

DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS DE PROJETOS DE INOVAÇÃO: UM ESTUDO EM UM POLO EMBRAPI

Bruno Netto Barbosa Paixão

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Polo de Inovação Campos dos Goytacazes - BR356, km 158 - Martins Lage - Campos dos Goytacazes - RJ
brunonbpaixao@gmail.com

Leonardo Barroso da Silva

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Polo de Inovação Campos dos Goytacazes - BR356, km 158 - Martins Lage - Campos dos Goytacazes - RJ
leo.barroso.silva@gmail.com

Marcus Vinicius das Neves Toffano

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Polo de Inovação Campos dos Goytacazes - BR356, km 158 - Martins Lage - Campos dos Goytacazes - RJ
mtoffano@uol.com.br

Henrique Rego Monteiro da Hora

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Polo de Inovação Campos dos Goytacazes - BR356, km 158 - Martins Lage - Campos dos Goytacazes - RJ
henrique.dahora@iff.edu.br

Rogério Atem de Carvalho

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Polo de Inovação Campos dos Goytacazes - BR356, km 158 - Martins Lage - Campos dos Goytacazes - RJ
ratem@iff.edu.br

RESUMO

Contexto: Atualmente a informação é um dos ativos mais importantes das organizações, porém é preciso transformar dados em conhecimento. Deste modo a mineração de dados (MD) tem um importante papel nas análises das informações e na descoberta de padrões. Neste trabalho foram utilizadas técnicas de MD com o auxílio do *software* WEKA para seleção de atributos, classificação e agrupamento de informações de projetos do pólo de inovação do Instituto Federal Fluminense. **Objetivo:** Identificar padrões dentro do conjunto dos dados dos projetos de inovação realizados pelo IFF de modo a gerar conhecimento e otimizar a gestão de projetos futuros. **Metodologia:** Inicialmente foi realizada a extração das informações a partir da base de dados de projetos do Polo de Inovação de Campos dos Goytacazes. Em seguida foram selecionados atributos relevantes para mineração, com a adequação dos dados para obtenção de melhores resultados. Por fim, houve a execução do algoritmo de classificação e geração da árvore de decisão que permitiu revelar informações dos projetos antes escondidas. **Resultados:** Os resultados descritos nos diversos ramos da árvore de decisão variam em termos de perfis de usuário, tipos de tarefas e tempo gasto. São reveladas interações entre perfis de usuário específicos, como Gerente/Coordenador e Bolsista/Estagiário, organização de atividades de autoria própria dos Pesquisadores e possível mudança no rol de tarefas de desenvolvimento, apontando muitas atividades sem registro de horas. **Conclusões:** Após a seleção dos atributos e geração das árvores de decisão, foi possível interpretar e descrever as relações entre os atributos das tarefas, revelando informações sobre os projetos. Os resultados podem ser comparados com artefatos de projetos existentes ou ainda servir de entrada para novos projetos do escritório de projetos. Novas descobertas de conhecimento na mesma base de dados podem ser experimentadas se novos cenários forem aplicados e novos algoritmos, como o K-means.

Palavras-chave: Banco de dados; Gerenciamento de projetos; j48; Mineração de dados; Weka.

ABSTRACT

Context: Information is currently one of the most important assets of organizations, but it is necessary to transform data into knowledge. Data mining (DM) plays an important role in information analysis and pattern discovery. In this paper we used DM techniques with the help of WEKA *software* for attribute selection, classification and information grouping of projects of the IFF Innovation Hub. **Objective:** To identify standards within the data set of innovation projects carried out by IFF in order to generate knowledge and optimize future project management. **Methodology:** Initially, information was extracted from the project database of the Campos dos Goytacazes Innovation Hub. Then, mining-relevant attributes were selected, with data adequacy for better results. Finally, the decision tree classification and generation algorithm was executed, which allowed revealing information from previously hidden projects. **Result:** The results described in the various branches of the decision tree vary in terms of user profiles, task types, and time spent. Iterations are revealed between specific user profiles, such as Manager / Coordinator and Scholars / Interns, organization of Researchers' own authoring activities and possible change in the development task list, pointing to many activities without time recording. **Conclusions:** After the selection of attributes and generation of decision trees, it was possible to interpret and describe the relationships between task attributes, revealing information about the projects. The results can be compared with existing project artifacts or as input to new projects from the project office. New knowledge discoveries in the same database can be tried if new scenarios are applied and new algorithms such as K-means.

Key-words: Data Base; Project Management; j48; Data mining; Weka.

