

Guia do usuário – CRISP-DM

Sumário

Guia do usuário do CRISP-DM	3
1 - Entendimento do negócio	4
2 - Entendimento dos dados	13
3 - Preparação de dados	18
4 - Modelagem	24
5 - Avaliação	29
6 - Implantação	33

Guia do usuário do CRISP-DM

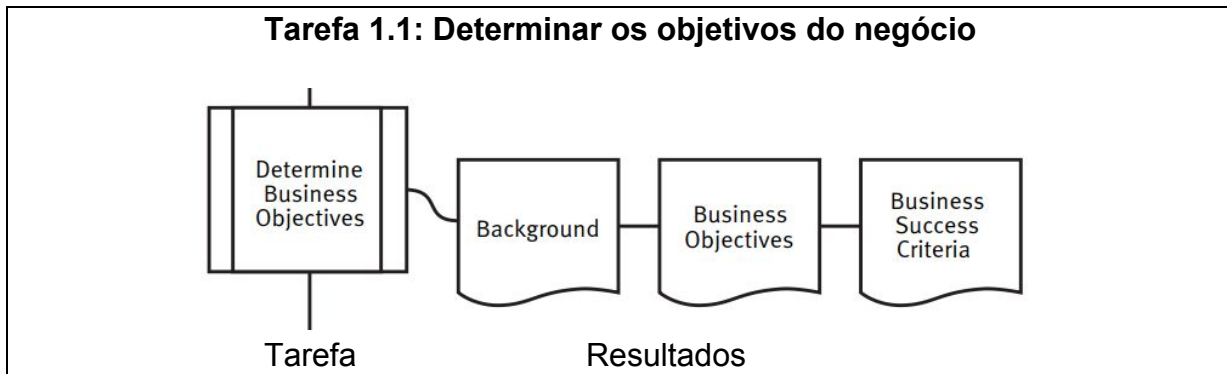
Horizontalmente, a metodologia CRISP-DM distingue entre o modelo de referência e o guia do usuário. O modelo de referência apresenta uma visão geral das fases, tarefas e resultados, e descreve o que fazer em um projeto de mineração de dados. O guia do usuário fornece dicas e dicas mais detalhadas para cada fase e cada tarefa dentro de uma fase, e descreve como realizar uma mineração de dados projeto.

Esse guia detalha todas as atividades que devem ser realizadas dentro das 6 fases do processo CRISP-DM, ilustradas na Figura 1, para o planejamento do projeto.



Figura 1: Fases do CRISP-DM

1 - Entendimento do negócio



Tarefa 1.1 – Determinar os objetivos do negócio: O primeiro objetivo do analista é entender completamente o que o cliente realmente quer. Muitas vezes, o cliente tem muitos objetivos e restrições concorrentes que devem ser adequadamente equilibrados. O objetivo do analista é descobrir fatores importantes no início do projeto que possam influenciar o resultado final. Uma consequência provável de negligenciar este passo seria gastar um grande esforço produzindo as respostas corretas para as perguntas erradas.

Saída 1.1.1 – Background: Coletar as informações conhecidas sobre a situação de negócio da organização no início do projeto. Esses detalhes não só servem para identificar mais de perto os objetivos de negócios a serem alcançados, mas também para identificar recursos, tanto humanos como materiais, que podem ser usados ou necessários durante o curso do projeto.

Atividades

Em relação a organização:

- Desenvolver organogramas identificando divisões, departamentos e grupos de projetos. O gráfico também deve identificar nomes e responsabilidades dos gerentes
- Identificar pessoas-chave no negócio e seus papéis
- Identificar um patrocinador interno (patrocinador financeiro e principal usuário / especialista em domínio)
- Indique se existe um comitê de direção e membros da lista
- Identificar as unidades de negócios que são afetadas pelo projeto de mineração de dados (por exemplo, Marketing, Vendas, Finanças)

Em relação a área de problema:

- Identificar a área do problema (por exemplo, marketing, atendimento ao cliente, desenvolvimento de negócios, etc.)
- Descreva o problema em termos gerais
- Verifique o estado atual do projeto (por exemplo, verifique se já está limpo dentro da unidade de negócios que um projeto de mineração de dados deve ser executado)

ou se a mineração de dados precisa ser promovida como uma tecnologia chave no negócio)

- Esclarecer os pré-requisitos do projeto (por exemplo, qual é a motivação do projeto? A empresa já usa mineração de dados?)
- Se necessário, preparar apresentações e mineração de dados presente para o negócio
- Identificar grupos-alvo para o resultado do projeto (por exemplo, esperamos entregar um relatório para gerenciamento superior ou um sistema operacional para usuários finais ingênuos?)
- Identificar as necessidades e expectativas dos usuários

Em relação a solução atual:

- Descreva qualquer solução usada atualmente para resolver o problema
- Descreva as vantagens e desvantagens da solução atual e o nível a que é aceito pelos usuários

Saída 1.1.2 - Objetivos do Negócio: Descreva o objetivo principal do cliente, desde uma perspectiva de negócios. Além do objetivo principal do negócio, geralmente há uma grande quantidade de questões comerciais relacionadas que o cliente gostaria de abordar. Por exemplo, o principal objetivo comercial pode ser manter os clientes atuais ao prever quando são propensos a se mudar para um concorrente, enquanto um objetivo comercial secundário pode ser determinar se taxas mais baixas afetam apenas um segmento específico de clientes.

Atividades

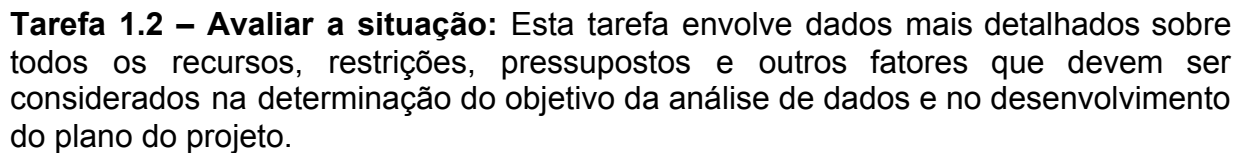
- Descreva informalmente o problema a ser resolvido
- Especificar todas as questões comerciais com a maior precisão possível
- Especificar outros requisitos de negócios (por exemplo, o negócio não quer perder nenhum cliente)
- Especificar os benefícios esperados em termos comerciais

Cuidado! Atenção com a definição de objetivos inalcançáveis tente torná-los tão realistas quanto possível!

Saída 1.1.3 - Critérios de sucesso do negócio: Descreva os critérios para um resultado bem sucedido ou útil para o projeto do ponto de vista do negócio. Isso pode ser bastante específico e facilmente mensurável, como a redução do *churn* do cliente para um determinado nível, ou geral e subjetivo, como "dar informações úteis sobre as relações". No último caso, certifique-se de indicar quem faria o subjetivo julgamento.

Atividades

- Ideia:** Antes de iniciar a avaliação da situação, você pode analisar experiências anteriores desse problema, tanto internamente, usando CRISP-DM, quanto externamente, usando soluções já existentes.



Atividades

Recursos de hardware

- Identificar o hardware base
- Estabelecer a disponibilidade do hardware básico para o projeto de mineração de dados
- Verifique se o cronograma de manutenção de hardware conflita com a disponibilidade do hardware para o projeto de mineração de dados
- Identifique o hardware disponível para a ferramenta de mineração de dados a ser usada (se a ferramenta for conhecida nesta fase)

Fontes de dados e conhecimento

- Identificar fontes de dados
- Identificar o tipo de fontes de dados (fontes on-line, especialistas, documentação escrita, etc.)

- Identificar fontes de conhecimento
- Identificar tipo de fontes de conhecimento (fontes on-line, especialistas, documentação escrita, etc.)
- Verifique as ferramentas e técnicas disponíveis
- Descreva o conhecimento de fundo relevante (informalmente ou formalmente)

Fontes pessoais

- Identificar o patrocinador do projeto
- Identificar o administrador do sistema, o administrador do banco de dados e a equipe de suporte técnico para mais perguntas
- Identificar analistas de mercado, especialistas em dados e especialistas em estatística e verificar sua disponibilidade
- Verifique a disponibilidade de especialistas do domínio para fases posteriores

Lembrar: o projeto pode precisar de equipe técnica em momentos estranhos ao longo do projeto, por exemplo durante a transformação de dados.

Saída 1.2.2 - Requisitos, suposições e restrições: Liste todos os requisitos do projeto, incluindo o cronograma de conclusão, compreensão e qualidade dos resultados e segurança, bem como questões legais. Como parte desta saída, certifique-se de que você tenha permissão para usar os dados.

Liste os pressupostos feitos pelo projeto. Estes podem ser pressupostos sobre os dados, que podem ser verificados durante a mineração de dados, mas também podem incluir pressupostos não verificáveis relacionados ao projeto. É particularmente importante listar estes últimos se afetarem a validade dos resultados.

