

UNIVERSIDADE FEEVALE

LUCAS ADAMS CAMARGO

DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS DE
ALTERAÇÕES EM SISTEMA DE GESTÃO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2019

LUCAS ADAMS CAMARGO

DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS DE
ALTERAÇÕES EM SISTEMA DE GESTÃO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Ciência da Computação pela
Universidade Feevale

Orientador: Gabriel da Silva Simões

Novo Hamburgo
2019

RESUMO

No processo de desenvolvimento de software frequentemente são inseridas inconsistências no sistema, o que acaba gerando diversos transtornos, tanto para a empresa, que precisa investir tempo em correções, quanto para o cliente, que não consegue utilizar o sistema apropriadamente. A empresa Rech Informática possui todas as alterações em seu sistema ERP documentadas de forma que é possível identificar se a alteração gerou alguma inconsistência ou não. Entretanto, não é possível analisar manualmente todos esses dados para extrair informações relevantes sobre essas inconsistências e evitar que elas ocorram. Dessa forma, esse trabalho trata da aplicação de técnicas de mineração de dados e *Machine Learning* sobre essa base de dados objetivando a geração de um modelo capaz de prever se uma determinada alteração no sistema irá gerar inconsistência. O processo de aprendizado de máquina será desenvolvido utilizando a linguagem de programação Python, que possui bibliotecas voltadas para a Ciência de Dados.

Palavras-chave: Inconsistências. Mineração de dados. *Machine Learning*.

SUMÁRIO

| | |
|---------------------|-----------|
| MOTIVAÇÃO | 5 |
| OBJETIVOS | 7 |
| METODOLOGIA | 8 |
| CRONOGRAMA | 9 |
| BIBLIOGRAFIA | 10 |

MOTIVAÇÃO

Cada vez mais a tecnologia faz parte do cotidiano, o que inclui o dia-a-dia das empresas, que frequentemente sustentam suas atividades utilizando sistemas de gestão para controlar o negócio. Esses sistemas são conhecidos pela sigla em inglês ERP (*Enterprise Resource Planning*), que se refere a um software de Planejamento de Recursos Empresariais. Segundo Davenport (2002), sistemas ERP podem ser definidos como aplicativos separados em pacotes que dão apoio às decisões da empresa fornecendo informações em um sistema integrado.

Uma das empresas que atua na área de desenvolvimento de sistema ERP é a Rech Informática, que está alocada na cidade Novo Hamburgo no estado do Rio Grande do Sul. O principal software desenvolvido pela empresa é o SIGER® (Sistema Integrado de Gestão Rech), que atende um total de aproximadamente 10.000 usuários (RECH, 2019). Conforme Haberkorn (1999) propõe, este tipo de empresa possui alguns processos centrais, como levantamento de requisitos e validação por exemplo, porém se destaca a atividade de desenvolvimento de software. No cenário da Rech Informática, todas as alterações no sistema estão registradas em uma ferramenta interna de controle chamada Sicla. Essa ferramenta consolida informações sobre atendimentos telefônicos realizados, compromissos em agendas dos técnicos, além do próprio registro de alterações no sistema ERP, que será alvo de estudo neste trabalho.

A base de dados do Sicla possui diversos dados sobre as alterações realizadas no sistema, e entre essas informações é possível verificar quais alterações levaram ao surgimento de inconsistências no sistema. Na realidade da empresa esses dados são aproveitados apenas para fins estatísticos de monitoramento, de forma que se tem o controle, por exemplo, se a quantidade total de horas investidas em correção de defeitos aumentou ou diminuiu em relação à anos anteriores. Sendo assim, existe um grande volume de informações relativas ao processo de desenvolvimento de software que poderiam ser aproveitadas para gerar conhecimento sobre essa atividade e assim possibilitar que ela ocorra de forma mais eficiente. Além disso, essas informações podem ser utilizadas não somente para monitorar o passado, mas também para prever futuras ocorrências de inconsistências.

Para solucionar esse tipo de situação que existe a área nomeada Descoberta de Conhecimento em Base de Dados ou, na sigla em inglês, KDD (*Knowledge Discovery in Databases*). Uma definição para este termo elaborada em 1996 é a seguinte: KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados (FAYYAD et al 1996 apud GOLDSCHIMIDT;PASSOS, 2005, p.9).

Como salientado anteriormente, o processo de descoberta de conhecimento é composto por várias etapas e, segundo Goldschmidt e Passos (2005), destacam-se três principais: pré-processamento, mineração de dados e pós-processamento. O pré-processamento abrange as operações de obtenção e preparação dos dados, enquanto que a etapa de mineração se refere à etapa na qual são obtidos os conhecimentos acerca da base de dados. A última etapa, o pós-processamento, objetiva avaliar a utilidade dos conhecimentos obtidos, buscando encontrar aplicações para os mesmos.

