

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Técnicas de Mineração de Dados para suporte à decisão
na Optimização da Qualidade de Serviço
aplicado à Indústria Cimenteira

Ricardo Manuel da Silva Abreu

Dissertação de Mestrado

2008

Técnicas de Mineração de Dados para suporte à decisão
na Optimização da Qualidade de Serviço
aplicado à Indústria Cimenteira

Ricardo Manuel da Silva Abreu

Dissertação apresentada à Universidade do Minho para obtenção do grau de Mestre em Informática,
elaborada sob orientação do Professor Doutor Orlando Manuel de Oliveira Belo.

2008

Agradecimentos

Este projecto é o culminar de um longo processo de aprendizagem que, a pouco e pouco, se foi tornando mais exigente e rigoroso implicando, naturalmente, maior concentração, dedicação e estímulo para nunca desistir mas sim, continuar a crescer.

Gostaria de agradecer ao Professor Doutor Orlando Belo, pelo apoio prestado e sua pela disponibilidade constante para os esclarecimentos e troca de ideias que se revelaram tão importantes para este trabalho.

À CIMPOR – Cimentos de Portugal, não só pela possibilidade de enquadrar este mestrado no seu seio mas, também, por me ter permitido, ao longo destes últimos anos, conhecer muito acerca da realidade extremamente interessante que é "*o mundo do cimento*".

Ao Eng. Cândido Martins pela confiança em mim prestada desde a primeira entrevista até hoje. Pelos seus conselhos e motivações, foi uma enorme fonte de saber não só profissional mas também, na forma de estar e relacionar com os outros.

A todos os colegas da Cachapuz, grandes "*companheiros de luta*" deste, de outros e novos caminhos a percorrer.

Não poderei deixar de referir a minha família, em particular os meus pais, Delfim Abreu e Rosa Madalena e meu irmão Rui Abreu. São elementos constantemente presentes na minha vida e sempre me servirão de guia.

Por fim, mas não por último, à minha esposa, Sofia, que pelo seu conforto, confiança e ajuda me deu forças para não desistir e continuar até aqui.

Muito obrigado a todos ...

Resumo

Técnicas de Mineração de Dados para Suporte à Decisão na Optimização de Qualidade da Serviço aplicado à Indústria Cimenteira

As necessidades actuais do mundo empresarial e dos seus diversos intervenientes humanos ultrapassa substancialmente aquilo que os sistemas operacionais providenciam. A simples análise de relatórios pré-definidos ou de *queries ad-hoc* mais ou menos sofisticadas não é suficiente, nem sequer se coaduna com as exigências dos mercados extremamente concorrenciais em que as empresas se posicionam.

Com este trabalho pretendeu-se demonstrar as vantagens da aplicação de técnicas de mineração de dados numa área de negócio estrutural das economias actuais: a indústria cimenteira. Através aplicação de tarefas de associação e classificação, analisaram-se relações ou modos operatórios de risco associados a um processo específico: carregamento de produto ensacado.

Os resultados obtidos demonstraram o enorme potencial deste tipo de abordagens no contexto da indústria cimenteira e em sistemas industriais como o Sistema Logístico de Viaturas podendo, assim, contribuir activamente para a optimização da qualidade de serviço deste sector.

Abstract

Data Mining Techniques for Decision Making in the Optimization of the Quality of Service applied to the Cement Industry

The current needs of the enterprise world and their several human interveners substantially overcome what the operational systems provide. The simple analysis of pre-defined reports or ad-hoc queries more or less sophisticated is not sufficient, and doesn't even adjust to the demands of the extremely competitive markets where the companies are positioned.

This work had the objective of demonstrating the advantages of the application of data mining techniques in a structural business area in the current economies: the cement industry. Through the application of association and classification techniques, it was possible to analyze relations or risky operational modes associated to a specific task: loading of bagged product.

The results achieved demonstrated the potential of these kinds of approaches in the cement industry context and in industrial systems like SLV (Logistics Vehicle System), proving its capability to actively contribute for the optimization of the quality of service in this sector.

Índice

Introdução	1
1.1 Contexto.....	1
1.2 Motivação	3
1.3 Objectivos.....	5
1.4 Estrutura do documento.....	6
Indústria Cimenteira	9
2.1 Enquadramento na realidade de negócio	9
2.2 Configuração das instalações.....	11
2.3 Conhecimento dos sistemas de informação	15
2.4 Dash-boards	18
2.4.1 Análise Temporal	19
2.4.2 Análise de Quantidades	23
Integração de técnicas de mineração.....	29
3.1 Processo de Descoberta de Conhecimento	30
3.1.1 Selecção de Dados.....	31
3.1.2 Pré-Processamento e Transformação de Dados.....	33
3.1.3 Aplicação de algoritmos de Data Mining	34
3.1.4 Avaliação de resultados	34
3.2 Tarefas e Técnicas de Mineração	35
3.2.1 Tarefas	35
3.2.2 Técnicas.....	37
3.3 Aplicações.....	41

Preparação dos Dados.....	45
4.1 Seleção de Dados.....	45
4.2 Limpeza de Dados	46
4.2.1 Tratamento de datas de operação	46
4.2.2 Tratamento de nulos pontos de operação.....	48
4.2.3 Ranking de Desvios	50
4.3 Janela de Oportunidade para execução do modelo	51
Qualidade de serviço aplicado ao carregamento a saco	53
5.1 Contextualização	53
5.2 Desenvolvimento de modelos	55
5.3 Análise de resultados	68
Conclusões e Trabalho Futuro.....	71
Bibliografia.....	75
Referências WWW	79

Índice de Figuras

Imagem 1 - Inquérito acerca de metodologia de mineração utilizado nas organizações (Fonte [8])	6
Imagem 2 - Implantação SAP na indústria cimenteira, segundo dados de 2004 (Retirado de [5])	10
Imagem 3 - Configuração de Instalações (Fonte [14])	13
Imagem 4 - <i>Data Mart</i> Global de Operações	17
Imagem 5 - <i>Dash-board</i> qualidade de atendimento temporal	20
Imagem 6 - Tempos atendimento operações de carga (sem. 26 a 28)	21
Imagem 7 - Tempos atendimento global (sem. 26 a 40)	21
Imagem 8 - Variação semanal dos índices de atendimento	23
Imagem 9 - Variação horária dos índices de atendimento	23
Imagem 10 – <i>Dash-board</i> qualidade de carregamento	24
Imagem 11 – Desvios de quantidade em operações de ensacado (sem. 36 a 40)	25
Imagem 12 – Desvios de quantidade por linha em operações de ensacado (sem. 36 a 40)	26
Imagem 13 - Variação horária de quantidades executadas (semana 36 a 40)	26
Imagem 14 - Variação semanal dos desvios executadas (Semana 36 a 40)	27
Imagem 15 - Variação horária dos desvios executadas (Semana 36 a 40)	27
Imagem 16 - Fases do processo de descoberta de conhecimento (Adaptado de [Fayad et al., 1996])	31
Imagem 17 - Árvore de Decisão para um problema de classificação do tempo de espera para carregamento de viaturas	38
Imagem 18 – Rede Neuronal da função $Z = 3X + 7Y$ (Adaptado de [Berry et al., 2000])	39
Imagem 19 – Ilustração da técnica algoritmos genéticos de <i>Data Mining</i> (Adaptado [Navega, 2001] e [Botica, 2004])	41
Imagem 20 - Tratamento de datas – Valores por defeito	47
Imagem 21 - Relação de dependências entre datas associadas a um processo de movimentação	48

Imagem 22 - Detalhe do interior de armazém de produto ensacado.....	54
Imagem 23 - Detalhe do tapete de abastecimento de sacos	54
Imagem 24 - Ranking de desvio absoluto associado a operação de carregamento a saco.....	56
Imagem 25 - Relação de rankings de desvios percentuais	56
Imagem 26 - Definição estrutura de mineração base	57
Imagem 27 - <i>Cube slice</i> para estrutura de mineração	58
Imagem 28 – Visão parcial de modelos de mineração implementados	59
Imagem 29 – Lista de regras de maior importância na aplicação do modelo <i>Carregamento_Assoc</i>	60
Imagem 30 - Lista de regras de menor importância na aplicação do modelo <i>Carregamento_Assoc</i>	60
Imagem 31 - Rede de dependências para o modelo <i>Carregamento_Assoc</i>	61
Imagem 32 – Rede de dependências para quantidades entre [36.97 – 40.45] toneladas.....	61
Imagem 33 - Dispersão de viaturas segundo intervalos de quantidades do modelo.....	62
Imagem 34 - Rede de dependências para ranking de desvio percentual "4"	62
Imagem 35 - Rede de dependências para ranking de desvio absoluto "5"	63
Imagem 36 - Rede de dependências para ranking de desvio absoluto "4"	63
Imagem 37 - Rede de dependências para ranking desvio absoluto "-5"	64
Imagem 38 - Rede de dependências para quantidades entre 17.06 e 27.16 toneladas	64
Imagem 39 - Relação de dependências relativas ao modelo <i>Carregamento_Class</i>	65
Imagem 40 - Árvore de decisão relativa ao <i>ranking desvio absoluto</i>	66
Imagem 41 - Regras de decisão para nodo " <i>Mat abs desv >= 0.435 and Mat abs desv < 0.757</i> ".	67
Imagem 42 - Detalhe do nodo quantidade [36.97 – 40.45[.....	68

Índice de Tabelas

Tabela 1 - Presença internacional do grupo Cimpor (Fonte [6])	11
Tabela 2 - Relação entre operações executadas e distribuição temporal	22
Tabela 3 - Normalização de Registos (Adaptado [Pinto et al., 2004]).....	33
Tabela 4 - Relação entre Técnicas e Tarefas de Mineração (Adaptado de [Teixeira, 2005])	35
Tabela 5 - Ranking de desvios percentuais de quantidade	50
Tabela 6 - Ranking de desvios absolutos de quantidade	51
Tabela 7 - Relação de movimentação para operações de carga.....	53
Tabela 8 - Definição da Estrutura de Mineração.....	58

Siglas

ACCC	Amreyah Cimpor Cement Company
AMCC	Amreyah Cement Company
CRISP-DM	Cross Industry Standard Process for Data Mining
ERP	Enterprise Resource Planning
ETL	Extraction, Transformation and Loading
KDD	Knowledge Discovery in Databases
PAC	Plataforma Analítica Cachapuz
SLV	Sistema Logístico de Viaturas
SLV - BI	Sistema Logístico de Viaturas – Módulo Business Intelligence
SQL	Structured Query Language

Capítulo 1

Introdução

1.1 Contexto

O sector da Construção Civil influencia de forma decisiva a estrutura económica de um país. É um sector que consome um grande volume de produtos de outros segmentos industriais e que tem uma ampla cadeia produtiva. Nas primeiras fases do desenvolvimento económico, o sector da construção cresce normalmente de forma mais rápida relativamente a outros sectores, em virtude do processo de industrialização e urbanização pelos quais passam os países em fase de desenvolvimento [Kureski et al. 2007].

A indústria do cimento apresenta-se como um dos pilares base deste sector onde, nos últimos anos, tem concentrado numerosos investimentos em áreas como a automação industrial e controlo de processos, procurando a redução de custos de produção, além dos tradicionais e actuais investimentos em controlo de poluição e gestão ambiental. Comparada com outras indústrias transformadoras, o ciclo de produção do cimento é curto, mas envolve grandes lotes e assenta essencialmente em três grandes vertentes [Wang et al. 2007]:

- Plano Produtivo - A empresa é desenhada para a produção em massa onde a única forma de redução de custos é o funcionamento contínuo em plena carga.

- Recepção e Expedição - Os processos de produção são executados de modo sequencial e contínuo. Para a correcta articulação entre a entrada de matérias-primas e a expedição de produto é fundamental que os níveis de stock se mantenham dentro dos limites aceitáveis assegurando-se um processo de produção constante.
- Qualidade de produto - A qualidade do produto está intimamente ligada com a qualidade das matérias-primas utilizadas no seu processo de fabrico. Devido à constante variação na qualidade das matérias-primas recebidas, o processo produtivo é continuamente ajustado e refinado como forma de garantir os índices de qualidade desejados.

O produto final, ao nível dos diversos fabricantes mundiais, apresenta alta homogeneidade de especificações, não existindo portanto grande diferenciação em termos de qualidade. A concorrência centra-se essencialmente no preço e qualidade no serviço de atendimento, sendo que, neste sentido, custos de transporte, localização das empresas e tempos de atendimento nas instalações, são factores com grande relevo no processo. Neste contexto, é importante para as cimenteiras oferecer um serviço atractivo e fiável para os seus clientes, fornecedores e transportadores. Assim, para fazer face a esta realidade, as empresas tendem a apostar em soluções focalizadas na automatização da logística interna da empresa que permitam:

- Redução de custos relacionados com pessoal, fretes e transportes contratados.
- Operacionalidade do sistema real, 24 sobre 24 horas, sem recurso a operadores.
- Garantia de índices de atendimento de qualidade com redução do tempo médio de serviço.
- Previsão de situações de congestionamento que possam colocar em causa o normal funcionamento das instalações.

O não cumprimento destes requisitos conduz normalmente a um aumento da insatisfação dos clientes, dos fornecedores ou dos transportadores, com natural prejuízo para a empresa. Para tal, torna-se fundamental realizar regularmente análises comparativas dos tempos de operação de cada uma das fases de movimentação das viaturas na fábrica, bem como a análise dos picos de afluência, para que se possam aplicar medidas que assegurem os índices de atendimento esperados. Considerando a organização tradicional das fábricas para expedição de cimento, o processo de carregamento passa pela definição de zonas, ou unidades funcionais, tanto na fábrica como no parque, nas entradas e saídas e nas zonas de carregamento. O motorista chega à fábrica e estaciona a viatura em parque. Nesta fase, será criado o registo de chegada às instalações

ficando o motorista a aguardar a autorização de entrada nas instalações, evitando-se assim congestionamentos em zonas críticas da fábrica, como entradas e saídas, onde a circulação de viaturas é grande e sujeita a acidentes de circulação. A autorização de entrada estará dependente de vários parâmetros de sistema, como horários de recepção de produtos, limites definidos para as zonas de carregamento, etc. O motorista dirige-se para a entrada onde é registada a entrada efectiva na fábrica, podendo em seguida dirigir-se para o local de carregamento. Dada a natureza lenta do processo de carregamento, as fábricas tendem geralmente a disponibilizar diferentes pontos de operação alternativos como forma de permitir o carregamento de várias viaturas em simultâneo. Após o carregamento, o motorista dirige-se para a saída das instalações onde é emitida toda a informação necessária.

Para os gestores, a obtenção de uma relação de equilíbrio de funcionamento não é simples, nem imediata. Se, por um lado, pretendem evitar congestionamentos no interior da fábrica, por outro, a existência de um número elevado de viaturas em parque pode ter impacto sobre a imagem da empresa. Contudo, poderá não existir uma relação directa entre o tempo de espera em parque e a taxa de afluência de viaturas às instalações. Ineficiências no processo de carregamento, como avarias, poderão traduzir-se num aumento do tempo de carregamento reflectindo-se nas fases anteriores, nomeadamente nos tempos de espera em parque. Face a estes cenários, os gestores necessitam de informação adequada com formatos específicos, que lhes permitam, face ao histórico da instalação, gerir e reagir em tempo útil aos parâmetros de funcionamento, focalizado na obtenção de índices de qualidade de serviço aceitáveis com redução de custos. Definição de número de pontos de carga a disponibilizar em cada momento, limites de viaturas no interior da fábrica, são exemplos de parâmetros efectivos, que podem contribuir para o índice da qualidade de serviço da fábrica. Tendo em conta as necessidades das empresas que operam nesta área, poderão ser usadas técnicas de mineração de dados para extracção da informação necessária para os gestores.

1.2 Motivação

A automatização e sistematização dos processos de cargas e descargas em unidades industriais com grande movimentação são uma necessidade fundamental para o sucesso do negócio das empresas. Tempos de espera prolongados, congestionamentos no interior da unidade industrial

juntamente com deficiências na execução dos processos de carga ou descarga, conduzem inevitavelmente a prejuízos e descontentamento dos clientes. Deste modo, o objectivo do sistema SLV é automatizar logística de movimentação interligada com o *ERP (Enterprise Resource Planing)* da unidade industrial, na medida em que os *ERP's* actuais não estão desenhados para solucionar este tipo de problemas. Toda a gestão de entidades, produtos, encomendas e validações de negócio serão da responsabilidade do ERP, enquanto a gestão de movimentação, parque e chamada, operações de carga ou descarga serão da responsabilidade de soluções, como por exemplo, o SLV (*Sistema Logístico de Viaturas*) desenvolvido pela Cachapuz – Bilanciai Group [14].

Afirmando-se como uma solução inovadora no âmbito da automatização logística dos processos, o *SLV* faz uma ruptura com os modos de funcionamento tradicionais baseados na intervenção de operadores. Introduzindo o conceito de auto-serviço, permite aos motoristas efectuarem a operação de carga ou descarga de forma autónoma e com redução de erros. Este pequeno passo, criou uma enorme necessidade de indicadores de processo e decisão sobre os quadros de gestão, onde as ferramentas analíticas se tendem a afirmar como factor diferenciador e de sucesso. Caracterizando-se pela sua simplicidade de utilização, estas ferramentas colocam à disponibilidade dos gestores mecanismos de relacionamento e correlação de dados que, aliados à flexibilidade na criação de perspectivas sobre os dados, tempos de resposta imediatos, independentemente da complexidade ou perspectiva desejada, fomentam a criatividade e astúcia dos gestores na busca e definição de vistas específicas sobre o seu negócio.

Neste sentido, o *SLV* tem vindo a ser estruturado para permitir a incorporação deste tipo de tecnologias na sua Framework base. Para tal, têm vindo a ser efectuadas uma série de etapas na definição da futura *Plataforma Analítica Cachapuz*. Tendo consciência que o sucesso dos projectos passa pela sua aceitação e adesão por parte dos clientes, o processo de concepção de protótipos passou pelo envolvimento de clientes chave nas diferentes áreas, onde o seu contributo acelerou o processo de integração na solução final e testes reais nas suas instalações. A *CIMPOR – Cimentos de Portugal*, é considerada pela *Cachapuz* não um cliente mas um parceiro que tem, ao longo das últimas décadas, participado nesta dinâmica de processo e permitindo à *Cachapuz* adquirir um forte *know-how* desta área industrial. Como resultado, a visibilidade em termos de mercado global, com a instalação da solução nas realidades do mercado nacional ou novos mercados como o do Egipto, onde é notório o crescimento e competitividade do sector. Por outro lado, o facto de o *SLV* se encontrar instalado a nível nacional em todos os operadores do sector e ter como parceiro a

CIMPOR que faz parte do top 10 a nível mundial, confere à solução um nível de alta credibilidade, com potencial para divulgação e alargamento para outros *players*, dentro do sector ou em sectores de movimentação de mercadorias, onde a automatização de operações e funcionamento em regime de auto-serviço 24 H / dia, 365 dias / ano são encarados como a chave para o sucesso.

Inserido neste contexto altamente favorável, foram criadas condições para o desenvolvimento do projecto *SLV-BI*, ou seja, implementação dos processos de *ETL (Extraction, Transformation and Loading)*, que irão servir de suporte às funcionalidades de exploração de dados e afirmação da componente analítica do sistema. No entanto, o *SLV-BI* representa uma parte dos objectivos da *PAC*, faltando a introdução dos processos de mineração. Actuando sobre modelos de análise específicos, buscam padrões nem sempre evidentes ou de fácil análise aos gestores, disponibilizando-os de uma forma simples e intuitiva, nos quais se enquadram os objectivos deste trabalho tendo, também, em consideração uma estratégia de evolução e consolidação do Sistema Logístico de Viaturas desenvolvido pela *Cachapuz* com consequente projecção e divulgação num mercado global.

1.3 Objectivos

Através de um caso de estudo, pretende-se demonstrar possíveis vantagens imediatas com a adopção de técnicas de mineração de dados permitindo, de forma prática e objectiva, contribuir para a melhoria da qualidade de serviço e atendimento de unidades industriais. Focalizando um sector específico, a indústria cimenteira, pretende-se elevar os sistemas de suporte à decisão para um patamar superior. Procurando uma plataforma analítica enriquecida com novos paradigmas de descoberta de conhecimento, pretende-se disponibilizar meios efectivos aos seus gestores no sentido de avaliarem a qualidade e índices de satisfação das unidades industriais. Neste sentido, foram traçados os seguintes pontos específicos:

- Desenvolvimento de modelos de mineração de dados, com ênfase nas técnicas de classificação e associação, que permitam identificar automaticamente os locais e horas onde ocorrem desvios sistemáticos do índice de qualidade de serviço.
- Estudo de possíveis causas que possam dar origem a esses desvios, nomeadamente a afluência de viaturas ou congestionamento nos locais de carga, e desenvolvimento de

modelos de mineração de dados, que permitam relacionar automaticamente estas causas com os seus efeitos na qualidade de serviço.

De modo a assegurar o sucesso deste trabalho, foi efectuada uma análise preliminar sobre ferramentas e metodologias existentes. Embora o leque de escolhas não seja muito alargado, e tendo por base o inquérito apresentado na Imagem 1, a metodologia *CRISP-DM* [9] apresentou-se como a selecção natural para este projecto, dado:

- Ser uma metodologia amplamente adoptada em projectos deste género.
- Ter uma forte preocupação, dada a sua génese, nas realidades e requisitos da indústria.
- Encontrar-se suportada por documentação de boa qualidade.

