

A UTILIZAÇÃO DE CRITÉRIOS DE ESPECIALISTAS EM SELEÇÃO DE IDEIAS NA TAREFA DE MINERAÇÃO DE IDEIAS

Luiz Fernando Spillere de Souza¹;

Alexandre Leopoldo Gonçalves²

Abstract: *Idea mining aims to capture and distinguish an idea within a text. The ideal condition is that it can perform the task of an expert using all the criteria that are used in the selection of ideas. This paper demonstrates how Machine Learning techniques and selection criteria considered by experts can positively impact the mining of ideas during the identification of potential ideas. The results obtained through two classifiers presented a gain of accuracy in the classification of texts as ideas when incorporating selection criteria. In this way, it has been demonstrated through practical tests that the use of expert criteria in the selection of ideas improves the accuracy of the automatic methods of idea mining.*

Keywords: *Idea Mining, Machine Learning, Idea Selection Criteria.*

Resumo: *A mineração de ideias tem como objetivo captar e distinguir uma ideia dentro de um texto. A condição ideal é que se possa executar a tarefa de um especialista na sua área utilizando-se de todos os critérios que ele utiliza na seleção de ideias. Este trabalho demonstra como a utilização de técnicas de Aprendizado de Máquina e critérios de escolha considerados por especialistas podem impactar positivamente na mineração de ideias durante a etapa de identificação de potenciais ideias. Os resultados obtidos através de dois classificadores apresentaram um ganho de acurácia na classificação de textos como ideias ao incorporar critérios de escolha. Desta forma, demonstrou-se por meio de testes práticos que a utilização de critérios de especialistas na seleção de ideias melhora a acurácia dos métodos automáticos de mineração de ideias.*

Palavras-chave: *Mineração de Ideias; Aprendizado de Máquina, Critérios de Seleção de Ideias.*

¹ Doutorando no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC) - Universidade Federal de Santa Catarina (UFSC) - Florianópolis - SC - Brasil. Email: spillere@gmail.com

² Professor Titular no Departamento de Engenharia e Gestão do Conhecimento (dEGC) – Universidade Federal de Santa Catarina (UFSC) - Florianópolis – SC – Brasil. Email: a.l.goncalves@ufsc.br

1 INTRODUÇÃO

A mineração de ideias vem sendo utilizada pelas organizações de forma estratégica para promover inovação de produtos e serviços. No mundo globalizado onde a concorrência é contínua, lançar um produto inovador no mercado pode ser o diferencial entre a existência ou não de uma organização. Neste sentido, vários métodos automáticos de mineração de ideias vêm sendo empregados e atualizados constantemente, conforme cita a literatura.

Estes estudos inicialmente se preocuparam em registrar, classificar e disponibilizar as ideias dentro das organizações (Paukkeri, 2009). Posteriormente surgiram os métodos automáticos no domínio tecnológico (Thorleuchter & Van Den Poel, 2012; Thorleuchter, Van Den Poel, & Prinzie, 2010), no domínio social (Thorleuchter, Herberz, & Van Den Poel, 2011) e no domínio interdisciplinar (Thorleuchter & Van den Poel, 2016), voltados para web (Thorleuchter & Van Den Poel, 2013) e combinados com outras técnicas (Christensen, Norskov, Frederiksen, & Scholderer, 2017; Li, Li, & Chen, 2014; Sérgio, Souza, & Gonçalves, 2017; Vwen, Lee, Tan et al., 2017; Vwen, Lee, & Tan, 2017). Também são citados estudos que avaliam os métodos automáticos de mineração de ideias utilizando especialistas (Alksher et al., 2017; Klein & Garcia, 2015).

