Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção

UTILIZAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS NA ANÁLISE DAS INFORMAÇÕES DE UMA UNIVERSIDADE

Rudiney Marcos Herdt

Dissertação apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de Santa Catarina
como requisito parcial para obtenção
do título de Mestre em
Engenharia de Produção

Florianópolis
2001

Rudiney Marcos Herdt

UTILIZAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS NA ANÁLISE DAS INFORMAÇÕES DE UMA UNIVERSIDADE

Esta dissertação foi julgada e aprovada para a obtenção do título de **Mestre em Engenharia de Produção** no **Programa de Pós-Graduação em Engenharia de Produção** da Universidade Federal de Santa Catarina

Florianópolis, 26 de novembro de 2001.

Prof. Ricardo Miranda Barcia, Ph.D.
Coordenador do Curso

BANCA EXAMINADORA

Aran Bey Tcholakian Morales, Dr. Eng.
Orientador

Alexandre Leopoldo Gonçalves, M.Eng.

Oscar Ciro Lopes, Dr. Eng.

Ricardo Villarroel Dávalos, Dr. Eng.

Agradecimentos

A Deus, que é a fonte de vida e que me proporcionou saúde e força para a realização deste trabalho.

À minha esposa Simone, que sempre esteve ao meu lado, incentivando-me a realizar este curso de mestrado e compreendendo os momentos em que tive de ficar ausente.

Gostaria de deixar o meu agradecimento especial ao meu orientador, Aran Bey Tcholakian Morales, que participou em todas as fases na elaboração deste trabalho, possibilitando-me uma diferencial orientação.

Aos meus colegas de trabalho que de alguma forma colaboraram para o meu aprendizado.

À Assessoria de Avaliação Institucional da Unisul, na figura dos professores Jailson Coelho e Eduardo Búrigo, que forneceram o apoio necessário e os dados que serviram de fonte para a pesquisa.

A todos os professores e colegas que, durante o curso, foram amigos e souberam repartir suas experiências, abrindo novos horizontes e possibilitando o amadurecimento de idéias que fortaleceram a busca de novos conhecimentos.

Sumário

Lista de Ilustrações

Lista de Tabelas

Lista de Gráficos

Lista de Reduções

FESSC	- Fundação Educacional do Sul de Santa Catarina
KDD	- Knowledge Discovery in Databases
OLAP	- On-line Analytical Processing
SAD	- Sistema de Apoio à Decisão
SAE	- Setor de Apoio ao Estudante
SAEST	- Setor de Apoio aos Estágios
Unisul	- Universidade do Sul de Santa Catarina

Resumo

Uma realidade que está cada vez mais presente em todas as organizações é a necessidade na busca por informações. As técnicas de mineração de dados são apresentadas como revolucionárias porque geram suas próprias hipóteses e garantem informações com maior rapidez. O principal objetivo da mineração de dados é encontrar padrões de comportamento em um grande volume de dados. A mineração de dados é uma das etapas no processo de descoberta de conhecimento em banco de dados KDD. As principais etapas no processo de KDD são: definição de metas, seleção, pré-processamento, transformação, mineração de dados e interpretação. O *data warehouse* não é obrigatório, mas é de grande importância para uma mineração de dados eficiente. A mineração de dados pode ser utilizada na aplicação de várias técnicas. As principais são: clusterização, classificação e regras de associação. A Universidade do Sul de Santa Catarina (Unisul) possui um processo permanente de avaliação, denominado Avaliação Institucional, que é feito através de uma pesquisa com 65 questões, que são aplicadas aos alunos anualmente. Assim, o presente trabalho busca encontrar algum padrão de comportamento nos dados resultantes desta pesquisa.
Foram utilizadas para análise uma ferramenta estatística e duas técnicas de mineração de dados: clusterização e regras de associação. Este trabalho procura utilizar os recursos de cada uma dessas técnicas para encontrar conhecimentos úteis na tomada de decisão bem como fazer uma comparação entre elas.

Palavras-chave: [mineração de dados, descoberta de conhecimento em banco de dados, estatística]

Abstract

A reality that is each more present in the organizations is the necessity in the fetching for information. The data mining techniques are presented as revolutionary because they generate its proper hypotheses and they guarantee information with bigger rapidity. The main objective of the data mining is to find standards of behavior in a great volume of data. The data mining is one of the stages in the process of discovery of knowledge Discovery Database - KDD. The main stages in the KDD process are: definition of goals, selection, daily pay-processing, hashing, data mining and interpretation. The date warehouse is not obligator, but it is of great importance for an efficient data mining. The data mining can be used in the application of several techniques. The main ones are: clustering, sorting and rules of association. The University of the South of Santa Catarina (Unisul) have a permanent process of evaluation, called Institucional Evaluation, that is made through a research with 65 questions, that are applied the pupils annually. Thus, the present work searchs to find some standard of behavior in the resultant data of this research.
They had been used for analysis a tool statistics and two techniques of data mining: clustering and association rules. This work looks for to use the features of each one of these techniques to find useful knowledge in the taking of decision as well as making a matching between them.

Key-word: [data mining, knowledge discovered database, statistics]

1 INTRODUÇÃO

1.1 APRESENTAÇÃO

As empresas estão percebendo que a informação tornou-se o seu maior bem. Na busca por qualidade, elas apostam na informação como a melhor estratégia de competitividade.

O desenvolvimento da tecnologia e uma considerável redução dos preços de equipamentos de armazenagem e processamento fazem com que o volume de dados cresça rapidamente. Pesquisas indicam que a quantidade da informação no mundo dobra a cada 20 meses (PILA, 2000). Essa enorme quantidade de dados armazenada em banco de dados muitas vezes supera a possibilidade de esses dados realmente serem utilizados, não se aproveitando, dessa forma, os benefícios diretos (Figueira 1999). Seu valor verdadeiro está na habilidade de se extrair de uma imensidão de informações aquelas que são verdadeiramente úteis para a tomada de decisão. Esses e outros fatores fazem com que as empresas invistam em tecnologia de informação voltada à descoberta de conhecimento. A mineração de dados, através de suas técnicas, surge como uma poderosa ferramenta capaz de encontrar em enormes quantidades de dados valiosas informações que possam detectar padrões de comportamentos em diferentes conjuntos de dados.

Essas mesmas informações também estão se tornando um diferencial no mundo empresarial das instituições educacionais, que vêm crescendo muito nestes últimos anos. Isso se dá pela alta demanda na busca de capacitação e de especialização profissional ou mesmo de novos conhecimentos, tentando encontrar um diferencial na briga por um posicionamento profissional destacado. Esse promissor mercado está fazendo com que uma grande quantidade de novas instituições educacionais venha fazer parte dele, bem como também faz com que as instituições existentes busquem o seu espaço para continuar

existindo. Nesse mundo competitivo, onde somente as mais ágeis terão seu lugar garantido, torna-se indispensável o conhecimento do seu negócio para a tomada de decisão rápida e precisa.

A Universidade do Sul de Santa Catarina (Unisul) possui os cursos oferecidos como seu principal produto e procura, entre seus objetivos, alcançar a qualidade desses cursos.

Buscando essa melhoria na qualidade, a Unisul criou uma assessoria para tratar da avaliação dos professores, dos coordenadores e das instalações físicas chamada Assessoria da Avaliação Institucional.

Por intermédio dessa assessoria, a Unisul realiza anualmente pesquisas de avaliação de alunos, professores, dirigentes e infra-estrutura. Estas pesquisas, vem acumulando uma grande quantidade de dados e somando com os demais dados oriundos dos sistemas operacionais existentes, gerou um significativo bando de dados.

O presente trabalho pretende, através das avaliações periódicas realizadas pela universidade sobre seus cursos, examinar as informações obtidas dos alunos. Os resultados dessas avaliações poderão ser utilizados como ferramenta de gestão, de modo a verificar em que pontos a universidade deve melhorar para satisfazer o seu principal cliente, o aluno.

1.2 JUSTIFICATIVA

A avaliação institucional é entendida como um processo contínuo de aperfeiçoamento das ações desenvolvidas pela Universidade na busca da qualidade de seus serviços de ensino, pesquisa, extensão e gestão.

O processo de avaliação institucional envolve a coleta de uma grande quantidade de informações referentes a cada disciplina cursada na Unisul. Esse processo gera um grande volume de dados, e o trabalho de retirada de informações úteis torna-se extremamente

custoso. Essas dificuldades fazem com que muitas possíveis informações gerenciais deixem de ser retiradas.

Diante dessas necessidades e levando-se em conta a tecnologia hoje disponível, a utilização de ferramentas especializadas na busca e interpretação de informações voltadas à tomada de decisão parece ser uma forte aliada de uma organização competitiva e de qualidade. Essas ferramentas vão encontrar padrões diversos com a utilização de diferentes técnicas de mineração dados.

1.3 OBJETIVO

Este trabalho tem como objetivo analisar, através de técnicas de mineração de dados, as informações obtidas dos alunos nas periódicas avaliações realizadas pela universidade, para posterior utilização como ferramenta de apoio às analises realizadas pela avaliação institucional.

1.3.1 Objetivos específicos

- Aplicar técnicas de analises exploratória de dados a base de dados utilizada na avaliação institucional de uma universidade.
- Utilizar o processo de descoberta de conhecimento como metodologia de análises da base de dados.
- Aplicar diferentes técnicas de mineração de dados na base utilizada para análise.
- Comparar e analisar os resultados encontrados nas diferentes técnicas utilizadas.
- Analisar o perfil dos alunos nos diferentes campi.

1.4 ESTRUTURA DO TRABALHO

Além do presente capítulo, este trabalho é composto de mais quatro capítulos. O capítulo corrente objetiva contextualizar e justificar como a aplicação de técnicas de mineração de dados irá auxiliar na resolução do problema de busca de informações gerenciais em universidades de ensino. A estrutura dos demais capítulos apresenta-se da seguinte maneira:

- **Capitulo 2**: apresenta uma introdução sobre mineração de dados, sua conceituação e um histórico de como surgiu essa tecnologia. É conceituado o processo de descoberta de conhecimento bem como também são apresentadas quais as principais etapas desse processo. Por último, serão apresentadas algumas das técnicas de mineração de dados e os algoritmos utilizados neste trabalho para cada técnica.

- **Capítulo 3**: este capítulo trata da Avaliação Institucional na Unisul. Será apresentada a logística de captura de dados, os formulários preenchidos pelos alunos e o significado das questões, assim como os resultados alcançados nessa avaliação.

- **Capitulo 4**: neste capítulo serão apresentadas com detalhes a metodologia proposta para a busca de informações gerenciais e a descrição de como foi resolvida cada etapa no processo de descoberta de conhecimento.

- **Capitulo 5**: capítulo em que serão relatadas as conclusões do trabalho bem como algumas recomendações e projetos para trabalhos futuros.

2 MINERAÇÃO DE DADOS

2.1 INTRODUÇÃO

O mundo dos negócios tem armazenado uma quantidade incrível de dados, principalmente no que diz respeito ao perfil dos consumidores. No cenário atual, as empresas reagem rapidamente às mudanças no mercado, o que faz com que a necessidade da informação torne-se cada vez mais indispensável. É nesse contexto que as técnicas de mineração de dados estão sendo cada vez mais utilizadas para converter grandes massas de dados em informações estratégicas. São usadas em diversas áreas como análise de riscos, marketing direcionado, controle, controle de qualidade, análise de dados, etc. Essa tecnologia está sendo utilizada para descrever características do passado, assim como predizer tendências para o futuro. Sua utilização permite avanços tecnológicos e descobertas científicas, além de garantir vantagem competitiva.

Uma diferença significativa entre a mineração de dados e outras ferramentas de análise está na maneira como estas exploram as interrogações dos dados. As diversas ferramentas de análise disponíveis dispõem de um método baseado na verificação, isto é, o usuário constrói hipóteses sobre interrogações específicas e então verifica a resposta através do sistema. Esse método torna-se dependente da intuição e da habilidade do analista em propor hipóteses interessantes. Já o processo de mineração de dados fica responsável pela geração de hipóteses, garantindo dados apurados e com maior rapidez.

2.2 MINERAÇÃO DE DADOS: DEFINIÇÃO

Mineração de dados é uma tradução literal do termo inglês "data mining" que significa garimpar imensas quantidades de dados armazenados, em que uma busca

incansável e persistente possibilita encontrar uma minúscula pepita de ouro em uma montanha de entulho de dados (SANTOS, 1999).

Mineração de dados consiste na análise de dados históricos para identificação de padrões que possam esclarecer o presente. Ela não só responde à previsão de negócios como também pode revelar os atributos mais importantes que influenciam nessas previsões (THEARLING, 2000).

Mineração de dados é o processo de significante extração de informações previamente não conhecidas de enormes bases de dados, sendo essas informações utilizadas para se tomar importantes decisões nos negócios. A mineração de dados inclui avançadas tecnologias de software, metodologias inovadoras e, freqüentemente, serviços de consultas (DW BRASIL, 2000).

Mineração de dados caracteriza um conjunto de técnicas utilizadas de maneira automática para explorar exaustivamente relações complexas existentes em banco de dados de grandes dimensões e torná-las claras (WEISS, 1998), (FELDENS, 1999).

Na verdade, mineração de dados constitui um dos passos de um processo maior denominado Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases* – KDD), que é realizado por ferramentas computacionais em desenvolvimento para crescentes volumes de dados. O termo KDD foi formalizado em 1989 em referência ao amplo conceito de se procurar conhecimento em dados.

Assim, pode-se dizer ainda que mineração de dados é um passo no processo de KDD que consiste na aplicação de análise de dados e algoritmos de descobrimento que produzem uma enumeração de padrões (ou modelos) particular sobre os dados. (FAYYAD,1997).

O interesse de se peneirar montanhas de dados das empresas e transformá-las em informação de alto valor para a tomada de decisão já existia há muito tempo. Com o barateamento das tecnologias, a implantação desse processo foi viabilizada.

Sistemas existentes desde os anos 80 no nicho científico hoje têm condições de ser aplicados em áreas comerciais, permitindo a construção de base de dados de altíssimos volumes, capazes de refletir a história da transformação de um dado em tendências ou em pesquisas complexas (GIMENES,2000).

A tecnologia de mineração de dados é resultado de um processo de pesquisa e desenvolvimento de produtos. Essa evolução começou quando os dados eram primeiramente estocados em computadores, com melhoria de acesso e, mais recentemente, com o avanço de tecnologias que permitiram ao usuário navegar entre seus dados em tempo real. Com esse revolucionário processo, a mineração de dados permitiu ir além do acesso ao dado já armazenado como também fazer prospectivas e proativar a informação esperada.

A utilização de dados e informações comerciais vem sendo utilizadas a muito tempo e passou por diversas fases, passando de uma simples consulta tradicional até mineração de dados. A seguir o quadro 1 apresenta as etapas evolucionárias da mineração de dados.

Quadro 1 - Etapas na evolução do tratamento das informações nas organizações

Etapa evolucionária	Questão Comercial	Tecnologias Disponíveis	Fornecedores de produtos	Características
Coleção de dados (1960s)	"Qual foi minha receita total nos últimos cinco anos"?	Computadores, fitas e discos	IBM, CDC	Retrospectiva e distribuição de dados estática
Acesso a dados (1980s)	"Quais foram as vendas unitárias de São Paulo em março?"	Bancos de dados relacionais (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospectiva, distribuição de dados dinâmica a nível de registros
Data warehousing & Suporte à decisão (1990s)	"Quais foram as vendas unitárias de São Paulo em março? Avalie também Campinas."	On-Line Analytical Processing (OLAP), bancos de dados multidimensionais, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospectiva, distribuição dinâmica de dados a múltiplos níveis
Data mining (atualmente)	"Qual a previsão para as vendas de Campinas no próximo mês? Por quê?	Algoritmos avançados, computadores multiprocessados, bancos de dados massivos	Pilot Lockheed, IBM, SGI, E outras (novas empresas)	Prospectiva, distribuição de informação ativa

FONTE: UNIVERSIDADE ESTUDUAL DE MARINGA. GRUPO DE SISTEMAS INTELIGENTES, 1999

Podemos observar no quadro 1 que as consultas tradicionais em banco de dados contrastam com a mineração de dados simplesmente porque estão limitadas a questões simples, tais como "quais foram as vendas de suco de laranja em janeiro de 1995 em São Paulo?". A análise multidimensional, geralmente chamada de *On-line Analytical Processing* (OLAP), habilita os usuários a fazer consultas muito mais complexas, como, por exemplo, comparação das vendas programadas de diferentes regiões. A mineração de dados, por outro lado, através do uso de algoritmos de busca, tenta descobrir padrões, tendências e inferir regras. Essas regras auxiliam o usuário na tomada de decisões em alguma área comercial ou científica.

As companhias estão começando a perceber que o seu mais precioso patrimônio é a informação que eles possuem do consumidor e dos padrões de compra. O sucesso na competitividade depende da qualidade da tomada de decisão e de um processo de melhoria

dessa qualidade com base em transações e decisões anteriores. A habilidade de se melhorar o conhecimento sobre os consumidores e o mercado permitirá que gerentes direcionem melhor seus produtos e serviços.

A competitividade está exigindo aos sistemas de apoio à decisão que respondam, cada vez mais, perguntas mais complexas (THEARLING, 2001).

2.3 DATA WAREHOUSE

O *data warehouse* não é um fator indispensável no processo de KDD, mas se pode dizer que o potencial da mineração de dados é muito maior quando tem-se os dados apropriados, coletados e armazenados em um *data warehouse*. O Data Warehousing é uma poderosa técnica que torna possível a extração de dados operacionais e a eliminação de problemas de inconsistência entre formatos e dados legados (UNIVERSIDADE ESTUDUAL DE MARINGA. GRUPO DE SISTEMAS INTELIGENTES, 1999) . Como é possível a integração de dados independentes de localização e formatos, é possível também a incorporação de informações adicionais.

2.3.1 Características de um Data Warehouse

Segundo (INMON,1997), um *data warehouse* é um conjunto de dados baseado em assuntos integrados, não volátil, variável em relação ao tempo e de apoio às decisões gerenciais. Suas principais características são:

- **Organizado por assunto**: os dados em um *data warehouse* são organizados *com base* nos assuntos ou nos negócios de interesse das empresas (BISPO, 1999). Por exemplo, uma companhia de seguros que usa *data warehouse* organiza seus dados por clientes, prêmios e apólices, em vez de produtos diferentes (auto, vida, etc). Os

dados são organizados por assunto e contêm as informações necessárias para o processo de decisão.

- **Integrado**: quando o dado reside em diferentes aplicações no ambiente operacional, sua codificação normalmente é inconsistente (UNIVERSIDADE ESTUDUAL DE MARINGA. GRUPO DE SISTEMAS INTELIGENTES, 1999) . Um exemplo é o caso do campo Sexo que, em uma aplicação, pode ser "m" e "f", e em outra, "0" e "1". Quando os dados são integrados no *data warehouse* esses valores devem ser transformados para uma codificação única.
- **Variante em relação ao tempo**: o *data warehouse* contém dados de dez anos ou mais para ser utilizados em comparações, tendências e previsões. Esses dados não são atualizados (DILLY,2000).
- **Não-Volatilidade**: uma vez colocados no *data warehouse*, os dados não sofrem nenhuma atualização, somente são acessados (DILLY,2000).

2.3.2 Processo em Data Warehouse

O principal processo de armazenamento de dados é composto dos seguintes subprocessos:

- **Extração** - Este é o primeiro passo no processo de colocar dados para o D.W. Significa ler e entender a origem dos dados e copiar partes desses dados que serão necessários ser armazenados para trabalhos futuros.
- **Transformação** - Todos os dados que são extraídos podem passar por alguns procedimentos de transformação antes de ser armazenados.
  - **Limpeza** - Na limpeza dos dados, são resolvidos erros, conflitos de domínio ou formatos (KIMBALL, 1998).

- **Purging** - registros selecionados dos dados legados que não são utilizados pelo *data warehouse* (KIMBALL, 1998).
- **Combinação de dados** - Combinação de campos ou inclusão de informações textuais equivalentes de código de sistemas legados.
- **Chaves substitutas** - Criar chaves substitutas para cada dimensão de *record* em ordem de dependência com as chaves de sistemas legados. Deve haver uma integridade entre as tabelas de dimensão e as tabelas de fatos.
- Construir agregações para melhorar a performance de *queries* comuns.
- **Carregando e indexando** - Depois da transformação, os dados estão prontos para ser carregados para o *data warehouse*. Em seguida, os dados devem ser indexados para ganho de performance.

- **Cheque a qualidade** - Depois de todos os outros passo realizados, o último vai checar a qualidade dos dados. Todas as outras categorias de relatório devem ser preservadas e todos os contadores e totais devem ser satisfatórios. Todos os valores devem ser consistentes com as séries temporais dos dados de origem.
- **Release/publicações** - Depois dos dados carregados e da qualidade garantida, os usuários podem ser notificados de que novos dados já estão disponíveis (KIMBALL, 1998).
- **Atualização** - Ao contrário dos tradicionais *data warehouse*, os modernos *data marts* podem ser atualizados, às vezes, com freqüência. Geralmente são dados carregados através de cargas gerenciadas e não atualizações transacionais (KIMBALL, 1998).
- **Querying** - É o processo de uso do *data warehouse*, que envolve consultas de usuários finais, relatórios impressos, complexas aplicações de suporte à decisão e mineração de dados.

