

**Sistema Inteligente para Identificação e Monitoramento de Fontes de Informação:  
um Estudo de Caso na Companhia Energética de Minas Gerais**

**ERIC DE PAULA FERREIRA**

UNIVERSIDADE DO ESTADO DE MINAS GERAIS (UEMG)

**FREDERICO GIFFONI DE CARVALHO DUTRA**

UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)

**FERNANDO SILVA PARREIRAS**

UNIVERSIDADE FUMEC (FUMEC)

**WLADMIR CARDOSO BRANDÃO**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (PUC MINAS)

**DANIEL HENRIQUE MOURÃO FALCI**

UNIVERSIDADE FUMEC (FUMEC)

# **SISTEMA INTELIGENTE PARA IDENTIFICAÇÃO E MONITORAMENTO DE FONTES DE INFORMAÇÃO: um estudo de caso na Companhia Energética de Minas Gerais**

## **INTELLIGENT SYSTEM FOR IDENTIFICATION AND MONITORING OF SOURCES OF INFORMATION: a case study at Companhia Energética de Minas Gerais**

### **INTRODUÇÃO**

A eficiência na gestão organizacional vem sendo discutida ao longo do tempo na literatura, com o objetivo de identificar pontos de melhorias nos resultados empresariais. Entre as ferramentas de gestão, a gestão de processos tem sido utilizada com um enfoque administrativo, otimizando os processos das organizações de forma a facilitar o atendimento das necessidades das partes interessadas, assegurando melhores desempenho e assertividade.

Para Parreiras e Dutra (2019), ser competitivo é conhecer o ambiente externo e dominar o ambiente interno, definindo as estratégias de ação que revertam em sucesso para a organização. Segundo os autores, como o ambiente externo apresenta grande dificuldade em ser alterado, resta à organização monitorá-lo, para definir a melhor organização interna. Fazer isso sem o apoio de informações precisas e confiáveis é deixar à sorte, os caminhos da organização.

Em seus estudos, Dutra e Barbosa (2015) relatam que, com a reestruturação do setor elétrico brasileiro, foi necessário dividir as empresas de energia elétrica em áreas específicas, no intuito de promover a competição no setor. Dessa forma, os consumidores foram divididos entre cativos e livres. Enquanto os cativos estão sujeitos à tarifa regulada da concessionária de distribuição à qual está conectado, os consumidores livres têm a opção de compra de energia elétrica de qualquer fornecedor, por meio de contratos bilaterais com as empresas geradoras de energia elétrica ou com empresas exclusivas de comercialização.

A partir dessas reflexões e tendo como cenário a Companhia Energética de Minas Gerais - CEMIG, a pesquisa tem como objetivo desenvolver a arquitetura do sistema inteligente implementado no projeto P&D GT641 – “Proposta de uma Metodologia de Busca de Informações e Estimativa de Demanda e Consumo de Clientes Livres Potenciais”, que visa viabilizar o processo sistematizado de estimativa de demanda por consumo de energia elétrica de empreendimentos de grande porte.

O trabalho seguirá a seguinte estrutura: As Seções 2 e 3 apresentam a revisão teórica dos conceitos de Processo Manual de Inteligência Competitiva e Fontes de Informação; A Seção 4 apresenta os procedimentos metodológicos; o arcabouço conceitual proposto é apresentado na Seção 5; Seção 6 apresenta as considerações finais.

### **PROCESSO MANUAL DE INTELIGÊNCIA COMPETITIVA**

O processo de inteligência competitiva pode ser visto como um conjunto de ações coordenadas (obtidas legalmente) de busca, tratamento, distribuição e proteção de informação útil aos atores econômicos (TARAPANOFF, 2004). Esse processo permite o aumento da

competitividade da organização por meio de um entendimento maior, aprofundado e ético do ambiente competitivo como um todo.

Com base nos estudos de Gomes e Braga (2001), Dutra (2016) propôs um processo manual de inteligência competitiva composto em cinco etapas, conforme mostra a Figura 1.

**Figura 1: Modelo manual de inteligência competitiva**



Fonte: Dutra (2016)

## FONTES DE INFORMAÇÃO

Para Cunha (2001), as fontes de informação abrangem os manuscritos e as publicações

impressas, além de objetos, como amostras minerais, obras de arte ou peças museológicas, podendo ser divididas em três categorias: documentos primários, documentos secundários e documentos terciários. O Guia BVS (2003), por sua vez, conceitua fonte de informação como qualquer recurso que responda uma demanda por parte dos usuários, incluindo produtos e serviços de informação, pessoas ou rede de pessoas e programas de computador.

Quando levamos a temática para o ambiente organizacional, Choo (2006) em seus estudos sobre a gestão do conhecimento nas organizações classifica as fontes de informação em quatro categorias: externas e pessoais, externas e impessoais, internas e pessoais, e internas e impessoais. O autor ressalta que a informação é um componente intrínseco de quase tudo o que uma organização faz.

