

**INSTITUTO FEDERAL GOIANO – CAMPUS CERES
BACHARELADO EM SISTEMAS DE INFORMAÇÃO
LARA CAROLINE DAMACENO FARIA**

**ANÁLISE DO ENSINO À DISTÂNCIA DO INSTITUTO FEDERAL GOIANO
UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS**

**CERES – GO
2019**

LARA CAROLINE DAMACENO FARIA

**ANÁLISE DO ENSINO À DISTÂNCIA DO INSTITUTO FEDERAL GOIANO
UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS**

Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Sistemas de Informação do Instituto Federal Goiano – Campus Ceres, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação sob orientação do Professor Lucas José de Faria.

**CERES – GO
2019**



INSTITUTO FEDERAL
Goiano

Repositório Institucional do IF Goiano - RIIF Goiano
Sistema Integrado de Bibliotecas

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☐ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☒ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional | - Tipo: |

Nome Completo do Autor: Lara Caroline Damaceno Faria

Matrícula: 2016103202030273

Título do Trabalho: Análise do Ensino à Distância do Instituto Federal Goiano Utilizando Técnicas de Mineração de Dados

Restrições de Acesso ao Documento

Documento confidencial: ☒ Não ☐ Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: ____/____/____

O documento está sujeito a registro de patente? ☐ Sim ☒ Não
O documento pode vir a ser publicado como livro? ☐ Sim ☒ Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

1. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
2. obteve autorização de quaisquer materiais incluídos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
3. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Local

Data

Lara Caroline Damaceno Faria

Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:

Lucas José de Faria
Assinatura do(a) orientador(a)

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

F224a Faria, Lara Caroline Damaceno
ANÁLISE DO ENSINO À DISTÂNCIA DO INSTITUTO FEDERAL
GOIANO UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS /
Lara Caroline Damaceno Faria;orientador Lucas José
de Faria. -- Ceres, 2019.
79 p.

Monografia (em Bacharelado em Sistemas de
Informação) -- Instituto Federal Goiano, Campus
Ceres, 2019.

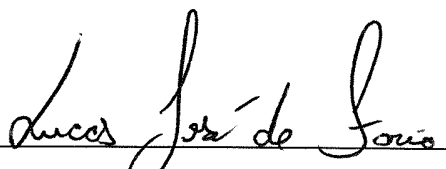
1. Ambiente Virtual de Aprendizagem. 2. Ensino à
Distância. 3. Mineração de Dados na Educação. I. Faria,
Lucas José de, orient. II. Título.

ANEXO IV - ATA DE DEFESA DE TRABALHO DE CURSO

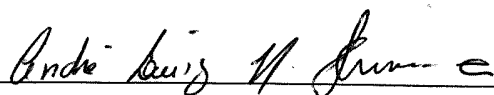
Ao(s) 22 dia(s) do mês de novembro do ano de dois mil e 19, realizou-se a defesa de Trabalho de Curso do(a) acadêmico(a) LARA CAROLINE DAMASCENO FARIAS, do Curso de Sistemas de Informação, matrícula 2016103202030273, cujo título é "Análise do Ensino à Distância do Instituto Federal Goiano Utilizando Técnicas de Mineração de Dados". A defesa iniciou-se às 20 horas e 30 minutos, finalizando-se às 20 horas e 05 minutos. A banca examinadora considerou o trabalho Aprovado com média 8,8 no trabalho escrito, média 9,5 no trabalho oral, apresentando assim média aritmética final 9,1 de pontos, estando o(a) estudante Apto para fins de conclusão do Trabalho de Curso.

Após atender às considerações da banca e respeitando o prazo disposto em calendário acadêmico, o(a) estudante deverá fazer a submissão da versão corrigida em formato digital (.pdf) no Repositório Institucional do IF Goiano – RIIF, acompanhado do Termo Ciência e Autorização Eletrônico (TCAE), devidamente assinado pelo autor e orientador.

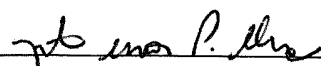
Os integrantes da banca examinadora assinam a presente.



Assinatura Presidente da Banca



Assinatura Membro 1 Banca Examinadora



Assinatura Membro 2 Banca Examinadora

Dedico esta conquista a todos que
me apoiaram e confiaram em mim.

AGRADECIMENTOS

Agradeço aos meus pais, Adriano e Lucilene, que me incentivam e me apoiam desde criança a estudar e buscar sempre o conhecimento. Aos meus familiares, meus tios e avós que me apoiam também desde criança e que acreditam em mim.

Agradeço ao meu orientador, Professor Lucas José de Faria, por todo apoio e principalmente pela confiança neste trabalho. Por apoiar minhas mudanças de tema, por auxiliar nas minhas indecisões e por toda a luz que me deu quando estava perdida.

Agradeço às minhas melhores amigas, Laura e Débora, que desde o ensino médio sempre me disseram “você nasceu pra isso” (mesmo que eu tenha demorado a aceitar isso e tenha tentado fugir de todas as maneiras), com quem pude compartilhar toda a luta, anseios e vitórias durante todo o percurso desse curso. Aos meus amigos de faculdade, em especial Thalia, Jeferson, Ianka, Gustavo e Luciano, com os quais compartilhei o dia a dia e que me suportaram também.

Agradeço ao meu namorado, Edson, pelo incentivo e pela compreensão à minha ausência, principalmente na reta final deste trabalho. Agradeço ao apoio, aos incentivos, à segurança e à confiança que me passou desde que inciei a minha jornada até aqui.

“Dobby agora é um elfo livre!”
Dobby, Harry Potter e a Câmara Secreta

RESUMO

O uso da tecnologia vem crescendo nas últimas décadas em todos os âmbitos. Seja na vida profissional, pessoal ou acadêmica, a utilização de dispositivos eletrônicos se torna cada vez mais necessária no cotidiano das pessoas. O ensino a distância (EaD) é cada vez mais buscado, visto que por meio de tecnologias, proporciona ao aluno o conforto de estudar onde quiser e quando quiser, não estando diretamente apegado a horários como ocorre no ensino totalmente presencial. Como meio de interação entre aluno e professor, o EaD, utiliza ambientes virtuais de aprendizagem (AVA). Nestes AVAs são produzidos diversos tipos de dados em grandes volumes que ao serem analisados podem retornar informações valiosas. Por se tratar de um grande volume de dados, é necessário a utilização de ferramentas que possibilitem esta análise, neste caso a mineração de dados. Desta maneira, o presente trabalho teve como objetivo realizar a análise dos dados da tabela de registros das interações de usuário, *mdl_logstore_standard_log*, do AVA *Moodle*, utilizado no curso Técnico em Administração do Campus Iporá do Instituto Federal Goiano. Foi utilizado o processo de descoberta de conhecimento em banco de dados e na técnica de mineração de dados, deste processo, foram aplicados os algoritmos de associação, agrupamento e classificação. Buscou-se compreender se o desempenho dos discentes tem relação com as variáveis de tempo gasto no questionário e/ou com a média da nota obtida em cada disciplina. Por fim os algoritmos apresentaram um desempenho fraco de forma que não demonstraram também se as variáveis implicam no desempenho do discente.

Palavras-chave: Ambiente Virtual de Aprendizagem. Ensino à Distância. Mineração de Dados na Educação.

ABSTRACT

The use of technology has been growing in the last decades in all scopes. Whether in professional, personal or academic life, the use of electronic devices becomes increasingly necessary in people's daily lives. Distance learning (DE) is increasingly sought, as through technologies, provides students with the comfort of studying where they want and when they want, not being directly attached to schedules as occurs in fully classroom teaching. As a means student-teacher interaction, e-learning uses virtual learning environments (VLE). In these VLEs, many types of data are produced in large volumes that, when analyzed, can return valuable information. Because it is a large volume of data, it is necessary to use tools that enable this analysis, in this case data mining. Thus, the present work aimed to perform the data analysis of the user interaction record table, *mdl_logstore_standard_log*, of the VLE Moodle, used in the Technical Course in Administration of Campus Iporá of the Goiano Federal Institute. The process of discovery of knowledge in database was used and in the data mining technique, from this process were applied the algorithms of association, grouping and classification. It sought to understand if the students' performance is related to the variables of time spent in the questionnaire and/or the average grade obtained in each subject. Finally, the algorithms presented poor performance so that they didn't demonstrate whether the variables imply student performance.

Keywords: Distance Learning. Educational Data Mining. Virtual Learning Environment.

LISTA DE ILUSTRAÇÕES

Figura 1 - Processo da descoberta de conhecimento em banco de dados.....	3
Figura 2 - Classificação do desempenho do aluno de acordo com a nota.....	10
Figura 3 - <i>Datasets</i> e colunas geradas para aplicação de algoritmos.....	10
Figura 4 - Total de Interações por Módulo.....	14
Figura 5 - 10 melhores regras geradas pelo algoritmo <i>Predictive Apriori</i> no <i>dataset</i> “A”	15
Figura 6 - 10 melhores regras geradas pelo algoritmo <i>Predictive Apriori</i> no <i>dataset</i> “B”	16
Figura 7 - Centroides gerados pelo algoritmo <i>K-Means</i> para o <i>dataset</i> “B”	18
Figura 8 - 10 melhores regras geradas pelo algoritmo <i>Predictive Apriori</i> aplicado no <i>dataset</i> “B” sem a coluna do usuário.....	21
Figura 9 - Centroides gerados pelo algoritmo <i>K-Means</i> para o <i>dataset</i> “B” sem a coluna de usuário.....	23

LISTA DE TABELAS

Tabela 1 - MCC e desempenho dos algoritmos de classificação aplicados no <i>dataset</i> “B”	19
Tabela 2 - MCC e desempenho dos algoritmos de classificação aplicados no <i>dataset</i> “B” sem a coluna de usuário.....	24

SUMÁRIO

1.INTRODUÇÃO.....	1
2.REVISÃO DE LITERATURA.....	3
3.METODOLOGIA.....	7
4.RESULTADOS E DISCUSSÃO.....	13
4.1.Verificação Quantitativa da Base de Dados.....	13
4.2.Resultados da Aplicação de Algoritmos de Associação no <i>Dataset “A”</i>	14
4.3.Algoritmos de Associação, <i>Clusterização</i> e Classificação no <i>Dataset “B”</i> . 16	
4.4.Algoritmos de Associação, <i>Clusterização</i> e Classificação no <i>Dataset “B”</i> sem a Coluna do Usuário.....	21
5.CONCLUSÃO.....	25
REFERÊNCIAS BIBLIOGRÁFICAS.....	27
APÊNDICES.....	30

1. INTRODUÇÃO

Com o avanço da tecnologia e a crescente utilização de sistemas computacionais nas mais diversas áreas, como nas interações sociais empregando os meios de comunicação mais dinâmicos e atuais, vêm-se produzindo um grande volume de dados. Estes dados podem apresentar informações valiosas quando analisados, porém a análise de grandes quantidades de dados demanda o uso de ferramentas que possam agilizar o processo. O processo de descoberta de conhecimento em banco de dados (do inglês, *Knowledge Discovery in Databases*) tem como técnica a mineração de dados, que é responsável por evidenciar conhecimentos a partir de um grande volume de dados, assim podendo revelar padrões que permitem realizar a previsão de comportamentos futuros (CARDOSO; MACHADO, 2008).

Além de várias formas de comunicação, a comunidade acadêmica conta ainda com a possibilidade de utilizar plataformas que possibilitam a capacitação mesmo fora do ambiente físico da instituição de ensino. Esses Ambientes Virtuais de Aprendizado (AVA) são, como o próprio nome diz, plataformas de aprendizado virtual, que para Santos (2002) consiste em um ambiente virtual que permite a interação de seres humanos e objetos técnicos que possibilitam a construção da aprendizagem. Atualmente a plataforma de aprendizado *Moodle* é utilizada por mais de 90 milhões de usuários em todo o mundo (MOODLE, 2018). O *Moodle* possui código aberto (do inglês, *Open Source*), permitindo que o usuário faça as alterações que desejar, possibilitando a adequação ao curso que será ministrado a partir da plataforma.

A plataforma *Moodle* gera inúmeros dados de registros de acesso, interações do usuário com sistema, apresenta respostas de teste e até mesmo o tempo gasto para respondê-los. Registram também as participações em *chats* e outros incontáveis parâmetros que ao serem estudados e analisados, podem gerar informações relevantes e valiosas para as instituições.

Se tratando de dados educacionais, existe uma área específica da ciência de dados conhecida como Mineração de Dados Educacionais (do inglês, *Educational Data Mining*) que Baker, Isotani e Carvalho (2011, p. 04) definem como “área de

pesquisa que tem como principal foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais”. A partir da mineração de dados educacionais no *Moodle* é possível “prever” evasões e identificar o perfil de cada aluno. A análise destes dados pode auxiliar também a tomada de decisões das instituições visando a melhora da qualidade do serviço prestado, assim como o aumento do índice de sucesso dos alunos desta modalidade de ensino.

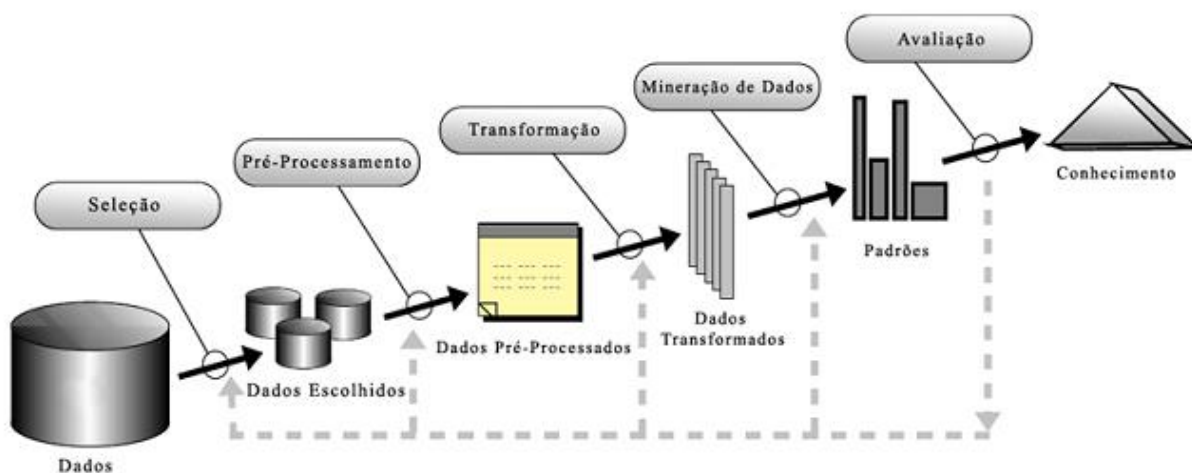
O ensino a distância (mais conhecido como EaD), permite que estudantes tenham horários e turnos flexíveis, assim se tornando uma nova alternativa de capacitação. A interação entre aluno e professor ocorre através de ferramentas online, utilizadas pela instituição que promove os cursos nesta modalidade (EAD, 2018).

O Instituto Federal de Ciência e Tecnologia, fornece além de cursos presenciais, a modalidade a distância por meio da plataforma *Moodle* como ambiente virtual de aprendizagem. Desta forma o presente trabalho tem como objetivo analisar o estudante de Ensino a Distância do Instituto Federal Goiano, por meio de métodos da mineração de dados a partir de *logs* de registro da plataforma *Moodle*. Especificamente, pretende-se verificar se o desempenho dos discentes tem relação com a forma que a plataforma é utilizada pelos mesmos e quais algoritmos têm a capacidade de realizar esta análise.

2. REVISÃO DE LITERATURA

O processo de descoberta de conhecimento em banco de dados (do inglês, *Knowledge Discovery in Databases* ou KDD) é definido por Fayyad (1996, p. 40) como “um o processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e, em última análise, compreensíveis”. Uma das etapas de KDD é a mineração de dados que é responsável pela análise de dados e aplicação de algoritmos para produzir os padrões dos dados analisados (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996). Outras etapas realizadas dentro da KDD (Figura 1) são a seleção dos dados, o pré-processamento dos dados selecionados, a transformação dos dados pré-processados para que finalmente seja feita a mineração de dados que gera os padrões que produzirão o conhecimento desejado.

Figura 1 - Processo da descoberta de conhecimento em banco de dados.



Fonte: <http://bit.ly/imagemkdd>. (2019)

Desde a década de 1990, a *Internet* vem se popularizando dia após dia, possibilitando a evolução da educação a distância que agora tem a alternativa de utilizar ambientes virtuais de aprendizagem (AVA) com o intuito de otimizar o processo educacional (LEONARDO; BHERING; GOMES, 2014). Franco, Cordeiro e Castillo (2003) apresentam quatro estratégias em que foram baseadas as primeiras versões de ambientes virtuais de aprendizagem sendo elas a incorporação de

correios eletrônicos e grupos de discussão, a possibilidade de gerenciamento de arquivos e cópias de segurança, criação de módulos para conteúdo e a avaliação para as atividades educacionais e também elementos para facilitar a administração acadêmica.

Desta forma, ambientes virtuais de aprendizagem tem a capacidade de gerar um grande volume de dados que podem apresentar informações valiosas, possibilitando tomadas de decisões tanto pedagógicas quanto administrativas. Por se tratar de um grande volume de dados, esta análise demanda um processo computacional, como a mineração de dados educacionais.

A mineração de dados educacionais consiste na adaptação de métodos da mineração de dados para analisar dados de contextos educacionais, gerados por ambientes virtuais de aprendizagem, como interação de professores e alunos com estes sistemas, por meio de métodos como predição, agrupamento e mineração de relações sendo estes os mais demandados (COSTA et al., 2012). Para predição, são desenvolvidos modelos que deduzem características específicas dos dados; no agrupamento, verifica-se os dados que se agruparão naturalmente assim podendo alocá-los em grupos; e na mineração de relações, são buscadas as relações entre as variáveis presentes no banco de dados (BAKER; ISOTANI; CARVALHO, 2011).

