

CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA
UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA
MESTRADO PROFISSIONAL EM GESTÃO E TECNOLOGIA EM
SISTEMAS PRODUTIVOS

REGINALDO PEREIRA DA SILVA

APLICAÇÃO DE *MACHINE LEARNING* PARA ANÁLISE DO *MARKET SHARE* E
PREVISÃO DE DEMANDA DE EQUIPAMENTOS IMPORTADOS PARA O MERCADO
DE SAÚDE

São Paulo

Junho/2020

REGINALDO PEREIRA DA SILVA

APLICAÇÃO DE *MACHINE LEARNING* PARA ANÁLISE DO *MARKET SHARE* E
PREVISÃO DE DEMANDA DE EQUIPAMENTOS IMPORTADOS PARA O MERCADO
DE SAÚDE

Dissertação apresentada como exigência parcial para a obtenção do título de Mestre em Gestão e Tecnologia em Sistemas Produtivos do Centro Estadual de Educação Tecnológica Paula Souza, no Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos, sob a orientação do Prof. Dr. Napoleão Verardi Galegale.

São Paulo
Junho/2020

S586a Silva, Reginaldo Pereira da
Aplicação de Machine Learning para análise do Market Share e
previsão de demanda de equipamentos importados para o mercado
de saúde / Reginaldo Pereira da Silva. – São Paulo: CPS, 2020.
90 f. : il.

Orientador : Prof. Dr. Napoleão Verardi Galeale
Dissertação (Mestrado Profissional em Gestão e Tecnologia em
Sistemas Produtivos). – Centro Estadual de Educação Tecnológica
Paula Souza, 2020.

1. Equipamentos Médico-Hospitalares. 2. Dados
Governamentais Abertos. 3. Market Share. 4. Machine Learning. 5.
Sistemas Produtivos. I. Galeale, Napoleão Verardi. II. Centro
Estadual de Educação Tecnológica Paula Souza. III. Título.

REGINALDO PEREIRA DA SILVA

APLICAÇÃO DE *MACHINE LEARNING* PARA ANÁLISE DO *MARKET SHARE* E
PREVISÃO DE DEMANDA DE EQUIPAMENTOS IMPORTADOS PARA O MERCADO
DE SAÚDE

Prof. Dr. Napoleão Verardi Galeale.

Prof. Dr. Adilson Carlos Yoshikuni

Prof. Dr. Marcelo Duduchi Feitosa

São Paulo, 1 de junho de 2020

A todos os estrategistas que buscam de maneira incessante, principalmente por meio das novas tecnologias, vantagens competitivas para seus modelos de negócios.

AGRADECIMENTO À 2BW ANALYTICS INTELIGÊNCIA DE NEGÓCIO

O presente trabalho foi realizado com o apoio da **2BW Analytics Inteligência de Negócios Ltda.** nas pessoas dos sócios Rafael Alfieri e Eduardo Leão, a qual destinou recursos para o desenvolvimento do protótipo da aplicação.

AGRADECIMENTOS

Agradeço primeiramente a Deus, em segundo à toda a minha família pelo apoio, exemplo de conduta, perseverança, resiliência, dedicação e empenho, aplicados durante o período em que estive trabalhando nesse projeto muito importante para o desenvolvimento educacional e pessoal.

Também a todos os professores com os quais tive a oportunidade de interagir durante o mestrado e especialmente ao meu orientador, Prof. Dr. Napoleão Verardi Galeale, por todo o apoio e incentivo na orientação da dissertação, para que fosse realizada com a máxima excelência e qualidade.

Ao Centro Paula Souza pela oportunidade e todo o apoio dos profissionais, principalmente da Débora A.C.P. Ricci, sempre disponível para ajudar.

Finalmente, a todos que, diretamente ou indiretamente, acreditaram, colaboraram e incentivaram durante o desenvolvimento dessa dissertação.

“Innovation is the central issue in economic prosperity. The best CEO's I know are teachers, and at the core of what they teach is strategy. The essence of strategy is choosing what not to do. The company without a strategy is willing to try anything.”

Michael E. Porter

RESUMO

SILVA, R. P. **Aplicação de *Machine Learning* para análise do *Market Share* e previsão de demanda de equipamentos importados para o mercado de saúde.** 91 f. Dissertação (Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2020.

Entender o comportamento econômico de um mercado ou de um segmento específico requer inúmeros processos. Estes processos abrangem desde a captura de cada fonte de dados, as normatizações tecnológicas que percorrem o formato do dado de sua origem até sua consolidação, sua homogeneização e ainda o enriquecimento que possa transformá-lo em informação de mercado e aferir a demanda e níveis adequados de inventários. O presente trabalho tem por objetivo desenvolver um protótipo e apresentá-lo como uma solução computacional com base em algoritmos de *Machine Learning*, que permita determinar o *Market Share* das indústrias de equipamentos importados para o mercado de saúde no Brasil e também melhorar a gestão da demanda e inventário destes sistemas produtivos. A natureza do objeto dessa pesquisa é de *Design Science Research* (DSR) para produzir o referido protótipo, que foi desenvolvido com base nos requisitos levantados por meio de um estudo de campo com empresas importadoras de equipamentos médico-hospitalares. A base de dados foi obtida por meio dos dados governamentais abertos (DGA) sobre de importação de produtos, disponíveis no website da Receita Federal Brasileira, utilizando recursos de *Robotic Process Automation* (RPA). O protótipo apresentou como resultados: 90% de acurácia no cálculo de *Market Share*, e um fator R^2 na regressão linear simples de 0,97 na previsão de demanda para um crescimento de mercado estimado de 50%.

Palavras-chave: Equipamentos Médico-Hospitalares, Dados Governamentais Abertos; *Market Share*; *Machine Learning*; *Data Analytics*; *Robotic Process Automation*; Sistemas Produtivos.

ABSTRACT

SILVA, R. P. **Machine Learning application to analyze Market Share and demand forecasting for the imported equipment to the healthcare market.** 91 p. Dissertation (Professional master's degree in management and Technology in Productive Systems). Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2020.

Understand the market economic behaviors of a market or a specific segment requires innumerable processes, which goes from each source of data capture, technological standardization. From the data format on the origin to the consolidation itself, its homogenization and even in the enrichment that can transform it into market information, as well as gauge the demand and adequate levels of inventories. The current work aims to develop a prototype and present it as a computational solution based on Machine Learning algorithms, which allows determining the Market Share of the equipment industries imported to the health market in Brazil as well as improving demand management and inventory of these productive systems. The nature of the object of this research is Design Science Research (DSR) to produce the referred prototype, which was developed based on the requirements raised through a field study with companies that import hospital medical equipment. The database was obtained through open government data (DGA) on product imports, available on the website of the Brazilian government website, using Robotic Process Automation (RPA) resources. The prototype presented as results: 90% accuracy in the calculation of Market Share, and a R^2 factor in the simple linear regression of 0.97 in the demand forecast, for an estimated market growth of 50%.

Keywords: Medical Products; Open Government Data; Market Share; Machine Learning; Data Analytics; Robotic Process Automation; Productive Systems.

LISTA DE QUADROS

Quadro 1:	Oito princípios de dados governamentais abertos	25
Quadro 2:	Formato da NCM	27
Quadro 3:	Formato da NCM com exemplo real	27
Quadro 4:	Perfil das pessoas entrevistadas	51
Quadro 5:	Principais concorrentes do protótipo e suas características	52
Quadro 6:	Diferenciais do protótipo comparados com os principais concorrentes	53
Quadro 7:	Dados da Tabela de Importações da RFB	56
Quadro 8:	Dados da Tabela de Palavras-chaves	57
Quadro 9:	Exemplo: algumas colunas do arquivo da RFB	63
Quadro 10:	Avaliação dos usuários chaves do protótipo	64
Quadro 11:	Tipos de Algoritmos de <i>Machine Learning</i>	91

LISTA DE TABELAS

Tabela 1:	Tabela 1 – Resultados (Outputs) de Pesquisa Design Science (DSR)	49
Tabela 2:	Tabela 2 – Dados da Dissertação utilizando DSR	49

LISTA DE FIGURAS

Figura 1:	Mapeamento da Cadeia de Valores Equipamentos Importados	18
Figura 2:	Diagrama da Representação Gráfica da Solução Computacional	20
Figura 3:	Estrutura da Dissertação	21
Figura 4:	Estrutura do Referencial Teórico	22
Figura 5:	Tela do site da RFB onde são extraídos os dados de importações por NCM .	26
Figura 6:	Exemplo de composição da NCM	28
Figura 7:	A Evolução do <i>Machine Learning</i> e IA	33
Figura 8:	O típico ciclo de <i>Machine Learning</i> (ML)	34
Figura 9:	Automação de Inteligência = RPA + ML	35
Figura 10:	O Processo de <i>Machine Learning</i> (ML)	36
Figura 11:	Regressão Linear Simples	39
Figura 12:	O Processo de <i>Big Data Analytics</i>	42
Figura 13:	Modelo de Processo de Pesquisa em Design Science (DSR)	46
Figura 14:	Fases dos Processos e Tecnologias da Solução Computacional	54
Figura 15:	Componentes da Solução Computacional	56
Figura 16:	Exemplo da Tabela de Palavras-chaves preenchida	58
Figura 17:	Exemplo da importação e transformação do arquivo da RFB com PDI	58
Figura 18:	Exemplo da Biblioteca de <i>Machine Learning</i> Python	60
Figura 19:	Demonstração preliminar da Solução Computacional	62
Figura 20:	FuncEx, concorrente da Solução Computacional	74
Figura 21:	Datagro, concorrente da Solução Computacional	75
Figura 22:	DataImport, concorrente da Solução Computacional	76

LISTA DE SIGLAS

BCG	<i>Boston Consulting Group</i>
DGA	Dados Governamentais Abertos
DSR	<i>Design Science Research</i>
DTT	<i>Deloitte Touche Tohmatsu</i>
ETL	<i>Extract/Transform/Load</i>
HBS	<i>Harvard Business School</i>
IA	Inteligência Artificial
ML	<i>Machine Learning</i>
NCM	Nomenclatura Comum do Mercosul
PDI	<i>Pentaho Data Integration</i>
PIMS	<i>Profit Impact of Market Strategies</i>
RFB	Receita Federal Brasileira
RPA	<i>Robotic Process Automation</i>
SI	Sistemas de Informações
SQL	<i>Structured Query Language</i>
TIC	Tecnologias de Informação e Comunicação

