



Avaliação de Ferramentas de Mineração de Dados: Uma Abordagem Com o Modelo TAM

Gleidison Santos Hora - gleidisonhora@hotmail.com

Jislane Silva Santos Menezes - jislane.menezes@academico.ifs.edu.br

Gilson Pereira dos Santos Júnior - gilson.junior@academico.ifs.edu.br

Almerindo Nascimento Rehem Neto - almerindo.rehem@academico.ifs.edu.br

Curso de Sistemas de Informação – Instituto Federal de Sergipe (IFS)

Divisão Tecnologia de Informação e Comunicação – Tema 01–Gestão da Informação e suas TI/Sis, Analytics, BigData

RESUMO

A mineração de dados é a área da ciência da computação que visa analisar grandes volumes de dados com o objetivo de descobrir informações úteis, ocultas e não triviais para as organizações. Para apoiar este processo de descoberta do conhecimento existem inúmeras ferramentas de mineração de dados como, por exemplo, *Weka*, *Knime*, *RapidMiner* e *Tanagra*. Entretanto, não se sabe como cientistas de dados, em formação, aceitam e utilizam tais ferramentas. Assim, o objetivo deste trabalho é avaliar a aceitação das ferramentas de mineração de dados (*Weka*, *Knime*, *RapidMiner* e *Tanagra*) através do modelo de aceitação de tecnologia (*Technology Acceptance Model* – TAM) com os discentes da disciplina de Mineração de Dados do curso de Sistemas de Informação de uma Instituição Pública. Para tanto, foi aplicado um questionário com 23 questões sobre Variáveis Externas (VE), Facilidade de uso percebido (FUP), Utilidade Percebida (UP) e Intenção Comportamental (IC) com 12 discentes do referido curso, durante o segundo semestre de 2016. A pesquisa mostrou as ferramentas analisadas são fáceis de serem usadas, com exceção da *Tanagra* que obteve 48,9% de aceitação neste quesito. Todas as ferramentas mostraram-se úteis para realizar as tarefas de mineração de dados. Além disso, os participantes demonstraram interesse em utilizar as tarefas diárias de mineração de dados, em particular o *RapidMiner*, *Weka* e *Knime*.

Palavras-chave: Mineração de dados; Ferramentas de mineração de dados; Modelo de aceitação de tecnologia.

INTRODUÇÃO

A competição entre organizações, devido à globalização, motiva o uso de ferramentas tecnológicas dentro do ambiente empresarial, a fim de obter vantagem competitiva e auxiliar no processo de tomada de decisão. Sabe-se que a ocorrência de falhas em sistemas de Tecnologia de Informação (TI) pode levar até mesmo organizações bem sucedidas a falência (DWIVEDI et al., 2014). Ou seja, uma TI mal gerenciada é um risco para qualquer negócio. Assim, surge a governança de TI, uma área da Informática responsável por garantir que a TI da empresa sustente e estenda as estratégias e objetivos da organização dentro dos seus processos e estruturas organizacionais (ITGI, 2007). É uma área de responsabilidade da alta administração, capaz de alinhar os requisitos de negócio a soluções informáticas de apoio,



garantindo continuidade de serviços e minimização da exposição do negócio aos riscos de TI. Um dos fatores que motivam a governança de TI é a integração tecnológica de softwares (FERNANDES; ABREU, 2014). Ela é caracterizada por tipos de sistemas como cadeias de suprimento, gestão da empresa e chão de fábrica, aplicações ERP (*Enterprise Resource Planning*), gestão de clientes CRM (*Customer Resource Management*) entre outros. O uso de tais sistemas, aliados a aplicações de *data warehouse*, *data mining* e inteligência organizacional, permitem a integração da gestão estratégica com a gestão tática e operacional das empresas, o que auxilia no processo de tomada de decisão.

A partir dos anos 90, houve a necessidade das organizações serem capazes de analisar, planejar e reagir às mudanças dos negócios o mais rápido possível. Por meio de grandes bases de dados transacionais, acumuladas ao longo dos anos, informações valiosas poderiam ser extraídas. A extração deste conhecimento motivou o surgimento do *Data Warehouse*, armazéns de dados que integram informações a partir dos banco de dados corporativos e fontes de dados externas a empresa, e de um arquitetura para maximizar a eficiência no processo de coleta, integração e análise dados, bem como no compartilhamento e extração de conhecimento, denominada de *Business Intelligence* (BI) (SASSI, 2010).

