

**DESENVOLVIMENTO DE UM SISTEMA DE APOIO À DECISÃO (SAD).
CONFRONTANDO A AVALIAÇÃO INSTITUCIONAL COM AS NOTAS
CURRICULARES E AS NOTAS DO ENADE – 2ª ETAPA.**

Mokwa, Nereu*

Riveros, Lilian**

Fernández, Carlos***

Resumo

Esta pesquisa busca apresentar os componentes utilizados na criação de um sistema de apoio à decisão, baseado na descoberta de conhecimento em bases de dados de uma instituição de ensino superior, que através do confronto das informações obtidas pela avaliação institucional, das notas curriculares dos alunos e das notas do exame nacional de desempenho do estudante, registra os conhecimentos encontrados e as ações possíveis para a melhoria na qualidade do ensino e faz o posterior acompanhamento dos resultados. Para tanto, demonstra-se como escolheu-se as ferramentas e como foram empregadas para chegar até o funcionamento do sistema. Abordou-se assim, os principais métodos tecnológicos utilizados para a criação do sistema de apoio à decisão.

Palavras-chave: Sistema de Apoio à Decisão. Avaliação Institucional. Enade. Descoberta de Conhecimento em Bases de Dados.

1 INTRODUÇÃO

O presente estudo denota a criação de um Sistema de Apoio à Decisão (SAD), que permita que uma Instituição de Ensino Superior (IES) possa planejar e acompanhar as melhorias na qualidade do ensino e com baixo custo. Através do uso de programas analíticos de informática, será extraído novos conhecimentos de suas bases de dados já existentes,

convertendo os custos com as avaliações, em investimentos e potencializando os resultados, pois esses programas permitem visualizar o conhecimento onde um homem demoraria anos para analisar esses dados.

No artigo anterior, levantou-se os objetivos, as justificativas, as hipóteses e os meios para se desenvolver um SAD que auxiliasse uma IES a descobrir conhecimento novo, possivelmente oculto em suas diversas e extensas bases de dados, com intuito de serem úteis na busca por melhorias para a qualidade de ensino.

Nossa intenção também, é oferecer recursos para que esses novos conhecimentos possam ser armazenados, utilizados e seus resultados avaliados e revistos sempre que necessário, transformando assim essa prática num círculo virtuoso, como ferramenta complementar as demais já utilizadas pelas Comissões Permanentes de Avaliação das IES's.

Enfatizou-se no artigo anterior, que a transitoriedade da sociedade é que suscita, que um Sistema de Apoio nessa área da melhoria na qualidade do ensino, seja interativo e iterativo simultaneamente.

Então, tendo como base essas qualidades, descrevere-se na primeira parte como será utilizado o modelo de prototipação e na segunda, porque escolheu-se a UML como ferramenta de escrita inicial para desenvolver o SAD.

Na terceira parte, sequencial a de modelagem, incorpora-se o levantamento dos requisitos não funcionais, dando forma e podendo então, visualizar de uma maneira abrangente, todo o esqueleto do sistema.

Prossegue-se com a incorporação, analisando as maneiras como os processos fariam a extração dos dados nas bases fornecidas pela IES e as demais obtidas externamente. Nessa quarta parte, necessitou-se também da UML para modelar alguns processos executados de forma quase que artesanal.

Em vista das inúmeras diferenças e ajustes a serem feitos nas bases de dados adquiridas, a fase de preparação descrita na quinta parte, se aproxima muito e justifica a utilização do termo Mineração de Dados.

Por fim, com os resultados obtidos, agora é possível transformar o conhecimento em ações e executá-las, tornando a medir a qualidade resultante e realimentando todo o sistema novamente.

2 DESENVOLVIMENTO

2.1 ENGENHARIA DE SOFTWARE

Para entender a escolha de iniciar este trabalho utilizando a Engenharia de Software, lembra-se que a engenharia de software é uma disciplina que se ocupa de todos os aspectos da produção de software, desde os estágios iniciais até a manutenção, quando o sistema já entrou em operação (SUMMERVILLE, 2003, p. 5).

Uma abordagem tradicional estabelece que eficiência, facilidade de manutenção, facilidade de uso e nível de confiança, são características essenciais para que um software seja considerado um produto de qualidade.