Dentre as técnicas utilizadas na mineração de dados educacionais está a clusterização, que consiste em apresentar dados que se agrupam de forma natural e posteriormente permitindo a classificação dos mesmos em grupos ou categorias. Esta técnica é utilizada por Pinheiro et al. (2014) com a finalidade de encontrar agrupamentos de alunos que têm o comportamento análogo por meio da análise de *logs*, do ambiente virtual de aprendizagem *Moodle*, utilizado pelos alunos.

Outra técnica utilizada é a de associação que possibilita a verificação de causa e efeito de determinada situação. Silva, Morino e Massahiro (2014) utilizaram a técnica de associação para apresentar a relação entre o desempenho no Exame Nacional do Ensino Médio (ENEM) e os fatores socioeconômicos de cada candidato. A partir dos dados socioeconômicos e de desempenho dos candidatos foi possível verificar que o desempenho do aluno pode ser influenciado pela baixa renda da família, por pais com nível baixo de escolaridade e pela quantidade elevada de pessoas que residem junto ao candidato.

Outro trabalho relevante utilizando as regras de associação é o de Santos et al. (2015) onde analisaram a relação entre padrões de comportamento e o desânimo do aluno, por meio de Regras de Associação da mineração de dados educacionais. Nesta análise, utilizou-se os dados de questionários proposto pelos autores aos alunos e os *logs* da plataforma *Moodle*, assim como em Pinheiro et al. (2014) mencionado anteriormente, para realizar a associação e verificar padrões de comportamento de um aluno em estado de desânimo. Oito das dez regras apresentaram o desânimo e os alunos que enviaram menos da metade das atividades, solicitadas pelo professor, se mostraram desanimados.

Além de *logs* de acesso, que são muito utilizados, outra variável que pode ser usada é a resposta em fóruns de discussões. Silva et al. (2015), em seu trabalho, mostraram que para auxiliar na tomada de decisões, é feita uma análise dos principais assuntos discutidos dentro do fórum de uma atividade utilizando algoritmos de agrupamento de dados e a partir do volume de postagens destes grupos é possível traçar ações estratégicas para melhorar a qualidade da gestão.

Assim como há a mineração de dados educacionais para tratar de conhecimentos educacionais, existe também a mineração de textos quando há a demanda da utilização de uma base de textos. Silva, Lopes e Filho (2014), propõem uma forma de melhorar o aproveitamento das atividades complementares dos discentes, por meio da construção de um armazém de dados. Este armazém de dados permite ao coordenador conhecer as áreas mais buscadas para a complementação do curso e ao aluno permite saber quantas horas foram cumpridas e em quais categorias de atividades. Mesmo que este trabalho tenha características voltadas a utilização da mineração de dados educacionais, os autores optaram pela utilização de conceitos de mineração de dados e mineração de textos.

A mineração de textos é definida por Webber, Cristofoli e Lima (2013) como “uma subárea que procura descobrir, de forma automática, informações (padrões e anomalias) em dados não estruturados, sendo a maioria textual”. Ainda em Webber, Cristofoli e Lima (2013) é apresentada uma ferramenta de mineração de textos intitulada *UnderMine Text Miner* que analisa artigos e faz a separação dos termos de forma em que possa auxiliar a área de pesquisa, facilitando para que o pesquisador consiga com facilidade e agilidade compreender os assuntos tratados em

determinado artigo. Desta forma, podemos concluir que a mineração de textos também pode ser utilizada em ambiente educacionais.

3. METODOLOGIA

Diante dos objetivos traçados, o presente trabalho consiste em uma pesquisa explicativa, pois pretende-se verificar se o modo como os discentes, do ensino a distância (EaD) do Instituto Federal Goiano, utilizam a plataforma influencia no desempenho dos mesmos. Também pretende-se identificar quais algoritmos podem realizar esta análise da melhor forma. É um trabalho com abordagem quantitativa pois visa a quantificação dos dados para se obter informações significativas.

Inicialmente, realizou-se a busca de artigos acadêmicos com o intuito de se obter um referencial teórico e o embasamento de quais técnicas e algoritmos poderiam ser utilizados. Por meio da palavra-chave “mineração de dados”, buscada na base de dados Google Acadêmico, foi encontrado um artigo que trata do levantamento dos principais trabalhos na área de mineração de dados educacionais no ano de 2014 (RODRIGUES et al., 2014). Ainda em Rodrigues et al. (2014) foi possível verificar eventos e revistas importantes que possuem publicações na área de mineração de dados educacionais, sendo selecionados, como base de busca para obtenção de artigos de referência tanto teórica quanto prática, o Simpósio Brasileiro de Informática na Educação (SBIE), a Revista de Novas Tecnologias na Educação e a Informática na Educação: Teoria & Prática.

Como meio de interação entre aluno e professor, os cursos EaD do Instituto Federal Goiano utilizam a plataforma *Moodle*. Sendo assim, os dados foram extraídos da base deste software. O curso escolhido para a análise foi o Técnico em Administração do Campus Iporá. A escolha deste curso se deu com base na pesquisa realizada na plataforma Nilo Peçanha, onde foi escolhido o curso com o mínimo de discentes com pendências, assim podendo concluir que este já havia sido finalizado (PNP, 2018). Por meio da página inicial do *site Moodle*, do próprio IF Goiano, foi possível verificar que o curso teve início no segundo semestre de 2017 e fim no segundo semestre de 2018 (IFGOIANO, 2019).

O órgão responsável pelo gerenciamento do *Moodle* é a reitoria do IF Goiano. Sendo assim, foi solicitado ao próprio órgão, os dados do *Moodle* referente ao curso Técnico em Administração do Campus Iporá. Foi solicitado somente uma tabela do banco de dados intitulada *mdl_logstore_standard_log* que possui colunas com dados significativos e que também não constituem dados considerados sensíveis de

acordo com o Art.5º da Lei Geral de Proteção de Dados Nº13.709 de 14 de Agosto de 2018, que os classifica como:

Dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural. (BRASIL, 2018)

Após a coleta dos dados, foi realizada a segunda fase de KDD que consiste no pré-processamento dos dados. Diferente da mineração de dados tradicional, a mineração de dados educacionais normalmente não precisa realizar a maioria dos processos de pré-processamento, visto que alguns AVAs tendem a armazenar tipos de dados para fins de análise, se tornando uma tarefa mais simples (GARCIA et al., 2007). Sendo assim o pré-processamento consistiu em transformação dos valores numéricos em dados categóricos (discretização de dados), criação e seleção de novos atributos de acordo com os existentes para a criação de um subconjunto, nivelção da informação a ser minerada a partir da criação de tabelas de sumarização, e por fim a transformação dos dados para adequação do formato requerido pelo algoritmo, sendo todas estas tarefas propostas também por Garcia et al. (2007).

No pré-processamento, foram discretizadas algumas colunas que não apresentavam relevância para este trabalho, sendo selecionadas somente as colunas *eventname*, *component*, *action*, *target*, *userid*, *courseid*, *timecreate* e *other*. A coluna *component* apresenta o módulo do sistema que foi acessado, como módulos de questionários, fóruns, pastas, entre outros. Na coluna *action* é possível verificar a ação realizada naquele módulo como visualização, criação, atualização, envios, etc. A coluna *target* apresenta o que foi afetado dentro daquele módulo como criação de *posts*, discussões iniciadas, páginas editadas, etc. Já a coluna *eventname* armazena a junção destas três colunas anteriores criando um tipo de endereço. Nas colunas *courseid* e *userid* é possível identificar, respectivamente, o identificador numérico individual do curso e do usuário. A coluna *timecreated* apresenta o tempo em formato *timestamp*. E por fim a coluna *other* apresenta alguns dados, de forma serializada pela linguagem de programação *Hypertext Preprocessor*

(PHP), que permitem verificar o nome dos cursos, notas de alunos e alguns outros identificadores numéricos individuais.

Todos os passos do pré-processamento, da transformação dos valores numéricos em categóricos à preparação para o formato do algoritmo, foram feitos utilizando a linguagem de consulta estruturada, mais conhecida como linguagem SQL (*Structured Query Language*), por meio do sistema de gerenciamento de banco de dados (SGBD) *MySQL*, com apoio da ferramenta *MySQL Workbench*.

O formato *timestamp* da coluna *timecreated*, não possibilita a visualização das datas de forma legível, sendo assim, foi necessário o tratamento deste campo no SGBD com a função “*from_unixtime()*”, para que o campo fosse convertido em data e hora. Para identificar o nome de cada disciplina, foi realizada a seleção de cada identificador e da coluna *other* onde foi possível verificar a criação daquela disciplina no módulo de aprendizagem. Como a coluna *other* se apresenta de forma muito “poluída”, para extrair somente o nome do curso, foram exportados os dados, a limpeza da coluna foi efetuada no software *Libre Office Calc* e após a limpeza os dados do curso foram inseridos em uma nova tabela no SGBD.

Para se obter as médias de tempo da realização do questionário, foram selecionadas colunas que correspondessem ao módulo questionário onde o mesmo tivesse sido iniciado e finalizado, de forma que o identificador destes fossem iguais. Após esta seleção foram criadas duas novas tabelas, uma com os tempos iniciais e outra com os tempos finais. Para que não fossem analisados dados de questionários que somente iniciaram e não terminaram, foram selecionados somente a diferença de tempo se houvesse o identificador do mesmo em ambas as tabelas, de início e fim. Após a obtenção do tempo gasto por aluno e por disciplina, foi calculada a média de tempo gasta por cada aluno em cada disciplina na realização do questionário.

Para a obtenção de notas, foram selecionados os dados em que a coluna *action* possuía o dado “*graded*”, correspondente a nota, a identificação do usuário, a disciplina e a nota do aluno apresentada na coluna *other*. O mesmo método de limpeza da coluna *other*, realizado para obter os nomes das disciplinas, foi utilizado também para a obtenção dos dados das notas, porém com a diferença que já na

seleção foi possível realizar uma pré limpeza com a função “*SUBSTRING()*” da linguagem *SQL*, em que é possível obter-se somente parte da coluna.

Após toda a limpeza, por meio de um comando *SQL*, foram selecionados os usuários, a média de notas, a classificação do desempenho de acordo com a nota, o nome da disciplina, a média de tempo gasto no questionário por disciplina e uma classificação nominal de faixa de tempo. O desempenho foi mensurado de acordo com as médias das notas obtidas nesta tabela, sendo classificado em quatro tipos como apresenta a figura 2. A faixa de tempo do questionário foi classificada inicialmente de 15 em 15 minutos dentro do período de 1 hora, após o período de 1 hora foi classificada de 3 em 3 horas até as 12 horas e por fim os tempos acima de 12 horas foram classificados na mesma faixa denominada acima de 12 horas.

Figura 2 - Classificação do desempenho do aluno de acordo com a nota.

Nota	Classificação do Desempenho
Menor que 2,5	Insuficiente
Entre 2,5 e 5,0	Fraco
Entre 5,0 e 7,5	Bom
Entre 7,5 e 10	Alto

Fonte: Própria (2019).

Foram construídos dois *datasets* (do inglês, conjunto de dados) para testes. O primeiro *dataset* foi criado com o identificador do usuário, data e a coluna de evento, com o intuito de identificar relação entre usuários, datas e eventos. O segundo *dataset* foi criado utilizando as colunas de identificador do usuário, média da nota, desempenho, disciplina, tempo do questionário e faixa de tempo do questionário, sendo as colunas desempenho e faixa de tempo do questionário, geradas a partir da média da nota e da média de tempo do questionário. A figura 3 apresenta os *datasets*, denominados em “A” e “B” e suas respectivas colunas:

Figura 3 - Datasets e colunas geradas para aplicação de algoritmos.

<i>Datasets</i>	Colunas
A	Identificador do Usuário
	Data de Acesso
	Evento Realizado
B	Identificador do Usuário
	Média da Nota
	Desempenho
	Disciplina
	Tempo do Questionário
	Faixa de Tempo do Questionário

Fonte: Própria (2019).

Após o pré-processamento e a transformação dos dados, foi realizada a aplicação dos algoritmos de mineração de dados. Para esta etapa foi utilizado o *software Weka*, na versão 3.8.3. O *software Weka* possibilita a aplicação de algoritmos de aprendizagem em um conjunto de dados, de forma simples, e contém os principais métodos da mineração de dados como: regressão, classificação, agrupamento, regras de associação e seleção de atributos; e possui também ferramentas de transformação de conjuntos de dados (FRANK et al., 2016).

Foram aplicados algoritmos de associação nos *datasets* “A” e “B”. No *dataset* “B” também foram aplicados algoritmos de agrupamento, conhecidos como “*clusters*”, e algoritmos de classificação. Para a utilização dos dados neste software foi necessária a conversão dos dados do *MySQL* para o formato *csv* (do inglês, *Comma Separated Values*). No *software Weka*, ao realizar a importação de cada *dataset*, foi necessário utilizar um filtro do próprio *software* para a conversão da variável do identificador do usuário de numérico para nominal, pois este identificador é utilizado para “nomear” cada usuário.

Para associação foram aplicados os algoritmos *Predictive Apriori* e *Apriori*. A diferença entre estes dois algoritmos é que no *Apriori* são utilizados parâmetros de suporte e confiança para mensurar a qualidade de cada regra de associação, enquanto no *Predictive Apriori*, estes mesmos dois parâmetros, são combinados e denominados como *predictive accuracy*, sendo este o parâmetro utilizado para gerar a regra de associação (AHER; LOBO, 2012). O *Apriori* gera regras muito específicas

e, estas regras, também podem ter um alto suporte e baixa confiança, enquanto o *Predictive Apriori* a confiança depende do suporte na tentativa de melhorar a precisão (PATIL; JOSHI; TOSHNIWAL, 2011).

Na classificação foram selecionados os algoritmos *Bayes Net*, *Naive Bayes* fundamentados no teorema de Bayes; *Decision Table* que utiliza tabelas de decisão; e por fim *J48* e *Random Forest* que utilizam árvores de decisão, sendo estes algoritmos propostos na análise do trabalho de Queiroga, Cechinel e Araujo (2015). Após a aplicação dos algoritmos realizou-se uma análise comparativa entre eles a partir dos índices de melhor classificação e melhor desempenho. Também foi realizada uma análise individual, a partir dos índices de Estatística Kappa, verdadeiro positivo e precisão, que indicam respectivamente a proporção de classificação correta de acordo com a classe verdadeira, a proporção de classificação de cada interação em cada classe e a proporção das interações que realmente pertencem a cada classe.

Na *clusterização* foi utilizado o algoritmo *K-means*, que tem como objetivo organizar objetos similares em grupo. O algoritmo localiza o centroide e atribui cada interação ao centroide mais próximo, sendo necessário atribuir um valor para “k”, sendo este o número de *clusters* a serem obtidos, e o ponto inicial de cada centroide (APON et al., 2006). Neste caso, o valor de “K” foi definido como 4, pois objetiva-se que ao fim possa se obter o agrupamento dos discentes por desempenho, e o ponto inicial é definido pelo próprio algoritmo do *software Weka* de forma aleatória.

Os testes foram realizados em uma única máquina, sendo esta, um *Notebook* da marca Dell, composta por um processador Intel Core i5-4200U, frequência de 1.60GHz, com 6GB de memória RAM e um SSD de 480GB para armazenamento. Como sistema operacional utiliza o *Windows 10 Pro* 2018 de 64 bits.

A seguir, realizou-se a avaliação dos resultados dos algoritmos e verificou se estes resultados são significativos e principalmente se condizem com a realidade. Espera-se que os resultados permitam compreender se a forma como os discentes utilizam a plataforma pode indicar o desempenho dos mesmos e quais algoritmos têm o melhor desempenho nesta análise.

4. RESULTADOS E DISCUSSÃO

Este capítulo é dividido em 4 sessões. Na primeira, são abordados os resultados da verificação quantitativa prévia da base de dados. A segunda sessão apresenta os resultados da aplicação dos algoritmos de associação no *dataset* “A”. Na terceira, os resultados da aplicação dos algoritmos de associação, *clusterização* e classificação no *dataset* “B”. Por fim, na quarta sessão, os resultados da aplicação dos mesmos algoritmos da terceira sessão também no *dataset* “B”, porém sem a coluna de identificação do usuário.

4.1. Verificação Quantitativa da Base de Dados

Utilizando o Sistema de Gerenciamento de Banco de Dados (SGBD) *MySQL* com a linguagem *SQL*, foi possível obter informações significativas sobre o curso. A tabela possui 113617 linhas e 20 colunas, em que cada linha representa uma interação do usuário com o sistema e as colunas apresentam informações em relação à interação, como o tipo de evento ocorrido, o componente acessado do sistema, ação realizada, a disciplina em que a interação foi realizada, o usuário que realizou, a data e hora, incluindo minutos e segundos.

