

Desenvolvimento de um Sistema de Apoio a Decisão para priorização de Pedidos de Desembolso no Estado de Pernambuco

Development of a decision system for prioritization of Disbursement Requests in the State of Pernambuco

Itallo Henrique de Santana Santos¹  orcid.org/0000-0003-1124-4150
Alexandre Magno Andrade Maciel¹  orcid.org/0000-0003-4348-9291

¹ Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

E-mail do autor principal: ihss@ecom.poli.br

Resumo

A Secretaria da Controladoria Geral do Estado (SCGE) analisa mensalmente despesas geradas pelos diferentes órgãos do Estado de Pernambuco com o objetivo de garantir os pagamentos daquelas que são mais sensíveis. Mais de 1 bilhão de reais em despesas ficaram pendentes no exercício de 2016, demonstrando importância de priorização dos pagamentos. Nesse contexto o artigo apresenta o processo de desenvolvimento de um Sistema de Apoio a Decisão (SAD) para a Secretaria da Controladoria Geral do Estado de Pernambuco. O sistema proposto tem a capacidade de classificar as despesas públicas por meio de árvore de decisão auxiliando o trabalho de análise dos gestores responsáveis na priorização de pagamentos. No artigo é detalhado a caracterização do problema, a fundamentação teórica usada no trabalho, a aplicação da metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) e o sistema. A utilização de árvore de decisão na classificação das despesas teve como resultado uma acurácia de 99%, mostrando que o uso desse tipo de modelo atendeu satisfatoriamente na solução do problema encontrado.

Palavras-Chave: Sistemas de Apoio à Decisão; Sistemas de informações; Árvore de Decisão; Mineração de Dados;

Abstract

The Secretaria da Controladoria Geral do Estado (SCGE) analyzes the monthly expenses generated by the different organs of the State of Pernambuco with the objective of pay sensitive expenses. Over one billion reais in expenses were outstanding in 2016, showing the importance of prioritizing payments. In this context the article presents the process of developing a Decision Support System (DSS) for the SCGE. The proposed system has the ability to classify public expenditures through a decision tree, assisting the analysis work of the responsible managers in the prioritization of payments. In the article is detailed the characterization of the problem, the theoretical basis used in the work, the application of CRISP-DM (*Cross Industry Standard Process for Data Mining*) and the system. The utilization decision tree in the classification of expenditures resulted in an accuracy of 99%, showing that the use of this type of model satisfactorily attended to the problem.

Key-words: Decision Support System; Information Systems; Decision trees; Data Mining.

1 Introdução

Nos últimos anos houve um grande crescimento do uso de tecnologias na área de gestão de organizações públicas ou privadas, elas hoje buscam cada vez mais garantir eficácia e eficiência nas suas atividades por meio de sistemas de informação. Laudon [1] define sistema de informação (SI) como um conjunto de componentes inter-relacionados que trabalham juntos na coleta, recuperação, processamento, armazenamento e distribuição de informações com a finalidade de facilitar o planejamento, o controle, a coordenação, a análise e o processo decisório em organizações. Os SI transformam as informações tornando mais fácil o acesso e entendimento pelos seus usuários.

Nesse contexto, Góes [2] comenta:

Governadores e prefeitos fazem inúmeras tentativas de 'revolucionar' na gestão dos seus Estados e Municípios, respectivamente, na ânsia de fazer de seu mandato um exemplo de gestão eficiente e com o maior atendimento possível ao interesse público. A 'nova' Administração Pública está em toda parte. Ela preocupa-se, essencialmente, com a gestão dos recursos financeiros e com o marketing público, cujos objetivos são a qualidade e a racionalidade, entendidas como sinônimos de economia e meios de redução de custos [2].

Os pagamentos das Despesas Públicas realizadas pelo Estado de Pernambuco seguem, normalmente, uma organização baseada em Pedidos de Desembolsos (PDs) inseridos no sistema eletrônico de gestão financeira do Estado. Pedidos de Desembolsos, conforme o nome intuitivo, são pedidos ou solicitações de pagamentos das despesas realizadas pelas unidades gestoras do Estado de Pernambuco. Diante do cenário de crise econômica, sobretudo a crise na arrecadação dos impostos estaduais, nem todos os PDs foram pagas dentro do prazo previsto. Tal situação demanda uma preocupação maior no controle dos gastos por parte do órgão de controle interno, no caso a Controladoria Geral do Estado de Pernambuco.

Em 2016, o governo do Estado de Pernambuco movimentou mais de 28 bilhões de reais com PDs objetivando o pagamento das próprias despesas, conforme informações extraídas da base de dados dos PDs fornecida pela Secretaria da Fazenda.