Estudiosos do tema como Taylor (1985), Lester e Waters (1989), Auster e Choo (1994), Barbosa (2002) e Farias (2007) seguem a mesma linha de raciocínio de Aguilar (1967), defendendo que as informações estratégicas para o ambiente de negócios da organização são sempre informações externas e quase sempre oriundas de fontes externas. Outros autores, como Gans (1980), Degent (1986) e Lesca e Almeida (1994), desenvolveram novas perspectivas, como a classificação de fontes de informação com base nos seus fluxos, nos setores do ambiente organizacional ou conforme a segmentação geográfica.

Com base nesses estudos, Dutra (2014) elaborou um modelo de abrangência holística para a classificação de fontes de informação, conforme mostra a Figura 2:

**Figura 2: Modelo de classificação de fontes de informação**



**Fonte: Dutra (2014)**

## PROCEDIMENTOS METODOLÓGICOS

O método utilizado para organizar a pesquisa segue a abordagem *Design Science Research* de acordo com o Framework da pesquisa proposto por Hevner (2014). O arcabouço conceitual (método proposto) e sua instanciação (sistema computacional) são validados por meio de estudo de caso. Este estudo de caso brevemente apresentado neste artigo tem caráter instrumental, pois usa um caso para obter insights sobre uma situação, neste caso a aplicação do arcabouço proposto.

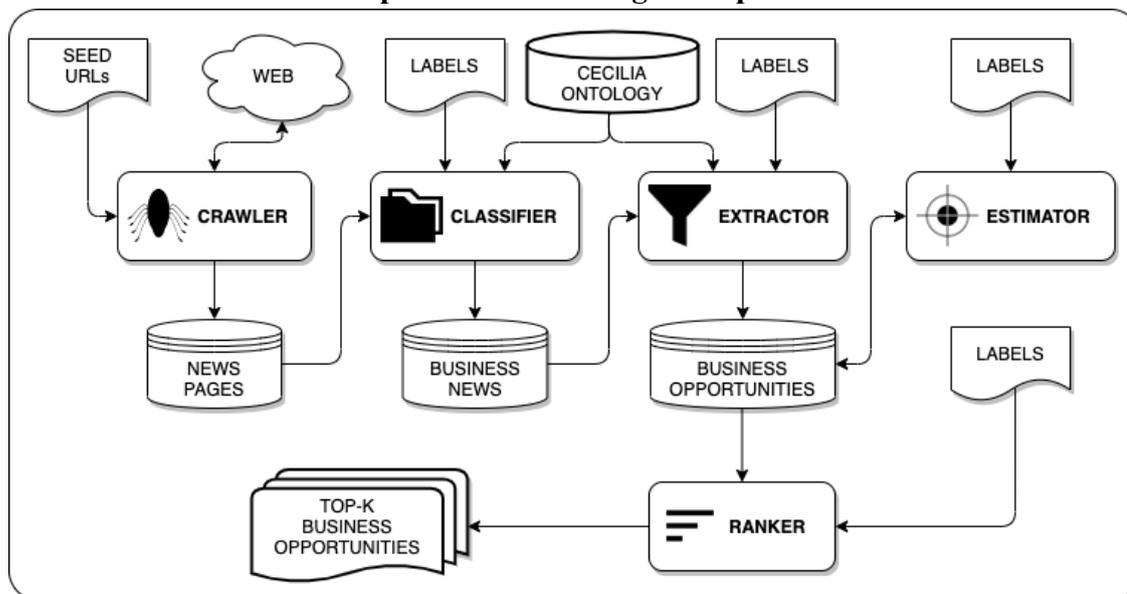
Para atingir o objetivo da pesquisa, primeiro foi realizado o mapeamento do processo manual de identificação e monitoramento de fontes de informação sobre empreendimento de grande porte utilizado na Cemig. Foram mapeadas todas as etapas do processo vigente até no momento, processo esse que é detalhado nos estudos de Parreiras e Dutra (2019), onde foram identificadas as etapas de inteligência que o sistema deveria ter para atender a proposta do projeto P&D GT641 – “Proposta de uma Metodologia de Busca de Informações e Estimativa de Demanda e Consumo de Clientes Livres Potenciais”.

Após a identificação e entendimento das etapas, foram realizadas revisões na literatura para identificar algoritmos considerados o estado da arte que atendam ao processo de sistematização. O arcabouço conceitual proposto, com o detalhamento das etapas de inteligência identificadas no processo de identificação e monitoramento de empreendimentos de grande porte é descrito na seção a seguir.

### ARCABOUÇO CONCEITUAL PROPOSTO

Com objetivo de viabilizar um processo sistematizado de estimativa de demanda por consumo de energia elétrica de empreendimentos de grande porte, foi proposta uma arquitetura composta de quatro componentes: *crawler* (coletor de notícias), *classifier* (classificador de notícias), *extractor* (extrator de evidências relacionadas à oportunidades de negócio) e *ranker* (ranqueador de oportunidades de negócio). A Figura 3 abaixo ilustra a arquitetura da abordagem proposta.