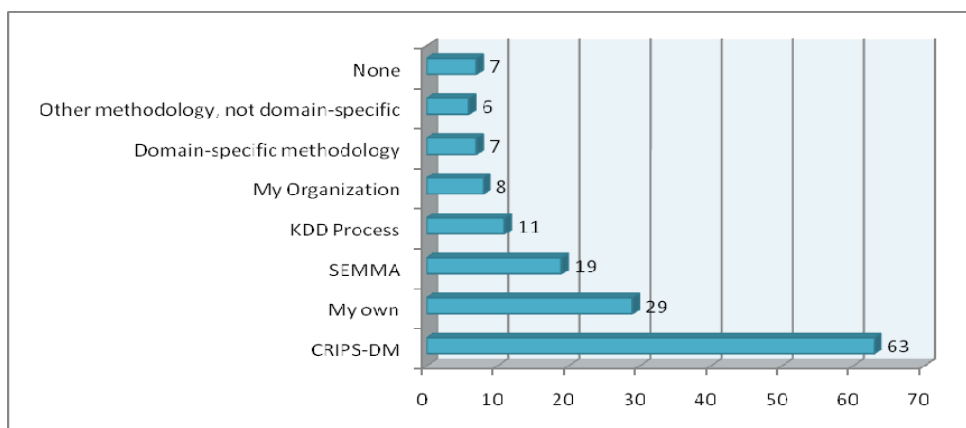


Imagem 1 - Inquérito acerca de metodologia de mineração utilizado nas organizações (Fonte [8])

1.4 Estrutura do documento

Este documento procura, em traços gerais, seguir as fases definidas na metodologia *CRISP-DM*. No entanto, face ao contexto específico em que esta tese decorre, foram efectuadas adaptações e ajustes na metodologia, de modo a que, sem colocar em causa a sua essência, permita ajustar e reflectir as necessidades da tese e os fundamentos da metodologia. Assim, o documento foi estruturado em cinco partes distintas, para além do capítulo actual:

- Indústria Cimenteira – Efectua-se uma contextualização do sistema *SLV* na realidade industrial, identificando-se os processos e a sua relevância na gestão logística, passando pela apresentação da componente *SLV-BI*, como base de trabalho, para a implementação de modelos de mineração. Por fim, termina-se com a apresentação de requisitos funcionais suportados por um conjunto de *dash-boards*, onde se agregam e sumarizam os principais indicadores relacionados com os objectivos da tese.
- Introdução de técnicas de mineração – Apresentam-se diferentes tarefas e técnicas de mineração, assim como os seus conceitos principais. Conclui-se com a uma visão focalizada na indústria cimenteira, onde se demonstra, a aplicabilidade de técnicas de mineração para processos de descoberta de conhecimento a este sector.
- Preparação dos Dados – Nesta secção será efectuada a identificação e apresentação das acções, no sentido de ajustar os dados aos requisitos específicos dos algoritmos e tarefas de mineração, dando especial ênfase às tarefas de limpeza, enriquecimento de dados e janelas de oportunidade para execução dos modelos desenvolvidos.
- Desenvolvimento de Modelos – Detalha-se o processo de implementação dos modelos projectados, utilizando ferramentas de modelação. Tirando partido, sempre que possível, da possibilidade do contacto de proximidade com gestores de negócio, pretendeu-se ajustar o modelo aos seus requisitos e, assim, torná-lo apto para a validação efectiva por parte dos agentes de decisão.
- Conclusões e Trabalho Futuro – Neste capítulo serão efectuadas considerações acerca da utilização de técnicas de mineração e potencialidade de incorporação no seio de soluções industriais projectos com as características do *SLV* lançando-se, também, possíveis linhas de orientação para trabalho futuro.

Capítulo 2

Indústria Cimenteira

2.1 Enquadramento na realidade de negócio

A indústria do cimento apresenta-se como um pilar central na dinamização e sustentabilidade das economias, com um leque de influências extremamente vasto. São vários os estudos e acções que incutem nesta indústria uma forte tendência de continuidade e, até aceleração, nas transformações a nível empresarial e fabril.

Numa vertente mais estratégica, é importante focalizar a globalização e as mudanças empresariais, bem como a expansão por diferentes países, aliada ao crescente avanço tecnológico de fabrico e tecnologias de informação. Isto irá conduzir-nos ao aumento de novos produtos, assim como a uma nova visão sobre os processos de negócios, e reforço das relações das empresas com os seus parceiros. Ainda neste contexto, [Prochnik et al. 1998] enquadra na globalização dos grandes grupos, modelos de distribuição como estratégias para assegurarem a qualidade de serviço e de atendimento ao mercado. Tratando o cimento como um produto diversificado mas ao mesmo tempo especializado para o sector da construção, revela a importância do conhecimento das necessidades de mercado, criação de uma imagem credível, crescente exigência no cumprimento de prazos de entrega, assim como fornecimento de produto com níveis de qualidade acima dos mínimos exigidos pelo cliente.

Tomando como exemplo o processo de expedição, a fidelização do cliente é claramente influenciada pelas condições oferecidas durante a estadia nas unidades industriais: respeito pelas normas de segurança, rigor na logística interna e eficiência nos processos de execução.

Para fazer face a estes requisitos, é imperioso que os gestores se ajustem às exigências do meio envolvente onde, através da adopção de soluções integradas, possam automatizar as unidades industriais, fornecendo indicadores chave, respondendo às rápidas mudanças do mercado, para assim garantir uma maior sustentabilidade do negócio. Dando especial relevância a processos de automatização e integração dos sistemas de informação, pretendem-se soluções verticais que vão desde a componente produtiva até à componente administrativa e de gestão. Relativamente à indústria cimenteira, apresenta-se como um dado importante a afirmação do ERP SAP como solução para a gestão do negócio, de acordo com o apresentado na Imagem 2 [5].

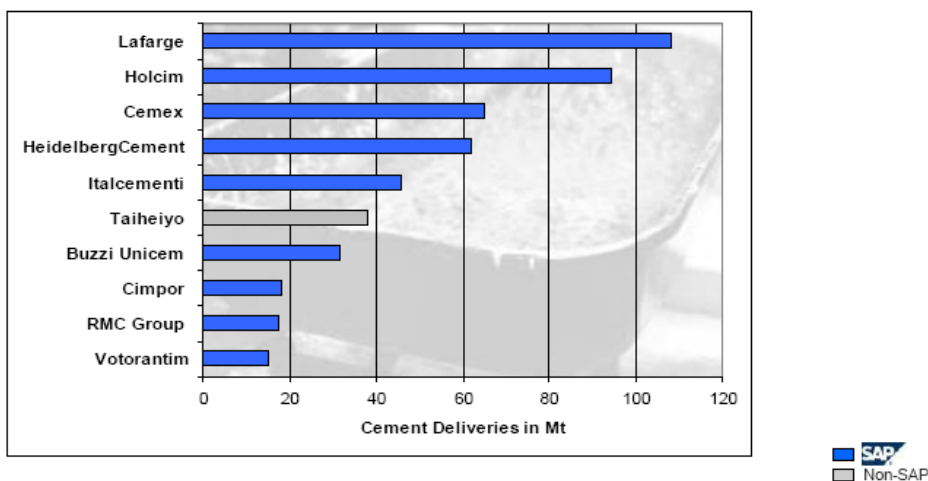


Imagem 2 - Implantação SAP na indústria cimenteira, segundo dados de 2004 (Retirado de [5])

Por outro lado, não pode ser descartada a noção de posicionamento e distribuição geográfica dos grandes grupos onde, tomando como exemplo o caso do *Grupo Lafarge*, já se encontra representado em 96 países, de acordo com o seu site institucional [9]. Esta tendência de globalização verifica-se, também, nos restantes grupos mundiais onde, a título de exemplo, se apresenta o *Grupo Cimpor* onde é notória a sua estratégia de expansão internacional, uma vez que está presente em 10 países para além de Portugal, tal como apresentado na Tabela 1.

Países	
Espanha	Moçambique
Marrocos	Brasil
Tunísia	Egipto
África do Sul	Cabo Verde
Turquia	China

Tabela 1 - Presença internacional do grupo Cimpor (Fonte [6])

Soluções como o *SLV* tendem a afirmar-se como uma mais-valia na realidade de negócio da indústria cimenteira, dado que, os *ERP's* como o *SAP*, não estão vocacionados para a automatização de processos de pesagem e movimentação de viaturas, tão específicas neste sector. Nos últimos anos, o *SLV* tem vindo a ser enriquecido com novas funcionalidades, fruto das inúmeras instalações efectuadas no mercado nacional e internacional, contribuindo significativamente para uma melhoria no desempenho dos processos de logística interna e movimentação das viaturas.

Nesta secção 2.2, detalha-se o modo de funcionamento de uma unidade industrial do cimento, focalizando-se na vertente de automatização logística de recepção e expedição, com uma forte evolução nos sistemas de informação, de forma a disponibilizar indicadores chave para o negócio.

2.2 Configuração das instalações

Os centros de produção procuram ajustar-se de forma a conseguirem responder eficazmente ao fabrico e expedição de cimento, que exige a utilização de grandes quantidades de matérias-primas como o calcário, o gesso, as margas ou mesmo os combustíveis para o funcionamento do forno de cozedura. [Prochnik et al. 1998] refere que estão a verificar-se mudanças nos modelos de transporte através das ligações ferroviárias ou marítimas com o objectivo de baixar os seus custos, no entanto, continua ainda a existir uma forte utilização do camião, sendo notório o seu impacto na organização das fábricas de cimento.

Tal como ilustrado na Imagem 3, as fábricas encontram-se muito moldadas para esta realidade, sendo possível identificar um conjunto de áreas funcionais que regulam as operações de movimentação no seu interior, garantindo níveis de segurança e desempenho necessários para o ciclo de produção e expedição:

1. Parqueamento – Zonas de concentração de viaturas que se encontram em espera para dar entrada nas instalações da fábrica. São geralmente de grandes dimensões iniciando o processo de ordenação de viaturas, dado que, após o aparcamento, se efectua o registo de viatura em espera na fábrica.
2. Portarias (Entradas e Saídas) – Zonas de controlo de acesso às instalações onde, mediante o tipo de operações, se poderá efectuar a obtenção de pesos. A título de exemplo, no carregamento a saco ou descarga de matérias-primas, onde os locais de operação não dispõem tradicionalmente de básculas, a obtenção de primeiro e segundo peso para cálculo do peso líquido processa-se sobre as básculas instaladas nestas zonas, tornando-as extremamente críticas e influentes no funcionamento da instalação. Qualquer perturbação nestas zonas reflecte-se imediatamente num aumento dos tempos de permanência sobre as básculas e, conseqüentemente, com criação de filas para entradas e saídas das instalações.
3. Carregamento a Granel – O processo de carregamento a granel caracteriza-se pelo doseamento "*on-line*" dos camiões cisterna, efectuando-se uma transferência directa de produto dos silos de armazenamento para as viaturas, através de complexos sistemas de automação. Erros na determinação da quantidade pedida, poderão levar a uma sobredosagem da viatura que, no limite, implicarão a descarga de produto em excesso, aumentando de forma directa o tempo da viatura nas instalações, assim como tempos de espera em parque. Devido aos limites da zona de carregamento, a viatura ocupa uma vaga não permitindo, assim, novas entradas até a sua saída.
4. Carregamento a Saco – Ao contrário do que se verifica no carregamento a granel, o cimento ensacado caracteriza-se por um doseamento prévio em unidades singulares (saco) que, posteriormente, será disponibilizado sobre a forma de diferentes tipos de acondicionamento (saco-a-saco, palete, pacotão) dependendo do tipo de automatização,

das características da empresa e dos requisitos do mercado. Erros no doseamento dos sacos, ou um controlo ineficiente dos processos de acondicionamento introduzirão, inevitavelmente, desvios nos processos de fornecimento.

5. Descargas de Matérias-Primas – Os processos de descarga de matérias-primas caracterizam-se, geralmente, por operações de curta duração e em grandes quantidades. As fábricas tendem a definir zonas isoladas para estas operações, de forma a não interferirem com outras operações e evitem congestionamentos no interior das instalações.

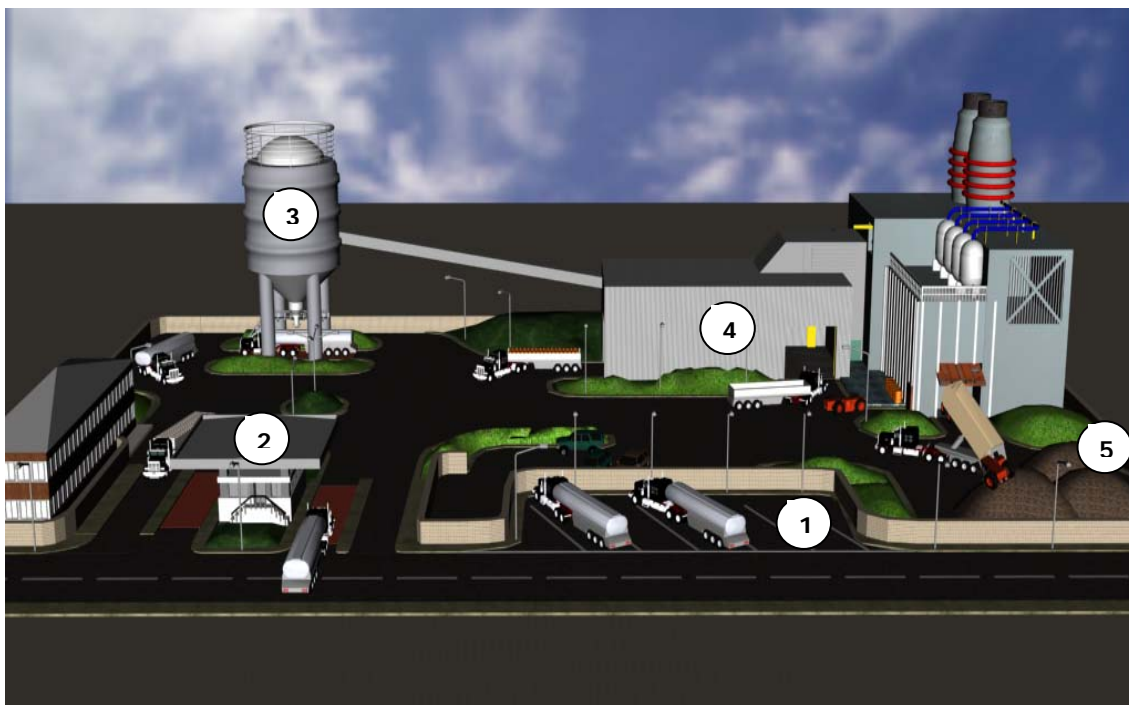


Imagem 3 - Configuração de Instalações (Fonte [14])

De facto, estas áreas funcionais desempenham um papel estrutural no processo de movimentação, visto que, da sua articulação, são idealizadas "*rotas ou circuitos*" que permitem aos gestores industriais agrupar e ordenar as viaturas, criando complexos processos paralelos de circulação de viaturas no interior das unidades. Através da definição de variáveis de controlo como horários de funcionamento, limites de viaturas nas zonas de operação e tempos previstos de atendimento, os gestores actuam de forma directa sobre o processo.

Focalizando um conjunto de objectivos de negócio estratégicos, a qualidade de serviço assume um papel de relevo, nomeadamente na oferta de um produto diferenciado num mercado tão homogéneo como o do cimento. No entanto, a definição de padrões de qualidade de serviço passa pela percepção, entre outros, dos índices de qualidade expectáveis pelos clientes, assim como o conhecimento de factores que possam ser a causa de análises deficientes. [Alcantara, 2000] identifica um conjunto de factores associados aos gestores que podem influenciar análises incorrectas:

- “Desconhecimento das expectativas dos clientes”.
- “Definição de padrões errados de qualidade”.
- “Falha na execução de serviço” devido à “utilização inadequada de recursos”.
- “Falta de articulação entre promoção de serviços e a execução dos mesmos”.

Em [Farah, 2002], as fábricas de cimento tendem a apresentar-se como um bloco central na cadeia logística, realçando que, ao longo do tempo, têm vindo a ampliar o seu espaço junto da actividade produtiva de uma infinidade de sectores. Deste modo, pretende-se que a criação de uma ligação mais efectiva entre comprador e ou vendedor, traga como resultado a qualidade de serviço representada por:

- Cumprimento do tempo de entrega.
- Precisão no atendimento.
- Qualidade do produto entregue.

Os centros de distribuição podem oferecer indicadores do seu grau de eficiência e de operacionalidade contudo, para cada tipo ou família de produtos, existem diferentes estruturas funcionais, fazendo com que, um centro de distribuição de cimento seja diferente daquele que distribui produtos congelados ou mesmo medicamentos.

Em [Alcantara, 2000], para se alcançarem ao máximo os benefícios estratégicos da logística, todas as suas actividades funcionais, como armazenagem, movimentação ou transportes, necessitam de ser realizadas com uma visão integral do sistema logístico. A excelência de cada aspecto funcional é somente relevante quando enquadrada no âmbito da sua contribuição para a melhoria total do sistema.

Neste sentido, e face ao contexto anteriormente apresentado, as fábricas de cimento tornam-se elementos activos nas cadeias logísticas, e tendem a identificar um conjunto de objectivos que permitam fomentar e fortalecer as ligações com os seus parceiros:

- Optimizar a qualidade de serviço de carregamento fazendo aproximar o desvio entre a quantidade pedida e quantidade efectiva para zero – Sendo o processo de fabrico deste tipo de produto fortemente normalizado, a sua qualidade tende a caracterizar-se por factores “externos” como a efectividade nos processos de carregamento.
- Uniformizar os tempos de permanência nas instalações evitando picos de afluência – Picos de afluência tendem a induzir tensões nos processos internos da instalação, conduzindo a ineficiências de processo e maiores preocupações com critérios de segurança face ao invulgar aglomerado de viaturas. Assim, o reajuste de viaturas para horas de menos afluência poderá aliviar de forma natural este tipo de constrangimentos.
- Evitar congestionamentos no interior das instalações mantendo os tempos e número de viaturas parque aceitáveis para os motoristas – Oferta de condições de segurança reflectem-se na satisfação dos motoristas e no reforço da imagem e projecção da organização.
- Optimizar recursos necessários para garantir os períodos de funcionamento contínuos sem incremento de custos – Incremento de qualificações e com melhor resposta aos requisitos dos parceiros de negócio como clientes, fornecedores ou transportadores.

Através da articulação destes factores, torna-se possível maximizar os benefícios estratégicos da logística, onde a satisfação do consumidor criada pela empresa e seus canais de distribuição, nomeadamente transportadores, está dependente do trabalho conjunto, do valor ou utilidade do produto.

2.3 Conhecimento dos sistemas de informação

Tal como referido anteriormente, a definição de índices de qualidade de serviço passa, entre outros factores, pelo conhecimento por parte dos gestores, das expectativas dos seus consumidores, assim como, por uma correcta articulação dos seus recursos na execução de serviço. Neste sentido, as organizações ao nível do mercado do cimento, tendem a adoptar

sistemas industriais para a automatização dos seus processos onde, a CACHAPUZ com a solução *SLV*, assume um papel de destaque no panorama nacional e internacional, complementados por sistemas de suporte à decisão, focalizados nas necessidades efectivas dos seus gestores, como é o caso da *PAC - Plataforma Analítica Cachapuz*.

A *PAC* surgiu como uma resposta às exigências dos gestores, na obtenção de índices e relatórios de negócio de uma forma simples, intuitiva e em tempo útil, para os quais, os tradicionais sistemas operacionais não estavam vocacionados, integrando assim funcionalidades convencionais de suporte à decisão, materializadas num sistema de *data warehousing* e de processamento analítico, complementadas com alguns mecanismos de mineração de dados, no qual se enquadra esta tese.

Durante a implementação da *PAC*, foi efectuado um levantamento exaustivo das funcionalidades da solução *SLV*, enquadrando-se a sua vertente de automação logística na movimentação de viaturas nas unidades industriais, culminando com a definição de uma *matriz de decisão*. A matriz de decisão foi um dos pontos-chave do processo, tendo-se identificado possíveis *data marts* que potencialmente irão integrar o futuro *data warehouse SLV*. Cada um dos *data marts* corresponderá a uma vista de dados para suporte às actividades de um ou vários agentes de decisão, integrando todos os elementos de dados relativos às necessidades de análise e, necessariamente, todas as perspectivas de análise (dimensões) comunicadas previamente (e documentadas) por esses agentes.

Do conjunto de *data marts* criado (*cargas, descargas, balanço cargas / descargas, logs processo*), foi efectuada uma análise mais aprofundada ao *data-mart global de operações* que agrega a informação relativa às operações de carga e descarga, assim como tempos envolvidos nas principais fases de processo (*total nas instalações, operação de carga ou descarga, tempo no interior das instalações*), dado oferecer uma visão global sobre as vertentes e objectivos de negócio identificados.

A Imagem 4 apresenta a estrutura base do referido *data mart*.