A avaliação humana é frequentemente necessária no processamento de linguagem natural para avaliar desempenho do sistema (Kishida, 2005), sendo importante considerar a seleção dos avaliadores, em termos de especialização e experiência (Alksher et al., 2017). Em um contexto ideal, um método automático de Mineração de Ideias deve incorporar todos estes critérios de seleção de ideias utilizados por especialistas, para que o sistema possa funcionar da forma mais semelhante possível com as suas escolhas. Estes critérios podem representar fatores quantitativos e qualitativos a serem considerados, tais como: compatibilidade de recursos, necessidade do mercado, superioridade do produto, singularidade, complexidade tecnológica e riscos sobre resultado de projetos (Mousavi, Torabi, & Tavakkoli-Moghaddam, 2013).

A partir do contexto acima exposto, o objetivo deste trabalho é responder a seguinte questão: a utilização de critérios de especialistas em seleção de ideias em determinado domínio pode melhorar a acurácia dos métodos automáticos de Mineração de Ideias?

O artigo está estruturado em cinco seções: a introdução com a contextualização do tema; a segunda seção apresenta o referencial teórico; a terceira seção expõe os procedimentos metodológicos, a quarta seção uma análise e discussão dos resultados; e, na quinta e última seção, são apresentadas as considerações finais do estudo.

2 FUNDAMENTAÇÃO TEÓRICA

As organizações empresariais acumulam muito conhecimento, incluindo grandes bancos de dados sobre produtos, clientes, concorrentes, mercados, entre outros, reunidos ao longo dos anos. As organizações também têm conhecimento arquivado no espaço de trabalho pessoal de seus colaboradores, como arquivos de texto, dados, livros e e-mails (Nonaka & Takeuchi, 2000).

A tendência é que esses bancos de dados cresçam rapidamente em volume e, desta forma, tornam-se impraticáveis análises manuais. Usando métodos avançados de mineração de texto é possível processar volumosas quantidade de dados e encontrar conexões semânticas entre eles. Para realmente apoiar a criatividade e inovações nas organizações existem ferramentas de *software* que coletam e armazenem ideias para então torná-las gerenciáveis e úteis em um processo de tomada de decisão (Paukkeri, 2009).

A mineração de ideias é uma nova tendência de pesquisa em engenharia e extração de conhecimento. Ele depende principalmente de informações latentes e suas mudanças dinâmicas para conduzir a criação de ideias, integração e avaliação durante o processo de criatividade sustentável. O processo de mineração de ideias é baseado na extração de termos latentes no texto a partir de texto não estruturado, utilizando padrões de caracterização da ideia (Alksher et al., 2017).

Existem na literatura vários métodos de mineração de ideias, porém não é o foco deste trabalho discutir, avaliar ou comparar estes métodos. Optou-se pelo estudo de *Machine Learning* (Aprendizado de Máquina) aplicado à Mineração de Ideias pelo fato deste já demonstrar sucesso em aplicações anteriores (Christensen, Liland et al., 2017; Christensen, Norkov et al., 2017; Vwen, Lee, & Tan, 2017), combinado com o fato de ser uma técnica atual e promissora.

2.1 MACHINE LERANING E APLICAÇÕES NA MINERAÇÃO DE IDEIAS

Machine Learning (Aprendizado de Máquina) é um subcampo da ciência da computação que desenvolve a capacidade dos computadores de aprender sem prévia programação (Samuel, 2000). Em outras palavras pode ser definido como ensinar computadores a reconhecer padrões (Christensen, Norkov et al., 2017).

As vantagens desta abordagem é uma acurácia comparável àquela alcançada por especialistas humanos representando uma economia considerável em termos de força de trabalho especialista, já que nenhuma intervenção de engenheiros de conhecimento ou especialistas de domínio é necessária para a construção do classificador ou para a sua portabilidade para um conjunto diferente de categorias (Sebastiani, 2002).

Normalmente, as técnicas de *Machine Learning* são divididas em dois ramos: técnicas de aprendizagem supervisionadas (como regressão, análise discriminante, árvores de decisão, redes neurais e máquinas de vetores de suporte) e técnicas de aprendizado não supervisionadas (análise de componentes principais, análise de agrupamentos e alocação latente de Dirichlet) (Christensen, Norskov et al., 2017).

