

Mineração de dados para analisar estratégia de negociação para o mercado de câmbio

Aurélio Miguel Machado da Silva¹, Raphael Araujo Bentes¹, Thiago Aparecido da Silva¹, Heloise Acco Tives Leão¹

¹Ciência da Computação - Centro Universitário Luterano de Palmas (CEULP/ULBRA)
Avenida Teotônio Segurado, CEP 77.019-900 – Palmas – TO – Brasil

{aumiguel.to, raphaelitb, thiagoaparecido2013, heloise.acco}@gmail.com

Resumo. *O trabalho em questão, tem como objetivo utilizar mineração de dados para analisar estratégia de negociação no mercado de câmbio. Neste trabalho será apresentado conceitos e características do mercado de câmbio, e por quais motivos e vantagens podemos utilizar a mineração de dados para explorar este mercado. Utilizando a mineração de dados foi realizada a análise utilizando o período de 18 anos de operações.*

1. Introdução

Mineração de Dados é um ramo da computação que teve início nos anos 80, quando os profissionais das empresas e organizações começaram a se preocupar com os grandes volumes de dados informáticos estocados e inutilizados dentro da empresa[De Amo 2004].

Nos dias atuais a mineração de dados tem tido um papel muito importante para o desenvolvimento científico e consecutivamente para os demais campos que possuem como base evolutiva, o conhecimento e pesquisas científicas. Uma das áreas que atualmente vem ganhando bastante notoriedade na mídia e vem sendo utilizado com mais frequência nas organizações, sendo utilizadas em diversos segmentos, porém ainda existem poucos estudos nessa área, se for comparar com o grande volume de dados que estão disponíveis para serem explorados.

Um dos campos em que as técnicas de mineração de dados possuem um amplo espaço para serem empregadas, são as áreas que envolvem os mercados financeiros. Essa área possui um volume extremamente grande de dados, tendo informações de grande relevância para que um investidor possa realizar a tomada de decisão de compra ou de venda de um determinado ativo. Tem informações ligadas a preço, notícias, relatórios contábeis, indicadores diversos, fluxos de ordens, entre outros.

Atualmente o mercado de câmbio é considerado o maior mercado que existe no mundo, por dia é movimentado em torno de 5,4 trilhões de dólares. Este movimento é maior do que o movimento financeiro de todas as bolsas de valores juntas. Conforme [Rossi 2010], o mercado de câmbio internacional (foreign exchange market, FX market, ou, simplesmente, Forex) é o lócus de negociação e troca entre moedas. Por definição, as transações nesse mercado estabelecem as taxas de câmbio spot e futura entre as diversas moedas do sistema internacional. O Forex é de longe o mercado o mais importante do mundo considerando como critério o volume de operações.

Com base nas possibilidades que são fornecidas pela mineração de dados, e pela capacidade de utilizar informações para gerar conhecimento e prever informações com relação ao preço de moedas. O artigo visa utilizar técnicas de mineração de dados, com o objetivo de analisar uma estratégia de negociação que contribua para o aumento de capital da carteira do investidor.

2. Revisão da Literatura

Cada vez mais, é possível notar a necessidade de que sistemas sejam agradáveis não apenas em seu manuseio, mas também que sejam úteis, trazendo comodidade aos seus usuários.

Para Stair (1998) os sistemas se configuram por conjuntos de elementos ou componentes que interagem entre si para atingir objetivos. Os sistemas baseados em computador disponibilizam meios de interação da máquina com homem, o qual por sua vez, utiliza os sistemas para realizar suas tarefas. Desta forma, a área Interação Homem-Computador (IHC), surge propondo consonância na ação exercida, de forma bilateral, entre o homem e a máquina (COSTA e RAMALHO, 2010).

Conforme Benyon (2011), sistemas interativos lidam cada vez melhor com a questão da interação homem-computador:

Sistema interativo é o termo que usamos para descrever as tecnologias com as quais o designer de sistemas interativos trabalha. Nesse termo pretende-se incluir componentes, dispositivos, produtos e sistemas de software, principalmente relacionados ao processamento da informação. Sistemas interativos são coisas que lidam com transmissão, exibição, armazenamento ou transformação de informação que as pessoas podem perceber. Eles são dispositivos e sistemas que respondem dinamicamente às ações das pessoas.