Apesar do esforço no cumprimento das obrigações financeiras, ainda restou mais de 1 bilhão de reais de PDs com competências referentes ao exercício de 2016, e que, antes, passaram a integrar como novos PDs do exercício de 2017. Diante de tal situação, houve a necessidade de priorizar o controle do pagamento de certos temas de PDs mais urgentes, como por exemplo, PDs de medicamentos, alimentação, programas sociais do estado e entre outros. E ainda priorizar o fomento do controle interno de PDs em unidades gestoras consideradas sensíveis, como por exemplo, o FEAS (Fundo Estadual de Assistência Social).

A Secretaria da Fazenda de Pernambuco (SEFAZ) detém o controle da base dos Pedidos de Desembolsos. Portanto, a SEFAZ busca melhorar o processo de priorização de PDs, inclusive com categorização e a criação de alternativas de Business Intelligence. Em reconhecimento e apoio a SEFAZ, a Secretaria da Controladoria Geral do Estado de Pernambuco (SCGE) inicia um estudo de análise das PDs, tendo em vista as atribuições de Controle Interno deste órgão segundo o artigo 70 da Constituição Federal e Lei Complementar nº 119 de 26/06/2008 do Estado de Pernambuco.

Atualmente, a SCGE utiliza a ferramenta Qlikview para gerar dados referentes aos PDs mensais no formato .xls, fornecida pela SEFAZ, para extração das informações onde é considerado os aspectos de unidades gestoras, temas dos gastos e valores considerados sensíveis ou fora da normalidade, de forma que no final sejam gerados alertas, com base nos aspectos citados, para que sejam encaminhados às autoridades competentes do executivo estadual assim tomando as medidas corretivas necessárias. Entretanto, o processo para geração dos alertas é considerado um problema por ser considerado ainda muito manual.

Nesse contexto, o artigo tem como objetivo desenvolver um sistema de apoio a decisão, descrevendo conceitos, técnicas e ferramentas usadas em seu desenvolvimento. O sistema será capaz de classificar Pedidos de Desembolso utilizando modelos de árvore de decisão.

2 Fundamentação Teórica

2.1 Sistemas de Apoio à Decisão

Os sistemas de apoio a decisão SAD são uma classificação dos SI. Esses sistemas tem como foco o suporte às decisões específicas de um processo, tem como objetivo melhorar a capacidade de compreensão da informação por meio de modelos e simulações permitindo o gestor ampliar conhecimento e esclarecimento de um problema. Vale salientar que os SADs estão limitados ao suporte e não na automatização das decisões.

Na literatura encontramos várias definições para os SADs. Sprague e Watson [3] definem SAD como sistemas computacionais que ajudam os responsáveis pela tomada de decisões a enfrentar problemas não-estruturais através da interação direta com modelos de dados e análises. Segundo Bidgoli [4], SAD é um sistema de informação baseado em computador, que consiste de *hardware* e *software* e elemento humano, para assistir qualquer decisão em qualquer nível, e que enfatiza tarefas não-estruturadas ou semi-estruturadas. Para Courtney (2001), os SAD são sistemas de gerenciamento de decisões interativos, baseados em computador, que ajudam os decisores a utilizar dados e modelos para resolver problemas não-estruturados.

Barbosa [5] complementa que conceito de SAD ainda não está completamente livre de divergências entre os pesquisadores, usuários e desenvolvedores, não havendo, portanto, nenhuma definição exata. Entretanto, podem ser destacados alguns aspectos em comum, tais como: conseguem juntar o pensamento humano e a informação automatizada; abrangem todas as fases do processo de decisão; objetivam ajudar e não substituir o decisor; utilizam modelos para análise de situações de tomada de decisão, possibilitar experiências com diferentes estratégias sob diferentes configurações facilitadas pela capacidade de modelagem permitem, e apresentam uma interface amigável [5].

2.2 Tipos de Sistema de Apoio à Decisão

Alter [6], com referência de dezenas de sistemas de informação, propôs um conjunto de operações genéricas que podem ser executadas por SAD, as quais se classificam em sistemas orientados a dados e sistemas orientados a modelos.