Os componentes de BI utilizam técnicas de mineração em bancos de dados para encontrar relações insuspeitas e resumir os dados de maneira compreensível e útil para o proprietário dos mesmos (HAND; MANNILA; SMYTH, 2001). Neste contexto, a mineração de dados (MD) é um processo de extração de informações ocultas e úteis para tomada de decisão a partir da classificação, da associação e do reconhecimento de padrões em grandes bases de dados (GOEBEL; GRUENWALD, 1999). A MD é aplicada em empresas, pesquisas e indústrias de diversas áreas de conhecimento com objetivo de melhorar processos ou analisar tendências. Na educação, por exemplo, pode ser usada para avaliação de estilos de aprendizagem (WEST, 2012). Já na área contábil, ela pode ser empregada na detecção automatizada de fraude (PHUA et al., 2005).

O interesse das organizações em descobrir informações úteis a partir de dados transacionais incentivou a construção de ferramentas e suítes para minerá-los, abrangendo todas as fases do processo de conhecimento, desde o pré-processamento de dados até a visualização (FAYYAD et al., 1996). Por outro lado, este interesse proporcionou uma demanda por profissionais que dominassem tal área de conhecimento e as ferramentas disponíveis no mercado, motivando o surgimento de cursos, disciplinas e pesquisas no meio acadêmico (BRITO et al., 2014). É comum, durante a formação destes profissionais, a utilização de ferramentas e suítes profissionais, tradicionalmente empregadas no mercado de trabalho, a fim de conhecer, entender e praticar as fases do processo de MD. Entretanto, há uma carência de estudos na literatura sobre o nível de aceitação de tais ferramentas pelos estudantes durante o processo de aprendizagem.

Assim, essa pesquisa busca mensurar o nível de aceitação de ferramentas de mineração de dados com base no modelo de aceitação de tecnologia TAM (*Technology Acceptance Model*) dentro do ambiente educacional, com intuito de selecionar qual ferramenta possui melhor usabilidade e utilidade para os estudantes durante a sua formação em Mineração de Dados. O TAM é considerado um dos modelos mais influentes e mais amplamente utilizado pelos pesquisadores para descrever a aceitação tecnológica pelos indivíduos, estudando a influência de fatores humanos na adoção das mesmas (DILLON; MORRIS, 1996; SILVA, 2005).

Para realizar a avaliação, foi utilizado um questionário com um grupo de alunos do curso de sistemas de informação do Instituto Federal de Sergipe (IFS), campus Lagarto, com o objetivo de mensurar o nível de aceitação das ferramentas avaliadas pelos entrevistados.



Com os dados obtidos nesta pesquisa procedeu-se um comparativo estatístico entre as ferramentas. Foram avaliadas quatro ferramentas de mineração de dados gratuitas, selecionadas após pesquisa bibliográfica em bases de periódicos: *Weka*, *RapidMiner*, *Tanagra* e *Knime* (BOSCARIOLI; VITERBO; TEIXEIRA, 2014), (RAMAMOHAN et al., 2012).

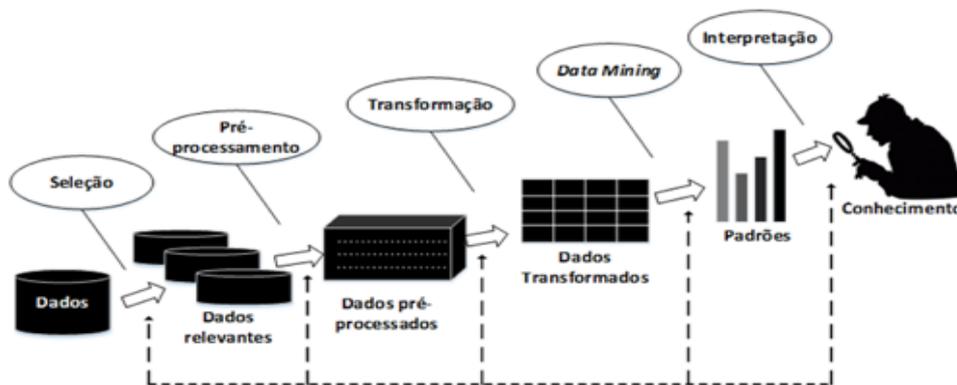
O PROCESSO DE MINERAÇÃO DE DADOS

O KDD (*Knowledge Discovery in Databases*) é uma área dinâmica e evolutiva que envolve outras áreas do conhecimento – Estatística, Inteligência Artificial e Banco de Dados. Este conjunto se constitui em uma tecnologia capaz de automatizar processos de busca por padrões úteis, também chamando de conhecimento, que podem estar oculto dentro de um amontoado de dados.

Para Fayyad et al. (1996) a descoberta de conhecimento é um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis. O processo não é trivial já que alguma técnica de busca ou inferência é envolvida, ou seja, não é apenas um processo de computação direta. Os padrões descobertos devem ser válidos com algum grau de certeza, novos, potencialmente úteis e compreensíveis.

