



Revista de Finanças Aplicadas

ISSN 2176-8854

WWW.FINANCASAPLICADAS.NET

MODELO DE PREVISÃO DE LUCROS DE COMPANHIAS LISTADAS NA BMF&BOVESPA BASEADO EM ANÁLISE DE BALANÇOS, INDICADORES MACROECONÔMICOS E MONITORAMENTO DE NOTÍCIAS

Leandro Guerra, Serasa Experian
Pós Graduado em Inteligência Analítica pela FIA
eng.leandroguerra@gmail.com

Regina da Silva Ornellas, FEA
Doutoranda em Administração de Empresas pela FEAUSP
reginaornellas@usp.br

MODELO DE PREVISÃO DE LUCROS DE COMPANHIAS LISTADAS NA BM&FBOVESPA BASEADO EM ANÁLISE DE BALANÇOS, INDICADORES MACROECONÔMICOS E MONITORAMENTO DE NOTÍCIAS

OBJETIVO

Desenvolver, com modelagem matemática, uma equação para estimar a variação percentual do lucro das companhias de capital aberto, identificando as variáveis relevantes para explicação deste fenômeno.

METODOLOGIA

Foram selecionadas empresas que possuem ação listada em bolsa, dos segmentos de matérias prima, siderurgia e bancário, pois, somados, representam mais de 50% do índice. Assim, foram criadas variáveis baseadas nos relatórios de balanço destas empresas. Também foram criadas variáveis baseadas em indicadores macroeconômicos e nas notícias correlatas à economia e às empresas. Assim, foi desenvolvido um modelo matemático analisar a influência destas variáveis na variação percentual do lucro destas companhias.

RESULTADOS E CONCLUSÕES

Como resultado, após o desenvolvimento de uma regressão para cada uma das empresas selecionadas, observa-se que as variáveis baseadas nas notícias de mídias eletrônicas (as variáveis não estruturadas) estão presentes e são relevantes para todos os modelos, ampliando o poder de explicação da variável dependente. Cada empresa apresentou um conjunto distinto de variáveis, o que evidenciou a influência das particularidades de cada segmento.

IMPLICAÇÕES PRÁTICAS

Criação de uma ferramenta que permita automatizar a análise de investimento, principalmente ao investidor pessoa física e pequenas empresas, conseguindo absorver as notícias e efeitos macroeconômicos nos resultados das empresas.

PALAVRAS-CHAVE

Decisão de Investimento, Mineração de Textos, Mercado Financeiro, Modelagem Matemática.

PREDICTIVE MODEL FOR PROFITS OF LISTED COMPANIES ON BM&FBOVESPA BASED ON ANALYSIS OF BALANCE SHEET, MACROECONOMIC INDICATORS AND MONITORING NEWS

OBJECTIVE

The goal is to develop, with mathematical modeling, an equation to estimate the percent change from earnings of companies, identifying the relevant variables that explain this phenomenon.

METHODOLOGY

We selected companies listed on the Brazilian stock exchange, from verticals such as raw materials, steel and banking, representing together more than 50% of the index. Thus, we created variables based on the reports of these companies. We also created variables based on macroeconomic indicators and related to the economy and the news about these companies. Thus, it was developed a mathematical model to analyze the influence of these variables on the percentage change of the earnings of these companies.

RESULTS AND CONCLUSIONS

As a result, after the development of a multi linear regression model for each of the selected companies, it was observed that the variables based on news from electronic media (the unstructured variables) are relevant to all models, expanding the power of explanation of the dependent variable. Each company presented a distinct set of variables, which showed the influence of the characteristics of each segment.

PRACTICAL IMPLICATIONS

This is the starting point to create a tool to automate the analysis of investments, mainly to individual investors and small business, absorbing the news and macroeconomic effects on business outcome.

KEYWORDS

Investment Decision, Text Mining, Financial Markets, Mathematical Modeling.

Este artigo tem como objetivo desenvolver, com modelagem matemática, uma equação para estimar a variação percentual do lucro das companhias de capital aberto, identificando as variáveis relevantes para explicação deste fenômeno.

INTRODUÇÃO

Tradicionalmente, no mercado de capitais, existem duas metodologias para a análise das ações de uma empresa: a análise técnica e a análise fundamentalista. A análise técnica consiste em realizar uma análise dos gráficos históricos das cotações das ações, juntamente com a análise de outros indicadores técnicos, para tentar antecipar o movimento futuro de uma ação (FORTUNA, 2011). Esta teoria, conhecida como teoria de Dow (desenvolvida pelo jornalista estadunidense Charles Henry Dow) possui como princípio o fato de que os índices de preços descontam tudo; todos os possíveis fatores que afetam a cotação dos preços dos ativos (ações) são descontados por esses índices que consideram todas as notícias, resultados contábeis e financeiros.

Por sua vez, a análise fundamentalista busca de modo simplificado, avaliar a saúde financeira das empresas, projetar seus resultados futuros e determinar o preço justo para as suas ações (FORTUNA, 2011). Para isso, os analistas levam em consideração os chamados fundamentos da empresa, isto é, todos os fatores macro e microeconômicos que influenciam no seu desempenho, bem como as perspectivas de lucro da companhia (ROSS, WESTERFIELD, JORDAN, 2011).

A pergunta que norteia este trabalho é: “É possível estimar qual será a variação no lucro de uma companhia listada em bolsa na próxima publicação de resultados, para auxiliar na tomada de decisão de investimentos?” Assim, este assunto foi investigado, pois se acredita que a sua proposta é inovadora para o mercado de capitais, unificando os princípios das duas principais teorias de análise do mercado de capitais através de um processo automatizado e fundamentado matematicamente. O estudo foi desenvolvido relacionando notícias de mídias eletrônicas, indicadores macroeconômicos, e balanços financeiros, através de uma modelagem matemática.

Outro fator de grande relevância é o auxílio que este modelo poderá dar para a identificação de empresas lucrativas, gerando maior retorno para o acionista, além de facilitar a tomada de decisão de investimentos para o pequeno investidor pessoa física. Isto se deve ao fato de o investidor nem sempre tomar decisões baseadas na racionalidade e a grande quantidade de informações disponíveis pode dificultar estas decisões. Desta forma, é interessante ao investidor ter ferramentas que o auxiliem avaliar um contexto mais amplo do mercado no momento de realizar seus investimentos (MALUF, 2010).