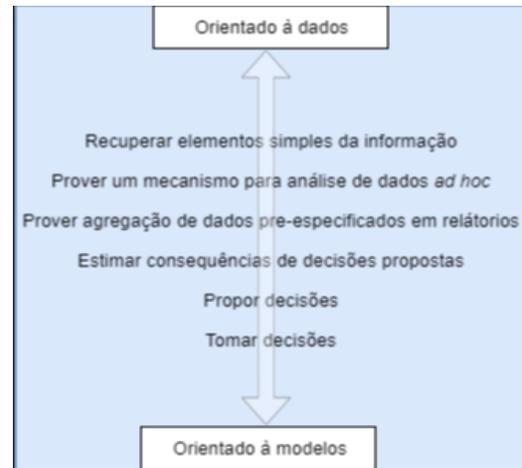


Figura 1 - Conjunto de operações genéricas
Fonte: adaptado de [6].

Além disso, Alter [6] com base nas operações genéricas classificou sete tipos diferentes de SAD. Barbosa [5] resume essas operações como:

- Sistemas de desenho de arquivos: representam basicamente uma versão automatizada dos sistemas de arquivamento manual, proporcionam maior segurança e rapidez na recuperação de informações.
- Sistemas de análise de dados: facilitam a análise a partir de arquivos com dados atuais ou históricos, geram indicadores.
- Sistemas de análise de informações: provêm acesso a uma série de dados orientados a decisão e pequenos modelos para prover informação gerencial, possibilitando a análise através do uso de dados internos.
- Modelos de contas: calculam a consequência de ações planejadas sobre a base de definições de contas e estimam variações das entradas nas fórmulas das contas.

- Modelos de representação: incluem modelos de simulação, estimam consequência de ações sobre a base de modelos, tais como probabilidades de ocorrências.
- Modelos de otimização: oferecem linhas de ação para uma solução ótima, consideram restrições; usados para decisões repetitivas que podem ser descritas matematicamente.
- Modelos de sugestão: consideram uma sugestão específica para uma decisão estruturada e repetitiva, substituindo procedimento menos eficientes.

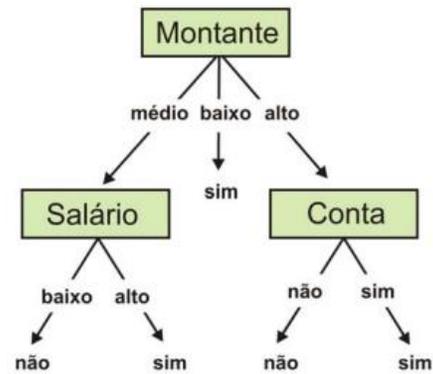


Figura 2 - Representação de uma árvore de decisão. Fonte: adaptado de [9].

2.3 Árvores de Decisão

Árvores de Decisão em Mineração de Dados (Data Mining) faz parte de métodos que envolvem classificação dos dados. Podem ser usadas como tecnologia de indução de regras [7]. Crepaldi et al. [7] complementa que:

A vantagem principal das Árvores de Decisão é a tomada de decisões levando em consideração os atributos mais relevantes, além de compreensíveis para a maioria das pessoas. Ao escolher e apresentar os atributos em ordem de importância, as Árvores de Decisão permitem aos usuários conhecer quais fatores mais influenciam os seus trabalhos.

Segundo Garcia [8], as Árvores de Decisão consistem de: nodos (nós), que representam os atributos, e de arcos (ramos), provenientes desses nodos e que recebem os valores possíveis para esses atributos (cada ramo descendente corresponde a um possível valor desse atributo). Nas árvores existem nodos folha (folha da árvore), que representam as diferentes classes de um conjunto de treinamento, ou seja, cada folha está associada a uma classe [8].

2.4 Representação de uma Árvore de Decisão

A Figura 2 apresenta um exemplo de árvore de decisão.

Na Figura 2, são analisados objetos que descrevem o recebimento de um empréstimo por uma pessoa. É considerado o valor do empréstimo (Montante), se ele é médio, baixo ou alto. Alguns objetos são exemplos positivos de uma classe sim, ou seja, os requisitos exigidos a uma pessoa, por um banco, são satisfatórios à concessão de um empréstimo, e outros são negativos, onde os requisitos exigidos não são satisfatórios à concessão de um empréstimo [9].

Com a árvore de decisão em mãos é possível gerar regras. Cada percurso na árvore (da raiz à folha) corresponde a uma regra de classificação [10]. Dois exemplos de regras são definidos a seguir baseado na figura (árvore).

*Se montante = médio e salário = baixo
então classe = não*
*Se montante = médio e salário = alto
então classe = sim*

2.5 Técnicas de Construção de Árvores de Decisão

A construção de uma Árvore de Decisão é dada pelo particionamento dos dados, onde a verificação de um atributo divide o conjunto de dados, criando um ramo para cada valor deste atributo. Para cada ramo é criado um nodo e um novo atributo é analisado para fazer o particionamento do subconjunto, esse processo é repetido sucessivamente, e tem como finalidade, separar os dados em classes onde diferentes classes tendem a fazer parte de um subconjunto diferente dos dados [10].

