



**UNIVERSIDADE CATÓLICA DO SALVADOR – UCSAL**  
**MBA EM TECNOLOGIAS E APLICAÇÕES DE BUSINESS INTELLIGENCE**

**BIG DATA E ANALITYCS PARA DESENVOLVIMENTO DE MODELO  
PREDITIVOS DE INADIMPLÊNCIA**

Disciplina: Análise de Dados e Data Mining.  
Prof.: Grimaldo Lopes  
Aluno: Camila da Silva Oliveira

**Salvador, Maio de 2019**

## Sumário

1.	INTRODUÇÃO .....	3
2.	INADIMPLÊNCIA NO SETOR DE ENERGIA ELÉTRICA .....	4
3.	PROBLEMA MAPEADOS .....	6
4.	SOLUÇÃO PROPOSTA.....	8
4.1.	<i>Sistema Big Data / Data Analytics</i> .....	8
4.2.	<i>Etapas de Desenvolvimento</i> .....	9
4.2.1.	<i>Entendimento do Negócio</i> .....	9
4.2.2.	<i>Entendimento dos Dados</i> .....	9
4.2.3.	<i>Preparação dos Dados</i> .....	9
4.2.4.	<i>Modelagem</i> .....	10
4.2.5.	<i>Avaliação</i> .....	10
4.2.6.	<i>Implementação</i> .....	10
4.3.	<i>Validação e Análise de Resultados</i> .....	10
5.	BENEFÍCIOS E RESULTADOS ESPERADOS .....	12
6.	BIBLIOGRAFIA .....	13

## 1. Introdução

Um dos maiores desafios enfrentados por todas as corporações no cenário atual é conseguir maximizar os resultados e, ao mesmo tempo, reduzir os custos para operacionalização dos processos diários. Para as empresas do setor de energia, esse desafio é ainda maior por se tratar de um setor regulamentado pelo governo, mais suscetíveis a mudança nos cenários políticos e econômicos. Ter acesso a informações com velocidade e confiabilidade, adquirir conhecimentos estratégicos a partir da grande variedade de dados armazenados na operação diária combinados com as informações disponibilizadas pelas redes sociais é um diferencial estratégico. Este diferencial é essencial para garantir competitividade e maior lucratividade. Neste cenário, se insere a Companhia de Eletricidade do Estado da Bahia – Coelba, uma das maiores do país.

Este trabalho tem como objetivo apresentar o desafio enfrentando pela concessionária de energia no que se refere ao combate a inadimplência, propor soluções utilizando as aplicações de Big Data e Analytics disponíveis no mercado para obtenção de novos insights para otimização dos recursos e obtenção de melhores resultados na arrecadação.

## 2. Inadimplência no Setor de Energia Elétrica

A Coelba é a empresa responsável pela distribuição de energia para 5,9 milhões de clientes, sendo a terceira maior concessionária do país em número de clientes e sexta em volume de energia, segundo o site da própria concessionária. A concessão de energia no Brasil é regulada pela Agência Nacional de Energia Elétrica (ANEEL), autarquia responsável por regular o setor elétrico brasileiro, sendo uma das atribuições estabelecer as tarifas cobradas aos consumidores.

Segundo a Cartilha Por Dentro da Conta de Luz divulgada pela ANEEL (2008, p.11), as tarifas são definidas de forma a “assegurar aos prestadores de serviços ganhos suficientes para cobrir custos operacionais eficientes e remunerar investimentos necessários para expandir a capacidade e garantir boa qualidade de atendimento”. De acordo com a ANEEL, dentro do cálculo da tarifa é contabilizado o valor referente ao índice de inadimplência – dívida não paga pelos consumidores. O valor repassado aos consumidores é parcial, o objetivo é evitar de onerar os clientes adimplentes e honestos, incentivando as concessionárias de investirem para a redução desses índices (ANEEL, 2008).

Um dos principais desafios das concessionárias é conseguir reduzir o índice de inadimplência de forma a manter a saúde financeira da empresa sem impactar no aumento dos custos operacionais, maximizando o lucro e tornando a empresa competitiva no mercado. A redução da inadimplência é um dos grandes desafios de todo serviço de energia. O Instituto Acende Brasil sinaliza que estão entre as principais questões que vêm comprometendo os resultados das revisões tarifárias de eletricidade, junto com o furto de energia, a inadimplência gera um prejuízo da ordem de R\$ 6 bilhões ao ano para as concessionárias do país.