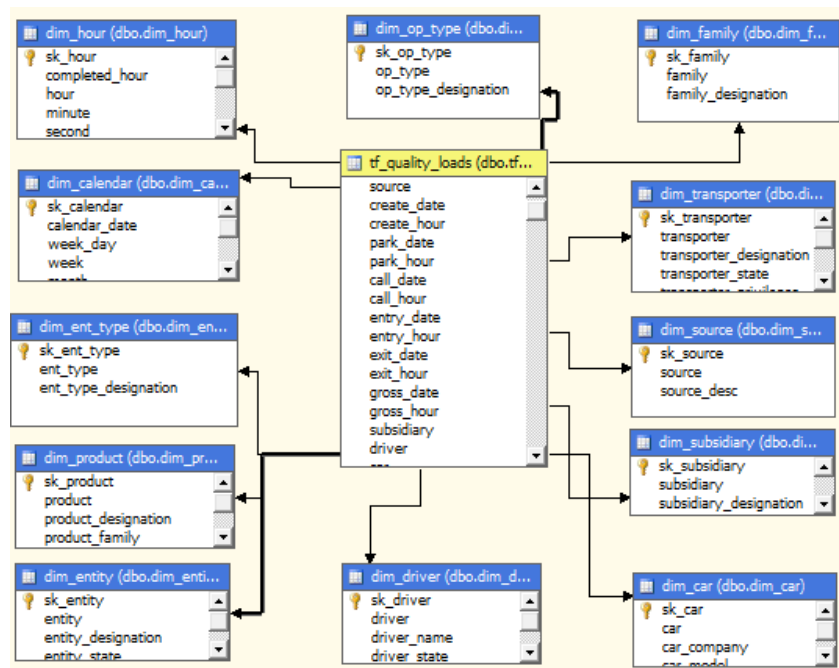


Imagem 4 - Data Mart Global de Operações

Assente numa tabela de factos que armazena informação relativa aos processos de carga e descarga, o seu grão representa o atendimento de uma viatura e permite medir o índice de qualidade de atendimento, focalizado em duas vertentes chave de negócio:

- Tempos de operação – Considerando os períodos de atendimento que são registados nas várias fase de movimentação das viaturas na instalação, desde a sua chegada até à saída das instalações.
- Efectividade no desvio de quantidade – Através do cálculo dos desvios em diferentes perspectivas, nomeadamente absoluta e percentual, entre a quantidade pedida e a efectivamente executada.

Face ao histórico de movimentação e taxas de afluência média nas instalações, prevê-se a uma média de 600 a 700 movimentos por dia que corresponderá a um crescimento médio anual na ordem de 227500 registos.

2.4 Dash-boards

Uma das razões que tem projectado e tornado atractivos os sistemas de suporte à decisão, está relacionada com a formatação prévia da informação a disponibilizar aos seus gestores, fornecendo-lhes, de forma imediata, uma visão geral da sua área de negócio. *Dash-boards* e *Score-cards* apresentam-se como ferramentas para gestão e sistematização de grandes volumes de informação. Através da combinação de poderosas técnicas de *reporting* e *layout*, disponibilizam de um modo visualmente atraente e intuitivo, a informação indispensável para identificar e responder em tempo útil a eventos críticos, assim como explorar e analisar tendências, sem necessidade de recorrer a dezenas de relatórios.

No âmbito da logística de movimentação interna nas cimenteiras, o histórico de instalações *SLV*, aliado ao forte relacionamento com os seus agentes de decisão, permitiu sistematizar um conjunto de requisitos e necessidades, com o objectivo de fornecer indicadores para a análise da qualidade de serviço segundo perspectivas chave. Neste sentido, foram desenhados e implementados *dash-boards* assentes no *data mart global de operações PAC*, focalizados nas componentes temporais de processo e eficácia de carregamento nas operações de movimentação de camiões.

Para este trabalho foi utilizada a base de dados de uma instalação do sistema *SLV* que a Cachapuz efectuou nas empresas do grupo CIMPOR, *Amreyah Cement Company (AMCC)* e *Amreyah Cimpor Cement Company (ACCC)* no Egipto em 2008. Apesar de se tratar de uma instalação recente, existiu um longo trabalho preliminar de preparação da componente operacional, onde rapidamente, foi notória a receptividade dos seus gestores e colaboradores, em participarem activamente no processo de análise para criação de modelos de mineração. Aliado ao trabalho já desenvolvido no projecto *PAC*, surgiu naturalmente o contexto ideal para se desenrolar este trabalho.

2.4.1 Análise Temporal

A eficiência de um centro de distribuição, pode ser avaliada pela ligação efectiva entre os clientes e fornecedores, com clara influência pelos índices de qualidade de serviço, ao nível de:

- Cumprimento do tempo e prazo de entrega.
- Precisão no atendimento
- Qualidade do produto entregue.
- Adequação dos recursos a alocar em função do volume e tipo de operações.

A análise temporal focaliza-se no índice de atendimento, baseado no diferencial tempo previsto e tempo efectivo, considerando as seguintes vertentes dos processos de movimentação:

- Tempo de atendimento – Reflecte o tempo que é necessário para o atendimento de uma viatura, desde a sua chegada até à sua saída das instalações, projectando-se o peso temporal do centro de distribuição na cadeia logística dos seus clientes / fornecedores.
- Tempo de espera – Indica o tempo necessário da chegada de viatura às instalações até à sua entrada.
- Tempo nas instalações – Tempo necessário desde a entrada da viatura até à saída das instalações. A conjugação deste valor com o tempo de espera permite numa primeira fase identificar qual das fases do processo está a influenciar desvios no tempo total de atendimento. A título de exemplo, tempos elevados de espera podem revelar ineficiências na gestão de parque, no entanto, o cruzamento com este valor pode indicar que na realidade estão a existir problemas internos na execução das operações.
- Tempo de operação – Reflecte o tempo para se executar o processo efectivo de carga ou descarga de viatura, permitindo avaliar, se possíveis desvios no tempo de atendimento interno, se devem a ineficiências nos processos de carregamento ou na logística de movimentação envolvente ao processo efectivo.
- Índice de atendimento – resulta da ponderação dos valores anteriormente calculados, permitindo aos seus gestores avaliarem globalmente a qualidade do serviço prestado, segundo a perspectiva temporal de execução.

De forma a possibilitar a compreensão dos valores e identificar possíveis causas, estes devem ser enquadrados com a realidade da instalação, surgindo normalmente novas variáveis a contemplar:

- Afluência de viaturas – Número total de operações, considerando o seu tipo (carga ou descarga) e a sua dispersão por portarias da instalação.
- Dispersão temporal – Dispersão de viaturas ao longo do dia (por hora).

Centrando-nos, agora, no *data-mart global de operações PAC*, foi desenvolvido o *dash-board qualidade de atendimento temporal*, onde se enquadram as variáveis anteriormente apresentadas e que são vocacionadas para os gestores da área logística de movimentação na fábrica. Neste sentido, a informação foi customizada e formatada de acordo com as suas necessidades, materializando-se através das vistas apresentadas, reduzindo as opções de filtragem apenas às suas necessidades efectivas.



Imagem 5 - *Dash-board* qualidade de atendimento temporal

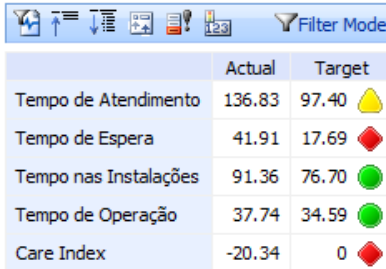
A compreensão das necessidades dos gestores e o seu perfil de gestão, foi um factor fundamental na determinação do número, tipo de relatórios ou *score-cards* e períodos temporais a incluir no relatório. Neste caso, pretendeu-se obter uma visão global do desempenho num curto prazo, ou seja, ser capaz de enquadrar a situação actual com um histórico recente, como meio de detecção de tendências. Neste sentido, o *dash-board* foi desenhado de forma a permitir uma abordagem *top-down* do desempenho.

Através da definição de um *score-card*, disponibiliza-se ao gestor, o seu conjunto de indicadores chave, reflectindo o estado de desempenho actual face ao esperado, recorrendo a três estados de análise:

- Verde – O desempenho está de acordo com os limites esperados.
- Amarelo – O sistema está ligeiramente fora do esperado, devendo ser lançado um estado de alerta.
- Vermelho – O desempenho encontra-se fora dos limites admissíveis e, como tal, deverão ser tomadas medidas urgentes de modo a repor a normalidade de funcionamento.

No exemplo apresentado, constata-se que no período da semana 26 a 28, para operações de carga, o índice de atendimento exige uma intervenção urgente, pois, os tempos globais de atendimento estão em estado de alerta e fortemente influenciados pelos graves desvios verificados ao nível dos tempos de espera em parque. Por outro lado, considerando um conjunto de operações mais abrangente (processos de carga e descarga para o período da semana 26 a 40), consta-se que os indicadores se encontram dentro dos limites definidos como aceitáveis à excepção do tempo no exterior das instalações. Este continua a revelar necessidade de intervenção urgente para reposição dos valores definidos como objectivo.

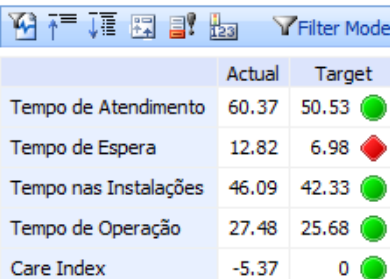
Tempos de Atendimento



	Actual	Target	
Tempo de Atendimento	136.83	97.40	▲
Tempo de Espera	41.91	17.69	◆
Tempo nas Instalações	91.36	76.70	●
Tempo de Operação	37.74	34.59	●
Care Index	-20.34	0	◆

Imagem 6 - Tempos atendimento operações de carga (sem. 26 a 28)

Tempos de Atendimento



	Actual	Target	
Tempo de Atendimento	60.37	50.53	●
Tempo de Espera	12.82	6.98	◆
Tempo nas Instalações	46.09	42.33	●
Tempo de Operação	27.48	25.68	●
Care Index	-5.37	0	●

Imagem 7 - Tempos atendimento global (sem. 26 a 40)

No sentido de permitir uma melhor compreensão sobre a dimensão dos valores apresentados, os gestores necessitam de conhecer o conjunto de análise, ou seja, o número de viaturas envolvidas, a sua distribuição temporal, logística nas instalações e o seu impacto na degradação ou melhoria da qualidade de serviço. Deste modo, pela Tabela 2, é possível verificar que face ao número de operações efectuadas, existe uma distribuição de afluência temporal não uniforme, com principal incidência a partir das 8:00 H. É, ainda, verificável que existem picos e quebras momentâneas, correspondentes à troca de turnos de trabalho operacionais, que implicam a paragem da unidade industrial em causa.

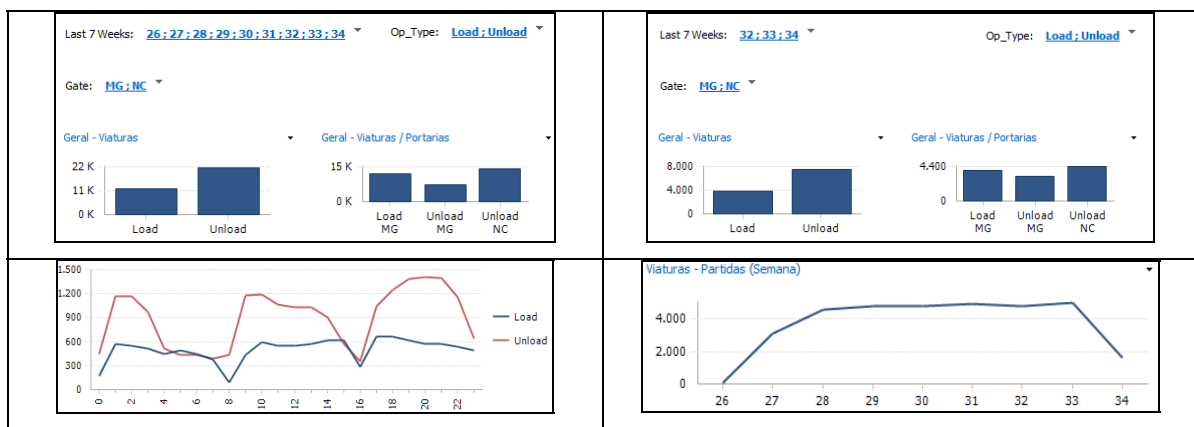


Tabela 2 - Relação entre operações executadas e distribuição temporal

Compreendido o comportamento de afluência de viaturas torna-se, então, necessário avaliar a reacção dos indicadores de desempenho ao longo do tempo. A Imagem 8 reflecte a evolução dos indicadores numa perspectiva semanal, constatando-se que, globalmente o sistema está a evoluir positivamente e aproximando-se dos objectivos de funcionamento definidos. No entanto, numa visão mais detalhada e considerando-se como eixo temporal a hora do dia, consta-se que os turnos tendem a criar desequilíbrios no funcionamento da instalação. É também evidente que, no período após as 16:00 H, os índices de atendimento são claramente inferiores aos objectivos fortemente influenciados pelos tempos de espera em parque.

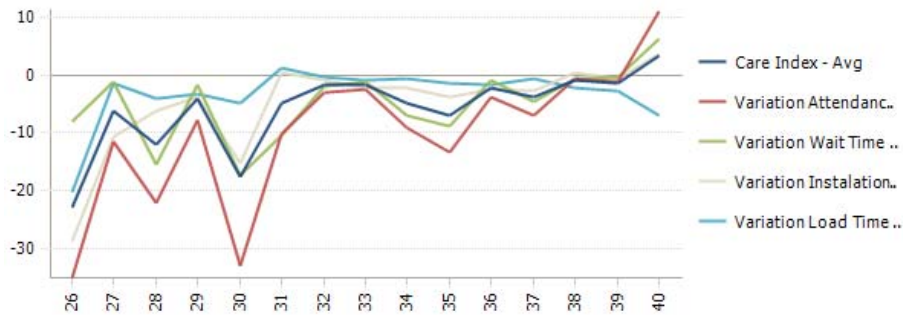


Imagem 8 - Variação semanal dos índices de atendimento



Imagem 9 - Variação horária dos índices de atendimento

2.4.2 Análise de Quantidades

Quando se trata de um produto tão uniforme e normalizado como o cimento, é fundamental a oferta de um serviço de qualidade superior, que possa ser reconhecido pelos consumidores como um factor extra e diferenciador perante os restantes fornecedores. Se, por um lado, a uma análise de desempenho temporal reflecte, o impacto do centro de distribuição na cadeia logística dos seus consumidores, é importante avaliar, também, a relevância de objectivos operacionais, o seu contributo para a minimização de custos e a melhoria na utilização dos seus recursos industriais. Face às características próprias do processo de expedição de produto ensacado, em que a quantidade transaccionada reflecte o número de sacos e não a quantidade de produto efectivamente transaccionada, o controlo rigoroso deste factor é fundamental para a garantia de fornecimento de um produto de qualidade para o cliente, sem colocar em causa os interesses da organização.

Neste sentido, a análise de quantidades focaliza-se no índice de carregamento, baseado no diferencial entre a quantidade pedida e a quantidade efectivamente executada nas suas zonas de carga, considerando as seguintes vertentes:

- Desvio percentual – Reflecte a percentagem de desvio, permitindo ao gestor responder aos padrões de qualidade com base no critério de análise do cliente.
- Desvio absoluto – Reflecte o desvio real, permitindo aos gestores quantificarem o impacto da ineficiência dos processos de carregamento na gestão de negócio, face ao desvio de produto efectivamente movimentado.

De forma a possibilitar uma análise mais global dos processos de carregamento, os gestores sentem a necessidade de “agregar” os desvios operacionais. Por tal, surge o conceito de *ranking* onde, através da definição de grupos de valores, são avaliados e identificados de uma forma simples, áreas de intervenção urgente.

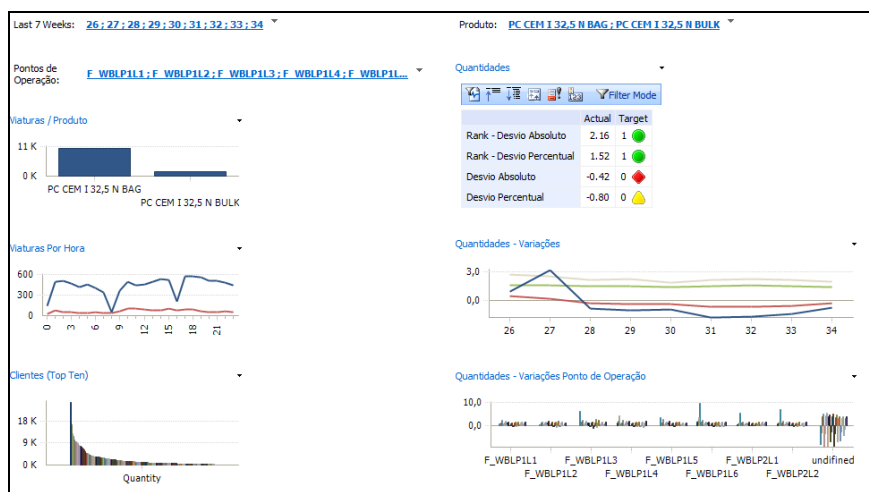


Imagem 10 – Dash-board qualidade de carregamento

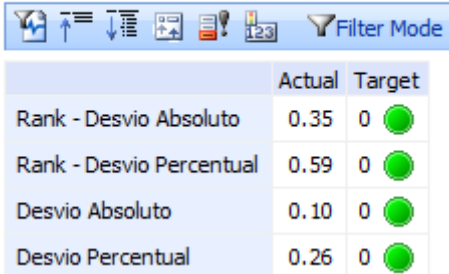
À semelhança do que foi referido no *dash-board* de *qualidade de serviço temporal*, foi desenvolvido o *dash-board* qualidade de carregamento onde, a configuração está directamente relacionada com as necessidades dos seus utilizadores. Neste caso, pretende-se obter uma visão de eficácia global de carregamento, enquadrando com cada um dos pontos de carregamento internos das instalações, para se identificarem possíveis focos de intervenção.

A utilização de *score-cards* foi novamente a melhor opção para disponibilizar aos gestores indicadores chave, segundo duas perspectivas:

- Global – Análise global de desempenho independente dos locais internos de carregamento.
- Individual – Análise detalhada por cada um dos locais internos de carga, identificando-se quais os locais que poderão estar a contribuir para a degradação do índice de carregamento.

Na Imagem 11 e Imagem 12, constata-se que, no período da semana 36 a 40 para operações de carregamento a sacco, os desvios verificados encontram-se dentro dos objectivos definidos, no entanto, constata-se que existe uma ligeira tendência de favorecimento do cliente representada por um desvio absoluto global médio de 100 Kg por operação. Numa análise mais detalhada sobre os desempenhos de cada um dos locais de carga, verifica-se que os diferentes locais estão a realizar um processo de complemento. Por exemplo, a linha 2 apresenta um desvio médio de – 60 Kg, enquanto a linha 4 apresenta um estado de alerta, dado estar fora dos limites estabelecidos como aceitáveis para os desvios absolutos. Embora não sendo alvo de análise deste *dash-board*, esta situação poderá estar relacionada com a afluência de viaturas que requisitam elevadas quantidades (ex: 60 toneladas) nas quais, o desvio absoluto poderá ser largamente ultrapassado, mesmo estando dentro dos limites percentuais.

Quantidades



	Actual	Target
Rank - Desvio Absoluto	0.35	0
Rank - Desvio Percentual	0.59	0
Desvio Absoluto	0.10	0
Desvio Percentual	0.26	0

Imagem 11 – Desvios de quantidade em operações de ensacado (sem. 36 a 40)

Quantidades - Ponto de Operação

	Actual	Target
F_WBLP1L1		
F_WBLP1L2		
Rank - Desvio Absoluto	-0.21	0
Rank - Desvio Percentual	-0.33	0
Desvio Absoluto	-0.06	0
Desvio Percentual	-0.15	0
F_WBLP1L3		
Rank - Desvio Absoluto	0.31	0
Rank - Desvio Percentual	0.57	0
Desvio Absoluto	0.10	0
Desvio Percentual	0.24	0
F_WBLP1L4		
Rank - Desvio Absoluto	0.49	0
Rank - Desvio Percentual	0.91	0
Desvio Absoluto	0.16	0
Desvio Percentual	0.39	0

Imagem 12 – Desvios de quantidade por linha em operações de ensacado (sem. 36 a 40)

Tal como no *dash-board* de qualidade de atendimento, os gestores têm acesso à distribuição temporal de viaturas nas instalações e o seu impacto nos índices de carregamento. Assim, procuram estabelecer uma relação entre a afluência e a degradação de desempenho, apresentando-se na Imagem 13, Imagem 14 e Imagem 15, a evolução de desempenho temporal do ponto de vista horário e semanal.

