

Avaliação da Qualidade de Software Através de Técnicas da Mineração de Dados

Simone Vasconcelos Silva

Coordenação de Informática – Centro Federal de Educação Tecnológica de Campos
(CEFETCampos) – Campos – RJ – Brasil

simonevsinfo@yahoo.com.br

Abstract: *This work approaches the evaluation of the quality of software based on the satisfaction of the users through the use of the data mining. The improvement of the quality of software is proposal in accordance with the extracted useful knowledge of diverse bases through the data mining. The present work uses the concepts of the two boarded areas, quality of software and data mining, besides considering a methodology of evaluation of the quality of software products, being this validated through a case study.*

Resumo: *Este trabalho aborda a avaliação da qualidade de software baseada na satisfação dos usuários através da utilização da mineração de dados. A melhoria da qualidade do software é proposta de acordo com os conhecimentos úteis extraídos de diversas bases através da mineração de dados. O presente trabalho utiliza os conceitos das duas áreas abordadas, qualidade de software e mineração de dados, além de propor uma metodologia de avaliação da qualidade de produtos de software, sendo esta validada através de um estudo de caso.*

1. Introdução

Atualmente a avaliação da qualidade de produtos ou serviços tornou-se uma necessidade, visto que os usuários estão cada vez mais exigentes quanto à qualidade dos produtos ou serviços adquiridos. E na visão do usuário um produto de qualidade é aquele que atende as suas necessidades, que seja fácil de usar e que funcione no seu ambiente organizacional [Silva 2003].

Na indústria de software é crescente a preocupação com metodologias de produção de software que permitam a qualidade dos softwares e a satisfação do usuário. Um dos importantes aspectos da qualidade de software, que tem merecido crescente atenção é a qualidade da interação entre o usuário e os softwares. A qualidade de software pode ser dividida em duas partes: qualidade do processo e qualidade do produto [Pressman 2001]. Neste trabalho será abordada a qualidade do produto.

Atualmente a indústria de software, em todo o mundo, vivência um processo de preocupação crescente com metodologias de produção de software que permitam o controle da qualidade dos programas e a satisfação do usuário, dentre outras características [Pressman 2001].

As técnicas de avaliação da qualidade de software podem vir a auxiliar o setor de desenvolvimento de software a alcançar um grau de satisfação maior dos usuários em relação aos produtos desenvolvidos. E assim, aproximar as características de seus produtos de software às demandas dos usuários.

A informação vem desempenhando um papel fundamental em todos os setores da sociedade. O desenvolvimento e sucesso das organizações baseiam-se na capacidade de coletar, tratar, interpretar e utilizar a informação de forma eficaz.

A rápida evolução dos recursos computacionais, ocorrida nos últimos anos permitiu que fossem gerados grandes volumes de dados. O crescimento do volume de dados tem gerado a necessidade de novas técnicas e ferramentas capazes de transformar dados em informações significativas e em conhecimento. Em resposta a essa necessidade, surgiu a Mineração de Dados (*Data Mining*).

Data Mining é uma tecnologia que emergiu da intersecção de três áreas: estatística clássica, inteligência artificial e aprendizado de máquina. *Data Mining* é parte de um processo chamado KDD (*Knowledge Discovery in Databases*), ou seja, Descoberta de Conhecimento em Bases de Dados, que, permite a extração de conhecimento previamente desconhecido e potencialmente útil de um banco de dados [Pampa 2003].

Este trabalho é uma aplicação das técnicas e ferramentas de extração inteligente de conhecimento em bases de dados na avaliação de qualidade de produtos de software.

2. Qualidade de Software

Qualidade de software (QS) é a conformidade do produto com os requisitos funcionais e de desempenho explicitamente declarados, a padrões de desenvolvimento claramente documentados e a características implícitas que são esperadas de todo software profissionalmente desenvolvido [Pressman 2001].

A avaliação da qualidade de software pode ser realizada em dois momentos: durante a geração do software e após este estar pronto para o uso, chamando esses dois momentos, respectivamente, de processo e produto [Rocha et. al. 2001]. No primeiro momento procura-se avaliar de que forma o software está sendo desenvolvido, identificando práticas que possam conduzir a problemas na qualidade do produto e desenvolvendo e/ou utilizando métodos e ferramentas que evitem esses problemas. No segundo momento, como produto concluído, procura-se avaliar a sua qualidade a fim de identificar deficiências e limitações em sua aplicabilidade como um produto final.

A norma internacional ISO/IEC 9126-1 define qualidade de software como “A totalidade de características de um produto de software que lhe confere a capacidade de satisfazer necessidades explícitas e implícitas” [ISO/IEC 9126-1 1999]. Necessidades explícitas são as condições e objetivos propostos por aqueles que produzem o software. A qualidade em uso é a satisfação das necessidades em um contexto específico. Muitas vezes a qualidade em uso vai de encontro às necessidades implícitas e devem permitir aos usuários atingir metas com efetividade, produtividade, segurança e satisfação em um contexto de uso especificado.