Vários algoritmos têm sido apresentados como os mais adequados para a classificação de documentos, porém, os algoritmos supervisionados *Support Vector Machines* (Máquinas de Vetores de Suporte) e *Naive Bayes* demonstram bons resultados na classificação de texto (Felden, Bock, Gräning, Molotowa, & Saat, 2006). Em Christensen, Norskov et al., (2017) também é demonstrado que o *Support Vector Machines* apresenta um desempenho superior comparando técnicas de classificação supervisionadas com ênfase na mineração de texto. Da mesma forma o desempenho do *Naive Bayes* demonstra ser um bom classificador para estas aplicações.

Desta forma, partindo do pressuposto acima, escolhe-se utilizar *Support Vector Machines* e *Naive Bayes* para os testes utilizados neste trabalho e demonstrados na seção 4.

2.1.1 *Support Vector Machines* e suas aplicações na Mineração de Ideias

O *Support Vector Machines* (SVM) é uma técnica de aprendizado supervisionada no campo da aprendizagem de máquina, aplicável a classificações e análise de regressão (Vapnik, 2013). Matematicamente, uma SVM gera uma decisão limite que melhor separa duas classes (por exemplo, no contexto deste trabalho, textos que representam ideias de textos que não representam ideias). Além disso, gera uma margem em torno do limite de decisão que tem sua posição e largura controlada e pode ser visto como um parâmetro que é usado para ajustar a sensibilidade dos classificadores (Christensen, Liland et al., 2017).

Com sua boa capacidade de generalização, os SVMs eliminam a necessidade de seleção de recursos, tornando a aplicação da categorização do texto consideravelmente mais fácil. Outra vantagem dos SVMs sobre os métodos convencionais é a sua robustez. SVMs mostram bom desempenho em todos os experimentos, evitando falhas catastróficas como

observado em alguns métodos tradicionais em determinadas tarefas. Além disso, as SVMs não exigem nenhum parâmetro de ajuste, pois eles podem encontrar boas configurações de parâmetros automaticamente (Joachims, 1998).

Os SVMs tiveram sucesso significativo em várias tarefas de aprendizado de máquina levando-se em conta aplicações reais (Tong & Koller, 2002). A análise de desempenho dos SVMs reconhecem as propriedades particulares do texto: (a) espaços de características de alta dimensionalidade, (b) características irrelevantes e (c) vetores esparsos (Joachims, 1998).

Estudos como o de Coussement e Van den Poel (2008) mostra que os SVMs apresentam bom desempenho de generalização quando aplicados à mineração de dados com bases de dados grandes e com muitos ruídos. Já (Christensen, Norskov et al., 2017) usaram uma SVM para testar se um classificador de aprendizado de máquina desta natureza poderia aprender o padrão de ideias escritas como texto. A comparação entre desempenho no conjunto de validação e desempenho no conjunto de teste demonstra a confiabilidade do classificador e o método utilizado.

2.1.2 Naive Baiyes e suas aplicações na Mineração de Ideias

O Naive Bayes (NB) é um classificador bastante simples se comparado a outros classificadores, na medida em que pressupõe que todos os atributos dos exemplos são independentes uns dos outros, dado o contexto da classe. Embora essa suposição seja claramente falsa na maioria das tarefas do mundo real, o NB geralmente realiza classificações muito boas (Mccallum & Nigam, 1998a)

O NB é um método clássico bem conhecido que tem sido amplamente utilizado na categorização de textos (Joachims, 1997; Mccallum & Nigam, 1998b). A tarefa de aprendizado de um classificador NB é usar um conjunto de documentos de treinamento para estimar parâmetros e em seguida, usar estimativas para classificar novos documentos. O classificador NB simplifica muito o aprendizado e compete bem com classificadores mais sofisticados, sendo indicado para situações onde é necessário escolher entre duas condições distintas de uma classe de parâmetros, ou seja, escolhas binárias (Rish, 2001).

Além disso, o NB é um dos mais eficientes e eficazes algoritmos de aprendizado indutivo para *Machine Learning* e Mineração de Dados. Seu desempenho competitivo na classificação é surpreendente, porque a suposição inicial de independência condicional em que se baseia, normalmente é verdade em uma série de aplicações do mundo real (Zhang, 2004).