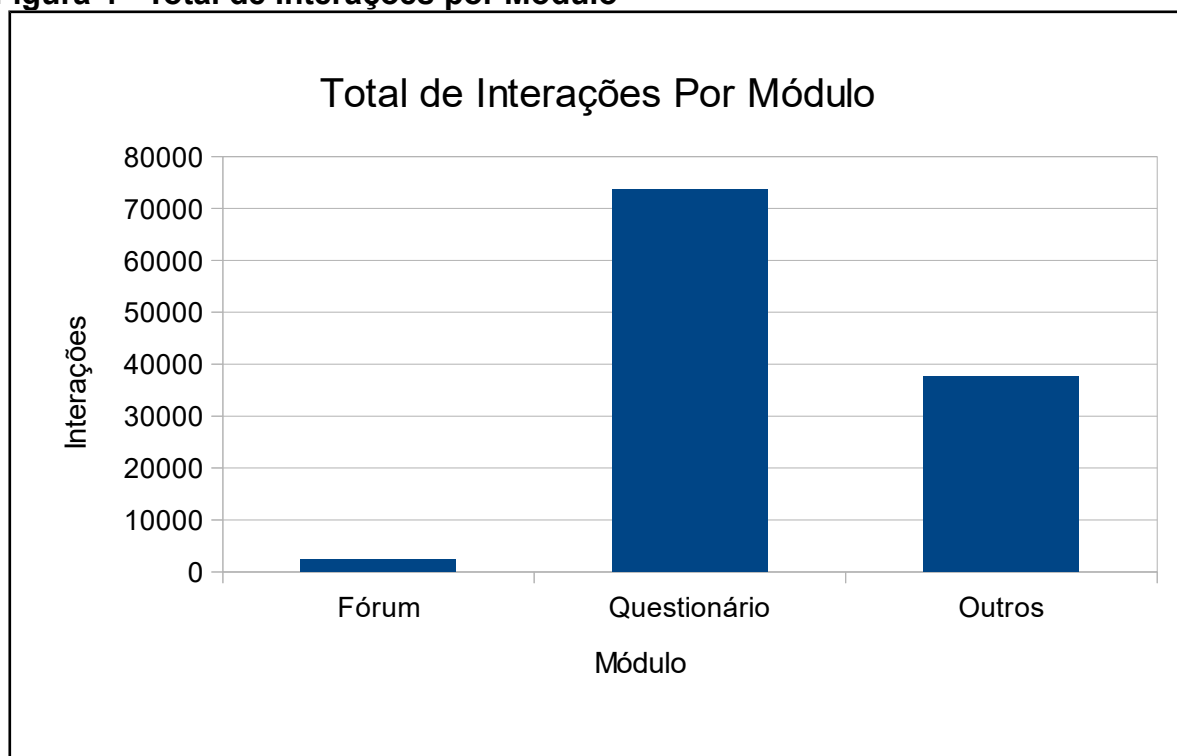
Com alguns comandos básicos de seleção no banco de dados foi possível verificar quantidade de usuários, disciplinas do curso e eventos mais realizados. Foram contabilizados um total de 68 usuários neste banco de dados, dentre eles administradores, professores e alunos, e 21 disciplinas neste curso. A coluna *eventname* desta tabela apresenta o evento realizado, sendo a “visualização de questionário” o evento que mais ocorreu.

O curso teve a duração entre Setembro de 2017 e Dezembro de 2018, tendo algumas interações também no primeiro semestre de 2019. Houveram 396 acessos em dias diferentes, com uma média de 286 interações por dia e o total máximo de interações em um dia ocorreu no dia 10 de novembro de 2017 com 4981 interações.

Nos trabalhos de Guércio et al. (2014) e Mendes e Silva (2016) foram utilizados dados em relação as interações com Fórum para analisar e prever o desempenho dos alunos. Na base de dados deste trabalho foram identificadas

poucas interações com Fóruns, um total de 2393 interações, das quais, 306 interações foram *posts* publicados no módulo. A maior parte delas correspondem ao módulo de questionário, totalizando 73636 interações, o que equivale a 64,81% das interações totais como apresenta o gráfico da figura 4.

Figura 4 - Total de Interações por Módulo



Fonte: Própria (2019)

4.2. Resultados da Aplicação de Algoritmos de Associação no Dataset “A”

No *dataset* “A”, com a data em formato “dia/mês/ano” foram aplicados somente algoritmos de associação, buscando encontrar relações entre os dados deste conjunto. O algoritmo *Apriori* não retornou nenhuma regra, enquanto o algoritmo *Predictive Apriori* retornou 100 regras com acurácias entre 99,49% e 97,58%. Todas as regras fizeram associação a algum usuário específico. A figura 5 apresenta as 10 primeiras regras geradas e o relatório completo desta análise conta no apêndice A:

Figura 5 - 10 melhores regras geradas pelo algoritmo Predictive Apriori no dataset “A”.

Melhores Regras
1. eventname=\core\event\grade_deleted data=2017-11-28 250 ==> userid=712 250 acc:(0.99495)
2. eventname=\mod_quiz\event\edit_page_viewed data=2018-08-13 126 ==> userid=753 126 acc:(0.99472)
3. eventname=\core\event\user_graded data=2018-06-21 107 ==> userid=753 107 acc:(0.9946)
4. eventname=\core\event\user_profile_viewed data=2017-11-08 97 ==> userid=655 97 acc:(0.99451)
5. eventname=\report_outline\event\report_viewed data=2017-11-08 90 ==> userid=655 90 acc:(0.99442)
6. eventname=\core\event\grade_deleted data=2018-02-01 90 ==> userid=712 90 acc:(0.99442)
7. eventname=\gradereport_grader\event\grade_report_viewed data=2017-11-26 162 ==> userid=755 161 acc:(0.99438)
8. eventname=\core\event\user_graded userid=768 86 ==> data=2018-03-19 86 acc:(0.99437)
9. eventname=\mod_quiz\event\edit_page_viewed data=2018-08-08 83 ==> userid=753 83 acc:(0.99432)
10. eventname=\core\event\user_graded data=2018-04-24 83 ==> userid=753 83 acc:(0.99432)

Fonte: Própria (2019)

Nas sete primeiras regras e nas duas últimas, o evento e a data foram associados diretamente ao usuário, enquanto somente na oitava regra o evento e usuário foram associados a data. Neste *dataset* existe uma quantidade muito variada de datas o que impossibilita que o algoritmo faça associações consideradas “relevantes”, mesmo que as associações realizadas tenham uma acurácia forte.

Um segundo teste foi realizado com o *dataset* “A”, tendo as datas em formato “mês/ano”, assim diminuindo a quantidade e a variedade de datas. Neste caso também, as associações foram geradas de maneira muito específica para os usuários, e também não possibilitou a visualização de informações relevantes, mesmo que o retorno das 100 regras do algoritmo tivesse a acurácia entre 99,44% e 86,40% (Apêndice B).

4.3. Algoritmos de Associação, *Clusterização* e Classificação no Dataset “B”

No *dataset* “B” foram realizados dois testes, no primeiro foram retiradas as colunas, média das notas e média de tempo do questionário, visto que estas já estão representadas nas colunas de desempenho e faixa de tempo do questionário. No segundo, além das colunas retiradas no primeiro, foi retirada também a coluna de identificação do usuário.

No algoritmo de associação *Apriori*, foi definido uma confiança de 0.7 equivalente a 70%, sendo os outros atributos automáticos do *Weka*. Houve o retorno de uma única regra em que verifica-se que o desempenho do aluno é ruim quando a faixa de tempo gasto no questionário é menor que 15 minutos, tendo 70% de confiança e *Lift* de 1.11. O *Lift* indica a frequência do desempenho ser ruim quando o tempo do questionário é menor que 15 minutos, ou seja, a frequência do evento B quando o evento A ocorre. No apêndice C é apresentado o relatório completo desta análise.

O algoritmo *Predictive Apriori* retornou 100 regras, com acurácias entre 99,37% e 72,21%. Das 100 regras, 64 associam o usuário específico a outras variáveis e 36 associam a disciplina também a outras variáveis. Na figura 6 são apresentadas as 10 primeiras regras e no Apêndice D é apresentado o relatório completo desta análise.

Figura 6 - 10 melhores regras geradas pelo algoritmo *Predictive Apriori* no dataset “B”.

Melhores Regras
1. userid=701 20 ==> faixa_tempo_quiz=Menos de 15 minutos 20 acc:(0.99373)
2. userid=669 desempenho=fraco 10 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.98825)
3. userid=695 desempenho=fraco 8 ==> faixa_tempo_quiz=Menos de 15 minutos 8 acc:(0.98441)
4. userid=699 desempenho=fraco 8 ==> faixa_tempo_quiz=Menos de 15 minutos 8

acc:(0.98441)
5. userid=670 desempenho=fraco 7 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.98138)
6. userid=687 desempenho=fraco 7 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.98138)
7. disciplina=Administração Financeira faixa_tempo_quiz=Menos de 15 minutos 25 ==> desempenho=fraco 24 acc:(0.97994)
8. userid=990 6 ==> faixa_tempo_quiz=Menos de 15 minutos 6 acc:(0.97707)
9. userid=986 desempenho=bom 6 ==> faixa_tempo_quiz=Menos de 15 minutos 6 acc:(0.97707)
10. disciplina=Ética faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==> desempenho=bom 6 acc:(0.97707)

Fonte: Própria (2019)

A primeira regra associou um usuário específico à faixa de tempo gasta no questionário “menor que 15 minutos”, tendo a acurácia de 97,70%. Já na sétima regra verifica-se que a disciplina de Administração Financeira e o tempo “menor que 15 minutos” estão associados ao desempenho fraco com acurácia de 97,99%. A nona regra associa diretamente um usuário específico ao tempo gasto no questionário “menor que 15 minutos”, tendo 97,70% de acurácia. Na décima regra verifica-se que a disciplina de Ética e a faixa de tempo do questionário “entre 15 e 30 minutos” estão associadas ao desempenho bom, também com 97,70% de acurácia.

Assim como alguns usuários são relacionados ao desempenho fraco quando utilizam o questionário por menos de 15 minutos, outros usuários apresentam também desempenho insuficiente quando utilizam o questionário menos que 15 minutos e bom desempenho quando utilizam entre 3 e 6 horas. No entanto, alguns usuários possuem desempenho alto associado ao tempo do questionário “menor que 15 minutos”.

Algumas disciplinas também são associadas ao desempenho fraco quando o tempo gasto é menor que 15 minutos ou entre 15 e 30 minutos como Administração Financeira e Empreendedorismo, sendo esta com desempenho insuficiente também associado. Enquanto Metodologia em EaD apresenta bom desempenho associado ao tempo gasto do questionário “entre 45 minutos e 1 hora” e o desempenho bom relacionado também ao tempo “entre 1 e 3 horas”.

É possível verificar dentre as regras fortes associações. Tanto o identificador de um usuário específico quanto a disciplina, quando associadas ao tempo gasto, podem determinar o desempenho. Pode-se inferir que, no geral, estas regras acabaram se tornando específicas para um usuário ou uma disciplina, em que foram associados diretamente o identificador do usuário com a faixa de tempo do questionário, a disciplina com a faixa de tempo e o desempenho.

O algoritmo de *clusterização* K-means foi programado para geração de 4 *clusters* baseado na variável de desempenho que é dividida em 4 tipos. O objetivo deste teste foi realizar o agrupamento dos alunos de acordo com o desempenho dos mesmos, determinado anteriormente na metodologia como “insuficiente”, “fraco”, “bom” e “alto”.

O algoritmo determinou os pontos iniciais de cada *cluster* aleatoriamente, e para cada um deste, definiu um identificador de usuário, uma disciplina e a faixa de tempo do questionário. No *cluster* 0 foram determinados o identificador de usuário “688”, a disciplina “Cenários Econômicos e Mercado” e a faixa de tempo “Menos de 15 minutos”; no *cluster* 1 o identificador “677”, a disciplina “Gestão de Pessoas” e o tempo “entre 30 e 45 minutos”; o *cluster* 2 “655”, “Comércio Eletrônico e Automação Comercial” e “Menos de 15 minutos”; e por fim o *cluster* 3 “988”, “Gestão de Qualidade” e “Menos de 15 minutos”. Na figura 7 são apresentados os centroides finais de cada *cluster* e cada um dos três atributos:

Figura 7 - Centroides gerados pelo algoritmo K-Means para o dataset “B”.

Atributo	Clusters			
	0	1	2	3
<i>userid</i>	688	703	665	988
disciplina	Metodologia em EaD	Gestão de Pessoas	Comércio Eletrônico e Automação Comercial	Gestão da Qualidade
faixa_tempo_quiz	Menos de 15 Minutos	Entre 30 e 45 minutos	Menos de 15 minutos	Menos de 15 minutos

Fonte: Própria (2019)

O tempo para geração do modelo foi de 0,01 segundos, considerado muito baixo. No *cluster* 0 estão localizados 81% das instâncias, o que representa a maior parte delas, enquanto os *clusters* 1 e 2 obtiveram 6% das instâncias e por fim o *cluster* 3 obteve 7%. O *cluster* 0 foi classificado pelo atributo de desempenho como fraco, o *cluster* 1 como alto, o 2 como insuficiente e o 3 como bom. Por fim, o algoritmo retornou uma taxa de erros de 58,63%, ou seja, não foi possível realizar a classificação dos dados em grupos de desempenho. O relatório completo desta análise está disponível no apêndice E.

Na classificação foram testados os algoritmos *Bayes Net*, *Naive Bayes*, *Decision Table*, *J48* e *Random Forest*. Todos os algoritmos utilizaram os parâmetros padrões, com 66% dos dados para treinamento e o restante para validação, a partir da variável de desempenho. A tabela 1 apresenta a média do MCC de cada teste de algoritmo e o desempenho do algoritmo. O MCC é o coeficiente de correlação de *Matthews* que mede a qualidade da classificação, quanto mais próxima de 1 melhor foi a predição. Já o desempenho do algoritmo está relacionado à quantidade de classificações corretas.

Tabela 1 - MCC e desempenho dos algoritmos de classificação aplicados no dataset “B”.

Algoritmo	MCC	Desempenho
Bayes Net	0,430	62,5%
Naive Bayes	0,427	63,6%
Decision Table	0,404	61,3%
J48	0,392	61,3%
Random Forest	0,297	54,1%

Fonte: Própria (2019)

Todos os algoritmos apresentam uma qualidade de classificação baixa, sendo o *Bayes Net* o que possui a maior qualidade e o *Random Forest* a menor qualidade dentre estes. Quanto ao desempenho, o algoritmo *Naive Bayes* é o que possui o maior e o algoritmo *Random Forest* apresentou o menor. Desta forma é possível inferir que a qualidade da classificação não necessariamente implica no

desempenho do algoritmo, visto que o algoritmo com maior índice de desempenho não possui o maior índice de qualidade de classificação e vice-versa.

O algoritmo *Bayes Net* classificou 65% das instâncias corretamente, porém a estatística *Kappa* obteve 0,4507, que significa uma baixa concordância de previsão em relação a verdadeira classe. As classes foram divididas em “fraco”, “bom”, “alto” e “insuficiente”, tendo respectivamente o verdadeiro positivo 0,752, 0,722, 0,240 e 0,375, que indica a proporção em que cada interação foi classificada em cada uma destas classes. A precisão obteve a média de 0,651, que apresenta dentre todos que foram classificados em determinada classe, a proporção dos que realmente pertencem àquela classe. O relatório completo desta análise consta no apêndice F.

O algoritmo *Naive Bayes* classificou 65,15% das instâncias corretamente, porém obteve 0,4393 na estatística *Kappa*. Foi classificado também com as classes “fraco”, “bom”, “alto” e “insuficiente”, com os valores de verdadeiro positivo próximo ao algoritmo anterior para “fraco” e “bom” e contrários para as classes “alto” e “insuficiente” sendo 0,320 e 0,292 respectivamente. A precisão média se manteve próxima, com valor 0,641. O relatório completo desta análise consta no apêndice G.

O algoritmo *Decision Table* classificou corretamente 61,36% das instâncias de forma correta e obteve o *Kappa* próximo aos dos algoritmos anteriores, em 0,4052. Gerou as classes da mesma maneira que os outros, porém obteve o verdadeiro positivo mais alto nas classes “alto”, 0,400, e “insuficiente”, 0,583, e um pouco mais baixo nas classes “fraco” e “bom”, 0,616 e 0,678. A média da precisão também não obteve diferença considerável, sendo 0,623. O relatório completo desta análise consta no apêndice H.

O algoritmo *J48* obteve 62,12% de classificações corretas e 0,4021 na estatística *Kappa*. Obteve média do verdadeiro positivo em 0,621, média da precisão em 0,622 e, por fim, gerou as classes da mesma maneira que os outros algoritmos (Apêndice I).

O algoritmo *Randon Forest* também não apresentou resultados muito discrepantes em relação aos anteriores, com 61% de classificações corretas, obtendo a menor estatística *Kappa* dentre os outros, 0,3893, o verdadeiro positivo com média de 0,617 e a precisão com média de 0,618 sendo também a menor dentre os outros (Apêndice J).

4.4. Algoritmos de Associação, *Clusterização* e Classificação no Dataset “B” sem a Coluna do Usuário

Neste segundo teste o algoritmo *Apriori* retornou a mesma associação, desempenho fraco associada a faixa de tempo do questionário, com 70% de confiança e 1.11 de *lift* (Apêndice K). Sendo assim, no caso deste algoritmo, não houve diferença no resultado entre os *datasets* com identificador de usuário e sem o identificador.

Neste teste o algoritmo *Predictive Apriori* também retornou 100 regras, porém a acurácia das regras entre 95,61% e 56,53%, sendo consideravelmente diferente do primeiro teste. Neste caso, como não haviam identificadores de usuários, 96 das 100 regras possuem uma disciplina específica. As 10 primeiras regras são apresentadas na tabela 6 e o relatório completa desta análise consta no apêndice L.

Figura 8 - 10 melhores regras geradas pelo algoritmo *Predictive Apriori* aplicado no dataset “B” sem a coluna do usuário.

Melhores Regras
1. disciplina=Administração Financeira faixa_tempo_quiz=Menos de 15 minutos 25 ==> desempenho=fraco 24 acc:(0.9561)
2. disciplina=Ética faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==> desempenho=bom 6 acc:(0.94358)
3. desempenho=insuficiente disciplina=Empreendedorismo 5 ==> faixa_tempo_quiz=Menos de 15 minutos 5 acc:(0.92733)
4. disciplina=Gestão de Pessoas faixa_tempo_quiz=Entre 15 e 30 minutos 5 ==> desempenho=fraco 5 acc:(0.92733)
5. disciplina=Empreendedorismo 34 ==> faixa_tempo_quiz=Menos de 15 minutos 31 acc:(0.87726)
6. disciplina=Administração Financeira 34 ==> desempenho=fraco 31 acc:(0.87726)
7. desempenho=alto disciplina=Direito 3 ==> faixa_tempo_quiz=Menos de 15 minutos 3 acc:(0.86733)
8. disciplina=Redação de Textos Oficiais faixa_tempo_quiz=Entre 45 minutos e 1 hora 3 ==> desempenho=bom 3 acc:(0.86733)

9. disciplina=Marketing faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 3 acc:(0.86733)
10. disciplina=Organização faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 3 acc:(0.86733)

Fonte: Própria (2019)

Na primeira regra a disciplina de Administração Financeira e a faixa de tempo “menor que 15 minutos” estão associadas ao desempenho fraco com a acurácia de 95,61% e na sexta regra a mesma disciplina está associada ao desempenho fraco. A disciplina de Ética e o tempo do questionário “entre 15 e 30 minutos” estão associados ao bom desempenho, na segunda regra. A disciplina de Empreendedorismo aparece nas regras três e cinco, sendo na três, junto ao desempenho insuficiente, associada ao tempo “menor que 15 minutos”, e na cinco, associada diretamente ao tempo “menor que 15 minutos”. A disciplina Gestão de Pessoas, Marketing e Organização, estão juntas ao tempo “entre 15 e 30 minutos” e associadas ao desempenho fraco. Por fim a disciplina de Direito e o desempenho alto associam-se ao tempo “menor que 15 minutos” e a disciplina de Redação de Textos Oficiais com o tempo “entre 45 minutos e 1 hora” se associam ao bom desempenho.