SUMÁRIO

INTRODUÇÃO	16
Questão de Pesquisa	18
Objetivos	18
Estrutura da Dissertação	20
1 FUNDAMENTAÇÃO TEÓRICA	22
1.1 Mercado de Equipamentos Médico-hospitalares	22
1.2 Dados de Produtos importados da Receita Federal Brasileira RFB	24
1.2.1 <i>Dados Governamentais abertos</i>	24
1.2.2 <i>Como obter os dados de importações da RFB</i>	25
1.2.3 <i>NCM – Nomenclatura Comum do Mercosul</i>	26
1.3 Market Share	28
1.4 Machine Learning (ML)	30
1.4.1 <i>Machine Learning – Aprendizado Supervisionado</i>	37
1.4.2 <i>Regressão Linear Simples</i>	38
1.4.3 <i>Análises e classificação de palavras-chaves</i>	40
1.5 Data Analytics e Big Data	41
1.6 RPA – Robotic Process Automation	43
2 METODOLOGIA	45
3 RESULTADOS E DISCUSSÃO	51
3.1 Protótipo da Solução Computacional	53
3.2 Componentes da Solução Computacional	55
3.2.1 <i>Exemplos das tabelas utilizadas na Solução Computacional</i>	56
3.2.2 <i>Ferramenta de ETL Open Source – Pentaho Data Integration ou Kettle</i>	58
3.2.3 <i>Biblioteca de Machine Learning Python</i>	59
3.2.4 <i>Solução Computacional em perspectiva</i>	62
CONCLUSÃO	66
REFERÊNCIAS	68
APÊNDICES	73
<i>Apêndice A – Roteiro de Entrevistas para determinar usabilidade e aplicação do protótipo</i>	73
<i>Apêndice B – Outras informações dos principais concorrentes do protótipo</i>	74
<i>Apêndice C – Avaliação das áreas usuárias do protótipo</i>	77
<i>Apêndice D – Propriedades e código fonte dos componentes do protótipo</i>	79
<i>Apêndice E – Tipos de Algoritmos de Machine Learning</i>	79

INTRODUÇÃO

Diversos segmentos de mercado, onde se incluem também as indústrias de equipamentos para a saúde, têm crescido de forma exponencial, ano após ano, nos mercados emergentes, como o Brasil. O principal foco desse mercado tem sido a priorização da modernização por meio de tecnologias como *Machine Learning*, Robótica, Inteligência artificial, entre outras, como aponta um estudo da consultoria Deloitte (*Global Healthcare Outlook*, 2020).

Entender a participação de mercado (*Market Share*) das empresas que comercializam equipamentos médicos no Brasil é fundamental para compreender o posicionamento dos concorrentes e com isso criar vantagens competitivas. Planejar de maneira mais assertiva a demanda e a necessidade de estoque desses equipamentos, de alto valor agregado, é fundamental para as empresas desse setor, além de um diferencial devido ao alto (*lead-time*) tempo de importação. Tendo como objetivo o ganho de *Market Share*, é necessário avaliar a relação dessa métrica com métricas de rentabilidade, já que os custos necessários para o aumento desse indicador podem ser altos a ponto de não compensar os investimentos (KOTLER; KELLER, 2012; FARRIS, 2013).

De maneira objetiva, a medida de *Market Share* representa a participação de uma empresa em relação aos seus concorrentes. Principalmente a partir de 1970, a medida passou a ter sua importância reconhecida e investigada em estudos acadêmicos, e tem se mostrado como uma métrica relevante também em estudos mais recentes (BRONNENBERG et al., 2000; SRINIVASAN; LESZCZYC e BASS, 2000; WILBUR e FARRIS, 2014). Nesse contexto, é importante colocar que a década de 1970 marcou o início do projeto Profit Impact of Market Strategies (PIMS), apoiado pelo Marketing Science Institute (MSI) e pela Harvard Business School (HBS). Buzzell, Gale e Sultan (1975) identificaram o *Market Share* como um dos fatores-chave de influência na lucratividade. As atenções começaram, portanto, a ser direcionadas para a medida de *Market Share*, cuja importância foi impulsionada, também, com a proposta da matriz Boston Consulting Group (BCG), proposta por Henderson na década de 1970 (HENDERSON, 1970, 1973).

A partir de então, os estudos têm buscado identificar a relação do *Market Share* com a métrica como satisfação e retenção de clientes (RUST e ZAHORIK, 1993), bem como os efeitos de decisões de marketing, como promoção, propaganda e distribuição no *Market Share* (ATAMAN, HEERDE, e MELA, 2010; BRONNENBERG et al., 2000; SRINIVASAN et al., 2000). Por um lado, as pesquisas têm motivado o entendimento do impacto do *Market Share*

em outras medidas não financeiras e, por outro lado, têm buscado compreender a relação entre os estímulos de marketing, vantagens competitivas, e o seu impacto no *Market Share*, bem como no planejamento da cadeia de suprimentos.

A participação de mercado dos equipamentos médico-hospitalares onde aproximadamente 75% dos produtos são importados (ANAHP, 2015), pode ser calculada por meio dos dados de importação, disponibilizados pela RFB (Receita Federal Brasileira) em seu website. Esses dados são considerados dados governamentais abertos.

Conforme Chopra e Meindl (2001), utilizando os dados de importação pode-se calcular a participação no mercado, gerando informações que são essenciais para que se possa tomar boas decisões. A tecnologia da informação proporciona ferramentas para reunir essas informações e analisá-las com o objetivo de tomar melhores decisões, referentes à cadeia de suprimentos, para que não haja uma demanda maior do que a disponibilidade de produtos, evitando assim a insatisfação dos clientes.

Segundo Sales (2005), a atividade de dimensionamento de estoques é responsável pela determinação da quantidade de produtos que devem ser mantidos em estoque e pelo controle de seus níveis, para que não haja excessos nem insuficiências. Tanto excessos quanto insuficiências de estoque prejudicam o fluxo de materiais da cadeia de suprimentos, resultando no aumento de custos e redução nos níveis de serviço.

A otimização dos estoques evita a superlotação das áreas de estocagem, facilitando o manuseio, a procura e localização dos itens e evitando o comprometimento dos níveis de qualidade; uma vez que a variação de produtos e quantidade do estoque estarão de acordo com as necessidades dos clientes, gerando um alto giro dos produtos (SALES, 2005).

Na Figura 1, observa-se o mapeamento dos processos da cadeia de suprimentos e sua importância para a satisfação dos clientes. O mapeamento alcança desde o conhecimento das informações do mercado, os ajustes à demanda em virtude de projeções de crescimento de participação de mercado, até o planejamento dos inventários por meio do processo de importação dos produtos, evitando a falta de produtos, que prejudicaria bastante a satisfação dos clientes, e também evitando o excesso, pois os equipamentos médico-hospitalares são produtos de alto valor agregado e de custos bastante elevados para a sua manutenção em estoque.

Figura 1 – Mapeamento da Cadeia de Valores Equipamentos Importados



Fonte: Resultado da Pesquisa

Questão de Pesquisa

Esse estudo procura responder a seguinte questão de pesquisa: Como, por meio de um protótipo de solução computacional, baseada em *Machine Learning*, pode-se calcular o *Market Share* de equipamentos importados para o mercado de saúde brasileiro, visando melhorar a gestão de demanda e os níveis de inventário?

Objetivos

Objetivo geral

O objetivo geral dessa pesquisa é desenvolver um protótipo de uma solução computacional, baseada em *Machine Learning*, que permita determinar o *Market Share* das empresas de equipamentos médicos hospitalares importados para o mercado de saúde no Brasil, por meio dos requisitos coletados por estudo de campo e entrevistas com potenciais usuários da solução, visando melhorar a gestão de demanda e os níveis de inventário, de acordo com o crescimento esperado de mercado.

Objetivos Específicos

Como desdobramento do objetivo geral e meio de responder à questão de pesquisa, foram definidos os seguintes objetivos específicos:

a) Realizar entrevistas com duas empresas importadoras, sendo uma de equipamentos médico-hospitalares, para identificação de requisitos para desenvolvimento de um protótipo de solução computacional que possa determinar o *Market Share* de equipamentos importados no mercado de saúde brasileiro, e outra importadora de produtos em geral, para verificar as semelhanças, para que se possa criar uma plataforma mais abrangente no futuro, permitindo expandir para outros segmentos de mercado.

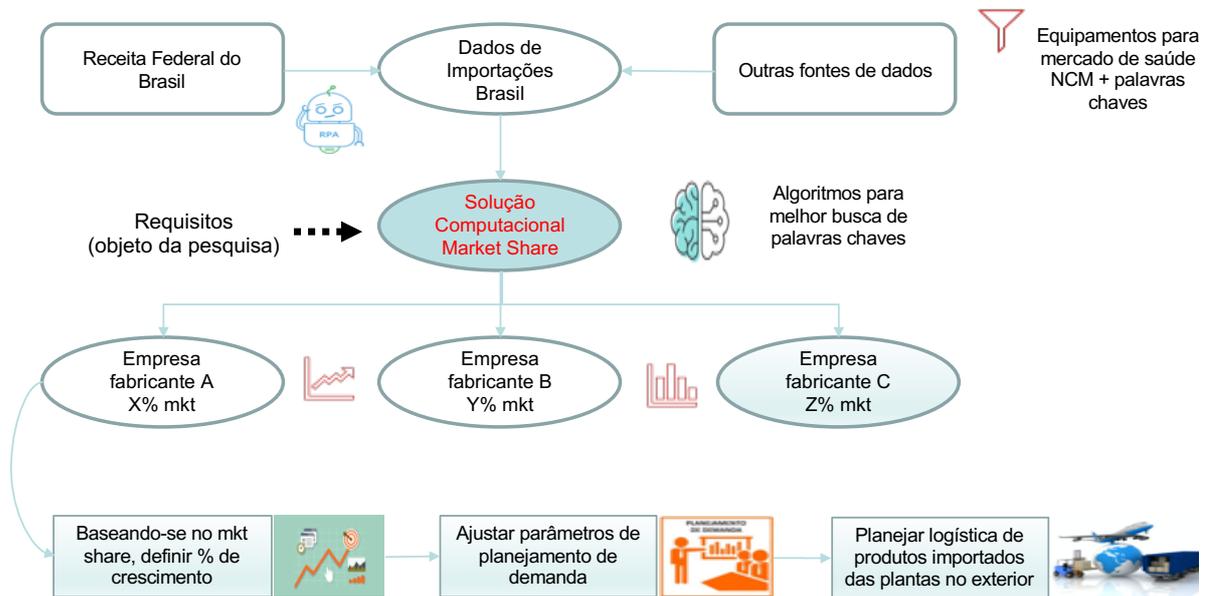
b) Desenvolver um protótipo de uma solução computacional, baseada em *Machine Learning*, que possa analisar as fontes de dados de importações feitas pelo Brasil, principalmente analisando os dados da RFB (Receita Federal do Brasil), com o objetivo de apresentar o *Market Share* das indústrias de equipamentos importados, gerando a possibilidade de análise dos dados para conhecimento e atuação dos concorrentes, e melhorar a gestão de demanda e os níveis de inventário.

c) Testar e avaliar os resultados, por meio da disponibilização do protótipo para a utilização em duas unidades de negócios de uma indústria de importação de equipamentos médico-hospitalares.