Summerville (2003, p.7), observa que o processo de software é considerado um conjunto de atividades e resultados associados, que possui quatro atividades básicas:

1. Especificação de software: definição de funcionamento e restrições.
2. Desenvolvimento de software: construção baseada nas especificações.
3. Validação do software: testes que garantam a qualidade do produto.
4. Evolução do software: melhorias que visam atender novas demandas.

Encontrou-se também uma definição das etapas de elaboração de software segundo Trebien (2002, p. 37), que afirma que independente do ciclo de vida ou paradigma de desenvolvimento de software, uma estrutura comum das fases de criação de um software compreende:

- a) fase 1 - definição de requisitos;

- b) fase 2 - projeto;
- c) fase 3 - implementação;
- d) fase 4 - teste;
- e) fase 5 - liberação para produção;
- f) fase 6 - liberação para comercialização.

Enfim, muitos são os modelos de processo de software utilizados pelo mercado hoje em dia mas, devido a complexidade da nossa tarefa de criar um SAD, mesclando recursos de Mineração de Dados e onde as fontes de informação possuem formatos diferenciados e envolverá processos manuais, como por exemplo o download dos arquivos com os resultados do Enade diretamente do site do Ministério da Educação; e para não perder muito tempo sem ver os resultados das etapas passo-à-passo, escolheu-se a abordagem chamada de Prototipação.

Segundo Sommerville (2003, p. 145):

Um protótipo é uma versão inicial de um sistema de software, que é utilizado para mostrar conceitos, experimentar opções de projeto e, em geral, para conhecer mais sobre os problemas e suas possíveis soluções. O desenvolvimento rápido de um protótipo é essencial para que (...) os usuários possam fazer experiência com o protótipo no início do processo de software.

Dessa maneira, como o presente trabalho trata-se de uma exploração por um domínio cujo resultado ainda é uma incógnita, à escolha deste modelo como ponto de partida pareceu ser a mais indicada, por permitir a integração de outros modelos durante o processo de desenvolvimento dos protótipos.

2.2 MODELAGEM UML

A escolha da UML como padrão de escrita para a modelagem do Sistema de Apoio à Decisão justifica-se por seus objetivos e recursos. Ela foi projetada para atender a alguns objetivos bastante específicos, para que possa verdadeiramente ser um padrão que resolva as necessidades práticas da comunidade de desenvolvimento de software (PENDER, 2004, p. 7).

Além disso, concorda-se com Pender (2004, p.10) que a UML foi desenhada para ser a base de integração entre qualquer quantidade de linguagens de modelagem, por exemplo: um modelo UML pode ser a entrada para um modelo de entidade-relacionamento e vice-versa.

Isto auxilia no trabalho de projetar, documentar e implementar o projeto, no acompanhamento do controle de qualidade e prepara o Sistema para o futuro, no ponto de vista que as implementações de novas funcionalidades ou mesmo correções ficam visíveis.

- Atores

Analizador

É o ator que agirá para adquirir os arquivos. Interagirá com outros atores e participará ativamente de praticamente todo o processo descoberta de conhecimento.

Tomador de Decisão

Fará suas críticas baseadas no trabalho do Analizador. Conhece bem a realidade da Instituição e o rumo das ações que devem ser adotadas em direção aos objetivos da Instituição.

- Casos de Uso:

Caso de Uso: Obter Dados

Descrição breve: O usuário, como analisador, deverá fazer os downloads dos arquivos do site do MEC e do Sistema de Avaliação Institucional, bem como solicitar e receber do GTIC da Unoesc de Joaçaba, os dados pertinentes as notas curriculares e organizá-los em pastas.

Descrição passo à passo:

Pré-condição: Identificação do usuário junto ao Sistema de Avaliação Institucional e autorização jurídica da Unoesc para manusear os dados acadêmicos.

Fluxo de eventos:

- No site do MEC são disponibilizados os arquivos em formato excell, com os resultados do Enade, que deverão ser baixados conforme o ano desejado;

- Após o login no Sistema de Avaliação Institucional, deverão ser acessados e salvos os arquivos em formato excell, correspondentes aos cursos e períodos desejados.
- Enviar requerimento ao GTIC de Joaçaba delimitando o período e campus que se deseja obter as informações dos alunos;
- Salvar os arquivos nas pastas corretamente, observando as nomenclaturas e ordens necessárias para posterior extração.

Pós-condição: Nenhuma.