A Figura 3 apresenta algoritmo genérico para construir Árvores de Decisão, em que S

representa o conjunto de exemplos aplicado a árvores de decisão, primeiramente S contém todo conjunto de dados treino.

Se todos os exemplos no atual conjunto de exemplos S satisfazem um critério de parada;

Então
Cria um nodo folha com algum nome da classe e para;

Senão

Seleciona um atributo A para ser utilizado como um atributo de particionamento e cria um nodo com o nome do atributo de particionamento.

Escolhe um teste sobre os valores de A , com resultados mutuamente exclusivos e coletivamente exaustivos R_1, \dots, R_k , e cria um ramo, a partir do nodo recentemente criado, para cada teste;

Particiona S nos subconjuntos S_1, \dots, S_k , tal que cada S_i , $i = 1, \dots, k$, contenha todos os exemplos em S com resultado R_i , do teste escolhido;

Aplica esse algoritmo recursivamente para cada subconjunto S_i , $i=1, \dots, k$;

Fim_senão

fim_se

Figura 3 - Algoritmo genérico para gerar árvores de decisão. Fonte: adaptado de [10].

2.6 Técnica de Seleção de Atributos

A seleção de atributos na geração de uma árvore de decisão consiste em determinar qual o atributo melhor se enquadra no nodo em análise, observando o aspecto de como esse atributo define o conjunto de dados. Logo é possível determinar a melhor partição a ser realizada [10]. A utilização do índice de Gini é amplamente utilizado nesse tipo de problema.

2.7 Índice de Gini

O índice de Gini, criado pelo matemático italiano Conrado Gini, mede o grau de heterogeneidade dos dados.

O índice de Gini em um determinado nó é:

$$Indice\ Gini = 1 - \sum_{i=1}^c p_i^2$$

Onde p_i é a frequência relativa de cada classe em cada nó e c é o número de classes. Se o índice de Gini for igual a zero, significa que o nó é puro, ao contrário, quando se aproxima de um o nó é considerado impuro, logo é possível particionar os dados de uma melhor forma com essa informação.

3 Metodologia

A metodologia utilizada para o desenvolvimento desse projeto é o CRISP-DM (Cross Industry Standard Process for Data Mining). O CRISP-DM tem com objetivo a padronização de conceitos e técnicas na busca de informações para a tomada de decisões (SILVA, 2002). O CRISP-DM é dividido em seis fases de maneira cíclica como mostra a Figura 4.

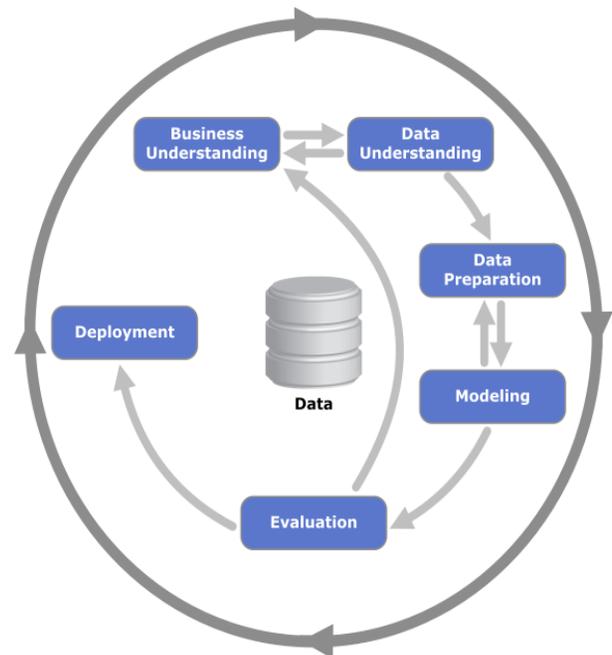


Figura 4 - Fluxo do CRISP-DM

3.1 Entendimento do Negócio

Nesta fase adquiriu-se entendimento de todo processo de análise dos Pedidos de Desembolso feito pela Secretaria da Controladoria Geral do Estado de Pernambuco. A Figura 5 o fluxo do processo de análise dos Pedidos.



Figura 5 - Processo de classificação de Pedidos de Desembolso feito pela SCGE.

A Secretaria da Controladoria da Geral do Estado mensalmente extrai os dados de Pedidos de Desembolso através da plataforma de BI Qlik View, onde um arquivo .xls é gerado, cada linha desse arquivo representa um Pedido de Desembolso que precisará ser analisado e classificado.

