MINERAÇÃO DE DADOS: UMA ANÁLISE SOBRE AS PUBLICAÇÕES DO SIEPE

Gustavo Cardozo Rodrigues [1]

Gustavo Rodrigues [2]

Tiago Furtado Drehmer Pinheiro [3]

Cristiano Tolfo [4]

Resumo:

A mineração de dados possibilita descoberta de novas informações que auxiliam na tomada de decisões. Este trabalho tem como objetivo realizar uma análise de trabalhos sobre data mining publicados no Salão Internacional de Ensino, Pesquisa e Extensão - SIEPE. O estudo também objetiva identificar as áreas de aplicação de data mining contidas nos trabalhos analisados. Para atender a estes objetivos, foi realizada uma pesquisa nos anais do SIEPE, considerando as publicações feitas no período de 2009 a 2017. Como resultado da pesquisa foram identificados 5 trabalhos que abordam o uso de data mining, o que representa uma quantidade considerada pequena, visto a importância do tema. Em virtude deste número de publicações, foi possível fazer uma análise em cada um dos trabalhos identificados. Considera-se então que esta quantidade pequena de publicações sobre o tema pode ser melhor explorada futuramente em trabalhos de iniciação científica, visto que a UNIPAMPA possui cursos das áreas de tecnologias e da computação.

Palavras-chave: mineração de dados, data mining, base de dados

Modalidade de Participação: Iniciação Científica

MINERAÇÃO DE DADOS: UMA ANÁLISE SOBRE AS PUBLICAÇÕES DO SIEPE

[1] Aluno de graduação. gustavocrod.unipampa@gmail.com. Autor principal

[2] Aluno de graduação. gustavocrod.unipampa@gmail.com. Apresentador

[3] Aluno de graduação. Tiago.drehmer@gmail.com. Co-autor

[4] Docente. ctolfo@gmail.com. Orientador

MINERAÇÃO DE DADOS: UMA ANÁLISE SOBRE AS PUBLICAÇÕES DO SIEPE

1 INTRODUÇÃO

Atualmente as organizações acumulam grandes quantidades de dados, obtidos através de operações diárias. Porém, a maioria ainda não usa adequadamente essa grande quantidade de informação para transformá-las em conhecimento (CÔRTES et al., 2002). A descoberta de novas informações através do reconhecimento de padrões, regras ou mudanças nessas grandes quantidades de dados é conhecida como mineração de dados (*data mining*) (ELMASRI; NAVATHE, 2005). O uso das técnicas de *data mining* permite então, que o conhecimento seja extraído de dados dispersos, possibilitando chegar em conclusões sobre o ambiente do qual eles foram recolhidos (CÔRTES et al., 2002).

Diante disso, este trabalho tem como objetivo realizar uma análise sobre os trabalhos publicados no Salão Internacional de Ensino, Pesquisa e Extensão - SIEPE que falam sobre *data mining*. O mesmo foi desenvolvido durante a disciplina Sistemas de Informação, ministrada no Curso de Ciência da Computação da Universidade Federal do Pampa - UNIPAMPA, Campus Alegrete.

O trabalho mostra um compilado com os artigos publicados e uma breve síntese de cada um para discernir as áreas que aplicam as técnicas de *data mining* no contexto da UNIPAMPA.

Este artigo está organizado da seguinte forma. Na seção 2, apresenta-se a metodologia aplicada no trabalho. Na seção 3 são mostrados os artigos encontrados e uma breve discussão sobre os mesmos. A seção 4 contém considerações sobre a pesquisa realizada.

2 METODOLOGIA

O levantamento do material para análise foi realizado sobre os anais da base dados do SIEPE, utilizando como palavras-chave: "mineração de dados" e "*data mining*". Após a seleção dos trabalhos que atendem ao objetivo da pesquisa, foi analisado qual a área de aplicação de cada um deles.

A busca envolveu as publicações realizadas no SIEPE no período de 2009 a 2017. A pesquisa visava identificar o número de publicações sobre *data mining* feitas no evento. Também pretendia identificar como estão sendo aplicadas as técnicas de mineração de dados nos campi da UNIPAMPA.