- **Feedback dos dados** - Neste processo, é feito um feedback dos dados do *data warehouse* com os dados legados e com os resultados obtidos nas consultas de mineração de dados.
- **Auditoria** - Aqui é importante saber de onde os dados vêm e como são calculados. Algumas técnicas de auditoria prevêem *links* diretos com os dados reais, possibilitando ao usuário verificar o registro de auditoria a qualquer momento.
- **Segurança** - Todo *data warehouse* passa por um dilema: se os dados ficarem acessíveis a todos, corre-se o risco de *hackers* ou espiões destruí-los; proteger-se totalmente dos *hackers* pode prejudicar o acesso aos dados. Deve-se, portanto, analisar o meio-termo entre acessível e seguro, de acordo com cada necessidade.
- **Backup e recover** - A questão aqui é definir quais os tipos variáveis que devem passar por uma atividade de *backup* e como será feito o *recover* desses dados quando for necessário.

2.3.3 Critérios para Data Warehouse

Alguns critérios devem ser observados em um projeto de *data warehouse*:

- **Performance de carregamento**: o *data warehouse* requer um carregamento com incremento de novos dados em um período de tempo bastante estreito. O *data warehouse* deve possuir um ótimo desempenho nesse carregamento para que não prejudique o grande volume de dados que deve receber.
- **Processamento de carga**: o carregamento de dados passa por várias etapas, incluindo conversões de dados, filtragens, reformatação, checagem de integridade, armazenamento físico, indexação e atualização de metadados. Essa etapa deve ser executada como uma única unidade de trabalho.

- **Qualidade de administração de dados**: o Warehouse deve assegurar consistência e integridade referencial, apesar de fontes sujas e banco de dados volumoso.
- **Desempenho de *query***: *queries* grandes, complexas para operações de negócios-chave devem ser completadas em segundos e não em dias (DILLY,2000).
- **Escalabilidade Terabyte**: os tamanhos dos *data warehouse* estão crescendo a taxas surpreendentes. Hoje variam de *gigabytes* a *terabyte*, e o bando de dados responsável pelo armazenamento não deve possuir nenhuma limitação nesse sentido. O desempenho das *queries* não deve depender do tamanho do bando de dados e sim da complexidade da questão (DILLY,2000).
- **Escalabilidade de usuário de massa**: o *data warehouse* deve suportar centenas de usuários simultâneos, sem perder o desempenho aceitável das *queries*.
- **Redes de Data Warehouse**: *data warehouse* raramente existem em isolamento. Sistemas de múltiplos *data warehouse* cooperam em uma rede maior de *data warehouse*. O servidor deve incluir ferramentas que coordenam o movimento de subconjuntos de dados entre warehouses. Os usuários devem ser capazes de olhar múltiplos warehouse de uma única workstation client e trabalhar com eles. Os gerentes de warehouse têm que gerenciar e administrar a rede warehouse de uma única localização física.
- **Administração de warehouse**: o *data warehouse* exige facilidades administrativas e flexibilidade. Deve possuir limites de recursos por usuário e ter priorização de *queries* em diferentes classes de usuários.
- **Análise dimensional**: o poder de visões multidimencionais é amplamente aceitável. O *data warehouse* deve suportar a criação rápida e fácil de resumos.

- **Funcionalidade de query avançada**: os usuários finais requerem cálculos analíticos avançados, análise seqüencial e comparativa, e acesso consistente para detalhar e resumir dados. Usar SQL em um ambiente de ferramenta point-and-click clinte/servidor às vezes pode não ser prático, ou até mesmo impossível. O *data warehouse* deve prover um conjunto de operações analíticas, incluindo operações seqüenciais e estatísticas.

2.3.4 Por que o Data Warehouse facilita a mineração de dados

Os *data warehouse* organizam o estágio para um efetivo processo de Data Warehousing. A mineração de dados pode ser feita sem o uso de um *data warehouse*, mas, este aumenta as chances do sucesso da mineração de dados (PILA, 2001).

Os *data warehouse* organizam os estágios de uma mineração de dados através de sua natureza, que incluem:

- dados Integrados;
- dados detalhados e resumidos;
- dados históricos; e
- metadados;

Cada um desses elementos melhora o processo da mineração de dados e os prospectos de sucesso. Abaixo seguem as respectivas descrições.

Dados integrados: sem os dados integrados, o minerador gastaria muito mais tempo limpando e condicionando os dados antes do processo da mineração de dados. Os *data warehouse* são integrados e têm todas essas tarefas (e muitas outras) feitas, portanto o minerador pode concentrar-se na mineração de dados em vez de limpar e integrar os dados.

Dados detalhados e resumidos: tanto os dados detalhados como os resumidos já existem

no d*ata warehouse*. Em alguns momentos, o minerador precisa analisar os dados em sua forma mais granular. Uma análise detalhada dos dados pode trazer informações importantes. Por outro lado, também é fundamental a utilização de dados resumidos para uma análise macro. Como esses dados resumidos também estão disponíveis no data warehouse, o minerador pode utilizar-se do trabalho de outros em vez de fazer todo o processo desde o início. Essa capacidade proporciona verdadeira e fácil disponibilidade de dados resumidos, salvando grandes volumes de trabalhos desnecessários de mineração.

Dados históricos: grandes quantidades de dados históricos são fundamentais para que o minerador de dados possa encontrar informações implicitamente guardadas neste depósito de dados. Um minerador que só possui dados atuais pode nunca detectar tendências e padrões de comportamento ao longo do tempo.

Metadados: são os dados sobre o dado, ou seja, o significado de cada dado encontrado no *data warehouse*. Esta informação é muito importante para o minerador, pois é através dela que ele pode descrever o contexto da informação. O metadado de uma informação está sendo examinado. Com o passar do tempo, torna-se tão importante quanto o próprio conteúdo. O minerador que usa como recurso os dados acomodados em um *data warehouse* conseqüentemente torna o seu trabalho mais fácil e mais eficiente.

2.3.5 Quem são os usuários do Data Warehouse

O usuário do *data warehouse* é uma pessoa que pode ser chamada de analista de SAD (Sistema de Apoio à Decisão). O analista de SAD é, antes de mais nada, uma pessoa de negócios e, além disso, um técnico. A principal tarefa do analista de SAD consiste em definir e descobrir informações usadas no processo corporativo de tomada de decisões.

O analista de SAD tem a postura de "me dê o que eu digo que quero e então eu poderei lhe dizer o que realmente quero" (INMOM, 1997). Em outras palavras, o analista de SAD age por meio de descobertas. Somente ao ver um relatório ou uma tela, ele pode começar a explorar o potencial do SAD.

2.4 NECESSIDADE DA MINERAÇÃO DE DADOS

Muitas empresas investem milhares de dólares em tecnologia de informação para ajudá-las na administração de seus negócios, melhorando sua eficiência e gerando competitividade. Há muitos anos que as empresas possuem grandes quantidades de dados armazenados, e espera-se um considerável aumento para um futuro próximo. Apesar dessa abundância de dados, muitas companhias têm sido incapazes de aproveitar o seu valor. Isso porque as informações implícitas das bases de dados não são facilmente obtidas.

Por exemplo, uma empresa pode guardar informações detalhadas de todas as compras dos clientes, mas ainda não consegue classificar ou verificar tendências de seus clientes. Similarmente, as empresas de seguros sempre armazenaram informações sobre as reclamações e ocorrências, e continuam com dificuldades de identificarem fraudes ou agruparem tipos de ocorrências.

Felizmente, avanços no campo da descoberta do conhecimento estão auxiliando os usuários a obter com maior facilidade os dados desejados, o que traz ganhos para a competitividade.

Os algoritmos complexos utilizados na mineração de dados estão ajudando o governo dos EUA no reconhecimento de padrões na análise de fraudes em impostos(UNIVERSIDADE ESTUDUAL DE MARINGA. GRUPO DE SISTEMAS INTELIGENTES, 1999). Essas ferramentas, até então, têm sido de domínio de várias grandes corporações, devido ao alto custo envolvido.

Avanços na coleta de dados científicos (por exemplo, sensores, satélites, processamento de código de barras) têm aumentado em muito o volume de dados, aliados aos avanços na área de armazenamento, com o uso extensivo de sistemas de gerenciamento de banco de dados e tecnologia de *Data Warehousing*. Assim, fatores têm sido combinados para manter a mineração de dados no foco das atenções do processo de decisão comercial. Segundo (UNIVERSIDADE ESTUDUAL DE MARINGA. GRUPO DE SISTEMAS INTELIGENTES, 1999), alguns desses fatores são:

a) valor nulo em grandes banco de dados, ou seja, base de dados inexploráveis;
b) consolidação de bases de dados, incluindo a concepção de *data warehouse*;
c) dramática redução do custo tanto para armazenamento como para processamento. Um *terabyte* de armazenamento custaria US$ 10 milhões em 1990; agora custa menos de US$ 1 milhão;
d) intensa competição de um mercado em crescente saturação;
e) habilidade de direcionar a manufatura, mercado e propaganda para um segmento de clientes; e
f) crescimento acelerado do mercado para produtos de mineração de dados.

2.4.1 Exemplos de aplicação da mineração de dados

As técnicas de mineração de dados são utilizadas quando o objetivo for solucionar problemas que envolvem:

- descoberta de padrões – entendimento de comportamento – planos estratégicos de marketing por tipo de cliente, técnicas de decisão por média, entendimento de tipos de clientes por item de consumo;

- descoberta de funções – relacionamentos numéricos – previsão de vendas, análises de informações efetivas, seleção de prováveis respostas para projeto de marketing; e
- descoberta de regras – relacionamentos expressos em linguagem – análise de marketing, projeto de promoções de vendas.

As empresas varejistas, de finanças, de saúde, de seguro, diariamente mantêm enormes quantidades de dados sobre as atividades de seus clientes. Implicitamente, esses dados revelam padrões de comportamento de seus clientes – comportamentos que podem ajudar na estratégia de marketing, reduzindo seus riscos.

Por exemplo, sabendo-se que 72% de seus clientes que compram uma certa marca de refrigerante podem comprar uma certa marca de batata frita (BERNARDINO,2000), ajudaria o varejista a determinar promoções apropriadas e otimizar o espaço físico de sua análise de afinidade. Pode fazer também com que o varejista decida não dar desconto na batata frita sempre que o refrigerante estiver em liquidação, quando a redução dos lucros mostrar-se desnecessária.

A mineração de dados pode ser aplicada em diversas áreas, com consideráveis benefícios. Veja abaixo alguns casos em que foram aplicados algoritmos de mineração de dados, sendo obtidos excelentes resultados.

Mercados

- Identificar o comportamento de compra dos clientes(BIGUS, 1997).
- Encontrar associação entre as características demográficas dos clientes.
- Prever quais clientes têm potencial para compra.
- Identificar regras comerciais de estoque e de venda de produtos através dos dados históricos de venda.

Bancos

- Detecção de padrões de fraudes no uso de cartões de crédito (BIGUS, 1997).
- Identificação dos clientes "honestos"(DILLY,2000).
- Prévia seleção de clientes para oferta de algum produto (cartão de crédito, seguro, título de capitalização, etc.).
- Encontrar relações ocultas entre diferentes indicadores financeiros (DILLY,2000).
- Ajudar no diagnóstico da saúde financeira das pessoas jurídicas.

Seguro e saúde

- Análise de solicitações – determinar quais procedimentos médicos são solicitados.
- Prever quais clientes adquirirão novas apólices.
- Identificar padrões de comportamento dos clientes de risco.
- Identificar comportamentos fraudulentos.

Transporte

- Determinar a distribuição dos horários entre os vários caminhos.
- Analisar padrões de sobrecarga.

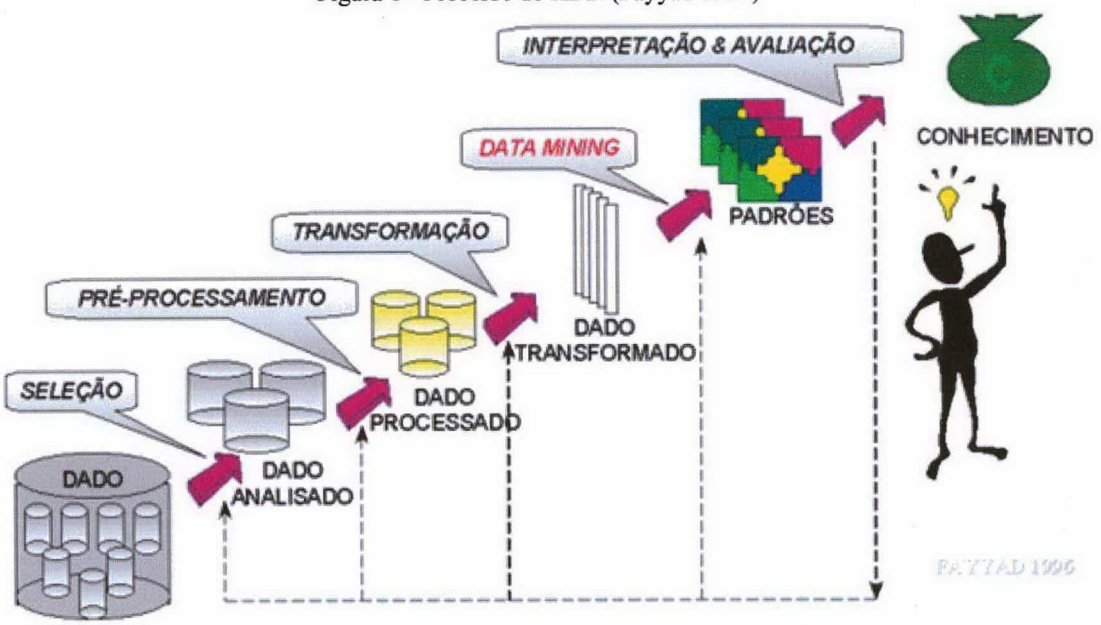
Medicina

- Caracterizar comportamentos de pacientes para prever as visitas ao consultório.
- Identificar terapias médicas de sucesso para diferentes doenças.

2.5 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS(KDD)

Para obtenção do conhecimento, o processo de KDD passa por diversas fases, que serão apresentadas abaixo

Figura 1 - Processo de KDD (Fayyad 1996)



a) **Definição de metas.** Definição do problema a ser resolvido através do processo KDD; Tal fase deve preocupar-se com critérios de desempenho, gargalos no domínio da aplicação e a interoperabilidade com o usuário final (Gonçalves 2000).

b) **Seleção.** É feita uma seleção ou um agrupamento de dados apropriados para análise (BIGOLIN,2000), de acordo com algum critério – por exemplo, todas as pessoas que possuam carro ou todos os clientes do sexo feminino. Dessa maneira, subconjuntos de dados podem ser determinados de acordo com a finalidade da aplicação (BIGOLIN,2000).

c) **Pré-processamento**. Eliminação de ruídos e erros, estabelecimentos de procedimentos para verificação de falta de dados, conversão dos dados para

construção de uma base de dados consistente(Gonçalves 2000). Nesta etapa, inclui-se a limpeza dos dados, ou seja, deve-se eliminar registros repetidos e problemas de tipagem. Essa etapa pode tomar até 80% do tempo necessário para todo o processo (Mannila, 1994) (BIGUS, 1997), devido às conhecidas dificuldades de integração da base de dados heterogênios.

d) **Transformação.** Os dados pré-processados devem ainda passar por uma transformação que os armazena adequadamente, visando facilitar o uso das técnicas de mineração de dados. Nesta fase, o uso de *data warehouse* se expande consideravelmente, já que nessas estruturas as informações estão alocadas de maneira mais eficiente. Em *data warehouse*, os dados não são voláteis, classificados por assunto e de natureza histórica, tendendo, portanto, a se tornarem grandes repositórios de dados extremamente organizados.

e) **Mineração de dados**. É basicamente a aplicação de um algoritmo de descoberta de padrões nos dados (Fayyad, 1997). É necessário escolher o algoritmo que mais se adapte ao objetivo do processo KDD: classificação, clusterização, regras de associação, etc(Figueira, 1999). A busca de padrões no processo KDD pode ser efetuada automaticamente pelo sistema ou interativamente com um analista responsável pela geração de hipóteses. Diversas ferramentas distintas, como redes neurais, indução de árvore de decisão, sistemas baseados em regras e programas estatísticos, tanto isoladamente como combinados, podem ser aplicadas ao problema.

f) **Interpretação/Avaliação.** Pode requerer a repetição de vários passos de iteração com os dados, mas normalmente é encarada como uma simples visualização dos dados. O KDD gera um relatório das descobertas, que passa a ser interpretado pelos analistas de mineração. Somente após a interpretação das

informações obtidas encontra-se conhecimento, que pode então ser utilizado para suporte à tomada de decisão humana.

2.6 TÉCNICAS DE MINERAÇÃO DE DADOS

Existem várias técnicas de mineração de dados, e cada uma delas é mais apropriada para a resolução de um determinado tipo de problema. Podemos citar: Clustering ou Agrupamento, Classificação, Regras de Associação, entre outras. Na seqüência, serão abordadas as técnicas utilizadas na análise que serviu de base para a elaboração deste trabalho.

2.6.1 Regras de associação

Dada uma coleção de itens e um conjunto de registro – cada qual contendo um número de itens da coleção, através de técnicas de associação – são realizadas operações sobre esse conjunto de registro, retornando afinidades entre a coleção de itens.

Tais afinidades são expressas por um conjunto de regras da forma "Um determinado percentual "x" de registros que contém os itens A, B e C também contém os itens D, E e F", onde A, B, C podem corresponder a qualquer número de itens (o mesmo para D, E e F) e onde o percentual "x" é chamado de fator de confiança. O objetivo da técnica é encontrar, a partir de um grande número de transações, tendências que possam ser usadas para entender e explorar padrões de compra.

Algoritmos de associação têm numerosas aplicações, incluindo supermercados, planejamento de estoque, mala direta para marketing direcionado e planejamento de promoções de vendas (BARBIERI, 2001). A regra de associação deriva a partir da mineração de dados de um banco de dados de transações, por exemplo, uma lista contendo

o conjunto de itens comprados pelo consumidor em uma visita a uma loja. A regra de associação poderia ser: "75% dos consumidores que compram coca-cola também compram batata-frita" (UNIVERSIDADE ESTUDUAL DE MARINGA , 1999).

Os algoritmos de associação produzem várias dessas regras, e cabe ao usuário selecionar as regras com maior grau de confiança. Podem existir também regras com múltiplas associações, tais como: "65% dos consumidores que compram coca-cola e batata frita também compram sorvetes" (UNIVERSIDADE ESTUDUAL DE MARINGA , 1999).

Essas regras ajudam o gerente de um supermercado a aumentar os seus lucros. Podemos, por exemplo, incrementar as vendas de sorvetes. O simples fato de conhecer a relação de venda entre vários produtos fornece condições para organizar os produtos nas prateleiras de tal forma que incentive o cliente à compra de todos os itens em que foi descoberta uma relação.

O algoritmo *Apriori* - Geração dos grandes conjuntos

O algoritmo *Apriori* gera um conjunto G dos grandes conjuntos de itens da base dados com suporte maior que um suporte mínimo especificado. De posse desse conjunto G, será possível extrair as regras de associação dos itens.

Cada regra possui dois atributos que determinam sua qualidade no conjunto de dados:

Suporte = (N.° de Registros com X e Y) / (total de transações)

Confiança = (N.° de transações com X e Y) / (N° de transações com X)

Fase 1 do algoritmo *Apriori*: Geração dos grandes conjuntos

1) G_1 = {Conjunto de grandes conjuntos de tamanho 1}

2) Para (k = 2; $G_{k-1} <> \varnothing$; K++ < número de itens) faça

3) início

4) C_k = GeraCandidatos(G_{k-1});

5) Para todas transações t $\in$ D faça

6) início

7) C_t = subconjunto (C_k, t);

8) Para todos os candidatos c $\in$ C_t do

9) c.count++;

10) fim

11) G_k = {c $\in$ C_k | c.count >= minsup}

12) fim

13) ConjuntoResposta = União de todos os $G_{k's}$;

A primeira ação do algoritmo é identificar todos os itens existentes, formando o conjunto G de conjuntos de tamanho igual a 1.

Após a geração do conjunto G_1, deve-se gerar os próximos conjuntos G de conjuntos de tamanho 2, 3, 4, ..., n. Assim, para constituir o conjunto G_2, é preciso gerar todas as combinações possíveis de dois itens, com os itens de G_1.

A partir de G_2, são gerados G_3, G_4, ..., G_k.

Um conjunto C_k é construído de acordo com o algoritmo *GeraCandidatos(G_{k-1})*, descrito a seguir.

A geração dos candidatos: o algoritmo *GeraCandidatos*

A idéia deste algoritmo é unir os elementos de G_{k-1}, 2 a 2 e conservar apenas aqueles em que todos os seus subconjuntos de tamanho k-1 pertençam a G_{k-1}. Este algoritmo se dá em dois passos: junção e poda.

União

Insert into C_k

Select p.item1 (= q.item1) $\cup$ ... $\cup$ p.itemk-2 (= q.itemk-2) $\cup$ p.itemk-1 $\cup$ q.itemk-1

From $G_{k-1}(p)$, G_{k-1} (q) {elementos p e q de G_{k-1}};

Poda

A idéia por trás da poda é eliminar todo *c* pertencente a C_k, tal que algum (k-1) subconjunto de *c* não pertence a G_{k-1}.

