

US 24 - Como analistas de dados o que acreditam que pode ser interessante ser investigado neste trabalho?

Como investidor e cliente, gostaria de saber a opinião dos analistas quanto a trabalhos futuros que não foram possíveis de serem estudados quanto ao escopo deste projeto e que futuramente pode representar algo interessante

A nossa sugestão está primariamente relacionada com a forma de correlacionar as notícias e o preço das ações.

1) Melhorias na Coleta das Notícias

Inicialmente deve-se rever a forma de coleta das notícias. A string de busca pode ser melhorada para cada empresa, considerando até várias buscas para uma mesma empresa de forma a maximizar os resultados (notamos que a maioria das empresas teve poucas notícias) e minimizar as páginas não relacionadas.

Uma sugestão muito mais interessante seria buscar no [Google News](#) utilizando todo o seu potencial de busca que há reúne notícias de várias fontes e permite a formação de strings de busca específicas e poderosas ([exemplo de busca](#) da palavra petrobras no Brasil)

2) Melhorias no Processamento das Notícias

Já que as cotações são valores numéricos, as notícias devem também ser expressadas em números. No projeto contou-se a quantidade de notícias em um dia por fonte de busca. Outra forma seria realizar a análise sentimental das notícias, o que não é realizado de forma trivial pois notícias são diferentes de tweets que tem *emotions*. Ou até manualmente, como sugerimos na próxima atividade.

3) Melhorias na Análise da Correlação

A análise de correlação do projeto não teve bons resultados e vários podem ser os motivos, então, para identificar as principais causas e tentar propor os novos passos sugerimos que seja realizada a seguinte atividade:

1. Seleciona-se algumas empresas e um período de tempo (dias com solavancos);
2. Seleciona-se as notícias para essas empresas nesse tempo;
3. Para cada notícia daria uma pontuação de -10 a +10 de acordo com o peso da notícia para a empresa;
4. Seleciona-se a cotação de cada empresas nesse tempo;
5. Compara a cotação com a média do peso das notícias para cada dia;
6. Compara a interpretação do ponto anterior com o resultado da correlação de pearson.

Essa atividade simples, poderia definir o curso do que deveria ser de fato, automatizado nessa análise. Respondendo algumas perguntas:

- Será se a análise de sentimento alcançaria essa análise que fizemos (esses valores já seria um conjunto de teste para o modelo)?
- Será se a notícia gera o solavanco no preço ou o solavanco no preço gera a notícia?
- Será se a correlação de pearson consegue capturar bem a correlação real? Ou deve-se buscar outra? Ou é melhor gerar uma nova função função de peso para esse contexto específico?

A ferramenta *Stocks* permite que seja analisada a série da cotação marcada com solavancos por empresa e que sejam visualizadas as notícias daquele dia. Então, a ferramenta pode ajudar, de certa forma, a realizar essa atividade.

4) Melhorias na Visualização

A visualização poderia ser melhorada na visualização da série temporal da cotação por empresa. Poderia ser adicionada também a série com o peso das notícias (quantidade, peso sentimental, etc.) logo abaixo com interação simultânea em ambas, tanto de *mouse hover* como na seleção de intervalo no eixo temporal. Assim o usuário poderia realizar a tarefa de análise da correlação entre ambas mais facilmente.