**Figura 3: Arquitetura da abordagem proposta para análise e estimativa de demanda de empreendimentos de grande porta**



Fonte: elaborado pelos autores

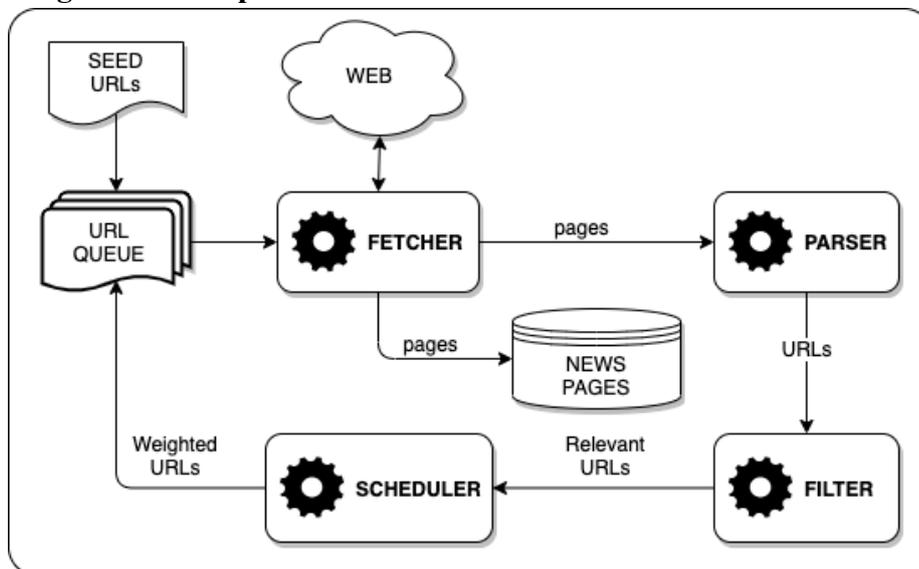
O processo consiste em coletar notícias sobre empreendimentos que são classificadas como úteis ou não úteis a partir da Web, para posteriormente extrair das notícias úteis oportunidades de negócio de acordo com ontologia previamente desenvolvida. O consumo de

demanda de energia elétrica de cada oportunidade de negócio é então estimado para que a oportunidade de negócio possa ser ranqueada a partir de sua importância para a CEMIG.

### Componente CRAWLER

O componente de coleta de notícias da Web utiliza coletores disponíveis no framework que possibilitam a implementação de métodos de coleta e extração de informação estruturada de sítios da Web (Castillo, 2005). Particularmente, a partir de uma semente de sítios relevantes na Web (URLs) fornecidos pela CEMIG, o componente varre a Web coletando páginas HTML dos sítios semente e de todos os sítios apontados a partir das páginas coletadas, se realimentando de maneira recursiva. A Figura 4 ilustra o funcionamento do componente CRAWLER.

**Figura 4: Componente CRAWLER de coleta de notícias da Web**

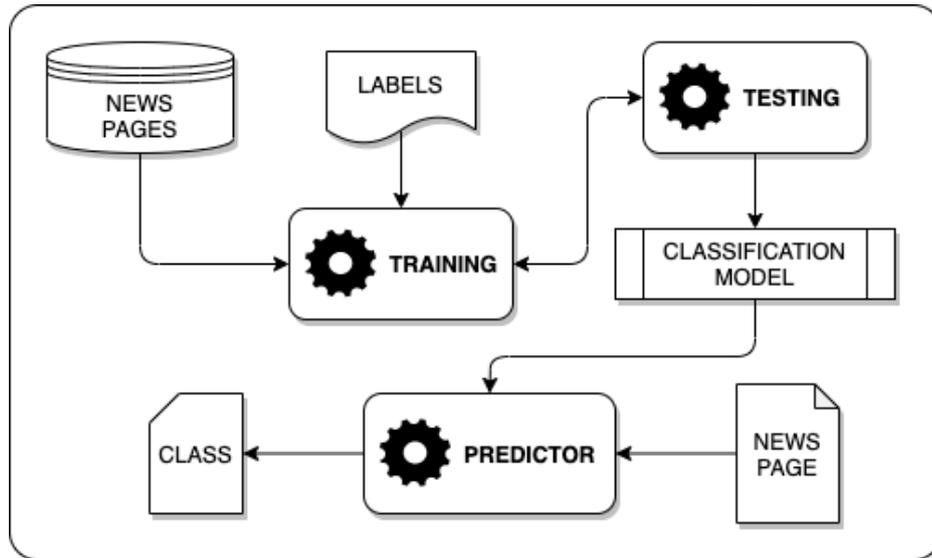


Fonte: elaborado pelos autores

### Componente CLASSIFIER

O componente para classificação de notícias utiliza algoritmos de classificação considerados estado da arte na literatura (Huang, 2015), possibilitando a classificação binária das páginas de notícias coletadas previamente, considerando a relevância de seu conteúdo como fonte de informação sobre oportunidades de negócio (empreendimentos) relevantes comercialmente para a CEMIG. A Figura 5 ilustra o funcionamento do componente CLASSIFIER.