O processo de KDD descrito anteriormente é composto por uma série de fases definidas, que são a seleção, o pré-processamento e a limpeza, a transformação, a mineração de dados, a interpretação/avaliação. Essa sequência compreende o ciclo que o dado percorre até se transformar em conhecimento útil, conforme a Figura 1.

Figura 1 – Etapas do processo KDD



Fonte: Adaptado de Fayyad et al. (1996).

Na fase de seleção são selecionados os atributos de interesse para a análise. Durante o pré-processamento os dados são tratados removendo informações faltantes ou inconsistentes. Alguns algoritmos de mineração necessitam que os dados estejam em formatos apropriados utilizando, por exemplo, operações de agregação. Neste momento é feita a transformação. É na etapa de mineração de dados (*Data Mining*) onde acontece a exploração e análise, de forma automática ou semi automática, de grandes bases de dados com objetivo de descobrir padrões e regras (FAYYAD et al., 1996). Ela dispõe de uma série de tarefas e técnicas que se adequa a cada tipo de situação exposta. As tarefas correspondem às especificações do que se deseja encontrar nos dados, ou seja, as categorias de padrões, enquanto as técnicas dizem respeito às especificações dos métodos que garantem como descobrir estes padrões. É



importante salientar que as tarefas utilizam diferentes técnicas para atingir seu objetivo. Como também uma técnica pode adotar diversos tipos de algoritmos para realizar uma tarefa.

As tarefas de Mineração de Dados mais conhecidas são: agrupamento, classificação, associação e regressão (PINTO, 2010). Estas tarefas farão parte da análise do questionário relacionada ao conceito utilidade.

FERRAMENTAS DE MINERAÇÃO

Para Rezende (2005), a área de extração de conhecimento de base de dados tem progredido nas organizações, ocorrendo um estímulo para o desenvolvimento de ferramentas capazes de facilitar a extração do conhecimento por meio de uma variedade de técnicas, com diferentes tipos de algoritmos voltados para a Mineração de Dados. Nessa pesquisa, busca-se avaliar 4 ferramentas de mineração de dados: *Weka*, *RapidMiner*, *Tanagra* e *Knime*.

Waikato Environment for Knowledge Analysis (WEKA) é uma ferramenta de KDD *open-source*, escrita em *Java*, que contempla uma série de algoritmos de diversas técnicas de mineração de dados, desenvolvida pela Universidade de *Waikato*, na Nova Zelândia que disponibiliza algoritmos para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização de dados provenientes de uma base de dados ou integradas a uma aplicação *Java* (WEKA, 2017).

Já o *RapidMiner* é um ambiente de desenvolvimento *open-source*, desenvolvido por pesquisadores da Universidade Técnica de *Dortmund*, que fornece uma interface gráfica do usuário (GUI) e *scripts* baseados em XML para realização de processos de *data mining*, *text mining*, análise de dados e *business intelligence*. A ferramenta possui centenas de algoritmos de preparação de dados e aprendizagem de máquina para suportar qualquer projeto de mineração de dados (RAPIDMINER, 2017).

O *Tanagra* é um programa gratuito de mineração de dados para fins acadêmicos e de pesquisa, desenvolvido em *Delphi* por pesquisadores da Universidade de *Lyon*, na França. O objetivo principal do projeto *Tanagra* é fornecer aos pesquisadores e estudantes um software de mineração de dados fácil de usar (TANAGRA, 2017). Conta com processos de classificação supervisionada e não-supervisionada, tais como clusterização, estatísticas parametrizadas e não parametrizadas e regras de associação. A ferramenta disponibiliza uma construção interativa e visual de árvores de decisão.

O *Knime* é uma ferramenta de mineração criada pela Universidade de *Konstanz*, na Alemanha, cujo propósito inicial era analisar os dados do segmento de mercado farmacêutico. Desde então, tornou-se uma plataforma geral analítica de dados, relatórios e integração. Escrito em *Java*, o *Knime* aproveita a capacidade de extensão do módulo do Eclipse por meio do uso de *plug-ins* (KNIME, 2017).

MODELO DE ACEITAÇÃO DE TECNOLOGIA

Estudos e pesquisas sobre aceitação de tecnologia, por indivíduos e organizações, têm apresentado um forte crescimento a partir da segunda metade da década de 90. Estes estudos são realizados com o intuito de buscar melhorias constantes, e identificar fatores intrínsecos e extrínsecos envolvidos nas decisões, intenções e satisfação dos indivíduos, quanto à aceitação e ao uso da tecnologia da informação, por meio de vários testes e métodos de avaliação (DIAS et al., 2003; SILVA, 2005; LÖBLER; VISENTINI; VIEIRA., 2006; VENKATESH et al., 2003).