2.1. Norma de Qualidade do Produto de Software

A norma ISO (*International Organization for Standardization*) / IEC (*International Electrotechnical Commission*) 9126 - Tecnologia da Informação – Características da Qualidade de Software e Métricas, propõe um enquadramento no qual é definido um conjunto de características que permitem avaliar a qualidade de um produto de software.

Essa norma é dividida em 4 partes: ISO/IEC 9126-1: Modelo de qualidade, ISO/IEC 9126-2: Métricas externas, ISO/IEC 9126-3: Métricas internas e ISO/IEC 9126-4: Métricas de qualidade em uso.

A ISO/IEC 9126-1 permite a avaliação da qualidade do produto final desenvolvido através de um conjunto de características e sub-características que devem ser verificadas, e certificam o produto de software quanto à sua qualidade. A norma apresenta um conjunto de seis características que devem estar presentes em um produto de software de qualidade, conforme mostra a tabela abaixo.

Tabela 1. Definição das características e sub-características

Característica	Conceito	Sub-característica
Funcionalidade	Funções que satisfazem as necessidades explícitas e implícitas.	Adequação, Acurácia Conformidade, Segurança de acesso e Interoperabilidade.
Confiabilidade	Capacidade do software de manter o nível de desempenho sob determinadas condições.	Maturidade, Tolerância à falhas, Recuperabilidade, e Conformidade.
Usabilidade	Esforço necessário para se utilizar o software, bem como o julgamento individual desse uso, por um conjunto de usuários.	Operacionalidade, Atratividade Apreensibilidade, Conformidade e Inteligibilidade.
Eficiência	Relacionamento entre o nível de desempenho do software e a quantidade de recursos usados.	Comportamento em relação ao tempo e em relação aos recursos, e Conformidade.
Manutenibilidade	Capacidade do software de ser transferido de um ambiente para outro.	Testabilidade, Estabilidade, Conformidade, Analisabilidade e Modificabilidade.
Portabilidade	Esforço necessário para fazer modificações especificadas no software.	Adaptabilidade, Capacidade para ser instalado, Coexistência, Capacidade para substituir e Conformidade.

A norma estabelece que cada característica seja dividida em sub-características, conforme a tabela acima, e que as sub-características sejam divididas em atributos ou critérios, os quais são definidos pela equipe de avaliação.

3. Mineração de Dados

A mineração de dados é a análise inteligente e automática de dados para descobrir padrões ou regularidades em grandes conjuntos de dados, através de técnicas que envolvem métodos matemáticos, algoritmos baseados em conceitos biológicos, lingüísticos e heurísticos [Coutinho 2003].

Pode-se dizer que *Data Mining* (ou mineração de dados) é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais [Fayyad e Uthurusamy

2002]. *Data Mining* (DM) vai além da simples consulta a um banco de dados, no sentido de que permite aos usuários explorar e inferir informação útil a partir dos dados, descobrindo relacionamentos escondidos no banco de dados. Pode ser considerada uma forma de descobrimento de conhecimento em bancos de dados (KDD - *Knowledge Discovery in Databases*), área de pesquisa de bastante evidência no momento, envolvendo Inteligência Artificial e Banco de Dados.

O *Data Mining* descende fundamentalmente de três linhagens: estatística clássica, inteligência artificial (IA) e o aprendizado de máquina (casamento entre a estatística e a IA) [Pampa 2003]. Essas técnicas são usadas juntas para estudar os dados e achar tendências e padrões nos mesmos. Hoje, o DM tem experimentado uma crescente aceitação nas ciências e nos negócios que precisam analisar grandes volumes de dados e achar tendências que eles não poderiam achar de outra forma.

3.1. Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery in Database - KDD*)

É o processo que envolve a automação da identificação e do reconhecimento de padrões em um banco de dados. KDD é a extração da informação interessante ou padrões dos dados em bases de dados grandes [Fayyad et. al 1996]. KDD é a tecnologia através da qual se pode extrair informação válida, previamente desconhecida e de máxima abrangência a partir de quantidades grandes de dados armazenados em bases de dados.

