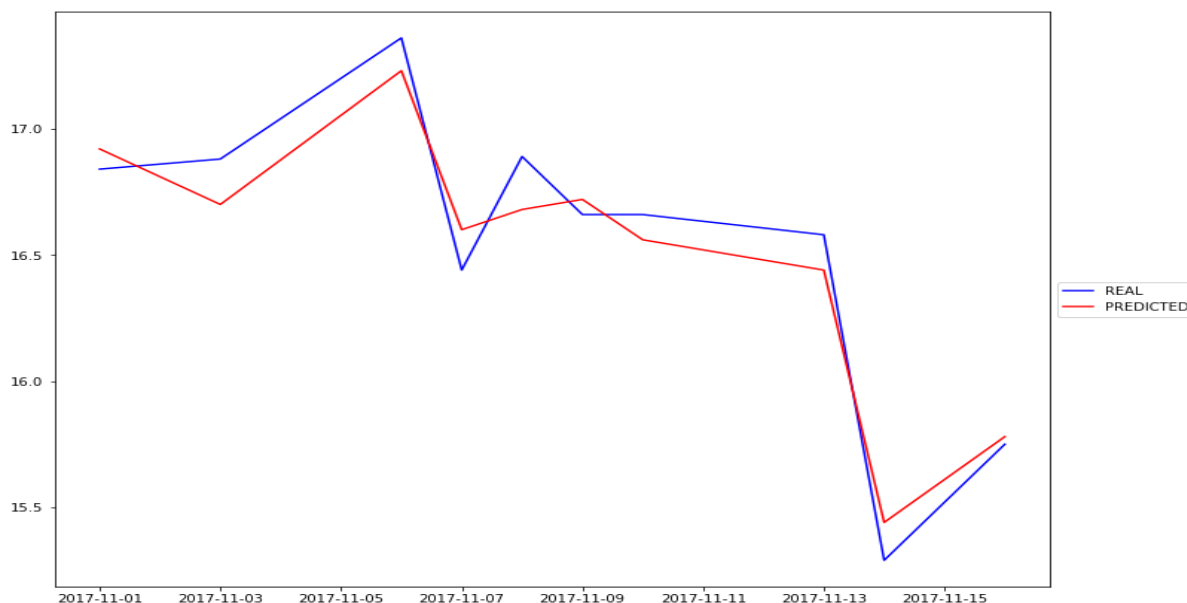


Imagem 9: Gráfico comparativo entre valores reais e preditos pelo modelo aplicado aos dados da PETR4.

5. Pós-Processamento

Para poder compreender os nossos resultados e entender se realmente foi possível extrair um indicador eficiente com o modelo de regressão linear utilizamos como métrica o coeficiente de determinação, também conhecido por R^2 .

O R^2 é uma medida de ajustamento de um modelo linear generalizado, como a regressão linear, indicando o quão bem o modelo explica os valores observados. Seu valor varia de 0 a 1 e, quanto maior, mais explicativo é o modelo.

O valor que obtivemos para essa métrica foi 0.9989744089142737

O valor de R^2 quando aplicado sobre nosso modelo de regressão assume o valor de 0.9989744089142737. Com isso, podemos concluir que houve o super ajuste do modelo, que é um resultado esperado, uma vez que o conjunto de dados utilizados para o ajuste do modelo pertence à mesma amostra que o conjunto de testes, ambos da PETR4. Se utilizarmos o modelo ajustado para a PETR4 para prever valores para outro ativo teríamos resultados muito ruins, pois os padrões extraídos pelo modelo mudam de acordo com o ativo.

6. Considerações Finais

Apesar do objetivo inicial de criar um preditor que atuasse em tempo real ter sido descartado, esse trabalho mostrou que é possível desenvolver um indicador para o valor de fechamento do pregão com certo grau de precisão, baseando-se nos valores de abertura, máxima, mínima e volume de ações negociadas para o respectivo ativo. Porém, dado a imprevisibilidade dos acontecimentos que podem influenciar os valores de máximo, mínimo e o volume de uma ação no decorrer no tempo, a aplicação apresentada, caso fosse integrada à uma aplicação real, poderia ser definida como um indicador, e não como preditor. O indicador forneceria uma projeção para o valor de fechamento do pregão, baseando-se no comportamento anterior do mercado para um determinado ativo.

É válido citar que durante o desenvolvimento do trabalho nos deparamos com diversas outras técnicas para tratar o problema de previsão na bolsa de valores, e algumas parecem realmente interessantes. Pensamos em uma abordagem em que primeiramente definiríamos diversas métricas estatísticas para a avaliação das cotações; em seguida, poderíamos utilizar a regressão linear para encontrar quais destas variáveis realmente são relevantes para a predição; e finalmente utilizaríamos séries temporais para realizar a predição do valor de fechamento de uma cotação.

7. Referências

[1] ORDONES, ARTHUR. Tem mais brasileiros na cadeia do que na bolsa: Por que tanta aversão ao mercado acionário?. 2013. Disponível em: <<https://www.infomoney.com.br/blogs/fora-do-mercado/blog-da-redacao/post/3071109/tem-mais-brasileiros-cadeia-que-bolsa-por-que-tanta-aversao>>. Acesso em: 11 nov. 2018.

[2] TAKAR, Téo. Estudo: brasileiro conhece Bolsa, mas fica na poupança ou não investe nada. 2018. Disponível em: <<https://economia.uol.com.br/financas-pessoais/noticias/redacao/2018/07/24/poupanca-investimento-preferido-e-bolsa-mais-lembrado.htm>>. Acesso em: 11 nov. 2018.

[3] XP Investimentos. 2018. Disponível em: <<https://www.xpi.com.br/>>. Acesso em: 11 nov. 2018.

[4] AZEVEDO, Rita. As ações mais negociadas na Bolsa no 2º melhor mês da história. 2018. Disponível em: <<https://exame.abril.com.br/mercados/as-acoes-mais-negociadas-na-bolsa-no-2-melhor-mes-da-historia/>>. Acesso em: 11 nov. 2018.