Contribuição da pesquisa

A contribuição dessa pesquisa é demonstrada por meio da representação gráfica apresentada na Figura 2, onde o objetivo principal é a disponibilização dos dados de *Market Share* das indústrias fabricantes de equipamentos para saúde no mercado brasileiro, para que seja efetuada a análise de crescimento de mercado, melhoria da gestão de demanda e níveis de inventários.

Figura 2 – Diagrama da Representação Gráfica do Protótipo



Fonte: Resultado da Pesquisa

As contribuições dessa pesquisa são apresentadas de duas maneiras: uma contribuição na área teórica, apresentando-se conhecimento acadêmico e de pesquisas recentes, sobre a utilização de tecnologias, como inteligência artificial; *Machine Learning*; *Robotic Process Automation* e *Big Data Analytics*, aplicados principalmente à área da saúde, foco principal dessa pesquisa. Na área prática, é apresentado um protótipo de uma solução computacional baseada em *Machine Learning*, que propõe o cálculo de *Market Share* de equipamentos médico-hospitalares e, utilizando-se de algoritmos de regressão linear e com simulações de crescimento e ganho de mercado, calcular o nível ideal de inventários para suprir a demanda.

Como parte das contribuições, foi efetuada uma pesquisa sobre produtos similares, que seriam concorrentes da solução computacional (ver apêndice B, onde são apresentados os três principais concorrentes).

Estrutura da Dissertação

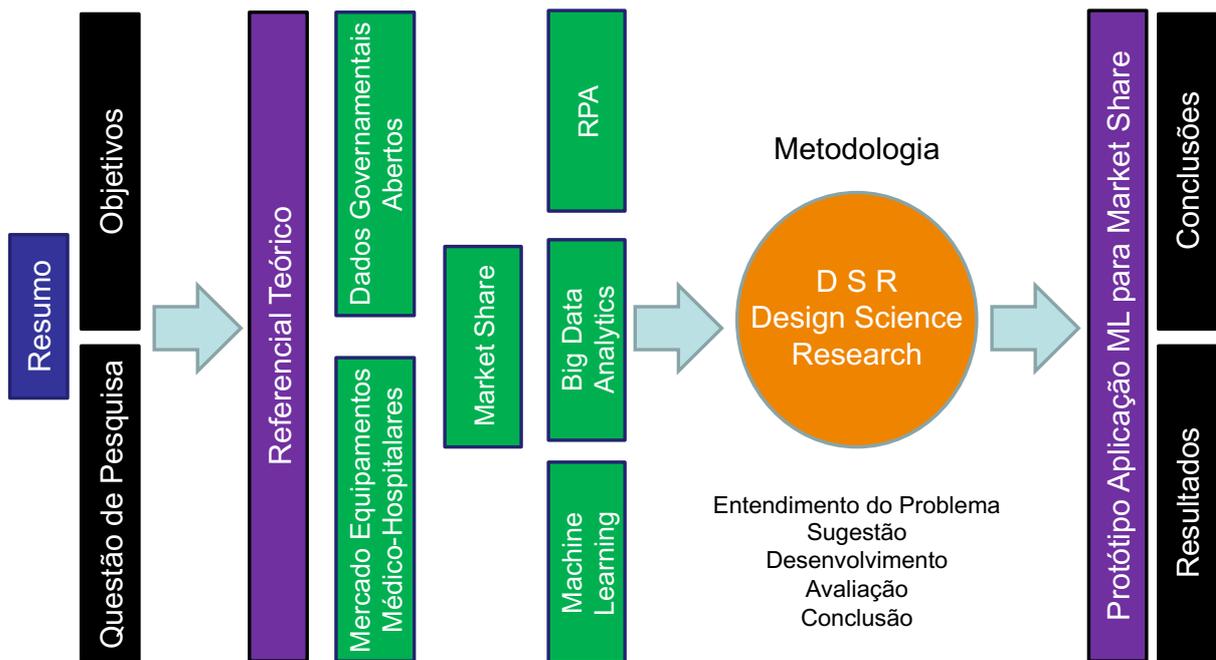
A estrutura da dissertação é demonstrada na Figura 3, onde serão apresentados os principais pontos que compõem a dissertação. Dessa maneira, temos, inicialmente: Resumo / Abstract, Questão de Pesquisa e Objetivos.

O Referencial Teórico, com os temas chaves da pesquisa: Mercado de Equipamentos Médico-Hospitalares, Dados Governamentais Abertos, *Market Share*, *Machine Learning*, *Big Data Analytics* e *Robotic Process Automation*.

A Metodologia aplicada na pesquisa foi *Design Science Research* (DSR), onde o principal objetivo é o desenvolvimento de um protótipo para resolver um problema de negócios. O protótipo desenvolvido foi uma Aplicação de *Machine Learning* para cálculo e análise do *Market Share* e previsão de demanda de equipamentos importados para o mercado de saúde.

Finalmente, serão apresentados os resultados obtidos por meio de avaliações realizadas com os usuários chaves do protótipo, bem como as conclusões com as limitações da pesquisa e sugestões para trabalhos futuros.

Figura 3 – Estrutura da Dissertação

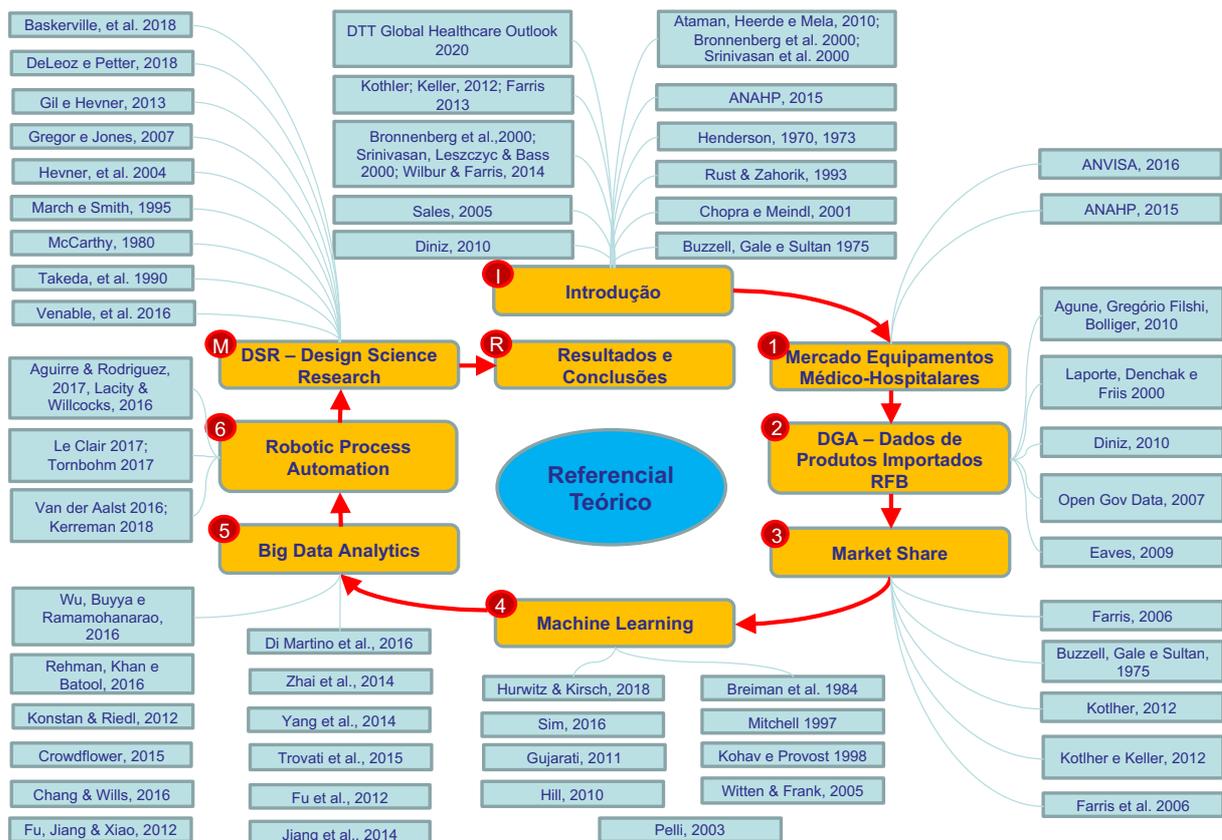


Fonte: Resultado da Pesquisa

1 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica está sendo realizada de acordo com os aspectos relevantes do projeto de pesquisa. A Figura 4 demonstra a estrutura do Referencial Teórico, com os autores mencionados nesse trabalho de pesquisa:

Figura 4 – Estrutura do Referencial Teórico



Fonte: Resultado da Pesquisa

O Referencial teórico foi elaborado por meio de pesquisas bibliométricas, e foram selecionados vários artigos relevantes, de acordo com o foco da pesquisa. Foi realizada a leitura de inúmeros resumos e foram selecionados os artigos de maior relevância e alinhados com a linha de pesquisa, efetuando-se uma contribuição teórica nas áreas de mercado de saúde, *Market Share*, *machine learnig*, *big data e analytics* e *robotic process automation*.

1.1 Mercado de Equipamentos Médico-hospitalares

Os equipamentos médico-hospitalares são essenciais para o diagnóstico, tratamento, prevenção e reabilitação da saúde, viabilizando a oferta de serviços de qualidade à vida do usuário, atentando às exigências da ANVISA no que diz respeito à qualidade e segurança (BRASIL, 2016).

As indústrias que compõem o setor de materiais possuem um elevado grau de conhecimento científico e inovação tecnológica, garantindo o desenvolvimento e aperfeiçoamento de produtos e, dessa forma, possuem significativa importância no suprimento de artigos tecnológicos voltados à saúde.

Os Estados Unidos concentram mais da metade (56,0%) dos produtores de dispositivos médicos. A Europa e o Japão também têm participação importante no mercado, ambos com 18,0% dos fabricantes, gerando um ambiente altamente dinâmico e competitivo, o que exige das empresas investimento na inovação (ANAHP, 2015). Nesse alinhamento reflexivo, as empresas brasileiras se encontram em maior nível de dificuldade, pois apresentam menos incentivo ao desenvolvimento tecnológico quando comparadas com as grandes empresas estrangeiras do setor. Esses fatores interferem no valor do produto, pois o material estrangeiro absorve grandes taxas tributárias, além de estar vinculado a fatores como a política da marca, concorrência e transparência no mercado (ANAHP, 2015). Entende-se, frente a essa constatação, como o processo de aquisição de equipamentos pelos serviços de saúde brasileiros é burocrático, funcionando por meio de importações, e de alto custo, devido à ausência de política de preço que regule essa modalidade. Trata-se, portanto, de uma área complexa, pois envolve recursos financeiros relativamente altos e interesses relacionados à indústria e aos participantes do serviço de saúde.