FUNDAMENTAÇÃO TEÓRICA

O Índice Bovespa - Ibovespa

No caso do mercado de ações, o Índice Bovespa é o mais popular e mais aceito indicador de desempenho do mercado brasileiro de ações. Sua metodologia de cálculo consiste em uma aplicação hipotética em uma carteira teórica de ações a partir de uma aplicação hipotética, sem nenhum investimento adicional, salvo o reinvestimento nos dividendos e a manutenção as ações recebidas ao título de bonificação. O índice é composto das ações que, de forma conjunta, representam mais de 80% das negociações e do volume financeiro do mercado brasileiro de ações. O peso ou participação destas ações no índice é ajustado de acordo com a representatividade deste título no mercado à vista, sendo que, quadrimestralmente, é feita uma reavaliação do índice e, se necessário, adequação de pesos e composição do mesmo. Assim, uma vez que a ação compõe o índice, ele apenas deixará de constar na composição deste se não conseguir atender aos critérios supracitados (FORTUNA, 2011).

O lucro para empresas

Existem várias etapas de apuração do lucro, até chegar-se ao lucro (ou prejuízo) líquido. O primeiro indicador é o lucro bruto, originado da diferença entre a venda das mercadorias ou serviços e o custo desta mercadoria ou serviço, sem considerar as despesas administrativas, de vendas e financeiras. Assim o lucro bruto é destinado à remuneração das despesas, ou seja, quanto maior for este número, maior será a remuneração dos administradores, diretores, massa salarial, dentre outros (MARION, 2004).

Na sequência, tem-se o lucro operacional, que é utilizado para medir o desempenho no negócio da empresa. Ele é contabilizado pela diferença entre o lucro bruto e as despesas operacionais. No âmbito não operacional, pode-se obter o lucro antes do imposto de renda (LAIR), subtraindo-se do lucro operacional as despesas não operacionais somadas as receitas operacionais. Do LAIR, chega-se ao lucro depois do imposto de renda subtraindo deste o imposto de renda e a contribuição social (MARION, 2004). Por fim, chega-se ao lucro líquido, que é o que será destinado aos sócios, acionistas e investidores – é o lucro depois do imposto de renda subtraído das doações e participações (remuneração destinada aos administradores e empregados) da empresa. Os acionistas decidirão qual será a parcela do lucro líquido que ficará retida e a parte que será distribuída na forma de dividendos (ROSS, WESTERFIELD, JORDAN, 2011) e (MARION, 2004).

Relações com investidores

Com a abertura do mercado brasileiro na década de 90, as companhias de maior porte listadas em bolsa puderam acessar o mercado externo para obter financiamento por meio do lançamento de ações ou títulos no exterior. Entretanto esta nova oportunidade proporcionou as empresas um contato com acio-

nistas habituados a investir em mercados com práticas de governança corporativa mais avançadas do que aplicadas no mercado brasileiro. Esse cenário provocou uma revitalização do mercado de capitais nacional dentro dos padrões mais exigentes e democratização do acesso ao mercado (IBRI, 2008).

As informações sobre os balanços e resultados financeiros das empresas listadas em bolsa podem ser encontradas na seção de Relação com Investidores (ou Relação com Investidor) disponível em cada sítio destas empresas na internet. A Lei 6.404/76 (Lei das Sociedades Anônimas) e a regulamentação da CVM, entre outras, determinam quais informações as empresas listadas em Bolsa devem divulgar ao mercado, aos acionistas e aos investidores (IBRI, 2008).

Modelagem matemática aplicada ao mercado financeiro

No dinâmico e moderno ambiente de economia global, existe uma abundância de dados nunca antes vista. Os tomadores de decisão com maior taxa de sucesso são aquele capaz de entender e processar a informação, usando-a de modo eficaz e direto. Em finanças, muitos consultores e operadores de mercado utilizam uma série de informações estatísticas para guiar suas recomendações de investimentos. No caso das ações, os consultores re-veem diversos dados financeiros incluindo relações entre preços e lucros e distribuição de dividendos. Comparando a informação para uma determinada ação com informações sobre a média do mercado de ações, um consultor financeiro pode tirar conclusões sobre se uma ação está ou não em um bom momento de compra (ANDERSON, SWEENEY, WILLIAMS, 2005).

Já em economia, um dos papéis desempenhados por economistas é fornecer previsões sobre a economia como um todo ou sobre algum aspecto ou segmento dela. Eles usam uma série de informações estatísticas ao fazer tais previsões, como a previsão de

taxas de inflação, que utiliza uma série de dados de indicadores como índice de preços ao produtor, a taxa de desemprego e a utilização da capacidade instalada. De modo muito frequente, esses indicadores são utilizados como variáveis preditivas de modelos de previsão computadorizados para calcular a taxa futura de inflação (ANDERSON, SWEENEY, WILLIAMS, 2005).

De modo mais explicativo, um método de aplicação de modelos de previsão na economia foi utilizado em (MELLO DE CARVALHO, CASTRO SOUZA, 2011). Neste artigo, foi estimada uma equação para as expectativas de inflação, seguido de uma estrutura a termo da taxa de juros. A expectativa utilizada é a da inflação doze meses à frente em função da política monetária, a meta de inflação para os próximos 12 meses, a medida de risco país, a própria expectativa de inflação do período anterior, a dinâmica da taxa de câmbio e da inflação corrente. A previsão de inflação é de apenas um mês à frente, mas já se pode entender a utilidade e aplicação de modelos estatísticos em economia e finanças como um aliado no momento de tomada de decisão de investimentos.

Assim, como sugerido por Gujarati e Porter (2011), a modelagem matemática em economia é utilizada para ir além das explicações qualitativas de teoria econômica e expressá-la na forma de equações matemáticas (modelos) de forma que possam ser testadas na prática, ou seja, verificar que aquela aproximação matemática é suficientemente próxima da realidade para o entendimento do fenômeno econômico que é estudado. A análise de um problema econômico pode ser realizada para prever os valores futuros de uma variável de previsão (seja um indicador como o produto interno bruto ou mesmo as projeções de lucro de uma empresa) com base em variáveis explanatórias, que sejam independentes da variável que se deseja prever.

Indicadores Econômicos

Uma definição para indicadores econômicos é que estes são grandezas de que expressam medidas da economia, expressas em valor numérico. Sua principal aplicação é a medição do nível de desenvolvimento de países, regiões e empresas, permitindo, também, efetuar comparações (Indicadores Económicos, 2014).

Existem diversos indicadores econômicos utilizados para mensurar o desempenho da economia brasileira, sendo que os que são escopo deste trabalho são a taxa de crescimento real do Produto Interno Bruto (PIB), a paridade do Real Brasileiro com o Dólar Americano (taxa de câmbio PTAX) e a inflação oficial do governo brasileiro (IPCA).

