

## **Mineração da percepção do aluno: Mineração de Dados Educacionais na geração de indicadores para a avaliação de práticas pedagógicas docentes**

**Fábio de Paula Santos<sup>1</sup>, Ismar Frango Silveira<sup>2</sup>, Carla Pineda Lechugo<sup>3</sup>**

<sup>1</sup>Instituto Federal de São Paulo (IFSP) e Centro Paula Souza, Salto-SP-Brazil

<sup>2</sup>Programa de Pós-Graduação em Eng. Elétrica e Computação - Univ. Presbiteriana Mackenzie - São Paulo-SP-Brazil

<sup>3</sup>Centro Paula Souza- Fatec Sorocaba- Sorocaba - SP- Brazil

fabiopsan@gmail.com, ismarfrango@gmail.com, carlalechugo@gmail.com

**Abstract.** *This paper proposes a conceptual model applying of Educational Data Mining techniques (MDE) with emphasis on Sentiment Analysis in an Institutional Evaluation to discover the perception of students about the teaching practices and allow the creation of new indicators and increasing the range of evaluation dimensions. The purpose to listen students is to understand their expectations, limitations and frustrations and, from that understanding, assign to the Institution and the teacher her role to ensure learning. The use of MDE in this discovery process allows a decision-making more accurate and faster.*

**Resumo.** *Este artigo apresenta um modelo computacional conceitual da aplicação de técnicas de Mineração de Dados Educacionais (MDE), com ênfase na Análise de Sentimentos, em uma Avaliação Institucional Docente, de forma a levantar a percepção dos alunos acerca das práticas docentes e permitir a criação de novos indicadores e desta forma, aumentar o leque de dimensões utilizadas para tal avaliação. Busca-se, ao ouvir os alunos, compreender suas expectativas, limitações e frustrações e, a partir deste entendimento, atribuir à Instituição e ao professor o papel de ajustar o que é possível e adequado para garantir o aprendizado. O uso da MDE neste processo permite uma tomada de decisão mais precisa e mais rápida.*

### **1. Introdução**

Com a evolução tecnológica e, mais precisamente, das atividades de Mineração de Dados Educacionais (MDE) abriu-se as portas para a análise de resultados de uma Avaliação Institucional Docente elaborada somente com perguntas abertas, tais como “Fale Bem ou “Reclame do professor” e desta forma, emergir visões das práticas pedagógicas vindas dos próprios alunos sobre as práticas pedagógicas docentes. Ora, se falamos em “alunos como coautores de um ambiente de aprendizagem” [Johnson et al. 2014], não seria óbvio ouvi-los?

Ao ouvi-los, também é possível melhorar os índices de aprendizagem e, em última instância, contribuir para a redução da evasão discente no Ensino Superior. Alunos que, durante o processo de ensino-aprendizagem vivenciaram situações das quais não se

sentiram satisfeitos (ou frustrados) são os que potencialmente desistem da escola, do professor e também de aprender [Lechugo 2016]. Ao aplicar um modelo computacional que se utiliza das técnicas de MDE, pretende-se fugir de pressupostos dos avaliadores, ou seja, fazer emergir novas categorias de análise e entender como o aluno interage com elas. Assim, este trabalho propõe uma diferente direção das avaliações institucionais, apontando para um modelo centrado no aluno para criar novos indicadores de práticas pedagógicas docentes.

Este trabalho está organizado desta forma: no capítulo 2, discute-se os conceitos de mineração de dados, de textos e de mineração de dados educacionais. O capítulo 3 apresenta uma introdução sobre a Análise de Sentimentos. Os capítulos 4 e 5 apresentam um modelo computacional e os resultados iniciais, e por fim, algumas considerações sobre a importância dos estudos da aplicação da MDE nas avaliações institucionais.

## **2. Mineração de dados, de textos e de dados educacionais**

Dados são símbolos ou signos não estruturados, sem significado, como valores numéricos depositados em um repositório; informação é o que está contida em uma descrição e que agrega significado e utilidade a estes dados, já o conhecimento é algo que permite uma tomada de decisão, e finalmente a Mineração de Dados é a ferramenta utilizada para este fim [Castro 2016].