O mercado de equipamentos médicos mantém contato direto com as instituições de saúde públicas e privadas. Todas as instituições necessitam de materiais médicos para suprir as demandas de trabalho e, nesse contexto, é importante ressaltar que o profissional da organização pública e privada, que está à frente das negociações, tem como principais atribuições, no desempenho do trabalho, a imparcialidade, a ética e a responsabilidade, visando sempre fazer mais e melhor com os recursos disponíveis.

1.2 Dados de produtos importados da Receita Federal Brasileira (RFB)

Para o cálculo da participação de mercado, serão utilizados os dados de importações, que são disponibilizados mensalmente no site da RFB. Esses dados são considerados dados governamentais abertos (DGA), de domínio público e de livre utilização pela sociedade.

Para Diniz (2010), a disponibilização de dados governamentais abertos permite que as informações sejam utilizadas da maneira e conveniência do interessado, de tal forma que elas possam ser misturadas e combinadas para agregar mais valor aos dados. Para ele, o objetivo de que as informações públicas sejam disponibilizadas segundo as regras dos dados abertos é superar as limitações existentes para que usuários de informações do serviço público possam facilmente encontrar, acessar, entender e utilizar os dados públicos segundo os seus interesses e conveniências.

1.2.1 Dados Governamentais Abertos

Dados governamentais abertos (DGA) ou governo aberto são termos utilizados mais recentemente para denominar a “disponibilização, por meio da internet, de informações e dados governamentais de domínio público para a livre utilização pela sociedade” (AGUNE; GREGÓRIO FILHO; BOLLIGER, 2010). Este conceito remete à defesa dos dados públicos como pertencentes ao cidadão, que deveria ter acesso irrestrito às informações governamentais, é defendido por vários autores, como Laporte, Demchak e Friis (2000), Agune, Gregório Filho e Bolliger (2010), Diniz (2010), entre outros. Também remete ao contexto do desenvolvimento recente das TICs (Tecnologias da Informação e Comunicação), onde a internet, ferramentas de bancos de dados, padrões e formatos abertos permitem e facilitam a disponibilização e acesso a essas informações.

Em 2007, um grupo de especialistas denominado OpenGovData desenvolveu os oito princípios dos dados governamentais abertos (OPEN GOV DATA, 2007). O Quadro 1 exibe os oito princípios que consideram os dados governamentais abertos.

Quadro 1 – Oito princípios de dados governamentais abertos

1. **Completo:** todos os dados públicos são disponibilizados. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou privilégio.
2. **Primário:** os dados são coletados na fonte, com o maior nível possível de granularidade, não em formas agregadas ou modificadas.
3. **Oportuno:** Os dados são disponibilizados o mais rápido possível para preservar o valor dos dados.
4. **Acessível:** os dados estão disponíveis para a maior variedade de usuários para a maior variedade de finalidades.
5. **Compreensíveis por máquina:** os dados são razoavelmente estruturados para permitir o processamento automatizado.
6. **Não discriminatório:** os dados estão disponíveis para qualquer pessoa, sem necessidade de registro.
7. **Não proprietário:** os dados estão disponíveis em um formato sobre o qual nenhuma entidade possui controle exclusivo.
8. **Livres de licença:** Os dados não estão sujeitos a nenhum regulamento de direitos autorais, patente, marca comercial ou segredo comercial. Restrições razoáveis de privacidade, segurança e privilégios podem ser permitidas.

Fonte: https://public.resource.org/8_principles.html adotado em dezembro de 2007.

Eaves (2009) apresentou as três leis dos dados governamentais abertos:

- a. Se o dado não for encontrado e indexado na web, ele não existe;
- b. Se não estiver aberto e disponível em formato compreensível por máquina;
- c. ele não pode ser aproveitado.

Se algum dispositivo legal não permitir sua replicação, ele é inútil.

Os benefícios da adoção dos DGA no campo da transparência e do controle social são, ao menos em tese, evidentes. A oferta de DGA tende a contribuir para o aumento da transparência do governo, criando melhores possibilidades de controle social das ações governamentais. No entanto, dada a relativa novidade do tema, ainda não se dispõe de pesquisas que demonstrem a totalidade desta possibilidade.

1.2.2 Como obter os dados de importações da Receita Federal Brasileira

Os dados de produtos importados estão disponíveis mensalmente na página de internet da RFB (Receita Federal Brasileira). Para fazer o download dos dados, basta seguir os seguintes passos:

- Acessar o link: <http://receita.economia.gov.br/dados/resultados/comercio-exterior>;
- Clicar em informações 2016 e selecionar o ano de referência, por exemplo: 2019;
- Para o foco da pesquisa, que é a indústria de equipamentos médico-hospitalares, baixar os seguintes NCMs: 27, 28, 29, 30, 32, 38, 39, 70, 82, 84, 85 e 90.

Figura 5 – Tela do site da RFB onde são extraídos os dados de importações por NCM

Menu - Sair

Seleção o Arquivo

Tipo Declaração: Importação Exportação

Ano Referência: 2019

Mês Ref Início: Janeiro

Mês Ref Fim: Outubro

Capítulo NCM: 85 Cap 85 - Máquinas, aparelhos e materiais elétricos

Baixar Arquivo CAPI85201901201910.zip

Últimos Arquivos Gerados

Importação Exportação

Capítulo 01 - Animais vivos	Out/2018	Nov/2018	Dez/2018	Fev/2019	Mar/2019	Abr/2019	Mai/2019	Jun/2019	Jul/2019	Ago/2019	Set/2019	Out/2019
Capítulo 02 - Carnes e miudezas, comestíveis	Nov/2018	Dez/2018	Jan/2019	Fev/2019	Mar/2019	Abr/2019	Mai/2019	Jun/2019	Jul/2019	Ago/2019	Set/2019	Out/2019
Capítulo 03 - Peixes e crustáceos, moluscos e outros invertebrados aquáticos	Nov/2018	Dez/2018	Jan/2019	Fev/2019	Mar/2019	Abr/2019	Mai/2019	Jun/2019	Jul/2019	Ago/2019	Set/2019	Out/2019
Capítulo 04 - Leite e laticínios; ovos de aves; mel natural; produtos comestíveis de origem animal, nã	Nov/2018	Dez/2018	Jan/2019	Fev/2019	Mar/2019	Abr/2019	Mai/2019	Jun/2019	Jul/2019	Ago/2019	Set/2019	Out/2019

Framework Demoiselle 2.5.0

Fonte: Website da Receita Federal Brasileira

Para cada NCM será criado um arquivo CSV com o seguinte nome: CAPI999999.csv, com dados dos processos de importação dos produtos classificados na NCM especificada. A Figura 5 exibe a tela da RFB onde são informados os parâmetros para que seja feito o download do arquivo com as informações especificadas nos mesmos. São parâmetros: Tipo de Declaração (Importação / Exportação), ano de referência, mês início e mês fim, e o capítulo da NCM, assunto que será abordado na próxima subseção.

1.2.3 NCM – Nomenclatura Comum do Mercosul

A NCM, Nomenclatura Comum do Mercosul, é uma metodologia adotada pelos países do Mercosul (Argentina, Brasil, Paraguai e Uruguai) em 1995, para permitir a identificação das mercadorias comercializadas no território sul-americano. O NCM foi baseado no Sistema Harmonizado de Designação e Codificação de Mercadorias (SH). O Sistema Harmonizado é adotado pela Organização Mundial das Alfândegas (OMA). Os seis primeiros dígitos do NCM são baseados no código SH. Isso permite que as mercadorias do Mercosul possam ser comercializadas nos 190 países que participam do acordo.

O método de identificação foi criado a partir do grau de participação do ser humano na produção da mercadoria, por isso o primeiro capítulo trata de animais vivos. Cada número ou par da NCM representa uma característica, conforme o Quadro 2, Formato da NCM:

Quadro 2 – Formato da NCM

Posição	Significado
2 primeiros dígitos	Capítulo
4 primeiros dígitos	Posição
6 primeiros dígitos	Sub-posição
7º dígito da NCM	Item
8º dígito da NCM	Subitem

Fonte: Dados da RFB

O NCM 3003.10.12, por exemplo, deve ser entendido conforme mostra o Quadro 3:

Quadro 3 – Formato da NCM com exemplo real

Posição	Significado
Capítulo 30	Produtos farmacêuticos
Posição 3003	Medicamentos (exceto os produtos das posições 30.02,030.05 ou 30.06) constituídos por produtos misturados entre si, preparados para fins terapêuticos ou

	profiláticos, mas não apresentados em doses nem acondicionados para venda a retalho;
Subposição 3003.10	Que contenham penicilinas ou seus derivados, com estrutura do ácido penicilânico, ou estreptomicinas ou seus derivados;
Item 3003.10.1	Que contenham penicilinas ou seus derivados, com estrutura do ácido penicilânico;
Subitem 3003.10.12	Amoxicilina ou seus sais

Fonte: Resultados da Pesquisa do Autor

Figura 6 – Exemplo de composição da NCM

30 Produtos farmacêuticos
3001 Glândulas e outros órgãos para usos opoterápicos, dessecados, mesmo em pó; extratos de glândulas ou de outros órgãos ou das suas secreções, para usos opoterápicos; heparina e seus sais; outras substâncias humanas ou animais preparadas para fins terapêuticos ou profiláticos, não especificadas nem compreendidas noutras posições.
3002 Sangue humano; sangue animal preparado para usos terapêuticos, profiláticos ou de diagnóstico; antisseros, outras frações do sangue e produtos imunológicos, mesmo modificados ou obtidos por via biotecnológica; vacinas, toxinas, culturas de microrganismos (exceto leveduras) e produtos semelhantes.
3003 Medicamentos (exceto os produtos das posições 30.02, 30.05 ou 30.06) constituídos por produtos misturados entre si, preparados para fins terapêuticos ou profiláticos, mas não apresentados em doses nem acondicionados para venda a retalho.
300310 Que contenham penicilinas ou seus derivados, com a estrutura do ácido penicilânico, ou estreptomicinas ou seus derivados
3003101 Que contenham penicilinas ou seus derivados, com a estrutura do ácido penicilânico
30031011 Ampicilina ou seus sais
30031012 Amoxicilina ou seus sais

exemplo-de-ncm

Fonte: Resultado da Pesquisa

A Figura 6 apresenta um exemplo real da composição da NCM. O exemplo apresentado refere-se a um produto farmacêutico, e descreve a composição desde o primeiro até o último nível.

1.3 *Market Share*

Sobre o seu cálculo, *Market Share* é a porcentagem de um mercado que uma empresa, marca ou produto possui em relação aos seus concorrentes. Esse cálculo pode ter tanto a perspectiva de volume de vendas quanto de faturamento, ou valor, que uma empresa possui em relação aos seus concorrentes em um mercado ou segmento específico (FARRIS et al., 2006).