1. INTRODUÇÃO

Nos dias atuais, devido à facilidade crescente de aquisição de imagens e coleta de dados das operações e atividades das empresas públicas e privadas para aplicações, o volume de dados armazenados cresce muito além da capacidade humana de processá-los para extrair informações relevantes e conhecimentos úteis à tomada de decisão (GUIDINI *et al.*, 2014). À medida que cada vez mais essa grande quantidade de dados está sendo coletado pelas organizações, técnicas que permitem mineração eficiente de grandes bancos de dados atraíram nos últimos anos interesse da academia e da indústria (CHRISTEN *et al.*, 2007).

Data Mining (ou em português, mineração de dados), é uma das alternativas mais eficazes para extrair conhecimento a partir de um grande volume de dados referentes (CARDOSO *et al.*, 2008). Ele é formado por um conjunto de ferramentas e técnicas que através do uso de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatística, são capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento. Esse conhecimento pode ser apresentado por essas ferramentas de diversas formas: agrupamentos, hipóteses, regras, árvores de decisão, grafos, ou dendrogramas (LAUDON *et al.*, 2011).

Na *Data Mining* aplica-se conhecimento multidisciplinar envolvendo as áreas de gerenciamento de bancos de dados, inteligência artificial e análise estatística, entre outras. Essa junção de áreas aumenta a complexidade dos algoritmos, exigindo capacidade computacional e critério na busca de padrões (MARKOV *et al.*, 2016). No campo da gestão, Laudon *et al.* (2011) definem como o uso da tecnologia da informação para descobrir regras, identificar fatores e tendências-chave, descobrir padrões e relacionamentos ocultos em grandes bancos de dados para auxiliar a tomada de decisões sobre estratégia e vantagens

competitivas.

De fato, Guidini *et al.* (2014) afirmam que sistemas de mineração de dados que implementam uma boa parte dos algoritmos desenvolvidos não são facilmente encontrados com código-fonte aberto e licença livre, porque fazem parte de aplicações comerciais. Como exemplos de sistemas livres os autores citam o Weka (Waikato Environment for Knowledge Analysis), que é um sistema de mineração de dados para realizar previsões e previsões através da utilização de dados, que foi desenvolvido na Universidade Waikato de Hamilton, Nova Zelândia, implementado em linguagem JAVA (WEKA), e o MLC++ (A Machine Learning Library in C++).

Data Mining se concentra no uso de aprendizado de máquina, reconhecimento de padrões e estatísticas para descobrir padrões nos dados (SILVA *et al.*, 2017). A aprendizagem pode ser de dois tipos: Aprendizagem supervisionada e aprendizagem não supervisionada. Em dados de aprendizado supervisionado é extraído/recortado dados de grandes bancos de dados, usando a classe de destino, onde como na aprendizagem não supervisionada, não há informações anteriores e tudo é feito dinamicamente (CHUG *et al.*, 2013).

A técnica de classificação da *Data Mining* é um método de aprendizado supervisionado baseado no conjunto de amostras de treinamento, usado para a previsão de defeitos de *software*. Ela é o processo de encontrar um conjunto de modelos que descrevem e distinguem classes, com o propósito de utilizar o modelo final (refinado) para predizer a classe de objetos que ainda não foram classificados. O modelo construído baseia-se na análise prévia de um conjunto de dados de amostragem ou dados de treinamento, contendo objetos corretamente classificados e o seu princípio é descobrir algum relacionamento entre os atributos preditivos e o atributo meta, de modo a descobrir um padrão que possa ser utilizado para previsões (LIBRELOTTO *et al.*, 2008). Os algoritmos de classificação incluem métodos que utilizam árvores de decisão (ID3, CART e C4.5), redes bayesianas, vizinhos mais próximos, algoritmos genéticos, lógica nebulosa, classificação baseadas em regras, entre outros (RODRIGUES *et al.*, 2012).

O classificador J48 é uma implementação escrita em Java do algoritmo C4.5, presente na ferramenta de Weka, e tem a finalidade de gerar uma árvore de decisão baseada em um conjunto de dados de treinamento, sendo este modelo usado para classificar as instâncias no conjunto de teste (NAKATANI *et al.*, 2012). A árvore de decisão C4.5, proposta por J. R. Quinlan em 1993, é uma extensão do algoritmo de classificação ID3 sendo uma melhoria deste, o que o tornou um dos mais clássicos algoritmos de árvore de decisão, trabalhando tanto com atributos discretos quanto contínuos. Também permite a utilização de atributos desconhecidos, representados por “?” (AMASYAL, 2008). É um algoritmo voraz que utiliza o método divisão e conquista para aumentar a capacidade de predição das árvores de decisão. Assim, sempre usa o melhor passo avaliado localmente, sem se preocupar se esse passo vai produzir a melhor solução, pega um problema e o divide em vários subproblemas sendo criadas sub-árvores entre a raiz e as folhas (CHUG *et al.*, 2013).