Por Produto Interno Bruto entende-se, segundo (Indicadores Económicos, 2014), que este é montante dos bens e serviços por ele produzidos em um determinado ano de referência. Corresponde a toda a produção realizada no país, seja por empresas nacionais ou estrangeiras. Segundo o Banco Central do Brasil, as taxas PTAX de compra e de venda correspondem, respectivamente, às médias aritméticas das taxas de compra e de venda das consultas realizadas diariamente (BANCO CENTRAL DO BRASIL, 2013). Já a taxa de inflação mede o acréscimo percentual médio nos preços dos bens e serviços produzidos pela economia. Neste trabalho, o índice de inflação utilizado é o Índice de Preços ao Consumidor Amplo, o IPCA.

Um exemplo de uso de indicadores econômicos na análise do comportamento ativos pode ser encontrado no trabalho de Gartner (Gartner, 2012), que utilizou as publicações sobre a taxa Selic para estudar o impacto desta no retorno das ações do sistema financeiro brasileiro, concluindo que esta informação pública não possui um efeito imediato no preço das ações do setor.

Notícias na internet

Em comunicação, define-se mídia como os canais ou ferramentas usadas para armazenamento e transmissão de informação ou dados. Mídia muitas vezes é usada como sinônimo de meios de comunicação de massa ou agências de notícias, mas pode se referir a um único meio utilizado para comunicar os dados para qualquer finalidade.

Segundo o relatório anual de jornalismo americano (Pews, 2010), a internet é a segunda fonte de notícias em popularidade, perdendo apenas para a tradicional televisão. Também, notícias de mídias online são o único meio que ainda cresce ano após ano – o rádio e as revistas ou jornais impressos possuem audiência declinante ano após ano. Esta relevância e notoriedade das notícias em meio digital fazem dela uma importante e influente fonte de dados para todos os mercados, incluindo o mercado de capitais e financeiros.

Teoria dos Mercados Eficientes

Essa teoria se baseia na hipótese de que toda informação disponível é rapidamente absorvida pelo mercado, por todos os seus participantes, sendo impossível que sejam auferidos ganhos significativamente acima da média do mercado. Esta característica é o conceito de mercado eficiente (FAMA, 1970). Ainda segundo (FAMA, 1991), existem três variantes para esta hipótese dos mercados eficientes. A primeira é a hipótese fraca, que considera que os mercados absorvem apenas as informações históricas que estão publicamente disponíveis. Uma hipótese média ou “semi forte” é baseada no princípio de que os preços dos ativos refletem estas informações públicas de forma instantânea. Fama ainda acrescenta a hipótese forte, dizendo que o mercado reflete instantaneamente até mesmo as informações consideradas privilegiadas.

METODOLOGIA

Seleções das empresas para análise

Foram selecionadas empresas que possuem ação listada em bolsa, dos segmentos de matérias prima, siderurgia e bancário, pois, somados, representam mais de 50% do índice. Assim, foram criadas variáveis baseadas nos relatórios de balanço destas empresas. Também foram criadas variáveis baseadas em indicadores macroeconômicos e nas notícias correlatas à economia e às empresas. Assim, foi desenvolvido um modelo matemático analisar a influência destas variáveis na variação percentual do lucro destas companhias.

Primeiramente, foi preciso escolher quais seriam as empresas que teriam sua variação de lucro modelada. Para isto foram estabelecidos os seguintes critérios:

1. Ação listada em bolsa que fazem parte da composição do índice Bovespa;
2. Três diferentes segmentos de atuação com grande expressão no índice: matérias prima, siderurgia e bancário. Estes segmentos somados representam mais de 50% do índice, segundo a composição descrita no site da BM&FBovespa (BMFBOVESPA – Índice Bovespa);
3. Caso o segmento apresentasse uma empresa de capital misto, mas controlada majoritariamente pelo governo federal, seria escolhida uma empresa com esta característica e, a outra, privada.

O primeiro critério se justifica, pois, como descrito anteriormente, o índice Bovespa representa uma média das ações com maior liquidez em bolsa e, conseqüentemente, empresas que possuem mais informações financeiras e também de notícias veiculadas nos meios de informação. Já o segundo foi utilizado para verificar e comparar possíveis diferenças e características entre diferentes seguimentos, ou seja, identificar se as variáveis que influenciam uma empresa do setor siderúrgico podem ser diferentes do setor bancário, por exemplo. O terceiro critério segue o mesmo raciocínio, mas no caso, permitirá uma comparação e percepção se há características diferentes entre empresas controladas pelo governo ou pela iniciativa privada.

Assim, seguindo os critérios estabelecidos, foram selecionadas as seguintes empresas:

- Matérias primas: CIA Vale do Rio Doce e Petrobrás;
- Siderurgia: Usiminas e Gerdau
- Bancos: Itaú Unibanco e Banco do Brasil

Levantamentos dos balanços e criação das variáveis relacionadas

Após a seleção das empresas, foi realizado um levantamento dos balanços trimestrais referentes ao período de março de 2010 a agosto de 2013 (totalizando 12 balanços por empresa). Este período foi escolhido para poder haver massa crítica de registros para o desenvolvimento do modelo, bem como histórico para a captação das notícias nas mídias eletrônicas. Os balanços foram analisados de acordo com os demonstrativos financeiros, para a construção das seguintes variáveis:

- Índice de Liquidez Corrente:

$$ILC = \frac{\textit{Ativo Circulante}}{\textit{Passivo Circulante}}$$

- Índice de Liquidez Seca:

$$ILS = \frac{\textit{Ativo Circulante} - \textit{Estoques}}{\textit{Passivo Circulante}}$$

- Margem de Lucro Bruto:

$$MLB = \frac{\textit{Lucro Bruto}}{\textit{Receita de Vendas}}$$

- Margem de Lucro Operacional:

$$MLO = \frac{\text{Resultado Operacional} + \text{Juros}}{\text{Receita de Vendas}}$$

- Margem de Lucro Líquido:

$$MLL = \frac{\text{Lucro Líquido Disponível aos Acionistas}}{\text{Receita de Vendas}}$$

- Lucro por Ação:

$$LA = \frac{\text{Lucro Líquido Disponível aos Acionistas}}{\text{Número de Ações}}$$

Todas as informações financeiras de resultados das empresas foram obtidas nos sites institucionais de relações com investidores.

A variável dependente, ou seja, a variação do lucro da empresa (VLE) também depende das informações encontradas nos demonstrativos financeiros e foi calculada segundo a equação:

$$VLE = \frac{\text{Lucro do trimestre atual}}{\text{Lucro do trimestre anterior}} - 1$$

O VLE é expresso em termos percentuais, e pode ser tanto positivo quando negativo, dependendo se a empresa auferiu lucro ou não, respectivamente.