Na etapa de transformações alguns campos do Pedido de Desembolso são verificados e/ou processados com o objetivo de transformar os dados de forma que caiba a análise de acordo com as regras de priorização definidas pela gestão. Através da ferramenta Excel é feito os processos a seguir.

Verificação da Unidade Gestora: Para cada Pedido é feita a verificação se a Unidade Gestora daquele Pedido de Desembolso é prioridade.

Verificação do Tema: Para cada Pedido é feita a verificação se o Tema daquele Pedido de Desembolso é prioridade.

Verificação do Valor: Para cada Pedido é feita a verificação se o valor em R\$ ultrapassa uma quantia específica. Pedido de Desembolso com valor muito alto são priorizados.

Extração da Data de Competência: Cada Pedido de Desembolso contém um campo textual que o descreve, nessa descrição pode estar contida a data que a despesa ocorreu. Essa data é importante, já que é possível saber quantos dias já se passaram desde que ocorreu a despesa. Se não for possível extrair essa data, a data de competência deverá ser a data que o Pedido de Desembolso foi adicionado ao sistema.

Inclusão da data Prazo Limite: Para cada Pedido de Desembolso é adicionada uma data que defina o prazo limite que o Pedido de Desembolso poderá atingir antes que ele precise ser priorizado.

Inclusão da data Geração: A data de Geração é a data que o Pedido de Desembolso está sendo analisado, a partir dessa data é possível saber se o Pedido de Desembolso passou do Prazo Limite.

Pedidos de Desembolso com data de Geração depois da data Prazo Limite são priorizados.

Na etapa de classificação é feita a análise das regras de priorização definida pela gestão. Para este projeto os Pedidos de Desembolso foram classificados pela Secretaria da Controladoria Geral do Estado em:

Alerta Tempo: Pedidos de Desembolso acima do prazo limite.

Alerta Unidade Gestora: Pedidos de Desembolso acima do prazo limite e de Unidades Gestoras consideradas prioritárias.

Alerta Tema: Pedidos de Desembolso acima do prazo limite e relacionados a Temas considerados prioritários.

Alerta Valor: Pedidos de Desembolso acima do prazo limite e com valor acima de uma quantia específica.

Aguardando Ordem Bancária: Pedidos de Desembolso que na análise em questão não são priorizados.

Com todo processo em mãos houve o entendimento que um método de classificação utilizando treinamento supervisionado com árvores de decisão tornaria o processo de classificação mais eficiente, pois reduziria a interferência humana no processo de classificação utilizando o excel que é uma ferramenta não ideal para fazer esse tipo de tarefa, além disso, padronizaria a classificação por meio de modelos fixos, é de se mencionar que com o modelo gerado o processo de classificação é feita de forma algorítmica.

3.2 Entendimento dos dados

3.2.1 Base de dados

A base de dados utilizada pelo SAD proposto é a base de dados dos Pedidos de Desembolsos fornecida pela Secretaria da Fazenda. Tal base contém as informações de todos os Pedidos de Desembolsos das unidades gestoras do Estado. A base existente possui 29333 registros com 12 atributos relativos mês de agosto do ano de 2016.

3.2.2 Dicionário dos dados

A Tabela 1 apresenta o dicionário dos dados.

Tabela 1 - Dicionário de dados.

Campo	Descrição	Tipo
UO SIGLA	Refere-se à unidade gestora que executa a despesa.	Char
Fonte Sintética	Refere-se à fonte de recurso público que será utilizada para pagamento da despesa.	Num
NU_PD_EXP	Data de inclusão do Pedido de Desembolso (PD) no sistema da Secretaria da fazenda	Char
CGS	Referem-se aos temas de gasto da despesa	Char
DT_INCLUSAO_DOCUMENTO_EXP	Data da inclusão do Pedido de Desembolso no sistema	DATE
PD RazãoSocial	Nome da empresa fornecedora do estado	CHAR
OBS_EXP	Campo aberto de observação. Contém detalhes sobre a despesa, além de poder incluir a data de competência.	CHAR
NU ORDEM BANCARIA EXP	Número da Ordem Bancária Expedida. Fato posterior ao Pedido de Desembolso que autoriza o pagamento da despesa pelo banco.	CHAR
DH_COMPETÊN CIA_AJUST	Data extraída do campo OBS_EXP, é a data de competência ajustada	DATE
Valor PD	Custo da despesa	NUM
SITUACAO	Classificação atribuída ao Pedido de Desembolso pelos analistas de acordo com as regras	CHAR

Transformação:

- Seleção dos dados que serão utilizados pelo modelo.
- Criação de novas colunas que serão utilizadas pelo modelo.
- Processo de limpeza (conversão de tipos).