$C_k = \{c \in C_k \mid \forall s$, com $|s| = k-1 \subset c$, s pertence $G_{k-1}\}$

Se $s \notin G_{k-1}$ então elimina c de C_k.

Geração das regras de associação

O segundo estágio do algoritmo toma cada um dos grandes conjuntos gerados na fase anterior e verifica se ele tem a mínima confiança especificada.

2.6.2 Agrupamento

O agrupamento é usado para segmentar uma base de dados em subconjuntos (THERLING,1999) – os grupos – como membros de cada grupo compartilhando um número de propriedades semelhantes.

Essa técnica refere-se ao agrupamento de elementos que utilize algum critério que determine as distâncias entre esses elementos e aqueles com distância menor que ficam no mesmo *cluster* (grupo) (UNIVERSIDADE ESTUDUAL DE MARINGA , 1999).

Abordagens de agrupamento direcionam-se a problemas de segmentação e consultam os dados de registros com um grande número de atributos, num conjunto relativamente pequeno de grupos ou segmentos. Esses conjuntos de dados podem ser criados estatisticamente ou por meio do uso de métodos de Redes Neurais Artificiais ou Técnicas de Indução. Esse processo de associação é executado automaticamente por algoritmos de agrupamentos que identificam as características distinguíveis de um conjunto de dados, através de limites de dados encontrados naturalmente. Não existe a necessidade de identificar os grupos de atributos que serão utilizados para segmentar o conjunto de dados.

O agrupamento é freqüentemente utilizado como um dos primeiros passos na análise dos dados feita pela mineração de dados. Ele identifica grupos de registros relacionados, os quais podem representar classes potenciais, e essas classes podem ser usadas como ponto de partida para explorar outros relacionamentos.

No uso de técnicas de agrupamento não se possui um conjunto de dados predefinidos, isto é, esse uso caracteriza-se pelo uso de um método de aprendizado não-supervisionado, no qual existe maior autonomia do algoritmo.

Um exemplo de agrupamento é olhar para um grande número de consumidores inicialmente desconhecidos e tentar ver, se eles possuem um agrupamento natural. Este é um exemplo de Mineração de Dados indireto, onde o usuário não tem uma ferramenta que

retorne um significado estruturado. O algoritmo de agrupamento vai encontrar o melhor particionamento dos registros dos consumidores (neste exemplo) e fornece uma descrição do centróide de cada grupo de dados. Em muitos casos, estes grupos tem uma interpretação óbvia da característica de cada grupo (KIMBALL, 98).

O algoritmo K-Means é um algoritmo não-supervionado, e sua função é gerar *cluster* através da similaridade dos dados. Os principais passos são os seguintes:

Passo 1 – particionar os itens em K *clusters* iniciais;

Passo 2 – percorrer a lista de itens, atribuindo cada item a um *cluster* cujo centróide é o mais próximo; Neste caso, o centróide é escolhido aleatoriamente;

Passo 3 - repetir o passo 2 até que não haja mais atribuições.

3 AVALIAÇÃO INSTITUCIONAL NA UNISUL

3.1 CARACTERIZAÇÃO DA UNISUL

A Universidade do Sul de Santa Catarina (Unisul) é uma fundação pública de direito privado, de caráter comunitário e regional, organizada por transformações da Fundação Educacional do Sul de Santa Catarina (FESSC). Foi reconhecida como Universidade pela portaria ministerial MEC n. 028, de 27 de janeiro de 1989.

Segundo dados do Setor Pessoal, a Unisul conta com 1.742 colaboradores, sendo 1.197 professores e 545 funcionários de áreas administrativas, técnicos e de apoio. O número de alunos é de 19.620 no ano 2000, com previsão de 23.130 em 2002 e de 25.190 em 2004. Em 1999 o faturamento foi de R$ 56 milhões, com previsão de R$ 70 milhões em 2000. Mantém colégio de ensino fundamental e médio, além de cursos de graduação, pós-graduação e extensão.

A Unisul desenvolve atividades nos seguintes campi e locais:

- Campus de Tubarão
  - ✓ Centro de Pós-Graduação (Tubarão)
  - ✓ Colégio Dehon (Tubarão)
  - ✓ Unidade de Imbituba
  - ✓ Unidade de Braço do Norte
  - ✓ Unidade de Laguna

- Campus de Araranguá
  - ✓ Unidade de Içara

- Campus da Grande Florianópolis

- ✓ Cidade Universitária Pedra Branca
- ✓ Ponte do Imaruím (Palhoça)
- ✓ Centro Internacional de Pós-Graduação (Florianópolis)
- ✓ Colégio Catarinense (Florianópolis)
- ✓ Unidade de Jurerê Internacional (Florianópolis)

Alguns outros cursos são oferecidos em cidades (Armazém, Garopaba e Grão-Pará), além dos cursos oferecidos nos campi e respectivas unidades.

Além das unidades de ensino citadas, a Unisul possui algumas outras unidades de negócios, ligadas a uma Fundação:

- Farmácia-escola
- Livraria
- Laboratório de Análises Clínicas
- Centro Tecnológico
- Escritório-Modelo de Advocacia
- Serviços de Psicologia Aplicada
- Serviços de Assistência Integrada à Saúde
- Mecanografia
- TV Educativa
- Rádio Universitária
- Editora Universitária

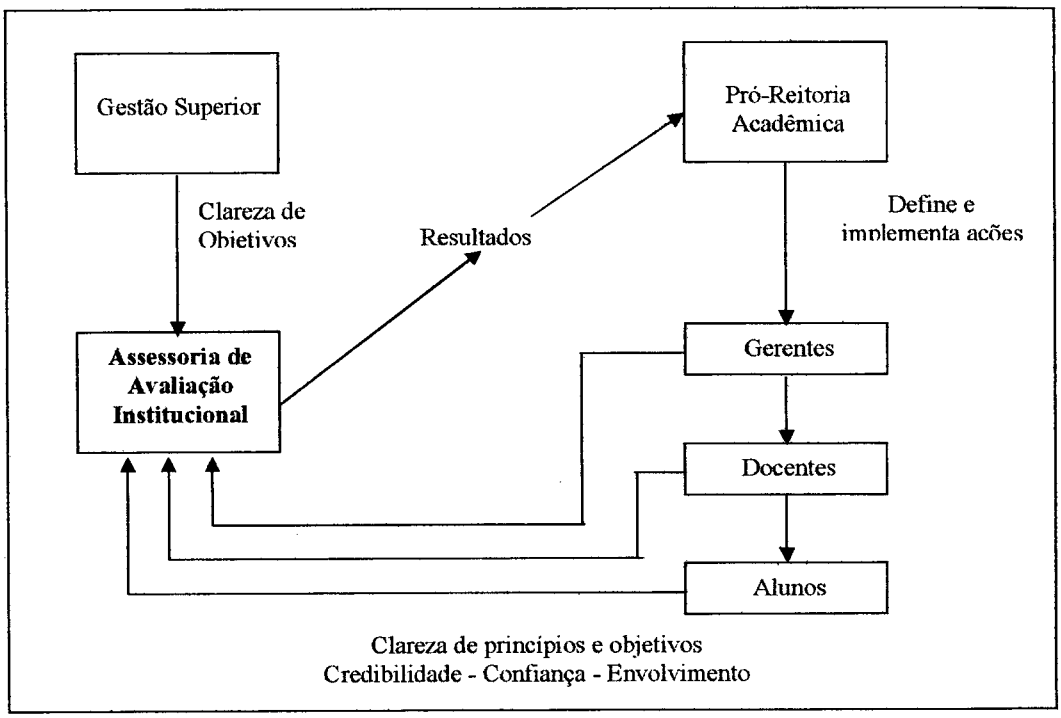
A principal atividade e fonte de renda da Unisul é a graduação, que no ano de 2000 alcançou o número de 19.620 alunos.

3.2 AVALIAÇÃO INSTITUCIONAL

Na busca da qualidade de seus serviços de ensino, pesquisa e extensão, a Unisul criou uma assessoria de Avaliação Institucional, que possui um processo permanente de avaliação.

Estruturalmente, a Avaliação Institucional da Unisul está assim colocada como mostra a figura 2:

Figura 2 - Estrutura da Avaliação Institucional



Fonte: Carvalho et al,1998

O processo de avaliação institucional tem como meta obter informações que venham a auxiliar o corpo dirigente na tomada de decisão, tendo como parâmetro informações colhidas no meio acadêmico (Carvalho et al., 1998).

Os dados colhidos para análise abrangem aspectos culturais, pedagógicos, econômicos e sociais.

Os aspectos pedagógicos referem-se à avaliação do corpo docente, dos gestores dos cursos e centros de ensino, e da satisfação dos alunos como consumidores.

Os aspectos econômicos referem-se ao aproveitamento de recursos humanos e de infra-estrutura disponível.

Os aspectos culturais referem-se à integração da Universidade dentro da comunidade.

Os aspectos sociais referem-se à formação de agentes de mudança e a contribuições para o ser humano e para a sociedade em que vive.

Para abranger todos esses aspectos, a Assessoria de Avaliação Institucional aplica um questionário com 65 questões(Anexo A), divididos em alguns grupos conforme segue abaixo:

- Quanto ao curso da Unisul
- Quanto ao coordenador do curso no seu campus
- Quanto ao seu desempenho como aluno
- Questões adicionais

As respostas para as questões pode assumir os valores de 1 a 5 (1 – Sempre/Sim, 2 – Quase sempre/Praticamente sim, 3 Raramente/Praticamente não, 4 – Nunca/Não, 5 Sem resposta/Não se aplica).

Para as questões adicionais os valores assumidos são os seguintes:
Idade: idade do alunos (1 – para idade até 20 anos, 2 – para idade entre 21 e 25 anos, 3 – para idade entre 26 e 30 anos, 4 – para idade acima de 30 anos).

Curso: código do curso do aluno.

Curso - Disc: código do curso da disciplina freqüentada pelo aluno.

Campus: Campus do aluno (1 – Tubarão, 2 – Araranguá, 3 – Palhoça).

Cada aluno recebe um formulário com este questionário para cada disciplina matriculada naquele semestre corrente. Depois de respondidos, esses formulários são lidos

através de um equipamento de leitura ótica, o qual fornece um arquivo-texto com as respostas de todos os formulários separados por campus. Depois da obtenção dos dados, estes precisam ser minerados e analisados para que possam ser úteis à tomada de decisão.

3.3 ANÁLISES JÁ REALIZADA PELA ASSESSORIA DE AVALIAÇÃO INSTITUCIONAL

A Assessoria de Avaliação Institucional realiza anualmente análises estatísticas multivariadas com os dados coletados dos alunos. Seu objetivo é avaliar individualmente cada professor, coordenador, curso e também os dados gerais da Unisul.

Os professores são avaliados individualmente, em cada disciplina e campus que lecionam, e são comparados com a média geral do curso e da Unisul. Essa análise permite que cada professor possa situar-se no contexto da instituição, no que se refere ao ensino.

Os indicadores para os professores foram agrupados na seguinte ordem:

Q1 – Produtividade (32 a 36)

Q2 – Conteúdo (37, 38, 39 e 45)

Q3 – Exigência (40, 41, 42, 49 e 50)

Q4 – Método (43, 44 e 52)

Q5 – Avaliação da Aprendizagem (45 a 48)

Q6 – Ética (58 a 60)

Q7 – Relação Professor/Aluno (51 e 53)

Q8 – Relação Professor/ Curso (54 e 55)

Q9 – Relação Professor/Unisul (56 e 57)

Os coordenadores também recebem a sua avaliação individual e a média geral da Unisul, com o mesmo objetivo de sua auto-avaliação.

Essa mesma análise também é feita com dados referentes ao curso, para que, através de uma comparação com dados gerais da Unisul, tanto o coordenador como os demais dirigentes possam avaliar a situação de cada curso.

4 BUSCA DE INFORMAÇÕES GERENCIAIS

4.1 DESCRIÇÃO DO PROBLEMA

Como se pôde observar, o processo de avaliação institucional gera uma grande quantidade de dados que precisam de um suporte tecnológico para serem aproveitados na tomada de decisão. Em conformidade com o objetivo deste trabalho, no presente capítulo serão utilizadas algumas técnicas de descoberta de informação para verificar a ocorrência de grupos de dados similares, ou seja, para identificar algum padrão de comportamento. Essas informações podem ser de grande valia na tomada de decisão, de forma a buscar uma melhor qualidade dos serviços prestados como também se elaborarem estratégias de conquista de novos clientes.

A metodologia utilizada neste trabalho está baseada no processo de descoberta do conhecimento (KDD) e envolve as etapas descritas a seguir.

- **Seleção dos dados.** A partir de um arquivo com as respostas das 65 questões aplicadas aos alunos no semestre 2000/A, foram selecionadas algumas variáveis com a ajuda da avaliação realizada pela Assessoria de Avaliação Institucional. Essas variáveis foram:

  Idade: Idade do aluno.

  - Qualidade-curso (Q1): considera que seu curso tem qualidade?
  - Conhece-curso (Q3): conhece o projeto do seu curso na sua totalidade?
  - Exigência (Q4): o seu curso está exigindo o suficiente para a sua formação profissional?
  - Qualidade-Unisul (Q5): considera a Unisul uma universidade de qualidade?
  - Imagem-Unisul (Q6): você tem orgulho de estudar na Unisul?
  - Imagem-curso (Q7): seu curso tem uma boa imagem na sociedade?

- Ambiente (Q20): o ambiente do campus é próprio e propício ao desenvolvimento das atividades acadêmicas?
- Gerencia-curso (Q27): o coordenador do seu curso revela ser um líder democrático e entusiasmado pelas questões do curso?
- Imagem-coord (Q31): indicaria seu coordenador como modelo de atuação? Produtividade (Q36): está desenvolvendo o plano de ensino?

- **Pré-processamento.** Foi criado para cada campus em estudo (Tubarão, Palhoça e Araranguá) um arquivo com as informações referentes às variáveis selecionadas.

- **Limpeza.** Durante o processo de leitura dos dados dos formulários, foram geradas muitas linhas de dados nulos. Foi necessário realizar um trabalho de identificação e de eliminação dessas linhas.

- **Transformação.** Todas as respostas das perguntas já foram normalizadas no processo de leitura dos formulários (1 – Sempre/Sim, 2 – Quase sempre/Praticamente sim, 3 – Raramente/Praticamente não, 4 – Nunca/Não, 5 – Sem resposta/Não se aplica).

- **Mineração de dados**. Foram aplicadas três técnicas de mineração de dados:
  - estatística descritiva;
  - *clustering*, utilizando o algoritmo de *K-means*;
  - regras de associação, utilizando o algoritmo *Apriori.*

4.2 ESTATÍSTICAS

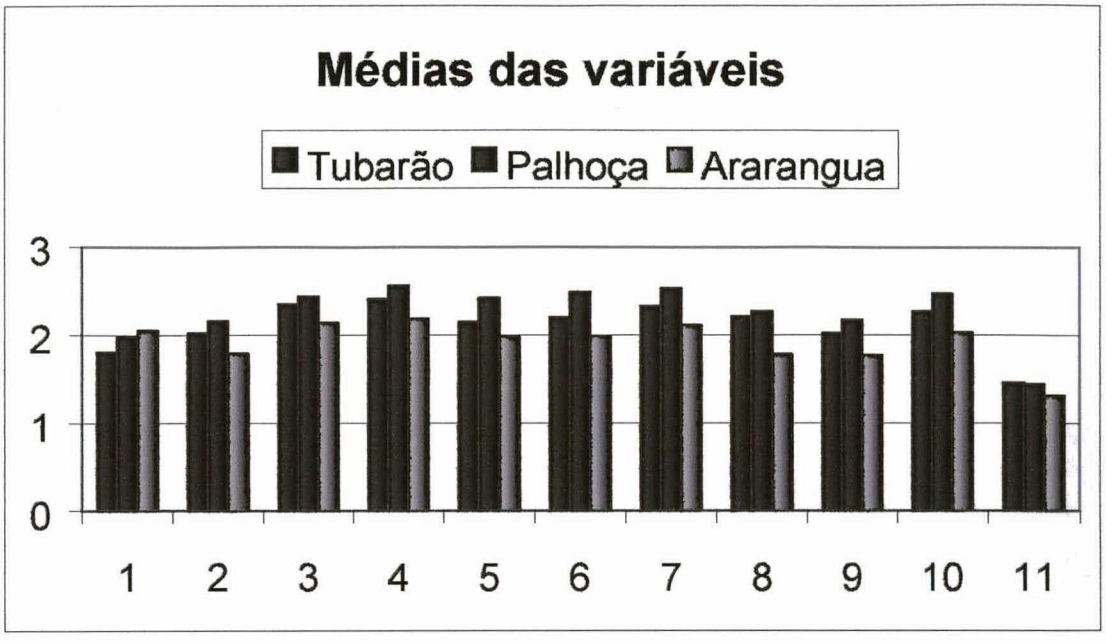
Antes de se aplicarem as técnicas de mineração de dados, foram realizadas análises estatísticas, as quais são apresentadas abaixo. Através dessas análises é possível compreender os dados utilizados nas técnicas de mineração de dados, como também ajudar na interpretação dos resultados apresentados pelas técnicas mencionadas.

4.2.1 Medidas de tendência central e de dispersão

Tabela 1 - Medidas de tendência central e de dispersão

		Média			Desvio-padrão			Variância		
	Variáveis	Tub	Pal	Ara	Tub	Pal	Ara	Tub	Pal	Ara
1	Idade	1,81	1,98	2,05	0,99	1,06	1,12	0,99	1,115	1,25
2	Qualidade-curso(Q1)	2,02	2,16	1,79	0,87	0,88	0,9	0,76	0,782	0,8
3	Conhece-curso(Q3)	2,35	2,44	2,14	0,99	1,01	0,94	0,98	1,014	0,89
4	Exigência(Q4)	2,41	2,56	2,18	1	1,02	1	0,99	1,033	1
5	Qualidade-Unisul(Q5)	2,15	2,42	1,98	0,97	1	0,98	0,94	0,993	0,96
6	Imagem-Unisul(Q6)	2,2	2,49	1,99	1,07	1,11	1,07	1,16	1,226	1,15
7	Imagem-curso(Q7)	2,33	2,53	2,11	1,15	1,17	1,12	1,32	1,37	1,24
8	Ambiente(Q20)	2,21	2,27	1,78	1,09	1,09	0,94	1,18	1,184	0,89
9	Gerencia-curso(Q27)	2,02	2,17	1,77	1,12	1,17	1,02	1,26	1,366	1,05
10	Imagem-Coord(Q31)	2,27	2,47	2,03	1,24	1,26	1,2	1,55	1,592	1,45
11	Produtividade(36)	1,46	1,44	1,31	0,93	0,88	0,75	0,87	0,767	0,57
	Total de dados	16384	12194	2879	16384	12194	2879	16384	12194	2879

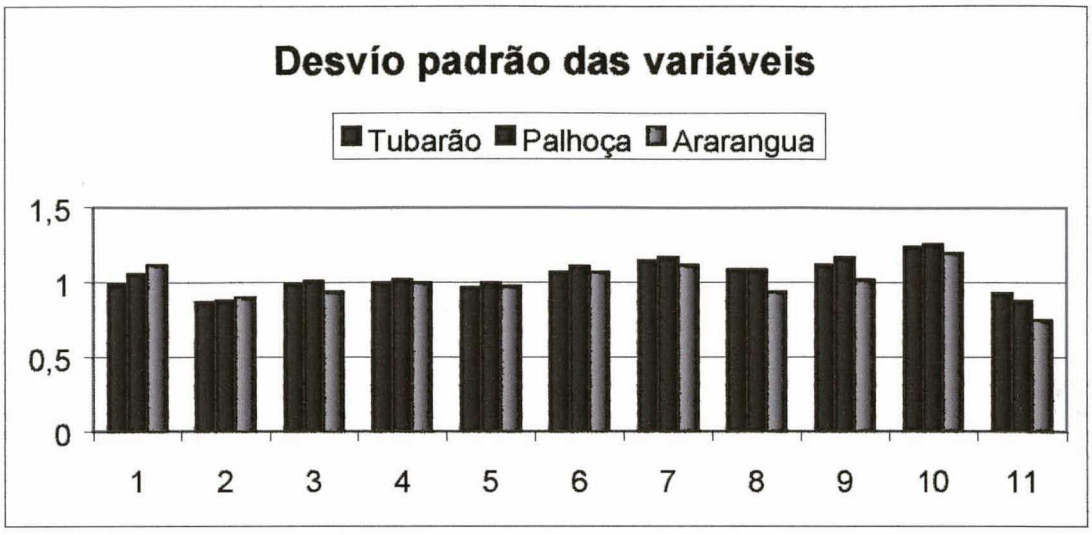
Gráfico 1 - Média de todas as variáveis



Sem considerar a variável *Idade*, pode-se observar que as médias no campus de Araranguá são sempre inferiores às médias dos campi de Tubarão e da Palhoça. Na variável *Idade* acontece o contrário: os alunos do campus de Araranguá possuem uma média maior. Parece existir uma tendência neste campus da variável *Idade* com relação às demais variáveis. Essa tendência não se justifica quando se comparam os campi de Tubarão e da Palhoça. Na variável *Idade*, a média no campus de Tubarão é menor que a do campus da Palhoça. No entanto, todas as outras médias (exceto a variável *Produtividade*) no campus de Tubarão são inferiores às do campus da Palhoça.

