

Histórico do assunto

Existem diversos indicadores provenientes de estudos baseados em análise técnica do mercado. Porém estes indicadores não consideram em suas análises efeitos econômicos e políticos, como troca de presidentes brasileiros, escândalos, variação de câmbio (dólar) e variações nos preços de matérias primas e mudanças econômicas em mercados internacionais.

No caso dos dados econômicos é possível medir em termos de variações e flutuações diárias (dólar e minério de ferro p.ex.), como se fazem com as ações sendo sua correlação forte com determinados ativos. A importância de incluir dados econômicos como os citados acima preenche uma lacuna dentro da análise puramente técnica. A principal motivação para este projeto é minha intensa participação no mercado de derivativos ligados a VALE e PETROBRÁS e tem seus preços diretamente influenciados pelos ativos(ações) das mesmas.

Descrição do problema

Este projeto tem por finalidade criar um preditor de preço de ações que leva dados de negociação diários em um determinado período como entrada e gera estimativas projetadas para determinadas datas de consulta. As entradas conterão métricas, como preço de abertura, maior preço negociado, volume de ações negociadas, preço de fechamento e desdobramentos e dividendos e volatilidade.

As empresas brasileiras com maior participação na BOVESPA são VALE e PETROBRÁS, sendo assim seus ativos e derivativos tem alta liquidez gerada pelo alto volume de negociação diária, pelo fato destas empresas estarem presentes três novas entradas também serão inseridas pois são importantes e tem forte influência no preço de suas ações, são elas, preço do minério de ferro no caso da VALE e preço do barril de petróleo e variação no câmbio real/dólar no caso da PETROBRÁS. O sistema irá prever o preço do fechamento através de um Web App Python.

Conjuntos de dados e entradas

Descrevendo o conjunto de dados:

- **Preço de Abertura** - Preço com o qual o mercado inicia as negociações da ação.
- **Maior preço negociado** - Preço máximo alcançado no decorrer do dia de negociações, ele pode variar dentro de um mesmo dia.
- **Volume de Ações negociadas** - O valor total, em dinheiro, negociado no período. Ou seja, este número é a soma de todos os valores pagos por compradores (e recebido por vendedores) no período da cotação. Ela fornece indícios a respeito da intensidade de movimentos relacionados ao preço.

- **Volatilidade** - Volatilidade é um termo geral usado para descrever oscilações diárias nos preços (independentemente da direção). Geralmente, mudanças na volatilidade tendem a causar mudanças nos preços.
- **Preço de Fechamento** - O valor negociado na última operação de compra/venda da ação.
- **Preço do minério de ferro** - Como matéria-prima básica para empresas do ramo siderúrgico seu preço tem impacto muito forte nos preços das ações.
- **Variação Cambial** - Para empresas exportadoras de commodities, como os preços externos são negociados em dólar uma variação no câmbio é um parâmetro importante na precificação das ações.
- **Preço Barril de petróleo** - Para empresas ligadas ao setor petrolífero, o preço do barril de petróleo é fortemente correlacionado com seus ativos.

Abaixo seguem algumas fontes:

- <https://finance.yahoo.com/>
- <https://www.bloomberg.com/professional/support/api-library/>
- <https://www.quandl.com/home-v3>

Descrição da solução

Para solução do problema será utilizada a linguagem Python 3.5 e suas bibliotecas aplicadas à Machine Learning como a NumPy, SciPy e Pandas. Com os dados de entrada poderá ser avaliada sua correlação com os preços das ações dentro de um intervalo de datas, treinando e testando os dados e efetuando uma comparação dos resultados com um conjunto de dados reais. Os preços de ações e outros ativos na bolsa apresentam padrões que se repetem ou tendem a se repetir no decorrer do tempo (minutos, horas, dias, meses, anos) e isso é utilizado por Traders como forma de predição (tentativa) dos preços dos ativos. Como se trata de um problema contínuo o algoritmo de aprendizagem supervisionada serão testados algoritmos com AdaBoost, Decision Tree e outros.

Modelo de referência (benchmark)

O modelo de referência utilizado será a Regressão Linear, comparando o resultado desta com os outros modelos de Regressão em aprendizagem supervisionada.

Métricas de avaliação

Uma execução básica do sistema envolveria uma chamada para a interface de treinamento e uma ou mais chamadas para a interface de consulta. Um ciclo de teste para medir o desempenho será implementado. Estes testes irão verificar a precisão da previsão para as datas de consulta em intervalos de datas após a data de término do treinamento, por exemplo, 7 dias depois, 14 dias depois, 28 dias depois, etc. O resultado da predição será considerado como bom caso o valor do preço da predição esteja em +/- 5% do valor real, considerando o valor real o valor de fechamento. Como forma de

avaliação dos modelos de Regressão será utilizado o R^2 - O coeficiente de determinação, também chamado de R^2 , é uma medida de ajustamento de um modelo estatístico linear generalizado - assim os resultados dos modelos serão comparados para decidir qual o melhor dentre os utilizados.

Design do projeto

- Implementando o preditor de ações:
 - o Uma interface de treinamento que aceita um intervalo de datas (data_início, data_final) e uma lista de símbolos (ex. VALE3, PETR4), e constrói um modelo de comportamento das ações. O código deverá ler o histórico de preços desejados de uma fonte de dados a escolher.
 - o Uma interface de consulta que aceita uma lista de datas e uma lista de símbolos, e mostra uma saída com os preços preditos dessas ações nas datas desejadas. As datas de consultas passadas como parâmetro devem ser posteriores ao intervalo de datas do treino e os símbolos do marcador devem estar contidos no conjunto dos símbolos que foram treinados anteriormente.

- Teste e medida de desempenho

Os testes para medida de desempenho detalhados na seção anterior serão realizados aqui através das interfaces de treinamento e entrada de dados (datas e símbolos).

- Construir uma Interface de Usuário

Uma vez que os testes atingiram resultados satisfatórios, ou seja, de acordo com o que foi determinado em relação a distância dos valores reais dos preditos, uma interface de usuário amigável será implementada na qual permitirá especificar as ações que se está interessado e fornecerá previsões em intervalos pré-definidos.

- Melhorias

Após as etapas acima, serão avaliadas possíveis melhorias no projeto, estendendo o sistema para sugerir boas ações para comprar e vender, e quando. O sistema também poderá manter um portfólio de ações sugeridas para o usuário. Estes pontos de melhoria serão considerados caso haja tempo hábil após a realização dos requisitos básicos.

Referências

- <https://pt.wikipedia.org/wiki/R%C2%B2>
- https://pt.wikipedia.org/wiki/Regress%C3%A3o_linear
- <https://pt.wikipedia.org/wiki/Benchmarking>
- Python for Finance_ Analyze Big Financial Data.pdf