



Uma Sugestão de um Algoritmo para Promover a Qualidade de Dados em um *Data Mining* para Campanhas de Marketing Direto

A Suggestion of an Algorithm to Promote Data Quality in Data Mining for Direct Marketing Campaigns

Recebido: 02/02/2023 | Revisado: 07/02/2023 | Aceito: 21/02/2023 | Publicado: 22/02/2023

Eduardo Alves de Freitas

Universidade Cidade de São Paulo - UNICID

<https://orcid.org/0000-0003-0814-4239>

eaf.freitas@gmail.com

Irapuan Glória Júnior

Fatec Santana de Parnaíba

<https://orcid.org/0000-0003-2973-3470>

ijunior@ndsgn.com.br

Ana Paula Freitas de Lima

Fatec Santana de Parnaíba

<https://orcid.org/0000-0001-7617-0047>

apfl2.lima@gmail.com

Resumo

Um dos grandes obstáculos a ser superado pelos profissionais especialistas em análise da informação é a transformação dos dados em informação. As grandes empresas geram cada vez mais dados de diferentes fontes, é plausível encontrar problemas em sua padronização. O marketing baseia-se em informação para a criação de campanhas e outras ações, desta forma, é imprescindível que esses dados estejam inseridos corretamente. Diante deste contexto, a presente pesquisa buscou identificar os principais requisitos necessários para a clarificação dos dados com a sugestão de um algoritmo a ser implementado de acordo com o ambiente da empresa que serviu como objetivo de estudo. A metodologia utilizada foi o *Design Science Research*. O resultado obtido foi um modelo computacional apresentado em diagramas UML que irão realizar a limpeza das informações de uma base de dados de uma empresa. A contribuição para a teoria é de apresentar os problemas oriundos de uma empresa em relação a qualidade dos dados. A contribuição para a prática é de apresentar uma sugestão de algoritmo para a limpeza dos dados aos gerentes de TI, analista de sistemas e programadores utilizarem em seus sistemas.

Palavras-chave: Clarificação, Banco de Dados, Marketing, Algoritmo



Abstract

One of the biggest obstacles to be overcome by professionals specialized in information analysis is the transformation of data into information. Large companies generate more and more data from different sources, it is plausible to encounter problems in their standardization. Marketing is based on information for creating campaigns and other actions, therefore, it is essential that this data is entered correctly. Given this context, this research sought to identify the main requirements for data clarification with the suggestion of an algorithm to be implemented according to the company's environment that served as the study objective. The methodology used was Design Science Research. The result obtained was a computational model presented in UML diagrams that will perform the cleaning of information from a company's database. The contribution to the theory is to present the problems arising from a company in relation to data quality. The contribution to the practice is to present an algorithm suggestion for data cleaning to IT managers, systems analysts and programmers to use in their systems.

Keywords: Clarification, Database, Marketing, Algorithm.

1. Introdução

A existência de mercados tão competitivos e dinâmicos a empresa é conduzida a implantar sistemas que auxiliam rápidos processos de tomada de decisão (Kotler et al., 2021). Atualmente uma grande parte das empresas procura conhecer as necessidades do consumidor e ter tais informações auxilia o profissional de marketing no processo de mineração de dados tanto para alavancar suas vendas como para fidelizar seus clientes (Armstrong et al., 2014; Kotler et al., 2021; Zanone, 2019).

Os avanços da tecnologia tanto de hardware quanto de comunicação têm produzido um problema de superabundância de dados, pois a capacidade de coletar e armazenar dados tem superado a habilidade de analisar e extrair conhecimento destes. Nesse contexto, é necessária a aplicação de técnicas e ferramentas que transformem, de maneira inteligente e automática, os dados disponíveis em informações úteis, que representem conhecimento para uma tomada de decisão estratégica nos negócios (Castro & Ferrari, 2017).

O processo de tomada de decisão em uma empresa envolve muitas variáveis e é extremamente importante a utilização de sistemas computacionais e profissionais qualificados para o levantamento e análise das informações necessárias, em que a correta utilização destes dois elementos indicará o sucesso no processo de tomada de decisão.



