

# **Programa de Pós Graduação de computação aplicada UDESC - JOINVILLE**

**Disciplina:** Metodologia da Pesquisa

**Mestrando:** Tiago Pastorello

Protocolo de revisão bibliográfica sistemática para realização de dissertação, na área de mineração de fluxo de dados parcialmente classificados com a presença de mudança de conceitos

## **1.1 Resumo**

A crescente necessidade na melhora do desempenho, dos algoritmos existentes de mineração de fluxo de dados e inteligência artificial, inspira a realização desse projeto de pesquisa científica intitulado: Mineração de fluxo de dados parcialmente classificados com a presença de mudança de conceitos.

A grande área que compreende o tema desta pesquisa é a mineração de dados. Dentro deste tema, o foco desta pesquisa científica compreende a mineração de fluxos de dados parcialmente classificados. Dentro do tema abordado o problema que buscamos elucidar com a realização desta pesquisa é como classificar fluxos de dados parcialmente classificados com presença de mudança de conceitos. Com isso a solução investigada no projeto de pesquisa é a aplicação de métodos de classificação semi-supervisionada, baseada no conceito de árvores de decisão.

**Keywords:**Semi-supervised learning, Iterative algorithms, Computer vision, Application software, Computational efficiency, Decision trees, Annealing, Error correction, Machine learning algorithms, Degradation

## **1.2 Introdução**

Cotidianamente somos expostos a grandes volumes de dados, e é notável que esse montante tem aumentado diariamente em grandes proporções. Esse grande número de dados estão disponíveis aos seres humanos sob as mais variadas e diversas formas, porém uma grande quantidade de dados não se traduz necessariamente em informação, pois a exposição/ acesso a estes dados “soltos” não significam necessariamente aquisição de conhecimento. Sob esse aspecto (ZLIOBAIT e et al., 2014) descreve que é possível a partir desses dados construir modelos preditivos para tentar entender o comportamento de vários tipos de sistemas em determinadas situações, e prever seus resultados com determinadas entradas, mais muitas vezes os modelos acabam ficando desatualizado e imprecisos porque são utilizados por muito tempo.

Outros fatores também influenciam na qualidade preditiva destes modelos, como tamanho infinito dos dados de entrada, dificuldade ou impossibilidade de classificação dos dados e mudanças na função objetivo do modelo, conhecido como "mudança de conceito".Com isso a utilização de algoritmos de aprendizado semi-supervisionado

demonstra-se uma alternativa. Nesta pesquisa buscamos gerar um modelo que seja capaz de classificar fluxos de dados parcialmente classificados que consiga lidar com possíveis mudanças de conceitos.

## 2.1 Trabalhos Relacionados

Para a realização desta pesquisa científica é de suma importância identificar o “estado da arte” do conhecimento em relação ao tema apresentado, para isso identificou-se a necessidade da realização de uma RBS/PBS, a fim de identificar artigos, periódicos e autores que possam agregar conhecimento e informações relevantes à pesquisa científica que se pretende realizar, além de ajudar a identificar possíveis propostas para o aperfeiçoamento dos métodos de classificação de fluxos de dados parcialmente classificados existentes que serão abordados na referida pesquisa. Neste contexto abaixo são citados alguns trabalhos relacionados ao tema desta pesquisa:

### **Semi-Supervised Random Forests," 2009 IEEE**

C. Leistner, A. Saffari, J. Santner and H. Bischof, *12th International Conference on Computer Vision*, Kyoto, 2009, pp. 506-513, doi: 10.1109/ICCV.2009.5459198

### **Semi-Supervised Learning Literature Survey Xiaojin Zhu Computer Sciences TR 1530 University of Wisconsin – Madison Last modified on July 19, 2008**

**Applying semi-supervised learning in hierarchical multi-label classification.** Araken Santos a , Anne Canuto b, <sup>†</sup> a Exact, Technology and Human Sciences Department, Federal Rural University of Semi-Arido, Angicos, RN 59515-000, Brazil Department of Informatics and Applied Mathematics (DIMAp), Federal University of RN, Brazil

## 3.1 Planejamento da RBS

Revisão bibliográfica sistemática é o processo de coletar, conhecer, compreender, analisar, sintetizar e avaliar um conjunto de artigos científicos com o propósito de criar um embasamento teórico-científico (estado da arte) sobre um determinado tópico ou assunto pesquisado. (Levy e Ellis, 2006).