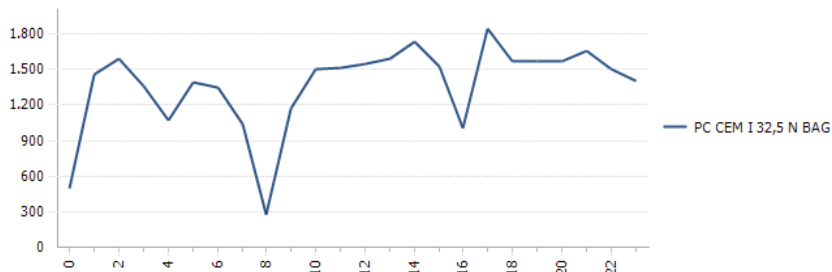


Imagem 13 - Variação horária de quantidades executadas (semana 36 a 40)

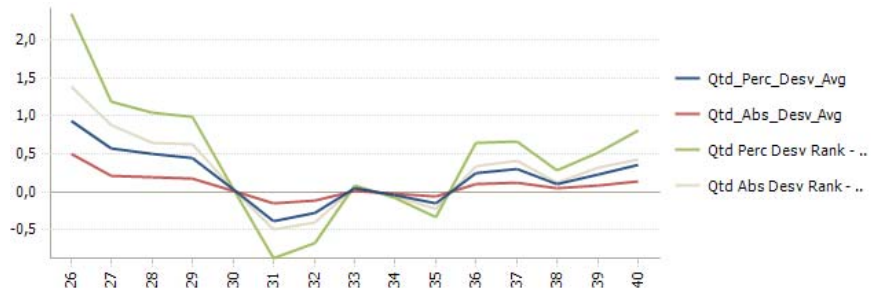


Imagem 14 - Variação semanal dos desvios executadas (Semana 36 a 40)

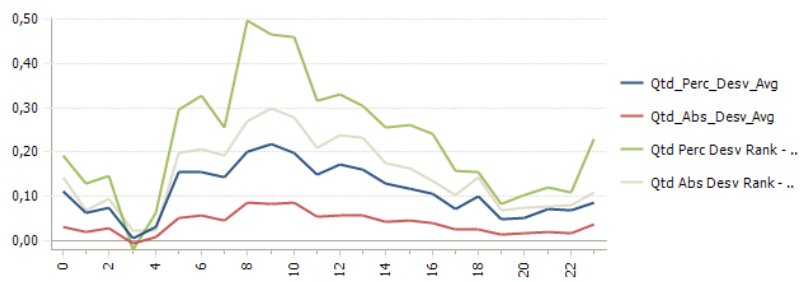


Imagem 15 - Variação horária dos desvios executadas (Semana 36 a 40)

Capítulo 3

Integração de técnicas de mineração

O contexto empresarial de alta competitividade está fortemente relacionado com a capacidade de decisão e resposta, em intervalos de tempo cada vez mais reduzidos, por parte dos seus gestores. Este facto tem impulsionado a sistematização e estruturação do conhecimento da organização conduzindo à implementação de sistemas de suporte à decisão. Através de complexos processos de *ETL (Extraction, Transformation and Loading)*, transformam-se grandes volumes de informação, geralmente dispersos por diferentes fontes e sistemas operacionais, em conhecimento consolidado, disponível para análise e manipulação efectiva em *data warehouses* empresariais.

Como apresentado anteriormente, a utilização de *dash-boards* permite o acesso simplificado à informação, disponibilizando uma visão geral e imediata do negócio. Os seus gestores ou utilizadores, sintetizam a informação de negócio através de consultas *ad-hoc*, comparações, vistas personalizadas, navegação por diferentes eixos de análise com respostas imediatas, independentemente do tamanho e complexidade dos dados em análise. Em [Berry et al., 2005], esta abordagem revela as enormes potencialidades de utilização dos *data warehouses* como uma "*memória*" da empresa, acrescentando que as suas vantagens se tornam limitadas sem incorporação de inteligência.

Com o contínuo aumento dos sistemas de informação, os gestores começam a sentir dificuldades na exploração de grandes volumes de informação. Torna-se difícil a extracção de informação essencial para o negócio, a compreensão de padrões, tendências, sazonalidade, o que leva a dificuldades para reagir atempadamente, definindo e traçando acções para o futuro.

Assim, surge a oportunidade para o campo de pesquisa de extracção de informações a partir de bases de dados, denominado Descoberta de Conhecimento em Base de Dados (*KDD - Knowledge Discovery in Databases*). Em [Fayyad et al. 1996], a descoberta de conhecimento em base de dados é o "*processo não-trivial de identificação de padrões válidos, até então desconhecidos, potencialmente úteis e de possível entendimento em grandes bases de dados*".

A descoberta de conhecimento em base de dados, implica um processo interactivo com múltiplas fases (selecção e preparação de dados, procura de padrões, avaliação e interpretação de conhecimento, refinamentos e utilização de conhecimento), com um forte envolvimento dos analistas e gestores conhecedores da realidade de negócio. Por "*não trivial*", entende-se que não se trata de uma computação directa de valores ou quantidades de valores comum aos sistemas de computação tradicionais. Trata-se da aplicação de modelos na procura, inferência de padrões e ou estruturas que não são evidentes ou identificáveis por seres humanos como, a título de exemplo, o cálculo da média de um conjunto de dados. As expressões "*até então desconhecidos, potencialmente úteis*", traduzem que, para o conhecimento ter valor, é necessário que revele algo de novo com algum grau de relevância, potencialmente útil e com benefícios para um determinado utilizador ou sistema. Por fim, os dados deverão ser entendidos e compreendidos pelos seus gestores ou profissionais envolvidos nos processos de tomada de decisão, caso contrário, deixarão de ser utilizados, tornando-se como "*não úteis*".

3.1 Processo de Descoberta de Conhecimento

O conceito *Data Mining* encontra-se usualmente associado ao processo de descoberta de conhecimento, no entanto, este representa apenas uma das etapas do processo global, que corresponde à utilização de modelos específicos para a identificação de padrões e informação relevante existente nos dados em análise [Fayyad et al. 1996]. A Imagem 16 ilustra as várias fases do processo de descoberta de conhecimento, onde é evidente a dinâmica e interactividade no decorrer de cada uma das suas etapas.

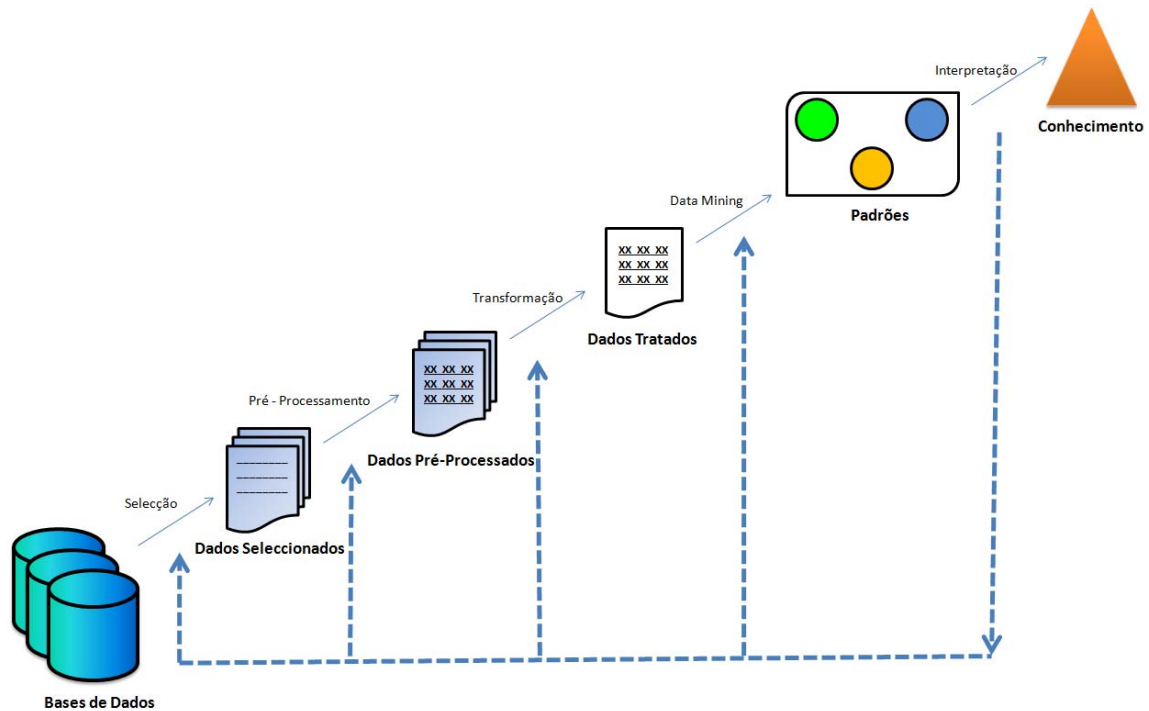


Imagem 16 - Fases do processo de descoberta de conhecimento (Adaptado de [Fayad et al., 1996])

A incorporação de conhecimento é transversal a todo o processo desenvolvendo-se em várias etapas, havendo, geralmente, a necessidade de serem executadas de forma repetida, em função da tomada de decisões e do novo estado de conhecimento obtido.

3.1.1 Selecção de Dados

A selecção dos dados corresponde à primeira fase do processo de descoberta de conhecimento. Tem como principal objectivo, após o estudo e compreensão do domínio de aplicação, limitar o espaço de pesquisa para aquele que se julga ser relevante para o processo, direccionando o foco para subconjuntos de variáveis ou de dados, onde será realizada a descoberta de conhecimento.

O conhecimento do domínio é fundamental para que se possa efectuar uma correcta selecção da informação. Em [Anand et al., 1995], o domínio do conhecimento consiste na informação já disponível sobre os dados, que resulta de outros processos de descoberta ou da incorporação de conhecimento de especialistas da área. Neste sentido, o envolvimento de gestores e conhecedores

da área com dados e objectivos de trabalho, permitirá otimizar a análise e selecção de dados para o processo.

Em [Botica, 2004], analisam-se e detalham-se alguns aspectos a considerar nesta fase, nomeadamente ao nível de:

- Relevância de Dados – As fontes de dados disponíveis para o processo de descoberta de conhecimento, encontram-se muitas vezes dispersos e em formatos diversificados, havendo assim, a necessidade de seleccionar os atributos efectivamente relevantes para o processo, descartando aqueles de carácter meramente informativo [Fayyad, 1996].
- Representatividade dos dados – O facto de, as bases de dados serem um repositório com elevado número de informação não focalizada na descoberta de conhecimento, faz com que haja a necessidade de se analisar a representatividade dos dados a seleccionar. A utilização de ferramentas estatísticas para análise de amostras ou o historial de dados recolhidos da organização, poderão permitir a identificação de eventos ou acontecimentos anómalos a descartar, de forma a não influenciarem a amostra.
- Volume de dados necessários – Definida a relevância e garantida a representatividade dos dados, é necessário avaliar o volume de dados disponível para a identificação do modelo. O volume de dados necessários para a criação de um modelo sólido, está altamente relacionado com o número de atributos seleccionados e respectivos valores associados. Ou seja, quanto maior for o número de atributos seleccionados, maior será o volume de dados necessário. Contudo, apesar da utilização de mais dados parecerem garantir maior representatividade e facilidade de aprendizagem dos modelos, poderão também causar algumas ameaças, tais como:
 - Elevado tempo gasto na construção do modelo.
 - A diluição de ocorrências pouco frequentes face ao elevado número da amostra, quando aplicado em modelos que visam realçar essas particularidades.
- Periodicidade da recolha de dados – Devido ao crescimento contínuo das bases de dados, bem como possíveis alterações nos dados, surge a necessidade de se prever a forma e a periodicidade da recolha dos dados. É, pois, necessário garantir que essas alterações são reflectidas nas relações produzidas, assegurando assim a sua representatividade, bem como a actualização do modelo.

3.1.2 Pré-Processamento e Transformação de Dados

Um dos factores críticos para o sucesso dos processos de descoberta de conhecimento em base de dados consiste na qualidade dos dados. As bases de dados dispõem, normalmente, de estruturas inadequadas ao processo *data mining*, apresentando problemas como dispersão de dados, dados omissos, volatilidade dos dados ou mesmo inconsistência, que podem tornar os padrões anteriormente encontrados inválidos.

O objectivo principal do pré-processamento de dados prende-se com a melhoria da qualidade dos dados a utilizar, de forma a otimizar significativamente a eficiência e precisão da fase *Data Mining* [Teixeira, 2005]. Em [Botica, 2004], os trabalhos de pré-processamento podem consumir até 80% do esforço total do projecto, sendo que, os problemas mais comuns correspondem a informação insuficiente e, ainda, a dados caracterizados por possuírem ruído ou por estarem incompletos. Existe, então, a necessidade de realização de procedimentos para a correcção de anomalias no conjunto de dados, que passam por operações de limpeza, de modo a remover esses ruídos e inconsistências dos mesmos.

Na transformação, os dados são ajustados e consolidados de acordo com o objectivo inicial, de modo a poderem ser aplicadas as técnicas de *Data Mining*. Consiste na última fase de preparação para submissão dos dados aos algoritmos de mineração, passando por:

- Normalização de registos – Redução dos dados a uma mesma escala, permitindo diminuir as variáveis para cada atributo, baseado no modo como são codificados e agregados. [Pinto et al., 2004] ilustra esta situação com o exemplo dos códigos postais e nomes de cidades onde, geralmente, aparecem definidos com múltiplas abreviações referindo-se ao mesmo registo, como por exemplo:

País	Código Postal	Cidade
PT	2400	Vila Nova de Gaia
PT	2400-230	Gaia

Tabela 3 - Normalização de Registos (Adaptado [Pinto et al., 2004])

- Remoção de atributos – Consiste na eliminação de atributos que não sejam relevantes para o estudo, podendo passar pela interacção com especialistas e gestores de área, ou então, pela utilização de árvores de decisão para identificação de atributos não considerados e, como tal, passíveis de serem removidos.
- Agrupamento de registos – Consiste na generalização de hierarquias conceptuais que representam o domínio do conhecimento, ou então, na discretização em classes definidas para o efeito.

3.1.3 Aplicação de algoritmos de Data Mining

Esta é a fase onde os dados seleccionados, processados e transformados, serão submetidos aos algoritmos e ferramentas para identificação e reconhecimento de padrões. Uma análise detalhada sobre os possíveis algoritmos a utilizar será apresentada em 3.2 Tarefas e Técnicas.

A selecção de técnicas que se adequem aos problemas e objectivos é essencial para a obtenção de resultados, podendo-se utilizar, sempre que possível, a combinação de várias técnicas, no sentido de se produzirem melhores resultados [Fayyad, 1996].

3.1.4 Avaliação de resultados

Da execução de modelos de mineração são gerados resultados que necessitam de uma análise para se identificarem padrões que representam conhecimento. Em [Botica, 2004], o conhecimento descoberto será caracterizado segundo as seguintes premissas: *válido*, *novo*, *útil* ou *interessante*, em função do seu grau de certeza, novidade para o domínio de estudo, relevância para os utilizadores ou domínio do caso de estudo. Em função desta fase, poderão ser reiniciados novos processos de refinamento face ao conhecimento descoberto, passando sempre pela consolidação e incorporação do conhecimento no domínio do caso de estudo.

3.2 Tarefas e Técnicas de Mineração

As tarefas e técnicas de Mineração são, sem dúvida alguma, a base de um processo de mineração. *Data Mining* está intimamente relacionado com a construção de modelos, onde um modelo é um algoritmo ou um conjunto de regras que se aplicam a um conjunto de dados de entrada para a obtenção de um resultado ou objectivo particular [Berry et al., 2005]. Em [Teixeira, 2005], cada modelo contém um conjunto de parâmetros que devem ser ajustados na fase de aprendizagem, de forma a ser possível extrapolar novas saídas (*outputs*), alimentando o modelo com novas entradas (*inputs*).

De notar, que não existe uma forma universal de solucionar todos os problemas associados à descoberta de conhecimento, uma vez que, cada modelo está vocacionado para tarefas específicas, onde a sua combinação poderá ser a chave na obtenção de melhores resultados no contexto associado à descoberta do conhecimento. A Tabela 4 ilustra a relação entre técnicas e tarefas de mineração.

Técnicas	Tarefas	Classificação	Associação	Clustering	Previsão
Árvores de Decisão		✓		✓	✓
Redes Neurais		✓	✓	✓	✓
Indução de Regras		✓	✓	✓	✓
Algoritmos Genéticos		✓		✓	✓

Tabela 4 - Relação entre Técnicas e Tarefas de Mineração (Adaptado de [Teixeira, 2005])

3.2.1 Tarefas

Em *Data Mining* uma tarefa corresponde a um tipo de problema de descoberta de conhecimento, para o qual se pretende determinar uma solução.

[Berry et al., 2005] apresenta as tarefas Classificação, Associação, *Clustering* e Previsão como um meio para a resolução de problemas de domínio intelectual, económico e de negócio. [Freitas et al., 2005], complementa referindo que uma tarefa de mineração pode ser classificada de acordo com o tipo de padrão descoberto, segundo duas vertentes:

- Preditiva – A que trata do uso de variáveis na base de dados para previsão de valores, tais como as tarefas de classificação e previsão;
- Descritiva – A que identifica os padrões em dados históricos, tais como, as tarefas de associação e clustering.

Classificação

A classificação é uma das tarefas mais comuns nos processos de mineração, na medida em que se aproxima da forma de ser e estar do próprio ser humano. Consiste na análise dos atributos de um determinado objecto, para a posterior associação a um conjunto de classes pré-definidas. Por tal, torna-se imprescindível uma correcta definição desse conjunto de classes, para que o processo seja capaz de classificar cada objecto introduzido de acordo com os seus atributos. As tarefas de classificação, caracterizam-se pela utilização de classes bem definidas e conjuntos de treino com exemplos pré-classificados, de forma a adequar os modelos para poderem ser aplicados a dados não classificados [Berry et al., 2005].

As tarefas de classificação são, normalmente, utilizadas nas seguintes situações:

- Classificações de aplicações de créditos de baixo, médio e alto risco.
- Selecção de conteúdos a apresentar em ambientes Web.
- Fraudes associadas a seguradoras.

Associação

Esta tarefa consiste na primeira aproximação à exploração de dados, sendo geralmente aplicados a um vasto conjunto de informação. Fornecem informação de grande valor, identificando correlações significativas existentes nos dados [Olsen, 2008].

A associação pretende determinar grupos de acontecimentos, tendo como objectivo a medição do relacionamento entre os objectos analisáveis, de forma a serem encontradas relações e ou correlações de acontecimentos, que normalmente ocorrem juntas num determinado conjunto de dados, às quais poderemos chamar de padrões.

As tarefas de associação encontram-se, normalmente, associadas às seguintes áreas:

- Retalho – Para análise de cabazes de compras.
- Páginas da *internet* – Análise de cliques em páginas da *internet* para medir associações ou relacionamentos entre páginas visitadas sequencialmente pelos visitantes de um site.

Clustering

Clustering consiste na segmentação de populações heterogéneas em subconjuntos com características homogéneas. Ao contrário do que acontece na classificação, esta tarefa não assenta na pré-definição de classes. Os registos são agrupados de acordo com as suas próprias similaridades e afinidades [Berry, 2005]. Esta tarefa é, geralmente, o ponto de partida para a aplicação de outras tarefas, tais como a classificação e a associação. Tomando como exemplo a projecção de campanhas no mercado, esta tarefa permite a divisão dos clientes em grupos com culturas, hábitos e interesses de compra similares, que irão permitir numa segunda fase, avaliar qual a promoção que terá melhores resultados para cada um dos grupos entretanto identificados.

Previsão

A Previsão acaba por ser muito semelhante à tarefa de classificação, diferindo apenas no facto dos registos serem classificados de acordo com um comportamento futuro previsto ou um valor futuro estimado [Berry, 2005]. Estas tarefas são normalmente utilizadas na previsão:

- Do lucro em função das vendas.
- Da fidelização de clientes num determinado período.

3.2.2 Técnicas

Uma técnica de mineração consiste num conjunto de processos baseados em conhecimento científico, de cálculo ou experimentação, utilizados para a obtenção de um determinado resultado [Teixeira, 2005]. Surgem como algoritmos ou metodologias que são aplicadas na implementação das tarefas apresentadas anteriormente.

Seguidamente, apresentam-se diversas técnicas de mineração.

Árvores de Decisão

Uma árvore de decisão define-se como uma estrutura onde se encadeiam diversas regras que apontam para uma classe ou valor. O seu principal objectivo prende-se com a separação de classes, em que os dados estão agrupados em nós, representando cada um deles, uma condição lógica simples sobre um determinado atributo [Botica, 2004]. As árvores de decisão são normalmente apropriadas para tarefas de classificação assentando em conjuntos de algoritmos, tais como ID3, CHAID e C5.0 [Teixeira, 2005].

Uma das características desta técnica prende-se com a facilidade de manipulação que, juntamente com a comunicação visual da árvore, permitem uma rápida compreensão e leitura por parte do utilizador.