Outro desafio a ser superado pelos analistas, é a unificação dos cadastros oriundos de diversas fontes e diversos aplicativos e ou sistemas operacionais em que o profissional deve estar apto a receber as bases de dados nos mais diversos formatos, como Excel, Access, SQL, texto, texto com delimitador de campo de registro e outros, pois nem sempre a fonte que disponibilizou os dados possui capacidade técnica suficiente para exportar corretamente seus dados, mesmo quando o aplicativo utilizado possui tais recursos (Pressman & Maxim, 2016; Sommerville, 2011).

Sendo assim, a preparação de dados é uma ferramenta importante no processo intenso de melhoria dos dados para as campanhas de Marketing (Kotler et al., 2021), pois os processos de normalização de dados, identificação de dados inconsistentes e limpeza de dados auxiliam de forma efetiva a mineração de dados (Elmasri & Navathe, 2019), que por sua vez pode ajudar as empresas a competir de maneira mais eficaz e a identificar importantes tendências de negócios que se transformaram no movimento para alcançar novas estratégias de mercado (Mintzberg, 2006).

Diante do exposto, este trabalho possui como questão de pesquisa: "Quais os processos que um algoritmo precisa possuir para auxiliar na melhoria da qualidade de dados de uma empresa?". Possui como objetivos: (1) Identificar as principais regras de limpeza de dados; e (2) Sugerir um algoritmo para a realização da limpeza de dados.

2. Referencial Teórico

2.1. Mineração de Dados

A mineração de dados é uma disciplina interdisciplinar e multidisciplinar que envolve conhecimento de áreas como banco de dados, estatística, aprendizagem de máquina, computação de alto desempenho, reconhecimento de padrões, computação natural, visualização de dados, recuperação de informação, processamento de imagens e



de sinais, análise espacial de dados, inteligência artificial, entre outras (Castro & Ferrari, 2017).

Um padrão pode ser definido como sendo uma afirmação sobre uma distribuição probabilística, na forma de regras, algoritmos, fórmulas e funções, entre outras, e encontrar padrões requer que os dados brutos sejam sistematicamente simplificados, desconsiderando o que é específico e privilegiando o que é genérico (Amaral, 2001), em que os dados padronizados podem ser utilizados em processo de mineração de dados.

As empresas estão cada vez mais impulsionadas pela análise de dados, portanto, há grande vantagem profissional em ser capaz de interagir com competência dentro e fora dessas empresas. Compreender os conceitos fundamentais, e ter estruturas para organizar o pensamento analítico de dados não só permitirá uma interação competente, mas ajudará a vislumbrar oportunidades para melhorar a tomada de decisões orientada por dados ou ver ameaças competitivas orientadas por dados (Provost & Fawcett, 2016).

2.2. Preparação de Dados

A preparação de dados é o conjunto principal de processos para integração de dados que reúnem dados de diversos sistemas de origem, transformam-nos de acordo com regras comerciais e técnicas e preparam-nos para etapas posteriores em seu ciclo de vida quando se tornam informações usadas por consumidores corporativos (Sherman, 2015).

A maioria das bases de dados reais, as etapas de pré-processamento consomem muito tempo e demandam bastante trabalho, mas o sucesso da mineração depende fortemente do cuidado dedicado a essa etapa do processo de descoberta de conhecimentos em bases de dados (Castro & Ferrari, 2017).

As etapas envolvidas na preparação de dados são: captura dos dados, reformatação, consolidar e validar, transformar (através de regras de negócio e conversões técnicas), analisa a qualidade dos dados, limpeza e armazenagem dos dados (Sherman, 2015).



Conhecer e preparar de forma adequada os dados para análise é uma etapa que pode tornar todo o processo de mineração muito mais eficiente e eficaz. Por outro lado, dados mal ou não pré-processados podem inviabilizar uma análise ou invalidar um resultado (Castro & Ferrari, 2017).

2.3. Marketing

O marketing é um processo social, onde grupos de pessoas obtêm aquilo que necessitam e desejam, diante da criação e oferta de produtos e serviços de valor (Kotler et al., 2021). Outra vertente afirma que é um conjunto de instituições e processos de comunicação, criação, entrega e troca de produtos e serviços que geram valor ao público-alvo (Armstrong et al., 2014; Kotler & Keller, 2006).