Como uma abordagem para apoiar a tomada de decisão baseada na computação, o *Data Mining* não é realmente uma tecnologia nova, mas empresta muitos algoritmos de estatísticas, inteligência artificial e outros campos. A consideração não se dá pelos algoritmos de mineração de dados, mas a ideia de obter automaticamente previsões e predições (MARKOV *et al.*, 2016).

O conhecimento de grandes bancos de dados é revolucionário. De acordo com Liu (2012), hoje em dia, grande quantidade de dados úteis e bem documentado, e as soluções de TI mais econômicas em termos de armazenamento e capacidade de processamento, tornam essa ideia mais realista: Novos algoritmos de centros de pesquisa e as universidades podem entrar em *software* comercial. Apesar de implementação da descoberta de conhecimento

totalmente automática a partir de banco de dados ainda está longe a idealidade esperada, este novo conceito e os contínuos esforços de pesquisa são a oportunidade para a revolução do futuro na tomada de decisões computacional.

O compartilhamento de grandes bancos de dados entre organizações também é de importância crescente em muitos projetos de Data Mining. Com muitos dados acumulados, muitas vezes as fontes têm que ser ligadas e agregadas de forma a melhorar a qualidade dos dados ou enriquecer os dados existentes com informações (CHRISTEN *et al.*, 2007).

Normalmente, aplicações práticas de *Data Mining* podem ser categorizadas de acordo com a tarefa que se pretende resolver. Além disso, os algoritmos de análise de dados utilizados nessas tarefas pressupõem a existência de um conjunto de dados na forma tabular (relacional). Esses conjuntos de dados, por sua vez, podem ser resultantes de uma fase prévia de pré-processamento para realização de transformações diversas sobre os dados obtidos de um ambiente transacional (GOLDSCHIMIT *et al.*, 2015).

Ao se falar sobre gestão de projetos ou gerenciamento de projetos, pretende-se enfatizar a importância desses instrumentos dentro das organizações, a necessidade deles e suas características. Os projetos, são parte integrante do processo decisório do planejamento organizacional, pois atuam como realimentadores e, assim, devem estar vinculados aos objetivos e metas da empresa. Ao verificar viabilidades, passa a ser um modelo da realidade (THIELMAM *et al.*, 2013).

Diferente dos tradicionalmente formais, elaborados somente para legitimar decisões já tomadas, os projetos inovadores estão, no bom sentido, revolucionando a área de empreendedorismo, encontrando na inovação o estímulo para a criação, desenvolvimento e obtenção dos benefícios advindos da inteligência do homem e das suas competências na construção de uma humanidade cada vez mais centrada no ecológico, no ético e no saudável, em prol do bem comum de seus membros (PAROLIN *et al.*, 2008).

Para Lopes (2017), as empresas já se conscientizaram de que promover a inovação tornou-se imprescindível para aumentarem sua competitividade e rentabilidade, mas que ainda precisam ter competência para inovar. Isto exige qualificação de seus recursos humanos, planejamento, equipes multidisciplinares, ferramentas diferenciadas de gestão, entre outros fatores.

Diante desta possibilidade aplicacional, o presente artigo objetiva apresentar uma análise de uma base de dados contendo informações de projetos do Polo de Inovação de Campos dos Goytacazes (PICG) utilizando o Data Mining, através do sistema WEKA e o algoritmo j48 como extrator de resultados dos mesmos.

2. METODOLOGIA

Para este artigo foi utilizada a metodologia de estudo de caso e aplicado o processo de descoberta de conhecimento em bancos de dados. As informações foram extraídas da base de dados do escritório de gerenciamento de projetos do Polo de Inovação de Campos dos Goytacazes (PICG). O estudo utilizou dados de fontes privadas como base para as análises, onde houve a definição de quais atributos poderiam ser divulgados. Entre as etapas predefinidas da técnica de descoberta de conhecimento em bancos de dados (DCBD) foram realizados os seguintes passos: seleção dos dados, análise preliminar dos dados, pré-processamento dos dados, transformação dos dados e data mining.

2.1. SELEÇÃO DOS DADOS

A etapa de seleção dos dados consistiu na extração de 164.732 registros de uma base de dados MySQL do sistema de gestão de projetos do polo de inovação do IFF. A partir da execução de uma query SQL foram definidos os atributos presentes na tabela abaixo:

Tabela 1: Dicionário de atributos de projetos.