Segundo Silva (2015), a Mineração de Dados pode ser definida como um processo automático ou semiautomático para explorar analiticamente grandes bases de dados, com a finalidade de descobrir padrões relevantes e importantes para embasar o conhecimento do usuário final (geralmente tomadores de decisão). Denominadas também de técnicas de *Data Science* ou de *Data Mining*, elas procuram encontrar padrões inesperados nos dados (padrões não óbvios) e desta forma, ir além do que é conhecido pelo usuário.

A Mineração de Textos é uma subárea da Mineração de Dados que procura descobrir, de forma automática, informações (padrões e anomalias) em textos [Webber, Cristofoli e Lima 2013].

Como uma especialização da Mineração de Dados, a Mineração de Dados Educacionais (MDE) se apresenta como um campo emergente que aplica técnicas de Mineração de Dados em informações obtidas de ambientes educacionais a fim de descobrir padrões ou evidências científicas sobre alunos e também sobre as formas de aprendizagem e que permite, dentre outras atividades, compreender de forma mais eficaz e adequada os alunos, como eles aprendem e o papel do contexto no qual esta aprendizagem ocorre [Baker, Isotani e Carvalho 2011].

É possível minerar dados de alunos para verificar a relação entre uma abordagem pedagógica e o aprendizado do aluno. Através desta informação o professor poderia compreender se sua abordagem realmente está ajudando o aluno e desenvolver novos métodos de ensino mais eficazes. Para alcançar tais objetivos, as ferramentas de Mineração de Dados aplicam algoritmos capazes de extrair conhecimentos dos dados através da análise descritiva (com uso de ferramentas para medir, explorar e descrever características intrínsecas aos dados), agrupamento, predição, associação e detecção de anomalias [Castro 2016].

O aumento do volume de dados educacionais em diferentes e heterogêneas origens de dados levou a necessidade de pesquisas com o foco e objetivos específicos ao

ambiente educacional, que foram elencados por Jindal e Borah (2013) e divididos em objetivos acadêmicos e administrativos. Os objetivos acadêmicos da MDE envolvem informações orientadas para pessoas (estudantes, professores), orientada para os departamentos e instituições (novos cursos, adequação às novas demandas) e orientadas ao domínio (relativo a uma determinada área de atuação). Os objetivos administrativos são relacionados à gestão, como por exemplo, ações de infraestrutura escolar.

O trabalho de Hegazi e Abugroon (2016) também faz um agrupamento das áreas mais relevantes da MDE dentro das Instituições de Ensino Superior (IES): pesquisas na área de gerenciamento de cursos (gestão de cursos e como melhorar o apoio ao aluno), pesquisas sobre o comportamento do aluno (prever a resposta do aluno para determinadas atividades propostas), sistemas de apoio à decisão (análise do processo educacional, tais como matrículas e seleção de cursos) e aplicações para evitar a evasão e reter o aluno, tal como prever qual aluno irá abandonar um curso ou prever quanto tempo um aluno irá retornar ao curso após trancá-lo [Hegazi e Abugroon 2016].

Os métodos de MDE podem ser divididos em dois grupos: os orientados para verificação (estatísticas, teste de hipótese, análise de variância, etc.) e os orientados à descoberta (predições, classificação, agrupamento), no entanto, segundo Baker (2010), as técnicas diferem frequentemente das técnicas da mineração de dados tradicional, pois devem explorar níveis de hierarquia e organização dos dados educacionais. A figura 1 abaixo apresenta uma taxonomia das principais subáreas de pesquisa em MDE.