É fundamental que as estratégias de marketing adotadas pelas empresas, independente do seu segmento ou porte; se destacando ao utilizar o marketing estratégico, estabelecendo um bom posicionamento de sua marca, ao gerar valor ao cliente e consequentemente aumentar sua lucratividade (Kotler et al., 2021).

Diante da Tecnologia da Informação (TI), é preciso o uso de tecnologias e ferramentas de gestão para a garantia do valor agregado ao produto ou ao serviço, que permita que as empresas se diferenciem perante a concorrência (Zanone, 2019).

A maior aplicação de técnicas de mineração de dados está no marketing, para tarefas como marketing direcionado, publicidade online e recomendações para venda cruzada. A mineração de dados é usada para gestão de relacionamento com o cliente para analisar seu comportamento a fim de gerenciar o desgaste e maximizar o valor esperado do cliente (Provost & Fawcett, 2016).

2.4. Qualidade de Dados

A noção de qualidade está diretamente relacionada com o nível de qualidade da coleta de dados percebida pelo cliente. Isso significa que clientes diferentes, mesmo



aqueles cujas necessidades são apenas sutilmente diferentes, podem classificar uma coleta de dados de maneira diferente. Sendo assim, os clientes exigirão que os dados sejam cada vez melhores, e melhor adaptados para atender às suas necessidades específicas (Redman, 2013).

As consequências da má qualidade dos dados muitas vezes são vivenciadas no dia a dia, mas, muitas vezes, sem fazer as devidas conexões com suas causas. A estimativa é que os problemas de qualidade de dados custam às empresas americanas mais de 600 bilhões de dólares por ano. Em organizações privadas, como empresas de marketing ou bancos, existem vários cadastros de clientes, atualizados com diferentes procedimentos organizacionais, resultando em informações inconsistentes e duplicadas (Batini & Scannapieca, 2006).

A baixa qualidade dos dados é um problema que afeta a maior parte das bases de dados reais. Assim, as ferramentas para a limpeza de dados atuam no sentido de imputar valores ausentes, suavizar ruídos, identificar valores discrepantes (outliers) e corrigir inconsistências (Castro & Ferrari, 2017).

3. Metodologia

Esta pesquisa possui, conforme apresentado na Tabela 1, a natureza qualitativa (Theóphilo & Martins, 2009), com o emprego da metodologia de *Design Science Research* (DSR) como meio de atender a requisição de criação um algoritmo (Nunamaker Jr et al., 1990) para realizar a limpeza de dados de um repositório de dados



Tabela 1 - Características da Pesquisa

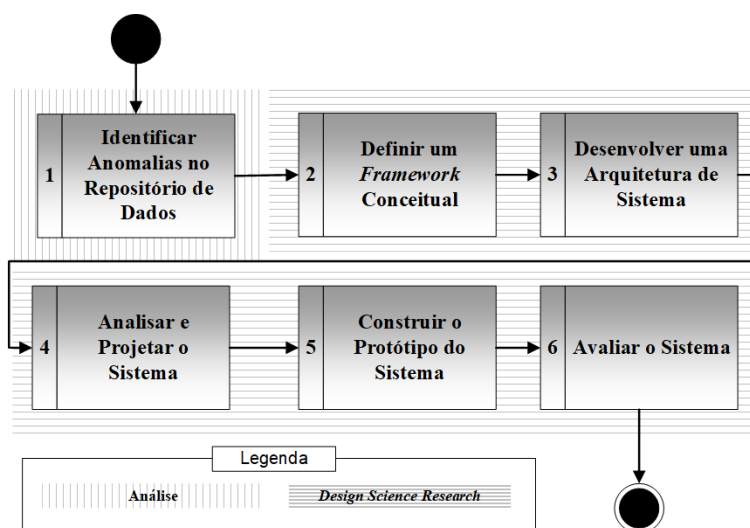
Item	Descrição	Autor(es)
Natureza	Qualitativa	Martins e Theóphilo (2009)
Metodologia	<i>Design Science Research</i>	Nunamaker Jr <i>et al.</i> (1990)
Unidade de Análise	Coleta de dados junto aos desenvolvedores	
Resultado Esperado	Sequenciamento de tarefas por meio de sugestão algorítmica	

