

O USO DE DATA MINING PARA A INTELIGÊNCIA DE ESTADO

– Um Estudo de caso

Glauber Nóbrega da Silva*

Resumo

O presente artigo discute como ferramentas de mineração de dados podem ser utilizadas para identificar padrões úteis à Inteligência de Estado. Para esse fim identifica e analisa padrões do Produto Interno Bruto - PIB de todos os países nos últimos quarenta anos de forma a demonstrar como é possível criar informações úteis à inteligência de Estado. Também demonstra como essas informações podem ser utilizadas para antever eventos com grande probabilidade de acerto.

I Introdução

A história mostra que o homem sempre recorreu a previsões para antever e planejar o futuro. Essa busca incessante possuiu e ainda possui apenas um objetivo: auxiliar em suas decisões. Prever e decidir são ações tão vitais e necessárias a nossa sobrevivência que todas as sociedades buscaram formas de subsidiá-las. Vários são os exemplos de grandes líderes que recorreram à magia, aos astros, oráculos e outras formas de previsão de futuro isentas de qualquer método científico. O oráculo de Delfos talvez seja uma das fontes de previsões mais difundidas na literatura mundial. No entanto, o verdadeiro subsídio ao processo decisório está no uso de uma metodologia que possibilite a obtenção, processamento e disseminação de informações em momento oportuno, o que conhecemos como Inteligência.

A Inteligência é considerada uma das profissões mais antigas do mundo e seus profissionais buscavam, acima de tudo, gerar informações adequadas a governantes ou chefes militares. A inteligência foi utilizada por Alexandre, o Grande, durante suas campanhas na Ásia, por Moisés durante o êxodo bíblico, por Napoleão durante suas campanhas da Europa e tantos outros até os dias de hoje (SANTOS et al, 2011).

2 Processos e métodos da Inteligência

A Inteligência é considerada como um misto de ciência e arte (Platt, 1974). Isso ocorre porque existem várias formas de coletar e tratar informações. Além disso, a informação pode estar contida em documentos impressos, mídias televisivas etc., a partir de origem tecnológica ou

* Tecnólogo em Meio Ambiente

humana. Cabe ao profissional de informações então identificar quais as suas melhores fontes e as formas mais adequadas de obtê-las e tratá-las de modo que estas sejam convertidas em algo útil ao tomador de decisão. Nesse contexto, é necessário que o profissional conheça as fases básicas do ciclo de produção do conhecimento, quais sejam: planejamento, coleta, processamento e difusão.

Quanto vale uma informação?

O planejamento é executado de acordo com os objetivos ou a missão dada ao profissional ou a equipe de Inteligência. É uma fase fundamental pois um erro ou omissão pode inviabilizar o produto a ser entregue. Durante o planejamento são definidas quais as informações necessárias, sua forma de obtenção, custos, prazos e resultados esperados. Ao planejamento segue-se a coleta ou atividade de pesquisa propriamente dita. Trata-se das ações de campo para coleta direta de informações ou de pesquisa indireta por meio de material bibliográfico e outros recursos digitais. O processamento ocorre com a reunião de todas as informações coletadas por meio de técnicas específicas, a maior parte destas baseadas na lógica com auxílio de ferramentas tais como sistemas de informações geográficas - SIG, estatística e outras. A difusão ocorre quando a informação devidamente tratada e validada é entregue ao usuário final.

2.1 Data mining como ferramenta de auxílio a análise de informações

O *data mining* ou a mineração de dados, como mais conhecida no Brasil, é uma ferramenta relativamente nova surgida apenas nas últimas décadas e graças a capacidade de processamento dos computadores. Essa ferramenta pode ser descrita como um filtro que elimina valores aleatórios de uma amostra. Essa amostra pode ser uma sequência de letras, números, textos, imagens, sons, etc. Para isso utiliza vários métodos, dentre eles a estatística, teorema de Bayes, redes neurais, algoritmos genéticos e outros que buscam e classificam padrões em grande quantidade de dados. Esses padrões auxiliam na descoberta de novas informações que podem auxiliar no desenvolvimento de hipóteses úteis ao ciclo da Inteligência. Mas como toda ferramenta, apenas auxilia o seu usuário. Cabe ao profissional de Inteligência a avaliação e interpretação dos resultados obtidos e a formulação das hipóteses necessárias à análise de informações.