Na Coelba o cenário é também desafiador, houve uma queda de 0,94 pontos percentuais na arrecadação de acordo com o release do site Bovespa (2018) comparando com o resultado acumulado de 12 meses do primeiro trimestre de 2018 com o resultado do primeiro trimestre de 2017. No demonstrativo de fechamento do ano de 2018, o resultado do índice de arrecadação (acumulado nos últimos 12 meses) foi de 97,02%, inferior ao seu comportamento no ano de 2017, que encerrou em 98,13%. Segundo o documento, o ambiente macroeconômico não foi favorável, o crescimento econômico foi abaixo do esperado e marcado por incertezas do mercado externo. Segundo Demonstrativo Financeiro da Coelba (2019, p. 3):

Conforme dados do boletim Focus, a previsão de aumento do PIB brasileiro no início de 2018 estava na faixa de 2,7%. Em junho, após a greve, a expectativa de crescimento já havia reduzido para 1,5% e, sem mostrar sinais de recuperação, fechou o ano com uma expectativa de apenas 1,3% de crescimento em relação ao ano anterior.

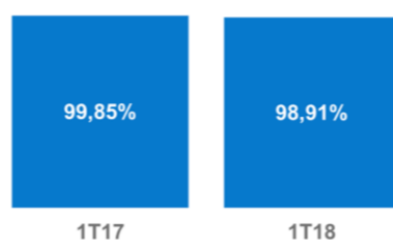


Gráfico 1: Release financeiro publicado no Bovespa

Para este ano, segundo o site da InfoMoney (2019), os primeiros indicadores revelam que o Brasil está crescendo a um ritmo mais lento que em 2018. Para manter ou melhorar os resultados da empresa no que se refere a recuperação da inadimplência a empresa deverá adotar medidas diferenciadas, uma das formas sugeridas seria a exploração dos dados e adoção de modelo analíticos para melhor assertividade das ações.

### 3. Problema Mapeados

Diante do cenário econômico desafiador, as concessionárias de energia precisam utilizar os recursos de forma a maximizar o resultado e reduzir os custos operacionais. Espera-se através de novas tecnologias disponíveis no mercado, processar todas as informações e através de análise estatística criar modelos que possibilite distinguir os diversos perfis de clientes, possibilitando direcionar as ações de cobrança de forma a obter a melhor recuperação, sem aumentar o custo com o processo de cobrança.

Para combate a inadimplência a empresa adota várias modalidades de ação de cobrança no intuito de incentivar o pagamento das faturas. O custo das ações varia de acordo com o grau de complexidade. As ações de cobrança adotadas atualmente pela Coelba de acordo com o release publicado em 2018 são:

- Ações administrativas:
  - SMS (Short Message Service / Serviço de Mensagem Curta) – Comunicação de dívida através de envio de mensagem para celulares;
  - URA (unidade resposta audível) – Contato através de ligações automáticas;
  - Negativação – Encaminhamento do documento fiscal para o cadastro ativo em um dos órgãos de proteção ao crédito.
  - Assessoria são empresas que trabalham com a intermediação entre devedores e credores.
- Negociação – parcelamento da dívida;
- Acionamento jurídica – acionamento jurídico para cobrança da dívida.
- Suspensão de fornecimento de energia – Ação mais conhecida como corte de energia, executada de acordo com os critérios definidos pela ANEEL.

Na operacionalização das suas atividades a concessionária reúne uma variedade significativa de informações dos clientes, como dados de consumo, dados técnicos das instalações, dados econômicos relacionados as contas de energia e informações das ações de cobrança adotadas para recuperação da inadimplência. Além dos dados diretamente ligado a operacionalização do serviço. A empresa possui vários canais de atendimento, mantém contato com os clientes através das principais redes sociais como Facebook e Instagram, uma fonte de dados não explorada. Pode-se trabalhar também com dados comprados das assessorias de cobrança para subsidiar com informações complementares do perfil do cliente no mercado.