No estudo realizado por Kim, Hahn e Zhang (2000) foram utilizados classificadores NB em um contexto de filtragem de texto. Uma das principais vantagens deste método demonstradas no estudo foi a utilização de pesos num modelo probabilístico fornecendo então uma medida natural para o cálculo de índices de confiança.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa tem por objetivo gerar conhecimento para aplicações em mineração de ideias visando solucionar problemas específicos. Pode se definir pesquisa como sendo a inquisição, o procedimento sistemático e intensivo, que tem por objetivo descobrir e interpretar os fatos que estão inseridos em uma determinada realidade (Lehfeld & Barros, 1991).

O método científico utilizado segue o método indutivo, uma vez que se baseia em teorias e práticas encontradas em estudos já publicados, no entanto, suas conclusões são elaboradas a partir das constatações particulares dos autores (Gil, 2008).

A abordagem é quantitativa, pois os resultados da pesquisa podem ser quantificados, recorrendo à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis (Fonseca, 2002).

Quanto à natureza é uma pesquisa aplicada, pois objetiva gerar conhecimentos para aplicação prática e dirigidos à solução de problemas específicos (Silva & Menezes, 2005).

E por último, quanto aos objetivos, é exploratória por proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou construir hipóteses (Gil, 2002).

Este trabalho apresenta o resultado de um estudo prático, sendo que para isto foram definidas as seguintes etapas:

- 1) Obtenção de uma base de dados de ideias já classificada por especialistas, separadas por ideias e não ideias;
- 2) Submissão desta base a dois classificadores distintos para testar sua acurácia;
- 3) Definição de três critérios utilizados por especialistas em seleção de ideias baseados em pesquisas na literatura;
- 4) Anotação da base de dados de ideias de um rótulo onde apareçam os critérios definidos no item anterior. Cabe ressaltar que, para tornar o sistema o mais próximo possível da realidade, estes rótulos foram adicionados a toda a base de dados, incluindo aqueles textos que não representam ideias;

- 5) Nova submissão da base de dados anotada aos mesmos classificadores para testar a acurácia;
- 6) Comparação dos resultados entre a base de dados de ideias inicial com a base de ideias anotada com critérios utilizados por especialistas.

4 DESENVOLVIMENTO E ANÁLISES

Este trabalho se utiliza de uma base de dados contendo textos representando ideias e não ideias, que foram selecionadas por especialistas. Esta base de dados então é submetida a uma aplicação de *Machine Learning* que utiliza os algoritmos SVM e NB a fim de testar a capacidade destes de selecionar as ideias. Em seguida, a base de dados é manualmente marcada a partir de 3 critérios de especialistas identificados na literatura. É importante salientar que as marcações ocorreram tanto nas ideias quanto nas não ideias, para tornar o processo o mais semelhante possível com a realidade. Após isso, a base de dados foi novamente submetida a mesma aplicação e algoritmos tendo seus resultados comparados aos resultados anteriores.

A base de dados utilizada neste estudo contém ideias disponibilizadas publicamente através do Portal Sinapse da Inovação[®], que é um programa de incentivo ao empreendedorismo inovador que tem por objetivo “transformar e aplicar as boas ideias geradas por estudantes, pesquisadores, professores e profissionais dos diferentes setores do conhecimento e econômicos em negócios de sucesso” (Sinapse, 2017). Esta base se constitui de 200 textos, sendo 104 considerados ideias e 96 considerados não ideias compostos por textos de notícias em diferentes áreas.