3 A análise econômica como um estudo de caso

A análise econômica é baseada em critérios diversos e pode possuir dois focos: a macroeconomia e a microeconomia. A primeira preocupa-se com o comportamento geral da economia enquanto a segunda com o seu comportamento restrito, o que pode ser uma empresa ou um nicho de mercado bem delimitado. Os dois focos são complementares e dificilmente trata-se de uma em modo sepa-

rado da outra. O tema foi escolhido para o estudo de caso por ser amplamente difundido para a sociedade e por possuir várias fontes de dados abertos, a maior parte das quais disponibilizadas por governos e órgãos oficiais.

Nesse momento pergunta-se: quanto vale uma informação? A resposta é: depende. Às vezes possuir a melhor informação não significa que os melhores resultados serão alcançados. Tudo dependerá de como o usuário final a utilizará. Um exemplo disso foram os prognósticos para a primeira crise do petróleo ocorrida em 1973. A Shell foi uma das únicas, senão a única empresa petrolífera que se antecipou àquela crise e com isso conseguiu evitar perdas substanciais de divisas. As informações estavam disponíveis, havia indícios, mas apenas a empresa Shell identificou e transformou suas previsões em ações de salvaguarda.

A atual conjuntura, com constantes quedas de bolsas de valores, evasão de divisas, falência de grandes corporações e até mesmo de Estados, impele os empresários e governos a buscar informações adequadas ao seu planejamento, tal como fez a Shell há quase quarenta anos. As análises macroeconômicas são úteis principalmente a empresas e governos que trabalham ou dependem de commodities como aço, minerais, produtos agrícolas e petróleo. Este artigo apresenta como a análise histórica do PIB dos países pode ser transformada em informação estratégica.

3.1 PIB como exemplo de fonte de informações para Inteligência

O PIB representa a soma de todos os bens e produtos produzidos por um país ao longo de um período. As séries temporais normalmente estão disponíveis em sites como o do Programa das Nações Unidas para o Meio Ambiente – PNUMA, programa financiado pela Organização das Nações Unidas – ONU.

O objetivo deste trabalho é apresentar o *data mining* como ferramenta para Inteligência; para isso optou-se por avaliar o comportamento econômico do PIB mundial e sua série com dados compilados desde o ano de 1970 até o de 2010. Os resultados anuais em dólares foram convertidos em valores qualitativos do tipo percentagem em relação ao ano anterior. Assim, a busca concentra-se na evolução individual de cada país e não na sua contribuição para a economia global. Essa abordagem tem duas razões de ser: demonstra a evolução do PIB em uma mesma unidade monetária e evita distorções causadas pela flutuação de moedas locais causadas, muitas vezes, por decisões políticas.

3.2 Ferramentas utilizadas

Para auxiliar na análise e interpretação dos dados optou-se pelo uso de mecanismos de mineração de dados – *data mining*, sistemas de informações geográficas, e estatísticas. O uso dessas ferramentas ocorreu sobre banco de dados (BD) desenvolvido a partir de fontes abertas de informações como o Google Public Data que contém os nomes dos

países e variação anual do PIB, a exemplo das informações contidas na Figura 1. O BD construído é bidimensional contendo

Os primeiros resultados surgiram sob a forma de *clusters*. As setas indicam a direção da causa e do efeito. A Figura 2

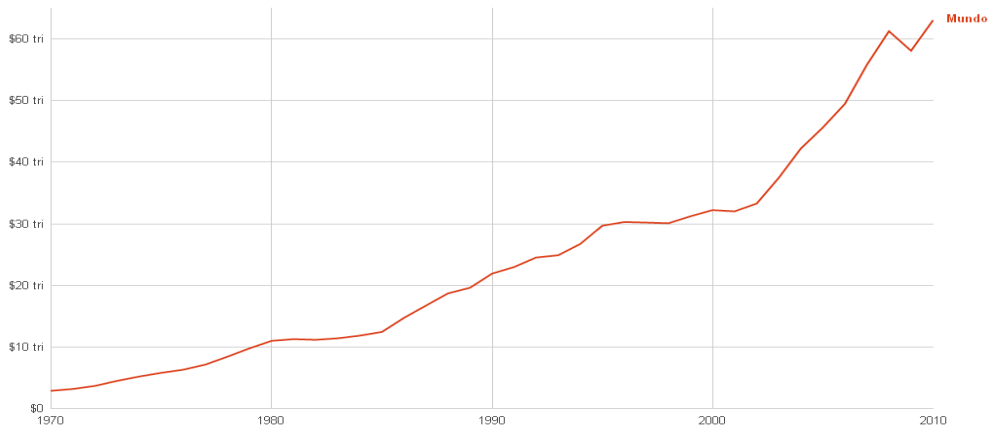


Figura 1 - PIB Mundial de 1970 a 2010 [Fonte: Google Public Data].