Carga:

- Base de treino transformada para classificador dos Pedidos de Desembolso
- Dados de Pedidos de Desembolso transformados para serem classificados
- Carga da base de treino para classificador dos Pedidos de Desembolso.
- Transformações de dados na base de treino com o objetivo de alimentar o algoritmo que gera árvore de decisão.

Nas figuras 6 e 7 encontram-se exemplos de ETL usado no sistema.

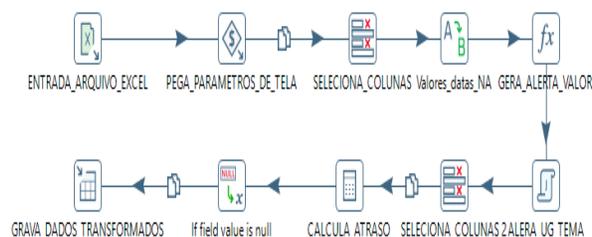


Figura 6 - Transformation que carrega dados a serem utilizados no treinamento do modelo.



Figura 7 - Job responsável por carregar dados transformados e gerar modelo de classificação.

3.3 Preparação dos Dados

3.3.1 Pentaho Data Integration

É uma ferramenta de ETL open source que permite manipular dados. Utilizada amplamente em aplicações de *Analytics*.

Através do PDI (*Pentaho Data Integration*) foram realizados os processos descritos a seguir:

Extração:

- Extração da base de treino
- Extração de dados de Pedidos de Desembolso que serão classificados

3.3.2 Modelagem

Nessa fase, foi utilizado a biblioteca *rpart* (*Recursive partitioning for classification, regression trees*) [11], disponível no ambiente de desenvolvimento integrado R [12]. O *rpart* implementa ideias encontradas no CART

(Classification and Regression Trees) de Reiman, Friedman, Olshen e Stone [13].

O rpart dá a possibilidade de criar modelos de classificação e regressão utilizando árvores binárias. O rpart utiliza a técnica TDIDT (Top-Down Induction of Decision Tree) [14]. O critério de seleção de atributos de particionamento escolhido foi o índice de gini, disponibilizado implementado pelo rpart.

O processo de ETL para etapa do treinamento e validação do modelo preparou os dados da seguinte forma:

Seleção da variável alvo: A coluna SITUACAO é o alvo do modelo, essa coluna descreve qual a situação do Pedido de Desembolso já classificado pela Secretaria da Controladoria da Geral do Estado de Pernambuco.

Criação e seleção de colunas predictoras: Para alimentação do modelo foram criadas colunas baseada nas regras de classificação de Pedidos de Desembolso são elas:

- UG PRIORIDADE: Essa coluna informa se a unidade gestora de um Pedido de Desembolso em questão é prioritária.
- TEMA PRIORIDADE: Essa coluna informa se o tema de um Pedido de Desembolso em questão é prioritário.
- VALOR PRIORIDADE: Essa coluna informa se o valor do Pedido de Desembolso em questão está acima de uma certa quantia.
- ATRASO: Essa coluna informa quantos dias o Pedido de Desembolso está atrasado.

Unidade Gestora Prioridade	Tema Prioridade	Valor Prioridade	Atraso	Situação
SIM	NÃO	NÃO	60	AGUARDANDO OB
SIM	NÃO	NÃO	100	ALERTA UNIDADE GESTORA
NÃO	NÃO	SIM	130	ALERTA VALOR

Figura 8 - Exemplo de variáveis de entrada para treino do modelo.

3.4 Avaliação

Para o processo de validação foi selecionado a base para treino e validação do modelo de agosto de 2016 com 27 mil onde cada registro está classificado com uma situação.

Foram selecionados de forma aleatória 80% dos dados para treinamento, e 20% para

validação. O balanceamento da base de treino por meio do método oversampling não aumentou significadamente o grau de informação das classes minoritárias. A árvore gerada pelo rpart é mostrada na Figura 9. Para a avaliação do resultado foi utilizado a matriz de confusão, conforme mostra a Figura 10.

Como resultado tivemos a classificação correta 99% dos dados de validação. A especificidade que representa a proporção de verdadeiros negativos e a sensibilidade que representa a proporção de verdadeiros positivos para cada classe obtiveram um resultado acima dos 99%.

O resultado mostrou que para as regras de priorização de Pedidos de Desembolso utilizada na base de agosto de 2016 a árvore binária conseguiu se adequar quase perfeitamente para cada tipo de situação.