Liste as restrições feitas no projeto. Essas restrições podem envolver a falta de recursos para realizar algumas das tarefas no projeto no tempo necessário, ou pode haver restrições legais ou éticas sobre o uso dos dados ou a solução necessária para realizar a tarefa de mineração de dados.

Atividades

Requisitos

- Especificar perfil de grupo de destino
- Capture todos os requisitos no agendamento
- Requisitos de captura de compreensão, precisão, habilidade de implantação, capacidade de manutenção e repetibilidade do projeto de mineração de dados e o (s) modelo (s) resultante (s)
- Requisitos de captura de segurança, restrições legais, privacidade, relatórios e cronograma do projeto

Premissas

- Esclarecer todos os pressupostos (incluindo os implícitos) e torná-los explícitos (por exemplo, para resolver a questão comercial, é necessário um número mínimo de clientes com idade acima de 50)
- Listar suposições sobre a qualidade dos dados (por exemplo, precisão, disponibilidade)
- Listar pressupostos sobre fatores (ex, modelo e os resultados podem ser apresentados à gerência / patrocinador sênior)

Restrições

- Verifique as restrições gerais (por exemplo, questões legais, orçamento, prazos e recursos)
- Verifique os direitos de acesso às fontes de dados (por exemplo, restrições de acesso, senha necessária)
- Verifique a acessibilidade técnica dos dados (sistemas operacionais, sistema de gerenciamento de dados, arquivo ou formato de banco de dados)
- Verifique se o conhecimento relevante está acessível
- Verificar restrições orçamentárias (custos fixos, custos de implementação, etc.)

Lembrar: A lista de premissas também inclui pressupostos no início do projeto, ou seja, qual o ponto de partida do projeto.

Saída 1.2.3 - Riscos e contingências: Liste os riscos, isto é, os eventos que podem ocorrer, impactando o cronograma, o custo ou o resultado. Liste os planos de contingência correspondentes: quais ações serão tomadas para evitar ou minimizar o impacto ou se recuperar da ocorrência dos riscos previstos.

Atividades

Identificar riscos

- Identificar riscos empresariais (por exemplo, o concorrente apresenta melhores resultados primeiro)
- Identificar riscos organizacionais (por exemplo, o projeto que solicita o departamento não possui financiamento para o projeto)
- Identificar riscos financeiros (por exemplo, o financiamento adicional depende dos resultados iniciais da mineração de dados)
- Identificar riscos técnicos
- Identificar riscos que dependem de dados e fontes de dados (por exemplo, má qualidade e cobertura)

Desenvolver planos de contingência

- Determine as condições em que cada risco pode ocorrer
- Desenvolver planos de contingência

Saída 1.2.4 - Terminologia: Compile um glossário de terminologia relevante para o projeto. Isso deve incluir pelo menos dois componentes:

- (1) Um glossário da terminologia comercial relevante, que faz parte da compreensão comercial disponível para o projeto
- (2) Um glossário de terminologia de mineração de dados, ilustrado com exemplos relevantes para o problema comercial em questão

Atividades

- Verificar disponibilidade prévia de glossários; de outra forma, começar a elaborar glossários
- Fale com especialistas do domínio para entender sua terminologia
- Familiarizar-se com a terminologia do negócio

Saída 1.2.5 - Custos e benefícios: Prepare uma análise custo-benefício para o projeto, comparando os custos do projeto com os benefícios potenciais para o negócio se for bem sucedido

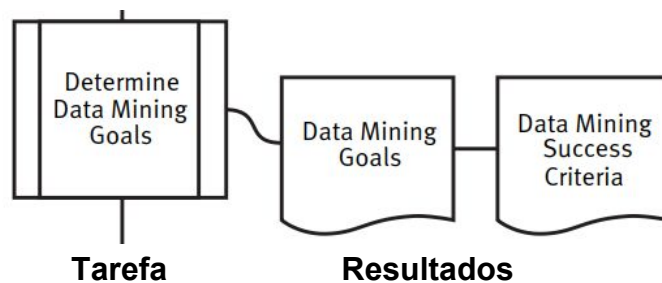
Atividades

- Estimar custos para a coleta de dados
- Estimar os custos de desenvolvimento e implementação de uma solução
- Identificar benefícios (por exemplo, melhor satisfação do cliente, ROI e aumento na receita)
- Estimar custos operacionais

Ideia: A comparação deve ser tão específica quanto possível, pois isso permite que um melhor negócio seja feito.

Cuidado! Lembre-se de identificar custos ocultos, como extração e preparação de dados repetidos, mudanças nos fluxos de trabalho e tempo necessário para treinamento.

Tarefa 1.3: Determinar os objetivos de mineração de dados



Tarefa 1.3 – Determinar os objetivos da mineração de dados: Um objetivo de negócios indica objetivos em termos de negócios; um objetivo de mineração de

dados indica os objetivos do projeto em termos técnicos. Por exemplo, o objetivo do negócio pode ser: "Aumentar as vendas de catálogo para clientes existentes", enquanto um objetivo de mineração de dados pode ser: "Preveja quantos widgets um cliente comprará, dado suas compras nos últimos três anos, informações demográficas relevantes e o preço do item".

Saída 1.3.1 - Resultados de mineração de dados: Descreva os resultados pretendidos do projeto que permitem a consecução dos objetivos de negócios. Note-se que estes são normalmente saídas técnicas.

Atividades

- Traduzir as questões empresariais para os objetivos de mineração de dados (por exemplo, uma campanha de marketing requer segmentação de clientes para decidir quem abordar nesta campanha, o nível / tamanho dos segmentos deve ser especificado).
- Especifique o tipo de problema de mineração de dados (por exemplo, classificação, descrição, previsão e agrupamento).

Ideia: Pode ser sensato redefinir o problema. Por exemplo, a modelagem da retenção de produtos, em vez da retenção de clientes, quando a focagem de retenção de clientes fornece resultados muito tarde para afetar o resultado.

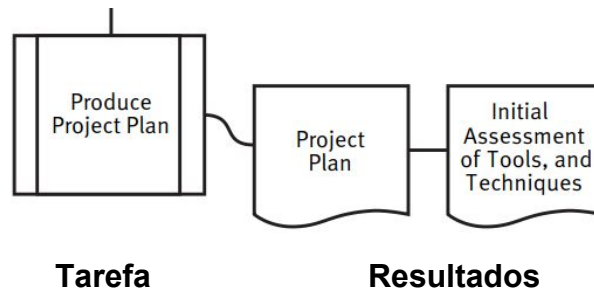
Saída 1.3.2 - Critérios de sucesso da mineração de dados: Defina os critérios para um resultado bem-sucedido para o projeto em termos técnicos, por exemplo, um certo nível de precisão preditiva ou um perfil de propensão para compra com um determinado grau de "elevação". Tal como acontece com os critérios de sucesso comercial, pode ser necessário descreva estes em termos subjetivos, caso em que a pessoa ou pessoas que fazem o julgamento subjetivo devem ser identificadas.

Atividades

- Especificar critérios para avaliação do modelo (por exemplo, precisão do modelo, desempenho e complexidade)
- Definir benchmarks para critérios de avaliação
- Especificar critérios que abordem critérios de avaliação subjetiva (por exemplo, capacidade de explicação do modelo e dados e visão de marketing fornecidos pelo modelo)

Cuidado! Lembre-se de que os critérios de sucesso da mineração de dados são diferentes dos critérios de sucesso comercial definidos anteriormente. Lembre-se de que é sábio planejar a implantação desde o início do projeto.

Tarefa 1.4 - Produzir o plano de projeto



Tarefa 1.4 – Produzir o plano de projeto: Descreva o plano pretendido para alcançar os objetivos de mineração de dados e assim atingir os objetivos comerciais.

Saída 1.4.1 - Plano do projeto: Liste as etapas a serem executadas no projeto, juntamente com a duração, recursos necessários, entradas, saídas e dependências. Sempre que possível, explique as iterações em larga escala no processo de mineração de dados, por exemplo, repetições das fases de modelagem e avaliação. Como parte do plano do projeto, também é importante analisar dependências entre cronograma e riscos. Marque os resultados dessas análises explicitamente no plano do projeto, idealmente com ações e recomendações para ações se os riscos se manifestarem.

Embora esta seja a única tarefa em que o plano do projeto é nomeado diretamente, ele deve ser consultado continuamente e revisado ao longo do projeto. O plano do projeto deve ser consultado no mínimo sempre que uma nova tarefa for iniciada ou uma nova iteração de uma tarefa ou atividade seja iniciada.