3.1. Procedimentos Metodológicos

Os procedimentos aplicados na pesquisa e apresentados na Figura 1 são:

- ✓ **Passo1: Identificar as Anomalias no Repositório de Dados.** Diante do repositório de dados fornecido pela Empresa-Alpha foram feitas leituras de dados, de forma randômica, para identificar as principais falhas;
- ✓ **Passo2: Definir um *Framework* Conceitual.** Foi elaborado, a partir das falhas identificada nos dados, os requisitos a serem considerados no processo de normatização dos dados;
- ✓ **Passo3: Desenvolver uma Arquitetura de Sistema.** Foram apresentados os processos e a arquitetura de software;
- ✓ **Passo4: Analisar e Projetar o Sistema.** A solução foi apresentada por meio de um diagrama de *Unified Modeling Language* (UML) que é utilizado para a representação de sistemas computacionais (OMG, 2015);
- ✓ **Passo7: Construir o Protótipo do Sistema.** Foram apresentadas e explicadas as telas do sistema;
- ✓ **Passo8: Avaliar o Sistema.** Foi apresentada uma tabela com as relações entre os requisitos e os processos que os contemplam.

Figura 1 - Procedimentos Metodológicos



4. Análise e Interpretação dos Resultados

4.1. Definição do *Framework* Conceitual

Ao analisar o repositório de dados foi possível identificar (Tabela 2) que existem registros com conteúdo que apresentam falta de padronização dos caracteres (**REQ01**) com palavras iniciando em maiúsculo e outros todas em minúsculo, sendo que serão transformadas em maiúsculo.

Outra característica identificada é a retirada de caracteres acentuados para não-acentuados (**REQ02**) com o intuito de facilitar a identificação de logradouro idênticos. Exemplos de conversão são os caracteres "Á", "Ã" e "Â" para "A".

Existem vários caracteres que são incluídos em um logradouro e podem trazer uma melhor estética ao dado, entretanto como não existe um padrão na capturados dos dados, estes caracteres são considerados indesejáveis (**REQ03**) no processo de preparação dos dados. Constata-se que a virgula pode ser utilizada para separar o logradouro do número



do logradouro, o ponto pode ser usado na abreviação de algumas palavras e o traço para separar palavras de números. Todos estes caracteres são utilizados para que a informação tenha um melhor entendimento por parte do usuário, entretanto pode-se considerar que isto é irrelevante e que tais caracteres interferem no processo de identificação de igualdade da informação quando realizado por sistemas informatizados, sendo assim, estes elementos devem ser substituídos pelo caracter branco, por exemplo a substituição por um espaço em branco dos caracteres "," (vírgula), "." (ponto), "-" (traço), "/" (barra), "*" (asterisco) e "_" (*underline*).

Há também os casos em que há a concatenação de dados de tipos diferentes, desta forma, é necessária a separação de valores numéricos dos não numéricos (**REQ04**), com os exemplos de separação de caracteres numéricos e não numéricos são as strings "RUA JOAO CARLOS SILVA 0265 **BL5**" e "R JOAO CARLOS **SILVA265** BL 5" seriam convertidos, respectivamente, para "RUA JOAO CARLOS SILVA 0265 **BL 5**" e "R JOAO CARLOS **SILVA 265** BL 5".

A incorreta inclusão de zeros à esquerda sem necessidade interfere no processo de identificação de logradouros idênticos, desta forma, necessitam ser retirados (**REQ05**), como os encontrados nas *strings* "RUA JOAO CARLOS SILVA **0265**", "R JOAO CARLOS SILVA **265**", "RUA JOAO CARLOS SILVA **00265**" e "RUA JOAO CARLOS SILVA **265**" convertida em "RUA JOAO CARLOS SILVA **265**".

É correta a existência de um espaço em branco para a separação de duas palavras, entretanto o incorreto processo de captura dos dados pode ocasionar excesso de espaços em branco entre duas palavras, assim, é necessário eliminar o excesso de espaços em branco entre as palavras (**REQ06**), como por exemplo "RUA ALAGOAS" seria convertido para "RUA ALAGOAS".



A possibilidade de existência de números no nome do logradouro (**REQ07**) pode permitir dubiedade, desta forma, devem ser indicados por extenso como em "RUA 15 DE NOVEMBRO" seria convertida para "RUA QUINZE DE NOVEMBRO".