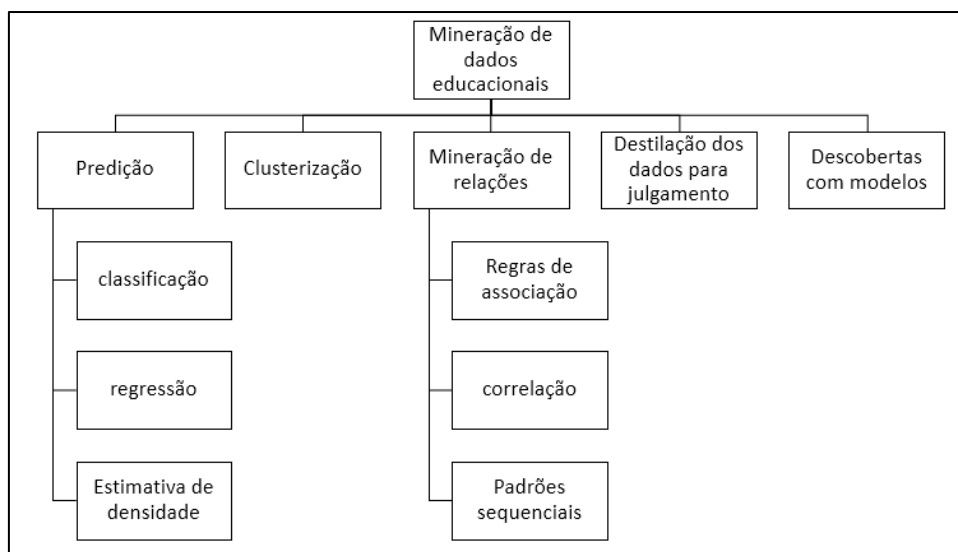


Figura 1 - Taxonomia das principais subáreas de pesquisa em MDE (Baker, 2010).

As três primeiras categorias são de interesse tanto da área de MDE quanto da área de mineração de dados em geral. Segundo Baker *et al.* (2011), a Destilação dos Dados pode ser muito útil para analisar a curva de aprendizagem dos alunos, já uma aplicação que usa Descobertas com Modelos, envolve duas etapas principais, um modelo de constructo que é desenvolvido usando (por exemplo) Aprendizado de Máquina, e outra etapa que é a sua validação [Hershkovitz *et al.* 2013].

Em uma avaliação institucional, as respostas dos alunos são obtidas através de formulários impressos ou digitais. Dentre as técnicas aplicáveis, destaca-se a classificação

de textos, que busca rotular um documento de acordo com suas características. Dentro do contexto da mineração de textos, este modelo propõe o uso da Análise de Sentimentos.

### 3. Análise de Sentimentos

Em termos da compreensão da linguagem natural, a Análise de Sentimentos é considerada como uma importante subárea da Análise Semântica com o objetivo de identificar não somente os assuntos que as pessoas falam e seu sentimento sobre esses assuntos, mas também se propõe a identificar características de um texto de acordo com um assunto específico, analisando e classificando-os em subgrupos pré-definidos, como sentimentos e emoções (feliz, triste), polaridade (positivo, negativo) ou em qualquer outro contexto previamente definido [Liu 2015].

As técnicas de Análise de Sentimentos podem ser resumidas na taxonomia da Figura 2 [Medhat, Hassan e Korashy 2014].

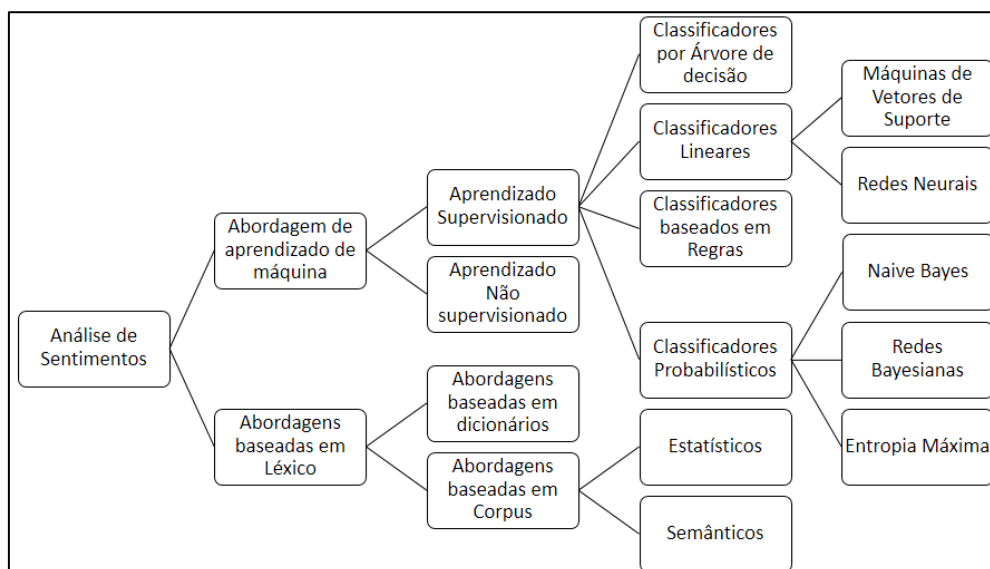


Figura 2 - Técnicas de Análise de Sentimentos. Adaptado de [Medhat, Hassan e Korashy 2014].