Captações de notícias na internet

Definição da fonte de dados

Primeiramente, foi selecionada a versão online do jornal Folha de São Paulo (<http://www.folha.uol.com.br/>) como fonte primária

de dados de notícias. A escolha foi baseada no fato de ser um jornal com boa credibilidade, de âmbito nacional, atualização frequente e também pela facilidade de acesso ao conteúdo público via programação e automatização da coleta de dados. O endereço utilizado para captação de dados foi o

[http://search.folha.com.br/search?q = ' + var + '&site = online](http://search.folha.com.br/search?q=' + var + '&site=online)

onde, no lugar da palavra var, é inserido o termo que se deseja realizar a busca pelas notícias. Além dos motivos elencados acima, este jornal foi selecionado devido a sua versatilidade e disponibilidade de consulta de dados históricos.

Definição dos termos de busca

Após a definição da fonte, foi necessário estabelecer quais os termos que poderiam ser úteis e relevantes para a criação de variáveis relacionadas às notícias publicadas. Os termos diretos, num total de 16, são mostrados no **Object 1** a seguir.

Object 1 - Termos Diretos	
Petrobrás	ITUB4
PETR3	Banco do Brasil
PETR4	BBAS3
Vale do Rio Doce	Gerdau
VALE5	GGBR4
VALE3	GGBR3
Itaú	Usiminas
Itaú Unibanco	USIM5

Na sequência, foram selecionados os termos que estiveram relacionados de maneira indireta com as empresas, ou seja, relacionados ao seu segmento de atuação, à situação macroeconômica do país, ao mercado de capitais, nomes de autoridades e aos

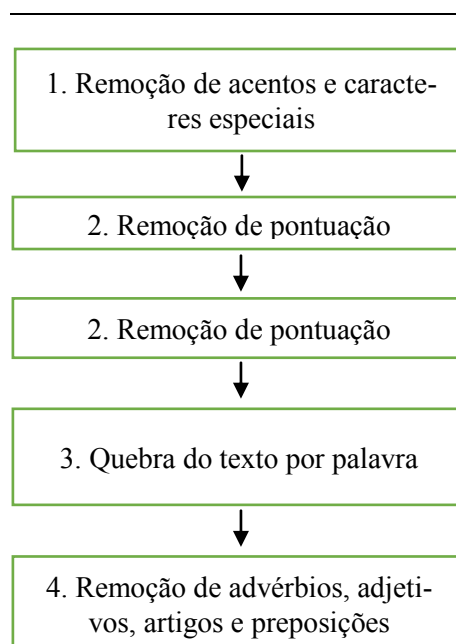
próprios indicadores econômicos, sejam termos língua portuguesa ou inglesa. Estes termos são apresentados no Object 2

Object 2 - Termos indiretos		
gasolina	minério de ferro	petróleo
combustíveis	inflação	dólar
cdi	selic	Dilma
commodities	curva de juros	NASDAQ
gasolina	minério de ferro	petróleo
combustíveis	inflação	dólar
cdi	Tombini	Mantega

Construção das variáveis não estruturadas

As variáveis não estruturadas são aquelas que não estão na forma de uma estrutura definida, tal qual uma tabela ou informações contidas em um formulário. Os títulos das notícias capturadas não possuem uma estrutura definida, por isso são classificados como informações não estruturadas. Por serem informações unicamente textuais, a criação de variáveis com estes dados exige, primariamente, o tratamento descrito no Object 3 a seguir.

Object 3 – Fluxo de Tratamento de dados das informações textuais



O desenvolvimento das variáveis foi realizado após estas etapas de tratamento de dados e foi baseada na construção de trigramas. Trigramas são a criação de uma estrutura que contém 3 palavras na sequência (Ventura,2008).

Depois de identificados os trigramas, os mesmos são relacionados ao tipo de evento ocorrido na variação de lucro. Cada trigrama é classificado de acordo com a ocorrência de uma variação positiva ou negativa no lucro. Como exemplo, tem-se o trigrama “bovesp_fech_qued” que é a combinação das palavras “Bovespa fecha em queda”. Sua ocorrência total quando uma busca é realizada envolvendo a Petrobras é de 328 (ou seja, em 328 notícias encontrou-se esta expressão), sendo que em 5 trimestres, ele é encontrado associado à uma viração positiva no balanço (identificada neste trabalho com o termo “BOM”) e em 8 trimestres foi associado à uma variação negativa (identificada neste trabalho com o termo “MAU”). No Object 4 abaixo temos um exemplo da construção de cinco trigramas para a Petrobras.

Object 4 – Exemplo de construção de trigramas para a Petrobras

TRIGRAMA	QTDE	BOM	MAU	TX BOM	TX MAU	TXB_TXM
elev_previsa_inflaca	99	2	8	4,85213E-05	8,93485E-05	0,54
dol_cai_r	97	4	6	9,70426E-05	6,70114E-05	1,45
negoci_dol_ating	96	3	3	7,2782E-05	3,35057E-05	2,17
bovesp_acentu_perd	95	4	4	9,70426E-05	4,46743E-05	2,17
abert_dol_val	94	2	3	4,85213E-05	3,35057E-05	1,45

Na sequência ilustrada no quadro acima, temos os trigramas relacionados as expressões:

- “elevação da previsão de inflação”
- “dólar cai a R\$”
- “negociação do dólar atinge”
- “Bovespa acentua perdas”
- “na abertura dólar vale”

Para completar a análise e construir as variáveis, cada trigrama é classificado de acordo com sua “taxa de BOM” (coluna TX_BOM), ou seja, pega-se a quantidade de ocorrências daquele trigrama e divide-se pelo total de ocorrências de trigramas que aparecerem relacionados à uma variação positiva no balanço (BOM). O mesmo raciocínio é utilizado para construir a “taxa de MAU” (coluna TX_MAU). Em seguida, é calculada a relação entre TX_BOM e TX_MAU, resultando em TXB_TXM. Essa última relação é utilizada para agrupar os trigramas com mesmas taxas. Cada grupo ou conjunto de trigramas com igual taxa de bom em relação à taxa de mau foi considerada uma variável. Esta variável assume o valor 1 caso se verifique a ocorrência no trimestre de algum destes trigramas, e assume o valor 0 caso não haja ocorrência. Cada empresa teve o seu conjunto de trigramas associado ao respectivo trimestre de publicação do balanço finan-

ceiro. Como existem os termos diretos para cada empresa, os trigramas foram diferentes para cada uma, bem como sua relação com períodos de alta ou baixa na variação de lucro.