**Figura 5: Componente CLASSIFIER para a classificação de notícias**

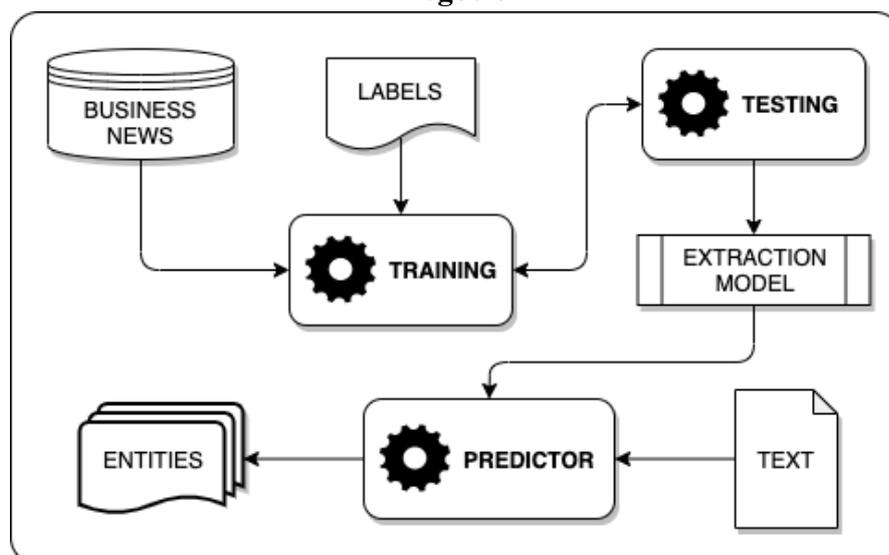


Fonte: elaborado pelos autores

### Componente EXTRACTOR

O componente para extração de evidências (*features*) relacionadas à oportunidades de negócio a partir de páginas de notícias coletadas da Web utiliza algoritmos de aprendizagem supervisionada considerados estado da arte na literatura, o que possibilita a implementação de métodos de reconhecimento e classificação de entidades - *Named Entity Recognition and Classification* (NERC) - para extração das evidências relevantes (Nadeau, 2007). A Figura 6 ilustra o funcionamento do componente EXTRACTOR.

**Figura 6: Componente EXTRACTOR para extração de evidências de oportunidades de negócio**

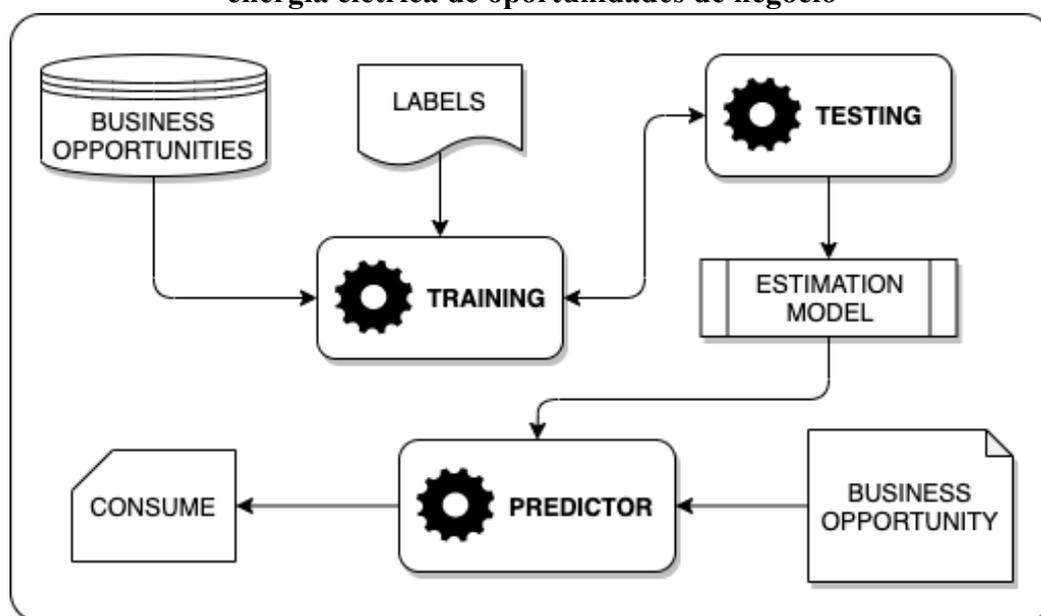


Fonte: elaborado pelos autores

### Componente ESTIMATOR

O componente para estimativa de consumo utiliza algoritmos de aprendizagem supervisionada disponíveis no *framework scikit-learn* (Pedregosa, 2011), o que possibilita a implementação de métodos de regressão para estimativa de dados ausentes, em particular para predição de potencial de consumo de energia relacionada à oportunidade de negócio considerando evidências relacionadas à própria oportunidade de negócio. Os métodos de regressão implementados são *Linear Regression*, *Random Forest Regression* e *Gradient Boosting Regression*. A Figura 7 ilustra o funcionamento do componente ESTIMATOR.