A Análise de Sentimentos pode ser realizada através de diversas abordagens e que podem ser subdivididas em duas grandes áreas: técnicas de Aprendizado de Máquina e das abordagens baseadas em Léxico (Figura 2).

A área de pesquisa chamada de Aprendizado de Máquina visa desenvolver programas ou modelos computacionais capazes de - automaticamente - melhorar seu desempenho por meio de experiências. Modelos de Aprendizado de Máquina podem ser supervisionados ou não-supervisionados. No primeiro modelo, os algoritmos ajustam parâmetros de um modelo a partir do erro medido entre respostas obtidas e esperadas. No segundo, os parâmetros de um modelo são ajustados com base na maximização de medidas de qualidade das respostas obtidas [Brunialti *et al.* 2015], ou seja, o modelo deve aprender a rotular os dados.

Nos métodos baseados em léxico é criado um dicionário de palavras com sua orientação semântica conhecida (positiva, negativa ou neutra sobre o assunto) para realizar a classificação do documento, opinião ou frase. Nos métodos baseados em corpus,

faz-se uso de técnicas sintáticas ou estatísticas, como a verificação da ocorrência concomitante de palavras com outras que possuam sua orientação semântica conhecida.

A Análise de Sentimentos estuda principalmente as opiniões que expressam ou implicam sentimento positivo ou negativo. Opinião é um conceito abrangente que cobre sentimentos, avaliação, estima, ou ainda, a atitude e as informações associadas a um tópico ou uma opinião alvo e também à pessoa que detém uma determinada opinião. Para a Análise de Sentimentos, o termo “sentimento” significa apenas o sentimento positivo ou negativo subjacente implícito nesta opinião[Liu 2015].

Liu (2015) apresenta uma simplificação matemática da estrutura de dados para a Análise de Sentimentos através da seguinte quintupla:  $(e_j, a_{jk}, s_{ijkl}, h_i, t_l)$  onde:

- $e_j$  representa a entidade que está sendo analisada (um político, uma empresa, um celular);
- $a_{jk}$ , é o aspecto em questão (confiança, pontualidade, beleza);
- $s_{ijkl}$  representa o sentimento da entidade em relação à questão, que pode assumir positivo (+), negativo (-) ou neutro;
- $h_i$  permite classifica o detentor da opinião (cliente, crítico);
- $t_l$  armazena a data da pesquisa de opinião.

Trata-se de uma estrutura de dados essencial para a Análise de Sentimentos e a falta de um destes elementos pode tornar a análise problemática. A ausência de  $t_l$  (data) pode causar uma falha na análise pois, as opiniões mudam ao longo do tempo. A falta de  $h_i$  (detentor) pode fazer muita diferença na análise de um discurso (esta opinião é do presidente dos EUA ou do presidente do Congresso Nacional Brasileiro?).

Nos últimos anos, surgiu um considerável interesse no desenvolvimento de iniciativas de ensino e aprendizagem para aprimorar a retenção e progressão dos alunos [Glenda *et al apud* Novak e Cowling 2011] e tais iniciativas passam pelo interesse em saber o que o aluno pensa destas práticas. O trabalho de Altrabsheh, Cocea e Fallahkhair (2014) descreve um modelo de um sistema em tempo real para analisar o *feedback* de alunos durante aulas e em seguida apresentar seus resultados ao professor usando Análise de Sentimentos. Também realiza um comparativo entre as diferentes técnicas de aprendizado de máquina, diferentes combinações de n-gramas, níveis de pré-processamento dos dados e do uso da classe neutra.