Atividades

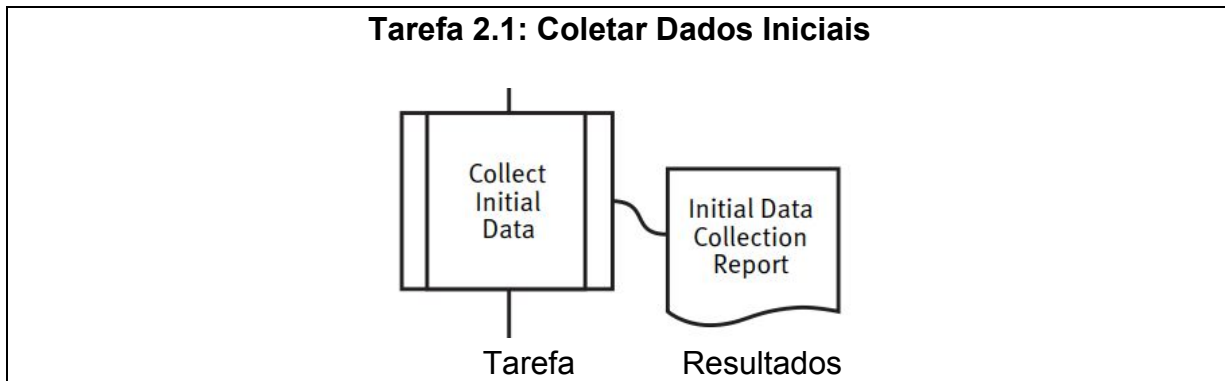
- Defina o plano de processo inicial e discuta a viabilidade com todo o pessoal envolvido
- Combinar todos os objetivos identificados e técnicas selecionadas em um procedimento coerente que resolva as questões comerciais e atenda aos critérios de sucesso comercial.
- Estimar o esforço e os recursos necessários para alcançar e implementar a solução. (É útil considerar a experiência de outras pessoas ao estimar prazos para projetos de mineração de dados. Por exemplo, muitas vezes postula-se que 50-70 % do tempo e esforço em um projeto de mineração de dados é usado na Fase de Preparação de Dados e 20-30% na Fase de Compreensão de Dados, enquanto apenas 10-20% são gastos em cada uma das Fases de Compreensão de Modelagem, Avaliação e Negócios e 5-10% na Fase de Implantação.)
- Identificar passos críticos
- Marcar pontos de decisão
- Marcar pontos de revisão
- Identificar iterações importantes

Saída 1.4.2 - Avaliação inicial de ferramentas e técnicas: No final da primeira fase, a equipe do projeto realiza uma avaliação inicial de ferramentas e técnicas. Aqui, é importante selecionar uma ferramenta de mineração de dados que suporte vários métodos para diferentes estágios do processo, uma vez que a seleção de ferramentas e técnicas pode influenciar todo o projeto.

Atividades

- Crie uma lista de critérios de seleção para ferramentas e técnicas (ou use uma existente se disponível)
- Escolha ferramentas e técnicas potenciais
- Avaliar a adequação das técnicas
- Revisar e priorizar as técnicas aplicáveis de acordo com a avaliação de soluções alternativas

2 - Entendimento dos dados



Tarefa 2.1 - Coletar dados iniciais: Adquira os dados (ou acesso aos dados) listados nos recursos do projeto.

Saída 2.1.1: Relatório inicial de produção de dados - Descreva os dados utilizados e inclua quaisquer requisitos de seleção de dados. O relatório de coleta de dados também deve definir se alguns atributos são relativamente mais importantes do que outros. Lembre-se de que qualquer avaliação da qualidade dos dados deve ser feita não apenas nas fontes de dados individuais, mas também nos dados resultantes da fusão de diferentes fontes de dados. Devido a inconsistências entre as fontes, os dados agrupados podem apresentar problemas que não existem nas fontes de dados individuais.

Atividades

Planejamento de requisitos de dados

- Planejar quais informações são necessárias (por exemplo, apenas para determinados atributos ou informações adicionais específicas)
- Verifique se todas as informações necessárias (para resolver os objetivos de mineração de dados) estão realmente disponíveis

Critério de seleção

- Especificar critérios de seleção (por exemplo, quais atributos são necessários para os objetivos de mineração de dados especificados? Quais atributos foram identificados como irrelevantes? Quantos attr Podemos lidar com as técnicas escolhidas?)
- Selecione tabelas / arquivos de interesse
- Selecionar dados em uma tabela / arquivo
- Pense em quanto tempo um histórico deve usar (por exemplo, mesmo que 18 meses de dados estejam disponíveis, apenas 12 meses podem ser necessários para o exercício)

Cuidado!

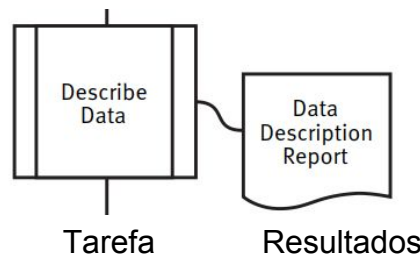
Esteja ciente de que os dados coletados de diferentes fontes podem dar origem a problemas de qualidade quando mesclados (por exemplo, arquivos de endereço mesclados com um banco de dados de clientes podem mostrar inconsistências de formato, invalidez de dados, etc.).

Inserção de dados

- Se os dados contiverem entradas de texto gratuitas, precisamos codificá-las para modelagem ou queremos agrupar entradas específicas?
- Como os atributos em falta podem ser adquiridos?
- Como podemos extrair os dados melhor?

Ideia: Lembre-se de que alguns conhecimentos sobre os dados podem estar disponíveis a partir de fontes não eletrônicas (por exemplo, de pessoas, texto impresso, etc.). Lembre-se de que pode ser necessário pré-processar os dados (dados da série temporal, médias ponderadas, etc.).

Tarefa 2.2 - Descrever os dados



Tarefa 2.2 – Descrever os dados: Examine as propriedades "brutas" dos dados adquiridos e relate os resultados.

Saída 2.2.1 - Relatório de descrição dos dados: Descreva os dados que foram adquiridos, incluindo o formato dos dados, a quantidade de dados (por exemplo, o número de registros e campos dentro de cada tabela), as identidades dos campos e quaisquer outros recursos de superfície que foram descobertos.

Atividades

Análise volumétrica de dados

- Identificar dados e método de captura
- Acessar fontes de dados
- Use análises estatísticas, se apropriado
- Relatório de tabelas e suas relações
- Verifique o volume de dados, o número de múltiplos, a complexidade
- Observe se os dados contêm entradas de texto gratuitas

Tipos e valores de atributos

- Verificar a acessibilidade e disponibilidade de atributos
- Verificar tipos de atributo (numérico, simbólico, taxonomia, etc.)
- Verificar limites de valores de atributo
- Analisar correlações de atributos
- Compreender o significado de cada atributo e valor do atributo em termos comerciais
- Para cada atributo, computa as estatísticas básicas (por exemplo, distribuição de computação, média, máx., Min, desvio padrão, variância, modo, skewness etc.)
- Analisar estatísticas básicas e relacionar os resultados com seu significado em termos comerciais
- Decida se o atributo é relevante para o objetivo específico de mineração de dados
- Determine se o significado do atributo é usado consistentemente
- Entrevista com especialistas do domínio para obter sua opinião de relevância atribuída
- Decida se é necessário equilibrar os dados (com base nas técnicas de modelagem a serem usadas)

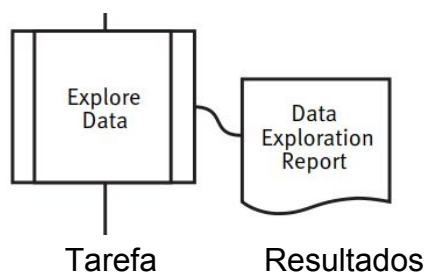
Chaves

- Analisar relacionamentos-chave
- Verifique a quantidade de sobreposições de valores de atributo de chave em tabelas

Revise os pressupostos / objetivos

- Atualizar a lista de premissas, se necessário

Tarefa 2.3 - Explorar os dados



Tarefa 2.3 - Explorar os dados: Esta tarefa aborda as questões de mineração de dados que podem ser abordadas usando técnicas de consulta, visualização e relatórios. Essas análises podem abordar diretamente os objetivos de mineração de dados. No entanto, eles também podem contribuir ou refinar a descrição dos dados e relatórios de qualidade e alimentar a transformação e outras etapas de preparação de dados necessárias antes que uma análise posterior possa ocorrer.

Saída 2.3.1 - Relatório de exploração de dados: Descreva os resultados desta tarefa, incluindo os primeiros achados ou hipóteses iniciais e seu impacto no restante do projeto. O relatório também pode incluir gráficos e gráficos que indicam

características de dados ou apontam para subconjuntos de dados interessantes que devem ser examinados.