Para Buzzell, Gale e Sultan (1975), o *Market Share* é identificado como um dos fatores-chave de influência na lucratividade. Em outras palavras, esse indicador dá um panorama geral

de como a empresa está situada no mercado, e também noções de como está a curva de aprendizado, da capacidade produtiva e do poder de distribuição das empresas.

De acordo com Kotler e Keller (2012), as grandes empresas, cada vez mais, não se limitam a analisar apenas a receita de vendas, passando a examinar também indicadores de participação de mercado. Além de estimar o potencial total e o potencial da área, é importante conhecer as vendas que ocorrem em seu mercado, identificando e monitorando o desempenho dos concorrentes.

Associações setoriais costumam levantar e publicar as vendas totais dos vários setores, porém não fornecem essas informações no menor nível às empresas. Com base nessas informações, cada empresa pode avaliar seu desempenho em relação ao setor como um todo. Por exemplo, caso as vendas de uma empresa cresçam 5% ao ano, enquanto as do setor crescem 10%, essa empresa está na verdade perdendo posição no setor.

Ademais, outra forma de se conseguir dados, é através da compra dessa informação por empresas especializadas, ou, no caso de nossa solução computacional, por meio de dados disponibilizados pelo governo (KOTLER; KELLER, 2012).

Segundo Kotler e Keller (2012), o líder de mercado detém a maior participação e geralmente lidera em variações de preços, lançamentos de produtos, cobertura de distribuição e intensidade promocional. Importante também destacar que a empresa deve monitorar não só *Market Share*, mas também os conceitos de *share of mind*, associado à questão: “Cite a primeira empresa que lhe vem à mente nesse setor”; e *share of heart*, ligado à questão: “Cite a empresa da qual você prefere comprar esse produto”.

Embora os três tipos citados sejam de extrema importância, o estudo proposto irá tratar apenas do ‘*Market Share*’ por se tratar de um indicador quantitativo, e, dessa forma, se torna possível apresentar de maneira precisa o objetivo da pesquisa.

Com isso, define-se então o *Market Share* como a porcentagem de um mercado que uma empresa, marca ou produto possui em relação aos seus concorrentes. Esse cálculo pode ter tanto a perspectiva de volume de vendas quanto de faturamento, ou valor, que uma empresa possui em relação aos seus concorrentes em um mercado ou segmento específico, conforme apresentado na Equação 1 (FARRIS et al., 2006). Assim,

$$\text{Market Share Volume (\%)} = \frac{\text{Vendas da Empresa para um produto ou marca (\#)}}{\text{Total de vendas da categoria do produto ou marca (\#)}} \quad (1)$$

E o respectivo cálculo em faturamento é dado pela Equação 2:

$$\text{Market Share Valor (\%)} = \frac{\text{Vendas da Empresa para um produto ou marca (\$)}}{\text{Total de vendas da categoria do produto ou marca (\$)}} \quad (2)$$

O *Market Share* é, portanto, um indicador que mostra a posição de uma empresa em relação aos seus concorrentes, podendo mostrar como está a situação total, ou mesmo de categorias de produtos específicas.

Exemplificando, o mercado de uma categoria de produtos fatura R\$100 milhões em um período determinado de 12 meses, a empresa ‘X’, nesse mesmo período, faturou um montante de R\$ 60 milhões nessa categoria, dessa forma a empresa ‘X’ possui 60% da participação de mercado nessa categoria, ou simplesmente 60% de *Market Share*. Como a participação da empresa ‘X’ é superior à 50% do mercado, podemos afirmar que ela é a líder de mercado.

1.4 *Machine Learning (ML)*

Segundo HURWITZ e KIRSCH (2018, p. 3), *Machine Learning* tornou-se um dos tópicos mais importantes nas organizações de desenvolvimento que procuram maneiras inovadoras de alavancar ativos de dados para ajudar a empresa a obter um novo nível de entendimento. Com os modelos apropriados de *Machine Learning*, as organizações têm a capacidade de prever continuamente as mudanças nos negócios, para que possam ter uma previsão do que virá a seguir. Como os dados são constantemente adicionados, os modelos de *Machine Learning* garantem que a solução seja atualizada constantemente. O valor é claro e evidente: se você usar as fontes de dados mais apropriadas e em constante mudança no contexto do *Machine Learning*, terá a oportunidade de prever o futuro.

Constantemente surgem avanços tecnológicos buscando resolver problemas dos mais básicos aos mais complexos existentes na humanidade. Hoje em dia, o mundo está envolto de tecnologias por todos os lados. No celular de cada um, por exemplo, existem milhares de aplicativos, alguns que auxiliam a treinar o cérebro, outros que melhoram fotos, que ajudam a cuidar da saúde, contar passos, entre outros; tudo isto cabendo no bolso. O avanço da tecnologia tem sido significativo, e um dos resultados desse avanço é o *Machine Learning*.

Para WITTEN e FRANK (2005, p. 29), *Machine Learning* tem a preocupação em formular modelos que podem ser generalizados, com intuito de responder questões a partir de possíveis hipóteses. O início do *Machine Learning* foi reconhecido nos trabalhos pioneiros de quatro estatísticos (BREIMAN et al., 1984), que publicaram um livro sobre “*classification and regression trees*” em meados de 1980, e ao longo dos anos 1970 e 1980, um pesquisador de

Machine Learning, J. Ross, estava desenvolvendo um sistema para inferir árvores de classificação por meio de exemplos. Conforme Mitchel (1997), *Machine Learning* envolve a busca de um espaço muito grande de hipóteses para determinar uma que melhor se ajuste aos dados observados.

A Figura 7 demonstra a evolução do *Machine Learning* e da inteligência artificial que, apesar de serem apresentadas como tendências de tecnologias, surgiram na década 1950, mas estão muito mais difundidas atualmente, principalmente em virtude da capacidade e velocidade de armazenamento e processamentos dos dados.

Baiju (2019), em seu artigo sobre a história do *Machine Learning*, mostra que tem o seu início na década 1950 e que vários eventos marcaram sua evolução até os dias atuais, conforme será descrito a seguir.

Década de 1950: O termo inteligência artificial nasceu durante o *Dartmouth Workshop*, em 1956, que é amplamente considerado o evento fundador da inteligência artificial como um campo. O *workshop* durou entre seis a oito semanas e contou com a participação de matemáticos e cientistas, incluindo os cientistas da computação John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon. Ainda na década de 50, em 1957, o notável *Perceptron* do psicólogo americano Frank Rosenblatt foi uma tentativa inicial de criar uma rede neural com o uso de um resistor rotativo (potenciômetro) acionado por um motor elétrico. A máquina poderia receber uma entrada (como *pixels* de imagens) e criar uma saída (como *labels*).

Década de 1960: A regra (vizinho mais próximo) *Nearest Neighbor* (NN) é um clássico no reconhecimento de padrões, que apareceu em vários trabalhos de pesquisa nos anos 60, especialmente em um artigo escrito por T. Cover e P. Hart, em 1967. O algoritmo mapeou uma rota para vendedores ambulantes, começando em uma cidade aleatória, mas garantindo que eles visitassem todas as cidades durante um breve passeio.

Década de 1970: O Conselho de Pesquisa Científica do Reino Unido publicou o relatório *Lighthill*, de James Lighthill, em 1973, apresentando uma previsão muito pessimista no desenvolvimento de aspectos centrais da pesquisa em IA. Ele afirmou que em nenhuma parte do campo as descobertas feitas até então haviam produzido o grande impacto que fora prometido. Como resultado, o governo britânico cortou o financiamento para a pesquisa em IA em todas, exceto duas universidades. Esse período de financiamento e juros reduzidos é conhecido como inverno de IA (*IA Winter*). Ainda em 1979, os estudantes da Universidade de Stanford inventaram um robô chamado Cart, ligado por rádio a um grande computador *mainframe*, que podia atravessar obstáculos em uma sala por conta própria. Embora toda a

travessia da sala tenha levado cinco horas devido a mapas e erros pouco adequados, a invenção era o estado da arte na época.

Década de 1980: Em 1981, Gerald Dejong introduziu o conceito de *Explanation Based Learning* (EBL), que analisava dados e criava uma regra geral que pode ser adotada ao descartar dados sem importância. Ainda em 1985, o professor Terry Sejnowski, de Francis Crick, inventou o NetTalk, um programa que aprendia a pronunciar texto escrito em inglês, sendo mostrado como entrada e correspondendo a transcrições fonéticas para comparação. A intenção era construir modelos simplificados que pudessem lançar luz sobre a aprendizagem humana. Então, em 1986, David Rumelhart e James McClelland publicaram o *Parallel Distributed Processing*, que avançou o uso de modelos de redes neurais para *Machine Learning*.

Década de 1990: Em 1992, o pesquisador Gerald Tesauro criou um programa baseado em uma rede neural artificial, capaz de jogar gamão com habilidades que correspondiam aos principais jogadores humanos. Em 1997, o *Deep Blue*, da IBM, se tornou o primeiro sistema de xadrez de computador a vencer um campeão mundial de xadrez. O *Deep Blue* usou o poder da computação na década de 1990 para realizar pesquisas em larga escala de possíveis movimentos e selecionar a melhor jogada.

Anos 2000: Geoffrey Hinton criou o termo *deep learning* para explicar novos algoritmos que ajudavam os computadores a distinguir objetos e texto em imagens e vídeos.

Em 2010: A Microsoft desenvolveu o dispositivo de entrada com detecção de movimento chamado *Kinect*, que podia rastrear 20 características humanas a uma taxa de 30 vezes por segundo. Isso permitiu que as pessoas interagissem com o computador por meio de movimentos e gestos.

Em 2011: O Watson da IBM venceu um jogo do *Jeopardy*, jogo de perguntas dos EUA, contra dois de seus campeões. No mesmo ano, o *Google Brain* desenvolveu sua profunda rede neural, capaz de descobrir e categorizar objetos da maneira que um gato faz.

Em 2012: O ano assistiu à publicação de um influente trabalho de pesquisa de Alex Krizhevsky, Geoffrey Hinton e Ilya Sutskever, descrevendo um modelo que pode reduzir drasticamente a taxa de erros nos sistemas de reconhecimento de imagens. Enquanto isso, o *X Lab* do *Google* desenvolveu um algoritmo de *Machine Learning* capaz de navegar autonomamente nos vídeos do *YouTube* para identificar os vídeos que continham gatos.

Em 2014: O Facebook desenvolveu um algoritmo de software *DeepFace*, que era capaz de reconhecer e verificar indivíduos em fotos com a precisão de um ser humano.