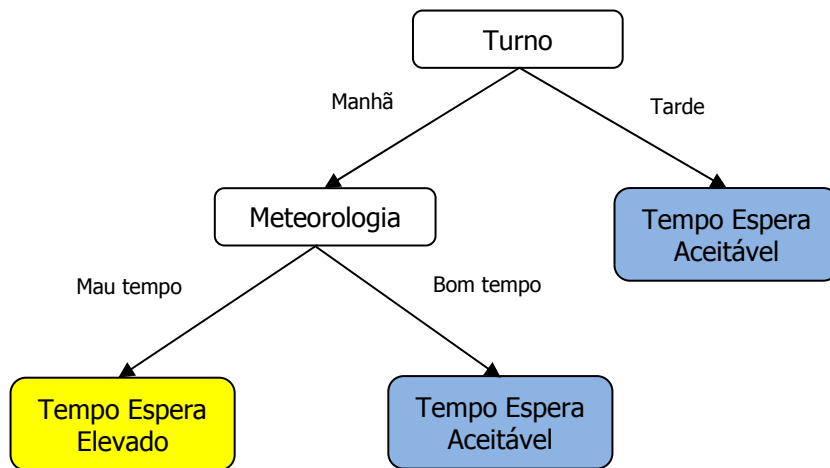


Imagem 17 - Árvore de Decisão para um problema de classificação do tempo de espera para carregamento de viaturas

A Imagem 17 ilustra o conjunto de regras que permitem a classificação do tempo de espera para carregamento de viaturas, numa linguagem simples e perceptível para o utilizador, como *Tempo de Espera Aceitável* ou *Tempo de Espera Elevado*. É possível evidenciar um conjunto de regras extraídas da árvore de decisão, como por exemplo, *Se Turno = "Manhã" e Meteorologia = "Mau Tempo" Então o tempo de espera para carregamento de viaturas é elevado.*

Redes Neurais

Redes Neurais definem-se como uma classe de modelação de prognóstico, que trabalha por ajuste repetido de parâmetro. Estruturalmente, uma rede neuronal consiste num número de elementos interconectados (chamados neurónios), organizados em camadas e unidos por conexões [Cunico, 2005]. As Redes Neurais são modelos sub-simbólicos, e muitos investigadores acreditam que oferecem a melhor aproximação para a construção de verdadeiros sistemas inteligentes, com capacidade para ultrapassar a explosão combinatória associada à computação simbólica, baseada em arquitecturas de *Von Neumann* [Teixeira, 2005]. [Botica, 2004] define as redes neurais como modelos que simulam o funcionamento do sistema nervoso humano. Cada elemento da rede está ligado a um ou mais elementos do nível seguinte, através de ligações às quais é atribuído um peso. A partir dos valores de entrada é encontrado um valor de saída. A Imagem 18 ilustra uma rede neuronal da função $Z = 3X + 7Y$ onde, para qualquer valor de entrada X e Y , é encontrado um valor de saída Z :

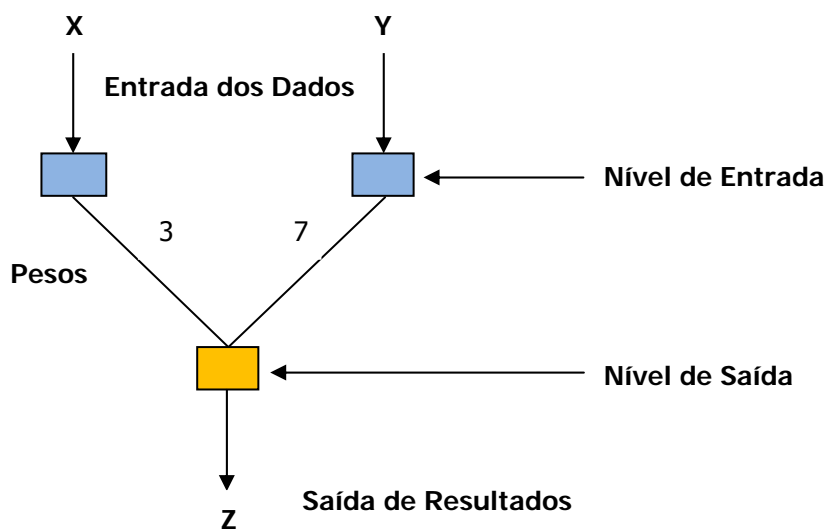


Imagem 18 – Rede Neuronal da função $Z = 3X + 7Y$ (Adaptado de [Berry et al., 2000])

Indução de Regras

A Indução de Regras é uma técnica de *Data Mining* que surge associada às árvores de decisão, já referenciadas anteriormente. São utilizadas como representação do conhecimento exposto nas

árvores de decisão. O seu objectivo passa, através da análise das probabilidades condicionais, pela descoberta de dependências entre os atributos. Os resultados são apresentados através de regras $X \rightarrow Y$, que significa que "se X está presente, então Y também tem probabilidade de estar presente". A variável X poderá ser constituída por uma combinação de vários atributos ou valores, tornando, assim, as regras bastante mais complexas [Teixeira, 2005].

De seguida apresenta-se um conjunto de regras extraídas da árvore de decisão, apresentada anteriormente, na secção de Árvores de Decisão:

Se Turno = "*Tarde*"

Então Tempo Espera Aceitável

Se Turno = "*Manhã*" e Meteorologia = "*Mau Tempo*"

Então Tempo de Espera Elevado

Algoritmos Genéticos

A técnica de algoritmos genéticos consiste no desenvolvimento de estratégias que optimizem algoritmos, baseada na evolução natural e genética das espécies, sendo as informações referentes a um determinado sistema, codificadas de forma idêntica à dos cromossomas biológicos. Perante um determinado conjunto de soluções de problemas, que competem entre si, serão seleccionados os melhores, para que se obtenha uma melhoria significativa na qualidade da solução [Botica, 2004]. Em [Navega, 2001], à semelhança do que acontece no mundo real, os dados iniciais são sujeitos a uma série de operações, tais como:

- **Seleccção** – Os elementos considerados como os melhores, serão escolhidos de acordo com uma função de avaliação definida para o domínio.
- **Cruzamento** – São escolhidos de uma forma aleatória pares de regras, dentro do conjunto seleccionado.
- **Mutação** – São introduzidas alterações esporádicas a um dado gene, podendo um caracter do alfabeto ser trocado por outro. Este processo é muito similar às mutações genéticas que ocorrem no DNA.

Como resultado é criada uma nova geração de soluções. O processo repete-se várias vezes até que seja atingida uma dada condição de paragem, definida por exemplo, pelo número máximo de gerações. A Imagem 19 ilustra um exemplo desta técnica.

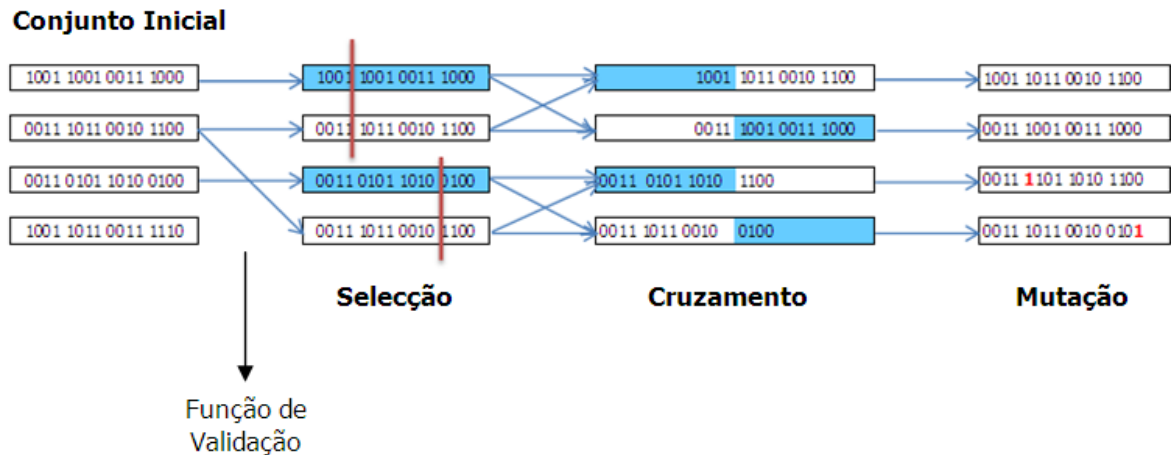


Imagem 19 – Ilustração da técnica algoritmos genéticos de *Data Mining* (Adaptado [Navega, 2001] e [Botica, 2004])

3.3 Aplicações

[Berry et al., 2005], num contexto vocacionado para o *marketing* e vendas, constata que, os padrões dos clientes tornam-se evidentes ao longo do tempo, na medida em que os seus comportamentos não são aleatórios, mas sim, revelam necessidades, gostos e preferências. A introdução de inteligência, através de técnicas de *data-mining* permitirá reagir de forma efectiva às suas necessidades futuras da organização. Assim, [Sumathi et al., 2006], revela que tem surgido um crescente interesse pelas organizações em adoptar técnicas de extracção e mineração de dados para identificação de correlações, anomalias e tendências não evidentes, através das convencionais análises de dados, em áreas de negócio extremamente diferenciadas, como por exemplo, medicina, transportes, banca ou manufactura. No entanto, a descoberta de conhecimento assenta em ferramentas e, como qualquer ferramenta, não é suficiente compreender como funcionam, mas também, como poderão ser utilizadas e aplicadas. Os resultados estão dependentes do domínio de aplicação e dos interesses dos gestores, passando

pela identificação das técnicas de mineração, que poderão oferecer as melhores respostas aos problemas apresentados.

Tomando como exemplo o caso da venda a retalho, a análise do cabaz de compras é fundamental para compreender as afinidades entre os produtos e serviços adquiridos. Tarefas como *associação* ou *classificação*, são geralmente utilizadas para detectar automaticamente importantes padrões e relações entre características e escolhas dos seus clientes. Num âmbito mais lato, a segmentação de clientes, utilizando tarefas de *clustering* na definição de grupos comportamentais comuns, tem vindo a ser explorada para detectar grupos de risco, nomeadamente de perda de clientes para concorrentes directos ou por outro lado, definir novas estratégias de marketing, no sentido de atrair um determinado perfil de clientes para a organização.

Um dos casos de maior sucesso e divulgação no âmbito da descoberta de conhecimento foi o da cadeia *Wal-Mart*, que identificou um hábito curioso dos seus consumidores. Com o objectivo de encontrar possíveis relações entre o volume de vendas e os dias da semana, chegou-se à conclusão que às sextas-feiras as vendas de cerveja cresciam na mesma proporção que as fraldas para bebés [11]. Posteriormente, uma investigação mais detalhada revelou que ao comprar fraldas para os bebés, os pais aproveitavam para comprar, também, cerveja para o fim-de-semana.

Como este, muitos outros casos de sucesso têm vindo a ser divulgados, tais como, *1-800-Flowers.com*, *IBM*, *Staples*, entre outros [10], em que inerente ao processo de descoberta de conhecimento, foi possível melhorar significativamente o desempenho do negócio, assim como a relação com o seus clientes.

No caso específico da Plataforma *PAC*, a utilização de *dash-boards* fornece aos gestores industriais meios efectivos para uma análise flexível e eficiente dos seus dados. No entanto, face ao crescimento do volume de dados e variáveis a analisar, surge a necessidade de se avaliar a integração de ferramentas de mineração de forma a identificar, realçar padrões e tendências relacionadas com os processos de movimentação de viaturas nas instalações.

Tomando como exemplo a indústria cimenteira, as tarefas de mineração poderão ter um impacto significativo na melhoria do seu desempenho e funcionalidade. Através de tarefas de segmentação, poderão ser criados grupos de clientes com comportamentos e tendências semelhantes ao nível da

periodicidade e hora de chegada às instalações, quantidades e tipos de produtos requisitados. Devidamente considerados, estes grupos poderão ser utilizados na otimização da disposição de cada um dos armazéns existentes, para que um dado cliente não tenha de se dirigir a vários locais para concluir a operação. Considerando outro tipo de características, como sendo o histórico de movimentos e hora de execução, através de tarefas de previsão, será possível estimar a afluência nas próximas horas, semanas ou meses, sendo como tal, um factor claramente diferenciador na forma como as fábricas poderão escalonar os seus turnos e recursos laborais. À semelhança do caso de sucesso da *Wal-Mart*, também aqui a compreensão de relações entre produtos, volume de vendas e épocas do ano poderá ser extremamente útil, no sentido de entender qual o comportamento de uma determinada zona ou região face aos períodos das chuvas, por exemplo. Tradicionalmente, nestas épocas existe uma forte tendência para a reestruturação de vias ou edifícios que está, por sua vez, intimamente relacionado com a utilização de produtos com determinadas características. Todavia, com base nos dados que temos disponíveis, serão utilizadas as tarefas de classificação e associação, para análise de desempenho e eficiência das linhas de carregamento de produto ensacado, cujos processos se apresentam detalhados nos capítulos seguintes.

A metodologia *CRISP-DM* servirá como guião funcional de tarefas e fases de processo, de forma a ser possível tirar partido da sua adequação aos mercados industriais, dada a sua fundamentação e génese estar intimamente relacionada com a indústria e ambiente empresarial. Em suma, procura-se com este passo, um complemento aos sistemas de análise, no sentido da obtenção de uma solução totalmente integrada e flexível, onde os utilizadores possam navegar por níveis e perspectivas sucessivas de informação de forma sistemática e precisa, enquanto os motores de mineração "*procuram*", "*sugerem*" e "*aceleram*" novas etapas do processo de negócio.

Capítulo 4

Preparação dos Dados

Compreendido o enquadramento da realidade de negócio, assim como o processo de descoberta de conhecimento, chegou o momento de ajustar os dados para as necessidades específicas dos algoritmos e técnicas de mineração. Nesta fase, serão então executadas tarefas no sentido de limitar e transformar a informação associada aos processos de movimentação SLV para futura submissão aos processos de mineração.

4.1 Selecção de Dados

Nesta fase, são seleccionados os dados relevantes para incluir nos processos de mineração, considerando-se as possíveis fontes como sistemas operacionais, *data warehouses* ou *data marts*, assim como o seu conjunto de dados ao nível de registos e atributos. O objectivo é identificar e limitar a informação efectivamente necessária, eliminando-se os dados que não tenham qualquer interesse para os processos de mineração. Embora existisse a possibilidade de utilização de dados directamente dos sistemas operacionais do *SLV*, este processo assentou na análise dos diferentes *data-marts SLV-BI* que, dada a sua génese, já contemplam toda a informação relevante para análise dos processos de cargas e descargas numa unidade cimenteira.

Do leque de *data-marts* existentes, cargas, descargas, consolidação de cargas e descargas, o *data-mart global de operações PAC*, disponibiliza todos os dados necessários para a aplicação dos modelos de mineração, no sentido de optimização da qualidade de serviço sendo, no entanto, necessário aplicar um conjunto de acções de limpeza como descrito na secção 4.2.

4.2 Limpeza de Dados

Com a limpeza de dados, pretende-se elevar a qualidade de dados para o nível necessário à aplicação das técnicas de modelação. Este passo corresponde ao processo no qual se identifica e corrige informação incorrecta e incompleta, assim como a eliminação de registos duplicados, que possam causar inconsistências na fase de mineração. Isto poderá envolver a selecção de conjuntos específicos de dados, a inserção de valores por defeito, ou a utilização de técnicas mais ambiciosas como sendo a estimativa de valores omissos por modelação [9].

Ao nível do *data-mart* seleccionado, foram detectadas algumas inconsistências que implicaram acções sobre os dados, de forma a eliminar inconsistências ao nível das datas de operação e catalogação de valores nulos, associados aos pontos de operação que nem sempre se encontravam preenchidos. [Botica, 2005] refere que existem diferentes formas de lidar com os dados omissos. Os modos de actuação passam por descartar registos com valores omissos, ou então, em contextos focalizados em áreas específicas de negócio, recorre-se a módulos de limpeza que possibilitam um tratamento mais eficaz com minimização de perdas de informação.

4.2.1 Tratamento de datas de operação

O processo de movimentação interno numa empresa industrial, como é o caso de uma cimenteira, envolve a execução simultânea de diferentes tipos de operação onde, como é compreensível, partilhando um mesmo espaço (a fábrica), cada processo tem um determinado comportamento, de modo a tornar a sua execução simples e eficiente. Assim, tomando como exemplo uma operação de carregamento a saco, face ao seu processo lento de carregamento e limitações físicas na zona de carregamento, é necessário contemplar um local de parque, de forma a organizar as filas de espera. Por outro lado, no caso das matérias-primas, o cenário é diferente, pois, o processo de descarga é extremamente rápido em locais que são tradicionalmente amplos e capazes de suportar um elevado número de viaturas. Esta pequena diferença, faz com que as unidades industriais criem processos de movimentação distintos entre cargas e descargas fazendo com que, por exemplo, existam operações que não executem o processo de espera em parque, podendo as viaturas dirigirem-se directamente para a entrada das instalações.

Um outro caso prende-se com o controlo do tempo de carga ou descarga, na medida em que, nem sempre, as fábricas têm as zonas de operação automatizadas e integradas com os sistemas de informação. Isto faz com que estes desconheçam o tempo efectivo do processo de carga.

Neste sentido, foi necessário contemplar um processo para o tratamento de datas de operação assegurando a definição dos indicadores: tempo de espera em parque, tempo dentro das instalações, tempo de operação e tempo total de atendimento. A primeira abordagem consistiu na definição de valores por defeito onde, tendo por base a data de criação, definiam-se tempos a adicionar em caso de falta de uma determinada data.

code	park_time_add	call_time_add	entry_time_add	exit_time_add	weight_time_add	gross_time_add
C	5	10	15	60	15	58
D	2	1	5	20	18	5

Imagem 20 - Tratamento de datas – Valores por defeito

Detalhando um pouco a Imagem 20, define-se que em caso de omissão de, por exemplo, a data de parque, esta seria calculada como sendo a data de criação do registo com a adição de 5 minutos. Caso se tratasse de uma operação de descarga, o tempo a considerar seria de 2 minutos.

Após uma análise mais exaustiva dos processos, esta abordagem mostrou-se insuficiente, dado não permitir uma diferenciação detalhada por produto ou circuito de movimentação. Por outro lado, mesmo analisando operações com o mesmo modo de funcionamento, o comportamento e desempenho da instalação são bastante variáveis, pelo que se definiu uma nova abordagem onde, o cálculo da data em falta está relacionado com o processo de movimentação do documento. Só em último caso, seria utilizada a regra apresentada na Imagem 20.

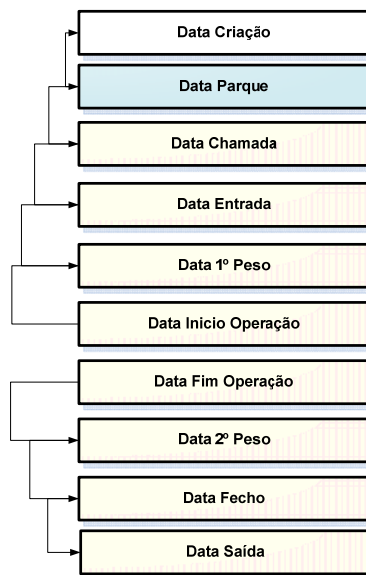


Imagem 21 - Relação de dependências entre datas associadas a um processo de movimentação

A Imagem 21 demonstra as relações de dependências entre cada um dos campos de movimentação onde, a título de exemplo, se pode apresentar o cenário de inexistência da data de chamada, que será característico nos processos de movimentação de matérias-primas. Nestes casos, será considerada a data de registo em parque que, no caso de não se encontrar definida, será considerada a data de criação do documento.

Com a aplicação destas regras, obteve-se uma relação de valores mais próximos do funcionamento real da instalação, e com uma reacção mais efectiva à existência de valores nulos nas datas. O tratamento tem por base o contexto de um movimento e não valores pré-determinados, que tendem a incluir desvios face ao contexto global em que os seus valores são calculados.

4.2.2 Tratamento de nulos pontos de operação

O controlo efectivo dos locais de operação (carga e descarga), está intimamente relacionado com o nível de automação que a unidade tem implementado. Tradicionalmente, o processo de automatização e integração com os sistemas inicia-se com o controlo dos processos de parque, entradas e saídas, ficando para uma segunda fase, a extensão para as zonas de operação de

cargas e descargas. Dada a multiplicidade de sistemas de automação e controlo envolvidos ser extremamente complexa e com custos de integração de sistemas relativamente elevados, as fábricas tendem a recorrer, numa primeira fase, a processos assistidos por operadores, realizando um controlo suportado por mecanismos de alto nível, nomeadamente através de relatórios de balanço. Os gestores garantem assim, um controlo de processo mínimo requerendo, também, que os sistemas operacionais estejam preparados para a futura integração dos processos e componentes de hardware necessários para a automatização dos pontos de operação.

Do ponto de vista operacional, esta situação conduziu à necessidade de se contemplarem mecanismos de transição, à medida que os pontos de operação vão sendo automatizados e integrados nos sistemas de informação. Gradualmente, passam a registar os locais e datas de início ou término de operação sendo, então, necessário incorporar mecanismos para a normalização da informação relativa aos pontos de operação.

Dado que o sistema SLV contempla a noção de circuitos, onde os pontos de operação para cargas e descargas estão caracterizados como "*pontos principais*", o mapeamento do posto de operação utilizará as seguintes regras:

- Determinação através do posto principal – Cada processo de movimentação encontra-se relacionado com um circuito de movimentação específico. Caso no circuito exista apenas um ponto principal capaz de satisfazer o pedido associado, assume-se que esse foi o ponto de operação utilizado.
- Atribuição de valor "*undefined*" – Para os casos onde a regra anterior não tenha retornado qualquer valor, será considerado um valor por defeito, no sentido de tornar evidente esta falha.

4.2.3 Ranking de Desvios

Do ponto de vista operacional, os gestores sentem a necessidade de definirem intervalos lógicos de funcionamento para normalizar os desvios da instalação. Neste sentido, foram definidos dois grupos denominados *rankings de desvio*, focalizados nas seguintes vertentes:

- Absoluto – Reflete o desvio efectivo entre a quantidade requisitada pelo cliente e a quantidade efectivamente disponibilizada.
- Percentual – Reflete o desvio percentual da operação no sentido de se definir uma medida para rejeição de fornecimento que se situa, no caso da instalação em análise, nos 2%.

Neste contexto, foram definidos intervalos de funcionamento de acordo com as expectativas dos gestores para a análise do seu negócio, tendo em consideração que é necessário contemplar uma região de funcionamento ideal e a partir daí, escalonar os desvios em intervalos consecutivos, contemplando ganhos e perdas.