Construções das variáveis baseadas nos indicadores macroeconômicos

As notícias e informações macroeconômicas, são referentes ao período de março de 2010 à agosto de 2013. Os indicadores captados foram: inflação oficial do governo brasileiro (IPCA), a taxa de crescimento real do Produto Interno Bruto (PIB) e da paridade do Real Brasileiro com o Dólar Americano (taxa de câmbio PTAX), todos em valores e taxas mensais. Os dados foram agrupados em intervalos de três meses para manter a mesma janela de tempo da publicação dos demonstrativos financeiros trimestrais.

ANÁLISE DOS RESULTADOS

Resultados para PETROBRAS

Após sequência descrita na metodologia, foram criados 43.876 trigramas relacionados aos termos diretos e indiretos da Petrobras. Em seguida, os 43.876 trigramas foram agrupados em relação a sua TXB_TXM, resultado em um total de 15 variáveis. Assim, somadas as demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 31 variáveis para o modelo da Petrobras. Na sequência, temos os 5 trigramas mais frequentes, relacionados as expressões:

- `pont_dol_cai` – “pontos dólar cai”
- `dia_agend_frac` – “dia de agenda fraca”
- `exteri_dol_vai` – “no exterior dólar vai a “

Como resultado, após o desenvolvimento de uma regressão para cada uma das empresas selecionadas, observa-se que as variáveis baseadas nas notícias de mídias eletrônicas (as variáveis não estruturadas) estão presentes e são relevantes para todos os modelos, ampliando o poder de explicação da variável dependente. Cada empresa apresentou um conjunto distinto de variáveis, o que evidenciou a influência das particularidades de cada segmento.

- descol_exteri_sob – descola do exterior e sobe”
- cai_quas_4 – “cai quase 4%”

Desta forma, após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com 5 variáveis significativas:

- MLL – Margem de Lucro Líquido
- LA – Lucro por ação
- ILS – Índice de Liquidez Seca
- PIB – Produto Interno Bruto
- RT1 – Conjunto de trigramas relacionados às notícias publicadas

No Object 5 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 5 – Coeficientes das variáveis finais do modelo de regressão para Petrobras

Attribute	Coef.
Intercept	0,712989
MLL	-29,965785
LA	-1,174705
ILS	2,135123
PIB	15,408123
RT1	1,207480

Este modelo possui um R^2 , que é o coeficiente de determinação da regressão, ou seja, o percentual da variância total da variável dependente que é explicado pela equação de regressão, de 0,882. Outra observação interessante é que o modelo é muito influenciado pela variação do PIB. Isto provavelmente se deve ao fato de a Petrobras ser uma empresa com relevância nacional e, além dis-

so, possuir seus produtos (petróleo, gás natural, gasolina, dentre outros) fortemente relacionados a diversos setores da economia brasileira. Apesar de parecer óbvio, o modelo serve para entender qual é o peso ou influência desta variável para o resultado da Petrobras.

Analisando-se a variável não estruturada RT1, tem-se que esta é composta de uma combinação de 9.655 trigramas. Dado a quantidade expressiva, destaca-se um exemplo, como “descol_exteri_sob” – “descola do exterior e sobe”, ou seja, houve um movimento da bolsa diferente das bolsas mundiais. Outra observação é que 447 trigramas, 4,63% do total desta regra, estão relacionados ao dólar americano; 150 trigramas, ou 1,55% do total desta regra são relacionados ao petróleo; 141 trigramas, ou 1,46% do total desta regra estão relacionados às notícias de Dilma Rousseff (Presidente da República) e 67 trigramas, 0,69% do total, às notícias do PIB brasileiro. Para avaliar a importância da variável não estruturada RT1, uma nova regressão foi feita retirando-a do modelo original. Como resultado, tem-se um coeficiente de determinação da regressão R^2 , de 0,753. Como este é inferior ao primeiro, de 0,882, conclui-se que a variável RT1 é responsável por um incremento de 17,1% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro da Petrobras.

Resultados para Vale do Rio Doce - VALE

Foram criados 28.836 trigramas relacionados aos termos diretos e indiretos da Vale. Em seguida, os 28.836 trigramas foram agrupados em relação a sua TXB_TXM, resultado em um total de 22 variáveis. Assim, somadas as demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 38 variáveis para o modelo da Vale. Na sequência, temos os 5 trigramas mais frequentes, relacionados as expressões:

- elev_projeca_inflaca – “eleva projeção de inflação”
- intervenca_bc_dol – “intervenção do BC dólar
- merc_intervenca_bc – “mercado com intervenção do BC”
- bovesp_mant_qued – “Bovespa mantém queda”
- merc_bols_pi – “mercado bolsa piora”

Após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com 3 variáveis significativas:

- MLO – Margem de Lucro Operacional
- IPCA_TEND – Tendência da variação do indicador de inflação
- RT21 – Conjunto de trigramas relacionados às notícias publicadas

No Object 6 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 6 – Coeficientes das variáveis finais do modelo de regressão para Vale

Attribute	Coef.
Intercept	0,939764
MLO	3,482470
IPCA_TEND	-1,104967
RT21	-1,989940

Este modelo possui um R^2 de 0,681, ou seja, 68,1% das observações de variações de lucro da Vale obtidas entre o primeiro trimestre de 2010 e o segundo trimestre de 2013 podem ser explicadas pelas variáveis acima descritas. Analisando-se a variável

não estruturada RT21, tem-se que esta é composta de uma combinação de 6 trigramas, descritos abaixo

- 'mund_vend_varej' – “mundo vendas no varejo”
- 'fech_alt_londr' – “fecha em alta em Londres”
- 'bols_eua_avanc' – “bolsa dos EUA avançam”
- 'indic_prec_consum' – “índice de preços ao consumidor”
- 'vend_varej_eua' – “venda no varejo dos EUA”
- 'mai_alt_desd' – “maior alta desde”

É possível observar que os trigramas estão relacionados ao avanço das bolsa internacionais e também ao consumo no varejo. Esta última relação é interessante dado que uma das variáveis relacionadas à inflação (IPCA_TEND) é significativa no modelo de regressão da Vale. Como a Vale é fornecedora de uma matéria prima que a base de produtos de consumo do mundo inteiro, o minério de ferro, a aparição de um elemento referente ao índice de preços e atividade no varejo mostra como a variável de mídia pode ajudar a explicar a variação de lucro.

Ainda na avaliação da importância da variável não estruturada RT21, uma nova regressão foi feita retirando-a do modelo original, preservando-se as demais. Como resultado, tem-se um coeficiente de determinação da regressão R^2 , de 0,433. Como este é inferior ao primeiro, de 0,681, conclui-se que a variável RT21 é responsável por um incremento de 57,3% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro da Vale.