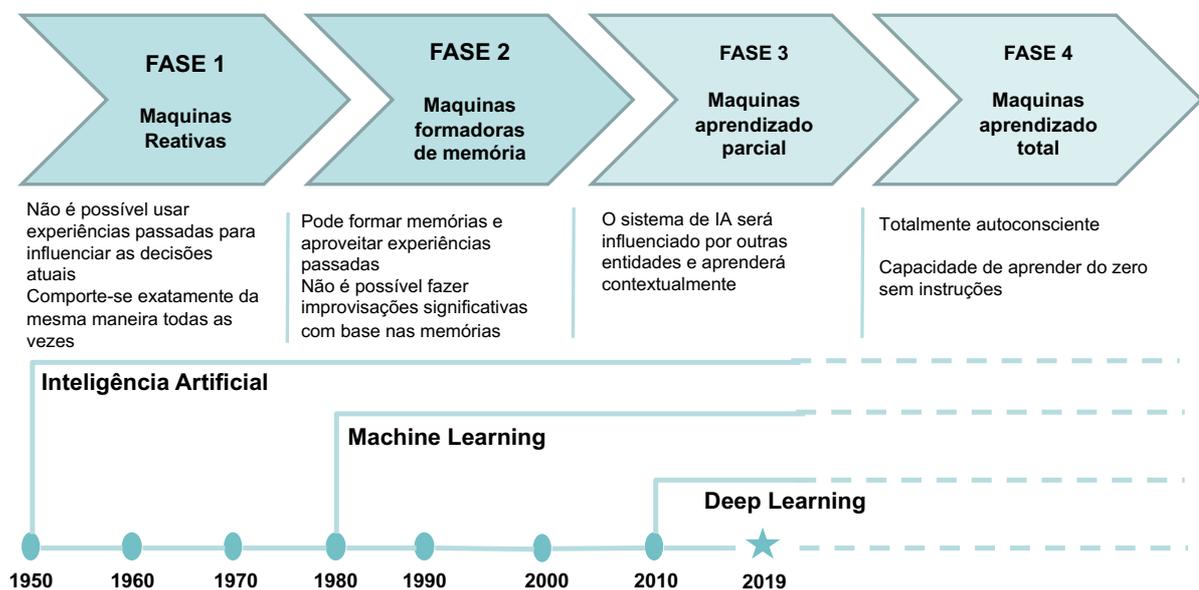
Em 2015: Andy Jassy, da AWS, lançou seus serviços gerenciados de *Machine Learning*, que analisavam os dados históricos dos usuários para procurar padrões e implantar modelos

predictivos. No mesmo ano, a Microsoft criou o *Distributed Machine Learning Toolkit*, que permitia a distribuição eficiente de problemas de *Machine Learning* em vários computadores.

Em 2016: O *AlphaGo*, criado por pesquisadores do *Google DeepMind* para jogar o antigo jogo chinês Go, venceu quatro de cinco partidas contra Lee Sedol, que é o melhor jogador de Go do mundo há mais de uma década.

Assim, foram apresentados fatos ocorridos nas décadas que marcaram a evolução do *Machine Learning*, como mostra a Figura 7, desde 1950 até os dias atuais.

Figura 7 – A Evolução do *Machine Learning* e da Inteligência Artificial (IA)



Fonte: Statista – Tendências Globais 2019, adaptado pelo autor

Segundo Kohavi e Provost (1998), o *Machine Learning* explora a construção e o estudo de algoritmos que podem aprender e fazer previsão de dados. O *Machine Learning* veio para contribuir de maneira significativa na análise de dados e no aperfeiçoamento dos resultados, afinal, quando a organização dá atenção à inteligência organizacional, esperam-se ótimos resultados, focando em decisões assertivas para alavancar o negócio.

A Figura 8 apresenta o típico ciclo de *Machine Learning*, alinhado com as questões de algoritmos que podem aprender e fazer previsão de dados. O ciclo é composto por 3 fases:

a) Detecção: nessa fase os dados são coletados de diferentes maneiras, uma vez que atualmente existem várias formas nas quais os dados podem ser capturados e armazenados. Esses dados serão utilizados como fonte para treinamento dos algoritmos de *Machine Learning*.

b) **Processamento**: segunda fase do ciclo, aqui os dados são agregados, harmonizados, enriquecidos, para em seguida possam ser processados pelo algoritmo de *Machine Learning*.

c) **Aprendizado**: nessa última fase do ciclo, padrões e tendências são revelados, gerando *insights*. Também são avaliados os resultados, possivelmente com ajuda humana ou não, dependendo do método de aprendizagem, para que sejam adaptados e melhorados os algoritmos.

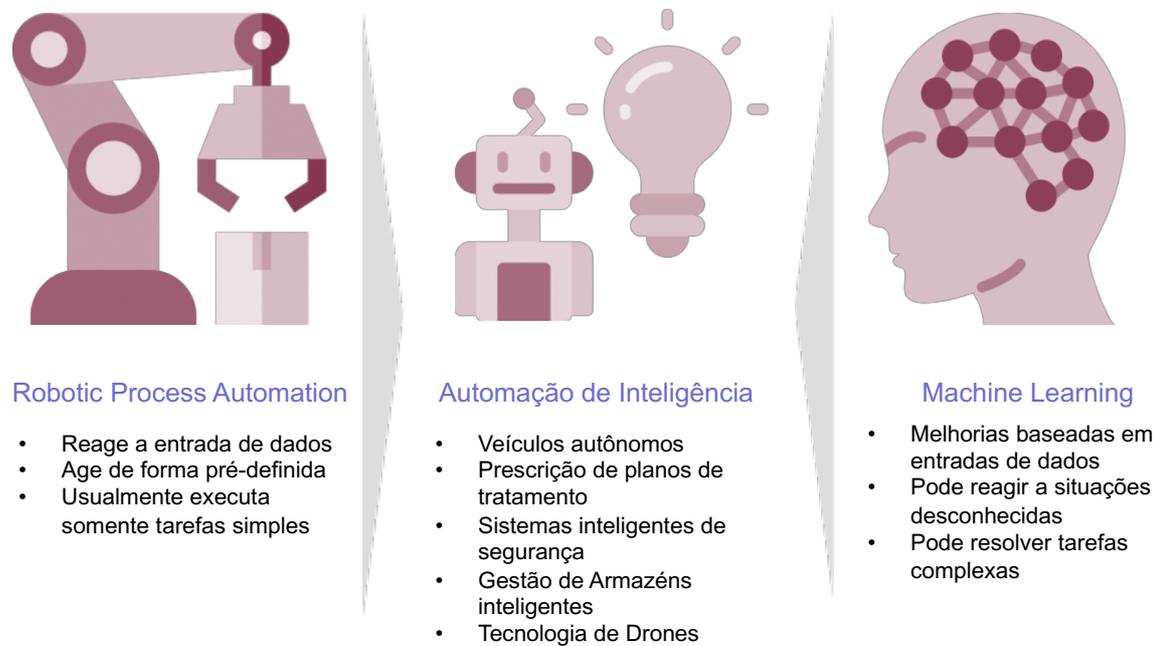
Figura 8 – O típico ciclo de *Machine Learning* (ML)



Fonte: Statista – Tendências Globais 2019, adaptado pelo autor

De acordo com a empresa de pesquisas Statista (Digital Economy Compass, 2019), embora o conceito de IA (Inteligência Artificial) pareça futurista, ele já é realmente utilizado por muitas empresas. Muitas empresas iniciantes e estabelecidas estão trabalhando para trazer aplicativos de IA ao mercado ou usá-los em seus próprios produtos e processos. Como mostra a Figura 9, Inteligência artificial é um conjunto de outras tecnologias que inclui, mas não se limita, a RPA (*Robotic Process Automation*) e *Machine Learning*.

Figura 9 – Automação de Inteligência = RPA + ML



Fonte: *Statista* – Tendências Globais 2019

No setor de tecnologia, 32% das empresas já usavam tecnologias de IA em 2017. Embora o uso da IA já seja alto em setores dependentes de tecnologia, provavelmente terá o maior impacto no setor de viagens, onde a IA pode aumentar a receita do desempenho da indústria em 11% até 2025.

Exemplos bem conhecidos de empresas que já usam a IA para seus produtos são a *Amazon*, com processamento de linguagem *Alexa*, ou *Tesla*, com seus recursos de direção autônoma. A plataforma chinesa de conteúdo de notícias e informações *Toutiao by Bytedance* usa a IA para criar e determinar a linha do tempo das notícias dos usuários com base no comportamento de uso anterior e atual, e também fornece conteúdo criado apenas pela IA.

A Figura 10 demonstra o Processo completo de *Machine Learning* (ML) e suas classes. Os métodos de *Machine Learning* podem ser divididos em três grandes classes, segundo o grau de controle dado ao usuário.

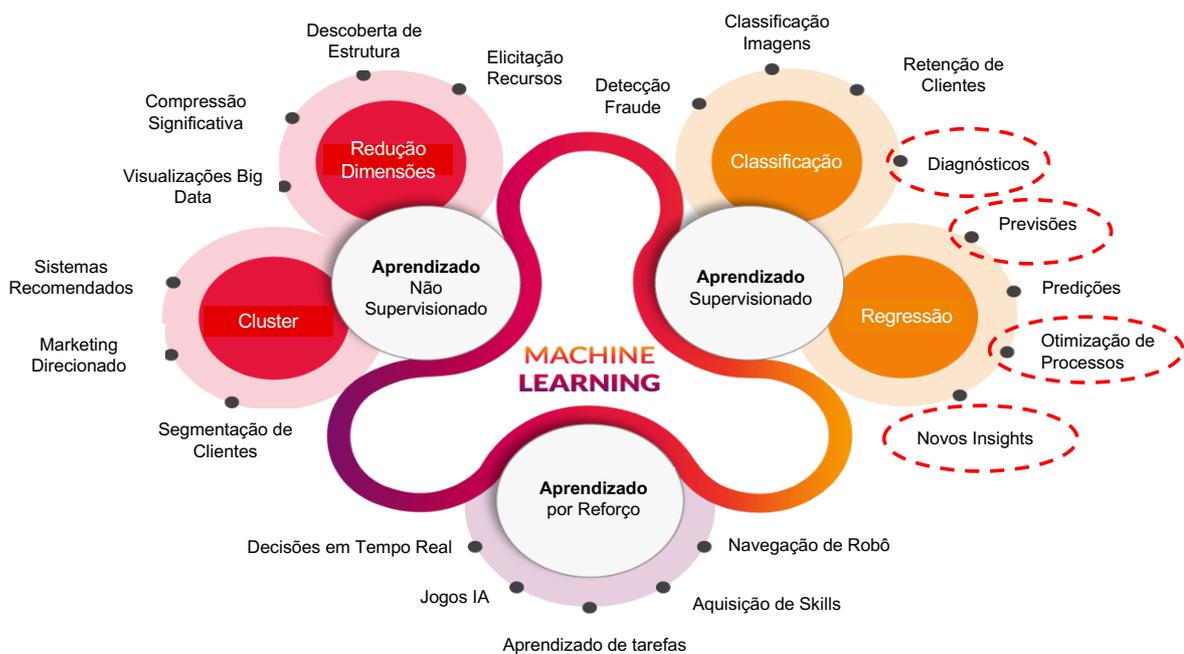
a) Aprendizado supervisionado: o usuário dispõe de um comportamento de referência preciso que ele deseja ensinar à rede. Sendo assim, a rede deve ser capaz de medir a diferença entre seu comportamento atual e o comportamento de referência, e então corrigir os pesos de

maneira a reduzir este erro (desvio de comportamento em relação aos exemplos de referência). Exemplo de aplicação: reconhecimento de caracteres em uma aplicação do tipo OCR (*Optical Character Recognition*), segundo Osório (1991).

b) Aprendizado por Reforço: o usuário possui apenas indicações imprecisas, como, por exemplo: sucesso/insucesso da rede sobre o comportamento final desejado. As técnicas de aprendizado por Reforço ou *reinforcement learning* são chamadas também de aprendizado semi-supervisionado, segundo Sutton (1998). Exemplo: aplicações em robótica autônoma, onde, supondo uma situação hipotética, sabemos que seguir em frente não é possível pois existe um obstáculo, mas, em compensação, não temos uma medida numérica que indique para que lado seguir e exatamente como devemos proceder para desviar deste obstáculo.

c) Aprendizado não supervisionado: os pesos da rede são modificados em função de critérios internos, tais como, por exemplo, a repetição de padrões de ativação em paralelo de vários neurônios. O comportamento resultante deste tipo de aprendizado é usualmente comparado com técnicas de análise de dados empregadas na estatística (*clustering*). Exemplo: diferenciar tomates de laranjas, sem, no entanto, ter os exemplos com a sua respectiva classe etiquetada (*self-organizing feature maps*), segundo Kohonen (1987).