A Tabela 5 reflecte as regras para definição dos rankings para desvio de peso percentual, onde:

- O *ranking* "0" representa a região ideal de funcionamento, agrupando as operações com desvios admissíveis máximos até 0.5 %.
- Os *rankings* "1" e "-1" representam um estado intermédio de alerta de funcionamento, admitindo desvios máximos até 1 %, antecedendo o nível de alerta severo representado pelos *rankings* "2", "-2", "3", "-3", que revelam uma aproximação ao limite máximo admissível que se situa nos 2 %.
- Os rankings "4", "-4", "5", "-5" reflectem níveis de intervenção urgente, dado se estar a operar fora dos limites aceitáveis para o funcionamento.

Ranking de desvios percentuais de quantidade												
-2.5 <=	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	>= 2.5
- 5	- 4	- 3	- 2	- 1	0	0	1	2	3	4	5	

Tabela 5 - Ranking de desvios percentuais de quantidade

Relativamente aos desvios absolutos de quantidade, as regras para definição de rankings de desvio encontram-se definidas na Tabela 6:

- O *ranking* "0" representa a região ideal de funcionamento, agrupando as operações com desvios admissíveis máximos até 100 Kgs e que representam uma margem até 2 sacos.
- Os *rankings* "1" e "-1" representam um estado intermédio de alerta de funcionamento, onde os desvios podem atingir os 200 Kgs, antecedendo o nível de alerta severo representado pelos *rankings* "2", "-2", "3", "-3", com desvio máximo admissível até 800 Kgs.
- Os *rankings* "4", "-4", "5", "-5" reflectem níveis de intervenção urgente, dado se estar a operar fora dos limites aceitáveis para o funcionamento.

Ranking de desvios absolutos de quantidade (toneladas)												
-0.8<=	-0.8	-0.6	-0.4	-0.2	-0.1	0	0.1	0.2	0.4	0.6	0.8	>= 0.8
-5	-4	-3	-2	-1	0	0	1	2	3	4	5	

Tabela 6 - Ranking de desvios absolutos de quantidade

4.3 Janela de Oportunidade para execução do modelo

A determinação da janela de oportunidade para extracção e consolidação de dados, é fundamental para se encontrar um compromisso entre as necessidades ou requisitos dos gestores e o impacto sobre os sistemas operacionais, resultante dos processos de *ETL* ou execução dos modelos de mineração. Em [Reinardi, 2008], os processos de extracção e execução de modelos são processados ciclicamente com intervalos regulares, como diários ou semanais. Se a recuperação de dados for executada uma só vez ou em períodos excessivamente alargados, os dados irão tornar-se obsoletos e inúteis. Por outro lado, em sistemas que variam pouco, não é necessário seleccionar dados relativos a pequenos períodos de tempo, porque se estaria a recolher informação repetida, de reduzido conteúdo informativo. Execuções excessivamente periódicas, irão sobrecarregar os sistemas operacionais sem mais valias evidentes para os gestores, dado que terão o seu próprio ritmo de acesso aos dados.

Por fim, é fundamental considerar as necessidades do negócio sobre o qual os processos de extracção e mineração irão actuar. Se estamos num contexto dinâmico, como por exemplo as campanhas de marketing focalizadas no cliente e nos seus interesses, os modelos devem ser executados várias vezes ao dia, para se evitar a notificação de clientes que tenham já anulado o interesse na subscrição de uma determinada campanha. Num outro contexto onde se pretende obter um relatório de rentabilidade semanal, os processos de extracção de informação das fontes de dados poderá ser executada uma vez por semana, antecedendo o momento previsto para disponibilização do relatório.

No contexto da industria cimenteira, é também possível identificar cenários e exigências diferentes relativas à execução dos modelos de extracção de dados para geração de relatórios assim como, o agendamento de execução dos modelos de *data mining* para descoberta de padrões que influenciem a qualidade de serviço pretendida.

Considerando a qualidade de serviço focalizada no processo, nomeadamente o processo de carregamento a saco que tem forte influência sobre os processos de negócio, os gestores sentem a necessidade de um acompanhamento periódico e contínuo, de forma a agir atempadamente a ineficiências ou comportamentos indevidos. Se, por um lado, a necessidade de controlo de desvios de quantidades possa ser encarada como um processo puramente operacional, por outro, a identificação de associações, relações de clientes com locais de carregamentos e existência de padrões de desvios de carregamento justificam, claramente, a utilização deste tipo de abordagens com tempos de actualização frequentes. Permite-se, assim que, durante a permanência de viatura nas instalações, se possa confrontar o comportamento da operação com o seu histórico de movimentação. Caso a viatura tenha um histórico que indica uma tendência para maiores desvios num determinado turno de trabalho em determinada linha, então durante os processos de carregamento poderão ser tomadas medidas de monitorização de proximidade, que rapidamente permitirão ajustar os desempenhos a valores esperados.

Relativamente à qualidade de serviço, focalizada na perspectiva temporal, degradação ou optimização destes valores é relativamente lenta sendo as análises efectuadas sobre períodos temporais alargados, onde a execução contínua de modelos não identificará variações significativas.

Capítulo 5

Qualidade de serviço aplicado ao carregamento a sacco

5.1 Contextualização

Nos processos de movimentação internos às companhias *AMCC* e *ACCC*, todos os processos são caracterizados pela execução de operações mono-produto, ou seja, em cada operação é transaccionado apenas um único produto em que, a força de vendas assenta no produto ensacado. A Tabela 7 reflecte o número de operações e quantidades movimentadas para as operações de carga.

Produto	Operações		Quantidade (Toneladas)	
	Total	Percent. (%)	Total	Percent. (%)
PC CEM I 32,5 N BAG	17.791	87	684.607,00	88
PC CEM I 32,5 N BULK	2.689	13	93.609,95	12

Tabela 7 - Relação de movimentação para operações de carga

Na Imagem 22 e Imagem 23, apresenta-se em detalhe o interior dos armazéns de ensacagem onde, após a recepção de viatura, os operadores iniciam o carregamento da viatura num processo demorado e árduo. Os operadores recebem os sacos, um a um, via um tapete de abastecimento que vão distribuindo pela viatura até a satisfação da quantidade pedida aquando da recepção do pedido.

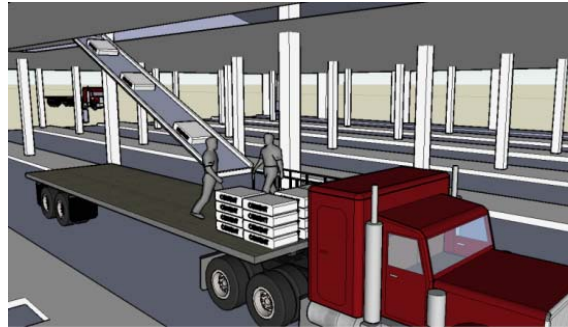


Imagem 22 - Detalhe do interior de armazém de produto ensacado



Imagem 23 - Detalhe do tapete de abastecimento de sacos

Face ao contexto específico no qual estas empresas se encontram, em que os processos para controlo de carregamento não estão automatizados e integrados directamente nos sistemas de informação, é fundamental contemplar processos de suporte à decisão que permitam aos gestores reagirem a ineficiências. Neste caso de estudo, foi dado especial ênfase ao processo de carregamento de produto ensacado, na medida em que representa 88% da quantidade total transaccionada e se encontram associados factores críticos de processo:

- Carregamento – O processo de carregamento em cada uma das linhas de operação é efectuado recorrendo a operadores. Estes, a diferentes níveis, serão responsáveis pelo carregamento, contagem e controlo dos sacos efectivamente carregados, assim como tempos aceitáveis para execução da operação.

- Unidade de Medida – Actualmente cada sacco corresponde a 50 Kg, reflectindo-se numa quantidade total movimentada de 13.692.140 de sacos.
- Controlo de Perdas – Dado os locais não se encontrarem automatizados, o controlo de perdas, nomeadamente de rebentamento de sacos à saída do tapete de carregamento, é de difícil determinação.

A articulação destes factores, tornam o controlo relativamente vulnerável e com perdas de informação em que, com o conhecimento de associações inerentes aos processos de carregamento, será possível a implementação de mecanismos de controlo mais eficientes, nomeadamente ao nível do controlo de quantidades e desempenho de cada uma das linhas. Assim, detalha-se em seguida, a aplicação de técnicas de associação e classificação ao processo específico de carregamento a sacco, fazendo-se no final uma análise crítica sobre as vantagens desta abordagem.

Tal como referido anteriormente, este trabalho encontra-se inserido no contexto da plataforma *SLV-BI* fortemente suportado por ferramentas *Microsoft*¹ e, como tal, a selecção das ferramentas *Microsoft SQL Server 2005 Analysis Services*, em particular o *Business Intelligence Development Studio*, surgiram naturalmente, permitindo desenvolver aplicações de *data mining* no mesmo ambiente de desenvolvimento dos restantes tipos de aplicações [15]. Para além das funcionalidades descritas, no que se refere ao desenvolvimento em si, o *SQL Server 2005 Analysis Services* foi melhorado face à versão *SQL Server 2000 Analysis Services*, através da incorporação de novos algoritmos, visualizadores e designers que simplificam o processo de desenvolvimento e visualização de relacionamentos nos processos de *data mining* e que respondem de forma efectiva aos requisitos e expectativas deste trabalho².

5.2 Desenvolvimento de modelos

A compreensão de relações entre desvios de quantidades e linhas de operação é fundamental para permitir acções precisas de correcção, nomeadamente, monitorização de proximidade aos processos de carregamento. A Imagem 24 reflecte a distribuição de desvios efectivos entre a quantidade pedida

¹ A Cachapuz é Microsoft Certified Partner desde 2006 facilitando o acesso a ferramentas de desenvolvimento

² Para informação mais detalhada poderá ser consultada a página do produto [14].

e a quantidade executada, verificando-se que apenas 30 % das operações se encontram no *ranking* ideal definido como "0". Os grupos intermédios são responsáveis por 70% desses desvios, concentrando-se em especial nos grupos de alerta "1, -1" e "2, -2".

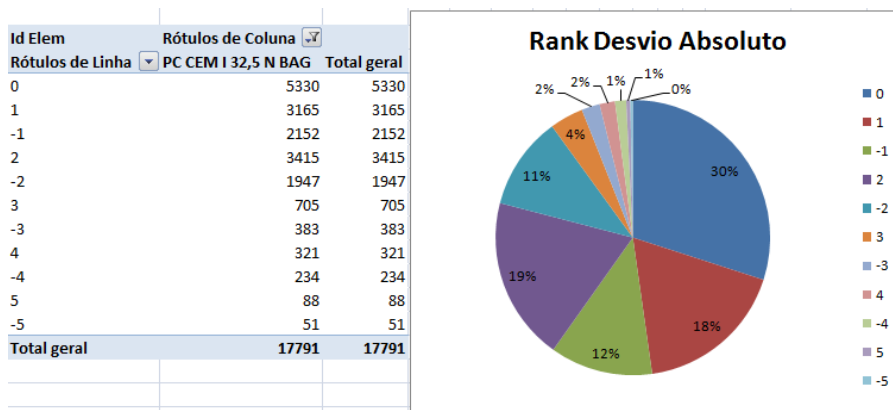


Imagem 24 - Ranking de desvio absoluto associado a operação de carregamento a sacco

Relativamente aos desvios percentuais, a Imagem 25 revela que 55 % das operações estão dentro do grupo estabelecido como ideal para o funcionamento. Os restantes 45 % estão distribuídos nos grupos intermédios. É de realçar, que cerca de 10 % das operações se encontrarem nos *rankings* "3, -3" ou superior indicando, à partida, uma forte tendência para operar nos limites admissíveis.

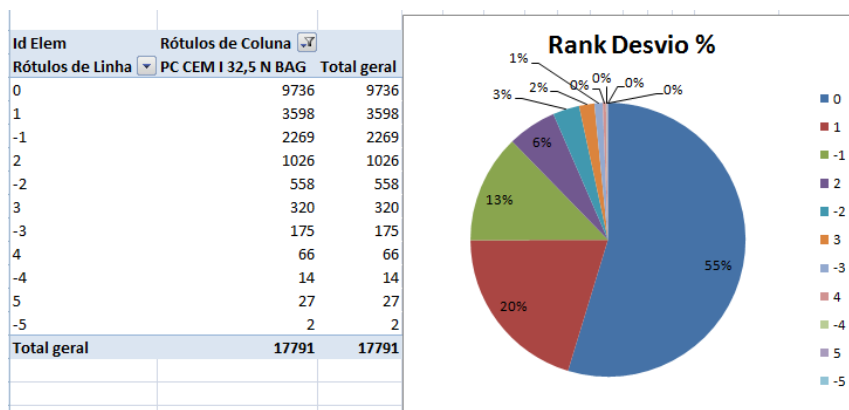


Imagem 25 - Relação de rankings de desvios percentuais

A estrutura de mineração desenvolvida teve por base o *data mart* global de operações, utilizando como fonte de dados "*Tf_quality_loads*", ilustrando-se, na Imagem 26 a estrutura de mineração que servirá de base a este caso de estudo.

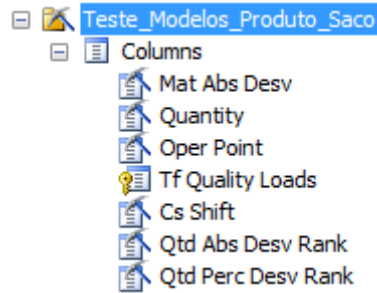


Imagem 26 - Definição estrutura de mineração base

No sentido de restringir o conjunto de dados ao estritamente necessário para a aplicação dos modelos, foram aplicadas regras de filtragem tal como apresentado na Imagem 27, de acordo com os seguintes pontos:

- Aplicação do modelo estritamente a operações de cargas a clientes onde o produto movimentado é "PC CEM I 32,5 N BAG". Procuram-se relações específicas ao processo de carregamento a sacco evitando-se, como tal, relacionamentos com operações paralelas como os processos de carga a granel.
- Eliminação de operações cujo ranking de funcionamento esteja dentro do intervalo ideal de funcionamento, ou seja, ranking de desvio percentual "0". Face ao histórico de operações relativamente recente, os 55% de operações dentro do perfil de normalidade de funcionamento, induzem baixos índices de relevância para as correlações entre os atributos em análise. Como tal, esse conjunto foi descartado nesta fase dado que, os restantes 45% representam um conjunto significativo para submissão aos modelos de mineração. Neste sentido, contemplando apenas os restantes níveis, os modelos exploraram o universo de operações relativas aos *rankings* de desvio percentual de alerta ou intervenção urgente, com resultados bastante mais atractivos.

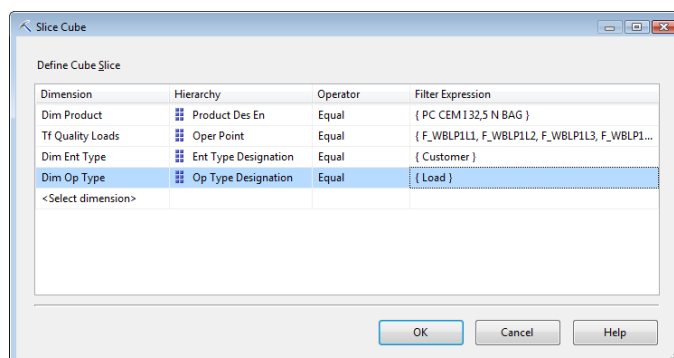


Imagem 27 - *Cube slice* para estrutura de mineração

Após a selecção e definição de campos a incluir no modelo, foi necessário caracterizar cada um dos campos assim como o seu comportamento. Em [13], disponibiliza-se informação mais detalhada acerca dos diferentes tipos que poderão ser associados às colunas da estrutura de mineração. Na Tabela 8 apresenta-se a relação de atributos e tipos de conteúdo considerados.

Campo	Descrição	Tipo de Conteúdo
Mat Abs Desv	Desvio absoluto médio associado a cada sacco da operação	Discretized
Qtd Abs Desv Rank	<i>Ranking</i> desvio de quantidade absoluta	Discrete
Qtd Perc Abs Rank	Ranking desvio de quantidade percentual	Discrete
CS Shift	Turno de operação que reflecte um intervalo horário de funcionamento na organização	Discrete
Oper Point	Ponto de operação para carregamento	Discrete
Quantity	Quantidade requisitada pelo cliente	Discretized

Tabela 8 - Definição da Estrutura de Mineração

A utilização do tipo de coluna "*discretized*" permitiu a geração de grupos representativos de valores onde, ao contrário do que foi realizado para os rankings de desvios de quantidades, não se pretendiam estabelecer previamente. Se na definição de rankings de desvios, o objectivo era definir claramente intervalos de funcionamento pré-determinados, nestes casos, pretende-se que durante a

execução dos modelos sejam dinamicamente criados grupos representativos, potenciando-se a descoberta de intervalos não previsíveis inicialmente.

Concluída a definição da estrutura de mineração, iniciou-se uma fase iterativa para definição de modelos de mineração e características de dados. A Imagem 28 ilustra o conjunto de modelos seleccionados onde cada modelo pretendia responder a problemas e questões específicas:

- *Carregamento_Assoc* - Obter os possíveis relacionamentos entre linhas de carregamento e os desvios registados por operação considerando-se, também, a quantidade movimentada e turno de operação.
- *Carregamento_Assoc_Class* – Classificar a origem dos desvios assim como compreender possíveis causas.

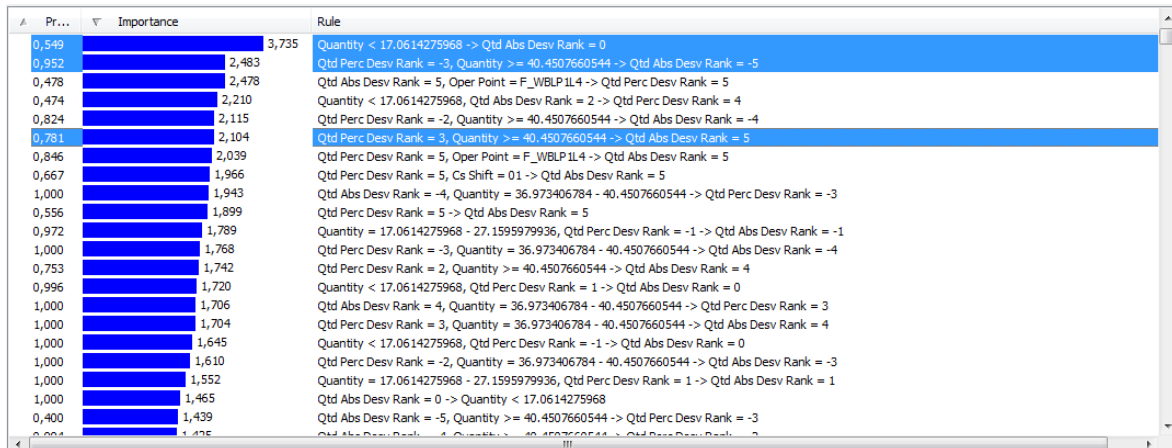
Structure	Carregamento_Assoc	Carregamento_Class
Cs Shift	Microsoft_Association_Rules	Microsoft_Decision_Trees
Mat Abs Desv	Predict	Predict
Oper Point	Ignore	Predict
Qtd Abs Desv Rank	Predict	Predict
Qtd Perc Desv Rank	Predict	Predict
Quantity	Predict	Predict
Tf Quality Loads	Key	Key

Imagem 28 – Visão parcial de modelos de mineração implementados

De seguida, detalha-se o modelo *Carregamento_Assoc* onde, a Imagem 29 demonstra que para quantidades requisitadas inferiores a 17 toneladas, o desvio absoluto de quantidade está dentro do intervalo ideal de funcionamento. No entanto, existe uma clara divergência em operações acima das 40 toneladas. Nestes casos, opera-se dentro dos *rankings* de *alerta severo* ou *intervenção urgente*, sendo também de notar que, existe uma situação de complemento através das regras "*Qtd Perc Desv Rank = -3, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = -5*" e "*Qtd Perc Desv Rank = 3, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = 5*". Isto é determinado pela importância das regras de associação onde os atributos estão presentes e pela probabilidade de estas regras se verificarem na amostra da informação existente. No modelo de associação, a importância significa a pontuação, ou *interesting score*, de uma determinada regra [Tang et al., 2005]. Quanto mais elevado for este valor, mais relacionados estão os acontecimentos.

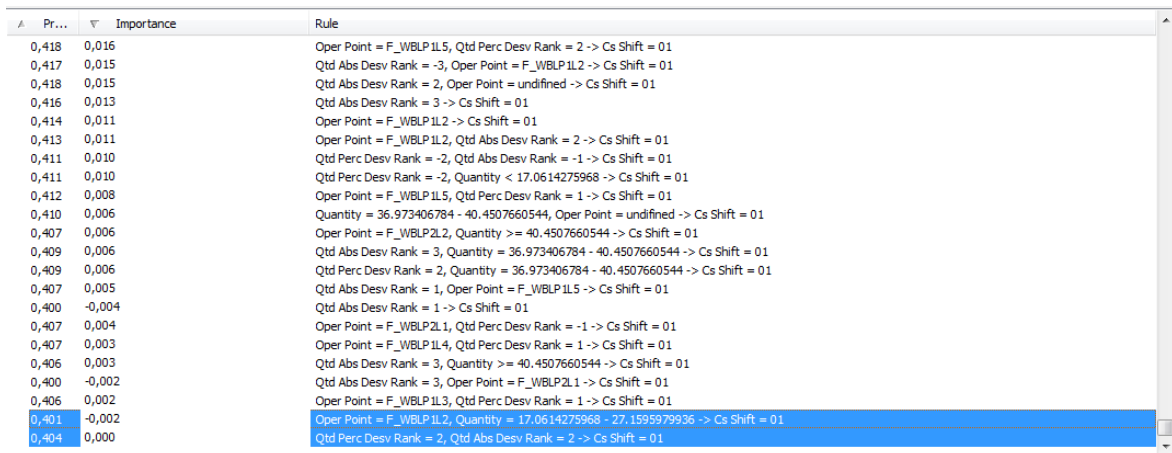
Qualidade de Serviço aplicado ao carregamento a saco

Quando não considerados devidamente, as relações entre *rankings* apresentados, tendem a anular-se, tornando-se imperceptíveis nas análises globais de desvios. Usualmente apenas se consideram somatórios globais, validando-se posteriormente se existem desvios face a valores esperados.