do 262 x 40 células, totalizando 10.480 campos. O BD foi avaliado em duas direções: a vertical contendo a comparação da variação do PIB ano a ano, independente dos países a quais pertencem; e a horizontal, onde a comparação dos PIBs ocorreu entre países.

apresenta a síntese desses resultados. Observar que cada ano está conectado a países ou a outros grupos de anos e pertence a uma classificação, podendo esta ser do tipo: causas, consequência, derivada e derivada 2.

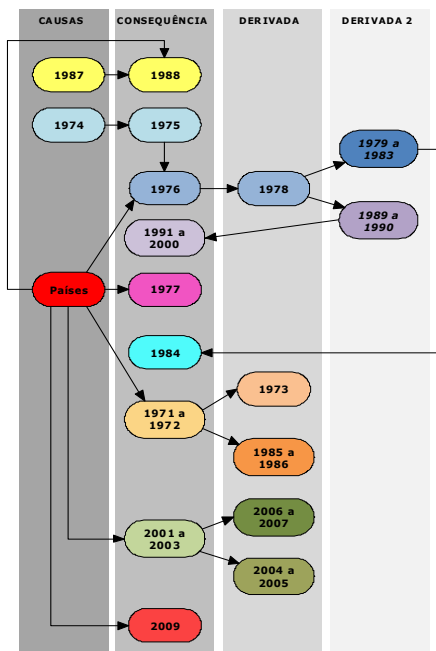


Figura 2 - Agrupamento dos anos obtidos a partir de uso de ferramenta *data mining*.

Observar também que no segmento 'Causas' existem três variáveis: Países, 1974 e 1987. A variável 'País', no BD, contém o nome de cada Estado ou Nação. A ligação dessa variável a qualquer outra variável indica que o tipo de evento foi comum a determinados grupos de países, ou seja, foi seletivo. No caso de 1987 e de 1974 não há a ligação com os países, indicando que os impactos foram globais e ocorreram em praticamente todas as economias. O ano de 1974, por exemplo, reflete os impactos da 1ª crise do petróleo ocorrida um ano antes. A partir de 1976 houve reações específicas de determinados países em relação a outros. A segunda crise do petróleo em 1979 ocorreu de forma semelhante

à de 1973 e seus efeitos foram sentidos durante quatro anos. O ano de 2009 surge isolado e foi ocasionado pela crise financeira causada pela bolha imobiliária norte-americana. Como demonstrado, é possível avaliar quebras de padrões, reações adversas e a influência de determinados eventos históricos na economia global.

3.3 Um padrão global

A análise horizontal gerou grupamentos de países, mas a grande quantidade de nações dificulta a visualização e interpre-

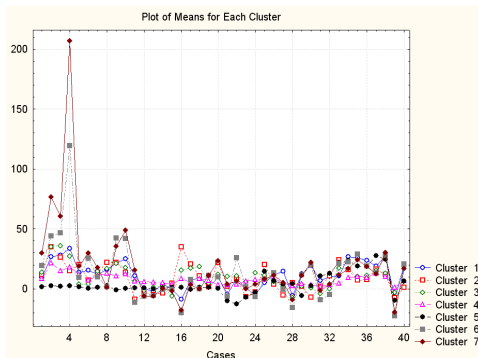


Figura 3 - Clusters com valores médios da variação dos PIBs mundiais obtidos em software estatístico.

tação dos resultados. Os países foram agrupados em sete *clusters* ou grupos

com características distintas. Os grupos possuem tamanho variado, indo de grupos menores com três países até maiores com mais de sessenta. Os países de cada grupo possuem comportamentos muito semelhantes, o que facilita a análise.

Na Figura 3, os Grupos 6 e 7, por exemplo, são formados basicamente por países exportadores de petróleo e pertencentes à Organização dos Países Exportadores de Petróleo (OPEP). As crises de 1973 e 1979 representaram grande impacto positivo nos seus PIBs, com maior impacto da 1ª crise e menor da segunda, contudo, à medida que o tempo evolui, o comportamento dos mesmos passam a ser distintos, fruto de reinvestimentos e de diversificação de fontes de renda, como no caso de Dubai que investiu em infraestrutura a partir da década de 1990 e tornou-se um centro financeiro do mundo árabe. Apesar de o agrupamento informar de forma direta quais os países pertencentes a cada *cluster*, convém utilizar o SIG para representá-los, como mostrado na Figura 4.