Atributo	Descrição
project_id	Id do projeto
issue_id	Id da tarefa
tempo_gasto	tempo gasto em uma tarefa (horas)
tarefa_sem_tempo_gasto	Sim/Não caso a tarefa não tenha registro de horas cadastradas
tipo_tarefa	Tipo da tarefa
responsável_id	Id do usuário configurado como responsável de uma tarefa
perfil_responsável_tarefa	Perfil do usuário no projeto responsável pela tarefa
autor_id	Id do usuário que criou a tarefa
perfil_autor_tarefa	Perfil do usuário no projeto que criou a tarefa
autor_responsável_iguais	Sim/Não caso os usuários donos e responsáveis pelas tarefas sejam os mesmos
data_inicio	Data de início da tarefa
data_previsao_termino	Data de previsão de término da tarefa
Status	Situação da tarefa
qtde_dias_tarefa	Diferença em dias entre a data_inicio e a data_previsao_termino
issue_eap_id	Id do pacote da EAP
issue_meeting_id	Id de cadastro de reunião no projeto
eap_parent_id	Pacote pai da EAP
eaps_data_inicio	Data de início do pacote da EAP
eaps_data_final	Data final do pacote da EAP

Fonte: Elaboração Própria.

Diante dos atributos dispostos acima foi realizada uma análise para definição de quais os elementos melhor se apresentavam na geração de classificações a partir do algoritmo J48 no *software* WEKA.

2.2. ANÁLISE PRELIMINAR DOS DADOS

Para execução desta etapa foram respondidos alguns questionamentos acerca dos dados de modo a facilitar o processo de análise:

1 - Existem padrões "óbvios" nos dados que podem ser úteis?

R: Não, nenhum padrão óbvio no conjunto de dados foi percebido.

2 - Existem especialistas que entendem bem os dados e com quem você pode conversar?

R: Sim, eu conheço especialistas que entendem os dados e posso falar com eles.

3 - Há ausência de valores para alguns atributos?

R: Sim, existem valores em falta nos atributos (data_previsao_termino, issue_eap_id, issue_meeting_id, eap_parent_id, eaps_data_inicio e eaps_data_final).

Quando o conjunto de dados é carregado no Weka, no lado direito podemos observar informações sobre o atributo selecionado, tais como seu valor e quantas vezes uma instância no conjunto de dados tem um determinado valor, média e desvio padrão amadurece

para valores numéricos. Em alguns casos, a classe atributo pode ser mapeada pelo Weka como sendo do tipo numérico. No entanto, para alcançar os requisitos de execução do algoritmo de classificação, foi alterada a classe de numérico para nominal através de instrução SQL antes da entrada dos dados via arquivo CSV. Adicionalmente, outras instruções foram aplicadas durante a fase de pré-processamento de dados.

2.3. PRÉ-PROCESSAMENTO DOS DADOS

Baseado nas informações inicialmente levantadas foi feita a exclusão de alguns atributos para otimização dos resultados obtidos, os elementos selecionados foram:

issue_tracker_id, issue_role_id, issue_project_id, issue_status_id, issue_assigned_to_id, issue_priority, issue_done_ratio, time_entry_hours, issue_eap_id, issue_meeting_id, eap_parent_id, eaps_data_inicio, eaps_data_final

2.4. TRANSFORMAÇÃO DOS DADOS

Foram convertidos os atributos que estavam definidos como numérico para o tipo texto de modo a potencializar os resultados no *software* weka. O processo de transformação dos dados ocorreu através de *software* de planilha eletrônica antes da importação do arquivo CSV.

2.5. DATA MINING

Foi adotado o *software* WEKA para realizar a mineração de dados por se tratar de um *software* gratuito e de código aberto que usa a GNU *General Public License* (GPL). O *software* foi escrito na linguagem Java™ e contém uma GUI para interagir com arquivos de dados e produzir resultados visuais. Ele também tem uma API geral, assim é possível incorporar o WEKA, como qualquer outra biblioteca, a seus próprios aplicativos para executar tarefas de mineração de dados automatizadas no lado do servidor.

3. DISCUSSÃO E RESULTADOS

3.1 DESCRIÇÃO DO OBJETO DE ESTUDO: GERENCIAMENTO DE PROJETOS NO POLO DE INOVAÇÃO DO IFF

O Instituto Federal de Educação, Ciência e Tecnologia Fluminense, também chamado de forma abreviada de Instituto Federal Fluminense, ou simplesmente IFF, é uma instituição ensino que faz parte da Rede Federal de Educação Profissional, Científica e Tecnológica. O IFF foi criado mediante transformação do CEFET-Campos em um Instituto Federal de Educação, Ciência e Tecnologia. É um dos 38 Institutos Federais, criados em dezembro de 2008, pelo Governo Federal, a partir dos CEFETs, escolas técnicas e agrotécnicas federais.