Pr...	Importance	Rule
0,549	3,735	Quantity < 17.0614275968 -> Qtd Abs Desv Rank = 0
0,952	2,483	Qtd Perc Desv Rank = -3, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = -5
0,478	2,478	Qtd Abs Desv Rank = 5, Oper Point = F_WBLP 1L4 -> Qtd Perc Desv Rank = 5
0,474	2,210	Quantity < 17.0614275968, Qtd Abs Desv Rank = 2 -> Qtd Perc Desv Rank = 4
0,824	2,115	Qtd Perc Desv Rank = -2, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = -4
0,781	2,104	Qtd Perc Desv Rank = 3, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = 5
0,846	2,039	Qtd Perc Desv Rank = 5, Oper Point = F_WBLP 1L4 -> Qtd Abs Desv Rank = 5
0,667	1,966	Qtd Perc Desv Rank = 5, Cs Shift = 01 -> Qtd Abs Desv Rank = 5
1,000	1,943	Qtd Abs Desv Rank = -4, Quantity = 36.973406784 - 40.4507660544 -> Qtd Perc Desv Rank = -3
0,556	1,899	Qtd Perc Desv Rank = 5 -> Qtd Abs Desv Rank = 5
0,972	1,789	Quantity = 17.0614275968 - 27.1595979936, Qtd Perc Desv Rank = -1 -> Qtd Abs Desv Rank = -1
1,000	1,768	Qtd Perc Desv Rank = -3, Quantity = 36.973406784 - 40.4507660544 -> Qtd Abs Desv Rank = -4
0,753	1,742	Qtd Perc Desv Rank = 2, Quantity >= 40.4507660544 -> Qtd Abs Desv Rank = 4
0,996	1,720	Quantity < 17.0614275968, Qtd Perc Desv Rank = 1 -> Qtd Abs Desv Rank = 0
1,000	1,706	Qtd Abs Desv Rank = 4, Quantity = 36.973406784 - 40.4507660544 -> Qtd Perc Desv Rank = 3
1,000	1,704	Qtd Perc Desv Rank = 3, Quantity = 36.973406784 - 40.4507660544 -> Qtd Abs Desv Rank = 4
1,000	1,645	Quantity < 17.0614275968, Qtd Perc Desv Rank = -1 -> Qtd Abs Desv Rank = 0
1,000	1,610	Qtd Perc Desv Rank = -2, Quantity = 36.973406784 - 40.4507660544 -> Qtd Abs Desv Rank = -3
1,000	1,552	Quantity = 17.0614275968 - 27.1595979936, Qtd Perc Desv Rank = 1 -> Qtd Abs Desv Rank = 1
1,000	1,465	Qtd Abs Desv Rank = 0 -> Quantity < 17.0614275968
0,400	1,439	Qtd Abs Desv Rank = -5, Quantity >= 40.4507660544 -> Qtd Perc Desv Rank = -3

Imagem 29 – Lista de regras de maior importância na aplicação do modelo *Carregamento_Assoc*



Pr...	Importance	Rule
0,418	0,016	Oper Point = F_WBLP 1L5, Qtd Perc Desv Rank = 2 -> Cs Shift = 01
0,417	0,015	Qtd Abs Desv Rank = -3, Oper Point = F_WBLP 1L2 -> Cs Shift = 01
0,418	0,015	Qtd Abs Desv Rank = 2, Oper Point = undefined -> Cs Shift = 01
0,416	0,013	Qtd Abs Desv Rank = 3 -> Cs Shift = 01
0,414	0,011	Oper Point = F_WBLP 1L2 -> Cs Shift = 01
0,413	0,011	Oper Point = F_WBLP 1L2, Qtd Abs Desv Rank = 2 -> Cs Shift = 01
0,411	0,010	Qtd Perc Desv Rank = -2, Qtd Abs Desv Rank = -1 -> Cs Shift = 01
0,411	0,010	Qtd Perc Desv Rank = -2, Quantity < 17.0614275968 -> Cs Shift = 01
0,412	0,008	Oper Point = F_WBLP 1L5, Qtd Perc Desv Rank = 1 -> Cs Shift = 01
0,410	0,006	Quantity = 36.973406784 - 40.4507660544, Oper Point = undefined -> Cs Shift = 01
0,407	0,006	Oper Point = F_WBLP 2L2, Quantity >= 40.4507660544 -> Cs Shift = 01
0,409	0,006	Qtd Abs Desv Rank = 3, Quantity = 36.973406784 - 40.4507660544 -> Cs Shift = 01
0,409	0,006	Qtd Perc Desv Rank = 2, Quantity = 36.973406784 - 40.4507660544 -> Cs Shift = 01
0,407	0,005	Qtd Abs Desv Rank = 1, Oper Point = F_WBLP 1L5 -> Cs Shift = 01
0,400	-0,004	Qtd Abs Desv Rank = 1 -> Cs Shift = 01
0,407	0,004	Oper Point = F_WBLP 2L1, Qtd Perc Desv Rank = -1 -> Cs Shift = 01
0,407	0,003	Oper Point = F_WBLP 1L4, Qtd Perc Desv Rank = 1 -> Cs Shift = 01
0,406	0,003	Qtd Abs Desv Rank = 3, Quantity >= 40.4507660544 -> Cs Shift = 01
0,400	-0,002	Qtd Abs Desv Rank = 3, Oper Point = F_WBLP 2L1 -> Cs Shift = 01
0,406	0,002	Oper Point = F_WBLP 1L3, Qtd Perc Desv Rank = 1 -> Cs Shift = 01
0,401	-0,002	Oper Point = F_WBLP 1L2, Quantity = 17.0614275968 - 27.1595979936 -> Cs Shift = 01
0,404	0,000	Qtd Perc Desv Rank = 2, Qtd Abs Desv Rank = 2 -> Cs Shift = 01

Imagem 30 - Lista de regras de menor importância na aplicação do modelo *Carregamento_Assoc*

Através da análise da rede de dependências, obtêm-se uma visão global de relacionamentos, apresentando-se na Imagem 31, as dependências para o modelo *Carregamento_Assoc*. No sentido de eliminar relações com um índice de importância marginal, foi ajustado o parâmetro de visualização para supressão de 25 % das relações menos significativas.

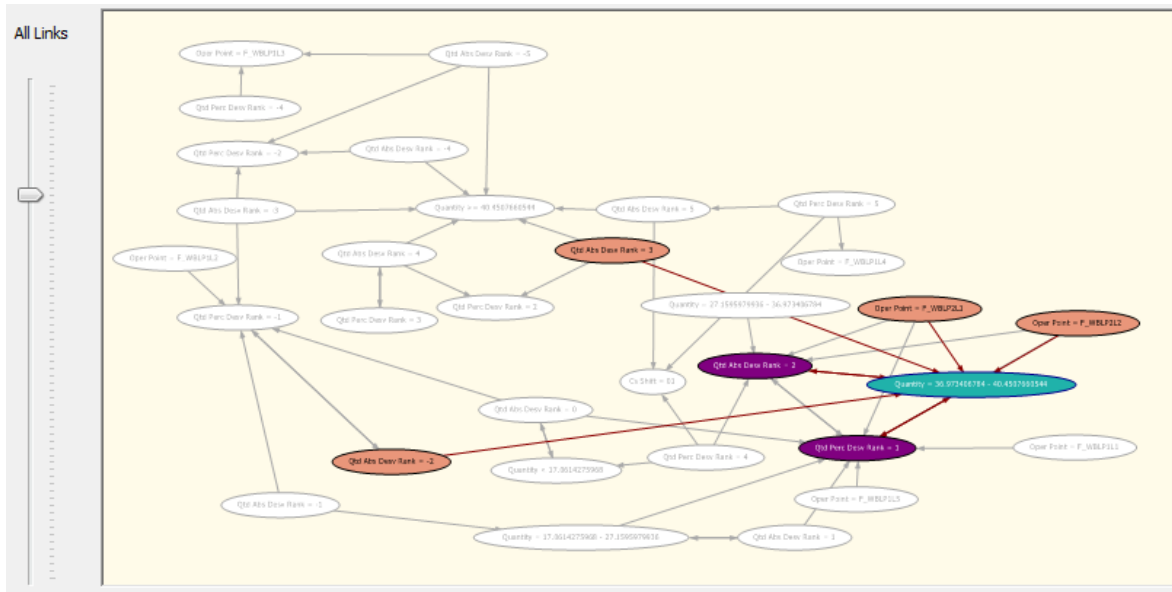


Imagem 31 - Rede de dependências para o modelo *Carregamento_Assoc*

Numa análise mais detalhada sobre os inúmeros relacionamentos sugeridos, é possível identificar um conjunto de associações que merecem melhor reflexão.

As linhas de carregamento "F_WBLP2L1" e "F_WBLP2L2", para operações cuja quantidade pedida se encontre entre [36.97 – 40.45] toneladas, apresentam um desempenho semelhante onde, o desvio absoluto se encontra no primeiro nível de alerta severo. Tendo em consideração que, do posto de vista operacional estas duas linhas são "alimentadas" por um ponto comum de ensacagem, esta relação poderá indiciar uma ligeira desregulação dos sistemas de enchimento.



Imagem 32 – Rede de dependências para quantidades entre [36.97 – 40.45] toneladas

Um outro ponto de relevante importância prende-se com o forte relacionamento entre o intervalo de quantidades [36.97 – 40.45] toneladas e os rankings percentuais "1" e absolutos de desvios "2". Atendendo a que, como apresentado na Imagem 33, este grupo de quantidades corresponde a 37%, poderá representar o comportamento sobre o qual a fábrica se encontra a operar.

Dispersão de Viaturas

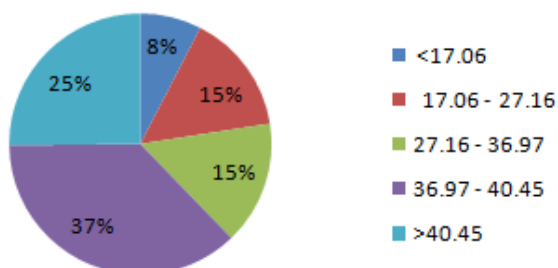


Imagem 33 - Dispersão de viaturas segundo intervalos de quantidades do modelo

Relativamente ao turno "01", constata-se que tem uma relativa propensão para se associar a desvios de risco, na medida em que, mesmo para operações de quantidades inferiores a 17.06 toneladas, é possível estabelecer uma relação entre o desvio percentual "4" e o turno "01".

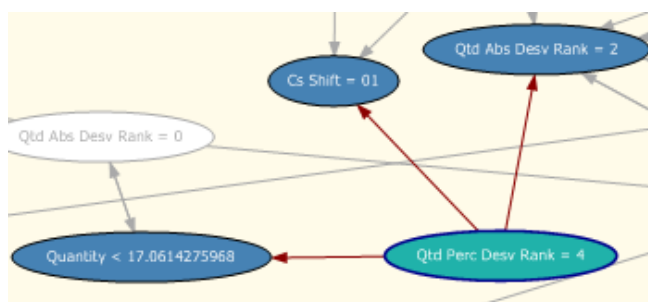


Imagem 34 - Rede de dependências para ranking de desvio percentual "4"

Para operações superiores a 40 toneladas a situação agrava-se, verificando-se que nestes casos, e lembrando que, estamos a considerar um conjunto de operações inferior a 1% do total de operações, os desvios absolutos representam o nível máximo "5" indicando a possibilidade de existência de

problemas de regulação das quantidades. Ainda neste contexto, destaca-se a linha "F_WBLP1L4" confirmando o indicador de *performance* da Imagem 12, que realçava indícios de ineficiências de desempenho.

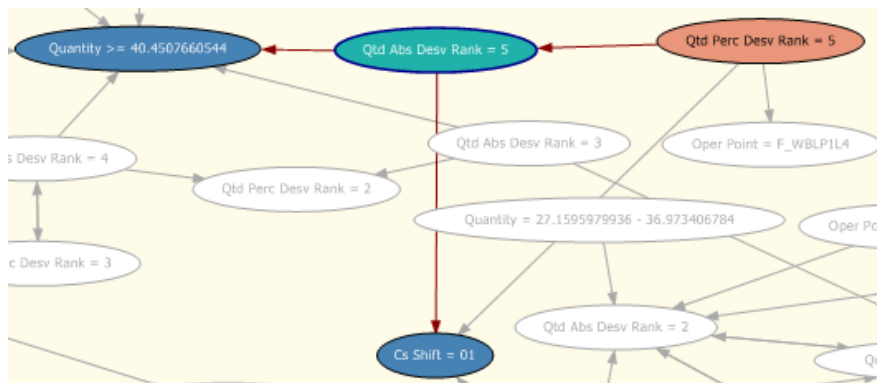


Imagem 35 - Rede de dependências para ranking de desvio absoluto "5"

Embora não associado a qualquer turno ou linha de operação, a possibilidade das linhas estarem a funcionar nos limites de autorização, assume alguma relevância para operações onde a quantidade movimentada se encontra acima das 40.45 toneladas se considerados os seguintes pontos:

- O ranking de desvio "3" corresponde ao limite a partir do qual, é operacionalmente necessária a autorização de saída de viatura por responsáveis da portaria.
- Embora de forma indirecta e não evidente, o ranking de desvio percentual "3" encontra-se relacionado com operações com quantidades superiores a 40.45 toneladas.

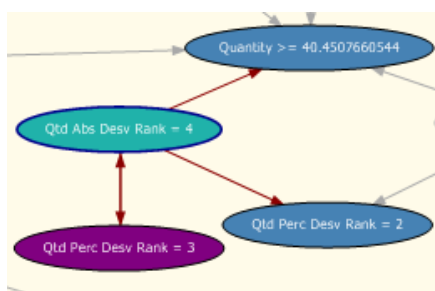


Imagem 36 - Rede de dependências para ranking de desvio absoluto "4"

Tal como verificado na Imagem 29, onde existia um complemento entre os *rankings* de desvios, na Imagem 37 verifica-se que o ranking de desvio absoluto "-5", também se encontra relacionado com

operações de quantidades superiores a 40.45 toneladas. Aqui, a linha de carga "F_WBLP1L3" realça um relacionamento de risco: o *ranking* percentual "-4" indicando que se tende a operar acima do limite máximo de autorizações.

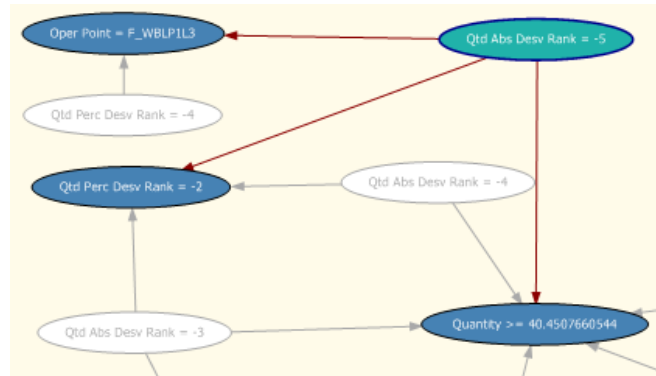


Imagem 37 - Rede de dependências para ranking desvio absoluto "-5"

Também para o grupo de quantidades entre [17.06, 27.16] toneladas se verifica um cenário de complementaridade onde, embora a relação mais forte seja os rankings de desvios positivos, na verdade, o ranking de desvio absoluto "-1" também apresenta alguma representatividade.



Imagem 38 - Rede de dependências para quantidades entre 17.06 e 27.16 toneladas

Neste ponto, sentiu-se a necessidade de aprofundar os critérios de relacionamento, contemplando-se no processo, o grupo de gestores responsáveis pela gestão das linhas de enchimento. Interessava perceber quais os desvios de quantidade por sacco e a sua relação com linhas de carga, desvios globais e turnos de operação. As relações entre a quantidade esperada das máquinas de enchimento e a quantidade efectiva na execução, permitirão confrontar os interesses de desempenho de cada uma das áreas de negócio, na procura de um valor de equilíbrio para os desvios. Se, por um lado, o ranking de desvio absoluto de quantidade representa os interesses de gestão ao nível global da

empresa, por outro, o desvio absoluto ao nível de cada sacco irá representar os interesses e eficiência na gestão dos processos de enchimento.

Assim, foi desenvolvido o modelo mineração *Carregamento_Class*, no sentido de classificar a origem dos desvios assim como, compreender possíveis causas. Através da indução de árvores de decisão, determinaram-se regras que de uma forma visual e intuitiva, descrevem o modo como as decisões de classificação foram tomadas.

A relação de dependências ilustrada na Imagem 39, revela um forte relacionamento entre quantidade a requisitada e os diferentes tipos de desvios considerando-se:

- Desvios ao nível do ranking absoluto de quantidades tendem a associar-se a turnos de operação.
- Desvios ao nível da quantidade individual de cada sacco tendem a associar-se a linhas de operação.

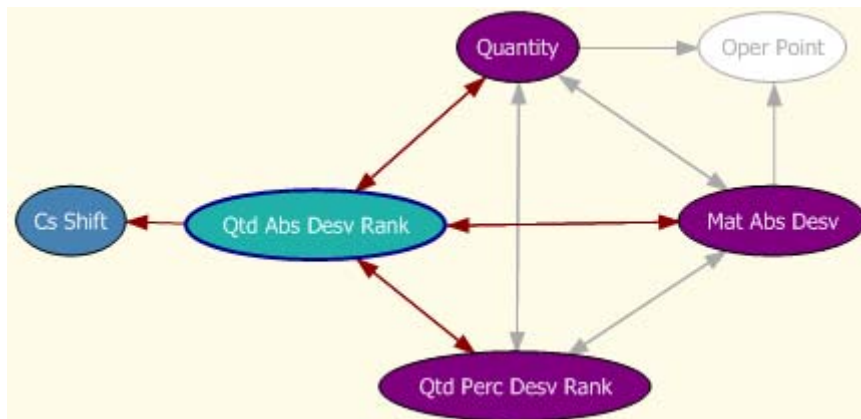


Imagem 39 - Relação de dependências relativas ao modelo *Carregamento_Class*

Compreendidos os relacionamentos entre cada um dos atributos em análise, apresenta-se na Imagem 40, a árvore de decisão relativa ao desvio absoluto de produto. Pretende-se compreender os factores que induzem o funcionamento nas regiões de alerta ou intervenção urgente.

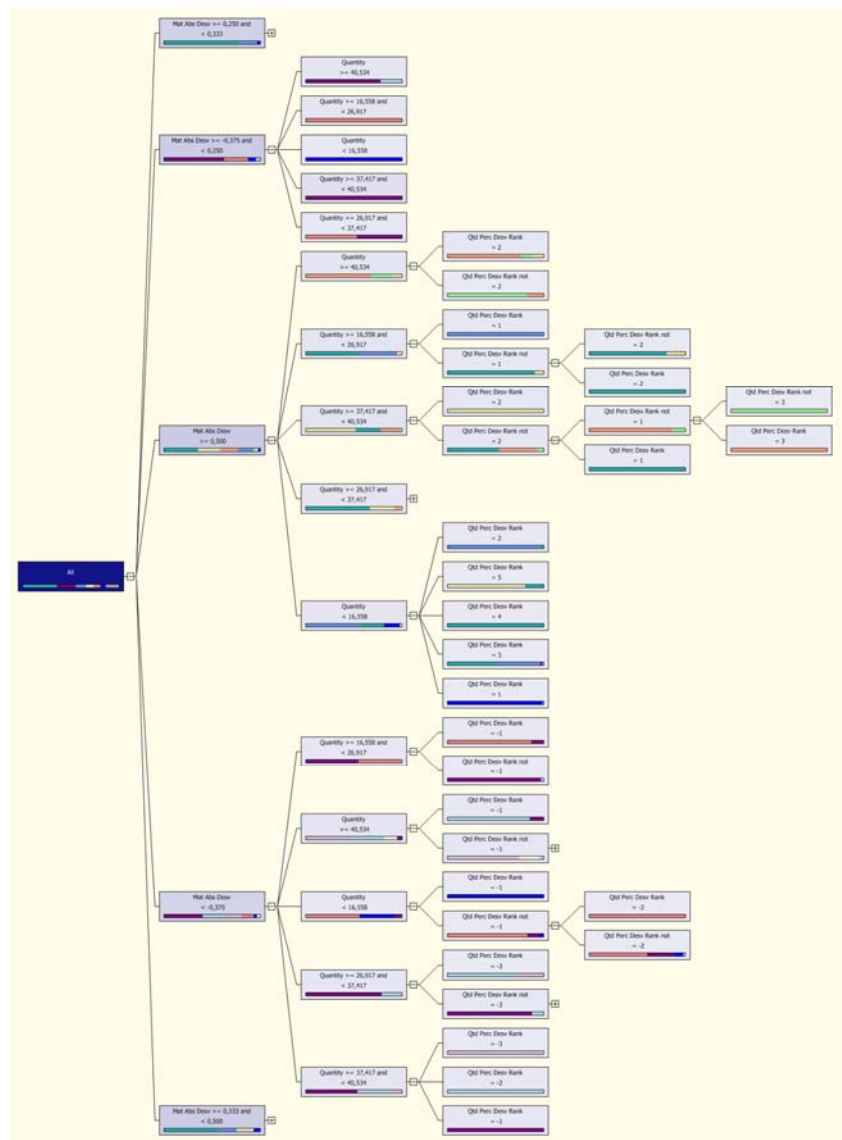


Imagem 40 - Árvore de decisão relativa ao *ranking desvio absoluto*

Do lote de regras geradas, foi dada especial atenção ao nodo "*Mat abs desv >= 0.435 and Mat abs desv < 0.757*", dado revelar uma disfunção total do processo de carregamento e com grande impacto ao nível do desvio de quantidades. Note-se que, o intervalo em causa equivale a cerca de 10 % do peso unitário de um saco de cimento, e está totalmente desfasado do valor de 2% máximo, definido operacionalmente para saída de viaturas sem autorização explícita.