A aplicação de *Machine Learning* utilizada foi o Lightside[®] por ser um aplicativo de código aberto destinado a tarefa de classificação aplicado à textos. Após a execução utilizando os algoritmos de SVM e NB, foram analisadas as matrizes de confusão e as métricas de avaliação de Acurácia e Kappa. A matriz de confusão, também conhecida como matriz de erro, representa um *layout* de tabela específico que permite a visualização do desempenho de um algoritmo, tipicamente um aprendizado supervisionado, mostrando as classificações corretas e incorretas. A acurácia é uma métrica simples informando o desempenho de um classificador ao usar esse recurso somente para prever rótulos de classe para seu conjunto de dados. Já a métrica Kappa mede o valor do recurso, variando de 0 a 1, para prever a classe na coluna anterior, em comparação com um valor aleatório. O valor de

Kappa de 1 significa que o característica é perfeitamente correlacionada com um rótulo (Mayfield & Rosé, 2012).

A atividade de seleção de ideias pode ser feita por especialistas em uma organização. Segundo Cooper (1985), um grupo pequeno multifuncional de gerentes pode se reunir periodicamente para analisar as ideias e avaliá-las por um sistema de pontuação, que determina critérios preestabelecidos para essa decisão. Se a ideia for rejeitada, o criador da ideia recebe um *feedback* justificando, a partir dos critérios, o porquê da não aceitação. Isso permite ao autor reformular a ideia para participar novamente do processo, garantindo um fluxo constante de novas ideias.

A partir de uma revisão da literatura, foram identificados vários critérios de especialistas que são levados em consideração em sua tarefa de selecionar ideias. Para a realização dos testes, foram escolhidos 3 destes critérios, listados e referenciados a seguir:

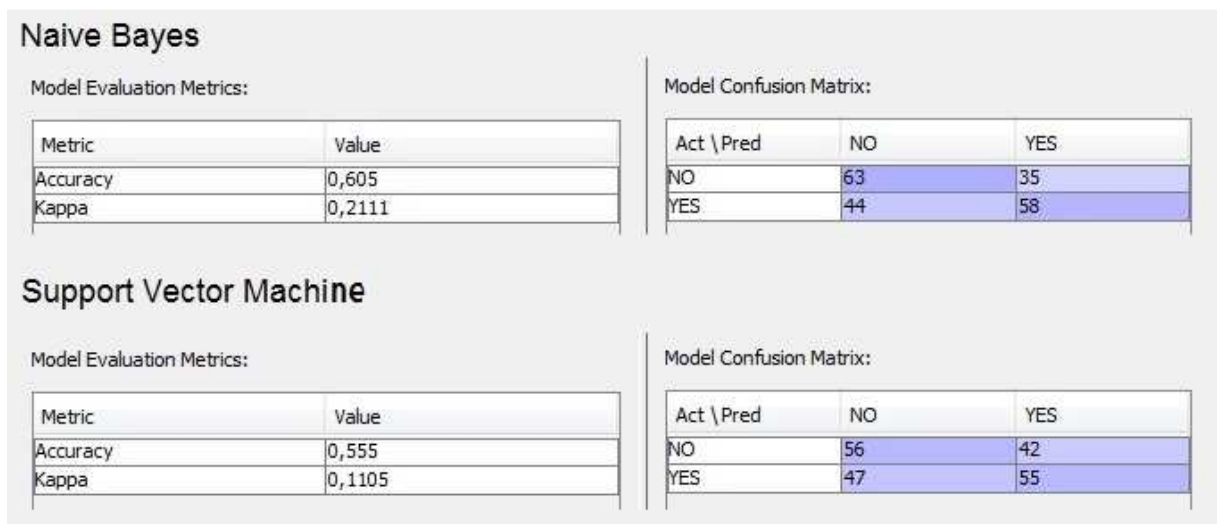
- Originalidade: representa a novidade e inovação, isto é, o quão incomuns e originais as ideias são a respeito do contexto em que se inserem (Magnusson, Wästlund, & Netz, 2016). O autor baseia-se em Amabile (1996), no qual defende que a originalidade é um conceito genérico e que as pessoas têm uma sensação intuitiva para o que é criativo;
- Produtibilidade: representa a perspectiva da empresa sobre a facilidade com que o serviço pode ser implementado e/ou produzido. Este critério assume a perspectiva da oferta (Magnusson et al., 2016);
- Viabilidade econômica: Para aprofundar o caráter econômico durante a avaliação da ideia, é essencial estimar o preço de venda e o benefício potencial. Também é importante avaliar se está de acordo com os objetivos da empresa e estimar os custos de produção e de desenvolvimento, bem como o tempo que vai levar para atingir o mercado (Ferioli, Roussel, Renaud, & Truchot, 2008; Ozer, 2004).