A descoberta de conhecimento em bases de dados de empresas desenvolvedoras de software já foi alvo de estudo em algumas ocasiões. Pode ser citado como exemplo o estudo realizado por Cruz e Ruiz (2008), no qual são analisados os apontamentos de atividades registradas durante o desenvolvimento e manutenção de software na empresa desenvolvedora. Esses apontamentos fornecem dados para a mineração de processos, que resulta em informações sobre como as atividades estão sequenciadas e que dependências existem entre elas. Uma questão relevante apontada por esse trabalho é a atenção para a distribuição dos dados disponíveis para análise, pois no cenário estudado existiam poucas informações sobre o processo de execução de projetos, pois são demandas que ocorrem com menos frequência. Da mesma forma, na base de dados de alterações no SIGER® existem um pequeno número de alterações que geram inconsistências se comparado à quantidade total de alterações feitas.

Porém, além de extrair conhecimento da base de dados também é possível gerar modelos capazes de prever o comportamento de novos cenários baseado nos dados antecessores através de um processo chamado *Machine Learning*. Como citado por Mitchell (2006), enquanto a Ciência da Computação focava inicialmente em programar manualmente computadores, o processo de *Machine Learning* é focado em como é possível fazer os computadores programarem a si próprios. Ou seja, esta área propõe técnicas de aprendizado que utilizam a experiência entregue à aplicação para gerar performance na execução de uma tarefa proposta (MITCHELL, 2006).

Sendo assim, o objetivo deste trabalho é analisar a base de dados de alterações realizadas para manutenção de software utilizando técnicas de *Machine Learning* visando gerar um modelo capaz de prever a ocorrência de inconformidades no sistema. Para isso, será necessário obter da base de dados as informações que serão analisadas, submetendo as mesmas à preparação necessária para utilização no processo de mineração de dados. A exploração dos dados será feita utilizando recursos da linguagem de programação Python. Após isso, serão estudados os resultados obtidos buscando validar a acurácia das regras e conhecimentos encontrados.

OBJETIVOS

Objetivo geral

Analisar a base de dados das alterações realizadas em um sistema de gestão empresarial aplicando técnicas de aprendizado de máquina para gerar um modelo capaz de prever a ocorrência de inconsistências no sistema.

Objetivos específicos

- Estudar técnicas de ciência de dados e mineração de dados;
- Extrair dados da base de dados de alterações no sistema ERP;
- Pré-processar informações da base de dados;
- Aplicar ferramentas de mineração de dados e aprendizado de máquina;
- Estudar resultados obtidos e avaliar confiabilidade dos mesmos.

METODOLOGIA

Neste trabalho será utilizada a natureza de pesquisa aplicada, pois serão utilizados conhecimentos e técnicas previamente estudados para gerar um modelo preditivo sobre as inconsistências no sistema ERP. No que se refere aos objetivos, a pesquisa será explicativa, haja vista que serão utilizadas técnicas de descoberta de conhecimento para auxiliar na compreensão do cenário estudado e caracterização do mesmo.

O procedimentos técnicos adotados para este trabalho serão a pesquisa bibliográfica e pesquisa experimental. Sendo assim, primeiramente será necessário buscar o referencial teórico para os processos de mineração de dados e *Machine Learning*, fundamentando a pesquisa e obtendo conhecimentos necessários para o desenvolvimento do trabalho. Após isso, serão colhidos os dados necessários da base de dados de alterações no sistema ERP, realizando o pré-processamento necessário para utilizá-los nos algoritmos de aprendizado de máquina. A parte experimental do estudo se dará por meio da aplicação dos algoritmos de aprendizado de máquina e posterior avaliação dos mesmos em relação à acurácia da predição do modelo.

CRONOGRAMA

Trabalho de Conclusão I

| Etapa | Meses | | | |
|--|-------|-----|-----|-----|
| | Mar | Abr | Mai | Jun |
| Redação do anteprojeto | ■ | ■ | | |
| Pesquisa bibliográfica sobre KDD e Machine Learning | | ■ | ■ | |
| Estudar linguagem Python e funções de aprendizado de máquina | | ■ | ■ | |
| Redigir TC I | | ■ | ■ | ■ |

Trabalho de Conclusão II

| Etapa | Meses | | | |
|---|-------|-----|-----|-----|
| | Ago | Set | Out | Nov |
| Extrair e pré-processar dados da base de alterações | ■ | ■ | | |
| Implementar e aplicar algoritmos de <i>Machine Learning</i> | | ■ | ■ | |
| Avaliação dos resultados obtidos | | | ■ | |
| Redigir TC II | ■ | ■ | ■ | ■ |

BIBLIOGRAFIA

CRUZ, John Igor B. da, RUIZ, Duncan Uma experiência em mineração de processos de manutenção de software. **In Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web (WebMedia '08)**. Vila Velha, ES. p. 247-253, out. 2008.

DAVENPORT, Thomas H. **Missão Crítica: obtendo vantagens competitivas com os sistemas de gestão empresarial**. Porto Alegre, RS: Bookman, 2002. 293 p.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: Um Guia Prático**. Rio de Janeiro: Elsevier Editora Ltda, 2005 – 4a Reimpressão.

HABERKORN, E. **Teoria do ERP**. São Paulo, Makron Books, 1999.

MITCHELL, Tom. *The Discipline of Machine Learning*. Pittsburgh: Carnegie Mellon University, julho de 2006. Disponível em:
<<http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>>. Acesso em: 30 março 2019.

Rech Informática, Nossos clientes. Disponível em: <<https://www.rech.com.br/clientes>>
Acesso em: 17 mar. 2019.