**Figura 7: Componente ESTIMATOR para estimativa de potencial de consumo de energia elétrica de oportunidades de negócio**

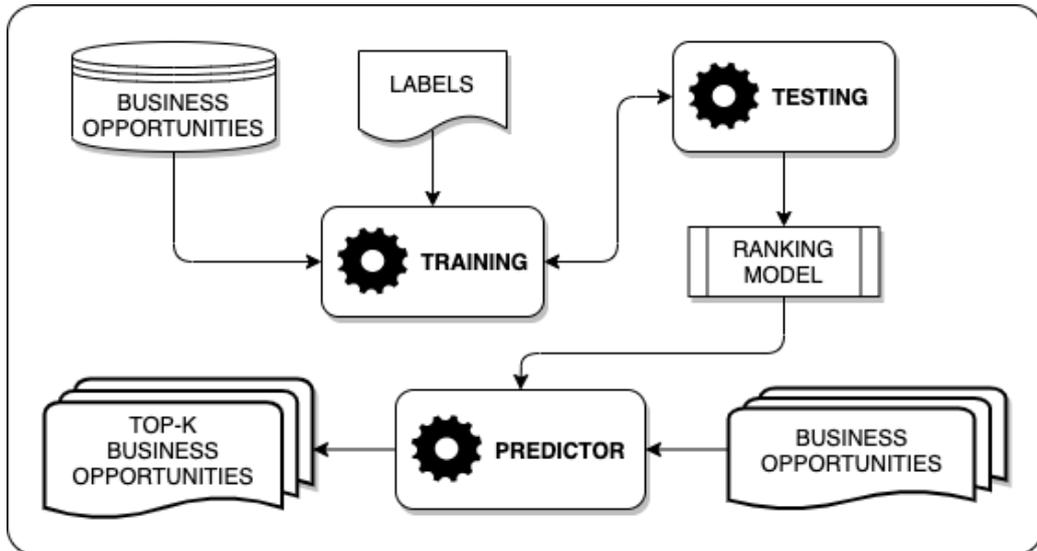


Fonte: elaborado pelos autores

### Componente RANKER

O componente para *ranking* de oportunidades de negócio utiliza algoritmos de aprendizagem supervisionada estado da arte na literatura para prever o consumo de energia elétrica relacionados a uma oportunidade de negócio. Além disso, utiliza algoritmos que possibilitam a implementação de métodos de *Learning to Rank* (LTR) em cenários de aprendizagem profunda (*deep learning*) para *ranking* das oportunidades de negócio (Pasumarthi, 2019). A Figura 8 ilustra o funcionamento do componente RANKER.

**Figura 8: Componente RANKER para ranking de oportunidades de negócio**

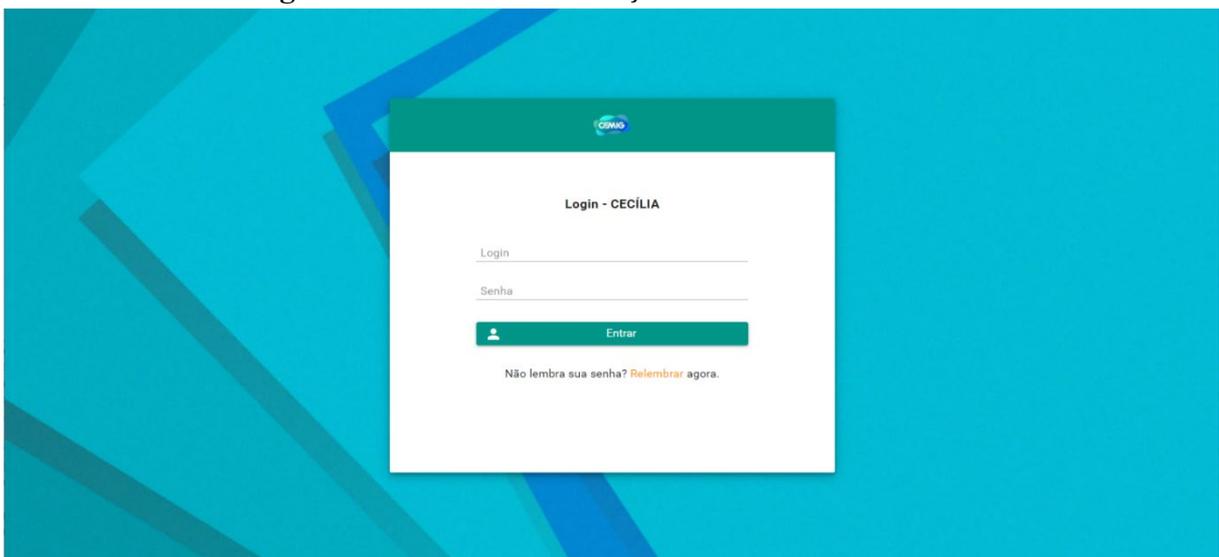


Fonte: elaborado pelos autores

## VALIDAÇÃO

A validação do arcabouço proposto, que é a sua instanciação, é feita por meio da evidência da implementação da funcionalidade. A tela de autenticação do sistema, que tem o nome de CECÍLIA (acrônimo que remete as entidades que viabilizaram a realização projeto) é ilustrada na Figura 9. A seguir, serão apresentadas algumas telas do sistema que caracterizam estas evidências.