Atividades

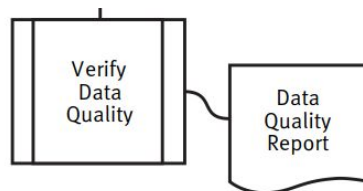
Exploração de dados

- Analisar as propriedades dos atributos interessantes em detalhes (por exemplo, estatísticas básicas, subpopulações interessantes)
- Identificar as características das subpopulações

Suposições de formulário para análise futura

- Considere e avalie informações e descobertas no relatório de descrições de dados
- Formar uma hipótese e identificar ações
- Transformar a hipótese em uma meta de mineração de dados, se possível
- Esclarecer os objetivos de mineração de dados ou torná-los mais precisos. Uma pesquisa "cega" não é necessariamente inútil, é preferível obter uma busca mais direta em direção aos objetivos comerciais.
- Execute análises básicas para verificar a hipótese

Tarefa 2.4 - Verificar a qualidade dos dados



Tarefa

Resultados

Tarefa 2.4 - Verificar a qualidade dos dados: Examine a qualidade dos dados, abordando questões como: os dados são completos (isso abrange todos os casos necessários)? É correto ou isso contém erros? Se há erros, como eles são comuns? Existem valores faltantes nos dados? Em caso afirmativo, como eles são representados, onde eles ocorrem e quão comuns são eles?

Saída 2.4.1 - Relatório de qualidade dos dados: Liste os resultados da verificação da qualidade dos dados; Se houver problemas de qualidade, liste possíveis soluções.

Atividades

- Identificar valores especiais e catalogar seu significado

Rever chaves, atributos

- Verifique a cobertura (por exemplo, se todos os valores possíveis estão representados)
- Verificar chaves
- Verifique se os significados de atributos e valores contidos se encaixam
- Identificar atributos ausentes e campos em branco
- Estabelecer o significado de dados ausentes
- Verifique se há atributos com valores diferentes que tenham significados semelhantes (por exemplo, baixo teor de gordura, dieta)
- Verifique a ortografia eo formato de valores (por exemplo, o mesmo valor, mas às vezes começa com uma letra minúscula, às vezes com uma letra maiúscula)
- Verifique os desvios e decida se um desvio é "ruído" ou pode indicar um fenômeno interessante
- Verifique a plausibilidade de valores (por exemplo, todos os campos que tenham os mesmos ou quase os mesmos valores)

Ideia: Analise todos os atributos que dão respostas que entram em conflito com o bom senso (por exemplo, adolescentes com altos níveis de renda). Use gráficos de visualização, histogramas, etc. para revelar inconsistências nos dados.

Qualidade de dados em arquivos planos

- Se os dados são armazenados em arquivos planos, verifique qual delimitador é usado e se ele é usado consistentemente em todos os atributos
- Se os dados são armazenados em arquivos planos, verifique o número de campos em cada registro para verificar se eles coincidem

Ruído e inconsistências entre fontes

- Verifique consistências e redundâncias entre diferentes fontes
- Planejar lidar com o barulho
- Detectar o tipo de ruído e quais atributos são afetados

Ideia: Lembre-se de que pode ser necessário excluir alguns dados, uma vez que eles não apresentam comportamento positivo ou negativo (por exemplo, para verificar o comportamento dos empréstimos dos clientes, excluir todos aqueles que nunca emprestaram, não financiam uma hipoteca de casa, aqueles cuja hipoteca é perto da maturidade, etc.).

Ideia: Avalie se os pressupostos são válidos ou não, atendendo às informações atualizadas sobre conhecimento de dados e negócios.

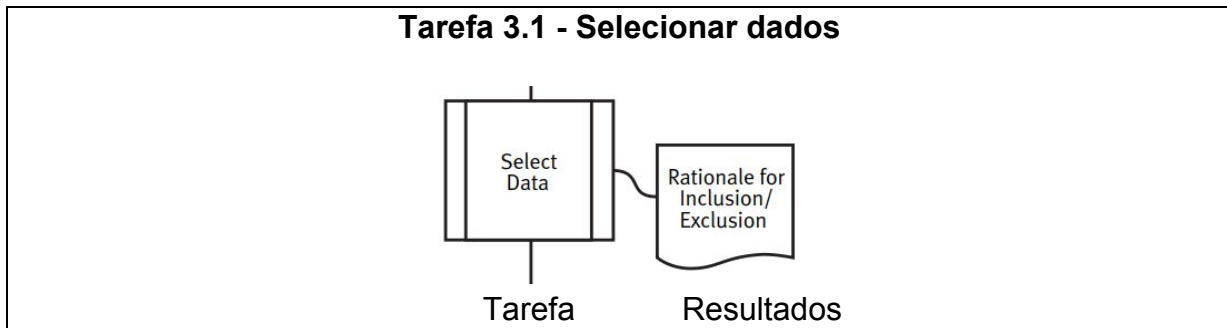
3 - Preparação de dados

Saída 3.1 - Conjunto de dados

Estes são os conjuntos de dados produzidos pela fase de preparação de dados, utilizados para modelagem ou para o trabalho de análise principal do projeto.

Saída 3.2 - Descrição do conjunto de dados

Esta é a descrição do (s) conjunto (s) de dados utilizados para a modelagem ou para o trabalho de análise principal do projeto.



Tarefa 3.1 - Selecionar dados: Decida que dados serão usados para análise. Os critérios incluem relevância para os objetivos de mineração de dados, qualidade e restrições técnicas, como limites de volume de dados ou tipos de dados.

Saída 3.1.1: Razão para inclusão / exclusão - Liste os dados que serão usados/ excluídos e os motivos para essas decisões.

Atividades

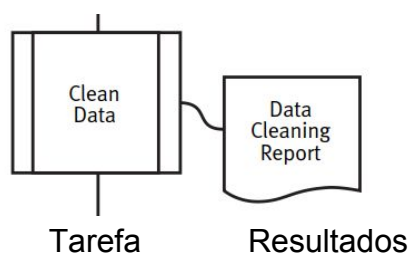
- Coletar dados adicionais apropriados (de diferentes fontes, tanto internamente como externamente)
- Execute testes de significância e correlação para decidir se os campos devem ser incluídos
- Reconsiderar critérios de seleção de dados (ver Tarefa 2.1) à luz de experiências de qualidade de dados e exploração de dados (ou seja, pode desejar incluir / excluir outros conjuntos de dados)
- Reconsiderar os critérios de seleção de dados (ver Tarefa 2.1) à luz da experiência de modelagem (ou seja, a avaliação do modelo pode mostrar que outros conjuntos de dados são necessários)
- Selecione diferentes subconjuntos de dados (por exemplo, atributos diferentes, apenas dados que atendem a certas condições)
- Considere o uso de técnicas de amostragem (por exemplo, uma solução rápida pode envolver testes de divisão e conjuntos de dados de treinamento ou reduzir o tamanho do conjunto de dados de teste, se a ferramenta não puder manipular o conjunto de dados completo. Também pode ser útil ter amostras ponderadas para dar

diferentes importância para diferentes atributos ou valores diferentes do mesmo atributo.)

- Documentar o raciocínio para inclusão / exclusão
- Verificar técnicas disponíveis para amostragem de dados

Ideia: Com base nos critérios de seleção de dados, decida se um ou mais atributos são mais importantes do que outros e ponderam os atributos em conformidade. Decida, com base no contexto (ou seja, aplicação, ferramenta, etc.), como lidar com a ponderação.

Tarefa 3.2 - Limpar dados



Tarefa 3.2 - Limpar dados: Elevar a qualidade dos dados ao nível exigido pelas técnicas de análise selecionadas. Isso pode envolver a seleção de subconjuntos limpos dos dados, a inserção de padrões adequados ou técnicas mais ambiciosas, como a estimativa de dados ausentes por modelagem.

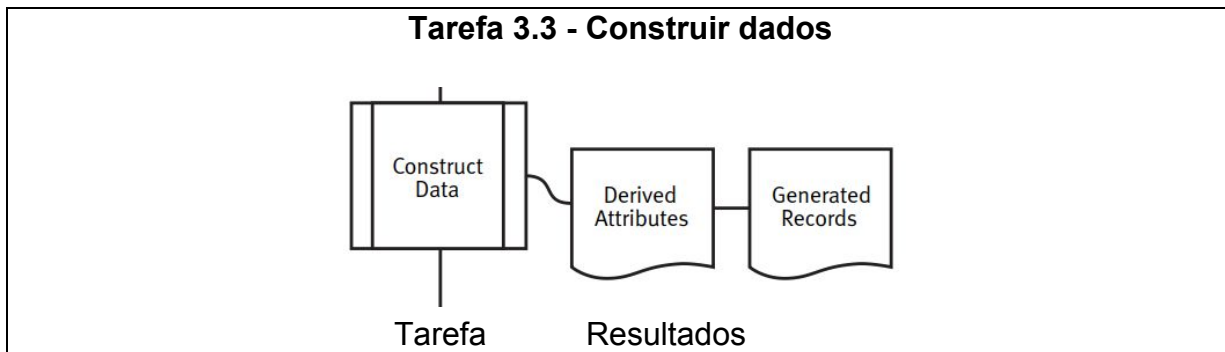
Saída 3.2.1 - Relatório de limpeza de dados: Descreva as decisões e ações que foram tomadas para resolver os problemas de qualidade de dados relatados durante a Tarefa Verificar Qualidade de Dados. Se os dados forem utilizados no exercício de mineração de dados, o relatório deve abordar problemas de qualidade de dados pendentes e o possível efeito que isso poderia ter sobre os resultados.