3 RESULTADOS e DISCUSSÃO

A pesquisa realizada nos anais do SIEPE resultou na identificação de 5 trabalhos que abordam a aplicação das técnicas de *data mining*.

Na Figura 1 é possível ver os trabalhos encontrados em relação ao longo dos anos em que foram publicados. Notável que após o primeiro ser publicado em 2009, se mantém uma frequência aproximada de uma publicação a cada dois anos, abordando os mais diferentes temas.

Figura 1- Trabalhos sobre mineração de dados publicados no SIEPE



Fonte: O autor

Na sequência é apresentada uma síntese de cada um dos trabalhos encontrados na pesquisa.

3.1 Processos De Data Mining Para A Formação De Conjuntos

No trabalho de Sechi et al. (2009), se visa introduzir alguns conceitos de Data Mining, a fim de retirar informações e compreender a estrutura de uma base de dados, para assim fornecer um resultado satisfatório e de fácil visualização para o usuário. Para isso foi utilizado uma tecnologia de mapas auto-organizáveis, gerados a partir do programa MATLAB, para analisar as características básicas e relevantes da tabela do campeonato brasileiro 2008. Sabendo que a mineração de dados é uma metodologia a qual visa identificar padrões semelhantes, ou seja, a partir de um banco de informações é executado o algoritmo, o qual é responsável pela organização e classificação dos clubes presente na competição em grupos. Com os grupos formado foi possível perceber que, times com dados semelhantes como São Paulo (Campeão) e Grêmio (Vice) se mantiveram juntos, justificando o fato deles terem tido as melhores campanhas da competição.

3.2 Desenvolvimento De Um Sistema Tutor Inteligente Para O Ensino De Nutrição Básica e Dietética

No trabalho de Pinto et al. (2017), é utilizado um sistema tutor inteligente para o ensino de nutrição básica e dietética, visando entender quais são as reais dificuldades dos alunos individualmente, para o aprendizado do conteúdo e assim determinar a melhor maneira que eles devem ser ensinado. No trabalho não são usados processos de data mining, porém como objetivos futuros eles definem que a utilização da mineração de dados, pode ser usada para auxiliar professores a interpretar os resultados obtidos na pesquisa realizada, de maneira que facilite a eles definir um perfil de ensinamento para cada aluno.

3.3 Uso de Avaliações Formativas no Ensino de Engenharias: Um Estudo de Caso

O trabalho apresentado por Santos et al. (2012), tem como objetivo, relacionar o desempenho de alunos matriculados em uma cadeira de engenharia da computação, sobre dados obtidos nas avaliações formativas proposta durante o decorrer da cadeira. Dados como percentual de assiduidade presencial; percentual de avaliações formativas respondidas; nota média das avaliações formativas; nota da avaliação somativa; rótulo da nota da avaliação

somativa (aprovado ou reprovado); e percentual de assiduidade total nas avaliação formativas; foram utilizados na mineração de dados para se chegar a uma conclusão sobre a eficácia do método. Para isso foi utilizado os algoritmos como REPTree e J48 disponibilizados pela ferramenta Weka. Com os resultados obtidos pelos algoritmos eles puderam constatar que quanto maior o comprometimento do aluno com a disciplina, participando das atividades formativas ou à presença em sala de aula, maior será sua chance de obter aprovação na avaliação somativa.

3.4 Desenvolvimento de um Software Utilizando Técnicas de Mineração de Dados para Análise de Evasão na UNIPAMPA

O trabalho proposto por Borin et al. (2013), tem como finalidade apresentar o software que estava sendo produzido ao qual o objetivo era definir padrões nos dados de alunos que largaram os cursos de engenharia da computação no campus bagé, a fim de justificar essa evasão. No trabalho foi utilizado a base de dados obtidos através do SIE (Sistema de Informações para o Ensino) da UNIPAMPA e do portal SiSU (Sistema de Seleção Unificado). Foi definido que sobre essa base dados existia a necessidade de um pré processamento que, consistiu em separar informações por alunos e a retirada de informações redundantes e irrelevantes. Após isso utilizando algoritmos da ferramenta weka seria aplicada a mineração de dados definindo assim o perfil de alunos que largaram o curso antes da graduação.