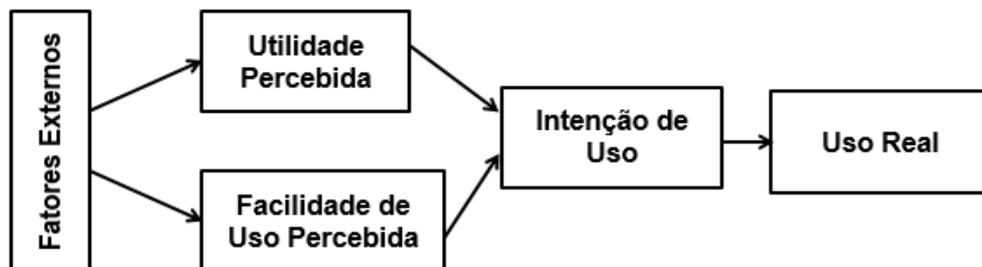
Vários modelos teóricos têm sido desenvolvidos e aplicados, entretanto dentre as diversas teorias, o modelo de aceitação de tecnologia, *Technology Acceptance Model* (TAM) é considerado um dos mais influentes e mais amplamente utilizado pelos pesquisadores (DILLON; MORRIS, 1996; SILVA, 2005).

O TAM foi proposto por Davis (1989) sendo uma adaptação do modelo da Teoria da Ação Raciocinada (TRA). Ele foi projetado para compreender a relação causal entre variáveis externas de aceitação dos usuários e o uso real do computador, buscando entender o comportamento do usuário através do conhecimento da utilidade e da facilidade de utilização percebida por ele (DAVIS, 1989).

Segundo o modelo, as pessoas tendem a usar ou não uma tecnologia com o objetivo de melhorar seu desempenho no trabalho – utilidade percebida. Porém, mesmo que essa pessoa entenda que uma determinada tecnologia é útil, sua utilização poderá ser prejudicada se o uso for muito complicado, de modo que o esforço não compense o uso – facilidade percebida.

A Figura 2 sugere que os indivíduos usarão a tecnologia se acreditarem que este uso fornecerá resultados positivos, focalizando-se na facilidade de uso percebida (*perceived ease of use*) e na utilidade percebida (*perceived usefulness*). Assim, o TAM normalmente é utilizado para entender o porquê que o usuário aceita ou rejeita a tecnologia de informação e como melhorar a aceitação, oferecendo, desse modo, um suporte para prever e explicar a aceitação.

Figura 2 – Modelo de Aceitação de Tecnologia (TAM)



Fonte: Davis et al., 1989.

METODOLOGIA

Nesse estudo, decidiu-se adotar a linha de pesquisa quantitativa, pois o modelo TAM é eminentemente quantitativo conforme vasta revisão bibliográfica sobre o assunto (DAVIS; BAGOZZI; WARSHAW, 1989). Além da vertente quantitativa, a pesquisa também se caracterizou como exploratória e descritiva, com levantamento e análise bibliográfica sobre o tema. Ela operacionaliza a medição dos seguintes construtos do modelo TAM: variáveis externas, facilidade de uso percebida, utilidade percebida e intenção comportamental de uso.

Um questionário *online* foi utilizado como instrumento de coleta de dados e sua estrutura foi construída a partir da revisão da literatura de questionários anteriores utilizados em estudos similares como (CARVALHO, 2015) e (MENEZES; RAMOS; SOARES, 2016). O questionário foi desenvolvido por meio do serviço *GoogleForms*, onde foi gerado um *link* e disponibilizado para os entrevistados durante o treinamento preliminar realizado em sala de aula.

O questionário foi submetido a um grupo de 12 (doze) alunos do curso de Sistemas de



Informação do IFS campus Lagarto, que experimentaram a versão 3.8.1 do *Weka*, a 5.3 do *RapidMiner*, a 1.4.50 do *Tanagra* e a 3.3.2 do *Knime* durante X dias. Para cada ferramenta analisada, os avaliadores responderam um questionário com 23 questões referentes a 4 variáveis em análise: Variáveis Externas (VE), Facilidade de uso percebida (FUP), Utilidade Percebida (UP) e Intenção Comportamental (IC). Os dados coletados foram submetidos à análise descritiva e inferencial envolvendo média (\bar{x}), desvio padrão (σ), mediana (Md), moda (Mo) e contagem de frequência. Para mensurar as variáveis, optou-se por utilizar uma escala *Likert* de cinco pontos para medir as atitudes percebidas dos participantes. A escala adotou os valores (1) Discordo Totalmente, (2) Discordo Parcialmente, (3) Neutro, (4) Concordo Parcialmente, (5) Concordo Totalmente.

A seguir, na Tabela 1, apresentam-se as variáveis e as questões utilizadas como afirmativas, nos questionários aplicados nesta pesquisa.