O processo de padronização do tipo do logradouro (**REQ08**) deve eliminar as abreviaturas com o intuito de eliminar possíveis ruídos no processo de identificação de igualdades, como nos exemplos das *strings* "R" e "R." para "RUA" e "AVEN" e "AV" para "AVENIDA".

O complemento do logradouro é importante para que se possa distinguir as diferentes unidades que integram do mesmo edifício ou terreno e assim identificar a utilização das partículas bloco, apartamento, lote, quadra, sala e outros, indicam a correta localização de uma unidade que faz parte de um logradouro, desta forma, não poderão ter abreviaturas incorretas (**REQ09**), como por exemplo as *strings* "R JOAO CARLOS SILVA 265 **BLOC** 5 AP 21", "R JOAO CARLOS SILVA 265 **BLOCO** 5 APTO 21", "R JOAO CARLOS SILVA 265 **BLO** 5 **APARTAMENTO** 21" e "R JOAO CARLOS SILVA 265 **BLOCO** 5 **APT** 21" convertidas para "R JOAO CARLOS SILVA 265 **BL** 5 **APTO** 21".

A localidade é uma informação essencial para o envio de uma correspondência, sendo assim, é importante que a localidade esteja normalizada para a correta utilização desta informação, desta forma, deve-se corrigir qualquer abreviatura ou incorreções gerando a padronização das localidades (**REQ10**), como nas *strings* "S B C" e "SAO B DO CAMPO" que são convertidas para "SAO BERNARDO DO CAMPO".

O bairro é considerado uma informação importante na montagem do endereço de um imóvel e devem estar padronizados (**REQ11**), como nos exemplos: "VL S JOSE", "VILA S JOSE" e "V S JOSE" que devem ser convertidos em "VILA SAO JOSE".

O Código de Endereçamento Postal (CEP) é o elemento que compõe o endereço de um imóvel e é uma informação fundamental para a correta identificação do local do



imóvel, sendo assim, deverá ser padronizado (**REQ12**) no formato de 8 dígitos contínuos, como nos exemplos "6020194", "06.020-194", "6.020.194", "06020-194" e "06.020.194" deverão ser convertidos em "06020194".

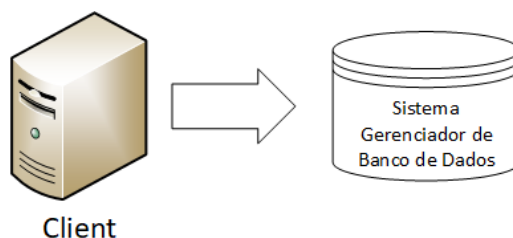
Tabela 2 - Requisitos

#	Requisitos
REQ01	Falta de Padronização dos Caracteres
REQ02	Eliminar Caracteres Acentuados
REQ03	Eliminar Caracteres Indesejáveis
REQ04	Separar de valores numéricos dos não numéricos
REQ05	Eliminação de zeros à esquerda de uma sequência de números
REQ06	Eliminação de excesso de espaços em brancos entre palavras
REQ07	Convertendo números no nome dos logradouros para extenso
REQ08	Padronização do tipo do logradouro
REQ09	Padronização correta da abreviação do complemento do logradouro
REQ10	Padronização das localidades
REQ11	Padronização do bairro
REQ12	Padronização do CEP

4.1. Desenvolver uma Arquitetura de Sistema

O algoritmo irá contribuir para clarificação dos dados do repositório de forma a ser executado sempre que necessário, conforme apresentado na Figura 2.

Figura 2 – Arquitetura da Solução



4.2. Analisar e Projetar o Sistema

A diagramação do sistema foi realizada por meio da UML (OMG, 2015) com a utilização do Diagrama de Atividades (DA_{tiv}) apresentado na Figura 3. A Figura 3a corresponde a três processos, sendo o primeiro a conexão com o Banco de Dados (**P01**), seguido do macro-processo de realização da limpeza dos dados (**P02**) e o último de encerramento da conexão com o banco de dados (**P03**).

Na Figura 3b está representado o detalhamento do processo macro de limpeza, com cada processo relacionado a um requisito funcional identificado no levantamento de dados.