Na Finlândia, país reconhecido pelas suas inovações no campo educacional, os professores são encorajados a encontrar novas soluções para melhorar o processo ensino-aprendizagem. O trabalho de Munzero *et al.* (2013) aplica técnicas de Análise de Sentimentos nos Diário de Aprendizagem dos alunos. Neste trabalho é possível observar que os obstáculos emocionais são conhecidos por atrapalharem o processo de aprendizagem: os alunos aprendem mais quando se sentem bem e confortáveis com determinada disciplina ou assunto e estes diários ajudam nesta direção.

#### 4. Desenvolvimento do Modelo

Na tentativa de ampliar o leque de dimensões a serem consideradas por uma avaliação institucional, em especial, a avaliação docente e mais precisamente, das práticas pedagógicas dos professores, este artigo propõe um modelo computacional para utilizar a Mineração de Dados Educacionais, em especial, a Análise de Sentimentos para obter

novas (ou confirmar as já existentes) dimensões de análise das práticas docentes a partir das percepções dos alunos.

Neste modelo, a captação dos dados é realizada por um formulário que é apresentado de forma aberta através de dois quadros onde o aluno deve falar “bem” ou “reclamar” de seus professores, e, onde as questões norteadoras do processo de avaliação deverão emergir das próprias respostas e opiniões dos alunos (Figura 3).

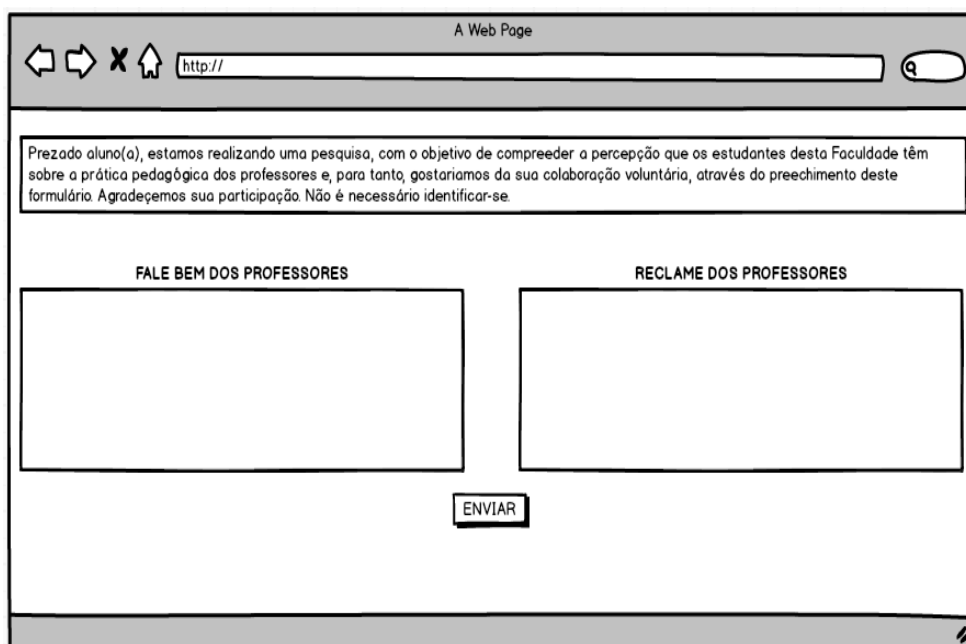


Figura 3 - Protótipo da interface do formulário. Baseado em [Lechugo 2016].

Apesar do formulário destacar que as opiniões positivas devem ser preenchidas no quadro da esquerda e as opiniões negativas devem ser expressas no quadro da direita, sabe-se que nem sempre isso poderá ocorrer (Figura 4).

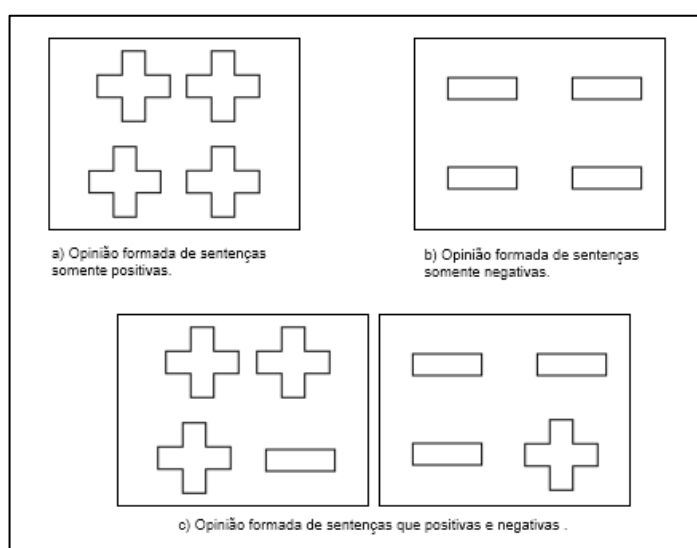


Figura 4 - Esquema das possíveis respostas de alunos. Fonte: autoria própria.

