

UTILIZANDO O KDD PARA CARACTERIZAÇÃO DO PERFIL DE SUBMISSÕES DE PROJETOS EXTENSÃO

Fausto Bernard Melo Soares

fausto.soares@ifs.edu.br

Otacílio Joaquim Rodrigues Cerqueira

otaciliocerqueira@gmail.com

Kaic de Oliveira Barros

kaicbarros@gmail.com

Maria Ylnah de Melo Santos

ylnahmelo@gmail.com

Willian Santos Farias

williansantos_141@hotmail.com

Resumo: Pensando no investimento feito pelo país e na condução da atividade de pesquisa e extensão, serão utilizadas técnicas de aprendizagem de máquina na busca por conhecimento em dados de submissão dos projetos de pesquisa e extensão do IFS ao longo dos anos de 2012, 2013, 2014, 2015, 2016 e 2017. Dessa forma será possível descobrir relações entre variáveis de que levaram a aprovação ou não dos projetos de pesquisa e extensão submetidos, bem como uma evolução ao longo dos anos dos programas das submissões realizadas para projetos de pesquisa e extensão do Instituto Federal de Sergipe

Palavras-Chave: Mineração de dados, KDD, x-means, k-means.

INTRODUÇÃO

Mediante o ascensão dos últimos anos da atividade de pesquisa e extensão desenvolvidas no IFS faz-se necessário um apoio à tomada de decisão da PROPEX no que diz respeito ao investimento nessas atividades, bem como suas ações de fomento das mesmas na sociedade. Assim, técnicas de aprendizagem de máquina podem ser lançadas em função de uma caracterização da pesquisa e extensão produzidas no IFS através dos projetos submetidos e executados nesta instituição. Dessa forma, algoritmos de classificação e

agrupamento de dados podem ser úteis no apoio a esta pró-reitoria para tomar futuras decisões quanto às suas ações na promoção da pesquisa e extensão no IFS.

MATERIAL E MÉTODOS

Inicialmente foi feito um levantamento dos programas de pesquisa e extensão da instituição, bem como o quantitativo de projetos submetidos, executados, cancelados no IFS em programas próprios ao longo dos anos de 2012 a 2017.

A seguir os dados foram tratados e limpos sendo subdivididos em grupos de acordo com o ano de submissão, área de conhecimento cuja qual o projeto é submetido, e grupo que submeteu o projeto (docentes e TAE's).

Foram então empregadas técnicas de aprendizagem máquina nos dados tratados na busca de conhecimento. Os resultados da aplicação de tais técnicas foram verificados para escolha do algoritmo que apresentou a melhor obtenção de conhecimento dentre eles. Logo após, características foram extraídas da mineração de dados efetuada nos grupos e validadas junto aos gestores de pesquisa e extensão do IFS.

Após a extração dos dados utilizou-se dos itens de avaliação dos artigos (RT – Relevância

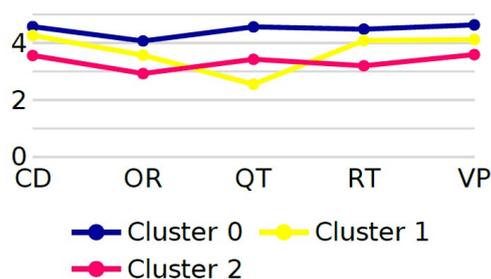
do Tema, Qualidade Técnico Científica QTC, Originalidade OR, Apresentação e Estrutura AE, Fundamentação Teórica FT, Viabilidade de Execução do Projeto VP, Contribuição para o Desenvolvimento CD) para aplicação dos algoritmos de aprendizagem de máquina não supervisionada x-means e k-means.

RESULTADOS E DISCUSSÃO

De acordo com a figura a seguir verifica-se que o quantitativo dos projetos de extensão da instituição por campus cresceu ao longo dos anos de 2012 a 2017.



Para os projetos de extensão o gráfico a seguir, criado com dados de submissão dos anos de 2013 a 2017 observa-se uma maior pontuação para os itens CD e VP, dentro de todos os grupos criados.



Fica claro também a existência de 3 grupos distintos resultante da mineração de dados: o cluster 0 onde estão os projetos submetidos que foram aceitos (em sua maioria), o cluster 1 com projetos na faixa de transição (aceitos e excedentes) e o cluster 2 com projetos com menores notas.

Observa-se dentro do grupo cluster 0 que as notas mantêm uma constância entre

os atributos. Fato que não ocorre no cluster 1, que por se tratar de um grupo mais heterogêneo apresentou a nota de QT a menor dentre os grupos gerados, proporcionando possivelmente a reprovação dos projetos que encontram-se nesse grupo.

E por fim, o cluster 3 apresenta um padrão constante entre as notas dos atributos.

CONCLUSÕES

O trabalho produzido ao longo do ano procurou caracterizar as submissões dos projetos através dos grupos gerados pelo algoritmo de agrupamento de dados k-means. Fornecendo assim com base nas características dos perfis obtidos subsídios a gestão para desenvolvimento da atividade de extensão na instituição.

REFERÊNCIAS

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery: an overview. In: *Advances in knowledge Discovery and data mining*, AAAI Press/The MIT Press, MIT, Cambridge, Massachusetts, and London, England, 1996

RUSSEL, S.; NORVIG, P.; *Inteligência Artificial*. 2 Ed. São Paulo: Editora Campus, 2004.

KUMAR, Vipin; STEINBACH, Michael; TAN, Pang. *Introdução ao datamining*. Rio de Janeiro: Ciência Moderna LTDA, 2009.