Com respeito à dispersão, existe uma similaridade entre os diferentes campi que não possibilita observar nenhuma tendência. Em algumas variáveis parece existir uma diferença maior, como é o caso da variável *Idade*, em que a variabilidade do campus de Araranguá é superior à dos outros campi e nas variáveis *Ambiente*, *Gerencia-curso* e *Produtividade*, em que a variabilidade do campus de Araranguá é inferior.

Gráfico 2 - Desvio-Padrão de todas as variáveis



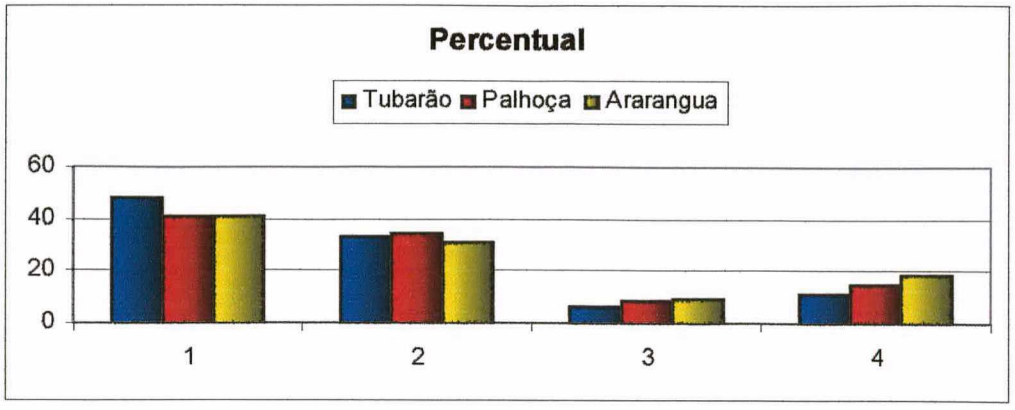
4.2.2 Tabelas de freqüência e histogramas

Idade

Tabela 2 - Freqüência e percentual da variável Idade por faixa de idade nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1 (<20)	7960	5039	1188	48,6	41,3	41,3	48,6	41,3	41,3
2 (21-25)	5441	4234	888	33,2	34,7	30,8	81,8	76,0	72,1
3 (26-30)	1053	1040	268	6,4	8,5	9,3	88,2	84,5	81,4
4 (>30)	1930	1881	535	11,8	15,4	18,6	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 3 - Percentual de faixa de idade nos campi



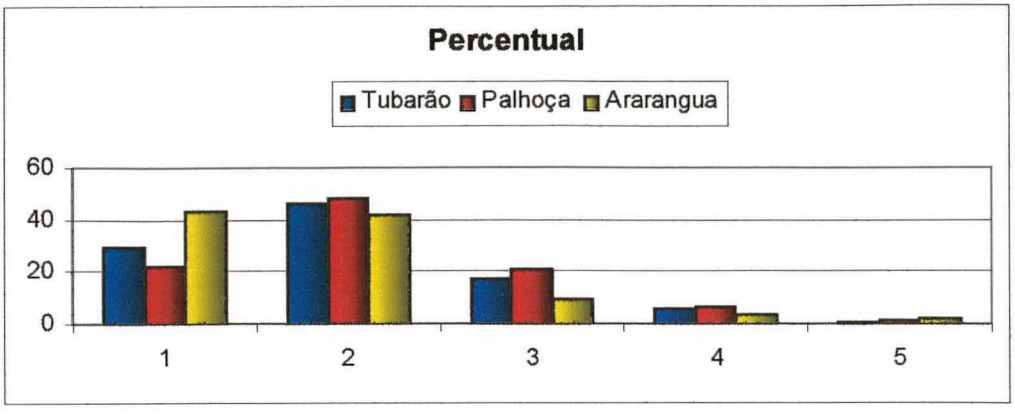
Observa-se na variável *Idade* uma diferença entre os campi de Tubarão e de Araranguá. Existe uma predominância de alunos mais novos em Tubarão e mais velhos em Araranguá.

Qualidade-curso (Q1)

Tabela 3 - Freqüência e percentual das respostas à variável Qualidade-curso nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	4840	2721	1240	29,5	22,3	43,1	29,5	22,3	43,1
2	7617	5935	1209	46,5	48,7	42	76	71	85,1
3	2876	2581	270	17,6	21,2	9,4	93,6	92,2	94,4
4	930	817	103	5,7	6,7	3,6	99,3	98,9	98
5	121	140	57	0,7	1,1	2	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 4 - Percentual de respostas à variável Qualidade-curso nos campi



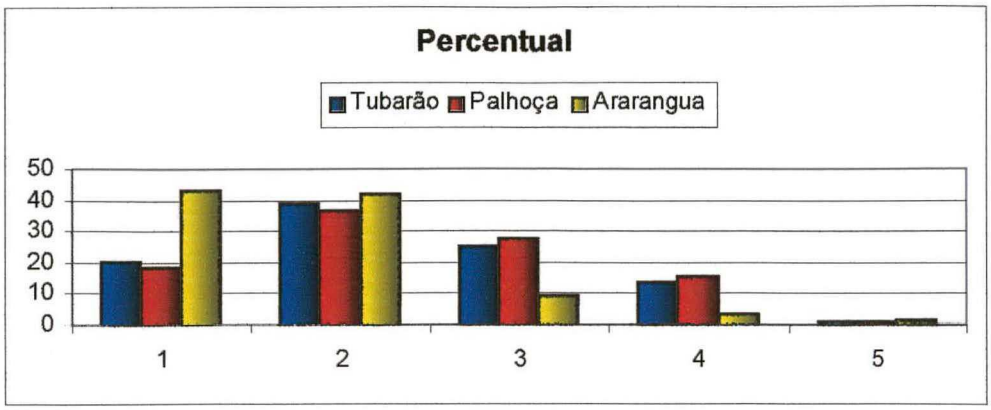
Com relação à variável *Qualidade-curso*, existe uma grande diferença entre o campus de Araranguá e o da Palhoça no que tange às respostas afirmativas. Pode-se observar que, nas respostas que correspondem aos itens 1 e 2 (ambas afirmativas), o campus de Araranguá possui uma avaliação da qualidade do curso significativamente melhor que o da Palhoça.

Conhece-curso (Q3)

Tabela 4 - Freqüência e percentual das respostas à variável Conhece-curso nos campi

	Freqüência			**Percentual**			**Percentual Cumulativo**		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	3378	2291	1240	20,6	18,8	43,1	20,6	18,8	43,1
2	6411	4479	1209	39,1	36,7	42	59,7	55,5	85,1
3	4180	3377	270	25,5	27,7	9,4	85,3	83,2	94,4
4	2251	1875	103	13,7	15,4	3,6	99	98,6	98
5	164	172	57	1	1,4	2	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 5 - Percentual de respostas à variável Conhece-curso nos campi



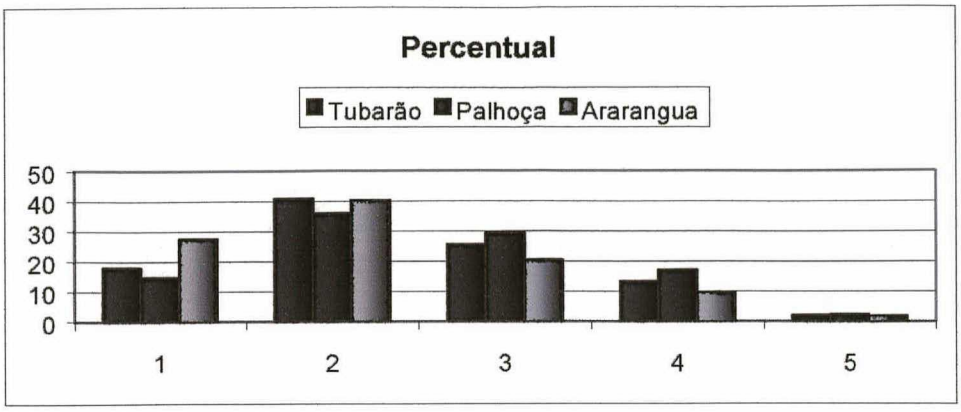
Com relação à variável *Conhece-curso*, a diferença nas respostas afirmativas do campus de Araranguá comparada aos demais campi é bastante acentuada, enquanto que as respostas negativas são encontradas com maior freqüência no campus da Palhoça.

Exigência (Q4)

Tabela 5 - Freqüência e percentual das respostas à variável Exigência nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	2952	1822	796	18	14,9	27,6	18	14,9	27,6
2	6703	4395	1164	40,9	36	40,4	58,9	51	68,1
3	4195	3583	589	25,6	29,4	20,5	84,5	80,4	88,5
4	2197	2103	278	13,4	17,2	9,7	97,9	97,6	98,2
5	337	291	52	2,1	2,4	1,8	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 6 - Percentual de respostas à variável Exigência nos campi



Ao se analisar a variável *Exigência*, a diferença do campus de Araranguá comparada aos demais campi é acentuada, apresentando respostas que correspondem ao item 1, que se refere às respostas afirmativas, ou seja, que possui exigência. Observa-se nesta variável que o percentual de resposta do item 1 diminui muito no que tange às variáveis anteriores no campus de Araranguá. No campus da Palhoça, o número de respostas negativas (itens 3 e 4) continua superior aos outros campi.

As mesmas características apresentadas na variável *Exigência* também são encontradas nas variáveis *Qualidade-Unisul, Imagem-Unisul* e *Imagem Curso*, conforme apresentados nos gráficos e nas tabelas abaixo.

Qualidade-Unisul (Q5)

Tabela 6 - Freqüência e percentual das respostas à variável Qualidade-Unisul nos campi

	Freqüência			**Percentual**			**Percentual Cumulativo**		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	4349	2123	1019	26,5	17,4	35,4	26,5	17,4	35,4
2	7224	5018	1224	44,1	41,2	42,5	70,6	58,6	77,9
3	2999	3162	358	18,3	25,9	12,4	88,9	84,5	90,3
4	1567	1612	225	9,6	13,2	7,8	98,5	97,7	98,2
5	245	279	53	1,5	2,3	1,8	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 7 - Percentual de respostas à variável Qualidade-Unisul nos campi

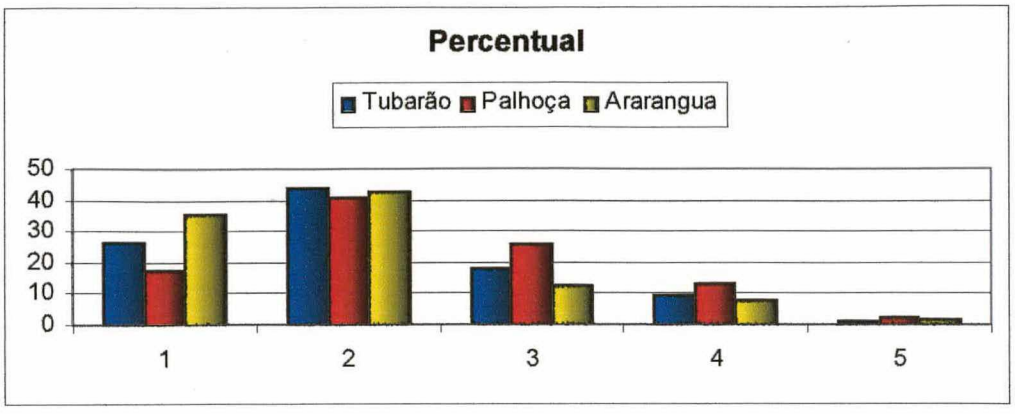


Imagem-Unisul (Q6)

Tabela 7 - Freqüência e percentual das respostas à variável Imagem-Unisul nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	4950	2497	1182	30,2	20,5	41,1	30,2	20,5	41,1
2	6005	4179	972	36,7	34,3	33,8	66,9	54,7	74,8
3	3128	3028	363	19,1	24,8	12,6	86	79,6	87,4
4	1843	2008	297	11,2	16,5	10,3	97,2	96	97,7
5	458	482	65	2,8	4	2,3	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 8 - Percentual de respostas à variável Imagem-Unisul nos campi

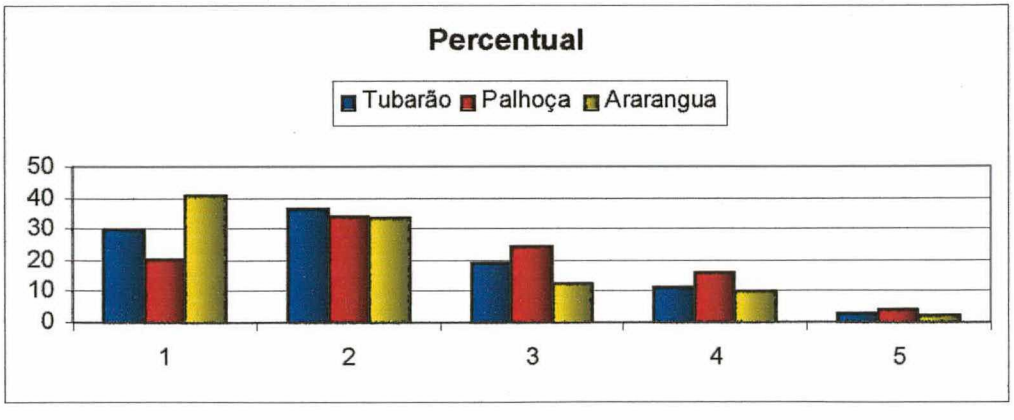
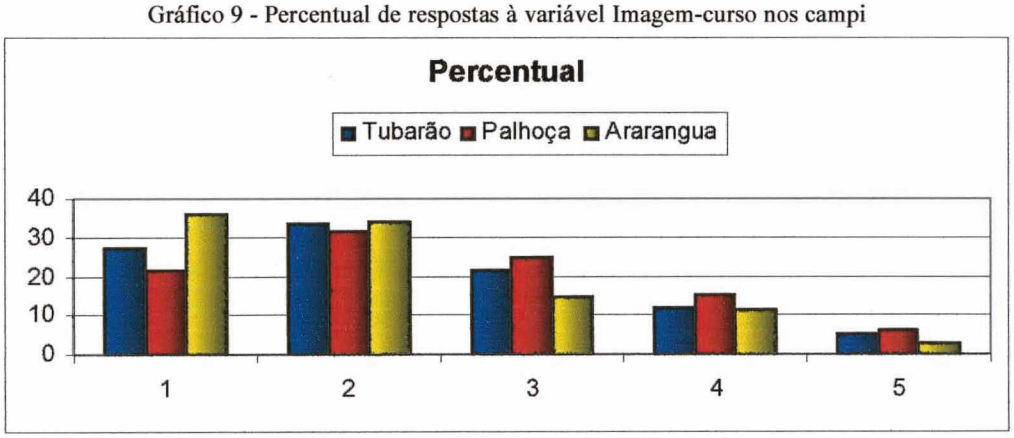


Imagem-curso (Q7)

Tabela 8 - Freqüência e percentual das respostas à variável Imagem-curso nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	4523	2661	1041	27,6	21,8	36,2	27,6	21,8	36,2
2	5498	3855	986	33,6	31,6	34,2	61,2	53,4	70,4
3	3579	3033	424	21,8	24,9	14,7	83	78,3	85,1
4	1949	1883	340	11,9	15,4	11,8	94,9	93,8	96,9
5	835	762	88	5,1	6,2	3,1	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 9 - Percentual de respostas à variável Imagem-curso nos campi

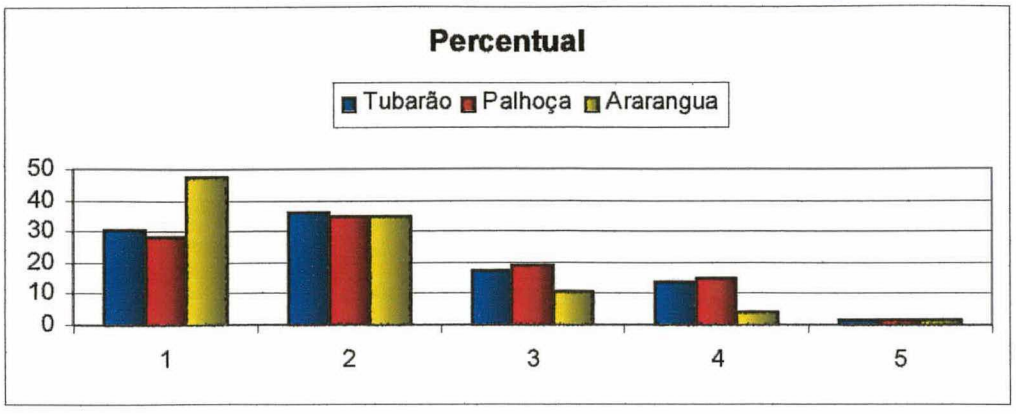


Ambiente (Q20)

Tabela 9 - Freqüência e percentual das respostas à variável Ambiente nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	5032	3463	1371	30,7	28,4	47,6	30,7	28,4	47,6
2	5913	4291	1012	36,1	35,2	35,2	66,8	63,6	82,8
3	2812	2379	314	17,2	19,5	10,9	84	83,1	93,7
4	2276	1815	127	13,9	14,9	4,4	97,9	98	98,1
5	351	246	55	2,1	2	1,9	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 10 - Percentual de respostas à variável Ambiente nos campi



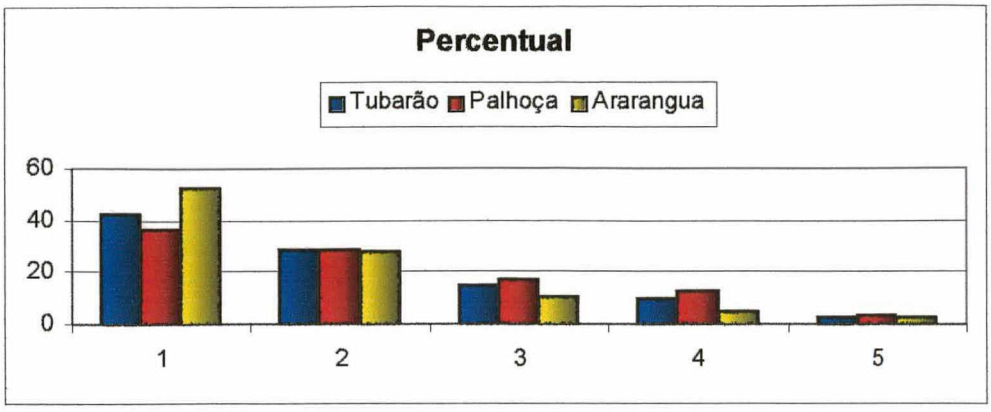
Com relação à variável *Ambiente*, a diferença do campus de Araranguá com respeito aos outros campi é muito acentuada, com respostas que correspondem ao item 1. Observa-se, neste item, a grande diferença entre as respostas afirmativas, enquanto Tubarão e Araranguá estão um pouco abaixo e possuem respostas bem semelhantes.

Gerência-curso (Q27)

Tabela 10 - Freqüência e percentual das respostas à variável Gerencia-curso nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	6992	4508	1521	42,7	37	52,8	42,7	37	52,8
2	4763	3566	808	29,1	29,2	28,1	71,7	66,2	80,9
3	2481	2091	317	15,1	17,1	11	86,9	83,4	91,9
4	1633	1568	154	10	12,9	5,3	96,9	96,2	97,3
5	515	461	79	3,1	3,8	2,7	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 11 - Percentual de respostas à variável Gerencia-curso nos campi



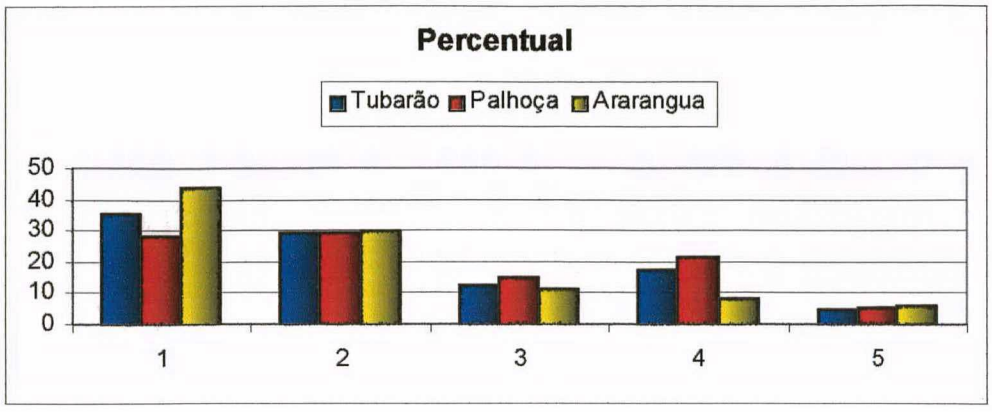
No tocante à variável *Gerencia-curso*, e logo abaixo à *Imagem do coordenador*, as mesmas diferenças entre os campi continuam existindo. Araranguá apresenta um maior número de respostas afirmativas.

