

III Workshop de Dissertações de Mestrado do PPgSI (2016)

FICHA DA PESQUISA

| DADOS GERAIS | | | | |
|--------------------------------------|--|---|---|--|
| Título do projeto de pesquisa | Análise comparativa de técnicas de classificação de dados aplicadas para identificação dos determinantes de decisão do consumo de filmes em salas de cinema no Brasil. | | | |
| Orientando | Nathalia Nieuwenhoff | | | |
| Orientador(es) | Marislei Nishijima | | | |
| Momento atual | <input type="checkbox"/> 2º semestre | <input type="checkbox"/> 3º semestre | <input checked="" type="checkbox"/> 4º semestre | <input type="checkbox"/> 5º semestre |
| Qualificação | <input type="checkbox"/> Qualificação já realizada em: / / | | <input type="checkbox"/> Qualificação planejada para: /09 /2018 | |
| Defesa | Prazo máximo para depósito: / / | | Depósito planejado para: /04/2017 | |
| Linha e Área de pesquisa | <input checked="" type="checkbox"/> Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input checked="" type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software <input type="checkbox"/> IHC | | <input type="checkbox"/> Inteligência de Sistemas: <input type="checkbox"/> IA <input type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico | |
| Área de aplicação | <input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input checked="" type="checkbox"/> Economia | <input type="checkbox"/> Educação <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Internet <input type="checkbox"/> Jogos | <input type="checkbox"/> Linguagem Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Química | <input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input type="checkbox"/> [outro – escrever] |
| Publicações | - | | | |

| DESCRIÇÃO DO PROJETO DE PESQUISA | |
|-------------------------------------|---|
| Contextualização / motivação | A evolução tecnológica resultou na mudança da lógica de consumo de bens de informação. A identificação de padrões de consumo neste contexto possibilita a avaliação mais precisa dos recursos envolvidos neste mercado visando o políticas mais adequadas para empresas e governo. |
| Problema de pesquisa | Utilização de técnicas de reconhecimento de padrões para classificação dos dados e identificação de padrões de consumo de um bem de informação, filmes em salas de cinema, tema complexo e pouco explorado no mercado brasileiro. |
| Objetivo geral da pesquisa | Analisar comparativamente o desempenho de 2 diferentes técnicas de reconhecimento de padrões para classificação dos dados e identificação de padrões do consumo de filmes em salas de cinema no Brasil. |
| Trabalhos relacionados | 1. Mas'ud, M.Z.; Sahib, S.; Abdollah, M.F.; Selamat, S.R.; Yusof, R., "Analysis of Features Selection and Machine Learning Classifier in Android Malware Detection," 2014; 2. Hong Y.; Xiaolei H.; Xiaorong H.; Hengwen C., "A Comparative Study on Data Mining Algorithms for Individual Credit Risk Evaluation", 2010; 3. Balachandran, K.; Anitha, R., "Ensemble based optimal classification model for pre-diagnosis of lung cancer," Computing, Communications and Networking Technologies (ICCCNT), 2013. |
| Justificativa e relevância | Os recursos públicos e privados cada vez possuem mais necessidade de identificar os padrões de consumo das pessoas para o estabelecimento de políticas públicas de acesso e para auxiliar as estratégias de venda e divulgação dos bens de informação para as empresas que atuam nesta área. Porém, a maioria dos estudos sobre tais padrões relacionados aos bens de informação se baseia em análises qualitativas sem o uso sistemático de informações quantitativas, bem como a combinação destas. |
| Proposta para Solução | Utilizar duas técnicas utilizadas na literatura de reconhecimento de padrões empregadas em aplicações com grande volume dados disponível para análises de identificação de padrões de consumo de cinema e geração de ferramenta para orientação de políticas para empresas do setor e governo. |
| Dados | Os dados que serão utilizados para o trabalho foram coletados no site do Instituto Brasileiro de Geografia e Estatística (IBGE) e se referem à Pesquisa de Orçamento Familiar (POF) realizada em um período de 12 meses entre os anos 2008 e 2009. |
| Validação | A validação da aplicação das técnicas de classificação de dados irá considerar as medidas relacionadas à precisão dos seus resultados e desempenho computacional, para identificação de características relevantes, vantagens e desvantagens de cada uma aplicadas no contexto em questão. Adicionalmente, serão utilizadas informações dentro e fora da amostra para validar o poder de classificação e previsão do modelo. |
| Limitações | O trabalho é uma aplicação multidisciplinar, direcionada para um bem de informação específico, filmes consumidos no mercado brasileiro de cinema. Adicionalmente, a análise comparativa será restrita a aplicação de duas técnicas de classificação de dados, sendo estas árvores de decisão (ID3) e redes neurais artificiais (Naive Bayes). |
| Resultados | Contribuições científicas: Identificar padrões de consumo de bens de informação e apresentar as principais vantagens e desvantagens da aplicação de técnicas de reconhecimento de padrões no contexto em questão. Adicionalmente, tem-se a contribuição da utilização de grande número de informação para um objeto ainda não investigado. A cultura no Brasil está sujeita a um conjunto de |