Em entrevista realizada na empresa, nem todos os dados são utilizados, em especial, informações oriundas de rede sociais. Os dados são armazenados em banco de dados Oracle, utilizando o sistema comercial CCS e de CRM da SAP. Recentemente, o departamento de cobrança adotou a tecnologia SAP HANA para melhorar a performance de análise e definição de estratégias de cobrança, no entanto, ainda não é empregada na área de cobrança aplicativos voltado para análises estatísticas. Apesar de empregar modelos, a manipulação de dados é realizada de forma descentralizada, não padronizadas e com limitação de dados para análise por haver limitação de recursos, impactando no resultado dos modelos desenvolvidos.

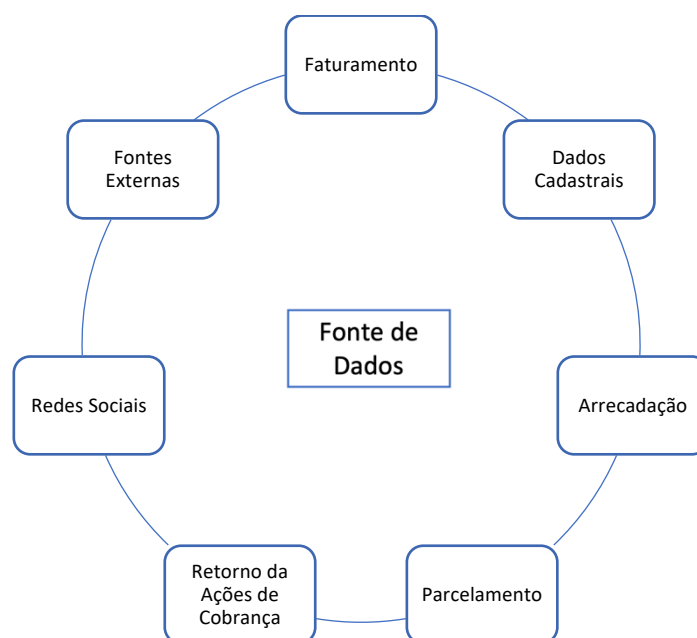


Imagem 1: Fonte de Dados – Área de Cobrança

Observa-se que o departamento de cobrança possui um volume significativo de dados, com variedade de informações e enfrenta problemas no que se refere a consistência, por se tratar de dados descentralizados com métricas calculadas sem padronização. Outra característica do setor é a velocidade que a informação precisa ser processada e disponibilizada. É uma das características da área de cobrança é a velocidade na mudança do cenário econômico, exigindo decisões rápidas e assertivas para tomada de decisão.

Os principais sistemas comerciais da Coelba são soluções SAP, uma das maiores empresas mundiais no setor de software empresarial, com amplo reconhecimento no mercado e uma das líderes no setor. A parceria é de longa data e a concessionária tem crescido e novas soluções foram sendo adquiridas, uma das mais recentes, voltada para o setor específico da cobrança foi o SAP HANA, tecnologia de banco de dados que armazena temporariamente as informações em memória volátil (RAM), coletando informações e desenvolvendo a capacidade de reagir em tempo real, possibilitando um desempenho superior em relação aos banco de dados disponíveis no mercado.

A solução foi adotada no setor de cobrança, na tentativa de atender a demanda de processamento dos dados para operacionalizar o processo de cobrança. Apesar da tecnologia avançada, a utilização da base de dados foi restrita, com acesso limitado as principais base de dados operacional para evitar sobrecarga na cadeia de produção das empresas. O uso da tecnologia trouxe vários ganhos para o departamento, os dados passaram a ser processado com mais velocidade e se tornaram mais disponíveis para os usuários. O resultado positivo incentivou a expansão para outra área da empresa com o acréscimo da solução SAP PA – SAP Predictive Analytics.

O setor de cobrança tem investido na adoção de novas tecnologias, mas o esforço tem sido direcionado para atender a demandas pontuais relacionada a transformação de dados em informação. É necessário revisar o processo para estruturar e incluir as etapas de forma organizada para exploração da captação dos dados, com inclusão das diversas fontes disponíveis, até a etapa da análise da informação, transformação desta informação em conhecimento e disponibilização do resultado de forma clara e estruturada.

## 4. Solução Proposta

A SAP é uma empresa de tecnologia de origem alemã criadora de softwares de gestão de empresas, é líder no mercado no ramo de softwares aplicativos empresariais. A organização chegou no Brasil em 1995 e está presente em inúmeros países. A partir de 2010, a empresa iniciou o investimento na tecnologia de banco de dados em memória, com objetivo de reduzir de forma significativa o tempo de processamento. Os demais serviços oferecidos pela empresa foram modernizados para rodar na arquitetura SAP HANA.