A associação das disciplinas apresenta o mesmo panorama que a associação dos discentes, no teste realizado da terceira sessão. Algumas disciplinas possuem o tempo “menor que 15 minutos” e são associadas ao desempenho fraco, porém outras disciplinas possuem a mesma faixa de tempo e são associadas até mesmo com o alto desempenho. No entanto, neste caso é possível visualizar com mais clareza o tempo gasto associado as disciplinas.

O algoritmo de *clusterização* foi testado da mesma maneira, para geração de 4 *clusters*, buscando o agrupamento dos alunos de acordo com o desempenho. Os pontos iniciais dos *clusters*, para os atributos disciplina e faixa de tempo foram:

- *Cluster* 0: 'Cenários Econômicos e Mercado', 'Menos de 15 minutos';
- *Cluster* 1: 'Gestão de Pessoas', 'Entre 30 e 45 minutos';
- *Cluster* 2: 'Comercio Eletrônico e Automação Comercial', 'Menos de 15 minutos';
- *Cluster* 3: 'Gestão da Qualidade', 'Menos de 15 minutos'.

Na figura 9 são apresentados os centroides de cada *cluster* para cada um dos dois atributos e no apêndice M consta o relatório completo desta análise:

Figura 9 - Centroides gerados pelo algoritmo *K-Means* para o *dataset* “B” sem a coluna de usuário.

Atributo	Clusteres			
	0	1	2	3
disciplina	Metodologia em EaD	Gestão de Pessoas	Comércio Eletrônico e Automação Comercial	Gestão da Qualidade
faixa_tempo_quiz	Menos de 15 Minutos	Entre 30 e 45 minutos	Menos de 15 minutos	Menos de 15 minutos

Fonte: Própria (2019)

O tempo para geração deste modelo foi de 0,01 segundos, bem como o teste feito na sessão anterior. No *cluster* 0 estão localizados 87% das instâncias, enquanto nos *clusteres* 1 e 2 estão localizados 4% em cada e no *cluster* 3 os 5% restantes. Cada *cluster* foi classificado de acordo com o atributo de desempenho e neste caso, *cluster* 0 corresponde ao grupo de desempenho “fraco”, o *cluster* 1 ao grupo de desempenho “bom”, o 2 “alto” e o 3 “insuficiente”. Em relação ao outro teste, houve diferença na ordem dos grupos, porém o grupo denominado como “fraco” em ambos os testes é o que possui a maior concentração de dados em ambos os casos. Por fim houve uma taxa de erro de 56,44%, com uma melhora de 2,16% em relação ao teste da sessão anterior, porém ainda demonstrando não ser capaz de realizar a classificação dos dados de acordo com o desempenho.

Os algoritmos de classificação, *Bayes Net*, *Naive Bayes*, *Decision Table*, *J48* e *Random Forest*, foram testados utilizando os mesmos parâmetros. A tabela 2 mostra o desempenho de cada algoritmo e o MCC. Os relatórios completos de cada algoritmo constam nos apêndices N, O, P, Q e R, correspondendo respectivamente aos algoritmos *Bayes Net*, *Naive Bayes*, *Decision Table*, *J48* e *Random Forest*.

Tabela 2 - MCC e desempenho dos algoritmos de classificação aplicados no dataset “B” sem a coluna de usuário.

Algoritmo	MCC	Desempenho
Bayes Net	0,430	62,5%
Naive Bayes	0,435	63,6%
Decision Table	0,404	61,3%
J48	0,402	62,1%
Random Forest	0,366	59,4%

Fonte: Própria (2019)

O algoritmo com melhor desempenho e com maior MCC foi o *Naive Bayes*, enquanto o algoritmo *Random Forest* teve o menor desempenho e o menor MCC. Os resultados se assemelham ao teste anterior, com exceção que neste caso o *Naive Bayes* apresentou também o maior MCC e no outro teste o *Bayes Net* teve o maior. O algoritmo *Random Forest* tem os menores índices, porém neste teste apresentou uma pequena melhora em ambos os parâmetros. Por fim, é possível concluir que os algoritmos possuem um desempenho razoável, porém as classificações não possuem uma boa qualidade.

5. CONCLUSÃO

A base de dados do *software Moodle* é muito rica em informações. A tabela utilizada neste trabalho possui 20 colunas e pode ser analisada de várias maneiras. Nota-se que neste curso específico, o recurso mais utilizado foi o questionário e que a maioria dos trabalhos em Mineração de Dados na Educação fazem análises sob a perspectiva de uso dos fóruns. Diante do fator de maior utilização do módulo de questionário, o presente trabalho buscou verificar a ligação do desempenho do aluno ao tempo de utilização do questionário e as notas obtidas.

Com base nos resultados obtidos na aplicação dos algoritmos de associação nos dois *datasets*, “A” e “B”, e algoritmos de classificação e *clusterização* no *dataset* “B” com e sem a coluna de identificação de usuário, conclui-se que não foi possível avaliar se de fato as variáveis de média de tempo gasto no questionário e média da nota, ambas por disciplina, implicam no desempenho do aluno.

Os testes realizados no *dataset* “A” não retornaram regras de associação no algoritmo *Apriori*, pois este é um algoritmo que gera regras muito específicas. Apesar do *Predictive Apriori* ter gerado 100 regras diferentes, estas não apresentaram informações significantes em relação ao uso do software pelos usuários; as regras foram geradas de forma muito específica para cada usuário e data.

A aplicação no *dataset* “B” do algoritmo de associação *Apriori*, retornou uma única regra indicando que o desempenho do aluno é ruim quando o tempo gasto no questionário é inferior à 15 minutos, sendo esta regra apresentada com um nível de confiança considerável. Quando o algoritmo retorna mais de uma regra, o índice do *lift* pode ser utilizado para verificar qual a melhor regra, sendo esta a que possui o *lift* mais alto. Contudo, neste caso, não foi possível verificar se seria a melhor regra pois foi a única retornada.

O algoritmo *Predictive Apriori* também retornou 100 regras na aplicação deste mesmo *dataset* “B”, porém as regras se mostraram muito específicas para um único usuário. No segundo teste, sem a coluna do usuário, as regras se tornaram muito específicas para cada disciplina. Neste caso não foi possível identificar também os casos gerais, pois em ambas as situações, as regras foram geradas de forma específica.

Na *clusterização*, utilizando o algoritmo *K-means*, o mesmo determinou os pontos iniciais da variável de tempo em três *clusteres*, culminando em centroides próximos também, o que pode ser um dos motivos pelo qual o algoritmo não foi capaz de realizar o agrupamento, com a taxa de erro de 58,63%. Ao realizar o mesmo teste sem a coluna do usuário, houve uma pequena diminuição na taxa de erro, mas o centroide da faixa de tempo se repetiu da mesma maneira. Sendo assim, o resultado indica que não há agrupamento de usuários.

Os algoritmos de classificação retornaram um desempenho regular, porém a qualidade de classificação de nenhum dos testes passou de 50%. No primeiro teste, os algoritmos demonstraram que não necessariamente o melhor desempenho implica no melhor classificador. No segundo teste, os algoritmos demonstraram uma pequena melhora tanto no desempenho quanto na qualidade de classificação. Mesmo verificando a análise detalhada de cada algoritmo, nenhum destes foi capaz de classificar os dados.

Por fim, conclui-se que diante dos objetivos traçados, de identificar se o desempenho tem relação com a forma com que os discentes utilizam a plataforma e os algoritmos que possuem o melhor desempenho nesta análise, não foi possível obter as respostas esperadas.

REFERÊNCIAS

- AHER, S. B.; LOBO, L. M. R. J. A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning. **International Journal of Computer Applications**, v. 39, n. 1, p. 48–52, 2012.
- APON, A. et al. Initial Starting Point Analysis for K-Means Clustering : A Case Study. **Clemson University - School of Computing**, 2006.
- BAKER, R. J. DE; ISOTANI, S.; CARVALHO, A. M. J. B. DE. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, p. 3–13, 2011.
- BRASIL. **Lei Geral de Proteção dos Dados nº13.709 de 14 de Agosto de 2018**. Brasil, 2018.
- CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, v. 42, n. 3, p. 495–528, 2008.
- COSTA, E. et al. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Jornada de Atualização em Informática na Educação - JAIE 2012**, p. 1–29, 2012.
- EAD. **O que é EaD?** Disponível em: <<https://www.ead.com.br/ead/o-que-e-ead.html>>. Acesso em: 26 mar. 2019.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, p. 37–54, 1996.
- FRANCO, M. A.; CORDEIRO, L. M.; CASTILLO, R. A. F. DEL. O ambiente virtual de aprendizagem e sua incorporação na Unicamp. **Educação e Pesquisa**, v. 29, n. 2, p. 341–353, 2003.
- FRANK, E. et al. **The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”**. 4. ed. [s.l: s.n.].
- GARCIA, E. et al. Drawbacks and solutions of applying association rule mining in learning management systems. **Proceedings of the International Workshop on Applying Data Mining in e-Learning 2007**, p. 13–22, 2007.
- GUÉRCIO, H. et al. Análise do Desempenho Estudantil na Educação a Distância Aplicando Técnicas de Mineração de Dados. **Anais dos Workshops do III Congresso Brasileiro de Informática na Educação (CBIE 2014)**, p. 641–650, 2014.

IFGOIANO. **Instituto Federal Goiano Educação a Distância**. Disponível em: <<https://eadmoodle.ifgoiano.edu.br/>>. Acesso em: 11 maio. 2019.

LEONARDO, E. DA S.; BHERING, L. S.; GOMES, S. G. S. Utilização Do Ambiente Virtual De Aprendizagem PVANet Para Formação de Professores. **Texto Livre: Linguagem e Tecnologia**, v. 7, n. 1, p. 14–25, 2014.

MENDES, R. DE Á.; SILVA, L. A. DA. Mineração de Dados Educacionais Aplicada à Análise Preditiva em Fóruns no Moodle. **V Congresso Brasileiro de Informática na Educação (CBIE 2016)**, p. 940–949, 2016.

MOODLE. **About Moodle**. Disponível em: <https://docs.moodle.org/36/en/About_Moodle>. Acesso em: 25 mar. 2018.

PATIL, B. M.; JOSHI, R. C.; TOSHNIWAL, D. Classification of type-2 diabetic patients by using Apriori and predictive Apriori. **Int. J. Computational Vision and Robotics**, v. 2, n. 3, p. 254–265, 2011.

PINHEIRO, M. F. et al. Identificação de Grupos de Alunos em Ambiente Virtual de Aprendizagem: Uma Estratégia de Análise de Log Baseada em Clusterização. **Anais dos Workshops do III Congresso Brasileiro de Informática na Educação (CBIE 2014)**, p. 582–591, 2014.

PNP. **Plataforma Nilo Peçanha**. Disponível em: <<http://resultados.plataformanilopecanha.org/2018/>>. Acesso em: 9 maio. 2018.

QUEIROGA, E. M.; CECHINEL, C.; ARAUJO, R. M. Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância. n. **CBIE-LACLO**, p. 1074–1083, 2015.

RODRIGUES, R. L. et al. A literatura brasileira sobre mineração de dados educacionais. **Anais dos Workshops do III Congresso Brasileiro de Informática na Educação (CBIE 2014)**, p. 621–639, 2014.

SANTOS, E. O. DOS. Ambientes Virtuais de Aprendizagem: Por Autorias Livres, Plurais e Gratuitas. **Revista FAEBA**, v. 11, p. 425–435, 2002.

SANTOS, F. D. et al. Análise de evidências do estado de ânimo desanimado de alunos de um AVEA: uma proposta a partir da aplicação de regras de associação. **Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE 2015)**, p. 1054–1063, 2015.

SILVA, L. A. et al. Mineração de Dados em publicações de Fóruns de Discussões do Moodle como geração de Indicadores para aprimoramento da Gestão Educacional. **Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE 2015)**, p. 1084–1093, 2015.

SILVA, L. A.; LOPES, F. S.; FILHO, A. R. DE A. V. FI. Uma proposta de Modelagem Analítica para o uso efetivo de Atividades Complementares no Ensino Superior. **Anais dos Workshops do III Congresso Brasileiro de Informática na Educação (CBIE 2014)**, p. 611–620, 2014.

SILVA, L. A.; MORINO, A. H.; MASSAHIRO, T. C. S. Prática de Mineração de Dados no Exame Nacional do Ensino Médio. **Anais dos Workshops do III Congresso Brasileiro de Informática na Educação (CBIE 2014)**, p. 651–660, 2014.

WEBBER, C. G.; CRISTOFOLI, L. G.; LIMA, M. DE F. W. DO P. UnderMine Text Miner – Uma Ferramenta de Mineração de Texto para Área Educacional. **Novas Tecnologias na Educação**, v. 11, 2013.

APÊNDICES

APÊNDICE A – RELATÓRIO DO ALGORITMO *PREDICTIVE APRIORI* NO DATASET A COM O CAMPO DATA NO FORMATO DIA/MÊS/ANO

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1
Relation: eventname_userid_data-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 113617
Attributes: 3
 eventname
 userid
 data

=== Associator model (full training set) ===

PredictiveApriori

=====

Best rules found:

1. eventname=\core\event\grade_deleted data=2017-11-28 250 ==> userid=712 250 acc:(0.99495)
2. eventname=\mod_quiz\event\edit_page_viewed data=2018-08-13 126 ==> userid=753 126 acc:(0.99472)
3. eventname=\core\event\user_graded data=2018-06-21 107 ==> userid=753 107 acc:(0.9946)
4. eventname=\core\event\user_profile_viewed data=2017-11-08 97 ==> userid=655 97 acc:(0.99451)
5. eventname=\report_outline\event\report_viewed data=2017-11-08 90 ==> userid=655 90 acc:(0.99442)
6. eventname=\core\event\grade_deleted data=2018-02-01 90 ==> userid=712 90 acc:(0.99442)
7. eventname=\gradereport_grader\event\grade_report_viewed data=2017-11-26 162 ==> userid=755 161 acc:(0.99438)
8. eventname=\core\event\user_graded userid=768 86 ==> data=2018-03-19 86 acc:(0.99437)
9. eventname=\mod_quiz\event\edit_page_viewed data=2018-08-08 83 ==> userid=753 83 acc:(0.99432)

10. eventname=\core\event\user_graded data=2018-04-24 83 ==> userid=753 83
acc:(0.99432)

11. eventname=\gradereport_singleview\event\grade_report_viewed data=2018-02-03 82 ==> userid=712 82 acc:(0.9943)

12. eventname=\report_outline\event\report_viewed data=2017-11-09 76 ==> userid=655 76 acc:(0.99419)

13. eventname=\mod_quiz\event\attempt_viewed data=2017-11-10 74 ==> data=2017-11-10 74 acc:(0.99414)

14. eventname=\mod_quiz\event\attempt_viewed data=2018-12-14 74 ==> userid=711 74 acc:(0.99414)

15. data=2018-06-16 65 ==> userid=3083 65 acc:(0.99389)

16. data=2018-05-12 64 ==> userid=703 64 acc:(0.99386)

17. data=2018-10-27 55 ==> userid=703 55 acc:(0.99345)

18. eventname=\core\event\grade_deleted data=2017-12-08 53 ==> userid=757 53
acc:(0.99333)

19. data=2018-07-04 52 ==> userid=655 52 acc:(0.99327)

20. data=2018-05-19 51 ==> userid=679 51 acc:(0.9932)

21. data=2018-06-06 51 ==> userid=1854 51 acc:(0.9932)

22. data=2017-10-03 49 ==> userid=757 49 acc:(0.99305)

23. eventname=\core\event\user_list_viewed data=2017-11-08 48 ==> userid=655 48
acc:(0.99297)

24. eventname=\core\event\course_viewed data=2018-08-18 46 ==> userid=688 46
acc:(0.99279)

25. eventname=\core\event\grade_deleted data=2018-02-04 46 ==> userid=662 46
acc:(0.99279)

26. eventname=\core\event\grade_deleted data=2018-02-03 45 ==> userid=712 45
acc:(0.99268)

27. eventname=\mod_quiz\event\attempt_reviewed data=2017-11-10 45 ==> userid=755 45
acc:(0.99268)

28. eventname=\gradereport_user\event\grade_report_viewed data=2018-11-12 44
==> userid=754 44 acc:(0.99258)

29. eventname=\core\event\grade_deleted data=2017-11-06 43 ==> userid=757 43
acc:(0.99246)

30. eventname=\core\event\grade_deleted data=2017-12-04 43 ==> userid=712 43
acc:(0.99246)

31. data=2018-05-26 42 ==> userid=703 42 acc:(0.99234)

32. data=2018-08-21 42 ==> userid=3048 42 acc:(0.99234)

33. data=2018-08-10 41 ==> userid=655 41 acc:(0.9922)

34. data=2018-10-12 41 ==> userid=703 41 acc:(0.9922)

35. eventname=\mod_quiz\event\attempt_viewed data=2018-05-15 40 ==> userid=682 40
acc:(0.99206)