Resultados para USIMINAS

Para analisar a variação de lucro da USIMINAS foram criados 30.219 trigramas relacionados aos seus termos diretos e indiretos. Após sua criação, os 30.219 trigramas foram agrupados em relação a sua TXB_TXM, resultado em um total de 15 variáveis não estruturadas para a Usiminas. Assim, somadas as demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 31 variáveis para o modelo da Usiminas. Na sequência, temos os 5 trigramas mais frequentes, relacionados as expressões:

- zon_eur_derrub – “zona do euro derruba”
- cresc_acim_med – “crescimento acima da média”
- ny_fech_qued – “NY fecha em queda”
- acim_r_2 – “acima de R\$2”
- 62_mil_pont – “62 mil pontos”

Após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com 3 variáveis significativas:

- MLB – Margem de Lucro Bruta
- PIB_TEND – Tendência da variação do indicador de PIB
- RT7 – Conjunto de trigramas relacionados às notícias publicadas

No Object 7 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 7 – Coeficientes das variáveis finais do modelo de regressão para Usiminas

Attribute	Coef.
Intercept	-2,208921
MLB	-15,777691
PIB_TEND	1,439903
RT7	2,920594

Este modelo possui um R^2 considerado elevado, de 0,891. Analisando-se a variável não estruturada RT7, tem-se que esta é composta de uma combinação de apenas um trigrama: fech_lev_alt – “fecha em leve alta”. Apesar de ser influenciada diretamente por apenas um elemento, quando avalia-se a importância da variável não estruturada RT7, realizando uma nova regressão retirando-a do modelo original, atinge-se como resultado um coeficiente de determinação da regressão R^2 , de 0,241. Assim, conclui-se que a variável RT7 é responsável por um incremento de 269,7% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro da Usiminas. Adicionalmente, caso a variável RT7 seja retirada, nenhuma variável fica significativa para o modelo.

Resultados para GERDAU

Para a Gerdau foram criados 8.801 trigramas relacionados aos seus termos diretos e indiretos. Após sua criação, os 8.801 trigramas foram agrupados em relação a sua TXB_TXM, resultado em um total de 17 variáveis não estruturadas para a análise da Gerdau. Somando-se às demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 33 variáveis para o modelo da Gerdau. Na sequência, temos os 5 trigramas mais frequentes, relacionados as expressões:

- bols_eua_atiing – “bolsa dos EUA atingem”

- bovesp_recu_035 – “Bovespa recua 0,35%”
- cresc_men_inflaca – “crescimento mensal da inflação”
- acim_70_mil – “acima dos 70 mil”
- aco_banc_pux – “ações de bancos puxam”

Após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com apenas 2 variáveis significativas:

- PIB_TEND – Tendência da variação do indicador de PIB
- RT4 – Conjunto de trigramas relacionados às notícias publicadas

No Object 8 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 8 – Coeficientes para variáveis finais do modelo de regressão para Gerdau

Attribute	Coef.
Intercept	-0,785405
PIB_TEND	0,528019
RT4	0,861763

É interessante notar que a tendência da variação do PIB é relevante, dado que o segmento de atuação da Gerdau, siderurgia, é essencial em diversos setores da economia e, caso esta esteja se desenvolvendo, é extremamente positivo para os negócios da empresa.

Este modelo não se mostrou bem ajustado como os demais, mesmo testando-se uma série de outras variáveis e suas combinações. O resultado final foi um R^2 considerado baixo, de 0,409. Entretanto, a influência da variável de mídia da web não deixa de ser relevante. Analisando-se a variável não estruturada RT4,

tem-se que, assim como no outro estudo de caso de siderurgia, esta é composta de uma combinação de apenas um trigrama: `ating_mai_nivel` – “atinge maior nível”. Apesar de ser influenciada diretamente por apenas um elemento, quando avalia-se a importância da variável não estruturada RT4, realizando uma nova regressão retirando-a do modelo original, atinge-se como resultado um coeficiente de determinação da regressão R^2 , de 0,241. Assim, conclui-se que a variável RT4 é responsável por um incremento de 269,7% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro da Gerdau. Adicionalmente, assim como no caso da Usiminas, caso a variável RT4 seja retirada, nenhuma variável fica significativa para o modelo.

Resultados para ITAÚ UNIBANCO

Nas próximas duas seções serão analisadas as regressões para o segmento bancário. Primeiramente, tem-se o caso do banco Itaú – Unibanco. Para analisar a variação de lucro, foram criados 8.021 trigramas relacionados aos seus termos diretos e indiretos. Após sua criação, os 8.021 trigramas foram agrupados em relação a sua `TXB_TXM`, resultado em um total de 19 variáveis não estruturadas para a análise do caso Itaú Unibanco. Assim, somadas as demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 34 variáveis. Na sequência, temos os 5 trigramas mais frequentes, relacionados às expressões:

- `bovesp_fech_lev` – “Bovespa fecha em leve”
- `alt_fech_r` – “em alta e fecha em R\$”
- `alt_otim_sobr` – “alta e otimismo sobre”
- `bbc_brasil_dilm` – “BBC Brasil Dilma”
- `bols_asia_avanc` – “bolsas da Ásia avançam”

É válido observar que, como o segmento de banco não possui estoque de mercadorias, não foi incluída na análise a variável de balanço do índice de liquidez seca, pois como o estoque é zero, o valor do ILS é o mesmo do índice de liquidez corrente. Após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com 4 variáveis significativas:

- MLB – Margem de Lucro Bruta
- LA – Lucro por ação
- RT2 – Conjunto de trigramas relacionados às notícias publicadas
- RT4 – Conjunto de trigramas relacionados às notícias publicadas

No Object 9 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 9 – Coeficientes das variáveis finais do modelo de regressão para Itaú Unibanco

Attribute	Coef.
Intercept	-1,393270
MLB	0,842001
LA	0,971851
RT2	0,096696
RT4	0,146402

É interessante notar que não houve influência significativa de nenhuma das variáveis relacionadas aos indicadores macroeconômicos. Entretanto, a influência de alguns indicadores macroeconômicos pode ser notada nas variáveis não estruturadas RT2 e RT4. Em RT2, tem-se a combinação de 4 trigramas:

- bols_asia_ca – “bolsas da Ásia came”

- bovesp_oper_alt – “Bovespa opera em alta”
- camb_diz_manteg – “câmbio, diz Mantega”
- derrub_bols_eua – “derrubam bolsas dos EUA”

Já em RT4 tem-se a combinação de 3 trigramas:

- bat_nov_record – “bate novo recorde”
- bols_nov_york – “bolsa de Nova York”
- dad_sobr_empreg – “dados sobre emprego”

Este modelo possui um R^2 considerado elevado, de 0,815. Analisando-se as duas variáveis não estruturadas RT2 e RT4 e realizando uma nova regressão retirando-as do modelo original, atinge-se como resultado um coeficiente de determinação da regressão R^2 , de 0,568. Assim, conclui-se que as duas variáveis são responsáveis por um incremento de 43,5% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro do Itaú Unibanco. Caso as variáveis sejam retiradas do modelo, MLB e LA ainda seriam significativas, mas, como visto, não teriam o mesmo poder explicativo.