Considerando estes critérios a base de dados foi anotada utilizando os rótulos <ideia_encontrada> e </ideia_encontrada> para a início e fim da ideia, respectivamente. Um exemplo de uma ideia encontrada a partir do critério “originalidade” é: “a automação e sistema do cultivo em abrigos protegidos a baixo custo, escalável e *plug-and-play*, composto de <ideia_encontrada> atuadores e sensores para monitoramento microclimático, *software* inteligente e aplicativo mobile, possibilitando o controle automático de irrigação à distância via Wi-Fi sem internet nos mais diversos cultivos e obtendo os dados em tempo real, </ideia_encontrada> indicando precisamente quando as plantas precisam de você, levando a uma cultura mais saudável e com maiores rendimentos”. Considerando o critério

“produtibilidade” tem-se o exemplo: “os resultados apontaram <ideia_encontrada> a crescente demanda do mercado por alimentos processados saudáveis, exóticos, com forte apelo visual e com reduzidos teores de açúcares e sódio, torna a estruturação de frutos uma atividade promissora<ideia_encontrada>. E, por fim, um exemplo do critério “viabilidade econômica”: “a proposta é aumentar a disponibilidade de dados, podendo assim, gerar informações mais precisas para o manejo dos cultivos, <ideia_encontrada> redução de custos de produção <ideia_encontrada> e otimização do consumo dos insumos agrícolas”.

Os resultados da análise inicial onde a base de dados de ideias sem anotação foi submetida ao aplicativo Lightside® são demonstrados na figura 1. Pode-se observar que o NB apresentou uma acurácia de 0,605, enquanto o SVM apresentou uma acurácia de 0,555. Os resultados são considerados baixos para um classificador de texto.

Figura 1 – Análise inicial da base de dados de ideias



Fonte: Autor

Em uma segunda análise a base de ideias foi anotada utilizando os critérios identificados neste estudo e submetida à aplicação Lightside®. Os resultados são demonstrados na figura 2. Pode-se observar que o NB apresentou uma acurácia de 0,825 e o SVM uma acurácia de 0,855. A utilização dos critérios de especialistas melhorou seu índice de acurácia em 26,66% para o NB e 36,36% para o SVM.

Figura 2 – Análise da base de dados de ideias após ser submetida aos critérios de especialistas

Naive Bayes		Model Confusion Matrix:		
Model Evaluation Metrics:		Act \ Pred	NO	YES
Metric	Value	NO	82	16
Accuracy	0,825	YES	19	83
Kappa	0,6501			

Support Vector Machine		Model Confusion Matrix:		
Model Evaluation Metrics:		Act \ Pred	NO	YES
Metric	Value	NO	77	21
Accuracy	0,855	YES	8	94
Kappa	0,7091			

Fonte: Autor

5 CONSIDERAÇÕES FINAIS

Este trabalho verificou através de testes práticos que a utilização de critérios de especialistas em seleção de ideias em determinado domínio melhora a acurácia dos métodos automáticos de mineração de ideias. Ao submeter uma base de dados anotada utilizando os rótulos (que correspondem aos critérios), os métodos *Naive Bayes* e *Support Vector Machine* apresentaram bons resultados na mineração de ideias, melhorando seu índice de acurácia em 26,66% e 36,36% respectivamente.

A limitação deste trabalho está na utilização de apenas uma base de dados para a realização de testes práticos, podendo o mesmo ser ampliado para testes em outras bases de dados. Com isso poderia ser obtida uma avaliação estatística do método permitindo o aperfeiçoamento dos resultados.