Espera-se, na entrada de dados “fale bem dos professores” encontrar a situação “a” da Figura 4 e na entrada de dados “reclame dos professores” a situação da figura “4b”. No entanto, por tratar-se de um texto aberto, onde o aluno poderá emitir qualquer opinião em qualquer quadro, apesar de saber-se da delimitação positiva e negativa de cada entrada de dados, será realizada a análise de cada sentença, mas será considerada a opinião do documento, ou seja, a opinião positiva, negativa ou neutra da composição de todas as sentenças da entrada de dados. Como requisito funcional de relevância deste formulário, pode-se destacar a livre entrada de dados, onde o aluno pode escrever o que desejar e sem nenhuma interferência do avaliador.

A partir das respostas obtidas através da interface proposta, inicia-se o processo de Mineração destes dados, com o objetivo de identificar categorias de análise que possam “saltar” das afirmações dos alunos, das práticas apontadas como mais relevantes nas práticas docentes e também qual a orientação semântica das respostas dos alunos em relação ao seu corpo docente, ou seja, se elas são positivas ou negativas.

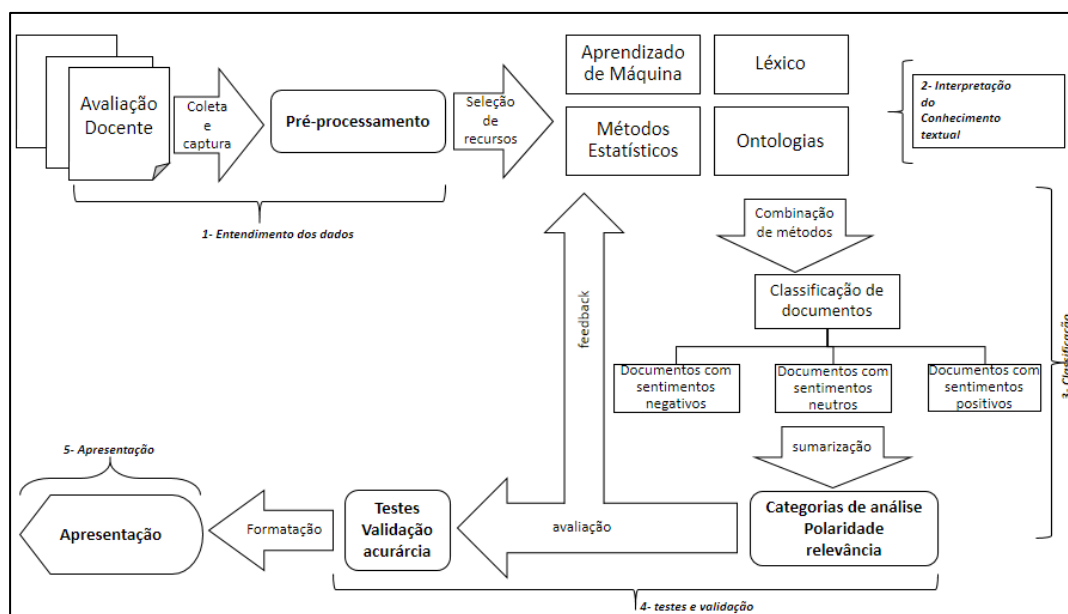


Figura 5 - Modelo Computacional para aplicação de Mineração de Dados Educacionais para Avaliação Institucional Docente. Fonte: Autoria própria.