Resultados para BANCO DO BRASIL

Seguido com os resultados do segmento bancário, tem-se o caso do Banco do Brasil. Para analisar a variação de lucro, foram criados 10.948 trigramas relacionados aos seus termos diretos e indiretos. Após sua criação, os 10.948 trigramas foram agrupados em relação a sua TXB_TXM, resultado em um total de 16 variáveis não estruturadas para a análise do caso do Banco do Brasil. Assim, somadas as demais variáveis descritas na metodologia deste trabalho, chega-se a um total de 31 variáveis. Na se-

quência, temos os 5 trigramas mais frequentes, relacionados as expressões:

- fech_lev_qued – “fecha em leve queda”
- eua_dol_cai – “EUA dólar cai”
- abr_alt_apo – “abre em alta após”
- oscil_pert_estabil – “oscila perto da estabilidade”
- lev_qued_dia – “leve queda no dia”

Após a análise de correlação entre as variáveis e aplicação do modelo de regressão linear múltipla, tem-se um modelo final com 6 variáveis significativas:

- PIB_TEND – Tendência da variação do indicador de PIB
- IPCA – variação trimestral da medida de inflação oficial do governo brasileiro
- RT1 – Conjunto de trigramas relacionados às notícias publicadas
- RT2 – Conjunto de trigramas relacionados às notícias publicadas
- RT9 – Conjunto de trigramas relacionados às notícias publicadas
- RT15 – Conjunto de trigramas relacionados às notícias publicadas

No Object 10 abaixo temos os respectivos coeficientes da regressão múltipla, calculados utilizando o software Tanagra.

Object 10 – Coeficientes das variáveis finais do modelo de regressão para Banco do Brasil

Attribute	Coef.
Intercept	0,075944
PIB_TEND	0,144403
IPCA	-13,702848
RT1	0,341919
RT2	-0,129832
RT9	0,305248
RT15	-0,295004

Houve uma forte influência da variável IPCA. Esta influência faz sentido pois, de acordo com informações do Banco Central do Brasil, o Banco do Brasil é o maior banco do Brasil em relação ao volume de operações de concessão de crédito, líquidas de provisões. Desta forma, variações no IPCA tendem a ter influência na taxa de juros, o que impacta significativamente o resultado financeiro de uma instituição bancária.

Analisando as variáveis não estruturadas tem-se que RT1 é resultado da combinação de 4.831 trigramas. Dado a quantidade expressiva, destaca-se um exemplo, como “decisa_sobr_jur” – “decisão sobre os juros”, que é um trigrama em linha com o IPCA. Já a variável RT2 é a combinação do trigrama “bovesp_abr_alt” – “Bovespa abre em alta”. Já a variável RT9 é a combinação de 73 trigramas como “americ_fech_alt” – americanas fecham em alta” – e “bovesp_lev_alt” – “Bovespa fecha em leve alta”. Por fim, a variável RT15 é resultado da combinação de 1.437 trigramas

Este modelo possui um R^2 considerado elevado, de 0,973. Foi o melhor modelo, em termos de poder explicativo desenvolvido neste trabalho. Este resultado se deve ao fato de que as variáveis escolhidas neste trabalho possuem forte influência explanatória

no estudo de caso do Banco do Brasil, diferentemente o ocorrido com o estudo de caso da Gerdau.

Analisando-se as variáveis não estruturadas e realizando uma nova regressão retirando-as do modelo original, atinge-se como resultado um coeficiente de determinação da regressão R^2 , de 0,252 e nenhuma das variáveis restantes seriam significativas no modelo. Assim, conclui-se que as duas variáveis são responsáveis por um incremento de 286,1% na explicação da equação de regressão, mostrando a relevância das informações de notícias na predição da variação de lucro do Banco do Brasil.

CONSIDERAÇÕES FINAIS

As variáveis construídas a partir dos dados capturados de notícias de mídias eletrônica possuem relevância na explicação da variação de lucro das empresas listadas em bolsa. Nos seis modelos de regressão linear múltipla desenvolvidos neste trabalho, estas variáveis não apenas fizeram parte da composição de todos os modelos, como também a ausência delas torna inexplicável o fenômeno em algumas empresas, como foi o caso do Banco do Brasil, Gerdau e Usiminas, quando consideradas as demais variáveis propostas neste trabalho.

Em relação às variáveis oriundas dos dados dos indicadores macroeconômicos (BANCO CENTRAL DO BRASIL, 2013), elas não compuseram apenas o modelo do banco Itaú. Apesar de parecer uma conclusão óbvia em termos econômicos, este trabalho quantifica e mostra, matematicamente, o quão influente é este comportamento da economia, traduzido na forma de indicadores, para as empresas. O indicador mais frequente encontrado neste trabalho foi o PIB, com suas variáveis de variação trimestral do PIB e sua tendência, ou seja, se a variação foi maior ou menor que a variação anterior.

As conclusões atingidas estabelecem o ponto inicial para a criação de uma ferramenta que permita automatizar a análise de investimento, principalmente ao investidor pessoa física e pequenas empresas, conseguindo absorver as notícias e efeitos macroeconômicos nos resultados das empresas.

Outro indicador macroeconômico relevante foi o IPCA, presente em dois modelos: Banco do Brasil e Vale. Como mostrado nos resultados, o IPCA tem forte influência no modelo do Banco do Brasil, dado que este é o maior banco do Brasil em relação ao volume de operações de concessão de crédito, líquidas de provisões. Desta forma, variações no IPCA tendem a ter influência na taxa de juros, o que impacta significativamente o resultado financeiro de uma instituição bancária. Já no caso da Vale, sua influência pode ser explicada pois um reajuste de preços no minério de ferro, o produto de maior influência no resultado da empresa, tem forte impacto no resultado financeiro.

Já o indicador PTAX, que mede a paridade do Real Brasileiro com o Dólar Americano (taxa de câmbio PTAX), não foi influente em nenhum dos modelos. Isto não significa que a paridade entre as moedas não influencia o resultado das empresas, mas a sua não influencia neste trabalho pode indicar que, durante o período de tempo analisado, não tenha havido tal influência nos resultados das empresas analisadas.