A interação homem-computador busca melhorar a qualidade dos sistemas que interagem conosco, pois seu desenvolvimento é voltado para o ser humano e não para a tecnologia. O design de sistemas focados em pessoas é muito importante para que os sistemas interativos sejam cativantes e que combinem adequadamente com as pessoas que as utilizam e seus estilos de vidas.

2.1 Trabalhos Relacionados

Um dos trabalhos relevantes para esta pesquisa, foi desenvolvido no Departamento de Ciência da Computação da Faculdade de Ciências Físicas da Universidade de Benin, onde explora várias áreas envolvidas com o contexto dos mercados, incluindo desenvolvimento de estratégias de negociação, utilização de dados históricos os que são tratados neste artigo. O trabalho, conforme o autor [Osunbor and Egwali 2016] descreve, o software proposto se encaixa na estratégia do trader. Fazendo um trabalho similar a um negociador, expert advisors analisam e integram enorme quantidade de dados presentes e históricos, processando-os muito rapidamente, para efeitos de previsão de movimentos futuros de preços, a fim de fornecer um expert advisors com compra, venda e recomendações que se enquadram dentro do estratégia.

Sobre sistemas de negociações automatizados, o autor [Barbosa and Belo 2010], descreve que, a fim de obter uma vantagem sobre seus concorrentes, instituições financeiras e outras os participantes têm dedicado e aumentando a quantidade de tempo e recursos para o desenvolvimento de sistemas de negociação automatizados. Por razões óbvias, a maioria desses sistemas é proprietária, e sua implementação é envolta em sigilo. É, portanto, muito difícil encontrar informações publicamente disponíveis sobre a arquitetura e desempenho de este tipo de sistemas.

Sobre técnicas de predição voltada para o mercado de câmbio, o autor [] descreve, Predição de preços de moedas se tornou uma tarefa desafiadora no Forex mercado desde o final da década de 1970 devido ao movimento de incerteza das taxas de câmbio. No trabalho em questão, utilizamos a equação de regressão linear para analisar a dados e descobrir os padrões de tendências no Forex. Esses padrões de tendências são modelados e aprendidos pelo algoritmo da rede neural artificial. O algoritmo Time Warping é usado para prever as tendências do futuro próximo. O experimento em questão mostra um resultado satisfatório usando a abordagem proposta.

3. Metodologia

Nesta seção serão descritos os métodos utilizados para realizar o projeto em questão, ferramentas para a coleta de dados, procedimentos para realizar a normalização e limpeza dos dados, programas e linguagem utilizadas para aplicar a mineração de dados e algoritmos para obter o resultado e modelo mais apropriado para o desenvolvimento da estratégia de negociação.

O processo de para realizar a extração de conhecimento escolhido foi o (KDD - Knowledge Discovery in Databases) Descoberta de conhecimento em bancos de dados. No contexto do artigo, o KDD foi escolhido por ser um processo com fluxo simples, bem definido e as etapas do processo se encaixaram bem nos dados do projeto desenvolvido.

A Figura 1 demonstra essas fases e como elas se integram. Utilizando o KDD, na etapa de seleção, foi escolhido o conjunto de dados para realizar a busca pelo conhecimento.

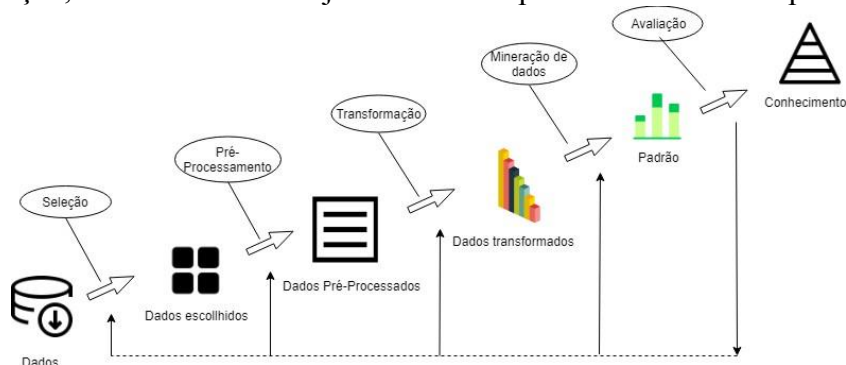


Figura 1. Fases do processo KDD

Os dados escolhidos nessa etapa foram as cotações de preço do par EUR/USD dos últimos 18 anos, o período de tempo escolhido para as análises iniciais foi o período diário, além das informações com relação a preço, outros dados utilizados fazem parte do data-set do projeto, o indicador RSI(Índice de força relativa).