Atividades

- Reconsidere como lidar com qualquer tipo de ruído observado
- Corrigir, remover ou ignorar o ruído
- Decida como lidar com valores especiais e seu significado. A área de valores especiais pode dar origem a muitos resultados estranhos e deve ser cuidadosamente examinada. Exemplos de valores especiais podem surgir através da obtenção de resultados de uma pesquisa em que algumas perguntas não foram feitas ou não foram respondidas. Isso pode resultar em um valor de 99 para dados desconhecidos. Por exemplo, 99 para o estado civil ou afiliação política. Valores especiais também podem surgir quando os dados são truncados, por exemplo, 00 para pessoas de 100 anos de idade ou todos os carros com 100.000 km no odômetro.

- Reconsiderar os critérios de seleção de dados (Consulte a Tarefa 2.1) à luz das experiências de limpeza de dados (ou seja, você pode querer incluir / excluir outros conjuntos de dados).

Ideia: Lembre-se de que alguns campos podem ser irrelevantes para os objetivos de mineração de dados e, portanto, o ruído nesses campos não tem significado. No entanto, se o ruído for ignorado por esses motivos, ele deve ser totalmente documentado, pois as circunstâncias podem mudar mais tarde.



Tarefa 3.3 - Construir dados: Esta tarefa inclui operações construtivas de preparação de dados, como a produção de atributos derivados, novos registros ou valores transformados para atributos existentes.

Atividades

- Verifique os mecanismos de construção disponíveis com a lista de ferramentas sugeridas para o projeto
- Decida se é melhor realizar a construção dentro da ferramenta ou externa (ou seja, que é mais eficiente, exato e repetitivo)
- Reconsiderar os critérios de seleção de dados (ver Tarefa 2.1) à luz das experiências de construção de dados (ou seja, você pode desejar incluir / excluir outros conjuntos de dados)

Saída 3.3.1 - Atributos Derivados: Os atributos derivados são atributos novos que são construídos a partir de um ou mais atributos existentes no mesmo registro. Um exemplo pode ser: área = comprimento * largura.

Por que precisamos construir atributos derivados ao longo de uma pesquisa de mineração de dados? Não se deve pensar que apenas os dados de bancos de dados ou outras fontes devem ser usados na construção de um modelo. Os atributos derivados podem ser construídos porque:

- O conhecimento de fundo convence-nos de que algum fato é importante e deve ser representado, embora não tenhamos nenhum atributo atualmente para representá-lo

- O algoritmo de modelagem em uso manipula apenas certos tipos de dados - por exemplo, estamos usando regressão linear e suspeitamos que existem certas não-linearidades que não serão incluídas no modelo
- O resultado da fase de modelagem sugere que certos fatos não estão sendo cobertos

Atividades

Atributos derivados

- Decida se qualquer atributo deve ser normalizado (por exemplo, ao usar um algoritmo de agrupamento com idade e renda, em certas moedas, a renda irá dominar)
- Considere adicionar novas informações sobre a importância relevante dos atributos adicionando novos atributos (por exemplo, pesos de atributos, normalização ponderada)
- Como os atributos faltantes podem ser construídos ou imputados? [Decidir tipo de construção (por exemplo, agregado, média, indução).]
- Adicionar novos atributos aos dados acessados

Ideia: Antes de adicionar Atributos Derivados, tente determinar se e como eles facilitam o processo do modelo ou facilitam o algoritmo de modelagem. Talvez "renda por pessoa" seja um atributo melhor / mais fácil de usar do que "renda por família". Não derivar atributos simplesmente para reduzir o número de atributos de entrada.

Ideia: Outro tipo de atributo derivado é a transformação de atributo único, geralmente realizada para atender às necessidades das ferramentas de modelagem.

Atividades

Transformações de atributo único

- Especifique as etapas de transformação necessárias em termos de instalações de transformação disponíveis (por exemplo, altere um binning de um atributo numérico)
- Execute etapas de transformação

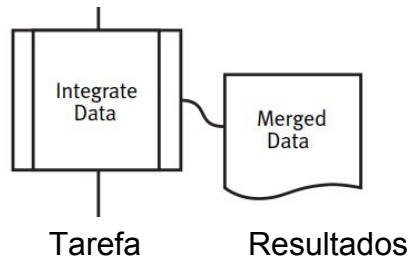
Ideia: Podem ser necessárias transformações para alterar os intervalos em campos simbólicos (por exemplo, idades para faixas etárias) ou campos simbólicos ("definitivamente sim", "sim", "não sei", "não") para valores numéricos. As ferramentas ou algoritmos de modelagem geralmente os exigem.

Saída 3.3.2 - Registros gerados: Os registros gerados são registros completamente novos, que adicionam novos conhecimentos ou representam novos dados que não estão representados de outra forma (por exemplo, tendo segmentado os dados, pode ser útil gerar um registro para representar o membro prototípico de cada segmento para posterior processamento).

Atividades

Verifique as técnicas disponíveis se necessário (por exemplo, mecanismos para construir protótipos para cada segmento de dados segmentados).

Tarefa 4: Integrar dados



Tarefa 3.4 - Integrar dados: Estes são métodos para combinar informações de várias tabelas ou outras fontes de informação para criar novos registros ou valores.

Saída 3.4.1 - Dados mesclados: Mesclar tabelas refere-se a juntar duas ou mais tabelas que têm informações diferentes sobre os mesmos objetos. Nesta fase, também pode ser aconselhável gerar novos registros. Também pode ser recomendado gerar valores agregados.

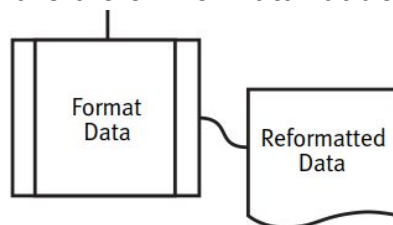
A agregação refere-se a operações em que novos valores são calculados resumindo informações de vários registros e / ou tabelas.

Atividades

- Verifique se as instalações de integração podem integrar as fontes de entrada conforme necessário
- Integrar fontes e resultados da loja
- Reconsiderar os critérios de seleção de dados (ver Tarefa 2.1) à luz das experiências de integração de dados (ou seja, você pode querer incluir / excluir outros conjuntos de dados)

Ideia: Lembre-se de que algum conhecimento pode ser contido em formato não-eletrônico.

Tarefa 3.5 - Formatar dados



Tarefa 3.5 - Formatar dados: As transformações de formatação referem-se principalmente a modificações sintáticas feitas aos dados que não alteram o seu significado, mas podem ser exigidas pela ferramenta de modelagem.

Saída 3.5.1 - Dados reformatados: Algumas ferramentas têm requisitos na ordem dos atributos, como o primeiro campo sendo um identificador exclusivo para cada registro ou o último campo sendo o campo de resultado que o modelo deve prever.

Atividades

Reorganizando atributos

■ Algumas ferramentas têm requisitos na ordem dos atributos, como o primeiro campo sendo um identificador exclusivo para cada registro ou o último campo sendo o campo de resultado que o modelo deve prever.

Reordenando registros

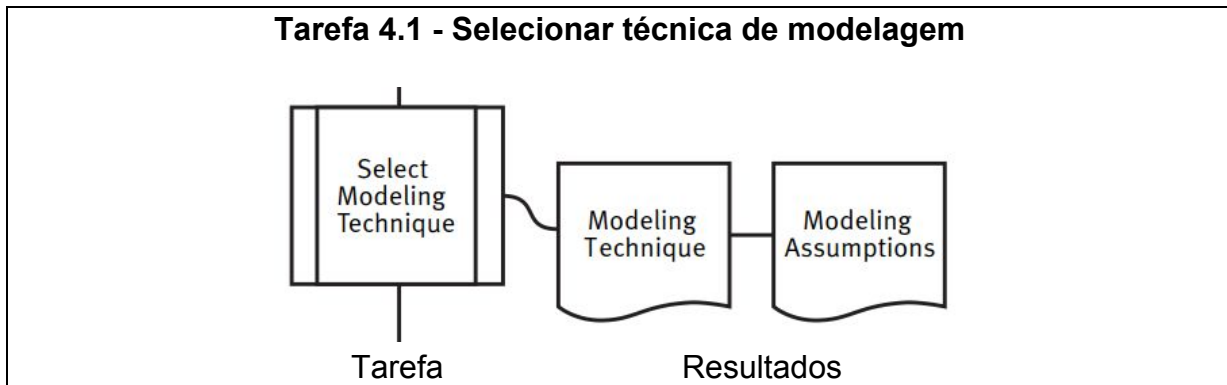
■ Pode ser importante alterar a ordem dos registros no conjunto de dados. Talvez a ferramenta de modelagem exija que os registros sejam classificados de acordo com o valor do atributo de resultado.