No modelo proposto na Figura 5, pode-se identificar cinco etapas principais:

- i. Etapa de entendimento dos dados: Extraídos das ferramentas de coleta de dados (em formulários impressos ou eletrônicos), as opiniões dos alunos passam por algumas tarefas de pré-processamento e de limpeza de dados. Dentre as tarefas desta fase, podemos citar as atividades de tokenização, remoção de palavras não discriminantes, stemização, entre outras.
- ii. Etapa do conhecimento textual: Nesta fase os *tokens* obtidos no pré-processamento serão transformados em vetores de entrada para os algoritmos de Análise de Sentimento e diversas abordagens serão utilizadas: aprendizado de máquina, léxicos ou dicionários, ontologias e abordagens estatísticas.
- iii. Classificação: As opiniões dos alunos serão agrupadas por categorias de análise, que serão levantadas através de métodos estatísticos, o que permitirá o surgimento

de novas categorias de análise das práticas docentes que não são contempladas sob a ótica atual (a do avaliador). A sumarização será responsável por identificar as categorias mais abordadas pelos alunos.

- iv. Testes e validação: Onde será definida a acurácia do modelo. Não atingido um certo nível de acurácia, o processo é refeito.
- v. Apresentação: Finalmente, se os dados de resultados atingirem o nível de acurácia satisfatório, eles serão exibidos da forma mais amigável através de ferramentas de visualização de dados.

## 5. Discussão dos resultados iniciais

Esta pesquisa ainda está em desenvolvimento e, neste momento, só é possível descrever a análise “manual”. Para apurar a acurácia deste modelo computacional, os resultados obtidos serão confrontados com os resultados encontrados em uma pesquisa de avaliação institucional docente realizada em uma Instituição de Ensino Superior Tecnológico [Lechugo 2016], onde não foram usados modelos estatísticos e de Mineração de Dados Educacionais. Nesta pesquisa, após serem coletados os formulários preenchidos manualmente pelos alunos, iniciou-se um processo de imersão na leitura dos mesmos, a fim de compreender as percepções dos alunos acerca dos professores. Após inúmeras leituras, as categorias de análise emergiram, ou seja, foram definidas a *posteriori* [Franco 2005]. As categorias identificadas nestas leituras foram: *i*) relacionamento professor-aluno, *ii*) conhecimento sobre os conteúdos de ensino, *iii*) capacidade de ensinar e *iv*) motivação do professor.

É importante ressaltar que as categorias elencadas acima, representam as características que definem a profissão docente pelo olhar do aluno. O fato do formulário estar em branco, e não haver previamente definidas as categorias para a análise do aluno, permitiu que ele falasse bem e reclamasse do professor sob qualquer aspecto, mas todos eles avaliaram a figura do professor pelos mesmos critérios, ou seja, existe uma expectativa prévia que define a profissão docente e pela qual ela será julgada.

Através de uma análise comparativa, espera-se que novas categorias de análise das práticas pedagógicas possam surgir a partir do modelo computacional proposto, assim como, de forma automática, as mesmas categorias que foram levantadas de forma “manual” sejam identificadas, corroborando com a utilidade deste formato de Avaliação Docente. Sendo assim, pretende-se ampliar os indicadores de avaliação da qualidade das IES, fomentando a reflexão e as discussões sobre as práticas docentes mais eficientes (ou não) no processo de aprendizagem do aluno.

## 6. Considerações Finais

Evitar a evasão no Ensino Superior é um grande desafio das IES e existem poucos instrumentos que detectem formalmente as razões pelas quais o aluno abandona a escola. Uma avaliação institucional, por participar ativamente da vida da instituição, é detentora de uma visão privilegiada e pode contribuir com os procedimentos acadêmicos, sendo um instrumento de correção de metas e objetivos [Santos Baggi e Lopes 2011]. Neste momento, a identificação e a organização destes dados com relação às satisfações e às insatisfações dos alunos se tornam tão necessária. Além disso, o gerenciamento destes dados de forma *on-line*, devem contribuir para que professores e gestores tomem decisões rápidas e efetivas para corrigir os erros do processo.