Como trabalhos futuros sugere-se a utilização de outras técnicas de mineração de ideias e a inserção de novos critérios de especialistas. Outra possibilidade é a criação de métodos automáticos suportados por estruturas de representação de conhecimento, como ontologias, para anotar os critérios dos especialistas em textos que representem potenciais ideias.

AGRADECIMENTOS

Este trabalho recebeu o apoio financeiro parcial da CAPES – Brasil.

REFERÊNCIAS

- Alksher, M. A., Azman, A., Yaakob, R., Kadir, R. A., Mohamed, A., & Alshari, E. (2017). A Framework for Idea Mining Evaluation. *16th International Conference on New Trends in Intelligent Software Methodology Tools, and Techniques, SoMeT 2017*.
- Amabile, T. M. (1996). *Creativity in context: The social psychology of creativity*. Boulder, CO: Westview. Recuperado de <https://scholar.google.fr/scholar?hl=fr&q=creativity+in+context+amabile&btnG=&lr=#2>
- Christensen, K., Liland, K. H., Kvaal, K., Risvik, E., Biancolillo, A., Scholderer, J., Nørskov, S., et al. (2017). Mining online community data: The nature of ideas in online communities. *Food Quality and Preference*, 62(December 2016), 246–256. Elsevier. Recuperado de <http://dx.doi.org/10.1016/j.foodqual.2017.06.001>
- Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In Search of New Product Ideas: Identifying Ideas in Online Communities by Machine Learning and Text Mining. *Creativity and Innovation Management*, 26(1), 17–30.
- Cooper, R. G. (1985). Selecting winning new product projects: Using the NewProd system. *The Journal of Product Innovation Management*, 2(1), 34–44.
- Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327.
- Felden, C., Bock, H., Gräning, A., Molotowa, L., & Saat, J. (2006). Evaluation von Algorithmen zur Textklassifikation. *Freiberg Working Papers*, 10.
- Ferioli, M., Roussel, B., Renaud, J., & Truchot, P. (2008). Evaluation of the Potential Performance of Innovative Concepts in the Early Stages of the New-Product Development Process (Npdp). *Design*, 1139–1148.
- Fonseca, J. J. S. da. (2002). *Metodologia da pesquisa científica*. São Carlos: Serviço de Biblioteca e Informação Recuperado de http://leg.ufpi.br/subsiteFiles/lapnex/arquivos/files/Apostila_-_METODOLOGIA_DA_PESQUISA%281%29.pdf
- Gil, Antonio Carlos. (2002). *Como elaborar projetos de pesquisa*. (S. Paulo, Org.) (5^o ed).
- Gil, António Carlos. (2008). *Métodos e técnicas de pesquisa social*. (Ediitora Atlas S.A., Org.).
- Joachims, T. (1997). A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. *Fourteenth International Conference on Machine Learning*, 143–151. Recuperado de <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA307731>
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1398, 137–142.
- Kim, Y.-H., Hahn, S.-Y., & Zhang, B.-T. (2000). Text filtering by boosting naive Bayes classifiers. *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. ACM*, 168–175.