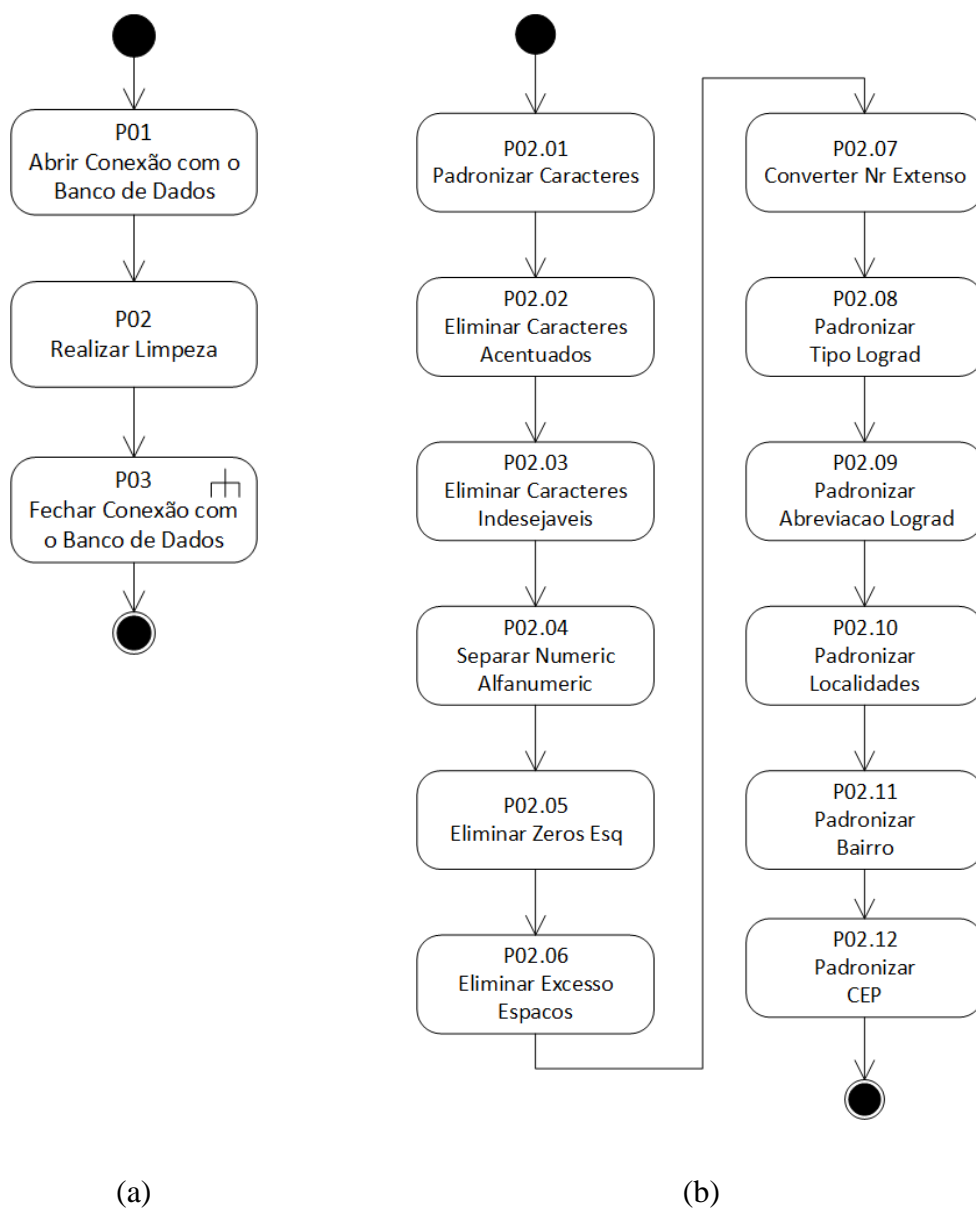
Os analistas e desenvolvedores da Empresa-Alpha poderão, a critério do gerente de TI, realizar a codificação dos processos de acordo com o conhecimento técnico da equipe, por exemplo, o primeiro processo é o Padronizar Caracteres (**P01**) que consiste em transformar em maiúsculos todos os caracteres, uma sugestão é de aplicar a função UPPER (Microsoft, 2022b) em todos os campos de todos os registros para que os dados fiquem em maiúsculo, conforme o Código 1:

Código 1 – Padronizar Caracteres (P01)

```
Update TB_Cliente  
Set Nm_Cliente = Upper(Nm_Cliente)
```

Caberá ao gestor escolher como realizar e em qual camada, se por banco de dados com funções e comandos T-SQL (Microsoft, 2022b) ou por meio de rotinas em linguagem de programação, com o .NET (Microsoft, 2022a).