Características:

Esta etapa do sistema de apoio à decisão é praticamente toda ela executada manualmente, devido ao fato de que os sistemas que fornecerão os dados de origem, não possuem integração entre eles, o que inviabiliza a automatização dessa fase. Independentemente disso, as ferramentas adotadas para o desenvolvimento suprem a necessidade de documentação e padronização de como ela deve ser executada.

Encerra-se por aqui a demonstração desta etapa, porque o resto dessa documentação segue esse mesmo padrão e seria repetitiva demais, além de não ser intenção deste artigo ensinar como utilizar a UML e nem suas ferramentas, apenas denotar sua importância para realização deste trabalho. Por isso, segue-se adiante com os demais processos de criação do SAD.

2.3 REQUISITOS NÃO-FUNCIONAIS

Após a modelagem, o desenvolvimento do sistema necessita nesta etapa, definir agora os requisitos inerentes à sua concepção prática. Isso é o que se chama de requisitos não-funcionais. Neste momento, é preciso delimitar toda a tecnologia que será empregada na programação, no banco de dados, na geração e tratamento dos resultados encontrados e seus registros para posterior consulta e acompanhamento evolutivo das ações tomadas.

No ambiente de desenvolvimento escolheu-se o Netbeans utilizando as hierarquias e bibliotecas de frameworks empregados para criação das páginas do sistema.

Descreve-se aqui, apenas os conceitos principais, pois cada um destes componentes consumiria um artigo inteiro se for visto com detalhes:

- Programas de extração: desenvolvidos para rodar no próprio desktop. Foi implementado utilizando a biblioteca JXL para leitura e gravação dos arquivos em excel. O objetivo era ler rapidamente os arquivos de origem e jogar para o banco de dados, o que deveria consumir pouco tempo para codificação mas, foi uma das tarefas mais complexas devido ao formato dos arquivos não seguirem um padrão, principalmente os oriundos da Avaliação da IES pois, a cada semestre, algumas perguntas do questionário mudam, outras são suprimidas, outras acrescentadas e com isso nenhum arquivo é igual ao outro.

Teve-se que utilizar estruturas condicionais complexas dentro de laços de repetição que varressem todo o arquivo e como uma pinça, retirasse apenas as informações necessárias. Apesar da necessidade de extrair uma vez só essas informações, é necessária absoluta certeza de que elas estão 100% corretas, caso contrário não servem para análise e quanto maior o grau de granularidade melhor para a mineração por isso, a preocupação em não desperdiçar nada, valorizando também o trabalho que se teve em adquirí-las.

Nas figuras 1, 2 e 3, têm-se alguns exemplos de como esses arquivos se encontram:

Na escolha seguinte, após criar os modelos de extração de dados, imagina-se que o sistema deveria permitir ser acessado de qualquer local, desta maneira, escolheu-se dentre os recursos disponíveis no mercado, a plataforma Java EJB com JSF, que segue a filosofia MVC e pareceu mais correto adotar o JPA como padrão de acesso aos dados. Para o banco de dados, escolheu-se o MySQL por ser um dos mais clássicos e para software não-comercial sua licença não necessita de pagamento.

Escolheu-se o JPA porque está se tornando a principal tendência em Java, uma vez que ele é menos dependente ainda dos softwares conhecidos como conectores de banco de dados, como EclipseLink e Hibernate. E o EJB porque permite a incorporação de comportamento de sistema desktop para um sistema web. É ele que irá facilitar a execução de rotinas de processamento de dados conforme as regras de negócio, diretamente na plataforma web.

2.4. MINERAÇÃO DOS DADOS

Para os procedimentos de mineração de dados, é necessário compilar toda a base já extraída, removendo dela todas as inconsistências que por ventura ainda existam.

No nosso caso analisado, descarta-se todas as linhas com informações faltantes e em outros casos, quando não havia todos os arquivos necessários, algumas colunas foram suprimidas.

O intuito é chegar num arquivo do formato compatível com a ferramenta Weka, que foi utilizada para fazer a mineração propriamente dita. Esta ferramenta utiliza um arquivo de extensão .arff.

A integração do Weka com o sistema é feita através das bibliotecas, que são as mesmas utilizadas pela interface própria. O Weka já foi construído dessa maneira para facilitar esse tipo de integração.

2.5 RESULTADOS

Nesta etapa descreve-se uma das análises feitas, com um pequeno conjunto de dados extraídos e preparados. Utilizou-se um arquivo contendo informações das notas curriculares de 92 alunos, distribuídas em 3 disciplinas de um mesmo curso, cruzadas com a avaliação institucional realizada no mesmo ano, para ver se as notas curriculares possuem algo em comum com a avaliação das disciplinas.