Tabela 1 – Construtos x Variáveis

Variáveis Externas (VE)	
VE1	Possui conhecimento prévio para utilização da ferramenta em análise?
VE2	O treinamento foi adequado para se fazer uso da ferramenta de mineração de dados?
VE3	O treinamento me deu confiança para utilizar a ferramenta?
Facilidade de Uso Percebida (FUP)	
FUP1	É fácil usar a ferramenta?
FUP2	A ferramenta me apresenta onde estou e como chegar aonde eu quero?
FUP3	A ferramenta é clara de ser compreendida?
FUP4	É fácil encontrar a informação que desejo na ferramenta?
FUP5	A ferramenta é intuitiva de modo que antes de clicar em algum botão eu sei a ação dele?
FUP6	A ferramenta facilita a lembrança de como realizar atividades?
FUP7	Utilizar a ferramenta é agradável?
Utilidade Percebida (UP)	
UP1	A ferramenta agrega valor na relação ensino aprendizagem?
UP2	A ferramenta disponibiliza visualização de relatórios na saída dos dados?
UP3	A ferramenta disponibiliza visualização de gráficos na saída dos dados?
UP4	A ferramenta disponibiliza operação de classificação?
UP5	A ferramenta disponibiliza operação de agrupamento?
UP6	A ferramenta disponibiliza operação de associação?
UP7	A ferramenta disponibiliza operação de regressão?
UP8	A ferramenta melhora meu desempenho durante as atividades?
UP9	A ferramenta produz os resultados que espero?
Intenção Comportamental de Uso (IC)	
IC1	Recomendo a utilização da ferramenta?
IC2	Estou motivado a utilizar a ferramenta?
IC3	Utilizar a ferramenta atendeu as minhas expectativas?
IC4	A ferramenta é apropriada para estudar mineração de dados?



RESULTADOS

As respostas coletadas foram organizadas por ferramenta e para cada ferramenta foram analisados os construtos Variáveis Externas (VE), Facilidade de Uso Percebida (FUP), Utilidade Percebida (UP) e por fim a Intenção Comportamental (IC).

Para a análise das respostas, os valores que representam 1 e 2 foram reunidos e substituídos, por indicativo de discordância (d), 3 para neutralidade (n) e 4 e 5 por indicativo de concordância (c). Em resultados onde a média está entre 4 e 5 o resultado é considerado positivo. Entre 1 e 3, negativo e na variação do 3, um resultado neutro, não concorda ou discorda.

As Tabelas 2, 3 e 4 apresentam os resultados para os construtos VE, FUP e UP. A quantidade total de respostas (N) é igual a 48 para as 4 ferramentas avaliadas pelos 12 entrevistados. Nas tabelas, “ σ ” significa o desvio padrão, “md” a mediana, “ \bar{x} ” a média, “Mo” a moda, “d” indicativo de discordância, “n” neutralidade e “c” indicativo de concordância. O resultado final do construto é dado pela razão de somatório dos seus itens sobre a soma do total de respostas N dos itens.

Tabela 2 – Estatística Descritiva de VE (N=48)

Variáveis Externas (VE) – questões de 1 a 3							
	c	n	d	\bar{x}	Md	σ	Mo
VE1	42	1	5	4,3	5	1,16	5
VE2	33	7	8	3,7	4	0,97	4
VE3	32	6	10	3,7	4	1,08	4

Neste trabalho observou-se que as variáveis externas influenciaram positivamente as atitudes pessoais dos entrevistados. Para V1, 42 respostas demonstraram que os entrevistados já tinham algum conhecimento prévio por ter usado alguma das ferramentas. 66,7% afirmaram que o treinamento aplicado foi adequado (VE3). No geral, 74,3% dos respondentes assinalaram positivamente (c) para o construto VE (VE1+VE2+VE3), como mostra a Tabela 2.

Tabela 3 – Estatística Descritiva de FUP (N=48)

Facilidade de Uso Percebida (FUP) – questões de 1 a 7							
	c	n	d	\bar{x}	md	σ	Mo
FUP1	31	2	15	3,5	4	1,19	4
FUP2	33	1	14	3,6	4	1,33	4
FUP3	32	0	16	3,5	4	1,38	4
FUP4	39	1	8	3,9	4	1,19	4
FUP5	28	1	19	3,4	4	1,42	5;4;2
FUP6	37	0	11	3,8	4	1,33	4
FUP7	32	4	12	3,7	4	1,25	5

Ao analisar os dados das Tabelas 3, 4 e 5 que apresentam as variáveis do modelo TAM, verifica-se que os participantes da pesquisa demonstraram uma boa aceitação em relação à maioria dos itens do questionário (md=4). Isso significa que eles acreditam que os aplicativos são fáceis de utilizar, são úteis e que melhoram seu desempenho. A avaliação dos entrevistados mostrou uma concordância sobre o total de 69% para FUP, 88,4% para UP e



66,7% para IC. Os itens da variável utilidade percebida mostraram os níveis mais elevados de aceitação.