36. eventname=\gradereport_grader\event\grade_report_viewed data=2018-08-16

39 ==> userid=755 39 acc:(0.9919)
 37. userid=755 data=2018-08-16 39 ==>
 eventname=\graderreport_grader\event\grade_report_viewed 39 acc:(0.9919)
 38. eventname=\core\event\user_graded data=2018-06-26 37 ==> userid=753 37
 acc:(0.99154)
 39. data=2018-11-02 36 ==> userid=703 36 acc:(0.99134)
 40. eventname=\core\event\grade_deleted data=2018-06-15 36 ==> userid=1854 36
 acc:(0.99134)
 41. data=2018-09-22 35 ==> userid=703 35 acc:(0.99112)
 42. eventname=\mod_quiz\event\attempt_viewed data=2018-02-17 35 ==>
 userid=991 35 acc:(0.99112)
 43. eventname=\core\event\grade_deleted data=2017-10-25 35 ==> userid=758 35
 acc:(0.99112)
 44. eventname=\mod_quiz\event\attempt_viewed data=2018-10-05 34 ==>
 userid=692 34 acc:(0.99089)
 45. eventname=\core\event\user_profile_viewed data=2018-08-31 33 ==>
 userid=755 33 acc:(0.99062)
 46. eventname=\core\event\user_profile_viewed data=2018-09-25 33 ==>
 userid=755 33 acc:(0.99062)
 47. eventname=\mod_quiz\event\attempt_viewed data=2017-11-15 33 ==>
 userid=703 33 acc:(0.99062)
 48. eventname=\mod_quiz\event\attempt_viewed data=2018-01-13 33 ==>
 userid=679 33 acc:(0.99062)
 49. eventname=\core\event\course_viewed data=2018-10-14 31 ==> userid=757 31
 acc:(0.99002)
 50. data=2017-12-26 73 ==> userid=711 72 acc:(0.98995)
 51. eventname=\mod_quiz\event\attempt_viewed data=2018-08-14 30 ==>
 userid=692 30 acc:(0.98968)
 52. eventname=\mod_quiz\event\attempt_viewed data=2018-08-24 29 ==>
 userid=680 29 acc:(0.9893)
 53. data=2018-10-13 28 ==> userid=672 28 acc:(0.98888)
 54. data=2018-06-10 27 ==> userid=655 27 acc:(0.98841)
 55. data=2018-07-29 27 ==> userid=753 27 acc:(0.98841)
 56. eventname=\mod_quiz\event\attempt_viewed data=2018-11-01 27 ==>
 userid=711 27 acc:(0.98841)
 57. eventname=\mod_quiz\event\edit_page_viewed data=2018-10-22 26 ==>
 userid=655 26 acc:(0.98789)
 58. eventname=\mod_quiz\event\edit_page_viewed data=2018-11-30 26 ==>
 userid=753 26 acc:(0.98789)
 59. data=2018-08-26 25 ==> userid=683 25 acc:(0.98732)
 60. eventname=\core\event\course_module_deleted data=2017-11-28 23 ==>
 userid=712 23 acc:(0.98596)

61. eventname=\mod_quiz\event\edit_page_viewed data=2018-10-23 23 ==>
 userid=655 23 acc:(0.98596)
 62. data=2017-09-21 22 ==> userid=757 22 acc:(0.98516)
 63. data=2018-07-27 22 ==> userid=3048 22 acc:(0.98516)
 64. eventname=\gradereport_grader\event\grade_report_viewed data=2018-09-16
 22 ==> userid=753 22 acc:(0.98516)
 65. eventname=\mod_quiz\event\edit_page_viewed data=2018-05-28 22 ==>
 userid=3083 22 acc:(0.98516)
 66. eventname=\mod_quiz\event\attempt_viewed data=2018-10-09 22 ==>
 userid=683 22 acc:(0.98516)
 67. eventname=\gradereport_grader\event\grade_report_viewed data=2017-10-21
 21 ==> userid=754 21 acc:(0.98426)
 68. eventname=\core\event\user_profile_viewed data=2018-10-03 21 ==>
 userid=3048 21 acc:(0.98426)
 69. eventname=\core\event\grade_deleted data=2018-02-19 21 ==> userid=662 21
 acc:(0.98426)
 70. eventname=\core\event\user_list_viewed data=2017-10-23 152 ==> userid=655
 149 acc:(0.984)
 71. data=2018-07-03 20 ==> userid=655 20 acc:(0.98325)
 72. eventname=\core\event\course_viewed data=2018-06-17 20 ==> userid=655 20
 acc:(0.98325)
 73. eventname=\mod_quiz\event\course_module_viewed data=2017-11-15 20 ==>
 userid=703 20 acc:(0.98325)
 74. eventname=\core\event\user_profile_viewed data=2018-09-19 20 ==>
 userid=755 20 acc:(0.98325)
 75. eventname=\mod_quiz\event\attempt_viewed data=2018-01-18 20 ==>
 userid=664 20 acc:(0.98325)
 76. eventname=\mod_quiz\event\attempt_viewed data=2018-02-27 20 ==>
 userid=699 20 acc:(0.98325)
 77. data=2017-10-14 52 ==> userid=758 51 acc:(0.98318)
 78. eventname=\core\event\course_viewed data=2018-02-03 52 ==> userid=712 51
 acc:(0.98318)
 79. userid=4679 19 ==> data=2018-10-30 19 acc:(0.98212)
 80. data=2017-10-15 19 ==> userid=757 19 acc:(0.98212)
 81. data=2018-11-28 19 ==> userid=692 19 acc:(0.98212)
 82. eventname=\gradereport_grader\event\grade_report_viewed data=2018-08-03
 19 ==> userid=655 19 acc:(0.98212)
 83. eventname=\gradereport_grader\event\grade_report_viewed data=2018-08-31
 19 ==> userid=755 19 acc:(0.98212)
 84. eventname=\core\event\course_module_deleted data=2018-02-03 19 ==>
 userid=712 19 acc:(0.98212)
 85. eventname=\mod_quiz\event\edit_page_viewed data=2018-03-26 19 ==>

userid=757 19 acc:(0.98212)
 86. eventname=\mod_quiz\event\edit_page_viewed data=2018-04-26 19 ==>
 userid=757 19 acc:(0.98212)
 87. eventname=\mod_quiz\event\edit_page_viewed data=2018-08-09 19 ==>
 userid=757 19 acc:(0.98212)
 88. eventname=\gradereport_grader\event\grade_report_viewed data=2017-11-08
 92 ==> userid=755 90 acc:(0.98205)
 89. data=2017-09-28 18 ==> userid=757 18 acc:(0.98083)
 90. eventname=\core\event\course_module_deleted data=2018-03-14 18 ==>
 userid=2 18 acc:(0.98083)
 91. eventname=\mod_forum\event\post_deleted data=2018-03-19 18 ==>
 userid=757 18 acc:(0.98083)
 92. data=2018-04-04 17 ==> userid=682 17 acc:(0.97937)
 93. eventname=\core\event\user_graded data=2018-09-16 17 ==> userid=753 17
 acc:(0.97937)
 94. data=2018-08-01 46 ==> userid=3048 45 acc:(0.979)
 95. data=2018-04-22 16 ==> userid=757 16 acc:(0.97771)
 96. data=2018-05-05 16 ==> userid=757 16 acc:(0.97771)
 97. eventname=\core\event\user_graded userid=686 16 ==> data=2017-11-10 16
 acc:(0.97771)
 98. data=2019-05-09 15 ==> userid=655 15 acc:(0.9758)
 99. eventname=\gradereport_grader\event\grade_report_viewed data=2018-08-24
 15 ==> userid=753 15 acc:(0.9758)
 100. eventname=\core\event\user_list_viewed data=2017-12-20 15 ==> userid=655
 15 acc:(0.9758)

APÊNDICE B – RELATÓRIO DO ALGORITMO *PREDICTIVE APRIORI* NO DATASET A COM O CAMPO DATA NO FORMATO MÊS/ANO

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1
Relation: usuario_eventname_data2-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances: 113516
Attributes: 3
 userid
 eventname
 ano_mes

=== Associator model (full training set) ===

PredictiveApriori

=====

Best rules found:

1. userid=758 eventname=\core\event\grade_deleted 99 ==>
ano_mes=outubro/2017 99 acc:(0.99441)
2. userid=709 eventname=\mod_quiz\event\attempt_viewed 97 ==>
ano_mes=novembro/2017 97 acc:(0.99438)
3. userid=712 eventname=\gradereport_singleview\event\grade_report_viewed 95
==> ano_mes=fevereiro/2018 95 acc:(0.99436)
4. userid=768 eventname=\core\event\user_graded 86 ==> ano_mes=março/2018
86 acc:(0.99421)
5. userid=712 eventname=\core\event\user_graded 77 ==>
ano_mes=fevereiro/2018 77 acc:(0.99401)
6. userid=686 eventname=\mod_quiz\event\attempt_viewed 74 ==>
ano_mes=novembro/2017 74 acc:(0.99392)
7. userid=779 eventname=\core\event\grade_deleted 58 ==>
ano_mes=dezembro/2017 58 acc:(0.99312)
8. userid=655 eventname=\mod_quiz\event\edit_page_viewed 49 ==>
ano_mes=outubro/2018 49 acc:(0.99207)
9. eventname=\mod_quiz\event\edit_page_viewed ano_mes=outubro/2018 49 ==>
userid=655 49 acc:(0.99207)
10. userid=709 191 ==> ano_mes=novembro/2017 188 acc:(0.98863)
11. eventname=\report_participation\event\report_viewed ano_mes=agosto/2018 34

==> userid=3048 34 acc:(0.98607)
 12. userid=3083 eventname=\core\event\grade_deleted 33 ==>
 ano_mes=junho/2018 33 acc:(0.98523)
 13. userid=758 eventname=\mod_quiz\event\attempt_viewed 29 ==>
 ano_mes=outubro/2017 29 acc:(0.98078)
 14. eventname=\report_participation\event\report_viewed ano_mes=junho/2018 28
 ==> userid=3048 28 acc:(0.97935)
 15. eventname=\gradereport_singleview\event\grade_report_viewed
 ano_mes=outubro/2018 26 ==> userid=753 26 acc:(0.97602)
 16. eventname=\core\event\grade_deleted ano_mes=novembro/2018 23 ==>
 userid=1854 23 acc:(0.96967)
 17. eventname=\core\event\calendar_event_created ano_mes=agosto/2018 22 ==>
 userid=753 22 acc:(0.96715)
 18. eventname=\gradereport_user\event\grade_report_viewed
 ano_mes=novembro/2018 45 ==> userid=754 44 acc:(0.9621)
 19. userid=4679 19 ==> ano_mes=outubro/2018 19 acc:(0.95828)
 20. eventname=\core\event\course_module_updated ano_mes=dezembro/2018 19
 ==> userid=753 19 acc:(0.95828)
 21. eventname=\mod_forum\event\post_deleted ano_mes=março/2018 18 ==>
 userid=757 18 acc:(0.95488)
 22. eventname=\report_log\event\report_viewed ano_mes=outubro/2018 17 ==>
 userid=754 17 acc:(0.95125)
 23. eventname=\report_outline\event\report_viewed ano_mes=novembro/2017 178
 ==> userid=655 170 acc:(0.9495)
 24. userid=686 eventname=\core\event\user_graded 16 ==>
 ano_mes=novembro/2017 16 acc:(0.94742)
 25. userid=655 eventname=\report_outline\event\report_viewed 179 ==>
 ano_mes=novembro/2017 170 acc:(0.94349)
 26. eventname=\report_outline\event\report_viewed ano_mes=agosto/2018 13 ==>
 userid=3048 13 acc:(0.93467)
 27. eventname=\mod_quiz\event\report_viewed ano_mes=maio/2018 13 ==>
 userid=1854 13 acc:(0.93467)
 28. eventname=\core\event\grade_deleted ano_mes=maio/2018 12 ==> userid=2 12
 acc:(0.93002)
 29. eventname=\report_participation\event\report_viewed 71 ==> userid=3048 68
 acc:(0.92563)
 30. userid=686 eventname=\mod_quiz\event\attempt_summary_viewed 11 ==>
 ano_mes=novembro/2017 11 acc:(0.92515)
 31. userid=3083 eventname=\mod_folder\event\course_module_viewed 11 ==>
 ano_mes=maio/2018 11 acc:(0.92515)
 32. userid=686 eventname=\mod_quiz\event\attempt_started 10 ==>
 ano_mes=novembro/2017 10 acc:(0.92002)

33. userid=686 eventname=\mod_quiz\event\attempt_submitted 10 ==>
ano_mes=novembro/2017 10 acc:(0.92002)

34. userid=768 eventname=\mod_forum\event\discussion_viewed 10 ==>
ano_mes=fevereiro/2018 10 acc:(0.92002)

35. eventname=\mod_quiz\event\attempt_viewed ano_mes=julho/2018 10 ==>
userid=753 10 acc:(0.92002)

36. userid=662 eventname=\mod_folder\event\course_module_viewed 9 ==>
ano_mes=novembro/2017 9 acc:(0.91452)

37. userid=758 eventname=\core\event\course_module_deleted 9 ==>
ano_mes=outubro/2017 9 acc:(0.91452)

38. userid=760 eventname=\mod_folder\event\course_module_viewed 9 ==>
ano_mes=maio/2018 9 acc:(0.91452)

39. eventname=\core\event\course_updated ano_mes=agosto/2018 9 ==>
userid=753 9 acc:(0.91452)

40. eventname=\gradereport_singleview\event\grade_report_viewed
ano_mes=agosto/2018 29 ==> userid=753 28 acc:(0.91142)

41. userid=236 eventname=\core\event\course_module_updated 8 ==>
ano_mes=março/2018 8 acc:(0.90849)

42. userid=655 eventname=\gradereport_overview\event\grade_report_viewed 8
==> ano_mes=junho/2018 8 acc:(0.90849)

43. userid=3083 eventname=\mod_quiz\event\attempt_viewed 8 ==>
ano_mes=maio/2018 8 acc:(0.90849)

44. eventname=\gradereport_overview\event\grade_report_viewed
ano_mes=junho/2018 8 ==> userid=655 8 acc:(0.90849)

45. eventname=\mod_quiz\event\report_viewed ano_mes=julho/2018 8 ==>
userid=655 8 acc:(0.90849)

46. eventname=\gradereport_singleview\event\grade_report_viewed
ano_mes=novembro/2018 8 ==> userid=655 8 acc:(0.90849)

47. eventname=\mod_forum\event\user_report_viewed ano_mes=junho/2018 8 ==>
userid=3048 8 acc:(0.90849)

48. eventname=\mod_quiz\event\attempt_reviewed ano_mes=janeiro/2018 27 ==>
userid=755 26 acc:(0.90485)

49. userid=754 eventname=\core\event\user_list_viewed 73 ==>
ano_mes=outubro/2017 69 acc:(0.90182)

50. userid=655 eventname=\report_log\event\report_viewed 7 ==>
ano_mes=novembro/2018 7 acc:(0.90157)

51. userid=702 eventname=\mod_folder\event\course_module_viewed 7 ==>
ano_mes=outubro/2017 7 acc:(0.90157)

52. userid=758 eventname=\mod_forum\event\course_module_viewed 7 ==>
ano_mes=outubro/2017 7 acc:(0.90157)

53. eventname=\core\event\calendar_event_created ano_mes=setembro/2018 7
==> userid=757 7 acc:(0.90157)

54. eventname=\report_log\event\report_viewed ano_mes=novembro/2018 7 ==> userid=655 7 acc:(0.90157)

55. userid=779 eventname=\mod_quiz\event\edit_page_viewed 25 ==> ano_mes=dezembro/2017 24 acc:(0.89869)

56. userid=1854 eventname=\mod_quiz\event\attempt_viewed 25 ==> ano_mes=março/2018 24 acc:(0.89869)

57. userid=686 161 ==> ano_mes=novembro/2017 149 acc:(0.89687)

58. userid=662 eventname=\core\event\user_graded 6 ==> ano_mes=março/2018 6 acc:(0.89308)

59. userid=675 eventname=\mod_forum\event\discussion_viewed 6 ==> ano_mes=abril/2018 6 acc:(0.89308)

60. userid=768 ano_mes=outubro/2017 6 ==> eventname=\core\event\course_viewed 6 acc:(0.89308)

61. userid=3083 eventname=\mod_folder\event\folder_updated 6 ==> ano_mes=maio/2018 6 acc:(0.89308)

62. eventname=\mod_quiz\event\edit_page_viewed ano_mes=agosto/2018 228 ==> userid=753 209 acc:(0.89149)

63. eventname=\core\event\calendar_event_updated ano_mes=dezembro/2018 21 ==> userid=753 20 acc:(0.88774)

64. userid=753 eventname=\core\event\grade_deleted 35 ==> ano_mes=junho/2018 33 acc:(0.8832)

65. userid=757 eventname=\mod_forum\event\post_created 19 ==> ano_mes=março/2018 18 acc:(0.88288)

66. userid=758 eventname=\core\event\course_module_updated 19 ==> ano_mes=outubro/2017 18 acc:(0.88288)

67. eventname=\core\event\role_unassigned 5 ==> userid=2 ano_mes=maio/2018 5 acc:(0.88163)

68. eventname=\core\event\user_enrolment_deleted 5 ==> userid=2 ano_mes=maio/2018 5 acc:(0.88163)

69. userid=236 ano_mes=fevereiro/2018 5 ==> eventname=\core\event\course_viewed 5 acc:(0.88163)

70. userid=667 eventname=\gradereport_user\event\grade_report_viewed 5 ==> ano_mes=outubro/2017 5 acc:(0.88163)

71. userid=705 eventname=\mod_forum\event\discussion_viewed 5 ==> ano_mes=outubro/2018 5 acc:(0.88163)

72. userid=712 eventname=\mod_quiz\event\edit_page_viewed 5 ==> ano_mes=setembro/2017 5 acc:(0.88163)

73. userid=758 eventname=\core\event\calendar_event_deleted 5 ==> ano_mes=outubro/2017 5 acc:(0.88163)

74. eventname=\core\event\course_section_updated ano_mes=julho/2018 5 ==> userid=3048 5 acc:(0.88163)