Primeiro, foi analisado através do algoritmo k-médias presente no Weka, para verificar a existência de algum padrão de agrupamento nos dados. A figura 4 mostrada a seguir, é o resultado do cruzamento do número de linhas de dados presentes no arquivo e as notas curriculares, o que não nos informa quase nada, no máximo que na segunda metade do arquivo têm-se menos notas abaixo da média do que na primeira parte.

A vantagem é que facilmente, pode-se ir trocando os cruzamentos que deseja-se ir verificando, em busca de algo mais significativo, como por exemplo agora, como as notas dos alunos que avaliaram melhor a disciplina na questão 5, são mais altas que a média:

Nesta análise pode-se notar que existe uma diferença na quantidade de alunos em cada disciplina e que conforme esta quantidade diminui, aumentou a quantidade de notas acima da média e também a avaliação da disciplina.

Na figura 6, quando se compara as notas com a questão nº 1 da avaliação da disciplina, observa-se uma distância muito grande entre as disciplinas 1 e 2 em relação a disciplina 3, que ficou abaixo das outras:

A ferramenta Weka demonstra sua versatilidade ao analisar também, dados nominativos, o que sugere que se pode, além das análises com base somente em números, substituir as notas curriculares por 3 faixas como: acima da média, abaixo da média e dentro da média. Desta maneira, tem-se uma nova iteração que levará a reanálise de toda à base, diversificando ou certificando a visão já analisada, abrindo o leque de oportunidade de se encontrar conhecimento relevante.

Enfim, serão essas análises que irão pouco à pouco compondo a base de dados de conhecimento, que irá auxiliar na tomada de decisões, permitindo que se registre as ações que podem melhorar a qualidade de ensino e posteriormente, como será visto na próxima parte, dar suporte ao acompanhamento dos resultados.

Após todo o trabalho de análise dos resultados da mineração de dados, tem-se um espaço descritivo no sistema que armazenará as críticas e

ações que poderão ser tomadas no sentido de melhorar a qualidade do ensino da Instituição.

Este registro faz-se necessário para que os gestores, responsáveis por tomar a decisão de implementar ou não as ações, que dependerão é claro, de estar em acordo com o planejamento já em curso, para que no futuro seus resultados possam novamente ser medidos.

3 CONCLUSÃO

Diante de tudo que apresentou-se até aqui, pode-se ver só um pouco do que pode resultar a mineração dos dados de uma base de uma Instituição de Ensino Superior e lançar novo olhar sobre essa técnica. Mais uma vez, observa-se que fazer milagre não é o campo da pesquisa científica, aqui, tudo depende de muito esforço para aprender a juntar as peças. E o Sistema de Apoio à Decisão aqui proposto, juntamente com a Mineração de Dados e a Descoberta de Conhecimento em Bases de Dados, demonstraram que são capazes de atestar cientificamente alguns achismos que havia-se imaginado no início. O resultado deste projeto foi atingido, ao se criar mais uma ferramenta de ajuda no rumo da melhora na qualidade de ensino de uma IES, agora permitindo se souber, de onde que se pode iniciar uma nova pesquisa.

O uso frequente de um sistema de apoio à decisão, dessa maneira, tende a formar uma estrutura de conhecimento baseado em casos, também conhecido na área de Inteligência Artificial como RBC, que poderá ser o aporte para novos trabalhos na área.

REFERÊNCIAS

PENDER, T. UML, a Bíblia. Rio de Janeiro: Elsevier, 2004.