Figura 3 - Diagramas de Atividade

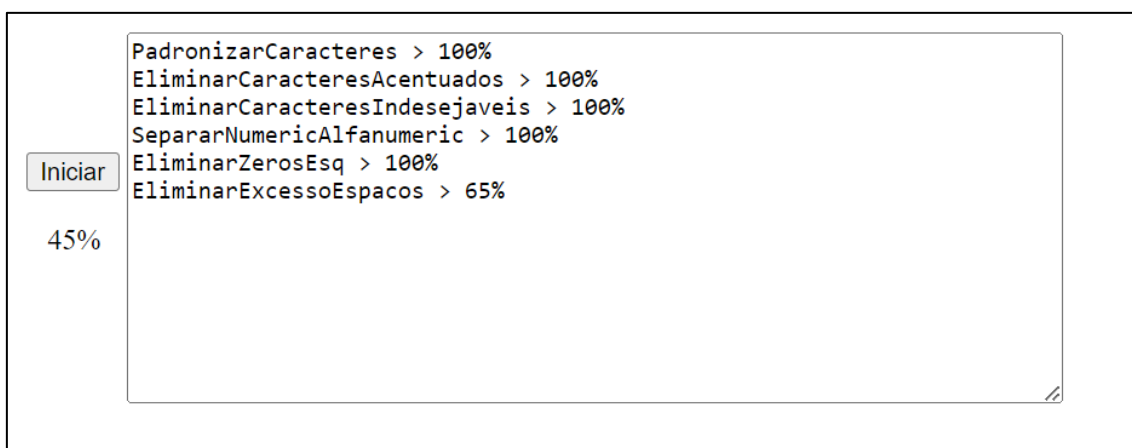


4.3. Construir o Protótipo do Sistema

A constituição da interface homem-máquina será por meio de uma tela com um botão de início (Iniciar) que, após ser pressionado, exibirá abaixo a porcentagem total concluída e no box lateral qual o processo que está sendo executado e a porcentagem concluída até o presente momento (Figura 4).

A execução deverá ser precedida de um *backup* conforme é recomendada nas diversas boas práticas de atuação em banco de dados (Elmasri & Navathe, 2019; Pressman & Maxim, 2016).

Figura 4 – Tela da Solução de Limpeza de Dados



4.4. Avaliar o Sistema

Baseando-se nos requisitos identificados na Empresa-Alpha para limpeza de dados e nos processos criados e apresentados na Figura 3a e Figura 3b, foi estabelecido uma relação de objetivos, entre requisito e processo, que culminou na Tabela 3.



Tabela 3 - Relacionamento dos Processos

Requisito	Descrição	Diagrama de Atividade	Processos
REQ01	Falta de Padronização dos Caracteres	P02.01	PadronizarCaracteres
REQ02	Eliminar Caracteres Acentuados	P02.02	EliminarCaracteresAcentuados
REQ03	Eliminar Caracteres Indesejáveis	P02.03	EliminarCaracteresIndesejaveis
REQ04	Separar de valores numéricos dos não numéricos	P02.04	SepararNumericAlfanumeric
REQ05	Eliminação de zeros à esquerda de uma sequência de números	P02.05	EliminarZerosEsq
REQ06	Eliminação de excesso de espaços em brancos entre palavras	P02.06	EliminarExcessoEspacos
REQ07	Convertendo números no nome dos logradouros para extenso	P02.07	ConverterNrExtenso
REQ08	Padronização do tipo do logradouro	P02.08	PadronizarTipoLograd
REQ09	Padronização correta da abreviação do complemento do logradouro	P02.09	PadronizarAbreviacaoLograd
REQ10	Padronização das localidades	P02.10	PadronizarLocalidades
REQ11	Padronização do bairro	P02.11	PadronizarBairro
REQ12	Padronização do CEP	P02.12	PadronizarCEP



4.5. Discussões

É possível observar que a maioria das ações que devem ser aplicadas para a limpeza dos dados está relacionada a um comando ou função T-SQL (Microsoft, 2022b). Em casos de que é necessário realizar uma rotina de *loop*, uma repetição de comandos, a *Engine* do MS-SQL pode ser muito mais lenta que um código similar executado em uma linguagem de programação, visto que o banco de dados costuma trabalhar no formato *Top-Down* (Microsoft, 2022b).