A Sede do IFF está localizada na cidade de Campos dos Goytacazes. O Instituto Federal Fluminense possui *campi* presentes em quatro cidades do Norte Fluminense (Campos dos Goytacazes, Macaé, São João da Barra e Quissamã), quatro no Noroeste Fluminense (Bom Jesus do Itabapoana, Itaperuna, Santo Antônio de Pádua e Cambuci), um na Região dos Lagos (Cabo Frio) e dois na Região Metropolitana (Itaboraí e Maricá).

O IFF conta com grupos que atuam em linhas de pesquisa e que compõem os projetos isolados e programas especiais. O Instituto é bastante competitivo na captação de recursos nas agências de fomento para as atividades de Ciência e Tecnologia e disponibiliza seus recursos humanos e infraestrutura para projetos em cooperação e consultorias nas mais diversas áreas de atuação.

Em seu planejamento estratégico, ações estão sendo implementadas para viabilizar um modelo de gestão eficiente da pesquisa, visando maximizar recursos materiais, humanos e financeiros, de modo a ampliar essa atividade e aumentar sua aplicabilidade e inserção na sociedade.

De modo a atender as demandas de programas e projetos, o IFF está estruturado e dispõe de uma diretoria de específica para gerenciar as ações estão sendo implementadas na área de inovação. O controle das atividades de pesquisa é feito pela Pró-reitoria de Pesquisa, que verifica se os projetos estão sendo apreciados e aprovados em assembleia departamental, se os departamentos estão estabelecendo um banco de projetos, entre outros.

3.2 RESULTADOS

A execução do algoritmo de classificação J48 teve como desfecho escolhido o atributo tipo tarefa. Através da árvore de decisão disposta na Figura 1, quando o perfil responsável pela tarefa é o Pesquisador, este tende a criar tarefas para si com o intuito de registrar tempo de estudo no projeto. Já quando o autor da tarefa é um Gerente ou Coordenador delegando atividades para o Pesquisador, a tendência é de o tipo de tarefa seja do tipo Gestão, podendo ainda culminar em subtipos como Projeto, Planejamento e Documentação.

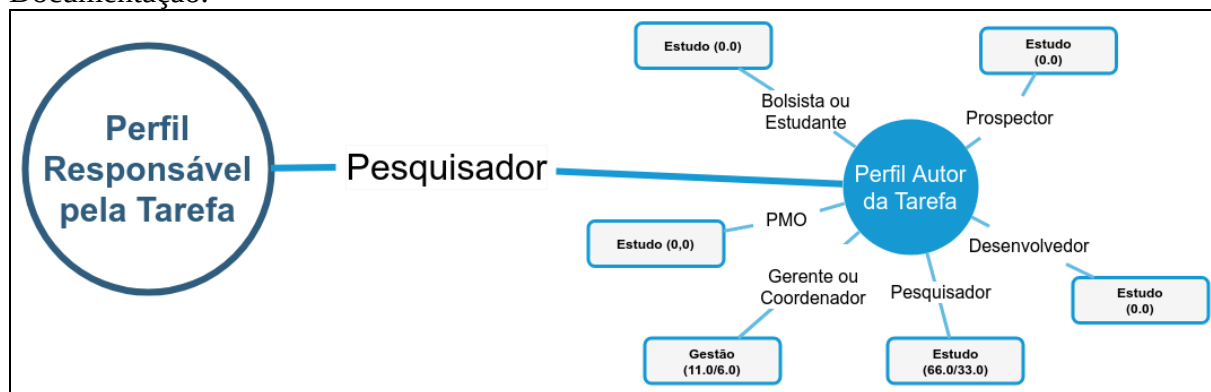


Figura 1. Árvore de decisão com informações do perfil Pesquisador.

Fonte: Elaboração Própria.

Bolsistas e/ou estagiários não criam tarefas, mas ficam responsáveis por atividades criadas diretamente pelos Gerentes e Coordenadores. As tarefas são do tipo Gestão e Execução e sua seleção varia conforme a quantidade de horas imputadas na tarefa. A árvore de decisão indica que caso o responsável pela tarefa registre mais de 10 horas de tempo gasto, a árvore de decisão indica que o tipo da tarefa tende a ser de execução. O Pesquisador é responsável por delegar tarefas de Implementação para Bolsistas/Estagiários (Figura 2).