PRESSMAN, R. S. Engenharia de Software. Rio de Janeiro: McGraw Hill, 2002. 5 ed.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Código da IES Arrega	Código da IES	IES	Sigla	UF	Organização	Dep. Administrativa	Código da Área	Área	Código da Subárea	Sub Área	Código do Município	Município	Número de Participantes Ingressantes	Número de Participantes Concluintes	Média FG_Conc	Média CE_Conc	Nota Enade Concluintes = Conceito Enade contínuo
1																		
2	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	1	ADMINISTRAÇÃO			510401703403	CUABA	182	98	46.2490	53.8102	4.06
3	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	2	DIREITO			510401703403	CUABA	82	96	58.7219	71.2781	4.41
4	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	8	COMUNICAÇÃO SOCIA	803	JORNALISMO	510401703403	CUABA	23	54	54.1833	53.5519	3.75
5	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	8	COMUNICAÇÃO SOCIA	803	JORNALISMO	510502076653	PONTAL DO ARI	36	0			
6	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	8	COMUNICAÇÃO SOCIA	804	PUBLICIDADE E PROF	510401703403	CUABA	22	56	55.0357	56.9518	3.57
7	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	8	COMUNICAÇÃO SOCIA	805	RADIALISMO	510401703403	CUABA	16	47	35.8979	37.6532	0.35
8	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	13	CIÊNCIAS ECONÔMICAS			510401703403	CUABA	38	127	39.0150	31.9055	1.97
9	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	18	PSICOLOGIA			510401703403	CUABA	69	0			
10	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	18	PSICOLOGIA			510502107602	RONDONOPOL	38	32	40.9219	48.9344	2.61
11	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	22	CIÊNCIAS CONTÁBEIS			510401703403	CUABA	73	126	40.9635	43.5722	3.64
12	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	22	CIÊNCIAS CONTÁBEIS			510502107602	RONDONOPOL	77	61	50.9607	34.3311	2.89
13	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	43	MÚSICA			510401703403	CUABA	29	48	30.1883	46.0875	1.42
14	1	1	UNIVERSIDADE FED	UFMT	MT	Universidade	PÚBLICA	66	BIBLIOTECOLOGIA			510502107602	RONDONOPOL	34	14	31.7786	50.6214	1.90
15	2	2	UNIVERSIDADE DE E	UNIB	DF	Universidade	PÚBLICA	1	ADMINISTRAÇÃO			530100100108	BRASILIA	175	136	62.3471	57.0893	4.80

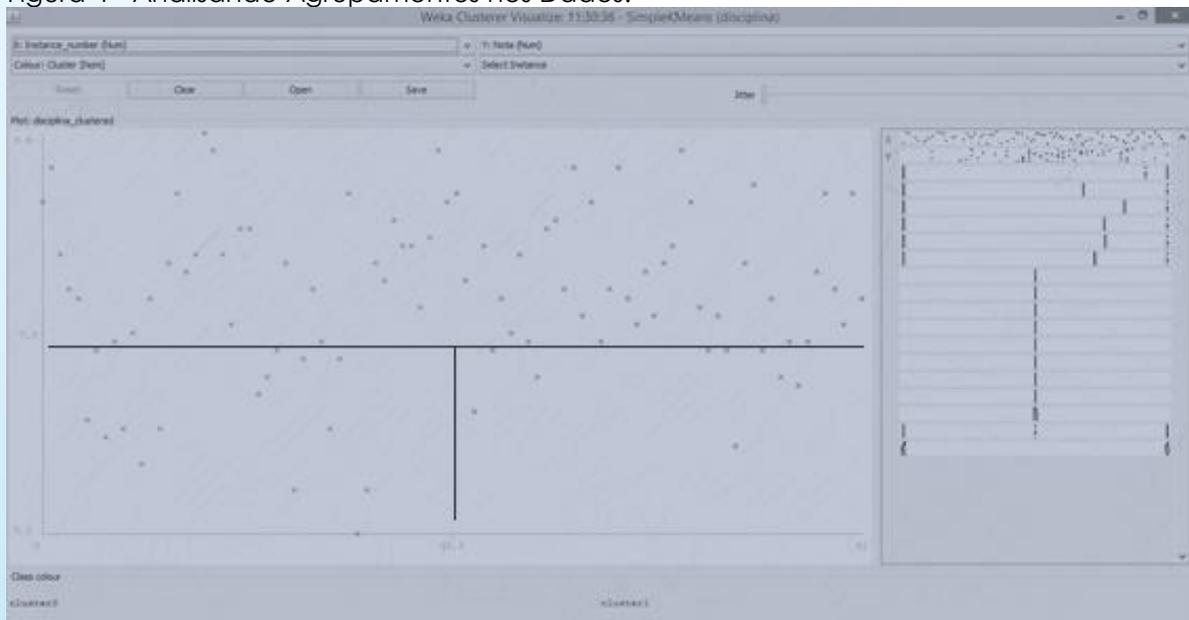
Fonte: O Autor

Figura 3 - Arquivo recebido com o histórico dos alunos.