Figura 10 – O Processo de *Machine Learning* (ML)



Fonte: Resultado da Pesquisa

Contudo, o foco da pesquisa é no aprendizado supervisionado, em modelos de classificação e regressão, como é apresentado em destaque na Figura 10.

1.4.1 *Machine Learning – Aprendizado Supervisionado*

O *Machine Learning* é um ramo da área de pesquisa em inteligência artificial. Uma categoria proeminente de aplicativos de *Machine Learning* supervisionado é a classificação (BISHOP, 2006). Segundo Hurwitz e Kirsch (2018, p. 20), o aprendizado supervisionado geralmente começa com um conjunto estabelecido de dados e um certo entendimento de como esses dados são classificados.

Pode-se encontrar tarefas de classificação em vários domínios que requerem julgamento de especialistas, como na área da saúde (por exemplo, determinar se alguém tem um tumor), lei (por exemplo, determinar se se deve condenar alguém por um crime) ou construção (por exemplo, determinar se uma certa construção seria estável). Com a disponibilidade de *big data*, este assunto será abordado com mais profundidade na seção 1.5, que aborda *big data* e *analytics*. Em determinados domínios de aplicativos, o potencial de aplicar o *Machine Learning* para classificação também aumentou. Por exemplo, Sim (2016) enfatiza a importância de tecnologias, como o IBM Watson, para várias tarefas de diagnóstico em um contexto médico. Também é necessário *big data* para treinar técnicas de *Machine Learning*, para que eles possam fornecer resultados precisos de classificação.

O aprendizado supervisionado tem como objetivo encontrar padrões nos dados que podem ser aplicados a um processo de análise. Esses dados rotulam recursos que definem o significado dos dados. Por exemplo, pode haver milhões de imagens de animais e ser incluída uma explicação sobre cada animal e, em seguida, você pode criar um aplicativo de aprendizado de máquina que distinga um animal do outro. Ao rotular esses dados sobre tipos de animais, você pode ter centenas de categorias de espécies diferentes. Como os atributos e o significado dos dados foram identificados, é bem compreendido pelos usuários que estão treinando os dados modelados para que se ajustem aos detalhes dos rótulos. Quando o rótulo é contínuo, é uma regressão; quando os dados provêm de um conjunto finito de valores, eles são conhecidos como classificação.

Em essência, a regressão usada para aprendizado supervisionado ajuda a entender a correlação entre variáveis. Um exemplo de aprendizado supervisionado é a previsão do tempo. Ao usar a análise de regressão, a previsão do tempo leva em consideração os padrões climáticos históricos conhecidos e as condições atuais para fornecer uma previsão do tempo.

Os algoritmos são treinados usando exemplos pré-processados e, neste ponto, o desempenho dos algoritmos é avaliado com dados de teste. Ocasionalmente, os padrões identificados em um subconjunto de dados não podem ser detectados na população maior de dados. Se o modelo estiver apto a representar apenas os padrões que existem no subconjunto de treinamento, você cria um problema chamado sobre-ajuste.

O excesso de ajuste significa que seu modelo está ajustado com precisão para os dados de treinamento, mas pode não ser aplicável a grandes conjuntos de dados desconhecidos. Para se proteger contra ajustes excessivos, o teste precisa ser realizado com dados rotulados imprevistos ou desconhecidos. O uso de dados imprevistos para o conjunto de testes pode ajudá-lo a avaliar a precisão do modelo na previsão de resultados e nos resultados.

Os modelos de treinamento supervisionados têm ampla aplicabilidade. Abrangem uma variedade de problemas de negócios, incluindo detecção de fraudes, soluções de recomendação, reconhecimento de fala ou análise de risco.

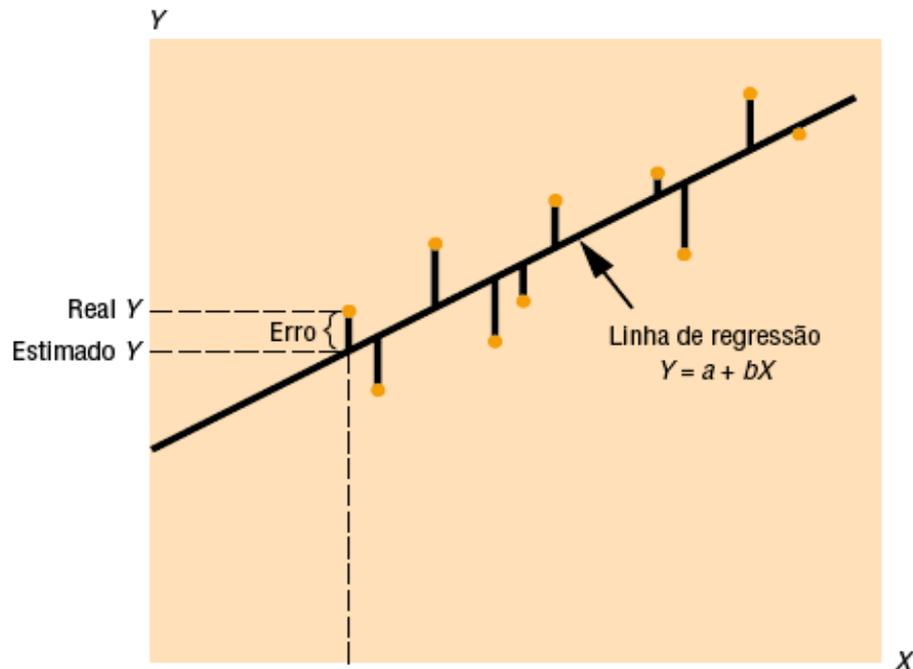
No desenvolvimento do protótipo da solução computacional, foco dessa pesquisa, foram aplicados algoritmos de classificação, utilizados no aperfeiçoamento das palavras-chaves, e algoritmo de regressão linear para calcular a demanda estimada de equipamentos necessários.

Serão abordados os temas referentes à regressão linear e ao algoritmo de *Machine Learning*, bem como algoritmos de classificação, que serão utilizados no enriquecimento das palavras-chaves utilizadas no protótipo da solução computacional.

1.4.2 Regressão Linear Simples

Segundo Gujarati (2011), a regressão é a análise da dependência de uma variável, a variável dependente, com uma ou mais variáveis, as variáveis explicativas, que tentam estimar ou prever a média (população) com bases naquilo que já se conhece, através de uma repetição de amostras. A análise de regressão é uma ferramenta estatística que estuda o relacionamento entre uma variável dependente e diversas outras variáveis independentes (HILL, 2010). Isso se dá através de um modelo matemático que associa a variação no comportamento da variável dependente com as variações existentes nas variáveis independentes, como é possível ver na Figura 11.

Figura 11: Regressão Linear Simples



Fonte: Garrison, Noreen e Brewer (2013)

No presente estudo serão utilizados os modelos de regressão linear simples e múltipla, em que as variáveis independentes (explicativas) ‘X’ são utilizadas para explicar o comportamento da variável dependente (explicada) ‘Y’, por meio da Equação 3, descrita a seguir:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (3)$$

Os ‘ β ’ são os parâmetros da regressão. Estes coeficientes podem ser determinados pelo método dos mínimos quadrados. Por fim, o ‘ ε ’ é a variável aleatória residual, representando todas as influências no comportamento de Y que não são explicadas pelas variáveis independentes.

Segundo Gujarati (2011), o objetivo da regressão é estimar o valor médio da variável dependente em termos das demais variáveis explicativas. Por esse motivo, quando o modelo é representado por uma reta, é necessário buscar aquela que se aproxima mais dos pontos, reduzindo dessa forma a dispersão.

Ao se fazer uma modelo regressão linear, é necessário realizar análises para verificar se os critérios utilizados na regressão são válidos. Para isso, são feitos alguns testes e avaliadas algumas premissas (PELLI, 2003). Nesse estudo, serão avaliados os critérios da correlação, o teste F, e outros fatores que também podem ser avaliados, por exemplo, a Heterocedasticidade.

1.4.3 *Análises e classificação de palavras-chaves*

A análise e classificação de palavras-chaves por meio de um algoritmo de *Machine Learning* é um tópico bastante relevante dessa pesquisa, pois, quanto maior a qualidade e mais específica for a palavra-chave, maior será a precisão no cálculo do *Market Share*, porque será mais bem relacionado o importador com o equipamento importado, aumentando a acurácia da informação.

Menger (2018) aplicou os algoritmos de *Machine Learning* para prever incidentes de violência entre pacientes de um hospital. Basicamente, os dados para serem processados foram os textos das análises clínicas e histórico dos pacientes. Em nosso protótipo utilizaremos a clássica *bag-of-words*, onde uma sentença é representada por meio de duas ou mais palavras, desconsiderando a gramática ou sequência das palavras. A partir desse saco de palavras será utilizada a técnica de *tokenization* (tokenização), onde as palavras ganham valores-chaves para poder ter uma representação em uma matriz e assim ser consumida pelos algoritmos. O algoritmo utilizado foi o *Naive Bayes*.

Pak e Paroubek (2010) utilizaram também um classificador *Naive Bayes* para categorizar tweets em positivo ou negativo, com base em N gramas e na classificação gramatical de partes do texto.

Davidov, Tsur e Rappoport (2010) buscaram categorizar os tweets utilizando apenas as mensagens com hashtags – palavras precedidas que representam o sentimento e/ou o assunto alvo daquela mensagem – e emoticons contidos no texto para treinar o classificador utilizado e categorizar tweets entre diversas emoções. Além disso, aspectos como pontuação e as palavras utilizadas também foram consideradas para a extração de sentimento.