Imagem-Coord (Q31)

Tabela 11 - Freqüência e percentual das respostas à variável Imagem-Coord. nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	5835	3474	1259	35,6	28,5	43,7	35,6	28,5	43,7
2	4795	3574	860	29,3	29,3	29,9	64,9	57,8	73,6
3	2097	1827	338	12,8	15	11,7	77,7	72,8	85,3
4	2861	2621	245	17,5	21,5	8,5	95,1	94,3	93,9
5	796	698	177	4,9	5,7	6,1	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 12 - Percentual de respostas à variável Imagem-Coord nos campi

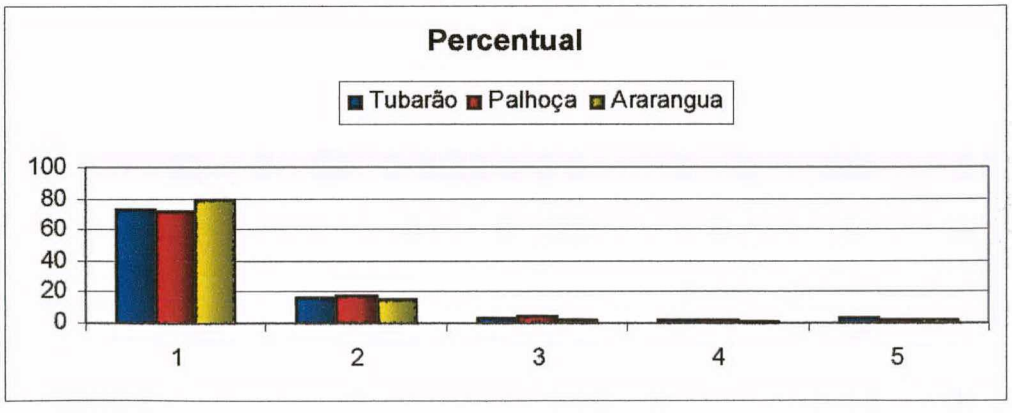


Produtividade (Q36)

Tabela 12 - Freqüência e percentual das respostas à variável Produtividade nos campi

	Freqüência			Percentual			Percentual Cumulativo		
	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang	Tubarão	Palhoça	Ararang
1	11956	8879	2281	73	72,8	79,2	73	72,8	79,2
2	2845	2179	450	17,4	17,9	15,6	90,3	90,7	94,9
3	656	569	59	4	4,7	2	94,3	95,4	96,9
4	382	264	29	2,3	2,2	1	96,7	97,5	97,9
5	545	303	60	3,3	2,5	2,1	100	100	100
Total	16384	12194	2879	100	100	100			

Gráfico 13 - Percentual de respostas à variável Produtividade nos campi



Com relação à variável *Produtividade*, observa-se o grande percentual de respostas afirmativas no tocante às outras variáveis.

4.2.3 Grupos de variáveis

As variáveis foram agrupadas para se fazer uma análise do percentual das respostas afirmativas (itens 1 e 2). O agrupamento foi feito com o auxílio da avaliação institucional, considerando as questões que abordam um mesmo tópico.

Tabela 13 - Percentual das respostas 1 e 1+2(1 – Sempre/Sim, 2 – Quase sempre/Praticamente sim) nos grupos das variáveis Curso, Imagem, Coordenador e Professor

	Curso								**Unisul**						**Coordenador e professor**					
	Q1		Q3		Q4		Q7		Q5		Q6		Q20		Q27		Q31		Q36	
	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2
Tubarão	29,5	76	20,6	59,7	18	58,9	27,6	61,2	26,5	70,6	30,2	66,9	30,7	66,8	42,7	71,7	35,6	64,9	73	90,3
Palhoça	22,3	71	18,8	55,5	14,9	51	21,8	53,4	17,4	58,6	20,5	54,7	28,4	63,6	37	66,2	28,5	57,8	72,8	90,7
Araranguá	43,1	85,1	43,1	85,1	27,6	68,1	36,2	70,4	35,4	77,9	41,1	74,8	47,6	82,8	52,8	80,9	43,7	73,6	79,2	94,9

Com relação às perguntas correspondentes ao *Curso*, observa-se uma diferença muito grande entre as repostas do item 1, dos campi de Araranguá e da Palhoça. A avaliação do curso é positiva em todos os campi, existindo como ponto mais fraco a variável Q4 (Exigência).

No tocante às perguntas referentes à *Unisul*, a análise é similar à anterior, notando-se que um índice muito baixo de alunos da Palhoça sentem orgulho da instituição (Q5) e também um baixo índice de alunos têm boa imagem da Unisul (Q6) quando comparados ao campus de Araranguá.

Já na avaliação de coordenadores e professores há um registro maior dos índices de aceitação.

Tabela 14 - Percentual das respostas 1 e 1+2 nos grupos das variáveis Imagem, Qualidade e Aula e serviços

	Imagem						Qualidade						Aula e serviços							
	Q6		Q7		Q31		Q1		Q5		Q27		Q3		Q4		Q20		Q36	
	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2	1	1+2
Tubarão	30,2	66,9	27,6	61,2	35,6	64,9	29,5	76	26,5	70,6	42,7	71,7	20,6	59,7	18	58,9	30,7	66,8	73	90,3
Palhoça	20,5	54,7	21,8	53,4	28,5	57,8	22,3	71	17,4	58,6	37	66,2	18,8	55,5	14,9	51	28,4	63,6	72,8	90,7
Araranguá	41,1	74,8	36,2	70,4	43,7	73,6	43,1	85,1	35,4	77,9	52,8	80,9	43,1	85,1	27,6	68,1	47,6	82,8	79,2	94,9

Com referência à Imagem, pode-se observar novamente uma grande diferença entre os campi de Araranguá e da Palhoça. Porém, de maneira geral, a *Imagem* está boa em todos os itens analisados. O campus da Palhoça apresenta uma imagem inferior à dos demais campi.

Quanto à *Qualidade*, observa-se uma fortíssima relação com a *Imagem*. Os valores daquela são praticamente um espelho dos valores desta.

Com respeito às variáveis de aula e de serviços, pode-se observar que a produtividade está bem avaliada em todos os campi e, conforme já foi citado anteriormente, o item *Exigência* é o mais deficitário.

4.2.4 Matriz de correlação

Através das Tabelas 14,15,16 e também dos Gráficos 14,15,16 apresentados abaixo, foi analisada a correlação, ou seja, o grau de similaridade de uma variável em relação às outras.

Tabela 15 - Correlação de todas as variáveis do campus de Tubarão

Tubarão		Idade	Q1	Q3	Q4	Q5	Q6	Q7	Q20	Q27	Q31	Q36
Correlação	Idade	1	0,052	0,009	0,003	0,006	-0,041	0,02	0,083	-0,042	-0,026	-0,025
	Q1	0,052	1	0,305	0,574	0,577	0,481	0,426	0,324	0,251	0,263	0,133
	Q3	0,009	0,305	1	0,397	0,298	0,259	0,268	0,21	0,207	0,208	0,099
	Q4	0,003	0,574	0,397	1	0,537	0,437	0,399	0,307	0,246	0,281	0,138
	Q5	0,006	0,577	0,298	0,537	1	0,69	0,438	0,369	0,244	0,263	0,121
	Q6	-0,041	0,481	0,259	0,437	0,69	1	0,434	0,336	0,257	0,273	0,105
	Q7	0,02	0,426	0,268	0,399	0,438	0,434	1	0,257	0,202	0,207	0,098
	Q20	0,083	0,324	0,21	0,307	0,369	0,336	0,257	1	0,177	0,204	0,088
	Q27	-0,042	0,251	0,207	0,246	0,244	0,257	0,202	0,177	1	0,726	0,11
	Q31	-0,026	0,263	0,208	0,281	0,263	0,273	0,207	0,204	0,726	1	0,111
	Q36	-0,025	0,133	0,099	0,138	0,121	0,105	0,098	0,088	0,11	0,111	1

Gráfico 14 - Correlação de todas as variáveis do campus de Tubarão

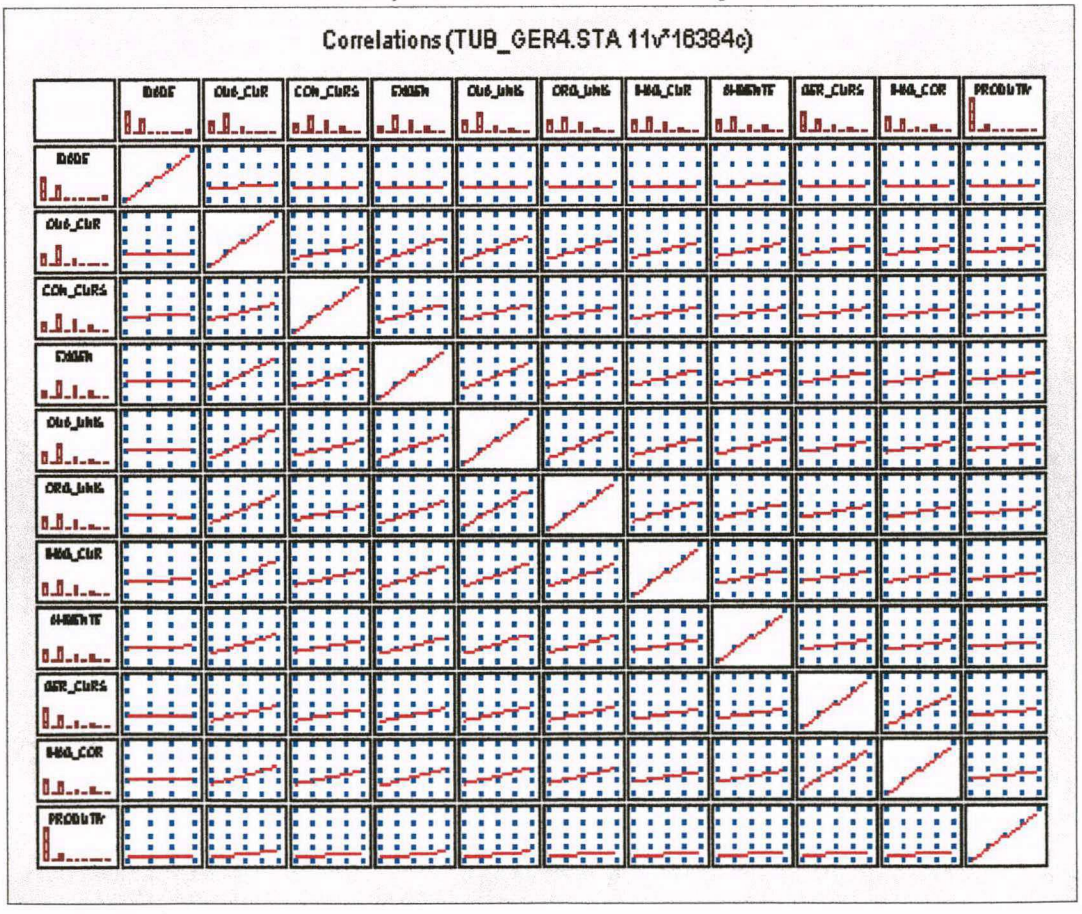


Tabela 16 - Correlação de todas as variáveis do campus da Palhoça

Palhoça	Idade	Q1	Q3	Q4	Q5	Q6	Q7	Q20	Q27	Q31	Q36
Correlation	Idade	1	0	-0,03	-0,05	-0,04	-0,12	-0,02	0,02	-0,06	-0,05	-0,06
	Q1	0,001	1	0,288	0,589	0,608	0,473	0,421	0,28	0,3	0,305	0,162
	Q3	-0,03	0,29	1	0,374	0,28	0,243	0,25	0,15	0,216	0,204	0,106
	Q4	-0,05	0,59	0,374	1	0,553	0,433	0,407	0,26	0,308	0,311	0,164
	Q5	-0,04	0,61	0,28	0,553	1	0,664	0,441	0,35	0,281	0,288	0,123
	Q6	-0,12	0,47	0,243	0,433	0,664	1	0,453	0,32	0,301	0,306	0,129
	Q7	-0,02	0,42	0,25	0,407	0,441	0,453	1	0,22	0,229	0,23	0,114
	Q20	0,023	0,28	0,147	0,258	0,346	0,32	0,223	1	0,203	0,222	0,082
	Q27	-0,06	0,3	0,216	0,308	0,281	0,301	0,229	0,2	1	0,735	0,134
	Q31	-0,05	0,31	0,204	0,311	0,288	0,306	0,23	0,22	0,735	1	0,133
	Q36	-0,06	0,16	0,106	0,164	0,123	0,129	0,114	0,08	0,134	0,133	1

Gráfico 15 - Correlação de todas as variáveis do campus da Palhoça

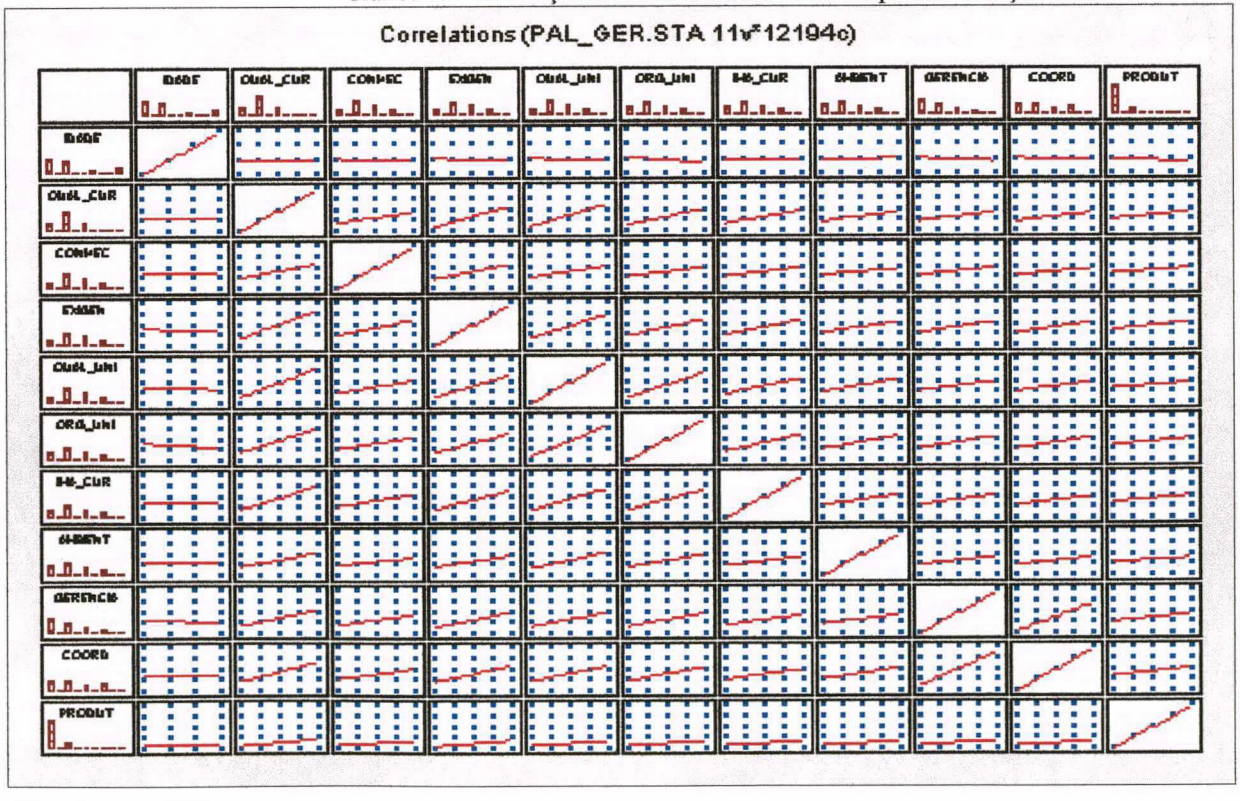
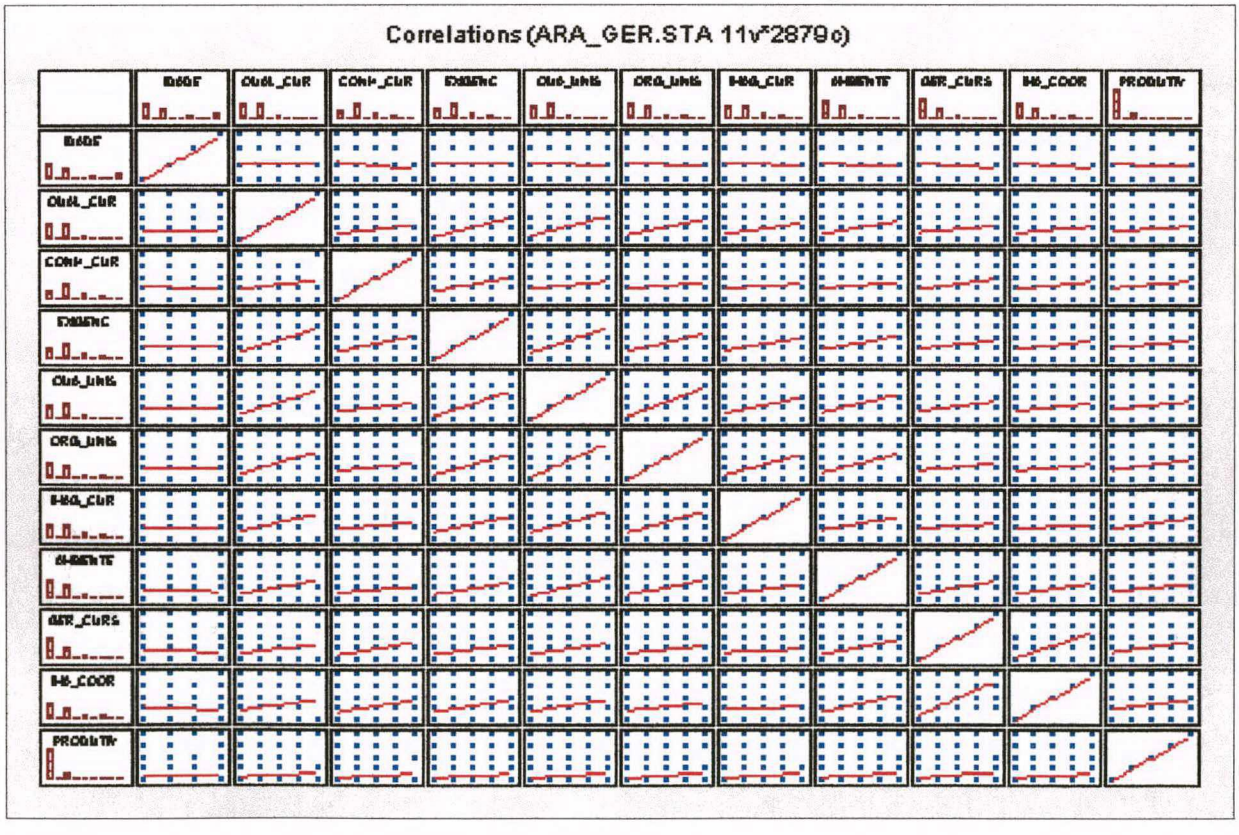


Tabela 17 - Correlação de todas as variáveis do campus de Araranguá

Araranguá	Idade	Q1	Q3	Q4	Q5	Q6	Q7	Q20	Q27	Q31	Q36
Idade	1	0,008	-0,109	0,003	-0,024	-0,038	-0,028	-0,055	-0,077	-0,105	-0,027
Q1	0,008	1	0,201	0,495	0,494	0,424	0,316	0,3	0,174	0,192	0,099
Q3	-0,109	0,201	1	0,334	0,174	0,166	0,125	0,138	0,264	0,226	0,116
Q4	0,003	0,495	0,334	1	0,583	0,456	0,327	0,302	0,196	0,222	0,113
Q5	-0,024	0,494	0,174	0,583	1	0,72	0,409	0,42	0,224	0,222	0,14
Q6	-0,038	0,424	0,166	0,456	0,72	1	0,478	0,396	0,171	0,169	0,164
Q7	-0,028	0,316	0,125	0,327	0,409	0,478	1	0,243	0,086	0,06	0,148
Q20	-0,055	0,3	0,138	0,302	0,42	0,396	0,243	1	0,324	0,306	0,13
Q27	-0,077	0,174	0,264	0,196	0,224	0,171	0,086	0,324	1	0,682	0,17
Q31	-0,105	0,192	0,226	0,222	0,222	0,169	0,06	0,306	0,682	1	0,14
Q36	-0,027	0,099	0,116	0,113	0,14	0,164	0,148	0,13	0,17	0,14	1

Correlation

Gráfico 16 - Correlação de todas as variáveis do campus de Araranguá



Pode-se observar, na matriz de correlação, a forte relação direta entre as variáveis *Qualidade-curso* (Q1), *Exigência-curso* (Q4), *Qualidade-Unisul* (Q5) e *Imagem-Unisul* (Q6). Observa-se outra forte correlação direta entre as variáveis *Gerencia-curso* (Q27) e *Imagem-Coord* (Q31). As relações são muito similares entre os três campi. Outra observação obtida das tabelas de correlação é a correlação praticamente nula entre a variável *Idade* e as outras variáveis analisadas.