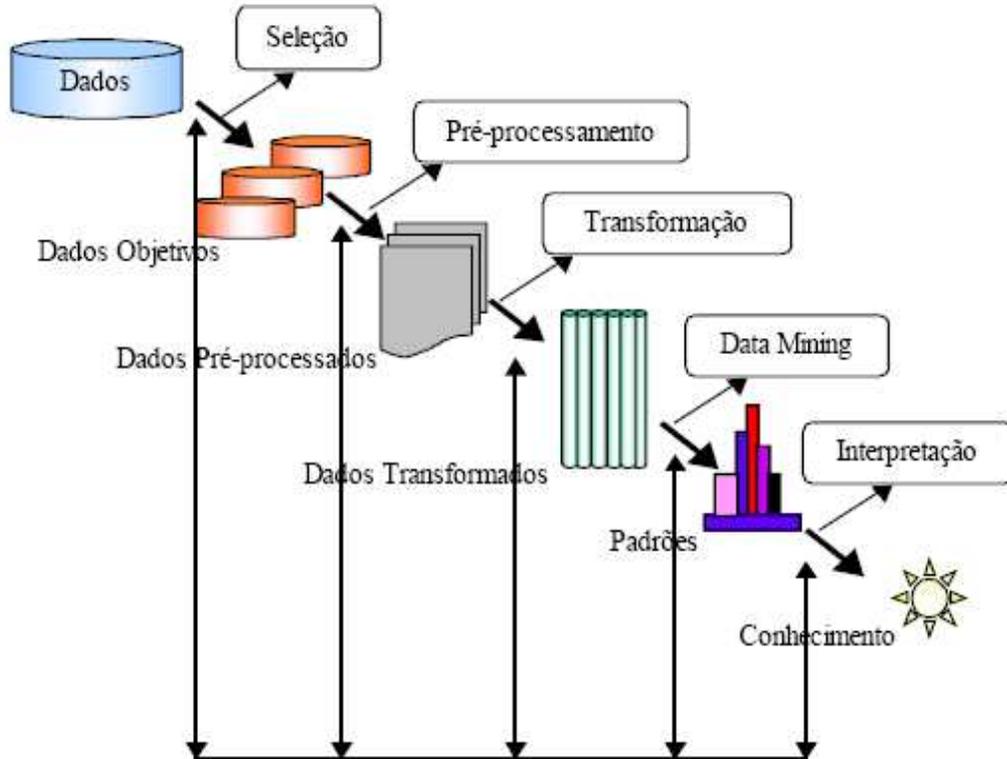


Figura 1. Etapas do KDD [Pampa 2003]

De acordo com Pampa (2003), as etapas do KDD são:

- **Seleção:** nesta etapa destaca-se o conhecimento prévio dos objetivos de um usuário final e o conhecimento do domínio, e nesta etapa é preciso compreensão do domínio, conhecimento dos dados relevantes, definição do problema em termos de domínio, definição de objetivos da aplicação e metas específicas, e definição do modelo da solução desejada.
- **Pré-Processamento:** atividades que visam gerar um grupo de dados representativos convenientemente organizados e estruturados para ser minerados pelo algoritmo selecionado, e esta etapa é composta por integração de dados (múltiplas fontes de dados heterogêneos podem ser integradas em um único BD), seleção de dado (identifica um subconjunto de atributos onde será minerado, facilitando o trabalho dos algoritmos de mineração), e limpeza de dados (estratégia adequada de manipulação de dados ruidosos, errôneos, perdidos ou irrelevantes).
- **Transformação de dados:** os dados são transformados ou consolidados em formatos apropriados para minerar, e consiste em discretizar dados (os valores contínuos dos atributos são divididos numa lista de intervalos, convertendo valores contínuos em valores discretos), com isso obtém uma melhora da compreensão do conhecimento, redução do tempo de processamento, diminuição do espaço de busca, facilitação do algoritmo de tomada de decisões, agregação de dados (agrega dados existentes nas informações de modo que essas agregações contribuam no processo de descoberta de conhecimento), derivação de dados (adicionam-se novos dados derivados por uma operação ou por séries de operações de dados existentes na tabela de dados), e redução de dados (reduz-se o número de variáveis a considerar).
- **Processamento:** os dados são analisados por um algoritmo e transformados em informações (resultados, padrões) úteis que serão avaliados no processo seguinte e consiste em escolha da função de mineração de dados, seleção de algoritmo de mineração de dados, e Data Mining (análise dos dados armazenados através de um programa computacional capaz de analisar os dados e encontrar padrões de interesse).
- **Pós-Processamento:** são feitas a avaliação e interpretação dos padrões, para serem representados em forma de conhecimento compreensível e confiável ao usuário, e para serem incorporadas ao conhecimento anterior e consiste em interpretação e avaliação dos resultados (identifica os padrões interessantes que representam o conhecimento baseado em algumas medidas de interesse), apresentação e incorporação de conhecimento (são usadas técnicas de representação e visualização para apresentar o conhecimento de forma compreensível ao usuário).

3.2. Tarefas do KDD

Dependendo dos objetivos da aplicação, a mineração de dados pode realizar múltiplas tarefas. Cada tarefa tem como base um conjunto de algoritmos que são utilizados na extração de relações relevantes em um, ou vários conjuntos de dados [Pampa e Aguayo 2003].