75. userid=754 eventname=\report_log\event\report_viewed 18 ==>

ano_mes=outubro/2018 17 acc:(0.88056)
 76. userid=702 eventname=\mod_quiz\event\attempt_viewed 137 ==>
 ano_mes=novembro/2017 125 acc:(0.87618)
 77. userid=662 eventname=\core\event\course_module_deleted 16 ==>
 ano_mes=fevereiro/2018 15 acc:(0.87595)
 78. eventname=\core\event\calendar_event_updated ano_mes=novembro/2018 16
 ==> userid=753 15 acc:(0.87595)
 79. userid=704 eventname=\core\event\course_user_report_viewed 15 ==>
 ano_mes=novembro/2017 14 acc:(0.87357)
 80. userid=779 eventname=\gradereport_singleview\event\grade_report_viewed 82
 ==> ano_mes=dezembro/2017 74 acc:(0.86954)
 81. eventname=\gradereport_singleview\event\grade_report_viewed
 ano_mes=fevereiro/2018 106 ==> userid=712 95 acc:(0.86839)
 82. userid=662 eventname=\mod_quiz\event\attempt_viewed 56 ==>
 ano_mes=fevereiro/2018 50 acc:(0.86713)
 83. eventname=\report_log\event\user_report_viewed ano_mes=novembro/2017 30
 ==> userid=655 27 acc:(0.86687)
 84. eventname=\core\event\user_list_viewed ano_mes=julho/2018 21 ==>
 userid=655 19 acc:(0.86622)
 85. userid=675 eventname=\mod_folder\event\course_module_viewed 12 ==>
 ano_mes=outubro/2017 11 acc:(0.86497)
 86. userid=2 eventname=\core\event\course_module_deleted 20 ==>
 ano_mes=março/2018 18 acc:(0.86489)
 87. userid=757 eventname=\mod_forum\event\post_deleted 20 ==>
 ano_mes=março/2018 18 acc:(0.86489)
 88. eventname=\core\event\course_module_deleted ano_mes=setembro/2017 20
 ==> userid=757 18 acc:(0.86489)
 89. userid=1854 eventname=\core\event\user_graded 121 ==> ano_mes=maio/2018
 105 acc:(0.86469)
 90. userid=0 4 ==> eventname=\core\event\calendar_event_updated 4 acc:
 (0.86404)

 91. userid=700 4 ==> ano_mes=outubro/2017 4 acc:(0.86404)
 92. userid=2 eventname=\mod_quiz\event\course_module_viewed 4 ==>
 ano_mes=novembro/2017 4 acc:(0.86404)
 93. userid=2 eventname=\core\event\course_module_updated 4 ==>
 ano_mes=novembro/2017 4 acc:(0.86404)
 94. userid=662 eventname=\gradereport_grader\event\grade_report_viewed 4 ==>
 ano_mes=fevereiro/2018 4 acc:(0.86404)
 95. userid=672 ano_mes=janeiro/2018 4 ==>
 eventname=\core\event\course_user_report_viewed 4 acc:(0.86404)
 96. userid=693 eventname=\mod_chat\event\sessions_viewed 4 ==>

ano_mes=outubro/2018 4 acc:(0.86404)
97. userid=695 eventname=\mod_resource\event\course_module_viewed 4 ==>
ano_mes=fevereiro/2018 4 acc:(0.86404)
98. userid=698 eventname=\core\event\course_user_report_viewed 4 ==>
ano_mes=maio/2018 4 acc:(0.86404)
99. userid=698 eventname=\mod_chat\event\sessions_viewed 4 ==>
ano_mes=março/2018 4 acc:(0.86404)
100. eventname=\mod_chat\event\sessions_viewed ano_mes=outubro/2018 4 ==>
userid=693 4 acc:(0.86404)

APÊNDICE C – RELATÓRIO DO ALGORITMO *APRIORI* NO PRIMEIRO TESTE DO DATASET B

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5

Instances: 776

Attributes: 4

userid

desempenho

disciplina

faixa_tempo_quiz

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.1 (78 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 2

Best rules found:

1. desempenho=fraco 345 ==> faixa_tempo_quiz=Menos de 15 minutos 242
<conf:(0.7)> lift:(1.11) lev:(0.03) [23] conv:(1.21)

APÊNDICE D – RELATÓRIO DO ALGORITMO *PREDICTE APRIORI* NO PRIMEIRO TESTE DO *DATASET B*

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5
Instances: 776
Attributes: 4
 userid
 desempenho
 disciplina
 faixa_tempo_quiz
=== Associator model (full training set) ===

PredictiveApriori
=====

Best rules found:

1. userid=701 20 ==> faixa_tempo_quiz=Menos de 15 minutos 20 acc:(0.99373)
2. userid=669 desempenho=fraco 10 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.98825)
3. userid=695 desempenho=fraco 8 ==> faixa_tempo_quiz=Menos de 15 minutos 8 acc:(0.98441)
4. userid=699 desempenho=fraco 8 ==> faixa_tempo_quiz=Menos de 15 minutos 8 acc:(0.98441)
5. userid=670 desempenho=fraco 7 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.98138)
6. userid=687 desempenho=fraco 7 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.98138)
7. disciplina=Administração Financeira faixa_tempo_quiz=Menos de 15 minutos 25 ==> desempenho=fraco 24 acc:(0.97994)
8. userid=990 6 ==> faixa_tempo_quiz=Menos de 15 minutos 6 acc:(0.97707)
9. userid=986 desempenho=bom 6 ==> faixa_tempo_quiz=Menos de 15 minutos 6 acc:(0.97707)
10. disciplina=Ética faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==> desempenho=bom 6 acc:(0.97707)

11. userid=684 desempenho=bom 5 ==> faixa_tempo_quiz=Menos de 15 minutos 5 acc:(0.97064)

12. userid=705 desempenho=insuficiente 5 ==> faixa_tempo_quiz=Menos de 15 minutos 5 acc:(0.97064)

13. desempenho=insuficiente disciplina=Empreendedorismo 5 ==> faixa_tempo_quiz=Menos de 15 minutos 5 acc:(0.97064)

14. disciplina=Gestão de Pessoas faixa_tempo_quiz=Entre 15 e 30 minutos 5 ==> desempenho=fraco 5 acc:(0.97064)

15. userid=664 desempenho=insuficiente 4 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.96042)

16. userid=677 desempenho=fraco 4 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.96042)

17. userid=688 desempenho=insuficiente 4 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.96042)

18. userid=696 desempenho=insuficiente 4 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.96042)

19. userid=667 faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 3 acc:(0.94281)

20. userid=680 faixa_tempo_quiz=Entre 3 e 6 horas 3 ==> desempenho=bom 3 acc:(0.94281)

21. userid=695 faixa_tempo_quiz=Entre 3 e 6 horas 3 ==> desempenho=bom 3 acc:(0.94281)

22. userid=988 desempenho=alto 3 ==> faixa_tempo_quiz=Menos de 15 minutos 3 acc:(0.94281)

23. disciplina=Administração Financeira faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 3 acc:(0.94281)

24. userid=686 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

25. userid=702 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

26. userid=709 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

27. userid=669 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

28. userid=669 desempenho=insuficiente 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

29. userid=672 desempenho=insuficiente 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

30. userid=672 faixa_tempo_quiz=Entre 3 e 6 horas 2 ==> desempenho=bom 2 acc:(0.90914)

31. userid=673 faixa_tempo_quiz=Mais que 12 horas 2 ==> desempenho=bom 2 acc:(0.90914)

32. userid=675 desempenho=bom 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

33. userid=675 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2

acc:(0.90914)

34. userid=689 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

35. userid=693 desempenho=insuficiente 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

36. userid=697 desempenho=fraco 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

37. userid=699 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

38. userid=704 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

39. userid=707 desempenho=alto 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

40. userid=707 desempenho=insuficiente 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

41. userid=707 faixa_tempo_quiz=Entre 15 e 30 minutos 2 ==> desempenho=fraco 2 acc:(0.90914)

42. userid=988 desempenho=insuficiente 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.90914)

43. userid=991 faixa_tempo_quiz=Mais que 12 horas 2 ==> desempenho=bom 2 acc:(0.90914)

44. disciplina=Metodologia em Ead faixa_tempo_quiz=Entre 45 minutos e 1 hora 2 ==> desempenho=bom 2 acc:(0.90914)

45. disciplina=Metodologia em Ead faixa_tempo_quiz=Entre 1 e 3 horas 2 ==> desempenho=bom 2 acc:(0.90914)

46. disciplina=Ética faixa_tempo_quiz=Entre 1 e 3 horas 2 ==> desempenho=bom 2 acc:(0.90914)

47. disciplina=Ética faixa_tempo_quiz=Entre 30 e 45 minutos 2 ==> desempenho=bom 2 acc:(0.90914)

48. disciplina=Sociologia Aplic. à Administração faixa_tempo_quiz=Entre 3 e 6 horas 2 ==> desempenho=bom 2 acc:(0.90914)

49. userid=675 11 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.88293)

50. userid=693 desempenho=fraco 11 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.88293)

51. userid=705 desempenho=fraco 11 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.88293)

52. disciplina=Empreendedorismo 34 ==> faixa_tempo_quiz=Menos de 15 minutos 31 acc:(0.86498)

53. disciplina=Administração Financeira 34 ==> desempenho=fraco 31 acc:(0.86498)

54. userid=669 20 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.84751)

55. userid=693 20 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.84751)

56. userid=696 20 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.84751)
 57. userid=705 20 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.84751)
 58. userid=689 desempenho=fraco 9 ==> faixa_tempo_quiz=Menos de 15 minutos 8 acc:(0.84546)
 59. disciplina=Ética faixa_tempo_quiz=Mais que 12 horas 9 ==> desempenho=bom 8 acc:(0.84546)
 60. disciplina=Metodologia em Ead 48 ==> desempenho=bom 41 acc:(0.83025)
 61. desempenho=fraco disciplina=Estatística Básica 24 ==> faixa_tempo_quiz=Menos de 15 minutos 21 acc:(0.82414)
 62. desempenho=fraco disciplina=Comercio Eletrônico e Automação Comercial 17 ==> faixa_tempo_quiz=Menos de 15 minutos 15 acc:(0.82328)
 63. userid=687 desempenho=bom 8 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.82182)
 64. userid=991 desempenho=fraco 8 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.82182)
 65. desempenho=insuficiente disciplina=Estatística Básica 8 ==> faixa_tempo_quiz=Menos de 15 minutos 7 acc:(0.82182)
 66. disciplina=Empreendedorismo 34 ==> desempenho=fraco 29 acc:(0.8197)
 67. disciplina=Gestão de Pessoas faixa_tempo_quiz=Menos de 15 minutos 22 ==> desempenho=fraco 19 acc:(0.81272)
 68. disciplina=Estatística Básica 36 ==> faixa_tempo_quiz=Menos de 15 minutos 30 acc:(0.80634)
 69. desempenho=insuficiente disciplina=Análise de Custo e Formação de Preços 21 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.80634)
 70. userid=664 20 ==> faixa_tempo_quiz=Menos de 15 minutos 17 acc:(0.79944)
 71. userid=677 20 ==> faixa_tempo_quiz=Menos de 15 minutos 17 acc:(0.79944)
 72. userid=687 20 ==> faixa_tempo_quiz=Menos de 15 minutos 17 acc:(0.79944)
 73. userid=699 20 ==> faixa_tempo_quiz=Menos de 15 minutos 17 acc:(0.79944)
 74. userid=988 20 ==> faixa_tempo_quiz=Menos de 15 minutos 17 acc:(0.79944)
 75. disciplina=Gestão de Pessoas 34 ==> desempenho=fraco 28 acc:(0.79669)
 76. userid=673 14 ==> desempenho=bom 12 acc:(0.79405)
 77. userid=670 desempenho=bom 7 ==> faixa_tempo_quiz=Menos de 15 minutos 6 acc:(0.79338)
 78. desempenho=bom disciplina=Matemática Financeira 19 ==> faixa_tempo_quiz=Menos de 15 minutos 16 acc:(0.79195)
 79. disciplina=Ética 43 ==> desempenho=bom 34 acc:(0.7724)
 80. desempenho=fraco disciplina=Gestão da Qualidade 16 ==> faixa_tempo_quiz=Menos de 15 minutos 13 acc:(0.76474)
 81. userid=670 20 ==> faixa_tempo_quiz=Menos de 15 minutos 16 acc:(0.76345)
 82. userid=684 20 ==> faixa_tempo_quiz=Menos de 15 minutos 16 acc:(0.76345)
 83. userid=689 20 ==> faixa_tempo_quiz=Menos de 15 minutos 16 acc:(0.76345)
 84. desempenho=fraco disciplina=Administração Financeira 31 ==>

faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.75295)
 85. disciplina=Comercio Eletrônico e Automação Comercial 34 ==>
 faixa_tempo_quiz=Menos de 15 minutos 26 acc:(0.7461)
 86. disciplina=Empreendedorismo 34 ==> desempenho=fraco
 faixa_tempo_quiz=Menos de 15 minutos 26 acc:(0.7461)
 87. disciplina=Matemática Financeira 38 ==> faixa_tempo_quiz=Menos de 15
 minutos 29 acc:(0.74576)
 88. userid=986 14 ==> faixa_tempo_quiz=Menos de 15 minutos 11 acc:(0.74039)
 89. desempenho=fraco disciplina=Planejamento Estratégico 14 ==>
 faixa_tempo_quiz=Menos de 15 minutos 11 acc:(0.74039)
 90. desempenho=fraco disciplina=Organização 25 ==> faixa_tempo_quiz=Menos de
 15 minutos 19 acc:(0.73813)
 91. userid=665 faixa_tempo_quiz=Menos de 15 minutos 10 ==> desempenho=fraco
 8 acc:(0.73544)
 92. userid=698 desempenho=fraco 10 ==> faixa_tempo_quiz=Menos de 15 minutos
 8 acc:(0.73544)
 93. userid=704 desempenho=fraco 10 ==> faixa_tempo_quiz=Menos de 15 minutos
 8 acc:(0.73544)
 94. desempenho=fraco disciplina=Contabilidade Básica 24 ==>
 faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.72951)
 95. disciplina=Contabilidade Básica faixa_tempo_quiz=Menos de 15 minutos 24 ==>
 desempenho=fraco 18 acc:(0.72951)
 96. disciplina=Análise de Custo e Formação de Preços faixa_tempo_quiz=Menos de
 15 minutos 24 ==> desempenho=insuficiente 18 acc:(0.72951)
 97. userid=704 20 ==> faixa_tempo_quiz=Menos de 15 minutos 15 acc: (0.72554)
 98. userid=707
 20 ==> faixa_tempo_quiz=Menos de 15 minutos 15 acc:(0.72554)
 99. disciplina=Planejamento Estratégico 34 ==> faixa_tempo_quiz=Menos de 15
 minutos 25 acc:(0.72214)
 100. disciplina=Administração Financeira 34 ==> faixa_tempo_quiz=Menos de 15
 minutos 25 acc:(0.72214)

APÊNDICE E – RELATÓRIO DO ALGORITMO *K-MEANS* NO PRIMEIRO TESTE DO DATASET B

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A

"weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5

Instances: 776

Attributes: 4

userid

disciplina

faixa_tempo_quiz

Ignored:

desempenho

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans

=====

Number of iterations: 3

Within cluster sum of squared errors: 1612.0

Initial starting points (random):

Cluster 0: 688,'Cenários Econômicos e Mercado','Menos de 15 minutos'

Cluster 1: 677,'Gestão de Pessoas','Entre 30 e 45 minutos'

Cluster 2: 665,'Comercio Eletrônico e Automação Comercial','Menos de 15 minutos'

Cluster 3: 988,'Gestão da Qualidade','Menos de 15 minutos'

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute	Full Data	0	1	2	3
	(776.0)	(627.0)	(46.0)	(49.0)	(54.0)
userid	664	688	703	665	988
disciplina	Metodologia em EaD	Metodologia em Ead	Gestão de Pessoas	Comércio Eletrônico e Automação Comercial	Gestão da Qualidade
faixa_tempo_quiz	Menos de 15 minutos	Menos de 15 minutos	Entre 30 e 45 minutos	Menos de 15 minutos	Menos de 15 minutos

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

```
0    627 ( 81%)
1     46 (  6%)
2     49 (  6%)
3     54 (  7%)
```

Class attribute: desempenho

Classes to Clusters:

```
0  1  2  3 <-- assigned to cluster
276 20 27 22 | fraco
222 18  4 25 | bom
 59  4  2  4 | alto
 70  4 16  3 | insuficiente
```

Cluster 0 <-- fraco

Cluster 1 <-- alto

Cluster 2 <-- insuficiente

Cluster 3 <-- bom

Incorrectly clustered instances : 455.0 58.634 %

APÊNDICE F – RELATÓRIO DO ALGORITMO *BAYES NET* NO PRIMEIRO TESTE DO DATASET B

=== Run information ===

Scheme: weka.classifiers.bayes.BayesNet -D -Q
weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E
weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5
Instances: 776
Attributes: 4
 userid
 desempenho
 disciplina
 faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Bayes Network Classifier
not using ADTree
#attributes=4 #classindex=1
Network structure (nodes followed by parents)
userid(48): desempenho
desempenho(4):
disciplina(20): desempenho
faixa_tempo_quiz(9): desempenho
LogScore Bayes: -7129.878950307306
LogScore BDeu: -8695.759177804652
LogScore MDL: -8325.922414719076
LogScore ENTROPY: -7331.126612951685
LogScore AIC: -7630.126612951685

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	174	65.9091 %
Incorrectly Classified Instances	90	34.0909 %
Kappa statistic	0.4507	
Mean absolute error	0.2315	
Root mean squared error	0.3436	
Relative absolute error	70.4057 %	
Root relative squared error	85.457 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,752	0,252	0,729	0,752	0,740	0,500	0,805	0,770	fraco
	0,722	0,241	0,607	0,722	0,660	0,464	0,827	0,658	bom
	0,240	0,025	0,500	0,240	0,324	0,302	0,901	0,502	alto
	0,375	0,029	0,563	0,375	0,450	0,417	0,895	0,433	insuficiente
Weighted Avg.	0,659	0,207	0,651	0,659	0,647	0,461	0,830	0,676	