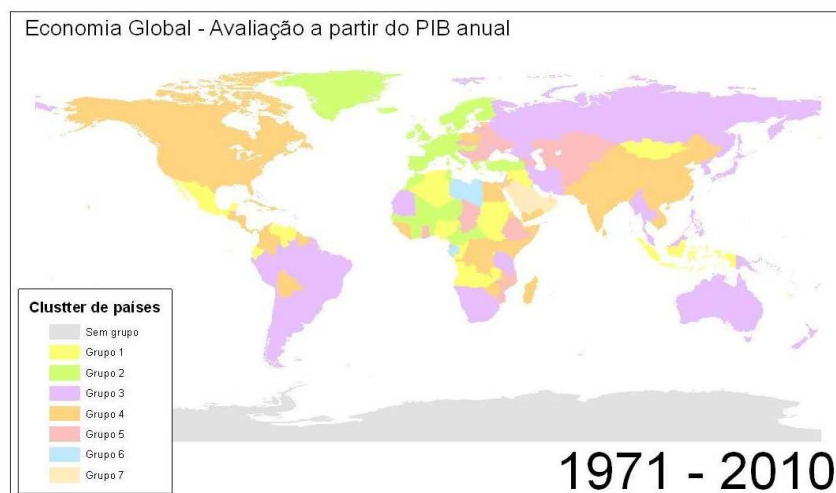


Figura 4 - Espacialização do Clusters de países em SIG.

A representação torna as associações mais lógicas, pois fica evidente a contribuição da posição geográfica de cada país e o grupamento a que pertence a exemplo do Grupo 5, formado principalmente por ex-repúblicas socialistas.

3.4 Semelhança entre comportamentos econômicos

Outra forma de visualizar a semelhança entre os países é apresentar uma planilha contendo a síntese de cada grupo. É conveniente nomear cada grupo com o nome de um país que o represente para facilitar a sua identificação. As tabelas são apresentadas nas tabelas 1 e 2.

da Romênia segue de forma isolada, ou seja, possui um padrão próprio e diferente dos demais. Nesse caso, o grupo segue a lógica de um cisne negro (TALEB, 2008), semelhante a uma aleatoriedade matemática, mas aqui provocado por efeitos culturais, sociais e econômicos cumulativos.

4 Padrões como ferramenta para projeções de futuros

Além de indicar semelhanças, é possível ainda utilizar os dados para avaliar a probabilidade de determinados eventos ocorrerem (AYRES, 2008). O primeiro passo é avaliar o coeficiente de deter-

Tabela 1 - Variação média do PIB (em %) de grupamento de países de 1971 a 1990.

Grupo	Ano																			
	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Grupo 1 (Venezuela)	11	32	28	37	17	17	14	17	23	28	13	1	3	6	0	9	4	8	1	11
Grupo 2 (Alemanha)	12	35	26	15	20	7	14	22	22	14	9	4	5	3	5	35	21	11	2	22
Grupo 3 (Brasil)	12	29	31	23	3	2	13	15	19	16	6	1	2	1	6	13	14	16	3	11
Grupo 4 (Estados Unidos)	8	20	14	18	12	7	10	13	11	13	6	5	6	6	3	9	6	10	7	4
Grupo 5 (Romênia)	2	3	2	3	2	1	1	3	1	1	1	1	0	1	2	1	1	1	1	0
Grupo 6 (Kuwait)	19	44	47	120	10	25	10	2	42	42	11	6	6	1	1	20	8	0	11	10
Grupo 7 (Arábia Saudita)	30	76	60	207	19	30	18	1	35	49	15	6	6	0	2	18	3	0	11	23

Tabela 2 - Variação média do PIB (em %) de grupamento de países de 1991 a 2010.