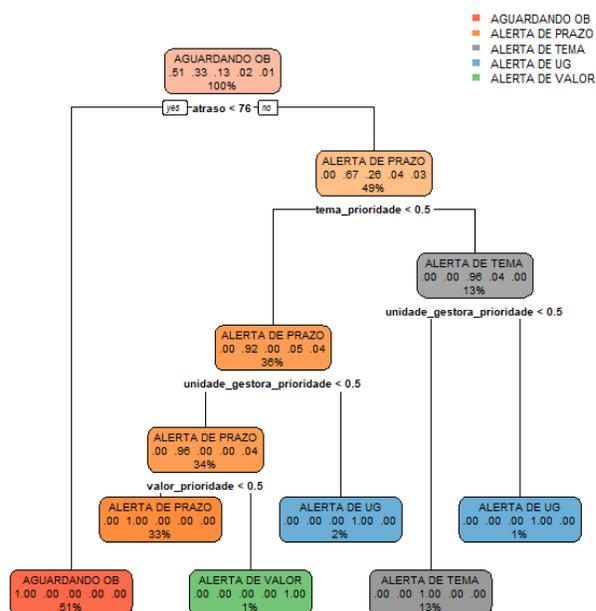


Figura 9 - Árvore de decisão gerada para classificação de Pedidos de Desembolso baseado na regra utilizada em Agosto de 2016.

	AGUARDANDO OB	ALERTA DE PRAZO	ALERTA DE TEMA	ALERTA DE UG	ALERTA DE VALOR
AGUARDANDO OB	3016	0	0	0	0
ALERTA DE PRAZO	0	1900	0	0	0
ALERTA DE TEMA	0	0	763	0	0
ALERTA DE UG	0	0	0	122	0
ALERTA DE VALOR	0	2	1	0	83

Figura 10 - Matriz de confusão.

3.5 Implantação

Para o desenvolvimento da aplicação foi utilizado a plataforma *Pentaho* que é uma plataforma de BI (*Business Intelligence*), o Pentaho dá a capacidade de desenvolver *dashboards* totalmente integráveis aos servidores Linux e Windows, facilitando o *deploy* da aplicação. Além disso ele prover segurança ao uso da aplicação, ideal para o sistema proposto.

O *pentaho* possibilita o desenvolvimento de dashboards através do Community Dashboard Editor (CDE). O CDE permite o rápido desenvolvimento de dashboards através da simplificação do processo de criação, design e integração de dados.

4 Ferramenta

Os objetivos do sistema de apoio a decisão proposto são descritos a seguir:

- Disponibilidade à usuários simultâneos.
- Capacidade de segurança de acesso.
- Automatização do processo de priorização.
- Adaptabilidade a novas regras de priorização.

4.1 Telas

O sistema tem duas telas, a primeira tela chamada de Tela Treino, é responsável, por gerar modelos de classificação. O usuário definirá a base de treino para gerar um modelo, assim como também irá definir quais são as regras de priorização que foram utilizadas para essa base de treino. Um nome para o modelo será definido. Este modelo será salvo no sistema para ser usado em uma classificação. A Figura 11 mostra a tela de treino.

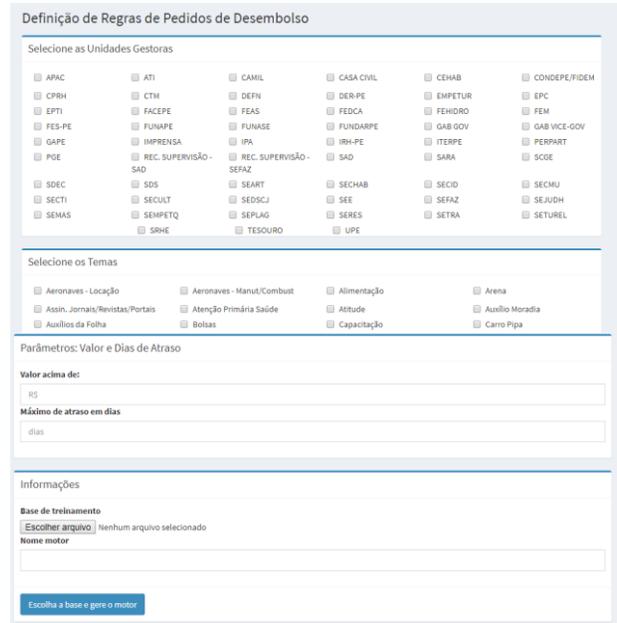


Figura 11 - Tela Treino.

A segunda tela chamada de Tela de Classificação será responsável por utilizar um modelo já salvo no sistema para classificar Pedidos de Desembolso. O usuário selecionará um modelo, assim como um arquivo de Pedidos de Desembolso para ser classificado em situações de alerta. A Figura 12 mostra a tela de classificação.