=== Confusion Matrix ===

```
a b c d <-- classified as
94 24 2 5 | a = fraco
19 65 4 2 | b = bom
1 18 6 0 | c = alto
15 0 0 9 | d = insuficiente
```

APÊNDICE G – RELATÓRIO DO ALGORITMO NAIVE BAYES NO PRIMEIRO TESTE DO *DATASET B*

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
 Relation: media_notas_tempoQuiz-
 weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
 weka.filters.unsupervised.attribute.Remove-R2,5
 Instances: 776
 Attributes: 4
 userid
 desempenho
 disciplina
 faixa_tempo_quiz
 Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	172	65.1515 %
Incorrectly Classified Instances	92	34.8485 %
Kappa statistic	0.4393	
Mean absolute error	0.239	
Root mean squared error	0.3429	
Relative absolute error	72.6776 %	
Root relative squared error	85.2839 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,752	0,266	0,718	0,752	0,734	0,485	0,806	0,769	fraco
0,700	0,230	0,612	0,700	0,653	0,457	0,827	0,661	bom
0,320	0,033	0,500	0,320	0,390	0,352	0,901	0,475	alto

	0,292	0,029	0,500	0,292	0,368	0,337	0,903	0,442	insuficiente
Weighted Avg.	0,652	0,210	0,641	0,652	0,641	0,449	0,831	0,674	

=== Confusion Matrix ===

```

a b c d <-- classified as
94 24 2 5 | a = fraco
19 63 6 2 | b = bom
1 16 8 0 | c = alto
17 0 0 7 | d = insuficiente

```

APÊNDICE H – RELATÓRIO DO ALGORITMO *DECISION TABLE* NO PRIMEIRO TESTE DO *DATASET B*

=== Run information ===

Scheme: weka.classifiers.rules.DecisionTable -X 1 -S
"weka.attributeSelection.BestFirst -D 1 -N 5"
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5
Instances: 776
Attributes: 4
 userid
 desempenho
 disciplina
 faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Decision Table:

Number of training instances: 776
Number of Rules : 20
Non matches covered by Majority class.
 Best first.
 Start set: no attributes
 Search direction: forward
 Stale search after 5 node expansions
 Total number of subsets evaluated: 8
 Merit of best subset found: 59.536
Evaluation (for feature selection): CV (leave one out)
Feature set: 3,2

Time taken to build model: 0.07 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	162	61.3636 %
Incorrectly Classified Instances	102	38.6364 %
Kappa statistic	0.4052	
Mean absolute error	0.252	
Root mean squared error	0.3442	
Relative absolute error	76.6165 %	
Root relative squared error	85.6188 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,616	0,230	0,706	0,616	0,658	0,391	0,793	0,751	fraco
	0,678	0,264	0,570	0,678	0,619	0,399	0,808	0,667	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,899	0,453	alto
	0,583	0,063	0,483	0,583	0,528	0,479	0,912	0,440	insuficiente
Weighted Avg.	0,614	0,208	0,623	0,614	0,614	0,404	0,819	0,666	

=== Confusion Matrix ===

```

a b c d <-- classified as
77 32 3 13 | a = fraco
21 61 6 2 | b = bom
1 14 10 0 | c = alto
10 0 0 14 | d = insuficiente

```

APÊNDICE I – RELATÓRIO DO ALGORITMO J48 NO PRIMEIRO TESTE DO DATASET B

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5
Instances: 776
Attributes: 4
 userid
 desempenho
 disciplina
 faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

disciplina = Metodologia em Ead: bom (48.0/7.0)
disciplina = Redação de Textos Oficiais: bom (47.0/18.0)
disciplina = Cenários Econômicos e Mercado: alto (44.0/18.0)
disciplina = Ética: bom (43.0/9.0)
disciplina = Sociologia Aplic. à Administração: bom (43.0/21.0)
disciplina = Direito: fraco (41.0/22.0)
disciplina = Gestão da Qualidade: bom (41.0/18.0)
disciplina = Marketing: fraco (41.0/17.0)
disciplina = Matemática Financeira
| faixa_tempo_quiz = Entre 3 e 6 horas: bom (1.0)
| faixa_tempo_quiz = Entre 15 e 30 minutos: fraco (6.0/2.0)
| faixa_tempo_quiz = Mais que 12 horas: fraco (1.0)
| faixa_tempo_quiz = Entre 6 e 9 horas: fraco (1.0)
| faixa_tempo_quiz = Menos de 15 minutos: bom (29.0/13.0)
| faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (0.0)
| faixa_tempo_quiz = Entre 1 e 3 horas: fraco (0.0)
| faixa_tempo_quiz = Entre 30 e 45 minutos: fraco (0.0)
| faixa_tempo_quiz = Entre 9 e 12 horas: fraco (0.0)
disciplina = Organização: fraco (39.0/14.0)
disciplina = Cooperativismo e Associativismo: fraco (39.0/15.0)

disciplina = Administração de Rec. Materiais e Patrimoniais e Logística: fraco (36.0/20.0)

disciplina = Contabilidade Básica: fraco (36.0/12.0)

disciplina = Estatística Básica: fraco (36.0/12.0)

disciplina = Comercio Eletrônico e Automação Comercial

| faixa_tempo_quiz = Entre 3 e 6 horas: fraco (0.0)

| faixa_tempo_quiz = Entre 15 e 30 minutos: insuficiente (4.0/1.0)

| faixa_tempo_quiz = Mais que 12 horas: insuficiente (3.0)

| faixa_tempo_quiz = Entre 6 e 9 horas: fraco (0.0)

| faixa_tempo_quiz = Menos de 15 minutos: fraco (26.0/11.0)

| faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (0.0)

| faixa_tempo_quiz = Entre 1 e 3 horas: fraco (1.0)

| faixa_tempo_quiz = Entre 30 e 45 minutos: fraco (0.0)

| faixa_tempo_quiz = Entre 9 e 12 horas: fraco (0.0)

disciplina = Planejamento Estratégico: insuficiente (34.0/15.0)

disciplina = Empreendedorismo: fraco (34.0/5.0)

disciplina = Gestão de Pessoas: fraco (34.0/6.0)

disciplina = Administração Financeira: fraco (34.0/3.0)

disciplina = Analise de Custo e Formação de Preços

| faixa_tempo_quiz = Entre 3 e 6 horas: insuficiente (0.0)

| faixa_tempo_quiz = Entre 15 e 30 minutos: fraco (4.0/1.0)

| faixa_tempo_quiz = Mais que 12 horas: fraco (4.0/1.0)

| faixa_tempo_quiz = Entre 6 e 9 horas: insuficiente (0.0)

| faixa_tempo_quiz = Menos de 15 minutos: insuficiente (24.0/6.0)

| faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (1.0)

| faixa_tempo_quiz = Entre 1 e 3 horas: insuficiente (0.0)

| faixa_tempo_quiz = Entre 30 e 45 minutos: insuficiente (1.0)

| faixa_tempo_quiz = Entre 9 e 12 horas: insuficiente (0.0)

Number of Leaves : 44

Size of the tree : 48

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	164	62.1212 %
Incorrectly Classified Instances	100	37.8788 %
Kappa statistic	0.4021	
Mean absolute error	0.2315	
Root mean squared error	0.3478	
Relative absolute error	70.388 %	
Root relative squared error	86.5223 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,672	0,273	0,689	0,672	0,680	0,399	0,784	0,725	fraco
	0,656	0,259	0,567	0,656	0,608	0,385	0,793	0,652	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,910	0,453	alto
	0,458	0,033	0,579	0,458	0,512	0,473	0,915	0,474	insuficiente
Weighted Avg.	0,621	0,224	0,622	0,621	0,619	0,402	0,811	0,651	

=== Confusion Matrix ===

```

a b c d <-- classified as
84 31 3 7 | a = fraco
24 59 6 1 | b = bom
1 14 10 0 | c = alto
13 0 0 11 | d = insuficiente

```

APÊNDICE J – RELATÓRIO DO ALGORITMO *RANDOM FOREST* NO PRIMEIRO TESTE DO *DATASET B*

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5

Instances: 776

Attributes: 4

userid

desempenho

disciplina

faixa_tempo_quiz

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.3 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.04 seconds

=== Summary ===

Correctly Classified Instances	163	61.7424 %
Incorrectly Classified Instances	101	38.2576 %
Kappa statistic	0.3893	
Mean absolute error	0.2549	
Root mean squared error	0.3444	
Relative absolute error	77.5195 %	
Root relative squared error	85.6606 %	

Total Number of Instances 264

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,672	0,266	0,694	0,672	0,683	0,407	0,793	0,768	fraco
	0,700	0,287	0,558	0,700	0,621	0,395	0,804	0,663	bom
	0,360	0,033	0,529	0,360	0,429	0,390	0,926	0,535	alto
	0,292	0,025	0,538	0,292	0,378	0,354	0,930	0,604	insuficiente
Weighted Avg.	0,617	0,229	0,618	0,617	0,610	0,396	0,822	0,695	

=== Confusion Matrix ===

```
a b c d <-- classified as
84 34 3 4 | a = fraco
20 63 5 2 | b = bom
0 16 9 0 | c = alto
17 0 0 7 | d = insuficiente
```


APÊNDICE K - RELATÓRIO DO ALGORITMO *APRIORI* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.1 -S
-1.0 -c -1

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5-

weka.filters.unsupervised.attribute.Remove-R1

Instances: 776

Attributes: 3

desempenho

disciplina

faixa_tempo_quiz

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.1 (78 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 2

Best rules found:

1. desempenho=fraco 345 ==> faixa_tempo_quiz=Menos de 15 minutos 242
<conf:(0.7)> lift:(1.11) lev:(0.03) [23] conv:(1.21)

APÊNDICE L - RELATÓRIO DO ALGORITMO *PREDICTIVE APRIORI* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5-
weka.filters.unsupervised.attribute.Remove-R1
Instances: 776
Attributes: 3
desempenho
disciplina
faixa_tempo_quiz

=== Associator model (full training set) ===

PredictiveApriori

=====

Best rules found:

1. disciplina=Administração Financeira faixa_tempo_quiz=Menos de 15 minutos 25
==> desempenho=fraco 24 acc:(0.9561)
2. disciplina=Ética faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==>
desempenho=bom 6 acc:(0.94358)
3. desempenho=insuficiente disciplina=Empreendedorismo 5 ==>
faixa_tempo_quiz=Menos de 15 minutos 5 acc:(0.92733)
4. disciplina=Gestão de Pessoas faixa_tempo_quiz=Entre 15 e 30 minutos 5 ==>
desempenho=fraco 5 acc:(0.92733)
5. disciplina=Empreendedorismo 34 ==> faixa_tempo_quiz=Menos de 15 minutos
31 acc:(0.87726)
6. disciplina=Administração Financeira 34 ==> desempenho=fraco 31 acc:
(0.87726)
7. desempenho=alto disciplina=Direito 3 ==> faixa_tempo_quiz=Menos de 15
minutos 3 acc:(0.86733)
8. disciplina=Redação de Textos Oficiais faixa_tempo_quiz=Entre 45 minutos e 1
hora 3 ==> desempenho=bom 3 acc:(0.86733)
9. disciplina=Marketing faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==>
desempenho=fraco 3 acc:(0.86733)

10. disciplina=Organização faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 3 acc:(0.86733)

11. disciplina=Comercio Eletrônico e Automação Comercial faixa_tempo_quiz=Mais que 12 horas 3 ==> desempenho=insuficiente 3 acc:(0.86733)

12. disciplina=Metodologia em Ead faixa_tempo_quiz=Menos de 15 minutos 25 ==> desempenho=bom 22 acc:(0.83823)

13. disciplina=Metodologia em Ead 48 ==> desempenho=bom 41 acc:(0.83779)

14. desempenho=fraco disciplina=Estatística Básica 24 ==> faixa_tempo_quiz=Menos de 15 minutos 21 acc:(0.83177)

15. disciplina=Empreendedorismo 34 ==> desempenho=fraco 29 acc:(0.82531)

16. desempenho=fraco disciplina=Comercio Eletrônico e Automação Comercial 17 ==> faixa_tempo_quiz=Menos de 15 minutos 15 acc:(0.82277)

17. disciplina=Gestão de Pessoas faixa_tempo_quiz=Menos de 15 minutos 22 ==> desempenho=fraco 19 acc:(0.81695)

18. desempenho=alto disciplina=Administração de Rec. Materiais e Patrimoniais e Logística 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.80993)

19. desempenho=insuficiente disciplina=Administração de Rec. Materiais e Patrimoniais e Logística 2 ==> faixa_tempo_quiz=Menos de 15 minutos 2 acc:(0.80993)

20. disciplina=Ética faixa_tempo_quiz=Entre 1 e 3 horas 2 ==> desempenho=bom 2 acc:(0.80993)

21. disciplina=Ética faixa_tempo_quiz=Entre 30 e 45 minutos 2 ==> desempenho=bom 2 acc:(0.80993)

22. disciplina=Sociologia Aplic. à Administração faixa_tempo_quiz=Entre 3 e 6 horas 2 ==> desempenho=bom 2 acc:(0.80993)

23. disciplina=Direito faixa_tempo_quiz=Mais que 12 horas 2 ==> desempenho=fraco 2 acc:(0.80993)

24. disciplina=Gestão da Qualidade faixa_tempo_quiz=Entre 3 e 6 horas 2 ==> desempenho=bom 2 acc:(0.80993)

25. disciplina=Organização faixa_tempo_quiz=Entre 3 e 6 horas 2 ==> desempenho=bom 2 acc:(0.80993)

26. disciplina=Contabilidade Básica faixa_tempo_quiz=Entre 30 e 45 minutos 2 ==> desempenho=fraco 2 acc:(0.80993)

27. disciplina=Estatística Básica faixa_tempo_quiz=Entre 15 e 30 minutos 2 ==> desempenho=fraco 2 acc:(0.80993)

28. desempenho=insuficiente disciplina=Análise de Custo e Formação de Preços 21 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.80849)

29. disciplina=Estatística Básica 36 ==> faixa_tempo_quiz=Menos de 15 minutos 30 acc:(0.80728)

30. disciplina=Ética faixa_tempo_quiz=Mais que 12 horas 9 ==> desempenho=bom 8 acc:(0.80323)

31. disciplina=Gestão de Pessoas 34 ==> desempenho=fraco 28 acc:(0.79461)

32. desempenho=bom disciplina=Matemática Financeira 19 ==> faixa_tempo_quiz=Menos de 15 minutos 16 acc:(0.78925)

33. disciplina=Ética 43 ==> desempenho=bom 34 acc:(0.76376)

34. desempenho=fraco disciplina=Gestão da Qualidade 16 ==> faixa_tempo_quiz=Menos de 15 minutos 13 acc:(0.75395)

35. desempenho=fraco disciplina=Administração Financeira 31 ==> faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.73941)

36. disciplina=Matemática Financeira 38 ==> faixa_tempo_quiz=Menos de 15 minutos 29 acc:(0.73236)

37. disciplina=Comercio Eletrônico e Automação Comercial 34 ==> faixa_tempo_quiz=Menos de 15 minutos 26 acc:(0.73196)

38. disciplina=Empreendedorismo 34 ==> desempenho=fraco faixa_tempo_quiz=Menos de 15 minutos 26 acc:(0.73196)

39. desempenho=fraco disciplina=Planejamento Estratégico 14 ==> faixa_tempo_quiz=Menos de 15 minutos 11 acc:(0.72509)

40. desempenho=fraco disciplina=Organização 25 ==> faixa_tempo_quiz=Menos de 15 minutos 19 acc:(0.72155)

41. disciplina=Gestão da Qualidade faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==> desempenho=bom 5 acc:(0.71995)

42. desempenho=insuficiente 93 ==> faixa_tempo_quiz=Menos de 15 minutos 68 acc:(0.71535)

43. desempenho=fraco disciplina=Contabilidade Básica 24 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.71196)

44. disciplina=Contabilidade Básica faixa_tempo_quiz=Menos de 15 minutos 24 ==> desempenho=fraco 18 acc:(0.71196)

45. disciplina=Análise de Custo e Formação de Preços faixa_tempo_quiz=Menos de 15 minutos 24 ==> desempenho=insuficiente 18 acc:(0.71196)

46. disciplina=Planejamento Estratégico 34 ==> faixa_tempo_quiz=Menos de 15 minutos 25 acc:(0.7057)

47. disciplina=Administração Financeira 34 ==> faixa_tempo_quiz=Menos de 15 minutos 25 acc:(0.7057)

48. disciplina=Organização 39 ==> faixa_tempo_quiz=Menos de 15 minutos 28 acc:(0.69363)

49. faixa_tempo_quiz=Entre 3 e 6 horas 28 ==> desempenho=bom 20 acc:(0.68586)

50. disciplina=Administração Financeira 34 ==> desempenho=fraco faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.68249)

51. disciplina=Análise de Custo e Formação de Preços 34 ==> faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.68249)

52. desempenho=fraco disciplina=Cenários Econômicos e Mercado 5 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.68039)

53. desempenho=fraco disciplina=Ética 5 ==> faixa_tempo_quiz=Menos de 15

minutos 4 acc:(0.68039)

54. desempenho=fraco disciplina=Sociologia Aplic. à Administração 5 ==> faixa_tempo_quiz=Menos de 15 minutos 4 acc:(0.68039)

55. disciplina=Redação de Textos Oficiais faixa_tempo_quiz=Entre 15 e 30 minutos 5 ==> desempenho=bom 4 acc:(0.68039)

56. disciplina=Estatística Básica faixa_tempo_quiz=Menos de 15 minutos 30 ==> desempenho=fraco 21 acc:(0.67608)

57. disciplina=Administração de Rec. Materiais e Patrimoniais e Logística 36 ==> faixa_tempo_quiz=Menos de 15 minutos 25 acc:(0.67471)

58. disciplina=Redação de Textos Oficiais faixa_tempo_quiz=Menos de 15 minutos 11 ==> desempenho=alto 8 acc:(0.66833)

59. disciplina=Redação de Textos Oficiais faixa_tempo_quiz=Mais que 12 horas 14 ==> desempenho=bom 10 acc:(0.66798)