A metodologia adotada para nortear a realização desta RBS, foi obtida com base em diferentes modelos já preconizados em outras áreas do conhecimento, sendo assim uma adaptação da proposta de outros autores, visando agregar o conteúdo trazido por estas metodologias para a realização desta RBS específica do tema da dissertação em questão, Mineração de fluxo de dados parcialmente classificados com a presença de mudança de conceitos.

Para a realização da execução da uma Revisão Bibliográfica (Sistemática, Levy e Ellis, 2006) defendem que esta realização seja feita por meio de um processo, onde o processo é uma sequência de passos e atividades para alcançar os resultados pretendidos. Os autores definem três fases principais neste processo, sendo estas entrada, processamento, e saída.

Na primeira fase denominada “entrada” estão compreendidas as informações preliminares que serão ainda processadas (como por exemplo: artigos clássicos , livros-texto, artigos de referência e artigos indicados por especialistas).

Na segunda fase do processo denominada de Processamento é onde se aplicam as seguintes etapas: conhecimento e compreensão da literatura, aplicação da revisão, análise dos resultados, síntese e avaliação dos resultados. Processos estes que ao serem executados irão definir as saídas de informações, representadas pelos autores como a fase 3 do referido processo.

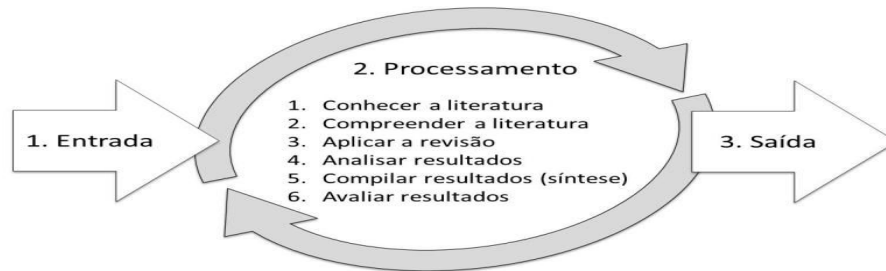


Figura 1. Fases de uma revisão bibliográfica efetiva. Fonte: Adaptado de Levy e Ellis (2006)

Outro modelo é apresentado por Biolchini et al. (2007) e aborda as principais fases de um processo para a realização de uma revisão sistemática, com foco na área de tecnologia de informação.

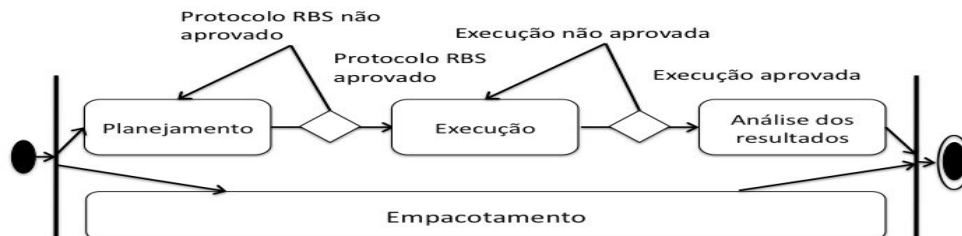


Figura 2. Representação do processo sistemático para revisão bibliográfica.

Fonte: Adaptado de Biolchini et al.(2007)

Como apresentado por Biolchini et al.(2007) este processo se inicia pela fase de planejamento. Nesta fase da RBS é definido o protocolo que detalha as questões centrais como tema da pesquisa, objetivos, palavras-chave, Strings de busca e o método de execução. Ainda nesta fase serão identificados, selecionados e avaliados os estudos primários, que passaram pelos critérios de inclusão e exclusão estabelecidos no protocolo da RBS. Após esta fase em que serão selecionados os artigos primários, os dados serão extraídos e compilados durante a fase de análise de resultados.