**Figura 9: Tela de autenticação do sistema CECÍLIA**

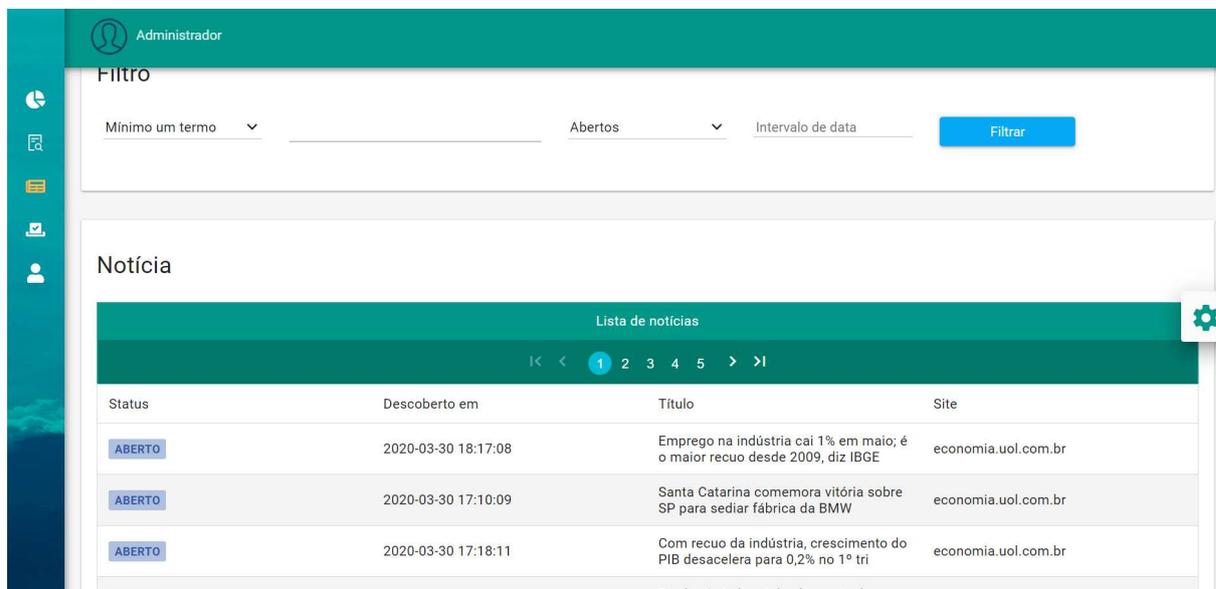


Fonte: elaborado pelos autores

A Figura 9 ilustra o funcionamento dos componentes CRAWLER e CLASSIFIER. As notícias são extraídas da web e apresentadas em uma lista. Ao clicar sobre uma notícia, o texto e as informações importantes são marcados. Após aprovar a notícia, ela será encontrada pela

filtragem do status Aprovado. Após reprovar a notícia, ela será encontrada pela filtragem do status Negado. O usuário pode filtrar por pelo menos um termo ou fazer uma pesquisa específica contendo todos os termos.

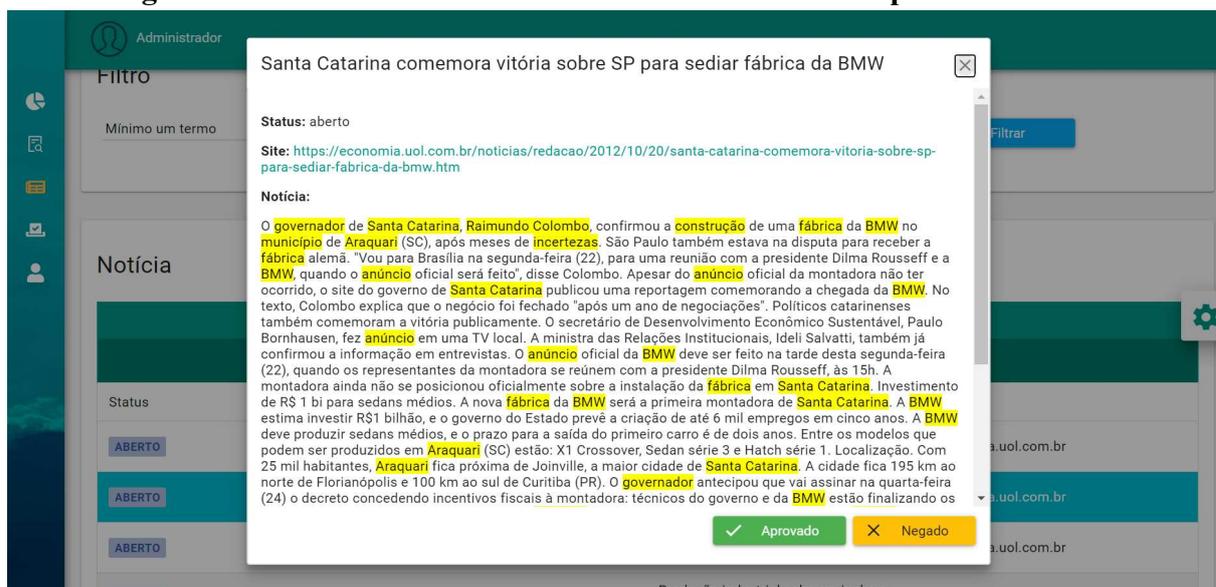
**Figura 9: Tela demonstrando o funcionamento dos Componentes CRAWLER e CLASSIFIER**