Através deste trabalho, espera-se divulgar o uso de técnicas de mineração de dados educacionais, em especial a aplicação da Análise de Sentimentos nas Avaliações Institucionais Docentes. Com isso, busca-se demonstrar a riqueza das análises não-amostrais que, a partir de todas as respostas dos alunos, e sem nenhum formato pré-concebido por parte do avaliador, torna possível identificar possíveis práticas pedagógicas que podem ser consideradas boas e outras nem tanto, sob a ótica dos alunos. Também, pretende-se contribuir com possíveis novas categorias de análises importantes para a avaliação do trabalho docente. É importante ressaltar que, as experiências positivas e negativas reportadas pelos alunos, não retiram do professor e da instituição a autoridade e responsabilidade da análise e tomada de decisão, até porque o objetivo da MDE em uma Avaliação Institucional Docente, não é “entender” os alunos para “atendê-los”, mas sim, de compreender a realidade que temos, para refletirmos sobre que alunos, que professores e, finalmente, que instituição queremos atuando na Educação Superior.

## Referências

- ALTRABSHEH, N.; COCEA, M.; FALLAHKHAIR, S. (2014) “Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback”. 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, p. 419–423.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. (2011) “Mineração de Dados Educacionais: Oportunidades para o Brasil”. Revista Brasileira de Informática na Educação, v. 19, n. 2, p. 3.
- BAKER, R. S. J. D. (2010) “Data mining for education”. International Encyclopedia of Education, v. 7, p. 112–118.
- BRUNIALTI, L. F. et al. (2015) “Aprendizado de máquina em sistemas de recomendação baseados em conteúdo textual uma revisão sistemática”. XI Brazilian Symposium on Information System. Goiania, GO: 2015 Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2015/029.pdf>>. Acesso em: 5 dez. 2015
- CASTRO, L. N. DE. (2016) “Introdução à Mineração de dados”. São Paulo: Saraiva.
- FRANCO, M. L. P. B. (2005) “Análise de conteúdo”. Brasília: Liber Livro Editora.
- HEGAZI, M. O.; ABUGROON, M. A. (2016) “The State of the Art on Educational Data Mining in Higher Education”. International Journal of Computer Trends and Technology, v. 31, n. 1, p. 46–56.
- HERSHKOVITZ, A. et al. (2013) “Discovery with models: A case study on carelessness in computer-based science inquiry”. American Behavioral Scientist, v. 57, n. 10, p. 1480–1499.
- JINDAL, R.; BORAH, M. D. (2013) “A Survey on Educational Data Mining and Research Trends”. International Journal of Database Management Systems (IJDBMS), v. 5, n. 3, p. 53–73.
- JOHNSON, L. et al. (2014) “Horizon Report Europe: 2014 Schools Edition”. Brussels, Belgium: [s.n.]. Disponível em: <[https://ec.europa.eu/jrc/sites/default/files/2014-nmc-horizon-report-eu-en\\_online.pdf](https://ec.europa.eu/jrc/sites/default/files/2014-nmc-horizon-report-eu-en_online.pdf)>.
- LECHUGO, C. P. (2016) “A Educação Tecnológica: sua história, seus professores e a

- percepção dos alunos sobre as práticas pedagógicas”. [s.l.] Universidade de Sorocaba.
- LIU, B. (2015) “Sentiment Analysis”. New York City, USA: Cambridge University Press.
- MEDHAT, W.; HASSAN, A.; KORASHY, H. (2014) “Sentiment analysis algorithms and applications: A survey”. *Ain Shams Engineering Journal*, v. 5, n. 4, p. 1093–1113.
- MUNEZERO, M. et al. (2013) “Exploiting sentiment analysis to track emotions in students’ learning diaries”. *Proceedings of the 13th Koli Calling International Conference on Computing Education Research - Koli Calling ’13*, p. 145–152.
- NOVAK, J.; COWLING, M. (2011) “The implementation of social networking as a tool for improving student participation in the classroom”. *ISANA International Academy Association Conference*, p. 1–10.
- SANTOS BAGGI, C. A. DOS; LOPES, D. A. (2011) “Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica”. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, v. 16, n. 2, p. 355–374.
- SILVA, L. A. (2015) “Mineração de dados: uma abordagem introdutória e ilustrada”. São Paulo: Editora Mackenzie.
- WEBBER, C. G.; CRISTOFOLI, L. G.; LIMA, M. DE F. W. DO P. (2013) “UnderMine Text Miner – Uma Ferramenta de Mineração de Texto para Área Educacional”. *Novas Tecnologias na Educação*, v. 11, n. 1.