	Q	R	S	T	U	V	W	X	Y	Z	
	ANO_BASE	PER_BASE	PERCENTUAL_FREQ	CARGA_HORARIA	FALTAS	MED_PARCIAL	MEDIA_FINAL	SITUACAO	'HISTÓRICOESCOLAR'	COD_CURSO_DISCIPLINA	CURSO_DISCIPLINA
61274	2012	1	0	60	0	8	8,00	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61275	2012	1	0	60	0	0	8,00	APROVADO	HISTÓRICO ESCOLAR	0	
61276	2012	2	0	60	0	8,7	8,70	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61277	2012	2	0	60	0	0	8,70	APROVADO	HISTÓRICO ESCOLAR	0	
61278	2013	1	0	30	0	7,8	7,80	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61279	2013	1	0	30	0	0	7,80	APROVADO	HISTÓRICO ESCOLAR	0	
61280	2012	1	0	30	0	8,8	8,80	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61281	2012	1	0	30	0	0	8,80	APROVADO	HISTÓRICO ESCOLAR	0	
61282	2013	1	0	30	0	5,7	5,10	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61283	2013	1	0	30	0	0	5,10	APROVADO	HISTÓRICO ESCOLAR	0	
61284	2011	2	0	60	0	8,2	8,20	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61285	2011	2	0	60	0	0	8,20	APROVADO	HISTÓRICO ESCOLAR	0	
61286	2012	2	0	60	0	5,1	6,60	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61287	2012	2	0	60	0	0	6,60	APROVADO	HISTÓRICO ESCOLAR	0	
61288	2012	1	0	60	0	9,3	9,30	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61289	2012	1	0	60	0	0	9,30	APROVADO	HISTÓRICO ESCOLAR	0	
61290	2011	2	0	60	0	8,3	8,30	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61291	2011	2	0	60	0	0	8,30	APROVADO	HISTÓRICO ESCOLAR	0	
61292	2011	1	0	30	0	8,6	8,60	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61293	2011	1	0	30	0	0	8,60	APROVADO	HISTÓRICO ESCOLAR	0	
61294	2012	1	0	60	0	8,4	8,40	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI
61295	2012	1	0	60	0	0	8,40	APROVADO	HISTÓRICO ESCOLAR	0	
61296	2012	2	0	60	0	9	9,00	APROVADO	DISCIPLINAS_CURSADAS	376	SUPERIOR DE TECNOLOGIA EM PRI