Fonte: elaborado pelos autores

A Figura 10 apresenta o funcionamento do componente EXTRATOR, que destaca os elementos identificados por meio de aplicação dos algoritmos de NERC na notícia extraída na web pelo componente CRAWLER e classificada como relevante pelo componente CLASSIFIER.

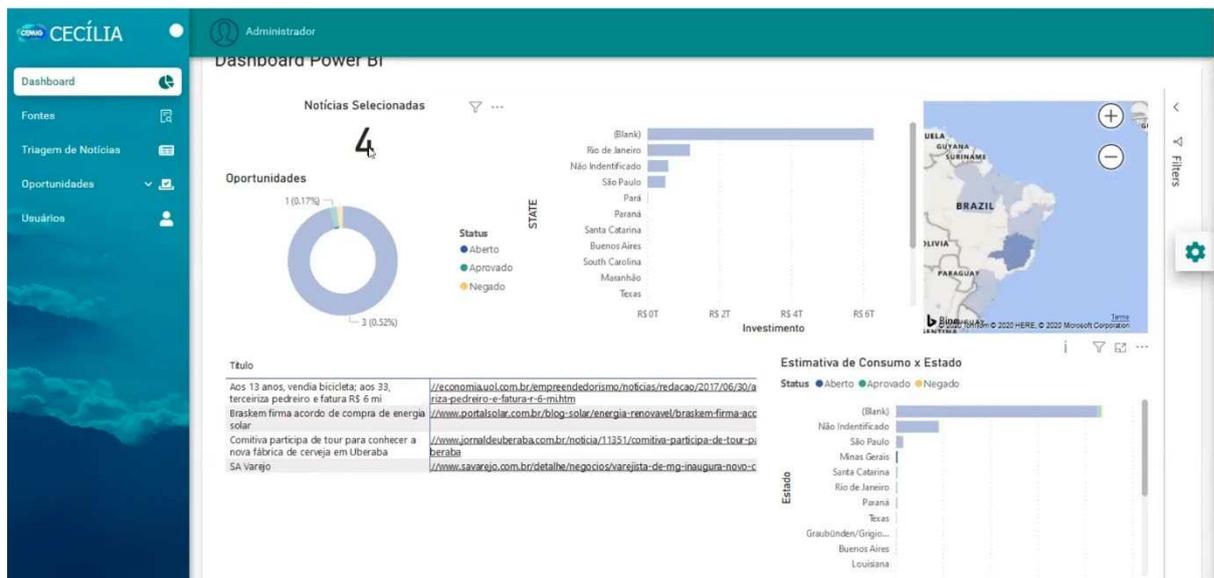
**Figura 10: Tela demonstrando o funcionamento do Componente EXTRACTOR**



Fonte: elaborado pelos autores

A Figura 11 apresenta um exemplo de resultado de aplicação dos algoritmos desenvolvidos para os componentes ESTIMATOR e RANKER. Nesta tela, o usuário pode visualizar oportunidades e filtrar. O usuário pode analisar a estimativa de consumo (resultado do componente ESTIMATOR), do lado direito, e a lista com as notícias ordenadas conforme aderência para prospecção (resultado do componente RANKER). Ao clicar sobre um ponto de oportunidade no mapa, o sistema apresenta o site, status e o título da oportunidade.

**Figura 11: Tela demonstrando o funcionamento dos Componentes ESTIMATOR e RANKER**



Fonte: elaborado pelos autores

## CONSIDERAÇÕES FINAIS

Tendo como cenário a Companhia Energética de Minas Gerais - CEMIG, a pesquisa teve como objetivo desenvolver a arquitetura de um sistema inteligente implementado no projeto P&D GT641 – “Proposta de uma Metodologia de Busca de Informações e Estimativa de Demanda e Consumo de Clientes Livres Potenciais”, que visa viabilizar o processo sistematizado de estimativa de demanda por consumo de energia elétrica de empreendimentos de grande porte.