É interessante observar que o modelo da Petrobras é o que mais possui influência das variáveis de balanço, sendo composto pela MLL (margem de lucro líquido), LA (lucro por ação) e o ILS (índice de liquidez seca). De maneira oposta, o modelo do Banco do Brasil e o modelo da Gerdau são os únicos que não possuem nenhuma variável oriunda das informações de demonstrações financeiras, sendo compostos apenas das variáveis de mídia e das variáveis dos indicadores macroeconômicos.

Na comparação entre empresas do mesmo segmento temos que apenas as empresas de siderurgia possuem em suas regressões alguma variável coincidente, que no caso é a variação de tendência do PIB, a PIB_TEND. Pode-se então concluir que, mesmo que as empresas pertençam a um mesmo setor da economia, é claro que podem existir elementos coincidentes, mas isto não é algo

mandatório e as empresas são organismos independentes que podem sofrer influências de diferentes variáveis e eventos, como é o caso observado em todas as empresas analisadas neste trabalho. A mesma inferência pode ser utilizada quando analisamos as duas empresas de capital misto, Petrobras e Bancos do Brasil. Conter o governo brasileiro como acionista majoritário não implica em alguma semelhança – relacionada a esta particularidade – nos fatores que são determinantes na variação do lucro. O mesmo pode ser dito quando se observa as demais empresas, que não possuem esta característica de possuir o governo como parte significativa na composição de seus quadros societários. No Object 11 a seguir é mostrada a composição dos modelos por empresa, de acordo com os grupos de variáveis: variáveis não estruturadas, macroeconômicas (indicadores) e demonstrações financeiras:

Object 11 – Composição dos modelos por empresa, de acordo com os grupos de variáveis.

Variável/Modelo	Petrobras	Vale	Gerdau	Usiminas	Itaú Unibanco	Banco do Brasil
Não estruturadas	Sim	Sim	Sim	Sim	Sim	Sim
Macroeconômica	Sim	Sim	Sim	Sim	-	Sim
PIB	Sim	-	Sim	Sim	-	Sim
IPCA	-	Sim	-	-	-	Sim
PTAX	-	-	-	-	-	-
Demonstrações Financeiras	Sim	Sim	Sim	Sim	Sim	Sim

No Object 12 tem-se um resumo dos resultados das regressões, com e sem as variáveis não estruturadas, bem como a informação da quantidade total das variáveis.

Object 12 – Resumo dos resultados das regressões

	Petrobras	Vale	Gerdau	Usiminas	Itaú Unibanco	Banco do Brasil
R^2 com não estruturadas	0,882	0,681	0,409	0,891	0,815	0,973
R^2 sem não estruturadas	0,753	0,433	0,112	0,241	0,568	0,252
Total de variáveis	5	3	2	3	4	6
Total de variáveis não estruturadas	1	1	1	1	2	4

Os resultados obtidos foram satisfatórios para entender como, se e qual a magnitude da influência das informações de mídias eletrônicas na predição da variação de lucro de empresas listadas em bolsa. Provou-se que não apenas há influência, mas também que em alguns casos é elemento fundamental na explicação do fenômeno.

Assim, abre-se precedente para a criação de uma ferramenta que permita automatizar esta análise de resultados e permitir, principalmente ao investidor pessoa física e pequenas empresas, uma ferramenta confiável e útil para auxiliar na tomada de decisão no momento de realizar um investimento.

REFERÊNCIAS BIBLIOGRÁFICAS

MALUF, Larissa Ayres (2010). O que Influencia a Mente do Investidor ao Tomar suas Decisões de Investimento. Revista de Finanças Aplicadas. Publicado em 17mai10, pp. 1-17.

BMFBOVESPA – Estudos para o desenvolvimento do mercado de capitais. Disponível em:

<http://www.bmfbovespa.com.br/pdf/relatorio2.pdf> - Acesso em janeiro de 2014.

VENTURA, João Miguel Jones. Extração de Unigramas Relevantes. Lisboa: Universidade de Nova Lisboa, 2008.

FORTUNA, Eduardo. Mercado Financeiro – Produtos e Serviços. São Paulo: Qualitymark, 2011.

ROSS, Stephen A., WESTERFIELD, Randolph W., JORDAN, Bradford D. Administração Financeira. São Paulo: Atlas, 2011.

BMFBOVESPA – Índice Bovespa – Ibovespa. Disponível em:

<http://www.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=IBOVESPA&idioma=pt-br>. Acesso em setembro de 2013.

MARION, José Carlos. Contabilidade Básica. São Paulo: Atlas, 2004.

Lei Nº 6.404/76 – das Sociedades por Ações. Disponível em:

http://www.planejamento.gov.br/secretarias/upload/Arquivos/dest/eventos/sociedade_por_acoes.pdf. Acesso em janeiro de 2014.

ANDERSON, David R., SWEENEY Dennis J., WILLIAMS Thomas A. Estatística Aplicada à Administração e Economia. São Paulo: Thomson Learning, 2005.

MELLO DE CARVALHO, Leonardo e CASTRO SOUZA, José Ronaldo de. Modelo para Previsão de Inflação. Diretoria de estudos Macroeconômicos do IPEA.

GUJARATI, Damodar N., PORTER Dawn C. Econometria Básica 5ªed. Porto Alegre: AMGH Editora, 2011.

GARTNER, Andrea. Análise da reação das ações do setor financeiro brasileiro as divulgações da Taxa Selic ocorridos entre 2004 e 2011. Revista Finanças Aplicadas. Publicado em 06/02/2012 pp. 12-12

INSTITUTO BRASILEIRO DE RELAÇÃO COM INVESTIDOR – IBRI (2008), Guia de relações com Investidores. Disponível em <http://www.ibri.com.br>. Acesso em janeiro de 2014.

Indicadores Económicos. In Infopédia. Porto: Porto Editora, 2003-2014. [Consult. 2014-02-02]. Disponível em [http://www.infopedia.pt/\\$indicadores-economicos](http://www.infopedia.pt/$indicadores-economicos).

BANCO CENTRAL DO BRASIL – Indicadores Econômicos consolidados. Homepage institucional. Disponível em: <http://www.bcb.gov.br/?INDECO>. Acesso em diversas datas entre junho e julho de 2013.

PEW RESEARCH CENTER'S – Pew's annual State of the Media. Homepage institucional. Disponível em: <http://stateofthemedias.org/>. Acesso em agosto de 2013.

FAMA, E. Efficient Capital Markets: a review of theory and empirical work, *Journal of Finance* 25, p. 383-417, 1969. Publicado em maio de 1970.

FAMA, E.F. Efficient capital markets: II. *The Journal of Finance*. Cambridge, v. 46, n. 5, p.1575-1618, 1991.