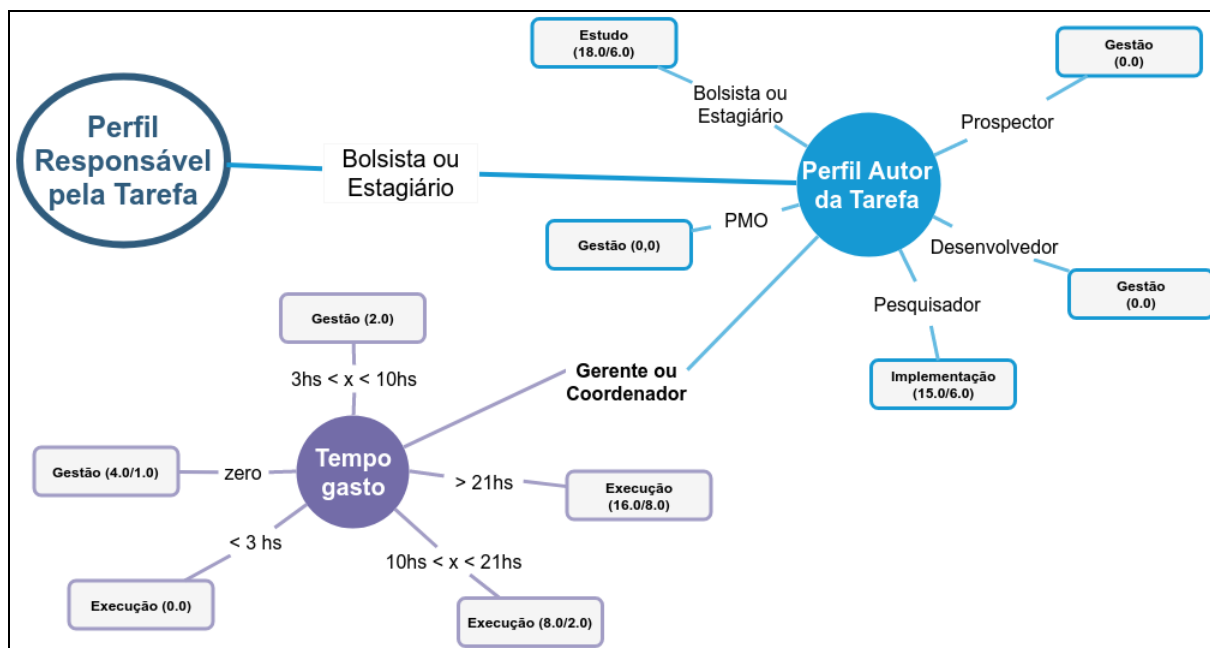


Figura 2. Árvore de decisão com informações do perfil Bolsista/Estagiário.

Fonte: Elaboração Própria.

Gerentes e Coordenadores possuem um montante de registro de atividades do tipo Reunião e Gestão. Para os registros de até 21hs, o tipo de tarefa indicada é Reunião. Para além deste patamar, tarefas de execução são apontadas como registros em blocos de horas. Tarefas de Gestão existentes não possuem registros de horas, indicando que tarefas com situação definida como “Fechada” não chegaram a ser executadas ou simplesmente não tiveram o tempo gasto registrado (Figura 3).

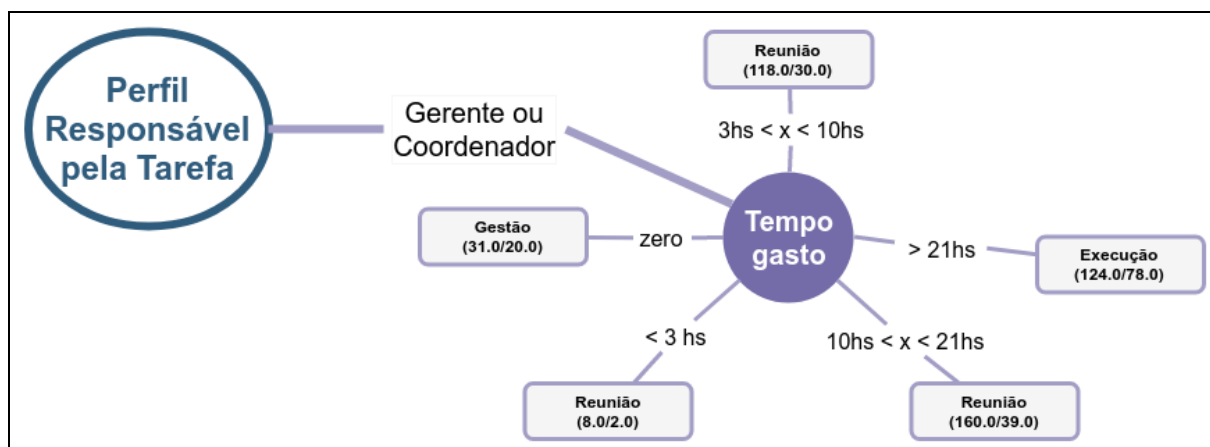


Figura 3. Árvore de decisão com informações do perfil Gerente/Coordenador.