Para atingir o objetivo da pesquisa, primeiro foi realizado o mapeamento do processo manual de identificação e monitoramento de fontes de informação sobre empreendimento de grande porte utilizado na Cemig e após isso, foram realizadas diversas revisões na literatura para identificar algoritmos considerados o estado da arte que atendam ao processo de sistematização. Com a recuperação dos algoritmos, foi proposto o arcabouço conceitual que é a base da sistematização de todo o processo.

O estudo se limitou a desenvolver a arquitetura de um sistema inteligente que visa viabilizar o processo sistematizado de estimativa de demanda por consumo de energia elétrica de empreendimentos de grande porte, sendo assim, a princípio, não é possível generalizar esse processo para outras áreas. Uma sugestão de pesquisas futuras seria a aplicação da mesma

metodologia para identificação e monitoramento de clientes em empresas de outros segmentos.

## **REFERÊNCIAS**

AGUILAR, F. J. Scanning the business environment. New York, NY: Macmillan, 1967.

AUSTER, E.; CHOO, C. W. How senior managers acquire and use information in environmental scanning. *Information Processing & Management*. v. 30, n. 5, p. 607-618, 1994.

BARBOSA, R. R. Inteligência empresarial: uma avaliação de fontes de informação sobre o ambiente organizacional externo. *Datagrama Zero*, v. 3 n. 6, 2002.

CASTILLO, Carlos. Effective web crawling. In: ACM SIGIR forum. New York, NY, USA: Acm, 2005. p. 55-56.

CHOO, Chun Wei. A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. 2. ed. São Paulo: SENAC, 2006.

CUNHA, M. B. Para saber mais: fontes de informação em ciência e tecnologia. Brasília: Briquet de Lemos Livros, 2001.

DEGENT, R. J. A importância estratégica e o funcionamento do serviço de inteligência empresarial. *Revista de Administração de Empresas*, v. 26, n. 1, p. 77-83, 1986.

DUTRA, F. G. C. Modelo de inteligência competitiva: uma proposta para monitoramento e prospecção de clientes na Companhia Energética de Minas Gerais. *Revista Inteligência Competitiva*, São Paulo, v. 6, n. 4, p. 224-243, out./dez. 2016.

DUTRA, F. G. C.; BARBOSA, R. R. Serviços de informação e monitoramento de clientes no setor elétrico. IN: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 2014, Belo Horizonte. GT 4: Gestão da Informação e do Conhecimento nas organizações, p. 1596-1614, 2014.

DUTRA, F. G. C.; BARBOSA, R. R. Serviços de informação e monitoramento de clientes no setor elétrico. *Revista Pensar Comunicação*, v. 2, p. 129, 2015.

FARIAS, G. B.; VITAL, L. P. Informação para negócios e políticas de informação. *Revista ACB: Biblioteconomia*, v. 12, n. 1, p. 87-98, jan./jun., 2007.

GANS, H. J. Deciding what's news. London: Constable Communication and Society Series. 1980.

GODOY, A. S. A Pesquisa Qualitativa e sua Utilização em Administração de Empresas. *Revista de Administração de Empresas*, São Paulo, v. 35, n. 4, 1995.

GUIA BVS. 2003. Operación de las fuentes de información. São Paulo: BIREME/OPS/OMS, 2003. p.1-10.

GOMES, E.; BRAGA, F. Inteligência competitiva: como transformar informação em um negócio lucrativo. Rio de Janeiro: Editora Campus, 2001.

HUANG, Zhiheng; XU, Wei; YU, Kai. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991, 2015.

LESCA, H.; ALMEIDA, F. C. Administração estratégica da informação. Revista de Administração, São Paulo, v. 29, n. 3, p. 66-75, jul./set.1994.

LESTER, R.; WATERS, J. Environmental scanning and business strategy. London, UK: British Library, Research and Development Department, 1989.

NADEAU, David; SEKINE, Satoshi. A survey of named entity recognition and classification. *Linguisticae Investigationes*, v. 30, n. 1, p. 3-26, 2007.

PARREIRAS, F. S.; DUTRA, F. G. C. Identificação e monitoramento de fontes de informação sobre empreendimentos de grande porte: o processo vigente na Companhia Energética de Minas Gerais. IN: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 2019, Florianópolis. GT 4: Gestão da Informação e do Conhecimento nas organizações, 2019.

PASUMARTHI, Rama Kumar et al. Tf-ranking: Scalable tensorflow library for learning-to-rank. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019. p. 2970-2978.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, v. 12, p. 2825-2830, 2011.

TARAPANOFF, K. Inteligência social e inteligência competitiva. *Encontros Bibli Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, n. especial, 1º sem. 2004.

TAYLOR, R. S. Information Values in decision context. *Information Management Review*. v. 1, n. 1, p. 47-55, 1985.

HEVNER, Alan R. et al. Design science in information systems research. *MIS quarterly*, p. 75-105, 2004.