Reformatado dentro do valor

■ São mudanças puramente sintáticas feitas para satisfazer os requisitos da ferramenta de modelagem específica

■ Reconsiderar os critérios de seleção de dados (ver Tarefa 2.1) à luz das experiências de limpeza de dados (ou seja, você pode querer incluir / excluir outros conjuntos de dados)

4 - Modelagem



Tarefa 4.1 - Selecionar técnica de modelagem: Como o primeiro passo na modelagem, selecione a técnica de modelagem inicial real. Se forem aplicadas múltiplas técnicas, execute esta tarefa separadamente para cada técnica. Lembre-se de que nem todas as ferramentas e técnicas são aplicáveis a todas e cada uma das tarefas. Para certos problemas, apenas algumas técnicas são apropriadas. "Requisitos políticos" e outras restrições limitam ainda mais as opções disponíveis para o engenheiro de mineração de dados. Pode ser que apenas uma ferramenta ou técnica esteja disponível para resolver o problema em questão - e que a ferramenta pode não ser absolutamente a melhor, do ponto de vista técnico.

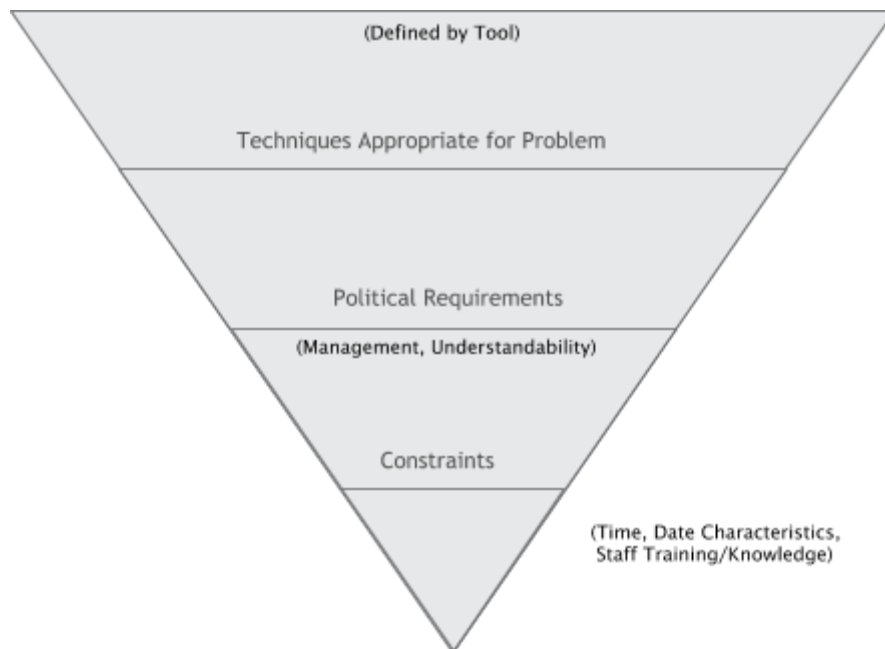


Figura – Universo de técnicas

Saída 4.1.1 - Técnica de modelagem: Registre a técnica de modelagem atual que é usada.

Atividades

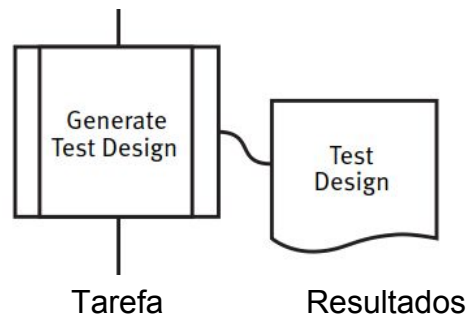
- Decida a técnica apropriada para o exercício, tendo em conta a ferramenta selecionada.

Saída 4.1.2 - Suposições de modelagem: Muitas técnicas de modelagem fazem suposições específicas sobre os dados.

Atividades

- Defina quaisquer pressupostos internos feitos pela técnica sobre os dados (por exemplo, qualidade, formato, distribuição)
- Compare estes pressupostos com aqueles no Relatório de Descrição de Dados
- Certifique-se de que estes pressupostos se mantenham e voltem para a fase de preparação de dados, se necessário

Tarefa 4.2 - Gerar design de teste



Tarefa 4.2: Gerar design de teste - Antes de construir um modelo, é necessário definir um procedimento para testar a qualidade e a validade do modelo. Por exemplo, em tarefas supervisionadas de mineração de dados, como a classificação, é comum usar taxas de erro como medidas de qualidade para modelos de mineração de dados. Portanto, o design de teste especifica que o conjunto de dados deve ser separado em conjuntos de treinamento e teste. O modelo é construído sobre o conjunto de treinamento e sua qualidade estimada no conjunto de teste.

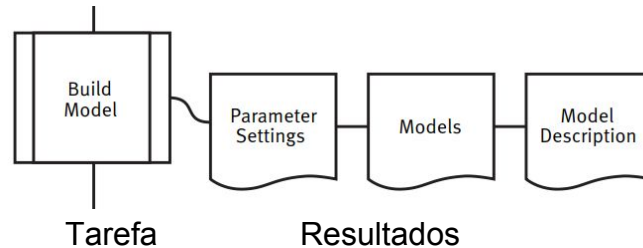
Saída 4.2.1 - Design de teste: Descreva o plano pretendido para treinar, testar e avaliar os modelos. Um componente primário do plano é decidir como dividir o conjunto de dados disponível em dados de treinamento, dados de teste e conjuntos de teste de validação.

Atividades

- Verifique os projetos de teste existentes para cada meta de mineração de dados separadamente

- Decida as etapas necessárias (número de iterações, número de dobras, etc.)
- Preparar os dados necessários para o teste

Tarefa 4.3 - Construir o modelo



Tarefa 4.3 - Construir o modelo: Execute a ferramenta de modelagem no conjunto de dados preparado para criar um ou mais modelos.

Saída 4.3.1 - Configurações dos parâmetros: Com qualquer ferramenta de modelagem, muitas vezes há um grande número de parâmetros que podem ser ajustados. Liste os parâmetros e os valores escolhidos, juntamente com o raciocínio para a escolha.

Atividades

- Definir parâmetros iniciais
- Documentar razões para escolher esses valores

Saída 4.3.2 – Modelos: Execute a ferramenta de modelagem no conjunto de dados preparado para criar um ou mais modelos.

Atividades

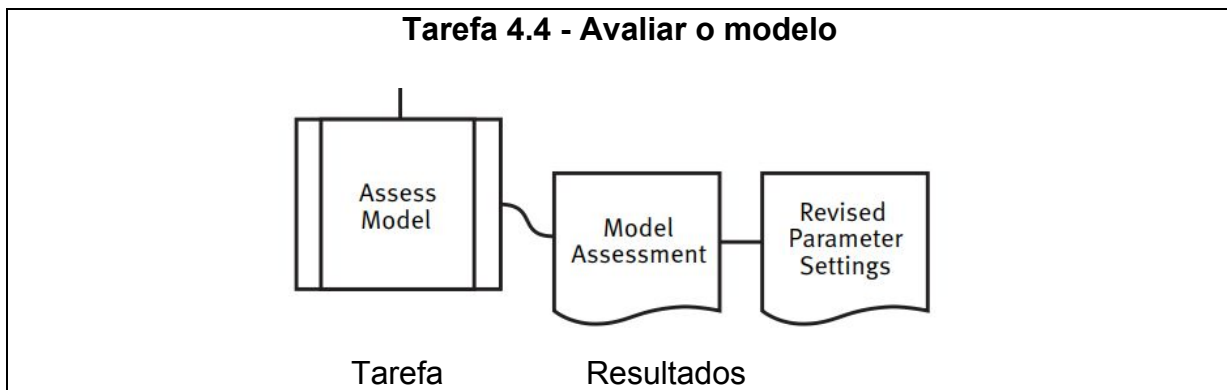
- Execute a técnica selecionada no conjunto de dados de entrada para produzir o modelo
- Resultados de mineração de dados pós-processamento (por exemplo, editar regras, exibir árvores)

Saída 4.3.3 - Descrição do modelo: Descreva o modelo resultante e avalie sua precisão esperada, robustez e possíveis falhas. Relatório sobre a interpretação dos modelos e quaisquer dificuldades encontradas.