Este trabalho apresentará algumas tarefas, tais como:

- **Associação:** Visa determinar relacionamentos entre conjuntos de itens ou regras de associação, identifica afinidades entre itens de um subconjunto de dados. Essas

afinidades são expressas na forma de regras. A porcentagem de ocorrência representa o fator de confiança da regra, e costuma ser usado para eliminar tendências fracas, mantendo apenas as regras mais fortes. Regras de Associação representam um padrão de relacionamento entre itens de dados do domínio da aplicação que ocorre com uma determinada frequência na base. A associação utiliza técnicas como algoritmos de extração de regras de associação.

- **Classificação:** É uma função que consiste na aplicação de um conjunto de exemplos pré-classificados para desenvolver um modelo capaz de classificar uma população maior de registros. É uma função de previsão que pode ser usada para encontrar um modelo que classifique um item de dado entre várias classes previamente definidas. Os algoritmos de classificação incluem árvores de decisão, estatística, redes neurais, algoritmos genéticos, etc. [Aurélio et. al. 1999]. A classificação identifica, entre um conjunto pré-definido de classes, aquela a qual pertence um elemento, a partir de seus atributos.

- **Agregação e segmentação (*Clustering*):** É a função descritiva cujo objetivo é identificar um conjunto de categorias finitas ou agrupamentos naturais. Estas categorias podem ser mutuamente exclusivas ou podem consistir em uma representação hierárquica. A proposta de segmentação é basicamente endereçada a problemas de agrupação, na qual se faz um “corte” de um grande número de atributos em um pequeno conjunto de grupos ou de segmentos relativos. A segmentação é realizada automaticamente por algoritmos que identificam características em comum e particionam o espaço n -dimensional definido pelos atributos. Análise de Clusters é o resultado da identificação de um conjunto finito de categorias (ou grupos - clusters) que contêm objetos similares. A clusterização utiliza técnicas como algoritmos de clusterização e redes neurais.

4. Metodologia para Avaliação da Qualidade do Software Através da Mineração de Dados

Com base tanto na literatura quanto em experiências práticas, propõe-se uma metodologia na qual a qualidade do produto de software será avaliada de acordo com o grau de satisfação dos usuários, ou seja, de acordo com a qualidade da interação entre os usuários e os produtos de software. E para extrair conhecimento desta avaliação será utilizada a Mineração de Dados.

O objetivo da metodologia é definir quais critérios precisam ser melhorados para garantir a qualidade do software e quais grupos de usuários possuem maior ou menor dificuldade para possível correção direcionada as reais necessidades.

A metodologia é baseada em três etapas, tais como, seleção dos critérios que serão avaliados no software; seleção da técnica de avaliação que será utilizada; e tratamento das etapas do KDD.

4.1. Critérios e Técnica de Avaliação Utilizados no Software

Seleciona-se as características e sub-características presentes na ISO 9126-1 que sejam facilmente percebidas pelos usuários comuns.

Estas características e suas sub-características selecionadas da norma, assim como os critérios atribuídos a cada característica utilizada para avaliação encontram-se na tabela abaixo.

Tabela 2. Características, sub-características e critérios para avaliação do software

Características	Sub-características	Critérios
Funcionalidade	Acurácia e Segurança de acesso	Precisão, Validação dos Dados e Segurança
Confiabilidade	Tolerância à falhas e Recuperabilidade	Recuperação de Dados e Resistência aos Erros.
Usabilidade	Operacionalidade, Atratividade, Apreensibilidade e Inteligibilidade	Help, Navegação, Documentação, Glossário, Aproveitamento de Dados, Mensagens e Padronização
Eficiência	Comportamento em relação ao tempo	Tempo de Processamento

A técnica de avaliação de satisfação do usuário que será utilizada é a pesquisa de opinião através de questionários, por apresentar baixo custo, rapidez de aplicação e dispensar a participação de especialistas. Os dados coletados através do questionário serão armazenados em um banco de dados. O questionário é composto das seguintes informações:

- Em relação ao usuário: nome, matrícula, setor, cargo e tempo de experiência com o software (em anos);
- Em relação aos critérios do software (conforme Tabela 2): cada critério poderá ser classificado pelo usuário como Muito Bom (MB), Bom (B), Médio (M), Ruim (R), Péssimo (P) ou Inexistente (I);
- Em relação a satisfação do usuário com o uso do software: o usuário poderá classificar o nível de satisfação como alto, médio ou baixo.

4.2. Etapas do KDD

- Seleção: domínio utilizado para geração de conhecimento, e serão utilizados os dados coletados através do questionário.
- Pré-Processamento: serão eliminados os campos referentes às informações dos usuários, como nome e matrícula, pois não influenciaram no conhecimento da qualidade do software utilizado. Eliminação dos registros que possuem campos com resposta nula, pois contribuem negativamente para a obtenção de resultados úteis.
- Transformação: O campo tempo de experiência que é medido em anos, deverá ser transformado em valor nominal obedecendo aos níveis (novato, médio, experiente). Para isso será necessário obter o intervalo entre o tempo mínimo e máximo da amostra e dividi-lo em três faixas. Na maioria dos casos a base de dados também sofre transformação no seu formato, de acordo com o software de mineração que será utilizado.
- Mineração de dados: para uma análise mais precisa e abrangente, propõe-se a utilização de três tarefas de mineração (classificação, clusterização e associação) com o objetivo de obter algum conhecimento útil, pois nem sempre uma única tarefa é capaz

de retirar de uma base os conhecimentos que se deseja ou necessita. A tabela abaixo define as tarefas, técnicas e algoritmos propostos para utilização.