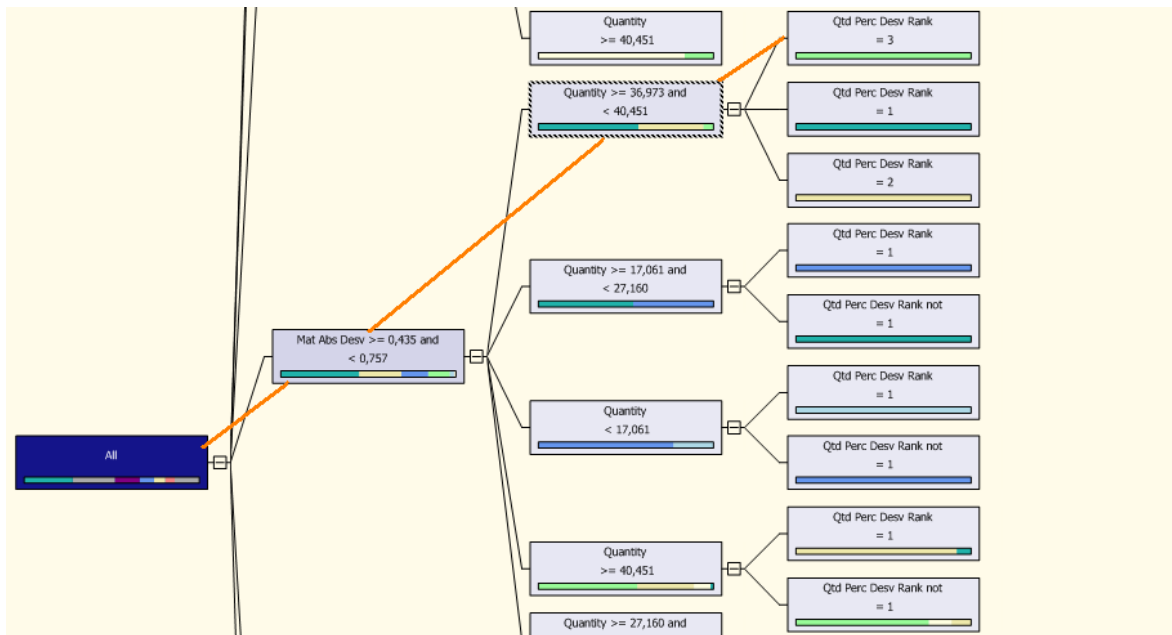


Imagem 41 - Regras de decisão para nó "Mat abs desv >= 0.435 and Mat abs desv < 0.757"

Embora possa parecer óbvia a relação apresentada pelos desvios de produto e globais (percentual ou absoluto), não deverá ser considerada a noção mais evidente: "Se existe um desvio elevado em cada sacco e a quantidade é superior a 40 toneladas, então, o desvio final vai ser elevado". Neste caso específico, revela-se um aspecto relevante e não evidente, que pode colocar em causa a qualidade de serviço inerente ao carregamento. Considerando a linha vermelha representada na Imagem 41, verifica-se:

- Cerca de 17 % do total de operações em estudo, foram executadas num regime de alerta relativamente ao controlo de ensacagem (representatividade do nó "Mat abs desv >= 0.435 and Mat abs desv < 0.757").
- Cerca de 7 % são relativas a operações entre [36.97 – 40.45[toneladas.
- Os *rankings* de desvio absoluto posicionam-se claramente na região de *alerta severo e intervenção urgente* como demonstrado em Imagem 42.
- Os *rankings* de desvio percentual não evidenciam cenários de intervenção urgente face à representatividade dos casos na totalidade de operações em estudo:
 - *Ranking* "1" → 4,13 %.
 - *Ranking* "2" → 2,73 %.
 - *Ranking* "3" → 0,43 %.
 - *Ranking* "4" ou superior → sem representatividade.

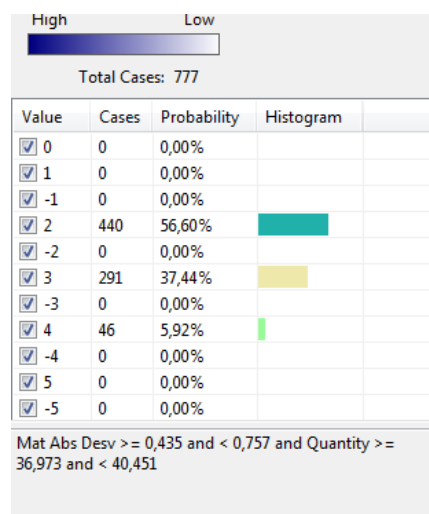


Imagem 42 - Detalhe do nodo quantidade [36.97 – 40.45]

O facto de, nestas condições não serem lançados alertas de desvios percentuais superiores a 2 %, e as análises de desvios operacionais tenderem a focalizar-se em relatórios globais de desempenho, este caso apresentado iria "diluir-se" no conjunto de operações em alerta, perdendo toda a expressividade. Por outro lado, uma análise operacional focalizada nos desvios absolutos, teria algumas dificuldades em fornecer uma informação tão precisa acerca do funcionamentos de risco relativo a desvios no produto ensacado como o apresentado.

5.3 Análise de resultados

O desenvolvimento e aplicação dos modelos protótipo apresentados, embora restritos a uma área específica dos processos logísticos de movimentação numa unidade cimenteira, permitiram validar a abordagem e ajuste das técnicas de mineração para optimização da qualidade de serviço. A abordagem proposta contemplou a utilização de duas tarefas:

- a) Associação de linhas de ensacagem a desvios de operação ou execução;
- b) Classificação de desvios focalizados nos desvios absolutos inerentes a cada operação logística.

A articulação de ambas, resultou como um processo natural na identificação de associações de risco traduzindo-se, assim, num procedimento útil para o controlo das áreas inerentes ao carregamento do produto ensacado.

Durante o processo de descoberta de conhecimento, foram identificados problemas relacionados não só com o conteúdo de alguns dados, mas também, com o reduzido número de registos. Problemas associados ao conteúdo de dados foram ultrapassados, de acordo com os *processos de preparação* apresentados no capítulo 4. No entanto, relativamente ao número de registos disponíveis para estudo, foi tomada a decisão de eliminação de casos, cujo ranking de funcionamento estava no nível ideal de funcionamento. Opções como aumento de registos na base de dados utilizando técnicas de clonagem balanceada, à semelhança do utilizado em [Botica, 2005] foram todavia considerados. O restante conjunto mostrou-se representativo para análise, obtendo-se bons resultados no final do processo de descoberta de conhecimento.

Os relacionamentos entre atributos e classes de valores suscitaram questões interessantes. Se por um lado, confirmaram o conhecimento consensual, por outro, poderão abrir horizontes a aplicações diversificadas, nomeadamente ao nível da implementação de sistemas inteligentes de controlo. Se considerarmos que os processos de automatização são graduais e faseados no tempo, este tipo de abordagem potenciará, num curto prazo, a optimização do processo, segundo duas vertentes distintas:

- Pró-actividade de sistemas – Soluções como SLV, poderão actuar de forma diferenciada mediante o contexto em que se encontram, passando por exemplo, pela criação automática de notificações para auditoria de qualidade de produto.
- Monitorização de proximidade – Os gestores poderão actuar sobre o processo, tendo por base variáveis totalmente definidas.

No entanto, a utilização poderá ser mais abrangente. Considerando que os grupos cimenteiros apresentam uma tendência de globalização, e que o *SLV-BI* está preparado para funcionar com várias fontes de dados, a abordagem proposta permitirá análises de desempenho globais, onde nestes campos, os *ERP's* como SAP não estão focados.

Capítulo 6

Conclusões e Trabalho Futuro

A dependência actual de soluções de descoberta de conhecimento, torna fundamental a integração de todos os fluxos de informação disponíveis numa empresa, tendo em vista o aumento dos índices de qualidade e satisfação de todos os intervenientes no seu ciclo de negócio. Nas últimas décadas tem-se assistido a uma informatização sistemática desses mesmos fluxos, o que conduziu, inevitavelmente, à geração e armazenamento de grandes volumes de informação, bem como, ao aparecimento de algumas técnicas especialmente focalizadas na exploração desses dados. Frequentemente, estas duas vertentes têm servido como motor de crescimento das empresas, onde a obtenção de informação necessária, assim como a possibilidade de visualização dos dados em diferentes perspectivas, e em tempo útil, permitem aos seus gestores suporte efectivo na tomada de decisões. Infelizmente, face à necessidade de informatização dos seus diferentes sectores, as empresas foram, também, adoptando diferentes sistemas operacionais, não integrados, e com um crescimento acelerado das suas bases de dados.

A excessiva dependência dos sistemas operacionais, aliada à dispersão de dados pelas empresas, não contribui para a rentabilização necessária do desempenho dos agentes de decisão e, conseqüentemente, para a efectividade dos processos de tomada de decisão. Face aos novos desafios de mercado, a tomada de decisões é um factor crítico que exige respostas rápidas dos seus gestores. Deste modo, a necessidade de uma visão global on-line das suas áreas de negócio, orientou as empresas para a implementação de sistemas de *Data Warehousing*, sistemas estes, desenhados especificamente para a análise de dados e suporte à decisão. A análise prévia dos requisitos dos

agentes de decisão, permite a materialização de todos os resultados com tempos de respostas praticamente imediatos, independentemente da complexidade ou eixo de análise que se considere, uma vez que todas as possíveis combinações de análises já se encontram calculadas e à distância de um “clique” de rato. Em complemento, a informação fica disponível de forma organizada e consolidada, podendo ser utilizada pelos analistas para extrair conhecimento útil para a empresa.

Com este trabalho pretendeu-se portanto, demonstrar as vantagens da aplicação de técnicas de mineração de dados numa área de negócio estrutural das economias actuais: a indústria cimenteira. Assim, e no sentido de se assegurar a viabilidade e visibilidade do modelo desenvolvido, existiu uma enorme preocupação inicial na procura de um contexto favorável para o acolhimento deste tipo de actividades. Através da forte relação de parceria entre as empresas *Cachapuz – Bilanciai Group* e *Cimpor – Cimentos de Portugal*, foi possível encontrar uma base de trabalho sustentada e com uma dinâmica própria permitindo o ajuste gradual dos modelos desenvolvidos. Por outro lado, o facto da Cachapuz estar em fase de evolução da sua plataforma analítica, permitiu criar um cenário de evolução e consolidação deste trabalho.

Os resultados obtidos permitiram demonstrar a aplicação e o enorme potencial desta abordagem na indústria cimenteira. Tomando como caso de estudo o processo de carregamento a saco, foi possível identificar cenários ou modos de funcionamento de risco, através da articulação conjunta de tarefas de associação e classificação. Devidamente acautelados, estes factores contribuirão para um claro aumento da qualidade e diferenciação do serviço prestado. Por outro lado, os resultados obtidos poderão potenciar uma melhor monitorização de proximidade por parte dos responsáveis industriais, assim como, facultar indicadores de processo que possam ser uma mais-valia no momento de tomada de decisão. Considerando-se novamente o processo de carregamento a saco, o facto do modelo apresentar um intervalo claro de desvio associado a cada saco, assim como a sua representatividade no conjunto total de operações em estudo, fornecerá ao gestor uma base sustentada para a tomada de decisões:

- Com base no intervalo apresentado, é possível para a fábrica tornar o processo mais eficiente. Neste caso, o gestor poderá despoletar um processo de controlo mais rigoroso como por exemplo: o controlo de proximidade; ou
- O limite apresentado corresponde ao ultrapassa os limites de recursos que a fábrica pode dispor. Neste caso, e face à representatividade dos valores, o gestor poderá equacionar a

remodelação ou introdução de novos processos e automatismos para controlo do processo de enchimento ou carregamento.

O modelo apresentado deverá ser encarado como o primeiro passo na evolução e incorporação de novos modelos, que possam responder às necessidades efectivas da logística de movimentação inerentes à indústria cimenteira. Tal como apresentado na secção 3.3, são várias as áreas onde as tarefas de mineração poderão ser aplicadas, com melhoria da qualidade de serviço e eficiência da organização.

Para concluir, por fim, podemos dizer que o retorno do investimento neste tipo de soluções, no caso particular da indústria cimenteira abordada nesta tese, é bastante interessante quando comparado com o esforço dispendido no estudo e implementação dos processos de mineração de dados em questão. Todavia, não se pense que este é mais um dos processos informáticos de fácil implementação. Isso não é verdade, uma vez que requer conhecimentos bastante avançados, não só ao nível dos processos de mineração de dados e modelos subjacentes, como também, na análise dos sistemas operacionais fontes, e em todos os processos de preparação e limpeza de dados, que normalmente são intervenientes importantes em tudo isto. De qualquer forma, com os conhecimentos certos sobre os modelos de negócio, as técnicas de mineração de dados e os dados alvo, temos com certeza os ingredientes certos para um excelente resultado. Só teremos que os dosear nas quantidades certas. Mas isso não é novidade.

Bibliografia

[Agarwal et al. 96] Sameet Agarwal, Rakesh Agrawal, Prasad M. Deshpande, Ashish Gupta, Jeffrey F. Naughton, Raghu Ramakrishnan e Sunita Sarawagi: "On the Computation of Multidimensional Aggregates". Proceedings of the 22th International Conference on Very Large Databases. Mumbai (Bombaim), Índia. Setembro, 1996.

[Anand et al., 1995] Sarabjot S. Anand., David A. Bell and John G. Hughes, University of Ulster, 1995.

[Agrawal et al. 97] Rakesh Agrawal, Ashish Gupta, Sunita Sarawagi. "Modeling Multidimensional Databases". 13th International Conference on Data Engineering. Birmingham, England. Abril, 1997.

[Alcantara, 2000] – Rosane Lucia Chicarelli Alcantara, "A integração das estratégias de Logística e Marketing maximizando o Serviço ao cliente: algumas reflexões", Universidade Federal de São Carlos, 2000.

[Berry et al., 2005] – Micheal J. A. Berry, Gordon S. Linoff, "Data Mining Techniques For Marketing Sales, and Customer Relationship Management Second Edition", Wiley, 2004.

[Botica, 2004] Natália Maria da Costa Botica, "Servator - modelo preditivo de apoio à prospecção arqueológica", Universidade do Minho, 2004.

[Chapman et al, 2000] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, "Crisp-DM 1.0: Step-by-step data mining methods", CRISP-DM Consortium, 2000.

[Cunico, 2005] Luis H. B. Cunico, "Técnicas em data mining aplicadas na predição de satisfação de funcionários de uma rede de lojas do comércio varejista.", Universidade Federal do Panamá, 2005.

[Farah, 2002] - Moisés Farah Jr., "Os desafios da logística e os centros de distribuição física", (revista FAE Business, n.2), Jun. 2002.

[Fayyad et al. 1996] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth: "From Data Mining to Knowledge Discovery", AAAI Press/MIT Press, 1996.

[Freitas et al., 2005] Cássio de Freitas, Davi Nascimento, Marconi Couto, Luiz Santana, "Data Mining: Conceitos, Técnicas e Aplicações", Universidade Federal da Bahia (UFBA), 2005.

[Kureski et al. 2007] Ricardo Kureski, Rossana Lott Rodrigues, Antonio Carlos Moretto Umberto Antonio Sesso Filho, Letícia Peret Antunes Hardt, "O macro sector da construção civil na economia brasileira em 2004", 2007.

[Navega, 2001] Sergio C. Navega, "Inteligência Artificial: Presente, Passado e Futuro", Anais do INFOIMAGEM 2001, 2001.

[Olsen, 2008] David L. Olson, "Advanced Data Mining Techniques, Springer, 2008.

[Pinto et al., 2004] Filipe Pinto, Manuel Filipe Santos, Paulo Cortez, Hélder Quintela, "Data Pre-processing for Database Marketing", IP Leiria, 2004.

[Prochnik et al. 1998], Victor Prochnik, Adriana Perez, Carla Maria de Souza e Silva, "A globalização na indústria do cimento", 1998.

[Reinardi, 2008] Vincent Reinardi, "BUILDING A DATAWAREHOUSE With SQL Server", APRESS, 2008.

[Sumathi et al., 2006] - S. Sumathi, S. Sivanandam, "Introduction to Data Mining and its Applications", Springer, 2006.

[Tang et al., 2005], Zhao Hui Tang and Jamie MacLennan, "Data Mining with SQL Server 2005", Wiley, 2005.

[Teixeira, 2005] Hélder Adalberto da Costa Quintela Teixeira, "Sistemas de conhecimento baseados em data mining: aplicação à análise da estabilidade de estruturas metálicas", Universidade do Minho, 2005.

[Torgal et al., 2005] F. P. Torgal, J. P. Castro Gomes, Said Jalali, "Ligantes geopoliméricos: uma alternativa ambiental ao cimento Portland no contexto da economia do carbono", APEB. Associação Portuguesa das Empresas de Betão Pronto, Maio 2005.

[Wang et al. 2007] Xiaohong Wang and Li Sun, "Study of MES for Cement Industry", IEEE, 2007.

Referências WWW

- [1] "CASE STUDY: Egypt", International Development Research Centre,
http://www.idrc.ca/in_focus_competition/ev-123832-201-1-DO_TOPIC.html#, 20 de Novembro de 2008.
- [2] "Business-Intelligence Dashboards Get Democratic", Computer World
<http://www.computerworld.com/databasetopics/data/story/0,10801,82064,00.html>, 20 de Novembro de 2008.
- [3] "IBM Cognos 8 Business Intelligence – Dashboards", Cognos IBM Company,
<http://www.cognos.com/products/cognos8businessintelligence/dashboards.html>, 20 de Novembro de 2008.
- [4] "Cement Manufacture", Cimpor – Cimentos de Portugal, S.A.
http://www.cimpor.pt/artigo_generico.aspx?sid=bbc0ec-7f1b-401e-9da3-254ea9c01c9a&cntx=7ZHUaMzZaU07mXxT4QpqRdcKKFeFqJ9LsTBK6qkJTwM%3D, 17 de Agosto de 2008.
- [5] "SAP Mill Industry – Overview, Vision and Strategy", SAP
http://www.sap.com/westbalkans/company/events/worldtour08/pdf/P1/Ahmed_Rasool_Mill_Products.pdf, 20 de Novembro de 2008.

- [6] "Presença Internacional", Cimpor – Cimentos de Portugal, S.A.,
http://www.cimpor.pt/listagem_generica.aspx?sid=373f4253-527d-45dc-8b57-9fb3b3097565&cntx=mUVEkcPdJaOmoM7vBGntN72a5IgcjN9ivv%2BcbVKpGEg%3D, 20 de Novembro de 2008.
- [7] "Lafarge dans le monde", Lafarge Groupe,
http://www.lafarge.fr/wps/portal/1_2_2-Lafarge_dans_le_monde, 24 de Novembro de 2008.
- [8] "Polls : Data Mining Methodology (Aug 2007)", KDnuggets,
http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm, 24 de Novembro de 2008.
- [9] "CRoss Industry Standard Processfor Data Mining", Crisp-DM web-site,
<http://www.crispdm.org>, 24 de Novembro de 2008.
- [10] "SAS – Case Studies", SAS,
<http://www.sas.com/success/indexByTechnology.html>, 24 de Novembro de 2008.
- [11] "Mineração de dados", Wikipédia,
http://pt.wikipedia.org/wiki/Data_Mining, acessido em 20 de Novembro de 2008.
- [12] "Unearth the New Data Mining Features of Analysis Services 2005", MSDN Magazine,
<http://msdn.microsoft.com/en-us/magazine/cc300503.aspx>, 28 de Junho de 2008.
- [13] "Content Types (Data Mining)", Microsoft Corporation,
[http://msdn.microsoft.com/en-us/library/ms174572\(SQL.90\).aspx](http://msdn.microsoft.com/en-us/library/ms174572(SQL.90).aspx), acessido a 26 de Novembro de 2008
- [14] "Cachapuz – SLV Cement", Cachapuz – Bilanciai Group,
<http://www.cachapuz.com/cachapuzsolutions/PortalRender.aspx?PageID={80594144-30b4-11dd-b84f-00112f582a69}>, acessido em 26 de Novembro de 2008.
- [15] "Cimpor Home Page", Cimpor – Cimentos de Portugal,
<http://www.cimpor.pt/>, acessido em 26 de Novembro de 2008.