60. desempenho=alto disciplina=Cenários Econômicos e Mercado 26 ==> faixa_tempo_quiz=Menos de 15 minutos 18 acc:(0.66763)

61. desempenho=fraco disciplina=Gestão de Pessoas 28 ==> faixa_tempo_quiz=Menos de 15 minutos 19 acc:(0.6586)

62. disciplina=Organização faixa_tempo_quiz=Menos de 15 minutos 28 ==> desempenho=fraco 19 acc:(0.6586)

63. disciplina=Contabilidade Básica 36 ==> desempenho=fraco 24 acc:(0.65388)

64. disciplina=Contabilidade Básica 36 ==> faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.65388)

65. disciplina=Estatística Básica 36 ==> desempenho=fraco 24 acc:(0.65388)

66. disciplina=Cenários Econômicos e Mercado 44 ==> faixa_tempo_quiz=Menos de 15 minutos 29 acc:(0.65084)

67. disciplina=Gestão da Qualidade 41 ==> faixa_tempo_quiz=Menos de 15 minutos 27 acc:(0.64951)

68. disciplina=Gestão de Pessoas 34 ==> faixa_tempo_quiz=Menos de 15 minutos 22 acc:(0.63756)

69. disciplina=Cooperativismo e Associativismo faixa_tempo_quiz=Entre 15 e 30 minutos 7 ==> desempenho=fraco 5 acc:(0.63639)

70. disciplina=Organização 39 ==> desempenho=fraco 25 acc:(0.63486)

71. desempenho=alto disciplina=Sociologia Aplic. à Administração 15 ==> faixa_tempo_quiz=Menos de 15 minutos 10 acc:(0.63302)

72. disciplina=Sociologia Aplic. à Administração faixa_tempo_quiz=Entre 15 e 30 minutos 4 ==> desempenho=bom 3 acc:(0.63084)

73. disciplina=Comercio Eletrônico e Automação Comercial faixa_tempo_quiz=Entre 15 e 30 minutos 4 ==> desempenho=insuficiente 3 acc:(0.63084)

74. disciplina=Análise de Custo e Formação de Preços faixa_tempo_quiz=Entre 15 e 30 minutos 4 ==> desempenho=fraco 3 acc:(0.63084)

75. disciplina=Análise de Custo e Formação de Preços faixa_tempo_quiz=Mais que 12 horas 4 ==> desempenho=fraco 3 acc:(0.63084)

76. desempenho=alto 69 ==> faixa_tempo_quiz=Menos de 15 minutos 43 acc:(0.62599)

77. disciplina=Redação de Textos Oficiais 47 ==> desempenho=bom 29 acc:(0.61664)

78. disciplina=Cooperativismo e Associativismo 39 ==> desempenho=fraco 24 acc:(0.61255)

79. disciplina=Análise de Custo e Formação de Preços 34 ==> desempenho=insuficiente 21 acc:(0.61239)

80. disciplina=Cenários Econômicos e Mercado faixa_tempo_quiz=Menos de 15 minutos 29 ==> desempenho=alto 18 acc:(0.61215)

81. disciplina=Sociologia Aplic. à Administração 43 ==> faixa_tempo_quiz=Menos de 15 minutos 26 acc:(0.60378)

82. disciplina=Marketing faixa_tempo_quiz=Mais que 12 horas 11 ==> desempenho=fraco 7 acc:(0.59834)

83. disciplina=Matemática Financeira faixa_tempo_quiz=Entre 15 e 30 minutos 6 ==> desempenho=fraco 4 acc:(0.59618)

84. desempenho=bom disciplina=Direito 18 ==> faixa_tempo_quiz=Menos de 15 minutos 11 acc:(0.59249)

85. disciplina=Cenários Econômicos e Mercado 44 ==> desempenho=alto 26 acc:(0.59026)

86. faixa_tempo_quiz=Entre 45 minutos e 1 hora 13 ==> desempenho=bom 8 acc:(0.5869)

87. disciplina=Direito 41 ==> faixa_tempo_quiz=Menos de 15 minutos 24 acc:(0.58368)

88. disciplina=Marketing 41 ==> desempenho=fraco 24 acc:(0.58368)

89. disciplina=Estatística Básica 36 ==> desempenho=fraco faixa_tempo_quiz=Menos de 15 minutos 21 acc:(0.57998)

90. desempenho=bom disciplina=Marketing 15 ==> faixa_tempo_quiz=Menos de 15 minutos 9 acc:(0.57802)

91. disciplina=Comércio Eletrônico e Automação Comercial faixa_tempo_quiz=Menos de 15 minutos 26 ==> desempenho=fraco 15 acc:(0.56873)

92. desempenho=insuficiente faixa_tempo_quiz=Entre 3 e 6 horas 3 ==> disciplina=Contabilidade Básica 2 acc:(0.56533)

93. disciplina=Cenários Econômicos e Mercado faixa_tempo_quiz=Entre 30 e 45 minutos 3 ==> desempenho=bom 2 acc:(0.56533)

94. disciplina=Sociologia Aplic. à Administração faixa_tempo_quiz=Entre 1 e 3 horas 3 ==> desempenho=bom 2 acc:(0.56533)

95. disciplina=Direito faixa_tempo_quiz=Entre 1 e 3 horas 3 ==> desempenho=fraco 2 acc:(0.56533)

96. disciplina=Direito faixa_tempo_quiz=Entre 30 e 45 minutos 3 ==> desempenho=bom 2 acc:(0.56533)

97. disciplina=Gestão da Qualidade faixa_tempo_quiz=Mais que 12 horas 3 ==>
desempenho=bom 2 acc:(0.56533)

98. disciplina=Organização faixa_tempo_quiz=Mais que 12 horas 3 ==>
desempenho=bom 2 acc:(0.56533)

99. disciplina=Administração de Rec. Materiais e Patrimoniais e Logística
faixa_tempo_quiz=Entre 15 e 30 minutos 3 ==> desempenho=fraco 2 acc:
(0.56533)

APÊNDICE M - RELATÓRIO DO ALGORITMO *K-MEANS* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A

"weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5-

weka.filters.unsupervised.attribute.Remove-R1

Instances: 776

Attributes: 3

disciplina

faixa_tempo_quiz

Ignored:

desempenho

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans

=====

Number of iterations: 3

Within cluster sum of squared errors: 907.0

Initial starting points (random):

Cluster 0: 'Cenários Econômicos e Mercado','Menos de 15 minutos'

Cluster 1: 'Gestão de Pessoas','Entre 30 e 45 minutos'

Cluster 2: 'Comercio Eletrônico e Automação Comercial','Menos de 15 minutos'

Cluster 3: 'Gestão da Qualidade','Menos de 15 minutos'

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute	Full Data	0	1	2	3
	(776.0)	(672.0)	(31.0)	(34.0)	(39.0)
disciplina	Metodologia em EaD	Metodologia em Ead	Gestão de Pessoas	Comércio Eletrônico e Automação Comercial	Gestão da Qualidade
faixa_tempo_quiz	Menos de 15 minutos	Menos de 15 minutos	Entre 30 e 45 minutos	Menos de 15 minutos	Menos de 15 minutos

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

```
0    672 ( 87%)
1     31 (  4%)
2     34 (  4%)
3     39 (  5%)
```

Class attribute: desempenho

Classes to Clusters:

```
0  1  2  3 <-- assigned to cluster
298 15 17 15 | fraco
236 10  1 22 | bom
66  2  0  1 | alto
72  4 16  1 | insuficiente
```

Cluster 0 <-- fraco

Cluster 1 <-- alto

Cluster 2 <-- insuficiente

Cluster 3 <-- bom

Incorrectly clustered instances : 438.0 56.4433 %

APÊNDICE N - RELATÓRIO DO ALGORITMO *BAYES NET* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.classifiers.bayes.BayesNet -D -Q
weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E
weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5-
weka.filters.unsupervised.attribute.Remove-R1
Instances: 776
Attributes: 3
desempenho
disciplina
faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Bayes Network Classifier
not using ADTree
#attributes=3 #classindex=0
Network structure (nodes followed by parents)
desempenho(4):
disciplina(20): desempenho
faixa_tempo_quiz(9): desempenho
LogScore Bayes: -4073.770497011008
LogScore BDeu: -4539.404012761952
LogScore MDL: -4463.378707182072
LogScore ENTROPY: -4094.073242311902
LogScore AIC: -4205.0732423119025

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	165	62.5 %
Incorrectly Classified Instances	99	37.5 %
Kappa statistic	0.4241	
Mean absolute error	0.235	
Root mean squared error	0.3386	
Relative absolute error	71.447 %	
Root relative squared error	84.2099 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,600	0,194	0,735	0,600	0,661	0,416	0,811	0,780	fraco
	0,756	0,276	0,586	0,756	0,660	0,458	0,826	0,685	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,897	0,467	alto
	0,500	0,063	0,444	0,500	0,471	0,415	0,915	0,499	insuficiente
Weighted Avg.	0,625	0,195	0,638	0,625	0,624	0,430	0,834	0,692	

=== Confusion Matrix ===

a b c d <-- classified as
75 34 3 13 | a = fraco
14 68 6 2 | b = bom
1 14 10 0 | c = alto

12 0 0 12 | d = insuficiente

APÊNDICE O - RELATÓRIO DO ALGORITMO *NAIVE BAYES* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5-

weka.filters.unsupervised.attribute.Remove-R1

Instances: 776

Attributes: 3

desempenho

disciplina

faixa_tempo_quiz

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	168	63.6364 %
Incorrectly Classified Instances	96	36.3636 %
Kappa statistic	0.4288	

Mean absolute error	0.242
Root mean squared error	0.3394
Relative absolute error	73.5763 %
Root relative squared error	84.428 %
Total Number of Instances	264

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,648	0,223	0,723	0,648	0,684	0,429	0,806	0,778	fraco
	0,756	0,276	0,586	0,756	0,660	0,458	0,827	0,685	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,891	0,434	alto
	0,375	0,033	0,529	0,375	0,439	0,400	0,915	0,499	insuficiente
Weighted Avg.	0,636	0,206	0,640	0,636	0,632	0,435	0,831	0,688	

=== Confusion Matrix ===

```

a b c d <-- classified as
81 34 3 7 | a = fraco
15 68 6 1 | b = bom
1 14 10 0 | c = alto
15 0 0 9 | d = insuficiente

```

APÊNDICE P - RELATÓRIO DO ALGORITMO *DECISION TABLE* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.classifiers.rules.DecisionTable -X 1 -S
"weka.attributeSelection.BestFirst -D 1 -N 5"
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5-
weka.filters.unsupervised.attribute.Remove-R1
Instances: 776
Attributes: 3
desempenho
disciplina
faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Decision Table:

Number of training instances: 776
Number of Rules : 20
Non matches covered by Majority class.
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 3
Merit of best subset found: 59.536
Evaluation (for feature selection): CV (leave one out)
Feature set: 2,1

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	162	61.3636 %
Incorrectly Classified Instances	102	38.6364 %
Kappa statistic	0.4052	
Mean absolute error	0.252	
Root mean squared error	0.3442	
Relative absolute error	76.6165 %	
Root relative squared error	85.6188 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,616	0,230	0,706	0,616	0,658	0,391	0,793	0,751	fraco
	0,678	0,264	0,570	0,678	0,619	0,399	0,808	0,667	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,899	0,453	alto
	0,583	0,063	0,483	0,583	0,528	0,479	0,912	0,440	insuficiente
Weighted Avg.	0,614	0,208	0,623	0,614	0,614	0,404	0,819	0,666	

=== Confusion Matrix ===

```

a b c d <-- classified as
77 32 3 13 | a = fraco
21 61 6 2 | b = bom
1 14 10 0 | c = alto
10 0 0 14 | d = insuficiente

```

APÊNDICE Q - RELATÓRIO DO ALGORITMO J48 NO SEGUNDO TESTE DO DATASET B SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: media_notas_tempoQuiz-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R2,5-
weka.filters.unsupervised.attribute.Remove-R1
Instances: 776
Attributes: 3
desempenho
disciplina
faixa_tempo_quiz
Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

disciplina = Metodologia em Ead: bom (48.0/7.0)
disciplina = Redação de Textos Oficiais: bom (47.0/18.0)
disciplina = Cenários Econômicos e Mercado: alto (44.0/18.0)
disciplina = Ética: bom (43.0/9.0)
disciplina = Sociologia Aplic. à Administração: bom (43.0/21.0)
disciplina = Direito: fraco (41.0/22.0)
disciplina = Gestão da Qualidade: bom (41.0/18.0)
disciplina = Marketing: fraco (41.0/17.0)
disciplina = Matemática Financeira
| faixa_tempo_quiz = Entre 3 e 6 horas: bom (1.0)
| faixa_tempo_quiz = Entre 15 e 30 minutos: fraco (6.0/2.0)
| faixa_tempo_quiz = Mais que 12 horas: fraco (1.0)
| faixa_tempo_quiz = Entre 6 e 9 horas: fraco (1.0)
| faixa_tempo_quiz = Menos de 15 minutos: bom (29.0/13.0)
| faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (0.0)
| faixa_tempo_quiz = Entre 1 e 3 horas: fraco (0.0)

| faixa_tempo_quiz = Entre 30 e 45 minutos: fraco (0.0)
 | faixa_tempo_quiz = Entre 9 e 12 horas: fraco (0.0)
 disciplina = Organização: fraco (39.0/14.0)
 disciplina = Cooperativismo e Associativismo: fraco (39.0/15.0)
 disciplina = Administração de Rec. Materiais e Patrimoniais e Logística: fraco (36.0/20.0)
 disciplina = Contabilidade Básica: fraco (36.0/12.0)
 disciplina = Estatística Básica: fraco (36.0/12.0)
 disciplina = Comercio Eletrônico e Automação Comercial
 | faixa_tempo_quiz = Entre 3 e 6 horas: fraco (0.0)
 | faixa_tempo_quiz = Entre 15 e 30 minutos: insuficiente (4.0/1.0)
 | faixa_tempo_quiz = Mais que 12 horas: insuficiente (3.0)
 | faixa_tempo_quiz = Entre 6 e 9 horas: fraco (0.0)
 | faixa_tempo_quiz = Menos de 15 minutos: fraco (26.0/11.0)
 | faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (0.0)
 | faixa_tempo_quiz = Entre 1 e 3 horas: fraco (1.0)
 | faixa_tempo_quiz = Entre 30 e 45 minutos: fraco (0.0)
 | faixa_tempo_quiz = Entre 9 e 12 horas: fraco (0.0)
 disciplina = Planejamento Estratégico: insuficiente (34.0/15.0)
 disciplina = Empreendedorismo: fraco (34.0/5.0)
 disciplina = Gestão de Pessoas: fraco (34.0/6.0)
 disciplina = Administração Financeira: fraco (34.0/3.0)
 disciplina = Analise de Custo e Formação de Preços
 | faixa_tempo_quiz = Entre 3 e 6 horas: insuficiente (0.0)
 | faixa_tempo_quiz = Entre 15 e 30 minutos: fraco (4.0/1.0)
 | faixa_tempo_quiz = Mais que 12 horas: fraco (4.0/1.0)
 | faixa_tempo_quiz = Entre 6 e 9 horas: insuficiente (0.0)
 | faixa_tempo_quiz = Menos de 15 minutos: insuficiente (24.0/6.0)
 | faixa_tempo_quiz = Entre 45 minutos e 1 hora: fraco (1.0)
 | faixa_tempo_quiz = Entre 1 e 3 horas: insuficiente (0.0)
 | faixa_tempo_quiz = Entre 30 e 45 minutos: insuficiente (1.0)
 | faixa_tempo_quiz = Entre 9 e 12 horas: insuficiente (0.0)

Number of Leaves : 44

Size of the tree : 48

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	164	62.1212 %
Incorrectly Classified Instances	100	37.8788 %
Kappa statistic	0.4021	
Mean absolute error	0.2315	
Root mean squared error	0.3478	
Relative absolute error	70.388 %	
Root relative squared error	86.5223 %	
Total Number of Instances	264	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,672	0,273	0,689	0,672	0,680	0,399	0,784	0,725	fraco
	0,656	0,259	0,567	0,656	0,608	0,385	0,793	0,652	bom
	0,400	0,038	0,526	0,400	0,455	0,411	0,910	0,453	alto
	0,458	0,033	0,579	0,458	0,512	0,473	0,915	0,474	insuficiente
Weighted Avg.	0,621	0,224	0,622	0,621	0,619	0,402	0,811	0,651	

=== Confusion Matrix ===

```
a b c d <-- classified as
84 31 3 7 | a = fraco
24 59 6 1 | b = bom
1 14 10 0 | c = alto
13 0 0 11 | d = insuficiente
```

APÊNDICE R - RELATÓRIO DO ALGORITMO *RANDOM FOREST* NO SEGUNDO TESTE DO *DATASET B* SEM O IDENTIFICADOR DO USUÁRIO

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: media_notas_tempoQuiz-

weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-

weka.filters.unsupervised.attribute.Remove-R2,5-

weka.filters.unsupervised.attribute.Remove-R1

Instances: 776

Attributes: 3

desempenho

disciplina

faixa_tempo_quiz

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.13 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	157	59.4697 %
Incorrectly Classified Instances	107	40.5303 %
Kappa statistic	0.3714	
Mean absolute error	0.2306	
Root mean squared error	0.3566	
Relative absolute error	70.1163 %	
Root relative squared error	88.7064 %	

Total Number of Instances 264

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,576	0,252	0,673	0,576	0,621	0,330	0,778	0,722	fraco
	0,678	0,310	0,530	0,678	0,595	0,351	0,781	0,622	bom
	0,520	0,046	0,542	0,520	0,531	0,483	0,900	0,444	alto
	0,458	0,029	0,611	0,458	0,524	0,489	0,905	0,543	insuficiente
Weighted Avg.	0,595	0,232	0,606	0,595	0,595	0,366	0,802	0,645	

=== Confusion Matrix ===

```
a b c d <-- classified as
72 43 4 6 | a = fraco
21 61 7 1 | b = bom
1 11 13 0 | c = alto
13 0 0 11 | d = insuficiente
```