Tabela 3. Tarefas, técnicas e algoritmos propostos

Tarefas	Técnicas	Algoritmos
Classificação	Árvores de Decisão	J48
Associação	Algoritmos de Regras de Associação	APRIORI
Clusterização	Algoritmos de Clusterização	SimpleKmeans

Os algoritmos propostos são baseados em suas inúmeras utilizações em casos práticos e citações na literatura. A ferramenta utilizada poderá ser única, caso ofereça todos os algoritmos citados acima, ou um conjunto de ferramentas diversas de modo que todos os algoritmos citados acima sejam executados.

- Pós-Processamento: análise e interpretação dos resultados, abordando apenas aqueles que sejam de interesse para avaliação da qualidade. Para a tarefa da classificação espera-se identificar o nível de satisfação do usuário por tempo de experiência no software, o nível de satisfação do usuário por cargo e o nível de satisfação do usuário por setor. Para a tarefa da associação espera-se identificar regras que relacionem os critérios ao nível de satisfação do usuário, o tempo de experiência no software ao nível de satisfação do usuário, setor o nível de satisfação do usuário por e setor o nível de satisfação do usuário. Para a tarefa de clusterização espera-se identificar os grupos de critérios mais próximos de modo que possa observar o grupo relacionado a satisfação alta, media ou baixa. Identificar os grupos de critérios mais próximos de modo que possa observar o grupo relacionado a experiência novato, médio, experiente. Identificar os grupos de critérios mais próximos de modo que possa observar o grupo relacionado a cada setor e a cada cargo.

5. Estudo de Caso

O estudo de caso foi utilizado apenas como um meio de ilustrar a metodologia apresentada no capítulo anterior. O produto de software avaliado foi o Sistema de Gestão Hospitalar (SGH) utilizado no Hospital Álvaro Alvin, localizado na cidade de Campos - RJ. Participaram da pesquisa 58 usuários do SGH. Os setores do hospital que participaram da pesquisa foram: Laboratório, Ultra, Farmácia, Ambulatório, Secretaria, Núcleo de Processamento de Dados (NPD), SAME (Setor de Atendimento aos Médicos), Compras, Financeiro e Prontuário. A base de dados que contém os dados coletados pelo questionário de avaliação é utilizada desde 2002.

5.1. Etapas do KDD

- Seleção: a base de dados que contém as informações dos questionários de avaliação.
- Pré-Processamento: foram eliminados oito registros que possuíam campos com respostas nulas O campo cargo foi eliminado pois os funcionários do hospital são divididos em quatro cargos (médicos, enfermeiros, direção e administrativos) e participaram da pesquisa apenas usuários que pertencem ao cargo administrativo.
- Transformação: o menor tempo na amostra foi de 1 ano e 2 meses e o maior de 12 anos. Então foi feito um intervalo de 1 a 12 anos, este intervalo dividido em três faixas,

gerou: $1 \leq \text{tempo} \leq 4$ (Novato), $4 < \text{tempo} \leq 8$ (Médio), e $8 < \text{Tempo} \leq 12$ (Experiente). Os dados do arquivo Questionário.mdb (criado em Access) foram exportados para um arquivo texto. Após a limpeza e transformação a base para mineração contém 16 atributos nominais e 50 instâncias. O arquivo texto foi transformado em um arquivo ARFF (Questionário.arff), que será utilizado na mineração de dados, pela necessidade criada pela ferramenta utilizada.