4.3 ANÁLISE DE CLUSTER

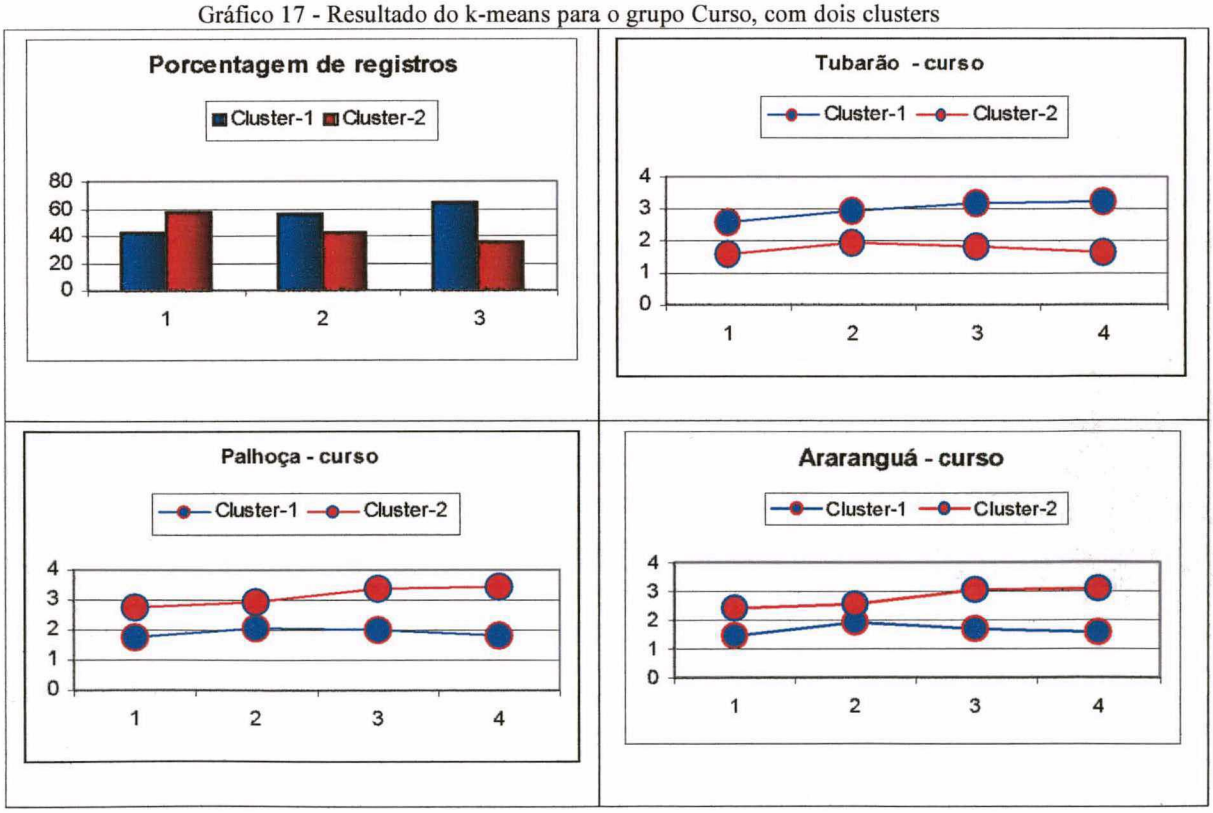
O processo de agrupamento foi feito utilizando-se o algoritmo K-means, escolhido por ser um método estatístico que usa uma medida de similaridade que separa dados semelhantes em grupos distintos. Os centros de *clusters* iniciais foram escolhidos aleatoriamente. A técnica de *clustering* tem a característica de transformar registros com grande número de atributos em pequenos conjuntos (MOXON, 1999). Essa técnica é largamente aplicada e recomendada em diversas publicações.

O algoritmo foi aplicado várias vezes na base de dados utilizando-se uma combinação de diferentes variáveis, de acordo com os grupos definidos nas análises anteriores (aplicados separadamente por campus). Os experimentos apresentados a seguir foram os que sugeriram informações mais significativas.

Tabela 18 - Resultado do k-means para o grupo Curso, com dois clusters

Tub	**Centro**		**%**	**Pal**	**Centro**		**%**	**Ara**	**Centro**		**%**
C1	Med	Des	42,59	C1	Med	Des	57,12	C1	Med	Des	64,29
Q1	2,61	0,85		Q1	1,73	0,62		Q1	1,45	0,56	
Q3	2,92	0,93		Q3	2,05	0,84		Q3	1,92	0,79	
Q4	3,16	0,85		Q4	1,97	0,73		Q4	1,69	0,67	
Q7	3,22	1,02		Q7	1,84	0,77		Q7	1,57	0,69	
C2			57,41	C2			42,88	C2			35,71
Q1	1,57	0,58		Q1	2,73	0,86		Q1	2,41	1,04	
Q3	1,93	0,80		Q3	2,95	0,98		Q3	2,55	1,04	
Q4	1,84	0,67		Q4	3,35	0,79		Q4	3,05	0,90	
Q7	1,67	0,70		Q7	3,44	0,97		Q7	3,09	1,07	

Gráfico 17 - Resultado do k-means para o grupo Curso, com dois clusters



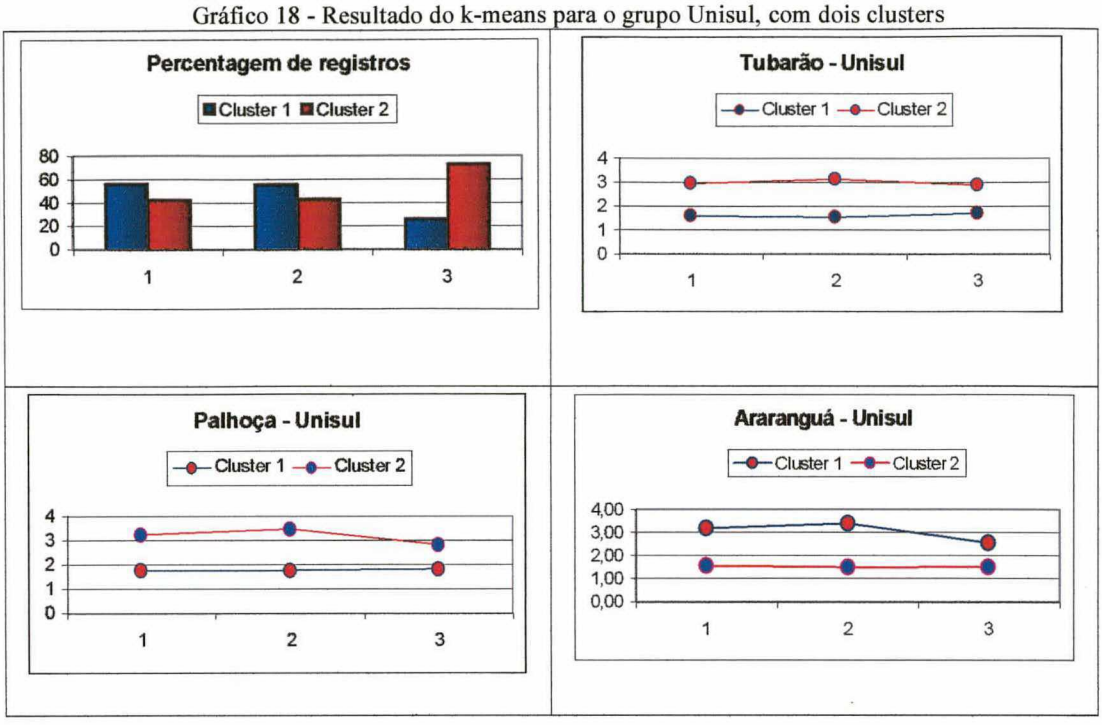
Observa-se uma similaridade muito grande entre os *clusters* dos três campi, na porcentagem de registro de cada *cluster*. Pode-se observar que um *cluster* agrupa as respostas afirmativas (itens 1 e 2) e o outro, as negativas (itens 3 e 4). No campus de Araranguá, as médias dos *clusters* sempre estão mais próximas das respostas afirmativas (item 1) do que no campus da Palhoça. O mesmo acontece com o *cluster* de respostas negativas: no campus da Palhoça, a média é maior que no de Araranguá. O campus de Tubarão encontra-se em ambos os *clusters* bem próximo dos valores da Palhoça.

Através dessas análises podemos concluir que os cursos do campus de Araranguá são mais bem avaliados e que existe uma relação entre as variáveis analisadas, o que significa que a situação de uma interfere no resultado de outra.

Tabela 19 - Resultado do k-means para o grupo Unisul, com dois clusters

Tub	**Centro**		**%**	**Pal**	**Centro**		**%**	**Ara**	**Centro**		**%**
C1	Med	Des	56,81	C1	Med	Des	56,20	C1	Med	Des	26,71
Q5	1,57	0,55		Q5	1,7728	0,58		Q5	3,18	0,90	
Q6	1,51	0,57		Q6	1,7446	0,64		Q6	3,38	0,87	
Q20	1,68	0,79		Q20	1,8498	0,88		Q20	2,54	1,15	
C2			43,18	C2			43,80	C2			73,29
Q5	2,92	0,87		Q5	3,2464	0,79		Q5	1,55	0,55	
Q6	3,10	0,90		Q6	3,4497	0,80		Q6	1,48	0,58	
Q20	2,89	1,05		Q20	2,8075	1,09		Q20	1,50	0,66	

Gráfico 18 - Resultado do k-means para o grupo Unisul, com dois clusters



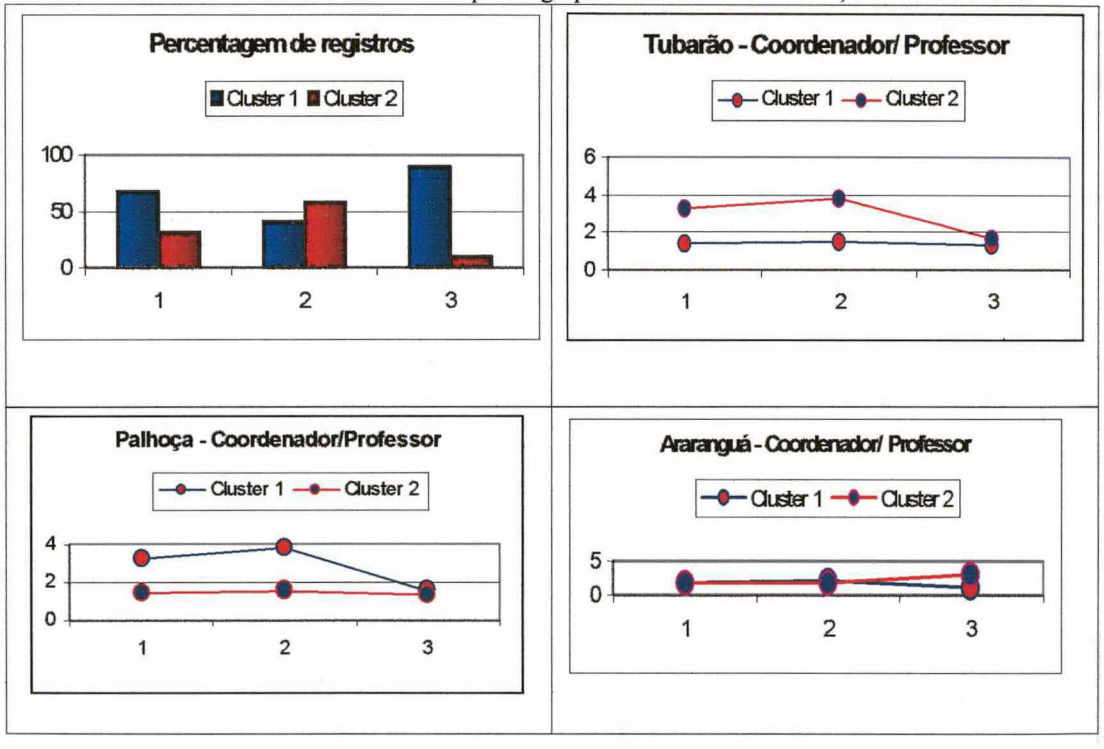
Nesse novo conjunto de variáveis, observa-se uma similaridade muito grande entre os *clusters* nos campi de Tubarão e da Palhoça: o *cluster* 1, com respostas predominantemente afirmativas, e o *cluster* 2, com respostas predominantemente negativas. No campus de Araranguá o *cluster* 2, representando as respostas afirmativas, obteve um alto percentual em relação ao *cluster* 1, com respostas negativas.

Essas análises mostram que a Unisul, sua imagem, qualidade e ambiente são bem avaliados e novamente uma variável influencia nos resultados das outras.

Tabela 20 - Resultado do k-means para o grupo Coordenador/Professor, com dois clusters

Tub	**Centro**		**%**	**Pal**	**Centro**		**%**	**Ara**	**Centro**		**%**
C1	Med	Des	68,13	C1	Med	Des	41,59	C1	Med	Des	90,00
Q27	1,44	0,59		Q27	3,21	0,98		Q27	1,77	1,03	
Q31	1,54	0,61		Q31	3,77	0,72		Q31	2,06	1,20	
Q36	1,37	0,84		Q36	1,56	0,99		Q36	1,12	0,32	
C2			31,87	C2			58,41	C2			10,0
Q27	3,25	0,99		Q27	1,44	0,59		Q27	1,78	0,99	
Q31	3,82	0,73		Q31	1,54	0,55		Q31	1,79	1,17	
Q36	1,65	1,07		Q36	1,35	0,77		Q36	3,03	1,19	

Gráfico 19 - Resultado do k-means para o grupo Coordenador/Professor, com dois clusters



Considerando as variáveis do grupo *Coordenador/Professor*, observa-se uma similaridade muito grande entre os *clusters* nos campi de Tubarão e da Palhoça. Nos dois campi, um dos *clusters* representa respostas afirmativas e o outro representa respostas negativas, com exceção da variável Q36, que é predominantemente afirmativa em ambos

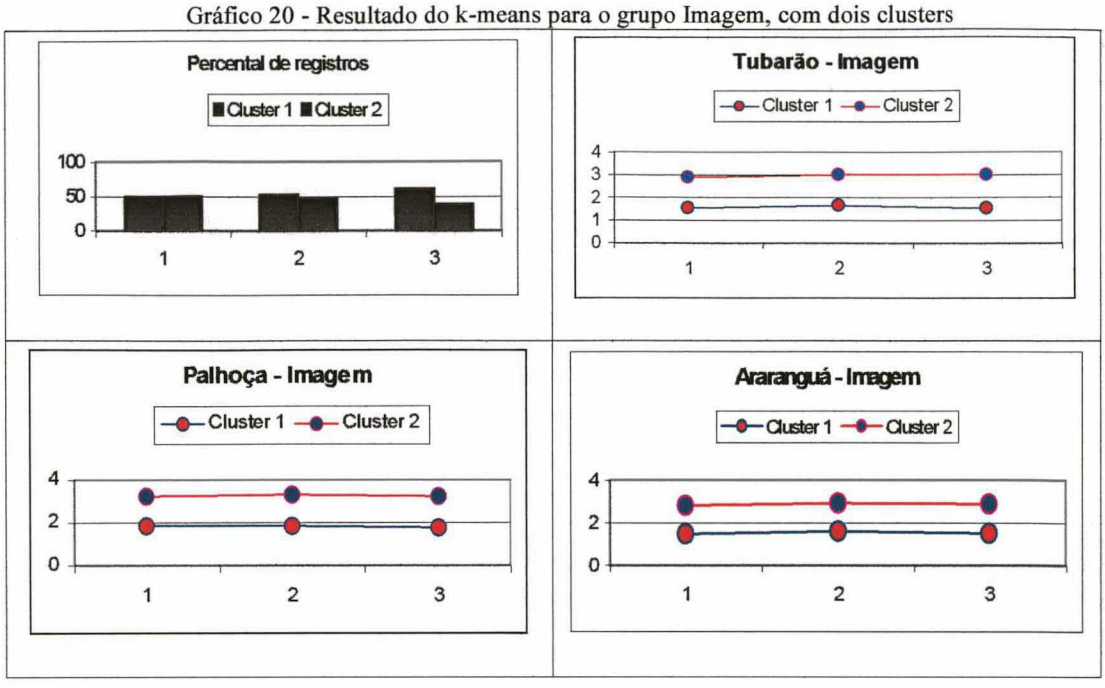
os *clusters*. No campus de Araranguá, os dois *clusters* são bem semelhantes e representam respostas afirmativas. A única exceção é na resposta da questão Q36, que é negativa, mas representa somente 10% dos dados neste *cluster*.

Esses dados mostraram que novamente o campus de Araranguá foi mais bem avaliado e a variável *Produtividade* foi sempre positiva, independentemente da imagem e da gerência do curso.

Tabela 21 - Resultado do k-means para o grupo Imagem, com dois clusters

Tub	**Centro**		**%**	**Pal**	**Centro**		**%**	**Ara**	**Centro**		**%**
C1	Med	Des	49,93	C1	Med	Des	52,78	C1	Med	Des	60,99
Q6	1,54	0,62		Q6	1,82	0,76		Q6	1,47	0,60	
Q7	1,65	0,69		Q7	1,86	0,80		Q7	1,60	0,68	
Q31	1,55	0,75		Q31	1,75	0,86		Q31	1,50	0,68	
C2			50,07	C2			47,22	C2			39,01
Q6	2,85	1,03		Q6	3,24	0,94		Q6	2,81	1,13	
Q7	3,02	1,11		Q7	3,28	1,06		Q7	2,92	1,18	
Q31	2,98	1,23		Q31	3,26	1,16		Q31	2,87	1,36	

Gráfico 20 - Resultado do k-means para o grupo Imagem, com dois clusters

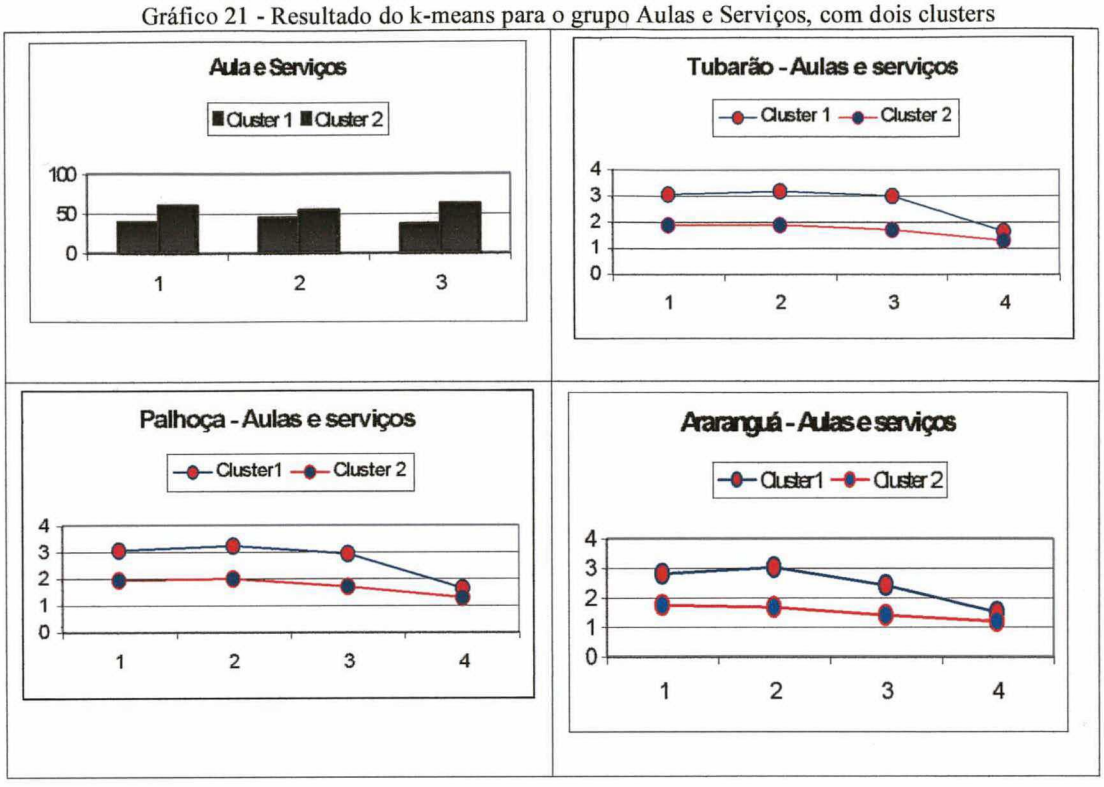


No grupo *Imagem*, observa-se uma similaridade muito grande entre os *clusters* dos três campi, na porcentagem de cada registro. Um dos *clusters* agrupa as respostas afirmativas (1 e 2) e o outro, as negativas (3 e 4). No campus de Araranguá, o número de respostas afirmativas destaca-se um pouco em relação ao de respostas negativas.