Fonte: Elaboração Própria.

O perfil Desenvolvedor possui uma particularidade na árvore de decisão referente às tarefas existentes de Des. de *Software* que não possuem registros de horas. A quantidade é significativa e pode indicar que as atividades que foram anteriormente planejadas não foram executadas por algum motivo, seja por decisão da gestão, seja por decisão da equipe de desenvolvimento.

Os subtipos de tarefas Des. de *Software* incluem Funcionalidade, Refatoração, Requisito, Defeito, Teste e Análise. Na faixa de horas além de zero e menor que 21, o tipo de

tarefa é Implementação. Quando o registro do tempo gasto se encontra acima de 21 horas, são definidas como atividades de Gestão (Figura 4).

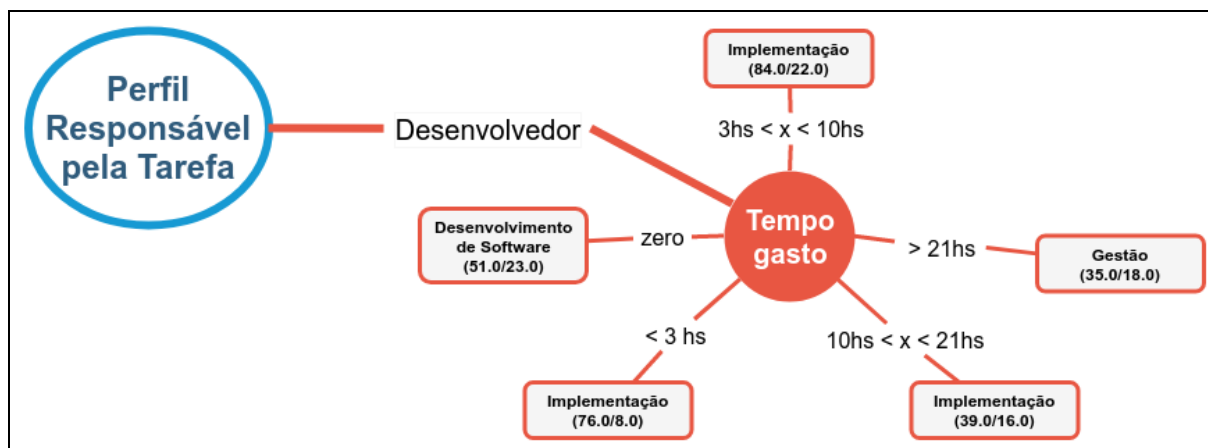


Figura 4. Árvore de decisão com informações para o perfil Desenvolvedor.
Fonte: Elaboração Própria.

Os resultados descritos nos diversos ramos da árvore de decisão são variados em termos de perfis de usuário, tipos de tarefas e tempo gasto. Enquanto uma parte da árvore aponta uma relação forte entre Gerentes/Coordenadores com Bolsistas/Estagiários, indicando proximidade na delegação de atividades do tipo Gestão, usuários Pesquisadores costumam organizar suas atividades em tarefas de própria autoria, com parcelas de atividades delegadas pelos Gerentes/Coordenadores.

Essa relação entre os recursos humanos identificada na prática pode ser comparada com a matriz de responsabilidade da estrutura analítica do projeto, artefato oriundo da fase de concepção do projeto. Para os projetos que não possuam a matriz, a identificação dos perfis e suas relações com horas gastas e tipos de tarefa poderão auxiliar na concepção de novos projetos.

Ainda sobre Gerentes/Coordenadores, a árvore indica alto volume de tarefas do Tipo Reunião, o que pode ser um indício de um alinhamento periódico com a equipe integrante do projeto. Já os Desenvolvedores, com liberdade para criar tarefas de desenvolvimento de sistemas, podem ter exagerado nas subtarefas, muitas delas fechadas sem registro de horas. Cabe ressaltar que a não exclusão das tarefas não desenvolvidas permitiu esse tipo de análise. Uma mudança no projeto pode ter forçado uma revisão das atividades previstas.

4. CONCLUSÃO

Este trabalho utilizou das técnicas de mineração com o auxílio do *software* WEKA para seleção, classificação e agrupamento de atributos de projetos da base de dados do Polo de Inovação de Campos dos Goytacazes.

Após a seleção dos atributos e geração das árvores de decisão, foi possível interpretar e descrever as relações entre os atributos de tempo gasto na tarefa, tipo da tarefa, perfil do responsável pela tarefa e perfil do autor da tarefa através da execução do algoritmo de classificação J48 no *software* Weka. Os perfis Desenvolvedor, Coordenador/Gerente e Pesquisador tiveram suas características de uso do sistema e relação com outros perfis reveladas. Da mesma forma com o Perfil Bolsista/Estagiário.