Além do foco na tecnologia em memória, a organização modernizou a tecnologia para utilizar a computação em nuvem – utilização de memória e da capacidade de armazenamento e cálculo de computadores em servidores armazenados em datacenters interligados pela internet, processamento de grande volume de dados, como tecnologia Big Data, Business Intelligence e, principalmente, mineração e análise de dados.

A tendência de investimento da SAP está alinhada com a necessidade e visão da Coelba para o futuro. A parceria já estabelecida facilitará uma futura transição, já que a SAP e suas parceiras possuem grande conhecimentos dos processos da Coelba e das bases de dados. Por conta disso, recomenda-se a manutenção da parceira e investimento dos produtos a empresa. Uma exceção é a visualização dos dados, a empresa já realizou estudos e optou pelo aplicativo de visualização Tableau, por ser mais desenvolvido e intuitivo, neste caso, sugere-se manter o aplicativo.

### 4.1. Sistema Big Data / Data Analytics

Para atender as demandas para processar os dados de fontes de distintas e estruturá-los para melhor utilização, recomenda-se a aquisição do SAP HANA Data Management Suíte. A solução possui a capacidade para integrar os dados de forma confiável, organizada e em nuvem permitindo maior flexibilidade da arquitetura.

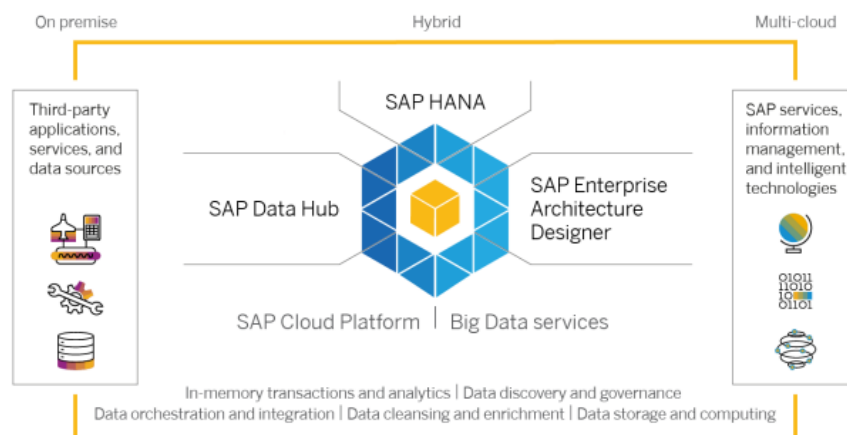


Imagem 2: SAP HANA Data Management Suit

O SAP Predictive Analytics é o aplicativo utilizado para desenvolvimento de modelos preditivos que permitem auxiliar na predição de comportamento, resultados futuros, servindo como base para orientar na tomada de decisões mais acertadas e lucrativas. A interface da solução é amigável, possibilitando o desenvolvimento de modelos produtivos sem a necessidade do conhecimento extenso de linguagens estatísticas. Também possui integração



com a linguagem R, possibilitando o uso mais avançado da ferramenta por estatístico e cientista de dados.

Para o problema mapeado é recomendado, além da aquisição das ferramentas, a adoção de modelos estatístico de segmentação comportamental dos clientes inadimplentes, possibilitando direcionar melhor as ações de cobrança, podendo-se definir ações mais brandas como SMS / URA para clientes com menor risco de inadimplência e ações mais enérgicas, como a suspensão de fornecimento. Desta forma, é possível gerar economia direcionando melhor as ações. Evita-se gerar despesa com atuação agressiva nos níveis mais brandos de risco, direcionando melhor os recursos, consequentemente, melhorando o resultado.

Complementar as ações, pode-se também efetuar o estudo para mapear o perfil de cliente que responde a uma determinada ação de cobrança, possibilitando definir melhor o direcionamento das atuações de cobrança e estudar novas estratégias de negócio para os casos identificados de consumidores que não respondem a qualquer ação de cobrança.