Outra análise também foi realizada sob uma perspectiva por ferramenta de mineração. A Figura 3 apresenta um comparativo quanto aos construtos VE, FUP, UP e IC por software.

Tabela 4 – Estatística Descritiva de UP (N=48)

Utilidade Percebida (UP) – questões de 1 a 9							
	c	n	d	\bar{x}	Md	σ	Mo
UP1	41	1	6	4,2	4,5	1,09	5
UP2	47	0	1	4,7	5	0,57	5
UP3	37	0	11	3,9	5	1,48	5
UP4	45	0	3	4,6	5	0,77	5
UP5	48	0	0	4,8	5	0,35	5
UP6	48	0	0	4,8	5	0,37	5
UP7	46	1	1	4,8	5	0,58	5
UP8	31	6	11	3,8	4	1,17	5
UP9	39	1	8	4,1	4,5	1,22	5

Tabela 5 – Estatística Descritiva de IC (N=48)

Intenção Comportamental (IC) – questões de 1 a 4							
	C	n	d	\bar{x}	md	σ	Mo
IC1	36	0	12	3,9	5	1,55	5
IC2	27	4	17	3,4	4	1,56	5
IC3	29	1	18	3,5	4	1,63	5
IC4	36	0	12	3,9	5	1,55	5

A Tabela 6 apresenta os resultados gerais por ferramenta, indicando nas colunas com a letra “c” o total de resultados em concordância e “d” os resultados em discordância explicados na seção Metodologia.

Analisando o construto VE observou-se que o índice de concordância da ferramenta *Knime* obteve o maior percentual com aproximadamente 83,3%. A *Weka* aparece em segundo lugar com 75% em nível de aceitação, seguida do *RapidMiner* e *Tanagra* que apresentaram o mesmo desempenho, 69,4%.

Tabela 6 – Resultado da concordância e discordância dos construtos por ferramenta

	VE (c)	VE (d)	FUP (c)	FUP (d)	UP (c)	UP (d)	IC (c)	IC (d)	Total (c)
<i>Weka</i>	27	5	64	17	101	4	39	8	231
<i>Knime</i>	30	5	57	26	102	3	42	4	231
<i>Tanagra</i>	25	6	41	40	74	29	8	40	148
<i>RapidMiner</i>	25	7	69	12	106	2	39	7	239

Para o construto FUP, *Weka* apresentou índice de aceitação de aproximadamente 76,2%. A ferramenta *Knime* obteve uma boa aceitação com 67,8% de aprovação. Este programa também obteve o maior percentual registrado para o construto VE o que contribuiu



para aceitação da FUP, haja vista, que a VE influencia a FUP.

Com 48,9% de aceitação, a ferramenta *Tanagra* apresentou dados inconclusivos para o construto FUP, ou seja, não se pode afirmar que a ferramenta é fácil de ser utilizada. Os itens FUP2, FUP3, FUP5 e FUP7 foram rejeitados, ou seja, a ferramenta não é clara, não é agradável e não é intuitiva.

A ferramenta *RapidMiner* apresentou altos índices de aceitação para todos os itens avaliados. Ela registrou um percentual de concordância de aproximadamente 82,1%, que permite concluir que a ferramenta é fácil de usar, agradável e bastante intuitiva.

No que se refere ao construto UP, a ferramenta *Weka* apontou altos índices de aceitação em todos os itens avaliados com cerca de 93,5%. A ferramenta melhora o desempenho na execução de tarefas de mineração e produz os resultados esperados.

A ferramenta *Knime* obteve elevados níveis de aceitação nos itens avaliados com aproximadamente 95,4%. Este resultado atesta que os entrevistados concordam que o software realiza as principais atividades de mineração de dados e disponibiliza visualização de gráficos na saída dos dados.

A ferramenta *Tanagra* obteve 68,5% de aceitação para o construto UP, pode-se observar que essa porcentagem se deu basicamente porque a ferramenta realiza as principais atividades de mineração de dados como: classificação, agrupamento, regressão e associação. Os itens UP3 e UP8 foram rejeitados, ou seja, a aplicação não apresenta visualização de gráficos na saída de dados e não melhora o desempenho dos avaliados.

Com 98,1% de aceitação, *RapidMiner* chegou próximo ao nível máximo de aceitação para o construto UP. Os entrevistados concordam que a ferramenta agrega valor em suas atividades, melhora o desempenho, oferece relatórios de saídas de dados e gera gráficos de saída de dados.