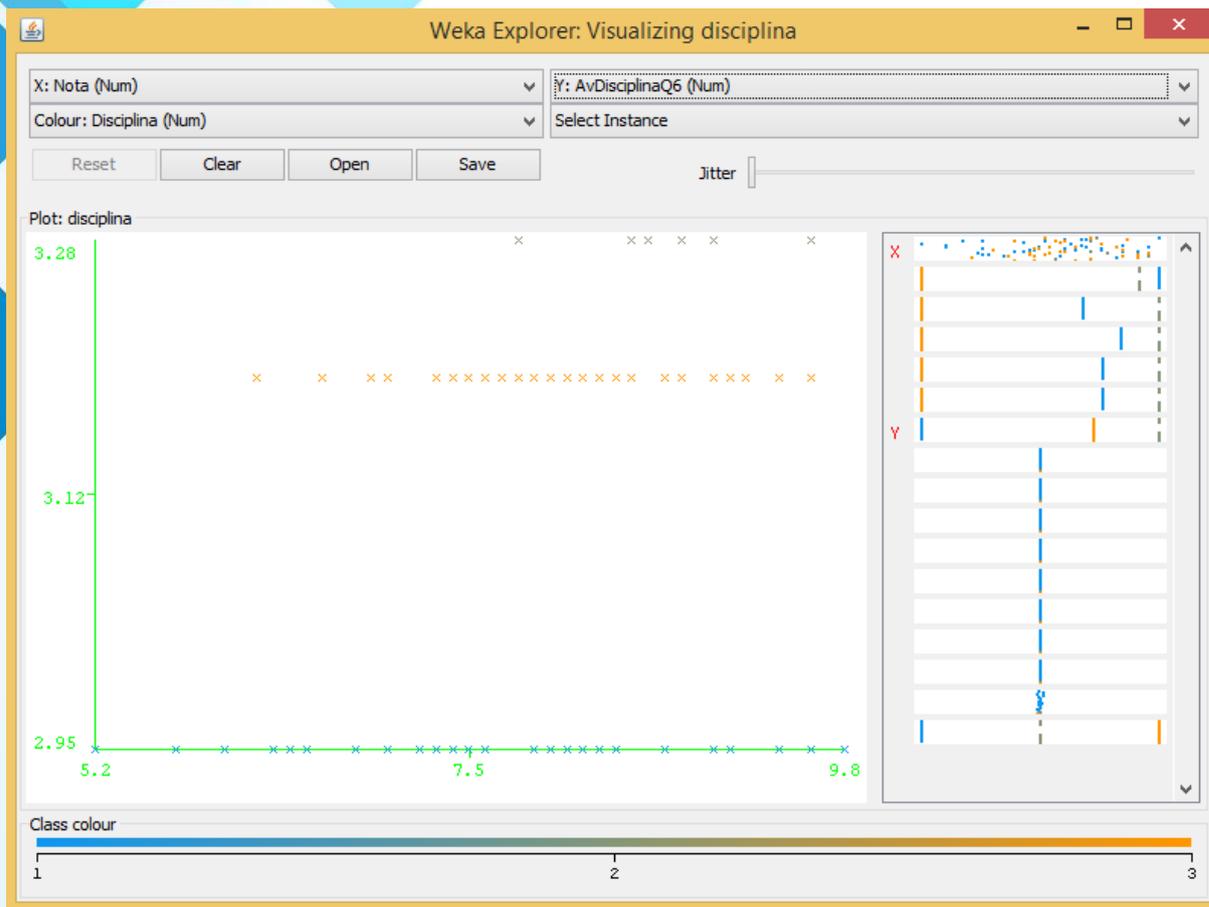
Fonte: O Autor

Figura 4 - Analisando Agrupamentos nos Dados.



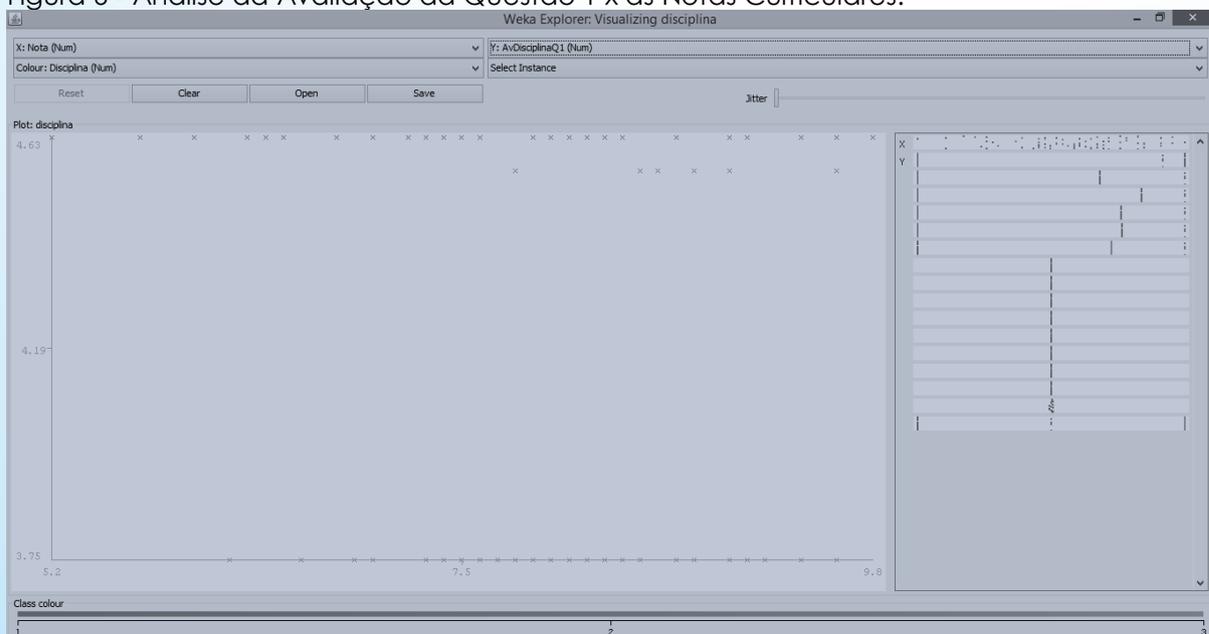
Fonte: O Autor

Figura 5 - Visualizando agrupamento Notas x Questão 6.



Fonte: O Autor

Figura 6 - Análise da Avaliação da Questão 1 x as Notas Curriculares.



Fonte: O Autor