Atividades

- Descreva quaisquer características do modelo atual que possam ser úteis para o futuro
- Gravar configurações de parâmetros usadas para produzir o modelo

- Dê uma descrição detalhada do modelo e de quaisquer características especiais
- Para modelos baseados em regras, liste as regras produzidas, além de qualquer avaliação da exatidão e cobertura padrão por cada regra ou modelo.
- Para modelos opacos, liste qualquer informação técnica sobre o modelo (como a topologia da rede neural) e quaisquer descrições comportamentais produzidas pelo processo de modelagem (como precisão ou sensibilidade)
- Descreva o comportamento e a interpretação do modelo.
- Conclusões do estado em relação aos padrões nos dados (se houver); às vezes, o modelo revela fatos importantes sobre os dados sem um processo de avaliação separado (por exemplo, que a saída ou conclusão é duplicada em uma das entradas)



Tarefa 4.4 - Avaliar o modelo: O modelo agora deve ser avaliado para garantir que ele atenda aos critérios de sucesso da mineração de dados e passa os critérios de teste desejados. Esta é uma avaliação puramente técnica baseada no resultado das tarefas de modelagem.

Saída 4.4.1 - Avaliação do modelo: Resumir resultados dessa tarefa, listar qualidades de modelos gerados (por exemplo, em termos de precisão) e classificar sua qualidade em relação um ao outro.

Atividades

- Avaliar os resultados em relação aos critérios de avaliação
- Resultado do teste de acordo com uma estratégia de teste (por exemplo: Treino e teste, validação cruzada, *bootstrapping*, etc.)
- Comparar resultados de avaliação e interpretação
- Criar classificação de resultados em relação ao sucesso e critérios de avaliação
- Selecione os melhores modelos
- Interpretar resultados em termos comerciais (na medida do possível nesta fase)
- Obter comentários sobre modelos por especialistas em domínio ou dados
- Verifique a plausibilidade do modelo
- Verificar efeito no objetivo de mineração de dados
- Verifique o modelo contra base de conhecimentos dada para ver se a informação descoberta é nova e útil

- Verifique a confiabilidade do resultado
- Analisar o potencial de implantação de cada resultado
- Se houver uma descrição verbal do modelo gerado (por exemplo, através de regras), avalie as regras: são lógicos, são viáveis, existem muitos ou muito poucos, eles ofendem o senso comum?
- Avaliar os resultados
- Obtenha informações sobre por que uma certa técnica de modelagem e determinadas configurações de parâmetros levam a resultados bons / ruins

Ideia: As "Tabelas de Elevação" e "Tabelas de Ganho" podem ser construídas para determinar o quão bem o modelo está prevendo.

Saída 4.4.2 - Configurações de parâmetros revisados: De acordo com a avaliação do modelo, revise as configurações dos parâmetros e ajuste-os para a próxima execução na tarefa de construir o modelo (Tarefa 4.3) até encontrar o melhor modelo.

Atividades

- Ajustar parâmetros para produzir melhores modelos.

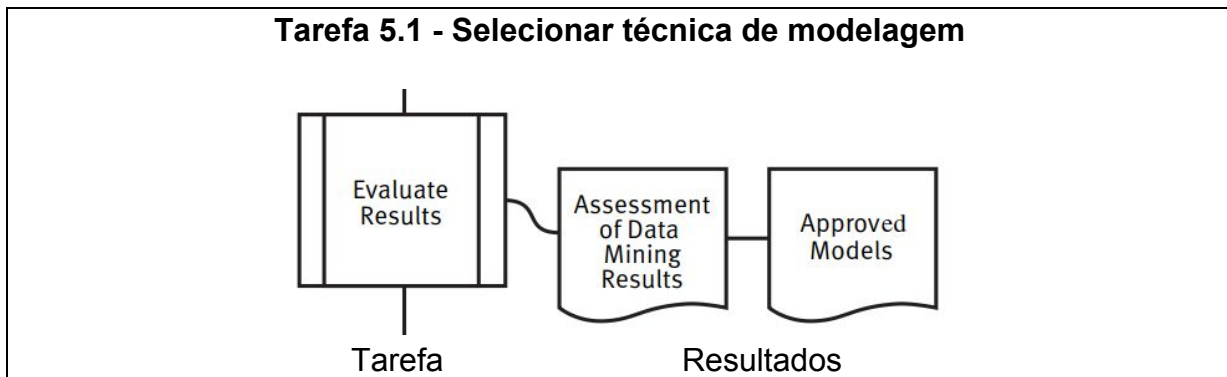
5 - Avaliação

As etapas de avaliação anteriores trataram de fatores como a precisão e a generalidade do modelo. Esta etapa avalia o grau em que o modelo atende aos objetivos de negócios e procura determinar se há algum motivo de negócios porque esse modelo é deficiente. Ele compara os resultados com os critérios de avaliação definidos no início do projeto.

Uma boa maneira de definir os resultados totais de um projeto de mineração de dados é usar a equação:

RESULTADOS = MODELOS + ENCONTROS

Nesta equação, estamos definindo que a produção total do projeto de mineração de dados não é apenas os modelos (embora sejam, é claro, importantes), mas também os achados, que definimos como qualquer coisa (além do modelo) que é importante no cumprimento dos objetivos do negócio ou importante para levar a novas questões, linhas de abordagem ou efeitos colaterais (por exemplo, qualidade de dados, problemas descobertos pelo exercício de mineração de dados). Nota: Embora o modelo esteja diretamente conectado às questões comerciais, os resultados não precisam estar relacionados a quaisquer questões ou objetivos, desde que sejam importantes para o iniciador do projeto.



Tarefa 5.1 - Avaliar os resultados: Esta etapa avalia o grau em que o modelo atende aos objetivos de negócios e procura determinar se há algum motivo de negócios porque esse modelo é deficiente. Outra opção é testar o (s) modelo (s) em aplicativos de teste no aplicativo real, se o tempo e as restrições orçamentárias o permitirem.

Além disso, a avaliação também avalia outros resultados de mineração de dados gerados. Os resultados da mineração de dados abrangem modelos que estão relacionados aos objetivos comerciais originais e a todas as outras descobertas. Alguns estão relacionados aos objetivos comerciais originais, enquanto outros podem desvendar desafios adicionais, informações ou sugestões para orientações futuras.

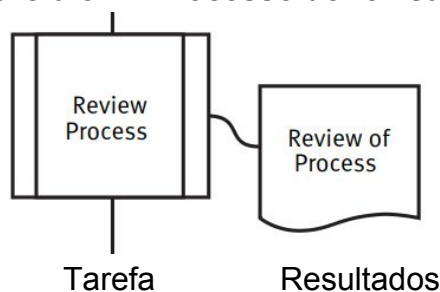
Saída 5.1.1 - Avaliação de resultados de mineração de dados: avaliar os resultados da mineração de dados em relação aos critérios de sucesso comercial. Resumir os resultados da avaliação em termos de critérios de sucesso comercial, incluindo uma declaração final relacionada ao fato de o projeto já atingir os objetivos comerciais iniciais.

Atividades

- Compreender os resultados da mineração de dados
- Interprete os resultados em termos da aplicação
- Verifique o efeito sobre o objetivo de mineração de dados
- Verifique o resultado da mineração de dados em relação à base de conhecimento fornecida para ver se a informação descoberta é nova e útil
- Avaliar e avaliar os resultados em relação aos critérios de sucesso comercial (ou seja, o projeto alcançou os Objetivos de Negócios originais)
- Comparar resultados de avaliação e interpretação
- Classificar os resultados em relação aos critérios de sucesso comercial
- Verifique o efeito do resultado no objetivo inicial da aplicação
- Determine se há novos objetivos comerciais a serem abordados mais tarde no projeto ou em novos projetos
- Recomendações de estado para futuros projetos de mineração de dados

Saída 5.1.2 - Modelos aprovados: Depois de acessar os modelos em relação aos critérios de sucesso do negócio, selecione e aprove os modelos gerados que atendem aos critérios selecionados.

Tarefa 5.2 - Processo de revisão



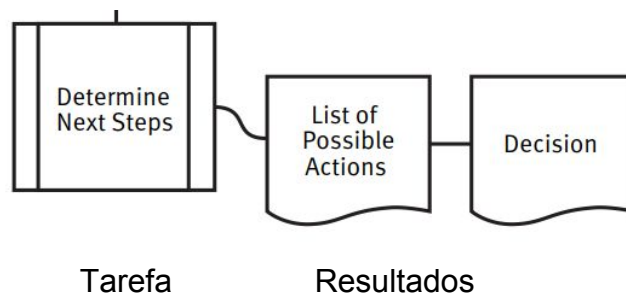
Tarefa 5.2 - Processo de revisão: Neste ponto, o modelo resultante parece ser satisfatório e parece satisfazer as necessidades do negócio. Agora é apropriado fazer uma revisão mais aprofundada do engajamento da mineração de dados, a fim de determinar se há algum fator ou tarefa importante que de alguma forma tenha sido negligenciado. Nesta fase do exercício de mineração de dados, a Revisão do Processo assume a forma de uma Revisão de Garantia da Qualidade.