Ao identificar correlação entre os dados, foi possível gerar conhecimento visando otimizar a gestão de projetos futuros. Os resultados podem ser comparados com artefatos de projetos existentes, como a matriz de responsabilidade do projeto, ou ainda servir de entrada para novos projetos do escritório.

Novas descobertas de conhecimento na mesma base de dados podem ser experimentadas, pois novos cenários podem ser aplicados selecionando outros atributos como desfecho da árvore de decisão. A base é extensa e podem ser incluídos novos algoritmos, como o K-means.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] AMASYAL, M. F. **New Machine Learning Methods and Drug Design Applications**, Ph.D. Thesis, Y_ld_z Technical University _Istanbul, Turkey. 2008.
- [2] CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Ver. Adm. Pública. V. 42, n. 3, p. 495-528. 2008.
- [3] CHRISTEN, P.; GOISER, K. **Quality and complexity measures for data linkage and deduplication**. In F. Guillet and H. Hamilton, editors, *Quality Measures in Data Mining*, v. 43 of *Studies in Computational Intelligence*. Springer. 2007.
- [4] CHUG, A.; DHALL, S. *Software defect prediction using supervised learning algorithm and unsupervised learning algorithm*. IET Conference Publications. (647 CP), p. 173-179. 2013.
- [5] GOLDSCHIMIT, R.; BEZERRA, E. **Data Mining: conceitos, técnicas, algoritmos, orientações e aplicações**. Elsevier Editora Ltda. Rio de Janeiro. 2ª Ed. 2015.
- [6] GUIDINI, M. P.; RIBEIRO, C. H. C. **Utilização da biblioteca TerraLib para algoritmos de agrupamento em Sistemas de Informações Geográficas**. GeoInfo. 2014. Disponível em: <http://www.geoinfo.info/proceedings_geoinfo2006.split/paper21.pdf>. Acesso em 19 de Nov. de 2018.
- [7] LAUDON. K.; LAUDO, J. *Sistemas de Informações Gerenciais: Fundamentos da inteligência de negócios/gestão da informação e de banco de dados*. 9º ed. São Paulo: ABDR. p. 159. 2011.
- [8] LIBRELOTTO, S. R.; MOZZAQUATRO, P. M. **Análise dos algoritmos de mineração J48 e a priori aplicados na detecção de indicadores da qualidade de vida e saúde**. Revista Interdisciplinar de Ensino, Pesquisa e Extensão, vol.1 nº1, p. 26-37. 2013.
- [9] LIU, Y. **A framework of Data Mining application process for credit scoring**. Institut für Wirtschaftsinformatik, Abteilung Wirtschaftsinformatik II, Georg-August-Universität Göttingen. 2012.
- [10] LOPES, B. **Gestão do Conhecimento e Inovação**. Edit. Poisson. Vol. 1. Belo Horizonte/MG, 2017. Pág. 1-201.
- [11] MARKOV, Z.; RUSSELL, I. **An Introduction to the WEKA Data Mining System**. ITiCSE'06. Bologna, Italy. p. 367-368. 2016.

- [12] NAKATANI, T.; KONDO, N.; SHIROGANE, J.; KAIVA, H.; HORI, S.; KATAMINE, K. **Toward the decision tree for inferring requirements maturation types**. IEICE Transactions on Information and Systems, v. 95-D, n. 4, p. 1021-1030. 2012.
 - [13] PAROLIN, S. R. H.; MORAES, D. C.; OLIVEIRA, H. C.; ZANON, S. L. M.; NARDELLI, T. **Elaboração de Projetos Inovadores na Educação Profissional**. Editora SENAI/ PR Departamento Regional. 2ª Edição. Vol. 1. . Curitiba: SESI/SENAI/PR, 2008.
 - [14] RODRIGUES, F. A.; AMARAL, L. R. Aplicação de Métodos Computacionais de Mineração de Dados na Classificação e Seleção de Oncogenes Medidos por Microarray. Revista Brasileira de Cancerologia. V. 58, n. 2. p.241-249. 2012.
 - [15] SILVA, N.; SOARES, J.; SHAH, V.; SANTOS, M. Y.; RODRIGUES, H. **Anomaly Detection in Roads with a Data Mining Approach**. Procedia Computer Science, v. 121, p. 415-422. 2017.
- THIELMAM, R.; SILVA, M. S. C. A Importância da Gestão de Projetos para Inovação em Empresas: o Caso do Setor de Energia. SEGET. Gestão e Tecnologia para a Competitividade. 2013.