

5

Filtragem Por Tópico

Algumas notícias coletadas apresentam informações sobre todo o mercado financeiro e não só sobre a Petrobras. Porém, esse trabalho visa classificar o sentimento das notícias com relação apenas a referida empresa petrolífera. O exemplo abaixo mostra uma notícia sobre o mercado.

Entre as ações com maior peso na carteira teórica Petrobras PN (PETR4) valorizava 0,47%, para R\$ 35,81; Vale PNA (VALE5) perdia 0,49%, aos R\$ 38,57; Itaú Unibanco PN (ITUB4) mostrava retração de 0,11%, a R\$ 35,96

Nesse exemplo, a notícia reporta ganho para a Petrobras, porém perda para as outras duas empresas. Como não estamos classificando o sentimento do mercado como um todo, e sim apenas o da Petrobras, a informação negativa sobre as outras companhias pode causar ruído no momento de classificar o sentimento da notícia e piorar a acurácia do experimento. A Petrobras não possui nenhum competidor cujo sentimento está inversamente ligado ao dela, dessa forma, a informação sobre outras empresas não é relevante para nosso estudo. Por isso, filtramos apenas as frases sobre a Petrobras. Nosso filtro é baseado em palavras chaves, filtrando automaticamente apenas frases em que essas palavras estão presentes. Após alguns experimentos, verificamos que apenas as seguintes palavras: *Petrobras*, *Petrolífera* e *petróleo*, são suficientes para coletar informações significativos sobre a empresa.