- **Mineração de Dados:** utilização da ferramenta WEKA para mineração de dados. A ferramenta WEKA (*Waikato Environment for Knowledge Analysis*) tem sido bastante utilizada na realização da etapa de *Data Mining*. Desenvolvida na linguagem Java, pela Universidade de Waikato na Nova Zelândia, esta ferramenta é de Domínio Público e trabalha com diversas Técnicas de *Data Mining*. Ela é composta de dois pacotes que podem ser embutidos em outros programas escritos em Java, permitindo que um desenvolvedor possa criar seu próprio *Data Mining Explorer*. A escolha da ferramenta se justifica por ela apresentar as seguintes características: software livre, facilidade de *download*, facilidade de instalação, usabilidade e quantidade de algoritmos. A ferramenta WEKA e a sua documentação encontram-se no site www.cs.waikato.ac.nz/ml/weka.

```
@relation SGH
@attribute Setor {Farm,Amb,Sec,SAME,Lab,NPD,Ultra,Finan,Fat,Comp,Pront}
@attribute TempoExp {Novato, Médio, Experiente}
@attribute Help {MB,B,M,R,P,I}
@attribute Mensagem {MB,B,M,R,P,I}
@attribute Navegação {MB,B,M,R,P,I}
@attribute Padronização {MB,B,M,R,P,I}
@attribute Precisão {MB,B,M,R,P,I}
@attribute Validação {MB,B,M,R,P,I}
@attribute AproveitarDds{MB,B,M,R,P,I}
@attribute RecuperarDds {MB,B,M,R,P,I}
@attribute ResistirErros{MB,B,M,R,P,I}
@attribute Segurança {MB,B,M,R,P,I}
@attribute TempoProcess {MB,B,M,R,P,I}
@attribute Documentação {MB,B,M,R,P,I}
@attribute Glossário {MB,B,M,R,P,I}
@attribute Satisfação {Alta,Média,Baixa}
@data
Farm,Médio,MB,B,B,B,B,M,M,R,R,MB,B,I,I,Média
Amb,Novato,B,MB,B,MB,B,M,M,R,P,MB,B,I,I,Média
...
```

Figura 2. Arquivo questionário.arff

- **Pós-Processamento:** Análise e interpretação dos resultados.

Resultado e Análise 1:

Tarefa	Técnica	Algoritmo
Clusterização	Algoritmo de Clusterização	SimpleKmeans

Com K=3 - Atributos ignorados: setor e tempo de experiência:

A Satisfação é considerada Baixa quando sete critérios (Help, Mensagem, Navegação, Padronização, Precisão, Segurança e Tempo de Processamento) são considerados Bons, dois critérios (Validação dos Dados e Aproveitamento de Dados) são considerados Médios, um critério (Recuperação de Dados) é considerado Ruim e um critério (Resistência aos Erros) é considerado Péssimo. A Satisfação é considerada

Média quando dois critérios (Help e Segurança) são considerados Muito Bons, cinco critérios (Mensagem, Navegação, Padronização, Precisão e Tempo de Processamento) são considerados Bons, dois critérios (Validação dos Dados e Aproveitamento de Dados) são considerados Médios e dois critérios (Recuperação de Dados e Resistência aos Erros) são considerados Ruins. A Satisfação é considerada Alta quando quatro critérios (Help, Mensagem, Precisão e Segurança) são considerados Muito Bons, três critérios (Navegação, Padronização e Tempo de Processamento) são considerados Bons, um critério (Validação dos Dados) é considerado Médio, dois critérios (Aproveitamento de Dados e Recuperação de Dados) são considerados Ruins e um critério (Resistência aos Erros) é considerado Péssimo. Pode-se verificar que os critérios Documentação e Glossário foram considerados inexistentes em todas as instâncias, e que a maioria dos usuários classificou o nível de satisfação como Médio (50%).

Com K=3 - Atributo Ignorado - Setor:

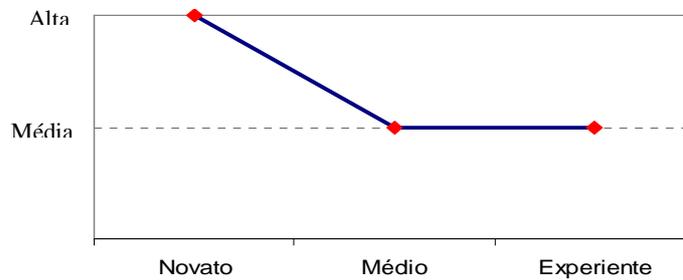


Figura 3. Nível de satisfação dos usuários em relação ao tempo de experiência

O nível de satisfação dos usuários em relação ao tempo de experiência dos mesmos, sofre declínio do Novato para o Experiente, devido a este conhecer mais detalhes do software e das tarefas, o que o torna mais exigente quanto a sua utilização, conforme a figura acima.

Com K=11 - Atributo Ignorado - Tempo de Experiência:

Verificou-se que não foram encontrados clusters que tivessem um significado relevante para a análise. Com o objetivo de prover algum tipo de informação útil, foi selecionado o critério setor e a partir daí repetido o algoritmo de clusterização. Com isso obteve-se resultados significativos, mas descaracterizou-se a tarefa de clusterização, a qual passou a ser apenas uma consulta. O resultado dos clusters apresentado abaixo segue os seguintes critérios respectivamente: Help, Mensagem, Navegação, Padronização, Precisão, Validação, AproveitarDds, RecuperarDds, ResistirErros, Segurança, TempoProcess, Documentação, Glossário e Satisfação.