Grupo	Ano																			
	91	92	93	94	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10
Grupo 1 (Venezuela)	5	8	0	5	3	10	2	2	3	9	2	5	11	19	16	15	18	18	8	13
Grupo 2 (Alemanha)	2	9	7	3	20	4	0	3	2	4	3	7	18	14	10	10	15	12	3	9
Grupo 3 (Brasil)	9	8	4	10	10	7	5	1	5	1	1	3	14	14	11	13	16	13	7	8
Grupo 4 (Estados Unidos)	3	5	8	9	9	7	5	1	4	3	3	7	13	17	13	13	17	16	4	9
Grupo 5 (Romênia)	11	13	7	3	11	8	2	0	2	1	3	6	15	16	13	11	17	19	1	7
Grupo 6 (Kuwait)	6	26	2	7	8	13	4	4	2	8	5	11	24	23	17	19	31	20	14	1
Grupo 7 (Arábia Saudita)	4	6	0	4	8	11	1	5	2	3	13	11	2	8	14	11	12	12	3	22

Observa-se que o comportamento do Grupo da Venezuela assemelha-se ao da Arábia Saudita; mas em que? Nesse caso, o que ocorre com o primeiro grupo tende a ocorrer com segundo, de forma bastante semelhante. Já o grupo

minação da amostra, ou seja, calcular a probabilidade de determinado evento ocorrer. Nesse caso, quanto mais próximo de 1 (certeza de 100%), maior a probabilidade de que o evento ocorra. A própria teoria ensina que o acerto dos

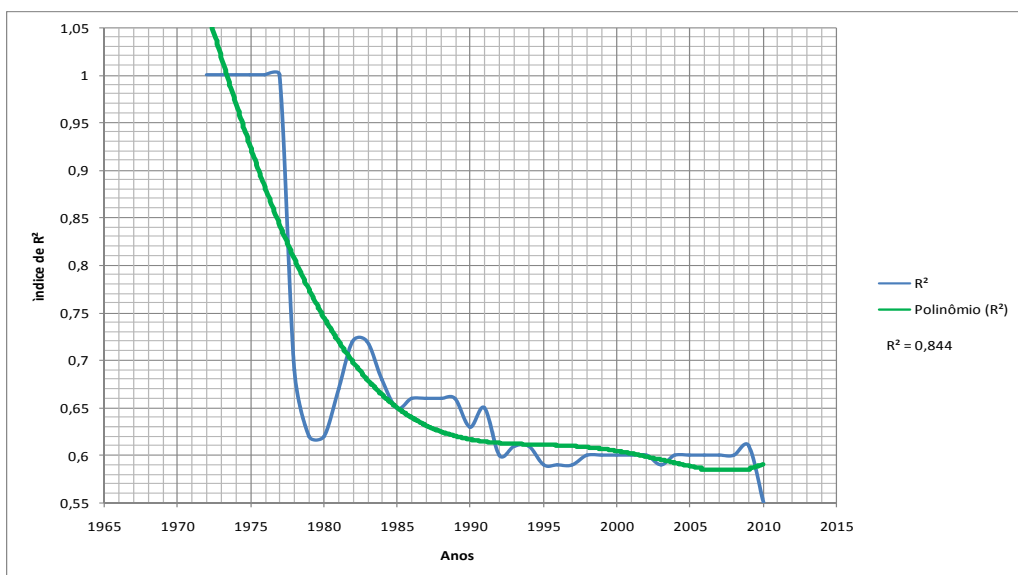


Figura 5 - Evolução do coeficiente de determinação da amostra

resultados torna-se mais difícil à medida que aumentamos a quantidade de variáveis, entretanto, e apesar de a economia mundial possuir várias destas, é possível avaliar que a probabilidade de prever séries futuras do PIB se estabiliza em $R^2 = 0.8$ (80% de margem de acerto), o que é uma probabilidade bastante razoável conforme comparação realizada entre a

variação mundial do PIB e os resultados previstos pelo modelo matemático, conforme gráfico presente na Figura 5.

Isso significa que é possível utilizar algoritmos para prever séries futuras com determinada margem de acerto. Para exemplificar isso, basta verificar que e a partir dos anos 2000 já era sinalizada uma