Tabela 22 - Resultado do k-means para o grupo Aula e Serviços, com dois clusters

Tub	**Centro**		**%**	**Pal**	**Centro**		**%**	**Ara**	**Centro**		**%**
C1	Med	Des	39,40	C1	Med	Des	45,37	C1	Med	Des	36,92
Q3	3,06	0,90		Q3	3,05	0,93		Q3	2,81	0,96	
Q4	3,19	0,87		Q4	3,26	0,84		Q4	3,03	0,90	
Q20	2,99	1,05		Q20	2,92	1,04		Q20	2,42	1,08	
Q36	1,67	1,11		Q36	1,63	1,35		Q36	1,51	0,96	
C2			60,60	C2			54,63	C2			63,08
Q3	1,90	0,74		Q3	1,93	0,75		Q3	1,75	0,67	
Q4	1,90	0,70		Q4	1,98	0,76		Q4	1,67	0,66	
Q20	1,69	0,76		Q20	1,73	0,79		Q20	1,40	0,58	
Q36	1,32	0,76		Q36	1,27	0,67		Q36	1,19	0,56	

Gráfico 21 - Resultado do k-means para o grupo Aulas e Serviços, com dois clusters



No grupo *Aulas e Serviços*, existe uma similaridade muito grande entre os *clusters* dos três campi, no que tange o porcentual de registro e comportamento. Pode-se observar que um dos *clusters* agrupa as respostas afirmativas (1 e 2) e o outro, as negativas (3 e 4), com exceção das respostas da questão 36 (produtividade), que é afirmativa em ambos os campi e *clusters*.

4.3 REGRAS DE ASSOCIAÇÃO

Para obtenção das regras de associação foi utilizado o algoritmo *Apriori* da ferramenta WEKA (WRITTEN, 2000). O algoritmo foi aplicado na base de dados utilizando-se uma combinação de diferentes variáveis, segundo os grupos definidos nas análises anteriores (aplicados separadamente por campus). As regras apresentadas são as que demonstram uma confiança acima de 0,50.

Tabela 23 - Regras para o grupo Curso do campus de Tubarão

N	Regras	Sup	Conf
1	Se q4_exigencia=2 6707 Então q1_qualidade_curso=2 4155	0,25	0,62
2	Se q7_imagem_curso=2 5498 Então q1_qualidade_curso=2 3036	0,19	0,55
3	Se q7_imagem_curso=1 4539 Então q1_qualidade_curso=1 2494	015	0,55
4	Se q1_qualidade_curso=2 7617 Então q4_Exigencia=2 4155	0,25	0,55
5	Se q7_imagem_curso=2 5498 Então q4_Exigencia=2 2837	0,17	0,52
6	Se q1_qualidade_curso=1 4840 Então q7_imagem_curso=1 2494	0,15	0,52
7	Se q3_conhece_curso=2 6411 Então q4_exigencia=2 3244	0,20	0,51
8	Se q3_conhece_curso=2 6411 Então q1_qualidade_curso=2 3212	0,20	0,5

Observa-se que as regras geradas são todas referentes às respostas afirmativas, ou seja 1 e 2. Pode-se perceber uma associação no comportamento das variáveis *Imagem do curso*, *Qualidade do curso* e *Exigência* , isto é, quando um dos itens não foi bem avaliado, por conseqüência os demais itens também não são. Essa informação é importante para o coordenador saber que a imagem do seu curso depende também de outras variáveis.

Regras similares também foram encontradas nos campi de Palhoça e de Araranguá, como podemos observar abaixo.

Tabela 24 - Regras para o grupo Curso do campus da Palhoça

N	Regras	Sup	Conf
1	Se q4_exigencia=2 4395 Então q1_qualidade_curso=2 2959	0,24	0,67
2	Se q7_imagem_curso=2 3855 Então q1_qualidade_curso=2 2342	0,19	0,61
3	Se q3_conhece_curso=2 4479 Então q1_qualidade_curso=2 2425	0,20	0,54

Tabela 24 - Regras para o grupo Curso do campus de Araranguá

N	Regras	Sup	Conf
1	Se q4_Exigencia=1 796 Então q1_Qualidade_curso=1 642	0,22	0,81
2	Se q1_Qualidade_curso=2 q3_conhece_curso=2 635 Então q4_Exigencia=2 435	0,15	0,69

3	Se q1_Qualidade_curso=2 q4_Exigencia=2 694 Então q3_conhece_curso=2 435	0,15	0,63
4	Se q3_conhece_curso=2 q4_Exigencia=2 711 Então q1_Qualidade_curso=2 435	0,15	0,61
5	Se q4_Exigencia=2 1164 Então q3_conhece_curso=2 711	0,25	0,61
6	Se q4_Exigencia=2 1164 Então q1_Qualidade_curso=2 694	0,24	0,6
7	Se q3_conhece_curso=1 742 Então q1_Qualidade_curso=1 434	0,15	0,58
8	Se q7_imagem_curso=1 1041 Então q1_Qualidade_curso=1 600	0,21	0,58
9	Se q1_Qualidade_curso=2 1209 Então q4_Exigencia=2 694	0,24	0,57
10	Se q7_imagem_curso=2 986 Então q4_Exigencia=2 544	0,19	0,55
11	Se q4_Exigencia=1 796 Então q7_imagem_curso=1 439	0,15	0,55
12	Se q3_conhece_curso=2 1310 Então q4_Exigencia=2 711	0,25	0,54
13	Se q7_imagem_curso=2 986 Então q3_conhece_curso=2 526	0,18	0,53
14	Se q1_Qualidade_curso=2 1209 Então q3_conhece_curso=2 635	0,22	0,53
15	Se q1_Qualidade_curso=1 1240 Então q4_Exigencia=1 642	0,23	0,52
16	Se q7_imagem_curso=2 986 Então q1_Qualidade_curso=2 510	0,18	0,52

Os dados do campus de Araranguá geraram várias regras, mas uma delas merece destaque pela sua confiança. Com 81% de confiança, podemos dizer que o curso que possui exigência é considerado um curso de qualidade.

Os resultados nos três campi são bastantes semelhantes, e todas as variáveis do grupo estão associadas, ou seja a avaliação de uma variável é boa quando as outras também são. No campus da Palhoça, o número de regras geradas foi reduzido bem como a confiança das regras encontradas.

Tabela 25 - Regras para o grupo Unisul do campus de Tubarão

N	**Regras**	**Sup**	**Conf**
1	Se q5_qualidade_unisul=1 4356 Então q6_imagem_unisul=1 3328	0,20	0,76
2	Se q6_imagem_unisul=2 6011 Então q5_qualidade_unisul=2 4273	0,26	0,71
3	Se. Q6_imagem_unisul=1 4966 Então q5_Qualidade_unisul=1 3328	0,20	0,67
4	Se q5_qualidade_unisul=2 7239 Então q6_imagem_unisul=2 4273	0,26	0,59

5	Se q20_ambiente=2 5926 Então q5_qualidade_unisul=2 3314	0,20	0,56
6	Se. Q6_imagem_unisul=1 4966 Então q20_ambiente=1 2604	0,16	0,52
7	Se. Q20_ambiente=1 5045 Então q6_imagem_unisul=1 2604	0,16	0,52

Tabela 26 - Regras para o grupo Unisul do campus da Palhoça

N	Regras	Sup	Conf
1	Se q6_imagem_unisul=2 4179 Então q5_qualidade_unisul=2 2834	0,23	0,68
2	Se q5_qualidade_unisul=2 5018 Então q6_imagem_unisul=2 2834	0,23	0,56
3	Se q20_ambiente=2 4291 Então q5_qualidade_unisul=2 2231	0,18	0,52

Tabela 27 - Regras para o grupo Unisul do campus de Araranguá

N	Regras	Sup	Conf
1	Se q5_qualidade_unisul -=1 1019 Então q6_imagem_unisul=1 858	0,30	0,84
2	Se q5_qualidade_unisul=1 E q20_ambiente=1 698 Então q6_imagem_unisul=1 584	0,20	0,84
3	Se q6_imagem_unisul=2 972 Então q5_qualidade_unisul=2 734	0,25	0,76
4	Se q6_imagem_unisul=1 E q20_ambiente=1 780 Então q5_qualidade_unisul=1 584	0,20	0,75
5	Se q6_imagem_unisul=1 1182 Então q5_qualidade_unisul=1 858	0,30	0,73
6	Se q5_qualidade_unisul=1 1019 Então q20_ambiente=1 698	24,2	0,68
7	Se q5_qualidade_unisul=1 E q6_imagem_unisul=1 858 Então q20_ambiente=1 584	0,20	0,68
8	Se q6_imagem_unisul=1 1182 Então q20_ambiente=1 780	0,27	0,66
9	Se q5_qualidade_unisul=2 1224 Então q6_imagem_unisul=2 734	0,25	0,6
10	Se q5_qualidade_unisul=1 1019 Então q6_imagem_unisul=1 E q20_ambiente=1 584	0,20	0,57
11	Se q20_ambiente=1 1371 Então q6_imagem_unisul=1 780	0,27	0,57
12	Se q20_ambiente=2 1012 Então q5_qualidade_unisul=2 533	0,19	0,53
13	Se. q20_ambiente=1 1371 Então q5_qualidade_unisul=1 698	0,24	0,51

As regras geradas são referentes às respostas positivas. Observa-se uma forte associação entre as variáveis *Qualidade da Unisul, Imagem da Unisul* e *Ambiente*. Nos casos em que uma variável é bem avaliada, a outra também o é.

No grupo *Unisul*, a imagem e a qualidade encontram-se associadas nos três campi. No campus de Tubarão, o bom ambiente está associado à boa imagem; no campus da Palhoça, o ambiente está associado à qualidade da Unisul; e no campus de Araranguá, o ambiente está associado tanto à imagem quanto à qualidade da Unisul. A confiança das regras encontradas no campus de Araranguá é um pouco superior à do campus de Tubarão e bem superior à do campus da Palhoça.

Tabela 28 - Regras para o grupo Coordenador/Professor do campus de Tubarão

N	Regras	Sup	Conf
1	Se q31_imagem_coord=1 E q36_prudutividade=1 4701 Então q27_gerencia_curso=1 3991	0,24	0,85
2	Se q31_imagem_coord=1 5846 Então q27_gerencia_curso=1 4917	0,30	0,84
3	Se q27_gerencia_curso=1 E q31_imagem_coord=1 4917 Então q36_prudutividade=1 3991	0,24	0,81
4	Se q31_imagem_coord=1 5846 Então q36_prudutividade=1 4701	0,29	0,8
5	Se q27_gerencia_curso=1 7014 Então q36_prudutividade=1 5525	0,34	0,79
6	Se q27_gerencia_curso=1 E q36_prudutividade=1 5525 Então q31_imagem_coord=1 3991	0,24	0,72
7	Se q31_imagem_coord=2 4795 Então q36_prudutividade=1 3396	0,21	0,71
8	Se q27_gerencia_curso=2 4763 Então q36_prudutividade=1 3351	0,20	0,7
9	Se q27_gerencia_curso=1 7014 Então q31_imagem_coord=1 4917	0,30	0,7
10	Se q31_imagem_coord=1 5846 Então q27_gerencia_curso=1 E q36_prudutividade=1 3991	0,24	0,68
11	Se q27_gerencia_curso=1 7014 Então q31_imagem_coord=1 E q36_prudutividade=1 3991	0,24	0,57
12	Se q27_gerencia_curso=2 4763 Então q31_imagem_coord=2 2547	0,16	0,53
13	Se. Q31_imagem_coord=2 4795 Então q27_gerencia_curso=2 2547	0,16	0,53

Tabela 29 - Regras para o grupo Coordenador/Professor do campus da Palhoça

N	Regras	Sup	Conf
1	Se q31_imagem_coorde=1 E q36_prudutividade=1 2843 Então q27_gerencia_curso=1 2425	0,20	0,85
2	Se. q31_imagem_coorde=1 3474 Então q27_gerencia_curso=1 2945	0,24	0,85
3	Se q27_gerencia_curso=1 E q31_imagem_coorde=1 2945 Então q36_prudutividade=1 2425	0,20	0,82
4	Se q31_imagem_coorde=1 3474 Então q36_prudutividade=1 2843	0,23	0,82

5	Se q27_gerencia_curso=1 4508 Então q36_prudutividade=1 3596	0,29	0,8
6	Se q31_imagem_coorde=2 3574 Então q36_prudutividade=1 2602	0,21	0,73
7	Se. q27_gerencia_curso=2 3566 Então q36_prudutividade=1 2560	0,21	0,72
8	Se q31_imagem_coorde=1 3474 Então q27_gerencia_curso=1 E q36_prudutividade=1 2425	0,20	0,7
9	Se q27_gerencia_curso=1 E q36_prudutividade=1 3596 Então q31_imagem_coorde=1 2425	0,20	0,67
10	Se q27_gerencia_curso=1 4508 Então q31_imagem_coorde=1 2945	0,24	0,65
11	Se q27_gerencia_curso=2 3566 Então q31_imagem_coorde=2 1929	0,16	0,54
12	Se q27_gerencia_curso=1 4508 Então q31_imagem_coorde=1 E q36_prudutividade=1 2425	0,20	0,54

Tabela 30 - Regras para o grupo Coordenador/Professor do campus de Araranguá

N	Regras	Sup	Conf
1	Se q31_Imagem_coorde=1 E q36_prudutividade=1 1081 Então q27_Gerencia_curso=1 985	0,34	0,91
2	Se q31_Imagem_coorde=1 1259 Então q27_Gerencia_curso=1 1120	0,39	0,89
3	Se q27_Gerencia_curso=1 E q31_Imagem_coorde=1 1120 Então q36_prudutividade=1 985	0,34	0,88
4	Se q27_Gerencia_curso=1 1521 Então q36_prudutividade=1 1320	0,45	0,87
5	Se q31_Imagem_coorde=1 1259 Entãoq36_prudutividade=1 1081	0,38	0,86
6	Se q31_Imagem_coorde=1 1259 Então q27_Gerencia_curso=1 E q36_prudutividade=1 985	0,34	0,78
7	Se q31_Imagem_coorde=2 860 Então q36_prudutividade=1 666	0,23	0,77
8	Se q27_Gerencia_curso=1 E q36_prudutividade=1 1320 Então q31_Imagem_coorde=1 985	0,34	0,75
9	Se q27_Gerencia_curso=1 1521 Então q31_Imagem_coorde=1 1120	0,39	0,74
10	Se q27_Gerencia_curso=2 808 Então q36_prudutividade=1 580	0,20	0,72
11	Se q27_Gerencia_curso=1 1521 Então q31_Imagem_coorde=1 E q36_prudutividade=1 985	0,34	0,65
12	Se q27_Gerencia_curso=2 808 Então q31_Imagem_coorde=2 470	0,16	0,58
13	Se Q36_prudutividade=1 2281 Então q27_Gerencia_curso=1 1320	0,46	0,58
14	Se q31_Imagem_coorde=2 860 Então q27_Gerencia_curso=2 470	0,16	0,55

As regras geradas são referentes às respostas positivas. Observa-se uma forte associação entre as variáveis *Gerência do curso, Imagem da Unisul* e *Produtividade.* Nos

casos em que uma variável é bem avaliada, as outras também são. A confiança das regras encontradas é bastante alta.

Os resultados nos três campi são idênticos. A confiança das regras do campus de Araranguá é um pouco superior em relação aos demais campi.

É interessante destacar que a gerência do curso é bem avaliada quando acompanhada de uma boa produtividade.

Tabela 31 - Regras para o grupo Imagem do campus de Tubarão

N	Regras	Sup	Conf
1	Se q7_imagem_curso=1 4539 Então q6_imagem_unisul=1 2548	0,16	0,56
2	Se q6_imagem_unisul=1 4966 Então q31_imagem_coord=1 2611	0,16	0,53
3	Se q6_imagem_unisul=1 4966 Então q7_imagem_curso=1 2548	0,16	0,51

Tabela 32 - Regras para o grupo Imagem do campus de Araranguá

N	Regras	Sup	Conf
1	Se Q7_imagem_curso=1 1041 Então q6_Orgulho_unisul=1 650	0,23	0,62
2	Se Q6_Orgulho_unisul=1 1182 Então q7_imagem_curso=1 650	0,23	0,55
3	Se Q6_Orgulho_unisul=1 1182 Então q31_Imagem_coorde=1 640	0,22	0,54
4	Se Q31_Imagem_coorde=1 1259 Então q6_Orgulho_unisul=1 640	0,22	0,51

As regras geradas são referentes às respostas positivas. Pode-se observar uma certa relação entre as imagens do curso e da Unisul e entre as imagens da Unisul e do coordenador.

No grupo *Imagem*, o campus de Tubarão teve o mesmo comportamento do campus da Palhoça. As regras geradas são todas de baixa confiança. O campus da Palhoça não gerou regras com confiança aceitável.

Tabela 33 - Regras para o grupo Aula e Serviços do campus de Tubarão

N	Regras	Sup	Conf
1	Se q4_exigencia=1 2952 Então q36_prudutividade=1 2466	0,15	0,84
2	Se q20_ambiente=1 5045 Então q36_prudutividade=1 4027	0,25	0,8
3	Se q3_conhece_curso=1 3382 Então q36_prudutividade=1 2689	0,16	0,8
4	Se q3_conhece_curso=2 E q4_exigencia=2 3244 Então q36_prudutividade=1 2503	0,15	0,77
5	Se q4_exigencia=2 6707 Então q36_prudutividade=1 5089	0,31	0,76
6	Se q3_conhece_curso=2 6411 Então q36_prudutividade=1 4844	0,30	0,76
7	Se q20_ambiente=2 5926 Então q36_prudutividade=1 4243	0,26	0,72
8	Se q3_conhece_curso=3 4186 Então q36_prudutividade=1 2825	0,17	0,67
9	Se q4_exigencia=3 4201 Então q36_prudutividade=1 2820	0,17	0,67
10	Se q3_conhece_curso=2 E q36_prudutividade=1 4844 Então q4_exigencia=2 2503	0,15	0,52
11	Se q20_ambiente=2 5926 Então q4_exigencia=2 3022	0,18	0,51
12	Se q3_conhece_curso=2 6411 Então q4_exigencia=2 3244	0,20	0,51

Tabela 34 - Regras para o grupo Aula e Serviços do campus da Palhoça

N	Regras	Sup	Conf
1	Se q3_conhece_curso=1 2291 Então q36_prudutividade=1 1831	0,15	0,8
2	Se q20_ambiente=1 3463 Então q36_prudutividade=1 2725	0,22	0,79
3	Se q4_exigencia=2 4395 então q36_prudutividade=1 3387	0,28	0,77
4	Se q3_conhece_curso=2 4479 Então q36_prudutividade=1 3359	0,26	0,75
5	Se q20_ambiente=2 4291 Então q36_prudutividade=1 3121	0,26	0,73
6	Se q3_conhece_curso=3 3377 Então q36_prudutividade=1 2316	0,19	0,69
7	Se q4_exigencia=3 3583 Então q36_prudutividade=1 2368	0,19	0,66

Tabela 35 - Regras para o grupo Aula e Serviços do campus de Araranguá

N	Regras	Sup	Conf
1	Se q4_Exigencia=1 E q20_ambiente_bom=1 569 Então q36_prudutividade=1 516	0,18	0,91
2	Se q3_conhece_curso=2 E q20_ambiente_bom=1 644 Então q36_prudutividade=1 563	0,20	0,87
3	Se q20_ambiente_bom=1 1371 Então q36_prudutividade=1 1189	0,41	0,87

4	Se q4_Exigencia=1 796 Então q36_prudutividade=1 688	0,24	0,86
5	Se q3_conhece_curso=1 742 Então q36_prudutividade=1 610	0,21	0,82
6	Se q3_conhece_curso=2 1310 Então q36_prudutividade=1 1068	0,37	0,82
7	Se q3_conhece_curso=2 E q4_Exigencia=2 711 Então q36_prudutividade=1 576	0,20	0,81
8	Se q4_Exigencia=2 1164 Então q36_prudutividade=1 917	0,32	0,79
9	Se q20_ambiente_bom=2 1012 Então q36_prudutividade=1 775	0,27	0,77
10	Se q4_Exigencia=1 E q36_prudutividade=1 688 Então q20_ambiente_bom=1 516	0,18	0,75
11	Se q4_Exigencia=3 589 Então q36_prudutividade=1 439	0,15	0,75
12	Se q4_Exigencia=1 796 Então q20_ambiente_bom=1 569	0,20	0,71
13	Se q4_Exigencia=1 796 Então q20_ambiente_bom=1 E q36_prudutividade=1 516	0,18	0,65
14	Se q4_Exigencia=2 E q36_prudutividade=1 917 Então q3_conhece_curso=2 576	0,20	0,63
15	Se q4_Exigencia=2 1164 Então q3_conhece_curso=2 711	0,25	0,61
16	Se 20_ambiente=2 1012 Então q3_conhece_curso=2 550	0,19	0,54
17	Se q3_conhece_curso=2 1310 Então q4_Exigencia=2 711	0,25	0,54
18	Se q3_conhece_curso=2 E q36_prudutividade=1 1068 Então q4_Exigencia=2 576	0,20	0,54
19	Se q3_conhece_curso=2 E q36_prudutividade=1 1068 Então q20_ambiente=1 563	0,20	0,53
20	Se q36_prudutividade=1 2281 Então q20_ambiente=1 1189	0,41	0,52

A partir das regras geradas, pode-se identificar que as respostas são predominantemente positivas. Observa-se com clareza que a boa produtividade está associada à exigência, ao bom ambiente e ao conhecimento do curso. A variável que mais está associada à produtividade é a exigência. No campus de Araranguá foi encontrada uma regra fora dos padrões dos outros campi, mostrando que mesmo com pouca exigência a produtividade é boa.