## 4.2. Etapas de Desenvolvimento

No mercado há três metodologias de desenvolvimento de modelos preditivos que são muito utilizadas a CRISP-DM (Cross Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model e Assess) e KDD (Knowledge Discovery in Databases). CRISP-DM descreve abordagens comumente usadas por especialistas em mineração de dados, SEMMA é a metodologia desenvolvida pela empresa SAS que foca principalmente nas tarefas de criação do modelo, deixando as questões de negócio de fora, por fim, KDD que foca em questões de negócio ou geração de modelos, mas sim na descoberta de conhecimentos a partir dos dados.

A Coelba possui estatísticos no quadro de negócio com conhecimento e experiência no emprego da metodologia CRISP-DM. O método é constituído de 6 etapas essenciais e será adotado na solução proposta.

### 4.2.1. Entendimento do Negócio

Nesta fase deverá se entender as características do negócio e ser definido os objetivos do projeto visando solucionar os problemas do negócio. Conforme Fernando Amaral (2016, pg.85) esta “é uma etapa crucial da qual o sucesso de implementação depende diretamente”. É recomendável envolver todos da equipe, em especial, os líderes.

### 4.2.2. Entendimento dos Dados

Uma vez entendido o negócio, deve-se entender os dados necessários à mineração, devendo ser avaliados, estruturados, analisados os relacionamentos, qualidade, quantidade e acesso. Nesta fase, as métricas devem ser mapeadas e documentadas facilitando para a necessidade de modificações futuras.

### 4.2.3. Preparação dos Dados

Para o processo de modelagem de dados é necessário dados organizados, limpos e selecionados. Nesta etapa, trabalha-se também com outras tarefas como a conversão dos dados numéricos e nominais.

#### 4.2.4. Modelagem

Esta fase é a construção dos modelos propriamente dito, é a aplicação de técnicas modelagem diferenciadas de acordo com o problema a ser desenvolvido. Nesta etapa constrói-se vários modelos, avaliando o resultado apresentado dentro de parâmetros aceitáveis.

#### 4.2.5. Avaliação

Avalia-se os modelos construído, compara-se o resultado caso tenham-se desenvolvidos mais de uma. É necessário analisar o desempenho está dentro dos parâmetros aceitáveis, caso positivo, segue-se para a próxima etapa, caso negativo, reinicia-se as etapas reavaliando os critérios adotadas na fase inicial.

#### 4.2.6. Implementação

Fase que se coloca o modelo em produção, para que se possa ser usado nas operações das empresas.

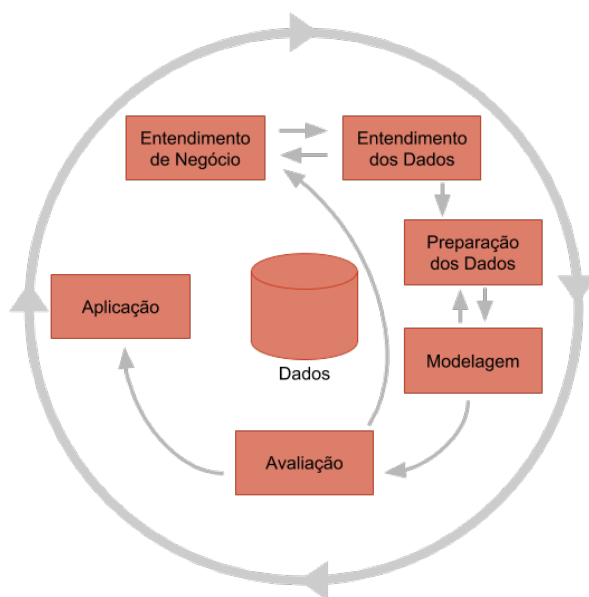


Imagem 3: Fases do modelo CRISP-DM

#### 4.3. Validação e Análise de Resultados

Para validação do resultado, compara-se o resultado dos valores obtidos e com os valores previstos pelo modelo obtendo a taxa de assertividade, neste caso aplica-se a matriz de confusão que irá alinhar as duas variáveis: valor previsto e valor observado. Com base nos resultados, também se apura as métricas de desempenho e predição, são elas: Acurácia, Sensibilidade e Especificidade.