Cluster 0 – Farmácia	(12%):	MB B B MB B M M R P MB B I I Média
Cluster 1 – Prontuário	(6%):	MB B B B MB M M R R B MB I I Média
Cluster 2 – Financeiro	(12%):	MB MB B B MB M R R P MB B I I Alta
Cluster 3 – SAME	(6%):	MB B MB B B M R R R B B I I Alta
Cluster 4 – Compras	(12%):	MB B B B B M M R R MB B I I Média
Cluster 5 – NPD	(8%):	B B MB B B M R R R B B I I Média
Cluster 6 – Ambulatório	(8%):	B MB B MB B M M R P MB B I I Média
Cluster 7 – Faturamento	(6%):	MB B B B M R R R B B I I Média
Cluster 8 – Ultra	(4%):	MB MB B B MB M R R P MB B I I Média
Cluster 9 – Secretaria	(10%):	MB B B MB B M M R P MB B I I Alta
Cluster 10 – Laboratório	(16%):	B B B B B P M R P B B I I Baixa

Figura 4. Resultado dos clusters

Pode-se verificar que a qualidade dos critérios variou de acordo com o setor de utilização do software. O nível de satisfação dos usuários, em utilizar o software, também variou de acordo com o setor. Pode-se verificar também a necessidade de melhoria no software, relacionada com as dificuldades apresentadas por cada setor.

A maioria dos setores classificou o nível de satisfação como Médio em relação ao uso do software, conforme figura abaixo.

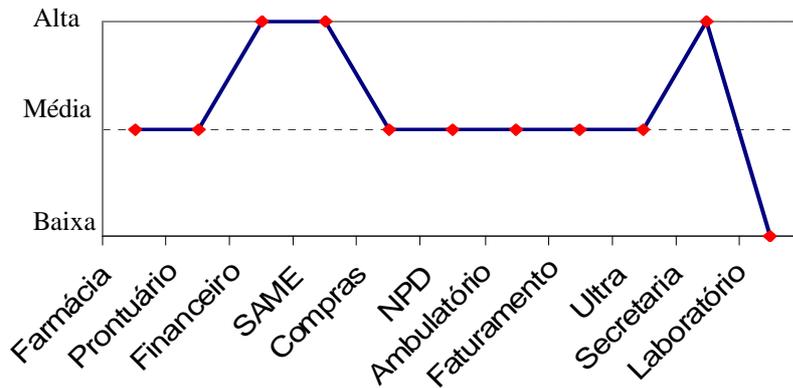


Figura 5. Qualidade dos critérios de acordo com o setor de utilização do software

Resultado e Análise 2:

Tarefa	Técnica	Algoritmo
Associação	Algoritmo de Associação	Apriori

As regras mais importantes foram:

- Satisfação=Alta ==> Help=MB conf:(1)
- TempoExp=Experiente ==> Satisfação=Baixa conf:(0.78)
- Satisfação=Alta ==> TempoExp=Novato conf:(0.77)

Analisando os resultados acima:

- 100% dos usuários que classificaram o nível de satisfação na utilização do software como Alto, consideram o Help do software Muito Bom.
- 78% dos usuários considerados Experientes classificaram o nível de satisfação na utilização do software como Baixo.
- 77% dos usuários que classificaram o nível de satisfação na utilização do software como Alto, são considerados usuários Novatos.

O help teve uma influência positiva na satisfação dos usuários. E o nível de satisfação dos usuários em relação ao tempo de experiência sofre declínio do Novato (Alto) para o Experiente (Baixo), devido a este conhecer mais detalhes do software e das tarefas, o que o torna mais exigente quanto a sua utilização.

Resultado e Análise 3:

Tarefa	Técnica	Algoritmo
Classificação	Árvore de Decisão	J48

Critérios: Tempo de Experiência e Satisfação:

Assim como na Clusterização e na Associação, através da figura abaixo, verifica-se que o nível de satisfação decresce do Novato para o Experiente também na Classificação.

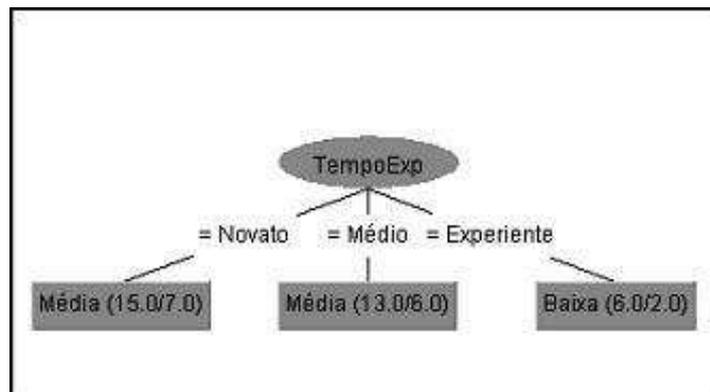


Figura 6. Árvore de decisão (tempo de experiência x satisfação)

Critérios: Setor e Satisfação:

Assim como na Clusterização, verifica-se que o nível de satisfação dos usuários varia de acordo com o setor também na Classificação. Sendo que na árvore de decisão houve um empate entre os usuários que consideraram o nível de satisfação com o software Médio e Alto, conforme figura abaixo.