- Kishida, K. (2005). Technical issues of cross-language information retrieval: A review. *Information Processing and Management*, 41(3), 433–455.
- Klein, M., & Garcia, A. C. B. (2015). High-speed idea filtering with the bag of lemons. *Decision Support Systems*, 78, 39–50. Elsevier B.V. Recuperado de <http://dx.doi.org/10.1016/j.dss.2015.06.005>
- Lee, A. V. Y., & Tan, S. C. (2017a). Discovering Dynamics of an Idea Pipeline: Understanding Idea Development within a Knowledge Building Discourse, 119–128.
- Lee, A. V. Y., & Tan, S. C. (2017b). Promising ideas for collective advancement of communal knowledge using temporal analytics and cluster analysis, 4(2107), 76–101.
- Lehfeld, N., & Barros, A. de J. P. (1991). *Projeto de pesquisa: propostas metodológicas*. (Vozes, Org.). Rio de Janeiro.
- Li, X., Li, L., & Chen, Z. (2014). Toward Extenics-Based Innovation Model on Intelligent Knowledge Management. *Annals of Data Science*, 1(1), 127–148. Recuperado de <http://link.springer.com/10.1007/s40745-014-0009-5>
- Magnusson, P. R., Wästlund, E., & Netz, J. (2016). Exploring Users' Appropriateness as a Proxy for Experts When Screening New Product/Service Ideas. *Journal of Product Innovation Management*, 33(1), 4–18.
- Mayfield, E., & Rosé, C. P. (2012). *LightSIDE: Text Mining and Machine Learning User 's Manual*. (C. M. University, Org.).
- Mccallum, A., & Nigam, K. (1998a). A Comparison of Event Models for Naive Bayes Text Classification.
- Mccallum, A., & Nigam, K. (1998b). Employing EM and Pool-Based Active Learning for Text Classification 2 Probabilistic Framework for Text Classification.
- Mousavi, S. M., Torabi, S. A., & Tavakkoli-Moghaddam, R. (2013). A Hierarchical Group Decision-Making Approach for New Product Selection in a Fuzzy Environment. *Arabian Journal for Science and Engineering*, 38(11), 3233–3248.
- Nonaka, Ikujiro; Takeuchi, H. (2000). The best Japanese companies offer aguide to the organizational roles, structures, and practices that produce consinuous innovarian. *The Knowledge-Creating Company* (Vol. 103, p. 411).
- Ozer, M. (2004). Managing the selection process for new product ideas. *Research Technology Management*, 47(4), 11.
- Paukkeri, M. (2009). Framework for Analyzing and Clustering Short Message, (September), 239–247.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3, 41–46.
- Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2), 206–226. Recuperado de <http://ieeexplore.ieee.org/document/5389202/>
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34.
- Sérgio, M. C., Souza, J. A., & Gonçalves, A. L. (2017). Idea Identification Model to Support Decision Making. *IEEE Latin America Transactions*, 15, 968–973.

- Silva, E. L. da, & Menezes, E. M. (2005). *Metodologia da Pesquisa e Elaboração de Dissertação* (4^o ed). Florianópolis: UFSC.
- Sinapse. (2017). Sinapse da Inovação. Recuperado junho 5, 2019, de <http://sc.sinapsedainovacao.com.br/>
- Thorleuchter, D., & Van den Poel, D. (2016). Identification of interdisciplinary ideas. *Information Processing and Management*, 52(6), 1074–1085. Elsevier Ltd. Recuperado de <http://dx.doi.org/10.1016/j.ipm.2016.04.010>
- Thorleuchter, D., & Van Den Poel, D. (2013). Web mining based extraction of problem solution ideas. *Expert Systems with Applications*, 40(10), 3961–3969. Recuperado de <http://dx.doi.org/10.1016/j.eswa.2013.01.013>
- Thorleuchter, Dirk, Herberz, S., & Van Den Poel, D. (2011). Mining social behavior ideas of Przewalski horses. *Lecture Notes in Electrical Engineering*, 121 LNEE, 649–656.
- Thorleuchter, Dirk, & Van Den Poel, D. (2012). Extraction of ideas from microsystems technology. *Advances in Intelligent and Soft Computing*, 168 AISC(VOL. 1), 563–568.
- Thorleuchter, Dirk, Van Den Poel, D., & Prinzie, A. (2010). Mining ideas from textual information. *Expert Systems with Applications*, 37(10), 7182–7188. Elsevier Ltd. Recuperado de <http://dx.doi.org/10.1016/j.eswa.2010.04.013>
- Tong, S., & Koller, D. (2002). Journal of Machine Learning Research 2001 Tong. *Journal of Machine Learning Research*, 1–22. Recuperado de <http://www.jmlr.org/papers/volume2/tong01a/tong01a.pdf>0Apapers://5e3e5e59-48a2-47c1-b6b1-a778137d3ec1/Paper/p1818
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- Zhang, H. (2004). The Optimality of Naive Bayes. *AA 1.2*.