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	<b>VP</b> Verdadeiro Positivo	<b>FP</b> Falso Positivo
	negativos	<b>FN</b> Falso Negativo	<b>VN</b> Verdadeiro Negativo

Imagem 4: Matriz de confusão para modelo de aplicação

Os resultados apresentados indicaram a efetividade do estudo realizados, caso o retorno seja negativo, o mesmo deverá ser reavaliado para alterações no processo, retornando as etapas anteriores da metodologia de desenvolvimento. O processo de validação e análise de resultado deverá ser aplicado periodicamente, revisando se o modelo está perdendo a aderência, quando necessário, novos estudos deverão ser realizados para garantir o retorno financeiro esperado pela empresa.

## 5. Benefícios e Resultados Esperados

Com a análise do Big Data e a adoção de modelos estatísticos de segmentação os clientes por risco de inadimplência, a Coelba poderá otimizar os recursos empregados nas ações de cobrança direcionando a ação mais adequada de acordo com o nível de risco de cada consumidor, dando maior inteligência ao processo. Com isso, a empresa poderá aumentar o alcance da cobrança aumento o sucesso na recuperação, antecipando o aumento da receita, consequentemente, poderá melhorar a imagem da empresa com os consumidores, pois haverá mais tolerância com os clientes que atrasam mais pagam.

Paralelamente, a empresa ganhará mais conhecimento do perfil de seus clientes, que poderá ser revertido em novas estratégias de cobrança, direcionamento da publicidade, desenvolvimento de novas ações, criando valor para a empresa. Os ganhos poderão ser maiores a depender do uso que a empresa fará com conhecimento obtido, conforme Pinheiro (2008, pg.14) “um importante passo para a criação de valor para as empresas é gerir de forma efetiva o conhecimento gerado por processos analíticos e de inteligência de negócio”.

## 6. Bibliografia

Pinheiro, Carlos André Reis. Inteligência Analítica – Mineração de Dados e Descoberta de Conhecimento. Rio de Janeiro: Editora Ciência Moderna Ltda., 2008.

Amaral, Fernando. Introdução a ciência de dados: mineração de dados e Big Data. Rio de Janeiro: Alta Books, 2016.

Release – Companhia de Eletricidade do Estado da Bahia  
<<http://bvmf.bmfbovespa.com.br/cias-listadas/empresas-listadas/ResumoDemonstrativosFinanceiros.aspx?codigoCvm=14524&idioma=pt-br>> Acessado em 19/05/19.

Demonstrações Financeiras Padronizadas - 31/12/2018 - V1. Relatório da Administração / Comentário do Desempenho  
<<https://www.rad.cvm.gov.br/ENETCONSULTA/frmGerenciaPaginaFRE.aspx?NumeroSequencialDocumento=80683&CodigoTipoInstituicao=2>> Acessado em 19/05/19.

Política Tarifária: Perdas e Inadimplência <[http://www.acendebrasil.com.br/br/sala-de-imprensa/pol%C3%ADtica\\_tarifaria\\_perdas\\_e\\_inadimplencia](http://www.acendebrasil.com.br/br/sala-de-imprensa/pol%C3%ADtica_tarifaria_perdas_e_inadimplencia)> Acessado em 19/05/19.

A Tarifa de Energia Elétrica < <http://www.aneel.gov.br/tarifas>> Acessado em 19/05/19.

Por dentro da Conta de Luz. Informação de utilidade pública.  
<[http://www2.aneel.gov.br/arquivos/PDF/Cartilha\\_1p\\_atual.pdf](http://www2.aneel.gov.br/arquivos/PDF/Cartilha_1p_atual.pdf)> Acessado em 19/05/19.

A economia brasileira em 2019: nova decepção?  
<<https://www.infomoney.com.br/blogs/economia-e-politica/um-brasil/post/8172785/a-economia-brasileira-em-2019-nova-decepcao>> Acessado em 19/05/19.

Gerenciamento de Dados < <https://www.sap.com/brazil/products/technology-platforms/data-management.html>> Acessado em 23/05/19.

CRISP-DM, SEMMA e KDD: conheça as melhores técnicas para exploração de dados  
<<https://paulovasconcellos.com.br/crisp-dm-semma-e-kdd-conheça-as-melhores-técnicas-para-exploração-de-dados-560d294547d2>> Acessado em 23/05/19.

Ciclo de vida dos dados <<https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4>> Acessado em 23/05/19.

Predição <<http://www.portalaction.com.br/analise-de-regressao/45-predicao>> Acessado em 23/05/19.