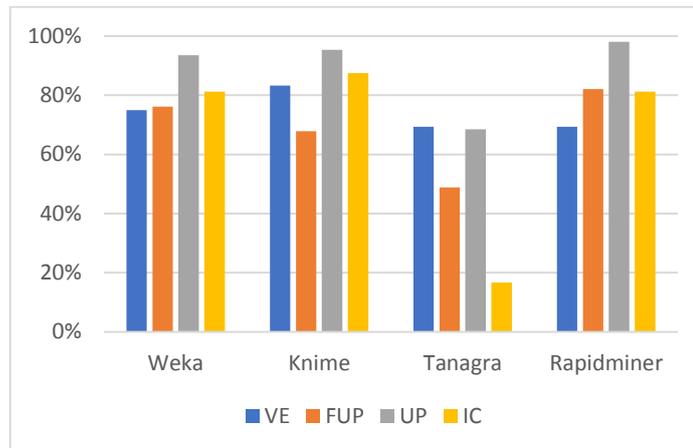
Quanto ao construto IC, a ferramenta *Knime* atendeu a expectativa dos respondentes com 87,5% de concordância alcançando o maior índice de aceitação. Em segundo lugar a ferramenta *Weka* foi considerada apropriada para realizar tarefas de mineração de dados com 81,2% de concordância.

O software *Tanagra* foi rejeitado neste item por cerca de 83,3% dos entrevistados. Conforme os resultados, a ferramenta não atende a nenhum dos itens avaliados, não é uma aplicação apropriada para as tarefas de mineração de dados, não atende às expectativas dos usuários e não é recomendada pelos entrevistados.

Com 81,2% de aceitação, a ferramenta *RapidMiner* mostrou ser uma aplicação apropriada para o estudo da mineração de dados, pois atende as expectativas dos entrevistados.

O gráfico da Figura 3 mostra um comparativo percentual do nível de concordância coletado pelo questionário por ferramenta. Este percentual foi calculado com base no valor atingido por cada ferramenta em comparação ao valor máximo possível por variável.

Figura 3 – Comparativo de concordância por ferramenta de mineração



CONSIDERAÇÕES FINAIS

As ferramentas de mineração de dados vêm ganhando importância tanto no meio acadêmico, como no meio corporativo. O objetivo destas ferramentas é facilitar e auxiliar os usuários nas tarefas de mineração de dados, logo, percebe-se a necessidade de utilizar ferramentas recomendadas que de fato ajudem os usuários em suas atividades. Desta maneira, aqui foi demandado um estudo com o objetivo de avaliar ferramentas comumente utilizadas e entender os motivos que levam a aceitação dessas tecnologias.

Com os resultados obtidos, verificou-se que as ferramentas *RapidMiner*, *Weka* e *Knime* foram aceitas pelos participantes da pesquisas porque se mostraram úteis no processo de descoberta do conhecimento e de fácil utilização nas tarefas realizadas. O *Tanagra* foi a única ferramenta não aceita pelos entrevistados, com a menor intenção de uso.

A aceitação das ferramentas foi determinada pela utilidade percebida, comprovando mais uma vez, o efeito direto da utilidade percebida na intenção de uso. Este construto possui maior relevância na avaliação da aceitação de sistemas, conforme teorizado por Davis (1989).

Por outro lado, também foi confirmada pelos resultados obtidos, que a facilidade de uso possui um efeito direto e positivo na utilidade percebida (DAVIS, 1989). A facilidade de uso é frequentemente citada como fator de satisfação. Importa, pois, conciliar a simplicidade com as possibilidades de uma utilização intuitiva e natural.

Cerca de 4 pontos limitantes permeiam a pesquisa: pequena quantidade de entrevistados, o questionário foi realizado com 12 alunos, esse número é relativamente baixo para nível de confiança, o que limita a pesquisa. Em termos de escopo, este estudo limitou-se a uma instituição de ensino, uma pesquisa futura poderia incluir outras instituições públicas e privadas. Quanto aos critérios baseados no modelo TAM, outras variáveis poderiam ser inseridas no modelo proposto, ou a análise poderia relacionar as variáveis com o perfil do usuário, por exemplo, se possui experiência ou não.

REFERÊNCIAS

BOSCARIOLI, C.; VITERBO, J.; TEIXEIRA, M. F. **Avaliação de Aspectos de Usabilidade em Ferramentas para Mineração de Dados**. Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro. Rio de Janeiro, v. 1, n. 1, p. 107-114, 2014.



BRITO, D. M.; JUNIOR, I. A. A.; QUEIROGA, E. V.; REGO, T. G. **Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina.** In Anais do Simpósio Brasileiro de Informática na Educação, v. 25, n.1, p. 882 – 890, 2014.