5. Conclusões

As empresas devem ter base de dados íntegros para que possam exercer as atividades de marketing para atingir seus clientes e alavancar suas vendas, independentemente do produto ou serviços que ofereçam. Existem soluções de mineração que podem auxiliar na identificação dos clientes em potenciais.

Este trabalho apresentou 12 processos que podem corroborar para a eliminação de preenchimentos errôneos de dados e a criação de padronização de informações numéricas e alfanuméricas.

A contribuição para a teoria é de apresentar os problemas oriundos de uma empresa em relação a qualidade dos dados. A contribuição para a prática é a de apresentar uma sugestão de algoritmo para a limpeza dos dados aos gerentes de TI, analista de sistemas e programadores utilizarem em seus sistemas, como em *eCommerce*. Em futuros trabalhos estão previstos a inclusão de padronização das localidades, o tratamento das UF inválidas e a implementação de um algoritmo de *match-code* com a junção do nome e endereço para buscas mais assertivas.



Referencial Bibliográfico

- Amaral, F. C. N. do. (2001). *Data Mining: Técnicas e aplicações para o marketing direto*. Editora Berkeley.
- Armstrong, G., Adam, S., Denize, S., & Kotler, P. (2014). *Principles of Marketing*. Pearson.
- Batini, C., & Scannapieca, M. (2006). Data quality dimensions. *Data Quality: Concepts, Methodologies and Techniques*, 19–49.
- Castro, L. N., & Ferrari, D. G. (2017). *Introdução à mineração de dados: Conceitos básicos, algoritmos e aplicações*. Saraiva.
- Elmasri, R., & Navathe, S. B. (2019). *Sistemas de Banco de Dados*. Pearson Universidades.
- Kotler, P., Kartajaya, H., & Setiawan, I. (2021). *Marketing 5.0: Tecnologia para a humanidade*. Sextante.
- Kotler, P., & Keller, K. L. (2006). *Administração de Marketing: A bíblia do marketing* (12º ed). Pearson.
- Microsoft. (2022a). *Documentação do C#*. <https://docs.microsoft.com/pt-br/dotnet/csharp/>
- Microsoft. (2022b). *SQL Server Technical Documentation*. Microsoft. <https://learn.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver16>
- Mintzberg, H. (2006). *O Processo da Estratégia*. Bookman Editora.
- Nunamaker Jr, J. F., Chen, M., & Purdin, T. D. (1990). Systems development in information systems research. *Journal of management information systems*, 7(3), 89–106.
- OMG. (2015). *Object Management Group—Unified Modeling Language 2.5*. OMG. <https://www.omg.org/spec/UML/2.5/PDF>
- Pressman, R., & Maxim, B. (2016). *Engenharia de Software Uma Abordagem Profissional* (8º ed). McGraw Hill Brasil.
- Provost, F., & Fawcett, T. (2016). *Data Science para negócios*. Alta Books.
- Redman, T. C. (2013). Data Quality Management Past, Present, and Future: Towards a Management System for Data. *Handbook of Data Quality: Research and Practice*, 15–40.
- Sherman, R. (2015). *Business Intelligence Guidebook: From Data Integration to Analytics*. Elsevier.



Journal of Technology & Information

- Sommerville, I. (2011). *Engenharia de Software* (9º ed). Pearson Education.
- Theóphilo, C. R., & Martins, G. de A. (2009). Metodologia da investigação científica para ciências sociais aplicadas. *São Paulo: Atlas*, 2(104–119), 25.
- Zanone, L. C. (2019). *CRM - Customer Relationship Management: Marketing de Relacionamento, Fidelização de Clientes e Pós-venda*. Actual.