## 3.2 Roteiro para a realização da RBS

Para a realização da RBS, será utilizado um roteiro com base no modelo de RBS Roadmap. (Conforto et al.,2011). Este é dividido em 15 etapas e 3 fases, entrada, processamento e saída.

### 3.3.1 FASE 1 – ENTRADA

**ETAPA 1.1, Problema** – O ponto de partida da RBS é a definição do problema, onde se busca elucidar uma ou mais questões. Nesta fase a formulação do problema deve se dar na forma de uma pergunta, deve ser claro, preciso e empírico. O problema em questão deve ser suscetível de resolução e ter uma dimensão viável.

**ETAPA 1.2, Objetivos** – Estes devem ser claros e factíveis pois a partir dos mesmos serão realizados os critérios de inclusão dos artigos para o estudo.

**ETAPA 1.3, Fontes primárias** – Consistem em artigos periódicos ou bases de dados que auxiliarão na definição das palavras-chave, e de autores e artigos relevantes. Nesta etapa é importante definir critérios de inclusão das fontes de pesquisa, e se atentar a condições como: o periódico está indexado em base de dados; apresenta trabalhos completos na área de investigação; se o trabalho foi indicado por especialistas, ou mesmo se foi identificado em estudos preliminares.

**ETAPA 1.4, Strings de busca** – Nesta fase é necessário identificar as palavras e termos referentes ao tema, também se faz necessário compreender as regras para se criar Strings de Busca utilizando operadores lógicos comumente aplicados em buscas avançadas ou buscas booleanas, se faz necessário testar combinações string para que funcione corretamente nos diferentes mecanismos de busca dos periódicos. O mesmo procedimento deve ser feito para conduzir as buscas nas bases de dados. Os operadores lógicos, tais como “OR”, “AND” são comuns na maioria dos mecanismos de busca.

**ETAPA 1.5, Critérios de inclusão** – Nesta etapa é necessário especificar e esclarecer quais são os aos objetivos da pesquisa a fim de possibilitar a escolha ou não de artigos para compôr a RBS, para isso utilizando-se de filtros de leitura. No primeiro filtro de leitura resumo e palavras-chave são lidos.

**ETAPA 1.6, Critérios de qualificação** – Nesta etapa é preciso observar elementos do artigo tais como: método de pesquisa utilizado, a quantidade de citações do artigo, o fator de impacto da revista que o artigo foi publicado, dentre outros. Alguns critérios como os exemplificados a seguir podem ser aplicados em diversos tipos de estudos: o método de pesquisa adotado; quantidade de citações que o artigo obteve; se o método adotado é do tipo “levantamento” (survey); se teve caráter explicativo ou descritivo; qual o tamanho e diversidade da amostra, entre outros.

**ETAPA 1.7, Método e ferramentas** – Deve-se definir as etapas para a condução das buscas, definir os filtros de busca, como será realizado a busca nos periódicos e bases de dados, como os resultados serão armazenados. Os métodos devem compreender refinamento de busca e buscas cruzadas, a partir de referências citadas nos artigos encontrados.

**ETAPA 1.8, Cronograma** – Definição do cronograma para realização da RBS, também dos equipamentos, planejar aquisição de softwares, definir parcerias, etc..

## FASE 2 – PROCESSAMENTO

**Etapa 2.1,** são realizadas as buscas.

**Etapa 2.2,** é realizada a leitura e análise dos resultados. Nesta etapa aplica-se o primeiro filtro de leitura onde o resumo e palavras-chave são lidos.

**Etapa 2.3,** neste passo utiliza-se o filtro 2 de leitura que consiste na leitura da introdução e conclusão dos artigos, posteriormente é realizada a documentação e arquivamento dos artigos selecionados nos filtros, bem como os resultados das buscas e filtros de leitura, compreende listagem da quantidade de artigos encontrados por periódico, quantidade de artigos excluídos, quantidade de artigos encontrados na busca cruzada, etc.. este processo é importantes para refinar as buscas e posteriormente serão úteis para argumentação teórica e embasamento da síntese da teoria sobre o assunto pesquisado.