Saída 5.2.1 - Revisão do processo: Resumir a revisão do processo e listar as atividades que foram perdidas e / ou devem ser repetidas.

Atividades

- Fornecer uma visão geral do processo de mineração de dados usado
- Analisar o processo de mineração de dados. Para cada etapa do processo, pergunte:
 - Era necessário?
 - Foi executado de forma otimizada?
 - De que maneiras poderia ser melhorado?
- Identificar falhas
- Identificar passos enganosos
- Identificar possíveis ações alternativas e / ou caminhos inesperados no processo
- Revisar resultados de mineração de dados em relação aos critérios de sucesso comercial

Tarefa 5.3 - Determinar os próximos passos



Tarefa 5.3 - Determinar os próximos passos: Com base nos resultados da avaliação e na revisão do processo, a equipe do projeto decide como proceder. As decisões a serem feitas incluem se pretende finalizar este projeto e passar à implantação, iniciar outras iterações ou configurar novos projetos de mineração de dados.

Saída 5.3.1 - Lista de ações possíveis: Liste possíveis ações adicionais, juntamente com os motivos para e contra cada opção

Atividades

- Analisar o potencial de implantação de cada resultado
- Estimar o potencial de melhoria do processo atual
- Verifique os recursos restantes para determinar se eles permitem iterações de processo adicionais (ou se recursos adicionais podem ser disponibilizados)
- Recomendar continuações alternativas
- Refinar o plano de processo

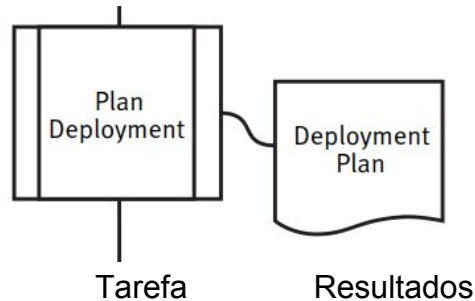
Saída 5.3.2 – Decisão: Descreva as decisões tomadas, juntamente com o raciocínio para elas.

Atividades

- Posicionar as possíveis ações
- Selecione uma das possíveis ações
- Documentar razões para a escolha

6 - Implantação

Tarefa 6.1: Planejar implantação



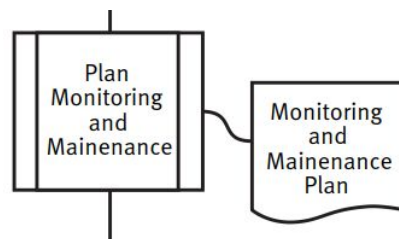
Tarefa 6.1 - Planejar implantação: Esta tarefa começa com os resultados da avaliação e conclui com uma estratégia para a implantação do (s) resultado (s) de mineração de dados no negócio.

Saída 6.1.1 - Plano de implantação: Resumir a estratégia de implantação, incluindo as etapas necessárias e como executá-las.

Atividades

- Resumir resultados desdobráveis
- Desenvolver e avaliar planos alternativos para implantação
- Decida por cada conhecimento de informação ou informação distinto
- Determine como o conhecimento ou a informação serão propagados para usuários
- Decida como o uso do resultado será monitorado e seus benefícios serão medidos (quando aplicável)
- Decida por cada modelo de modelo implantável ou resultado de software
- Estabelecer como o modelo ou o resultado do software serão implantados nos sistemas da organização
- Determine como seu uso será monitorado e seus benefícios serão medidos (quando aplicável)
- Identificar possíveis problemas durante a implantação (armadilhas a serem evitadas)

Tarefa 6.2: Planejar o acompanhamento e manutenção



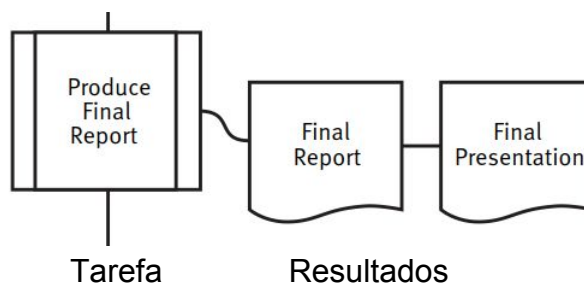
Tarefa 6.2 - Planejar o acompanhamento e manutenção: O monitoramento e a manutenção são questões importantes se os resultados da mineração de dados se tornarem parte do negócio do dia-a-dia e seu ambiente. Uma preparação cuidadosa de uma estratégia de manutenção ajuda a evitar períodos desnecessariamente longos de uso incorreto de resultados de mineração de dados. Para monitorar a implantação do (s) resultado (s) de mineração de dados, o projeto precisa de um plano detalhado para monitoramento e manutenção. Este plano leva em consideração o tipo específico de implantação.

Saída 6.2.1 - Plano de monitoramento e manutenção: Resumir a estratégia de monitoramento e manutenção, incluindo as etapas necessárias e como executá-las.

Atividades

- Verifique os aspectos dinâmicos (ou seja, o que as coisas podem mudar no ambiente?)
- Decida como a precisão será monitorada
- Determine quando o resultado da mineração de dados ou o modelo não devem mais ser usados. Identificar critérios (validade, limiar de precisão, novos dados, alteração no domínio do aplicativo, etc.) e o que deve acontecer se o modelo ou o resultado não puderem ser usados. (modelo de atualização, configurar um novo projeto de mineração de dados, etc.).
- Os objetivos comerciais do uso do modelo mudarão ao longo do tempo? Totalmente documentar o problema inicial que o modelo estava tentando resolver.
- Desenvolver plano de monitoramento e manutenção.

Tarefa 6.3: Produzir relatório final



Tarefa 6.3 - Produzir relatório final: No final do projeto, a equipe do projeto escreve um relatório final. Dependendo do plano de implantação, este relatório pode ser apenas um resumo do projeto e sua experiência, ou uma apresentação final do (s) resultado (s) de mineração de dados.

Saída 6.3.1 - Relatório final: No final do projeto, haverá pelo menos um relatório final no qual todos os tópicos são reunidos. Além de identificar os resultados obtidos, o relatório também deve descrever o processo, mostrar quais os custos que foram incorridos, definir quaisquer desvios do plano original, descrever planos de implementação e fazer recomendações para o trabalho futuro. O conteúdo detalhado real do relatório depende muito do público-alvo.

Atividades

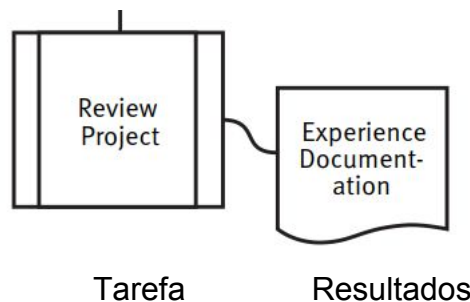
- Identificar quais relatórios são necessários (apresentação de slides, resumo de gerenciamento, detalhamento, explicação de modelos, etc.)
- Analise o quão bem foram alcançados os objetivos iniciais da mineração de dados
- Identificar grupos-alvo para o relatório
- Estrutura e conteúdo do relatório (s)
- Selecionar resultados a serem incluídos nos relatórios
- Escreva um relatório

Saída 6.3.2 - Apresentação final: Além de um relatório final, pode ser necessário fazer uma apresentação final para resumir o projeto - talvez para o patrocinador de gerenciamento, por exemplo. A apresentação geralmente contém um subconjunto da informação contida no relatório final, estruturada de forma diferente.

Atividades

- Decidir sobre o grupo-alvo para a apresentação final e determinar se já receberam o relatório final
- Selecione quais itens do relatório final devem ser incluídos na apresentação final

Tarefa 6.4: Revisar o projeto



Tarefa 6.4 - Revisar o projeto: Avalie o que correu corretamente e o que deu errado, o que foi feito bem e o que precisa ser melhorado.

Saída 6.4.1 - Documentação de experiência: Resumir a experiência importante adquirida durante o projeto. Por exemplo, armadilhas, abordagens enganosas ou dicas para selecionar as técnicas de mineração de dados mais adequadas em

situações semelhantes podem fazer parte desta documentação. Em projetos ideais, a documentação da experiência também cobre todos os relatórios que tenham sido escritos por membros individuais do projeto durante o projeto.

Atividades

- Entreviste todas as pessoas importantes envolvidas no projeto e pergunte-lhes sobre sua experiência durante o projeto
- Se os usuários finais da empresa trabalham com o (s) resultado (s) da mineração de dados, entrevistá-los: eles estão satisfeitos? O que poderia ter sido feito melhor? Eles precisam de suporte adicional?
- Resumir comentários e escrever a documentação da experiência
- Analisar o processo (coisas que funcionaram bem, erros cometidos, lições aprendidas, etc.)
- Documentar o processo específico de mineração de dados (Como os resultados e a experiência de aplicar o modelo podem ser devolvidos no processo?)
- Generalize a partir dos detalhes para tornar a experiência útil para projetos futuros