Posteriormente, na etapa de pré-processamento, foi realizado a limpeza dos dados, excluindo colunas irrelevantes para o problema, normalizando e limpando dados de variáveis como a de data.

Já na etapa de transformação, foi anexado ao mesmo data-set as informações obtidas a partir dos indicadores, pois na coleta dessas variáveis, foi realizado uma coleta separada, para que os dados dos indicadores viessem mais limpos, evitando a necessidade de se trabalhar a etapa de pré-processamento desses dados em questão.

Após os dados terem sido normalizados, limpos e transformados, será realizado a mineração dos dados. Para realizar a mineração, foi escolhido o classificador Naive Bayes. Na última fase, a partir dos resultados adquiridos com base no algoritmo descrito no parágrafo anterior, será realizado a etapa de interpretação de resultados, as informações obtidas foram primeiramente interpretadas, em seguida foi possível identificar novos padrões e modelos.

4. Desenvolvimento

As etapas de desenvolvimento foram desenvolvidas utilizando o processo KDD, conforme foi descrito na seção anterior, em todas as fases do processo foram gerados materiais, nesta seção será descrito com detalhe os procedimentos utilizados e os produtos, dados, informações e conhecimentos obtidos.

A primeira etapa do KDD tem como foco a escolha dos dados a serem trabalhados. No contexto do artigo os dados escolhidos foram as cotações diárias de preço do EUR/USD, dados diários do indicador RSI(Índice de Força Relativa). Para realizar a coleta dos dados foi desenvolvido um algoritmo em linguagem MQL5, utilizando-se da ferramenta de

back-test, disponível no software MetaTrader5 foi possível obter os dados desde o ano 2000, para realizar as demais fases do KDD.

Na fase de pré-processamento dos dados, foi necessário realizar limpeza e normalização dos dados, pois o algoritmo utilizado para a coleta dos dados foi desenvolvido com foco em capturar os dados históricos.

A transformação dos dados foi realizada integrando os dados das cotações com os dados dos indicadores, pois é necessário que os dados dos indicadores sejam referentes as determinadas datas de cada cotação.

Utilizando a ferramenta Weka e Metatrader, foi realizado uma série de experimentos com algoritmos voltados para predições, em todos os experimentos a variável alvo era: compra, venda ou neutro. as variáveis independentes foram as variáveis de cotação de preço, volume, variação de preço e os dados do indicador.

5. Resultados

Com a realização do trabalho em questão, um dos resultados e produtos mais relevantes foi o dataset construído para a realização dos experimentos, além de um grande volume de objetos coletados, a integração com variáveis obtidas a partir dos indicadores possibilita a realização de estudos futuros.

A partir dos experimentos realizados na etapa de desenvolvimento, foi possível extrair diversas informações sobre o comportamento do mercado e de como os indicadores podem auxiliar o ser humano no processo de tomada de decisão de compra, venda ou de aguardar uma melhor oportunidade.

A estratégia visa explorar momentos de sobre-compra e sobre-venda, esses momentos podem ser identificados quando o RSI cruza a zona de 70, indicando que o preço está sobre-comprado, quando o RSI cruza a zona de 30, indica que o preço está sobre vendido. Quando o preço cruza a zona de 70 é realizado uma venda, quando cruza a zona de 30 é realizado uma compra. É executado apenas uma operação por vez.

A partir do classificador Naive Bayes, utilizando informações de abertura, máxima, mínima, fechamento, volume e valor do RSI. As informações obtidas a partir do algoritmo podem ser vistos na figura abaixo.

Correctly Classified Instances	241	72.1557 %
Incorrectly Classified Instances	93	27.8443 %
Kappa statistic	0.4435	
Mean absolute error	0.3504	
Root mean squared error	0.423	
Relative absolute error	70.0958 %	
Root relative squared error	84.6099 %	
Total Number of Instances	334	

Figura 2. Resultado Classificador Naive Bayes

Com o objetivo de avaliar a performance da estratégia utilizando esse indicador, foi desenvolvido um robô para demonstrar o resultado dessa estratégia utilizando dados históricos. Segue abaixo a figura da performance da estratégia utilizando RSI para operações no tempo gráfico D1(Um dia), com o take-profit, stop-loss e stop-gain fixos, sendo eles de 1000 pips. O resultado de performance realizado é entre 01/01/2000 até 18/04/2018.