Nos três campi a produtividade depende das demais variáveis analisadas neste grupo. No campus de Tubarão, a variável que mais influi na produtividade é o ambiente, ao passo que na Palhoça e em Araranguá é a exigência. O campus de Araranguá gerou regras com maior confiança do que os de Tubarão e da Palhoça.

Quando as regras são analisadas para a tomada de decisão, elas podem direcionar investimentos na melhoria dos ambientes e conseqüentemente influenciar a produtividade em sala de aula.

4.5 ANÁLISE FINAL

4.5.1 Dados estatísticos

Pôde-se observar que cerca de 80% dos alunos são jovens até 25 anos de idade. Quanto à *Imagem*, à *Qualidade* e à *Produtividade*, a Unisul apresenta dados bastantes positivos. Aproximadamente 75% dos alunos consideram a Unisul uma universidade com qualidade, e 70% dizem ter uma boa imagem dessa instituição. Já 90% dos alunos acham que estão recebendo todo o conteúdo programático previsto.

Comparando-se esses dados nos campi analisados, observa-se que o campus de Araranguá possui uma média de idade superior aos campi de Tubarão e da Palhoça. Quando às outras variáveis analisadas, o campus de Araranguá possui um número de respostas positivas maior que os demais campi.

4.5.2 Utilizando Cluster

Analisando-se os *clusters* gerados com o algoritmo K-means, foram observados:

- **Imagem** – Nos três campi, utilizando-se dois *clusters,* o algoritmo conseguiu dividir os dados em *dois* grupos: o *cluster* 1 representa os alunos que responderam "sim" e "praticamente sim" quanto à imagem da Unisul, e o *cluster* 2 representa os alunos que responderam "não" e "praticamente não" à boa imagem da Unisul.

Podemos observar que as três variáveis *Imagem da Unisul, Imagem do curso* e *Imagem do coordenador* possuem forte relação entre si.

Nos campi de Tubarão e da Palhoça, a maioria dos alunos pertence ao *cluster* 1 (satisfeitos), enquanto que no campus de Araranguá a maioria dos alunos pertencem ao *cluster* 2 (insatisfeitos).

Analisando-se a *Imagem da Unisul* com relação à idade dos alunos, percebe-se que nos campi de Tubarão e de Araranguá os alunos satisfeitos são os de menos idade (até 25 anos). No campus da Palhoça a situação é inversa, visto que os alunos com maior idade (acima de 25) é que possuem uma boa imagem da Unisul.

A situação é bastante parecida na análise da imagem do curso e do coordenador com relação à idade.

- **Qualidade** - Analisando-se a qualidade da Unisul e a qualidade do curso percebe-se novamente uma forte relação entre essas duas variáveis. Nos três campi observou-se que existe um maior número de alunos concordando que a Unisul possui qualidade.

- **Exigência** – No campus de Tubarão, os alunos com idade até 25 anos estão satisfeitos com o nível de exigência. Os dados dos demais campi não geraram *clusters* significativos.

- **Ambiente do Campus** – O ambiente do campus e os serviços comparados com a idade não geraram *clusters* significativos. Pôde-se perceber que o número de alunos satisfeitos é um pouco maior em relação aos insatisfeitos.

- **Produtividade** - A produtividade ou o cumprimento do plano de ensino versus a idade não gerou *cluster* significativo. As opiniões sobre o cumprimento do plano de ensino são bastantes divididas entre os alunos que concordam e os que não concordam.

A técnica de *cluster* conseguiu dividir os dados em um grupo com respostas positivas e em outro com respostas negativas em quase todas as análises, o que significa que existe uma forte relação entre as variáveis analisadas. Esse comportamento se repetiu nos três campi, com a diferença do número de registros do *cluster* positivo sempre superior no campus de Araranguá. Por outro lado o número de registros dos *clusters* negativos foi superior no campus da Palhoça.

4.5.3 Utilizando regras de associação

- **Imagem** - As regras geradas confirmam a relação entre as variáveis analisadas – imagem da Unisul, do curso e do coordenador. O campus com melhor imagem é o de Araranguá, segundo as regras de associação geradas.

- **Qualidade** – As regras com maior confiança indicam a qualidade da Unisul em todos os campi, observando também uma relação entre qualidade do curso, da Unisul e do coordenador.

- **Exigência** - Os alunos com idade até 20 anos responderam que o nível de exigência quase sempre é adequado.

- **Produtividade** - A produtividade, ou seja, o cumprimento do plano de ensino está sendo executada para os alunos com idade até 20 anos.

- **Ambiente** - As regras geradas apresentam confiança bastante baixa.

Comparando-se os dados nos três campi, observou-se que as regras geradas são muito semelhantes, porém a confiança das regras geradas no campus de Araranguá é sempre ligeiramente superior ao campus de Tubarão e bastante superior ao campus da Palhoça.

Cada uma das técnicas citadas gerou resultados de acordo com as suas características. Com a análise estatística, pôde-se ter uma visão geral dos dados utilizados. Através da análise de *cluster,* o algoritmo dividiu esses dados em grupos, sendo possível observar neles algumas tendências. Essa técnica apresentou resultados bastante ricos em todos os grupos de variáveis analisados. Para aproveitar tais resultados, é necessário fazer uma interpretação do que significa cada um dos grupos gerados. Porém, essa interpretação às vezes se torna extremamente trabalhosa.

Por último, foi utilizada a técnica de regras de associação. Essa técnica gerou regras de tendência dos dados bem semelhantes às da análise de *cluster*. Já essas regras estão mais claras, não exigindo grande esforço na sua interpretação.

As análises feitas pela Assessoria de Avaliação Institucional se voltaram para a geração de dados estatísticos, a média e o percentual de satisfação dos alunos. Os resultados serviram para avaliar individualmente cada professor na disciplina bem como avaliar o coordenador e o curso. A análise deste trabalho foi efetuada de forma global por campus. O resultado encontrado foi a verificação de tendências comuns dos alunos em cada campus e nas variáveis analisadas.

Como resultado prático e útil para a tomada de decisão podemos destacar algumas descobertas:

- O campus de Araranguá foi o mais bem avaliado na maioria das variáveis analisadas:

- o ambiente do campus apareceu como importante variável, mostrando associações que influenciam a produtividade e conseqüentemente, refletem na imagem do curso e da Unisul, pois a mineração de dados gerou regras que denotam que a imagem depende da produtividade;
- os agrupamentos mostraram forte relação entre a Imagem da Unisul, a Imagem do curso e a Imagem do coordenador.
- o perfil dos alunos satisfeitos no campi de de Tubarão e de Araranguá são os de menor idade(até 25anos). Já em Palhoça, a situação é inversa, alunos com mais de 25 anos estão satisfeitos com a Unisul.

5 CONCLUSÕES E RECOMENDAÇÕES

5.1 CONCLUSÕES

A Unisul, através do processo de Avaliação Institucional e da Secretaria Geral de Ensino, reúne uma grande quantidade de informações. Tais informações precisam ser trabalhadas utilizando-se alguma técnica de extração de conhecimento, para que, dessa forma, esses conhecimentos possam servir à tomada de decisão.

Os algoritmos de *data mining* utilizados na análise do presente estudo proporcionaram (1) a descoberta de algumas regras de associação, (2) o relacionamento ou influência de uma variável em relação à outra e, ainda, (3) o comportamento dos dados através da divisão destes em grupos, considerando-se o seu grau de similaridade.

Durante o processo de limpeza dos dados, foi encontrada uma grande quantidade de dados nulos, o que tornou essa tarefa bastante difícil e ocasionou a perda de muitas linhas de dados, tendo, como conseqüência, prejuízos nos resultados obtidos.

A utilização da técnica de *clustering* é bastante adequada ao processo de descoberta do conhecimento, quando não se possui um conhecimento prévio, ou seja, não se dispõe de resultados já obtidos para fazer um treinamento supervisionado. Apesar de essa técnica requerer um trabalho minucioso na interpretação dos grupos gerados, ela mostrou-se bastante eficiente quando aplicada nos dados da avaliação institucional.

Ao ser aplicada a técnica de regras de associação, percebeu-se que as regras geradas foram muito semelhantes aos resultados obtidos na técnica de *clustering*. Pode-se dizer, então, que ambas as técnicas utilizadas mostraram-se eficientes na descoberta de conhecimento.

Os resultados encontrados foram úteis em diversos pontos, identificando variáveis que precisam ser trabalhadas, pois influenciam diretamente a qualidade dos serviços prestados.

5.2 TRABALHOS FUTUROS

Os conhecimentos extraídos a partir dos dados analisados foram bastante interessantes, porém, é possível trabalhar com outras técnicas de extração de conhecimentos nessa mesma base.

A utilização de um *data warehouse* que possibilite o armazenamento dos dados ao longo do tempo e a inclusão de algumas variáveis externas pode enriquecer a análise desses dados.

Os dados da avaliação institucional poderiam ser comparados aos da situação financeira e da procedência dos alunos antes do ingresso na universidade.

A Unisul já possui, em sua base de dados, as notas e o desempenho de cada aluno. Esses dados parecem ser muito ricos se comparados também aos dados da avaliação institucional.

O Provão é uma outra forma de avaliação das universidades que possibilita estimar o conhecimento do aluno após o término do curso. A utilização desses dados confrontados com os da avaliação institucional pode ser também uma rica fonte de informações.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, Maria Bernadete Martins; ARRUDA, Susuna Margareth. **Como fazer referências**: bibliográficas, eletrônicas e demais formas de documentos. Disponível em: <http://www.bu.ufsc.br/home982.htm>. Acesso em: out. 2001.

BARBIERI, Carlos. **BI – Business Inteligence:** modelagem e tecnologia, Axcell Books, 2001.

BERNARDINO, Giana da Silva et al. **Câncer on-Line Minner**. 2000. P 6-22.Trabalho de Conclusão de Curso(Ciências da Computação) Universidade do Sul de Santa Catarina - Unisul, Tubarão, 2000.

BIGOLIN, Nara Martini. **Data mining**: conceitos e técnicas. Escola Regional de Informática da SBC Sul: Ijuí, RS; Foz do Iguaçu, PR; Tubarão, SC. p. 15-19, maio 2000.

BIGUS, Joseph P. **Data Mining With Neural Networks.** Computing McGraw-Hill, San Francisco, 1997.

BISPO, Carlos Alberto Ferreira; CAZARINI, Edson Walmir. **Conceitos básicos e a elaboração do projeto lógico de um Data Warehouse.** Disponível em": <http://cazarini.cpd.eesc.sc.usp.br/Bisp/Art-03.hmt> Acesso em : out 1999.

CARVALHO, Eduardo B. et al. **A avaliação institucional**: um processo permanente. Publicação Científica da Universidade do Sul de Santa Catarina - Unisul: Tubarão. out. 1998.

DILLY, Ruth. **Data Mining:** Parallel Computer Centre . The Queen's University of Belfast, 1995. Possui material sobre alguns algoritmos de mineração de dados, tais como análise de grupos, redes neurais e indução de regras. Traz ainda um material interessante sobre a On-Line Analytical Processing (OLAP), comparando-a com a On-Line Transaction Processing (OLTP). Disponível em: <http://www-pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html>. Acesso em: jan 2000.

DW BRASIL, **Data Mining.** Possui uma introdução, conceito e técnicas de mineração de dados, Data Warehouse, CRM.
Disponível em : http://www.dwbrasil.com.br/html/dmining.html. Acesso em : set 2000.

FAYYAD, Usama. **Data Mining and Knowlege Discovery.**
Cambridge, Mass.: MIT Press, Editorial 1997

FELDENS, Miguel Artur; MORAES, Rodrigo Leal; PAVAN, Altimo. **Data Mining na gestão hospitalar.** Disponível em: < http://www.inf.ufgr.br/~feldens/datamining.html>. Acesso em: set 1999.

FELDEMS, Miguel Artur. **A descoberta de conhecimento aplicada à detecção de anomalias em bases de dados**. Porto Alegre: Instituto de Informática da UFRGS, 1996.

FELDENS, Miguel Artur. **Engenharia da descoberta do conhecimento em bases de dados**: estudo e aplicação na área da saúde. Porto Alegre: Instituto de Informática da UFRGS, 1997.

FIGUEIRA, Rafael. **Mineração de dados e banco de dados orientados a objetos.** Disponível em: < http://www.cos.ufrj.br/~rafael/mestrado/bdnc/monografia.html> acesso em: ago 1999.

GIMENES, Eduardo. **Data Mining.** Contém uma monografia sobre Data Mining. Disponível em: <http://br.geocities.com/dugimenes/.> Acesse em nov 2000.

GONÇALVES, Alexandre Leopoldo. **Utilização de técnicas de mineração de dados na análise dos Grupos de Pesquisa no Brasil**, 2000. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina- Departamento de Engenharia de Produção e Sistemas, Florianopolis, Agosto de 2000.

HOLSHEIMER, M.; SIEBES, A. **Data mining**: the search for knowledge in database. Amsterdam, Netherlands, jul. 1995. Disponível por FTP anônimo em ftp.cwi.nl no arquivo/puc/CWIreports/AA/CS-R9406.ps.Z.

INMON, William H. **Como construir o data warehouse.** Tradução de Ana Maria Netto Guz. Rio de Janeiro: Campus, 1997. Tradução da segunda edição

KIMBALL, Ralph. **The Data Warehouse:** Lifecycle ToolKit, Willey Computer Publishing , New York. 1998.

KIMBALL, Ralph. **Digging Into Data Mining.** Disponível em": <http://www.dbmsmag.com/9710d05.html > Acesso em : mar 1999.

MANNILA, H. **Dfinding Interesting Rules From Large Set of Discovered Association Rules,** 3rd Internacional Conference on Information and Knowledge Management, novembro 1994.

MENDONÇA, E. A. **Hycones III :** sistemas de apoio à UTI cardiológica. 1996. 103 p. Dissertação (Mestrado em Medicina com ênfase em Cardiologia) - Instituto de Cardiologia do Rio Grande do SUL e Fundação Universitária Cardiológica, Porto Alegre, 1996.

MOXON, Bruce. **Defining Data Mining**. The Hows and Whys of Data Mining, and It Differs from other Analytical Tecniques.
Disponível em: <http://www.dbmsmag.com/9608d53.htm>. Acesso em mai 1999

PILA, Adriano Donizete, **Datawarehouse.** Apresenta um artigo sobre Data Warehouse. Disponível em:
<http://www.igce.unesp.br/igce/grad/computacao/cintiab/datamine/datawarehous.html>. Acesso em jan 2001

PILA, Adriano Donizete, **Trabalho de Inteligência Artificial.** Apresenta uma introdução sobre Data Mining, técnicas e modelos, KDD, Data Warehouse.
Disponível em:
<http://www.igce.unesp.br/igce/grad/computacao/cintiab/datamining/introducao.html>.
Acesso em FEV 2000

SANTOS, Jose; HENRIQUES, Nunes; REIS, Vanda : **Data Mining/Data Warehousing.** Possui Conceitos, origem, funcionamento e técnicas de mineração de dados , como também uma comparação com ferramentas OLAP.
Disponível em : http://students.fct.unl.pt/users/nach/dmdw/. Acesso em : Nov de 1999

THEALING, Kurt. **An introduction do data mining**: discovering hidden value in your data warehouse. 2001. Possui uma boa introdução, incluindo fundamentos, evolução, exemplos, arquitetura e aplicações. Disponível em:
<http://www.santafe.edu/~kurt/text/dmwhite/dmwhite.shtml> Acesso em : jan 2001

THEARLING, Kurt . **An Introduction to Data Mining**. Disponível em: <
http://www3.shore.net/~kht/index.htm> Acesso em : novembro de 2000.

UNIVERSIDADE ESTUDUAL DE MARINGA. Grupo de Sistemas Inteligentes. **Mineração de Dados.** Disponível em http://din.uem.br/~ia/mineracao. Acesso em nov 1999.

WEISS, Sholom M.; INDURKHYA, Nittin. **Predictive data mining**: a practical guide. Morgan Kaufmann Publishers: San Francisco, California, 1998.

WITTEN, Ian H. and Eibe Frank - Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgam Kaufmann Publishers, 2000.

7 ANEXOS

Anexo A: Formulário de pesquisa

QUANTO AO CURSO E À UNISUL

01 - Considera que seu curso tem qualidade ?

02 - Tem informações regulares sobre o seu curso ?

03- Conhece o projeto do seu curso na sua totalidade ?

04 - O seu curso está exigindo o suficiente para a sua formação profissional ?

05 - Considera a Unisul uma universidade de qualidade ?

06 - Você tem orgulho de estudar na Unisul ?

07 - Seu curso tem uma boa imagem na sociedade ?

08 - Os formados pelo seu curso têm uma boa aceitação no mercado de trabalho ?

09 - Seu curso goza de prestígio junto aos estudantes da Unisul ?

10 - Você encontra nas biblioteca as fontes bibliográficas que procura ?

11 - A biblioteca oferece atendimento de qualidade ?

12 - Está satisfeito com o uso que faz do acervo disponível na biblioteca ?

13 - A secretaria geral e protocolo demonstram competência nas suas funções ?

14 - Os serviços de xerox e de mecanografia correspondem às suas necessidades ?

15 - O Setor de Apoio ao Estudante (SAE) realiza a divulgação de programas de bolsa em tempo hábil ?

16 - Você recebe informações sobre os outros programas do Setor de Apoio ao Estudante (SAE) ?

17 - Você conhece os programas do Setor de Apoio aos Estágios (SAEST) da Unisul ?

18 - Você dispõe de internet em casa ?

19 - Você está satisfeito com o serviço de informática e internet da Unisul ?

20 - O ambiente do campus é próprio e propício ao desenvolvimento das atividades acadêmicas ?

21 - As salas de aula são adequadas para o desenvolvimento das aulas ?

22 - As condições de uso das salas de aula são adequadas para o desenvolvimento das aulas?

23 - Você considera que existe segurança no campus e nas imediações ?

24 - Você está satisfeito com os serviços da lanchonete ?

QUANTO AO COORDENADOR DO CURSO NO SEU CAMPUS

25 - Oferece horários de atendimento acessível aos alunos ?

26 - Comunica-se com a sua turma de acordo com a necessidade ?

27 - Revela ser um líder democrático e entusiasmado pelas questões do curso ?

28 - Demonstra atividades extracurriculares visando melhorar a formação acadêmica ?

29 - Promove atividades extracurriculares visando a melhorar a formação acadêmica ?

30 - Informa sobre eventos significativos para o seu curso ?

31 - Indicaria seu coordenador como modelo de atuação ?

32 - É pontual, cumpre os horários das aulas ?

33 - É assíduo, comparece às aulas ?

34 - Administra o tempo em sala de aula de forma produtiva ?

35 - Apresentou o plano de ensino ?

36 - Está desenvolvendo o plano de ensino ?

37 - Revela domínio de conteúdo na sua disciplina ?

38 - Expressa o conteúdo das aulas com linguagem acessível aos alunos ?

39 - A bagagem intelectual é estimulante para os alunos ?

40 - Estimula a utilização de bibliografia atualizada ?

41- Estimula a produção científica promovendo atividades de pesquisa na disciplina ?

42 - Incentiva a leitura de livros, textos, jornais e revistas complementares às aulas ?

43 - Adota procedimentos facilitadores da aprendizagem ?

44 - Oportuniza reflexões críticas sobre os conteúdos ?

45 - Dá atenção às dúvidas apresentadas pelos aluno ?

46 - Redige provas ou verificações de aprendizagem de forma clara ?

47 - Informa sobre os critérios adotados na avaliação ?

48 - Planeja avaliações compatíveis com os objetos e conteúdos ministrados ?

49 - Trabalha a disciplina com nível de exigência suficiente ?

50 - Conduz o ensino da disciplina com qualidade ?

51 - Tem boa relação com os alunos ?

52 - Valoriza a participação do aluno nas aulas ?

53 - Estimula o bom relacionamento da classe ?

54 - Demonstra conhecer o projeto do curso ?

55 - Situa o aluno sobre a importância da sua disciplina para o curso ?

56 - Demonstra interesse e compromisso com a Unisul ?

57 - Demonstra conhecimento da Unisul no seu campus ?

58 - Respeita o aluno como pessoa ?

59 - Demonstra coerência entre o discurso e sua prática como professor ?

60 - Valoriza a coordenação e a solidariedade em sala de aula ?

61 - Indicaria este professor a outras turmas ?

62 - Você sente-se tranqüilo para opinar sobre este professor ?

QUANTO AO SEU DESEMPENHO COMO ALUNO

63 - É pontual nas atividades acadêmicas da disciplina ?

64 - É dedicado ao aprendizado da disciplina ?

65 - Está satisfeito com o próprio desempenho ?

As respostas para as questões acima podem ser um valor de 1 a 5 (1 – Sempre/Sim, 2 – Quase sempre/Praticamente sim, 3 Raramente/Praticamente não, 4 – Nunca/Não, 5 Sem resposta/Não se aplica).

QUESTÕES ADICIONAIS

Idade: idade do alunos (1 – para idade até 20 anos, 2 – para idade entre 21 e 25 anos, 3 – para idade entre 26 e 30 anos, 4 – para idade acima de 30 anos).

Curso: código do curso do aluno.

Curso - Disc: código do curso da disciplina freqüentada pelo aluno.

Campus: Campus do aluno (1 – Tubarão, 2 – Araranguá, 3 – Palhoça).