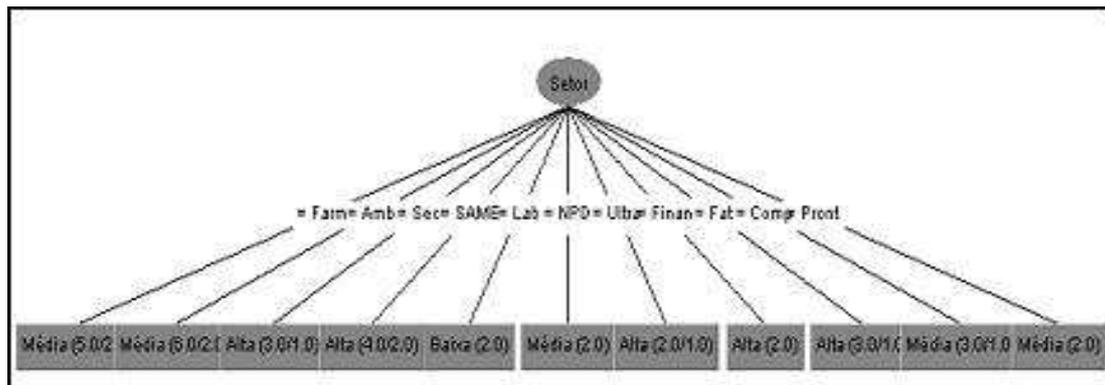


Figura 7. Árvore de decisão (setor x satisfação)

Resumo das análises:

Com as experiências adquiridas pelo processo de descoberta de conhecimento é possível diagnosticar os critérios do software que precisam ser melhorados, verificar as tarefas do software referentes a quais setores precisam de melhoria na interação com o usuário e propor resultados estratégicos para a equipe de desenvolvimento do software, orientando em tomadas de decisões que poderão garantir uma maior qualidade do software e satisfação de seus usuários.

6. Considerações Finais

As informações obtidas após a mineração dos dados servirão para subsidiar as tomadas de decisões e para aperfeiçoar os softwares avaliados, bem como para determinar a satisfação dos usuários, descobrindo falhas e identificando pontos onde os softwares necessitariam de melhorias.

O tratamento das informações coletadas deve conduzir a uma clara interpretação das reais necessidades de melhoria no software visando aumentar sua qualidade. A mineração de dados utilizada na avaliação da qualidade de software deverá identificar padrões e tomar decisões para garantia da qualidade do software, visando sua permanência no mercado.

Neste trabalho foi possível verificar que a utilização da mineração de dados na avaliação da qualidade de software é bastante eficaz, gerando conhecimento útil para:

- Diagnosticar a qualidade de diferentes produtos de software de acordo com as características de qualidade;
- Identificar a satisfação dos usuários em relação ao software;
- Identificar a necessidade de melhorias no software para garantia da qualidade, selecionando os critérios que precisam de alterações.

Referências

- Aurélio, M.; Vellasco, M.; Henrique L.C. (1999) “Descoberta de Conhecimento e Mineração de Dados”. Notas de aula, Engenharia Elétrica, PUC-RIO. Rio de Janeiro.
- Coutinho, F. (2003) “Data Mining”. Disponível em www.dwbrasil.com.br/dmining.
- Fayyad, U. M.; Uthurusamy, R. (2002) “Evolving Data Mining into Solutions for Insights”, *Communications of the ACM*: vol.45, Nº 8, p. 28-31.
- Fayyad, U. M.; Piatestky-Shapiro, G.; Smith P.; Uthurusamy, R. (1996) “Advances in Knowledge Discovery and Data Mining”. AIPress, The Mit Press.
- ISO/IEC 9126-1 (1999) “Information Technology, Software Product Quality, part 1: quality model”.
- Pampa N. R. (2003) “Técnicas e Ferramentas automáticas e inteligentes para a extração de conhecimento em bases de dados”. Dissertação de tese de mestrado. DSIF, FEEC, UNICAMP.
- Pampa N. R.; Aguayo M. T. (2003) “Aplicação de Data Mining a Dados de Avaliação da Qualidade de Produtos de Software.” Ministério da Ciência e Tecnologia. Centro de Pesquisas Renato Archer. São Paulo.
- Pressman, R. (2001) “Software Engineering - A practitioner's Approach”. 5ª edição, McGraw-Hill.
- Rocha, A., Maldonato, J., Weber, K. (2001) “Qualidade de Software – Teoria e Prática”. 1. ed. São Paulo: Prentice Hall.
- Silva, S. (2003) “Qualidade de Software – Uma Abordagem Baseada na Satisfação do Usuário”. Dissertação (Mestrado em Engenharia de Produção) – Campos-RJ, UENF, 172p.