Figura 3. Resultado RSI para tempo gráfico D1

Os resultados utilizando o indicador RSI para o período gráfico se mostrou insatisfatório, levando em consideração o acumulo de capital a longo prazo ou curto prazo. Tendo uma taxa de 47,3% de operações com lucro versus 52,69% de operações com perda.

A partir da estratégia definida anteriormente, diversos outros experimentos foram realizados, utilizando o mesmo algoritmo, fazendo alterações apenas no tempo gráfico.

Alterando o tempo gráfico foi possível identificar uma taxa de acerto um pouco maior em comparação ao primeiro experimento, porém no longo prazo foi possível obter resultados mais interessantes. Alterando apenas o tempo gráfico para o H1(Uma hora), obtivemos o seguinte resultado. A taxa de operações com lucro foi de 50,90% versus 49,10% de operações com perda.



Figura 4. Resultado RSI para tempo gráfico H1

A mineração que mostrou a melhor performance de saldo e taxa de acerto foi para o tempo gráfico M30(Trinta minutos). Mantendo a mesma estratégia o resultado, a taxa de operações com lucro foram de 51,74% versus 48,26%. Abaixo segue a imagem com a performance mantendo o mesmo período.



Figura 5. Resultado RSI para tempo gráfico M30

Para o período D1, a estratégia em questão realizou 520 operações, no período de H1, os resultados são com base em 2118 operações, para o período M30 foram 2443 negociações.

6. Conclusões

Os dados coletados a partir da mineração de dados utilizando o KDD, foram de grande valia para a realização do trabalho, e para o desenvolvimento eficiente da estratégia de negociação.

Com base nas informações colhidas a partir dos dados, foi possível verificar a validade e a tendência de movimento do preço do par EUR/USD, utilizando como auxilio para a análise o indicador RSI.

Conforme foi apresentado neste trabalho, a mesma estratégia performa de diversas formas diferentes quando as variáveis são alteradas, a estratégia utilizada para desenvolvimento desse trabalho, inicialmente se mostrou ineficiente para acumulo de capital, porém alterando apenas uma das variáveis, nesse caso o tempo gráfico, a mesma estratégia se tornou positiva.

Além dos produtos gerados, citados nessa seção. O trabalho em questão teve como produto um robô de negociação, o robô desenvolvido em MQL5, tem como objetivo auxiliar as pessoas em operações no mercado financeiro, possibilitando a abertura de operações durante 24 horas do dia, durante todos os dias da semana.

Conforme descrito na seção de resultados, a estratégia se mostrou efetiva tendo como objetivo o acumulo de capital no longo prazo. Como é de conhecimento do mercado, existem outros fatores não abordados nesse trabalho são importantes para a escolha de uma estratégia de negociações. A pesquisa aqui apresentada ainda está em fase inicial e como trabalhos futuros pretende-se continua a análise agregando novos fatores na mineração para o período já determinado

Referências

- Barbosa, R. P. and Belo, O. (2010). Multi-agent forex trading system. In *Agent and multi-agent technology for internet and enterprise systems*, pages 91–118. Springer.
- Campos, P. J. F. (2010). Abordagem comportamental e investimento racional no mercado cambial: o caso do mercado forex. PhD thesis, Instituto Politécnico do Porto. Instituto Superior de Contabilidade.
- De Amo, S. (2004). Técnicas de mineração de dados. *Jornada de Atualização em Informática*.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.
- Fayyad, U. M., Haussler, D., and Stolorz, P. E. (1996b). Kdd for science data analysis: Issues and examples. In *KDD*, pages 50–56.
- Guimarães, E. M. P. and Évora, Y. D. M. (2004). Sistema de informação: instrumento para tomada de decisão no exercício da gerência. *Ciência da informação*, 33(1).
- Osunbor, V. and Ekwali, A. (2016). Development of oseg: A forex expert advisor. *The Pacific Journal of Science and Technology*, 17(2).
- Prezepiorski Lemos, E., Arns Steiner, M. T., and Nievola, J. C. (2005). Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. *Revista de Administração-RAUSP*, 40(3).