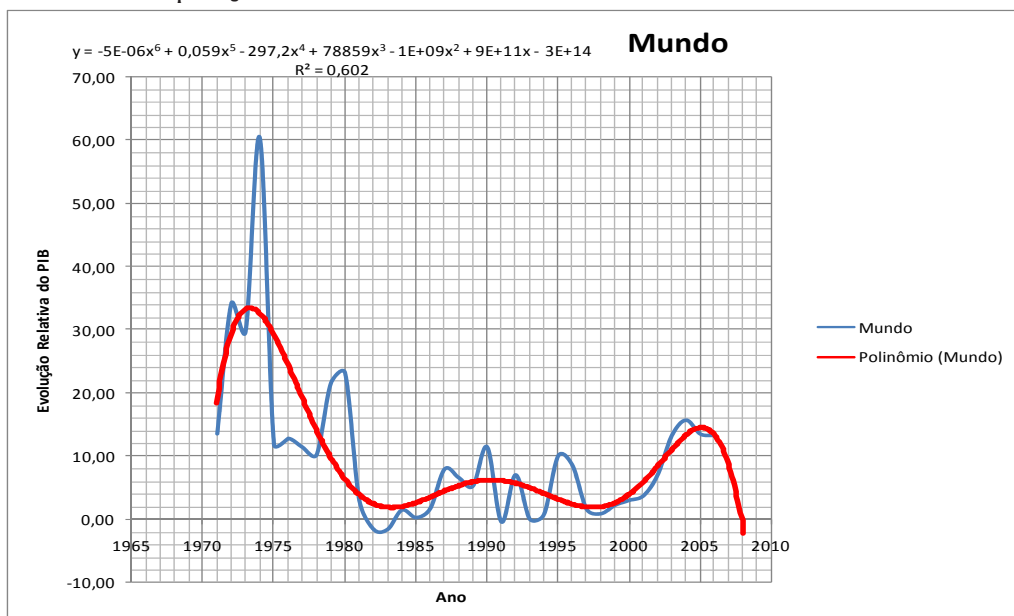


Figura 6 - Projeção de comportamento mundial.

queda do PIB para o biênio 2008-2009. A Figura 6 apresenta um dos resultados obtidos com o uso da série de dados até o ano de 2005 e sua projeção até 2008.

Como verificado, o índice de acerto é relativamente alto e a maior parte das inconsistências ocorre por causa da intensidade de aumento ou redução dos PIB dos países.

5 Conclusões

Todos os seres vivos possuem padrões de comportamento, ação e reação a determinadas situações. Esse mesmo mecanismo de defesa e sobrevivência também pode ser verificado no ser humano e em seus grupos, o que inclui a sociedade moderna e suas diversas formas de manifestações, a exemplo do PIB de cada país. A grande questão passa ser a identificação dos padrões implícitos de aliado ou inimigo. Nesse contexto, é essencial que os serviços de Inteligência delimitem quais padrões devem ser constantemente avaliados para auxiliar seus líderes e chefes de estado.

Como demonstrado, é possível estimar o futuro com certa probabilidade de acerto desde que se possuam os insumos adequados, softwares apropriados, objetivos

e missão institucionais bem definidos, e profissionais com capacidade para operacionalizar o ciclo da Inteligência. Quando adequadamente aplicadas, as técnicas de *data mining* podem gerar informações privilegiadas aos tomadores de decisão. Tudo depende do objetivo da instituição, da criatividade e das ferramentas presentes em seu corpo técnico.

6 Limitações e Sugestões

Esse artigo foi baseado em fontes abertas e utilizou sistemas de mineração de dados em tabela com duas dimensões (linhas e colunas). Contudo, um sistema de *data mining* com maiores dimensões (composto por sistemas de tabelas interligadas) e que inclua fontes negadas poderá identificar padrões que poderão se úteis à Inteligência. Qual o comportamento de países em períodos que antecedem um conflito? Quais os padrões de comportamento do crime organizado em determinado país ou região? E como prever potenciais ações de risco à segurança pública baseando-se em mudanças de padrões? Estes são exemplos de perguntas que podem nortear o desenvolvimento e emprego de tais sistemas.

Referências

- AYRES, Ian. *Super Crunchers: porque pensar com números é a nova maneira de ser inteligente*. Rio de Janeiro: Ediouro, 2008.
- NAPOLEONI, Loretta. *Economia bandida: a nova realidade do capitalismo*. São Paulo: Difel, 2010.
- PLATT, Washington. *A produção de informações estratégicas*. Rio de Janeiro: Agir, 1974.
- SANTOS, Marco A.; FRANCO, Jacinto Rodrigues. *A Atividade de Inteligência na Segurança para o Século XXI*. 2011. [s.l.: s.n.], 2011.
- TALEB, Nassim Nicholas. *A lógica do Cisne Negro: o impacto do altamente improvável*. 2008. Rio de Janeiro: Best Seller, 2008.