CARVALHO, J. Y. A. et al. **Adoção de um Software Público Brasileiro:** Estudo de Caso orientado pelo Modelo de Aceitação de Tecnologia. 2015. Disponível em: <<http://www.ibd.dcc.ufmg.br/colecoes/freerbase/2015/001.pdf>>. Acesso em: 29 maio 2017.

DAVIS, F. D. **Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.** MIS Quarterly, v. 13, n. 3, p. 319–340, 1989.

DAVIS, F. D.; BAGOZZI, R. P.; WARSHAW, P. R. **User acceptance of computer technology: a comparison of two theoretical models.** Management Science, Ann Arbor, v.35, n.8, p.982-1003, 1989.

DIAS, M. C.; ZWICKER, R.; VICENTIN, I. C. **Análise do modelo de aceitação de tecnologia de Davis.** RevistaSpei, Curitiba, v. 4, n. 2, p. 15-23, 2003.

DILLON, A.; MORRIS, M. **User acceptance of new information technology:** theories and models. Annual Review of Information Science and Technology, Medford (NJ), v.31, p.3-32, out/nov, 1996.

DWIVEDI, Y.; WASTELL, D.; LAUMER, S.; HENRIKSEN, H.; MYERS, M.; BUNKER, D.; ELBANNA, A.; RAVISHANKAR, M.; SRIVASTAVA, S. **Research on Information Systems Failures and Successes: Status Update and Future Directions.** Information Systems Frontiers, Springer US, p. 1–15, ISSN 1387-3326, 2014.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From data mining to knowledge discovery in databases.** AI magazine, v. 17, n. 3, p. 37, 1996.

FERNANDES, A.A.; ABREU, V.F. **Implantando a governança de TI:** da estratégia à gestão dos processos e serviços. 4. Ed. Rio de Janeiro: Brasport, 2014.

GOEBEL, M.; GRUENWALD, L. **A survey of data mining and knowledge discovery software tools.** ACM SIGKDD, San Diego, v. 1, n. 1, p. 20-33, set/out, 1999.

HAND, D; MANNILA, H; SMYTH, P. **Principles of Data Mining.** MIT Press, 2001.

IT GOVERNANCE INSTITUTE (ITGI). **Board briefing on IT governance.** 2nded. 2007.

KNIME. **KNIME Open Source Story.** 2017. Disponível em: <https://www.knime.com/knime-open-source-story> >. Acesso em: 13 nov. 2017.



LÖBLER, M. L.; VISENTINI, M. S.; VIEIRA, K. M. **A aceitação do comércio eletrônico explicada pelos modelos TAM e TTF combinados.** Anais do XXX Encontro da ANPAD, Salvador, set, 2006.

MENEZES, J. S.; RAMOS, D. G.; SOARES, M. S. **On Criteria to Choose a Content Management System: A Technology Acceptance Model Approach.** p. 559–564, ISSN 2325-9086, 2016.

PINTO, V. G. **Mineração de Dados Paralela e Distribuída Baseada No Ambiente Weka.** 2010. 58f. Dissertação (Graduação) - Universidade Federal de Santa Maria, Santa Maria, 2010.

RAMAMOCHAN, Y.; VASANTHARAO, K.; CHAKRAVART, C.K; RATNAM, A.S.K. **A study of data mining tools in knowledge discovery process.** Int. J. Soft Comput. Eng., 2, ISSN: 2231-2307, 2012.

RAPIDMINER. **RapidMiner:** About us. 2017. Disponível em: < <https://rapidminer.com/>>. Acesso em: 13 nov. 2017.

REZENDE, S. O. **Mineração de Dados.** Congresso da sociedade Brasileira de Computação, 25. 2005, São Leopoldo, Rs. A Universidade da Computação: Um agente de inovação e Conhecimento. São Leopoldo: Unisinos, p. 397 – 433, 2005.

SASSI, R. **Data Webhouse e Business Intelligence Operacional:** Revisitando a Tecnologia e analisando as tendências do armazém de dados, XXX Encontro Nacional de Engenharia de Produção, São Carlos, Brasil, 2010.

SILVA, A. L. M. R. **A influência do treinamento de usuários na aceitação de sistemas ERP em empresas no Brasil.** 2005. 118f. Dissertação (Mestrado em Administração) – Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

TANAGRA. **Tanagra Project.** 2017. Disponível em: < <https://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html> >. Acesso em: 13 nov. 2017.

WEKA. **Weka 3:** Data mining Software in Java. 2017. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/index.html>>. Acesso em: 13 nov. 2017.

WEST, D.M. **Big data for education: data mining, data analytics, and web dashboards.** 2012. Disponível em: <http://www.oxydiane.net/IMG/pdf/school_evaluation.pdf>. Acesso em: 07 out. 2017.