3.5 Mineração De Dados De Pesquisadores Usando O Scriptlattes

O trabalho proposto por Treviso e Kepler (2015), buscou melhorar o significado das informações e qualificação dos artigos da plataforma ScriptLattes, aplicando técnicas de mineração de dados. Através do Qualis, para qualificar os artigos, foram criados os métodos para quantificar as publicações que se enquadram em cada categoria do mesmo. Os autores aplicam uma refatoração sobre o analisador sintático (parser) da plataforma ScriptLattes para uma melhor filtragem dos dados, sendo que conseguiram melhorar o relatório com a integração do Qualis das publicações de artigos em periódicos de cada membro, para deixá-los mais significativo.

4 CONSIDERAÇÕES FINAIS

A pesquisa realizada neste trabalho atendeu ao objetivo de identificar os trabalhos publicados nos anais do SIEPE que abordam o tema *data mining*. Considerando a frequência de publicações e o grande benefício que a técnica traz, concluímos que esta é uma quantidade razoavelmente pequena e que pode ser melhor explorada futuramente. Podemos dizer também, que a técnica de mineração de dados vem sendo bastante útil para projetos acadêmicos nos campi da UNIPAMPA, sendo perceptível a aplicação dela em diferentes áreas, como nutrição, ensino, futebol, e modelagem de textos. Essas áreas representam uma pequena parcela do escopo que pode ser explorado com a mineração de dados.

trabalhos futuros:

Como estão as publicações de data mining (pesquisadores da unipampa) fora do contexto do SIEPE? Pesquisa em outras revistas e eventos - mapeamento

Pq existem poucas publicações dos pesquisadores da unipampa no siepe? - trabalhos futuros poderiam identificar fatores - correlação de areas de pesquisa dos docentes.

REFERÊNCIAS

BORIN, J. M. et al. Desenvolvimento de um Software Utilizando Técnicas de Mineração de Dados para Análise de Evasão na UNIPAMPA. In: V Salão Internacional de Ensino, Pesquisa e Extensão, 2. 2013, Bagé. **Anais**… Bagé: V SIEPE, 2013. p. 1.

DA COSTA CÔRTES, S.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados-funcionalidades, técnicas e abordagens**. PUC, 2002.

ELMASRI, R.; NAVATHE, S. **Sistemas de Bancos de Dados**. 4. São Paulo: Pearson Addison Wesley, 2005. 744 p.

SECHI, D. L. et al. Processos De Data Mining Para A Formação De Conjuntos. In: I Salão Internacional de Ensino, Pesquisa e Extensão, 1. 2009, Alegrete. **Anais…** Uruguaiana: I SIEPE, 2009. p. 1.

PINTO, T. H. D. el al. DESENVOLVIMENTO DE UM SISTEMA TUTOR INTELIGENTE PARA O ENSINO DE NUTRIÇÃO BÁSICA E DIETÉTICA, 1. In: 9º Salão Internacional de Ensino, Pesquisa e Extensão, 1. 2017, Itaqui. **Anais…** Sant'Ana do Livramento: 9º SIEPE, 2017. p. 1.

SANTOS, H. L. et al. Uso de Avaliações Formativas no Ensino de Engenharias: Um Estudo de Caso. In: IV Salão Internacional de Ensino, Pesquisa e Extensão, 1. 2012, Pelotas . **Anais…** Bagé: IV SIEPE, 2012. p. 1.

TREVISO, M. V.; KEPLER, F. N. In: VII Salão Internacional de Ensino, Pesquisa e Extensão, 2. 2015, Alegrete. **Anais…** Alegrete: VII SIEPE, 2015. p. 1.