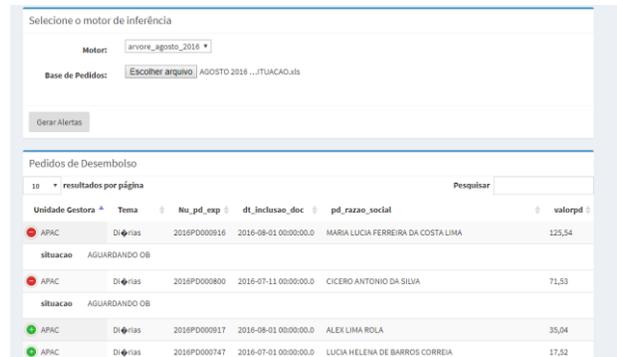


Figura 12 - Tela Classificação.

5 Conclusões

O seguinte trabalho alcançou o objetivo de criar um sistema que auxilie o processo decisório relacionado a um assunto importante que impacta diretamente todos na sociedade que são as despesas públicas. A aplicação de métodos, técnicas e ferramentas proporcionou de forma satisfatória a automatização do processo de

<http://dx.doi.org/10.25286/rep.v3i3.975>

classificação dos Pedidos de Desembolso do Estado de Pernambuco por meio de árvores de decisão.

Para o problema proposto de classificação uso de árvores de decisão se adequou bem, com um resultado satisfatório baseado na acurácia apresentada de 99%. O sistema desenvolvido dá a possibilidade da aplicação real de classificação dos Pedidos de Desembolso podendo ajudar ao processo de tomada de decisão e análise de dados pela Secretaria da Fazenda do Estado de Pernambuco (SEFAZ) e pela Secretaria Controladoria da Geral do Estado de Pernambuco (SCGE).

Referências

- [1] LAUDON, Kenneth C.; LAUDON, Jane Price. **Sistemas de informação com internet**. Rio de Janeiro: LTC, 1999.
- [2] GÓES, B. C. Administração Pública sob o Princípio da Eficiência. Artigo, **Escola da Magistratura do Estado do Rio de Janeiro**, Rio de Janeiro, 2014.
- [3] SPRAGUE, Jr.; WATSON, H. **Decision support systems: putting theory into practice**. USA: Prentice-Hall, 1989.
- [4] BIDGOLI, H. **Decision Support System - Principles and Practice**. New York: West Publishing Company, 1989.
- [5] BARBOSA, G. R. **Sistemas de Apoio A Decisão sob enfoque dos profissionais de tecnologia da informação e decisores**. Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife, 2003.
- [6] ALTER, S. **Decision Support Systems: current practice and continuing challenges**. California: Addison-Wesley Publishing Company, 1980.
- [7] CREPALDI G. Paola; Avila R. N. P; Oliveira J. P. M.; Rodrigues P. R.; Martins L. R. Um estudo sobre a árvore de decisão e sua importância na habilidade de aprendizado. Artigo, Faculdade Integrada Inesul Londrina, Londrina.
- [8] GARCIA, S.C.. O uso de árvores de decisão na descoberta de conhecimento na área da saúde. In: SEMANA ACADÊMICA, 2000, Rio Grande do Sul. **Anais...** Rio Grande do Sul:UFRGS, 2000.
- [9] PETERMANN, R.J. **Modelo de Mineração de Dados para Classificação de Clientes em Telecomunicações**. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2006.
- [10] STAIR, R. **Princípios de sistemas de informação: uma abordagem gerencial**, 2. Ed. Rio de Janeiro: LTC, 1998.
- [11] THERNEAU, T.; ATKINSON, B.; RIPLEY, B. Rpart: Recursive Partitioning and Regression Trees, 2012. R package. Disponível em: <<http://CRAN.R-project.org/package=rpart>>.
- [12] R Core Team, R: A Language and Environment for Statistical Computing Vienna, Austria, 2012. Disponível em: <<http://R-project.org/>>.
- [13] SILVA, E. M. **Descoberta de Conhecimento com o uso de Text Mining: Cruzando o Abismo de Moore**. Dissertação de Mestrado. Universidade Católica de Brasília, Brasília - DF. 2002.
- [14] BREIMAN, L. et al. **Classification and Regression Trees**. Chapman and Hall, 1984.
- [15] FRIZARINI, C. **Algoritmo para Indução de Árvores de Classificação para Dados Desbalanceados**. Dissertação de Mestrado. Universidade São Paulo, São Paulo, 2013.