**Etapa 2.4,** os artigos selecionados no filtro 2, são então submetidos à leitura completa (filtro 3).

**Etapa 2.5,** este é o início do processo iterativo de busca cruzada, onde são rastreados e identificados artigos relevantes por meio das citações dos autores. É o momento de ler por completo o trabalho, analisar e interpretar o texto.

**Etapa 2.6,** realiza-se a busca nas bases de dados. Esta busca visa encontrar artigos recentes, publicados em anais de congressos e eventos científicos que estão indexados em bases de dados. Para realizar esta busca e identificar um conjunto de artigos relevantes para o estudo usamos a String, o que reforça a necessidade do refinamento da mesma.

**Etapa 2.7,** catalogação e armazenagem dos artigos que passaram pelos filtros de leitura e irão compor as referências bibliográficas.

## FASE 3- SAÍDA

**ETAPA 3.1 Alertas,** realizar a inserção de “alertas” nos principais periódicos identificados durante a condução da RBS. Estas informações são úteis para priorizar os artigos e consequentemente identificar os principais periódicos para a área de estudo. Essa ação é útil para rastrear novos artigos e atualizar o repositório de artigos da pesquisa.

**ETAPA 3.2 Cadastro e arquivo,** os artigos selecionados nos filtros de leitura, analisados e interpretados serão incluídos no repositório de artigos da pesquisa.

**ETAPA 3.3 Síntese e resultados,** nesta etapa elabora-se um relatório que será uma síntese da bibliografia estudada. Este relatório consiste em um texto sobre o assunto estudado que poderá assumir o formato uma seção de revisão bibliográfica que posteriormente será inserida na tese ou dissertação

**ETAPA 3.4 Modelos teóricos,** As definições das hipóteses ou a construção dos modelos teóricos será embasada nos resultados da RBS, e a síntese do tema estudado. O modelo teórico é fruto do resultado final da RBS. Sendo este importante, para realização de estudos de caso, ou survey. Caso a RBS tenha sido embasada em hipóteses pré-definidas, nesse momento elas podem ser avaliadas e confrontadas com os resultados obtidos, verificando se podem ser refutadas ou corroboradas.

## 4.1 Referências Bibliográficas

Bereton et al. Lessons from Applying the Systematic Literature Review Process within the Software Engineering Domain. The Journal of System and Software, v. 80, p.571-583, 2007.

Biolchini, J.C.A., et al. Scientific research ontology to support systematic review in software engineering. Advanced Engineering Informatics, v.21, n.2, p.133-151, 2007.

Conforto, C.B; Amaral,D.C; Silva,S.L. Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. CDBGP Porto Alegre RS-Brasil, 12,13,14 Set 2011.

Cook, D.J.; Mulrow, C.D.; Haynes, R.B. Systematic reviews: synthesis of best evidence for clinical decisions. Annals of Internal Medicine, v.126, n.5, pp.376-380, 1997.

Dyba, T.; Dingsoyr, T. Empirical studies of agile software development: a systematic review. Information and Software Technology, v.50, pp.833-859, 2008.

Gil, A. Como elaborar projetos de pesquisa. Atlas: São Paulo, 2007.

Levy, Y.; Ellis, T.J. A system approach to conduct an effective literature review in support of information systems research. Informing Science Journal, v.9, p.181-212, 2006.

Portal de Periódicos CAPES. Disponível em: <http://novo.periodicos.capes.gov.br/>. Acesso em 09 julho 2020.

ZHU, X. Semi-Supervised Learning Literature Survey. 2005.

ŽLIOBAITė, I. et al. Active learning with drifting streaming data. IEEE Transactions on Neural Networks and Learning Systems, v. 25, n. 1, p. 27–39, Jan 2014. ISSN 2162-237X.