Outra vertente da análise do conteúdo gerado no Twitter é a extração de informação a partir do estudo de posts publicados nessa rede. Sakaki, Okazaki e Matsuo (2010) promoveram pesquisas cujo objetivo era detectar acontecimentos de forma mais rápida por meio da análise das mensagens divulgadas. Nesses casos, a intenção era criar uma ferramenta capaz de detectar terremotos e epidemias de gripe, respectivamente, antes mesmo de tais eventos serem anunciados oficialmente.

O classificador *Naive Bayes* é um modelo probabilístico simples, baseado na regra de Bayes, com uma forte suposição de independência. O modelo *Naive Bayes* envolve uma suposição de independência condicional simplificadora. Devido a uma classe (positiva ou negativa), as palavras são condicionalmente independentes uma da outra. Essa suposição não afeta muito a precisão na classificação do texto, mas cria algoritmos de classificação realmente rápidos, aplicáveis ao problema.

No nosso trabalho, a probabilidade máxima de uma palavra-chave (produto) pertencente a uma classe (fabricante) específica é dada pela expressão:

$$P(c / x) = \frac{P(x / c)P(c)}{P(x)}$$

$$P(c / x) = P(x_1 / c) * P(x_2 / c) * ... * P(x_n / c) * P(c) / (P_x)$$

Acima,

- $P(c | x)$ é a probabilidade posterior da classe (c, alvo) dada preditor (x, atributos)
- $P(c)$ é a probabilidade original da classe
- $P(x | c)$ é a probabilidade que representa a probabilidade de preditor dada à classe
- $P(x)$ é a probabilidade original do preditor

Em nosso protótipo, foi aplicado o algoritmo de *Naive Bayes*, que também é oferecido na biblioteca de *Machine Learning* de *Python*, *Scikit-Learn*, que é uma das bibliotecas de *Machine Learning* mais populares e utilizada.

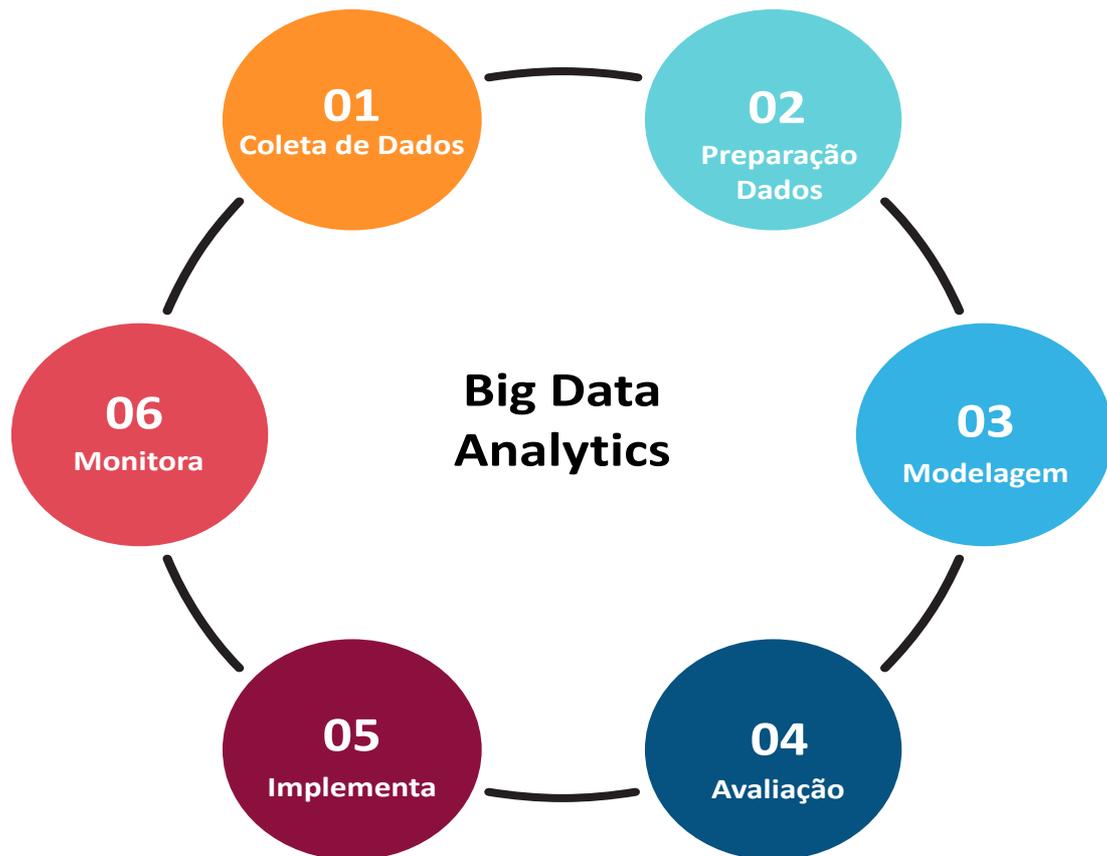
Na seção de resultados, onde será apresentado como o protótipo foi desenvolvido, será exibido o algoritmo de *Naive Bayes* como foi aplicado no protótipo.

1.5 ***Big Data e Analytics***

O *Big Data Analytics*, também conhecido como *Big Data Mining*, é o processo de descobrir padrões de conhecimento acionáveis a partir de *big data* (WU, BUYYA & RAMAMOCHANARAO, 2016). As empresas modernas se beneficiam dos processos de *big data*, pois fornecem informações de clientes e dados corporativos. A análise de *big data* ajuda a alcançar as metas de negócios, levando à retenção de clientes e maximização de lucros. Por exemplo, o Twitter usa a análise de *big data* como ferramenta de retenção de clientes, na qual

hashtags de tendências são exploradas para atrair clientes. Da mesma forma, empresas de comércio eletrônico, como Amazon e Netflix, usam ferramentas de análise de *big data* para recomendar produtos novos e relevantes aos usuários e maximizar sua geração de receita (KONSTAN & RIEDL, 2012).

Figura 12: O processo de *Big Data Analytics*



Fonte: Resultado da pesquisa

O processo de *Big Data Analytics*, como mostrado na Figura 12, é baseado em seis etapas principais: (1) *big data* é coletado de várias fontes de dados, diretas e indiretas, (2) operações de pré-processamento e integração de dados são realizadas para melhorar a qualidade do *big data*, (3) modelos de aprendizado são gerados usando métodos estatísticos e técnicas de mineração de dados baseadas em *Machine Learning*, (4) os modelos são avaliados usando dados de teste, (5) os modelos são implantados em aplicativos reais e (6) o desempenho dos modelos é monitorado em termos de precisão de previsão (REHMAN, KHAN & BATOOL, 2016).

Pesquisas mostram que 60% dos cientistas de dados passam a maior parte do tempo trabalhando na preparação de dados, aumentando assim a conservação de recursos nas empresas

(CrowdFlower, 2016). A adoção de estratégias eficazes de redução de dados facilita a otimização da carga de trabalho. Da mesma forma, as empresas podem minimizar o custo financeiro dos serviços de armazenamento de dados (Chang & Wills, 2016).

Os provedores de serviços em nuvem podem reduzir os custos operacionais, otimizando os serviços de armazenamento e a movimentação mínima de dados na rede (FU, JIANG & XIAO, 2012). Outra perspectiva para a redução de dados é a de que o conhecimento histórico sobre os comportamentos e operações da empresa dos clientes deva ser preservado, ao invés de processar iterativamente os mesmos dados brutos.

As violações da política de segurança (GALEGALE et al., 2017) e os compromissos de privacidade no final das empresas foram observados recentemente, colocando assim sérias ameaças aos clientes e diminuindo a confiança entre empresas e clientes. As empresas se beneficiam dos métodos de redução de *big data* de várias maneiras. São executadas informações de pré-processamento para reduzir fluxos de *big data* antes de entrar em sistemas de computação em nuvem (DI MARTINO et al., 2014).

São executados métodos de redução de dimensão para abordar a maldição da dimensionalidade e determinar os conjuntos de *big data* substancialmente relevantes (ZHAI et al., 2014). São também executados métodos de compactação e descompactação para reduzir a utilização da largura de banda na rede em *data centers* em nuvem (YANG et al., 2014).

Os métodos baseados na teoria de redes são usados para descobrir o relacionamento semântico entre pontos de dados no *big data*, para otimizar as operações de armazenamento e processamento em sistemas de computação em nuvem (TROVATI, 2015). Métodos de eliminação de redundância são usados para remover dados duplicados, para melhorar o valor do *big data* (FU et al., 2012). Os métodos de mineração de dados e aprendizado de máquina são usados para descobrir os padrões de conhecimento, para utilização lateral ao invés do processamento iterativo de dados brutos (JIANG et al., 2014).

1.6 Robotic Process Automation (RPA)

O *Robotic Process Automation* (RPA) é uma resposta industrial à enorme quantidade de trabalho manual que os indivíduos realizam diariamente, semanalmente ou mensalmente para oferecer suporte a uma ampla gama de processos de negócios de alto volume (AGUIRRE & RODRIGUEZ, 2017; Lacity e Willcocks, 2016). O RPA está associado principalmente ao nível da tarefa. As áreas de aplicação incluem finanças e contabilidade, manutenção de infraestrutura de TI e processamento de *front-office*. Os chamados robôs são programas de

software que interagem com sistemas, como planejamento de recursos empresariais e sistemas de gerenciamento de relacionamento com clientes. Os robôs podem coletar dados dos sistemas e atualizá-los, imitando manipulações manuais baseadas em tela.

Existem fornecedores dedicados de RPA, como AutomationEdge, Automation Anywhere, Blue Prism, Kryon Systems, Softomotive e UiPath, que oferecem apenas software RPA (Le Clair 2017; Tornbohm 2017). A promessa é a de que, com o uso das técnicas de Inteligência Artificial (IA) e *Machine Learning* (ML), tarefas mais complexas e menos definidas possam ser suportadas. Os seres humanos aprendem fazendo e aprendem com um treinador. O objetivo é que as ferramentas RPA aprendam da mesma maneira. Há também um vínculo óbvio com a mineração de processos (VAN DER AALST, 2016; KERREMANS, 2018).

Será utilizado no protótipo a tecnologia de RPA para o monitoramento da disponibilidade do arquivo da RFB e a transferência desse arquivo para o ambiente de processamento do protótipo.

2 METODOLOGIA

A Metodologia de Pesquisa em *Design Science Research* (DSR) foi a metodologia escolhida para orientar esse trabalho de pesquisa. O DSR fornece um modelo de processo para pesquisa em Sistemas de Informação (SI) e outras disciplinas de recursos aplicados, bem como um modelo mental para os revisores avaliarem os pesquisadores, segundo K. Peffers et al. (2007).

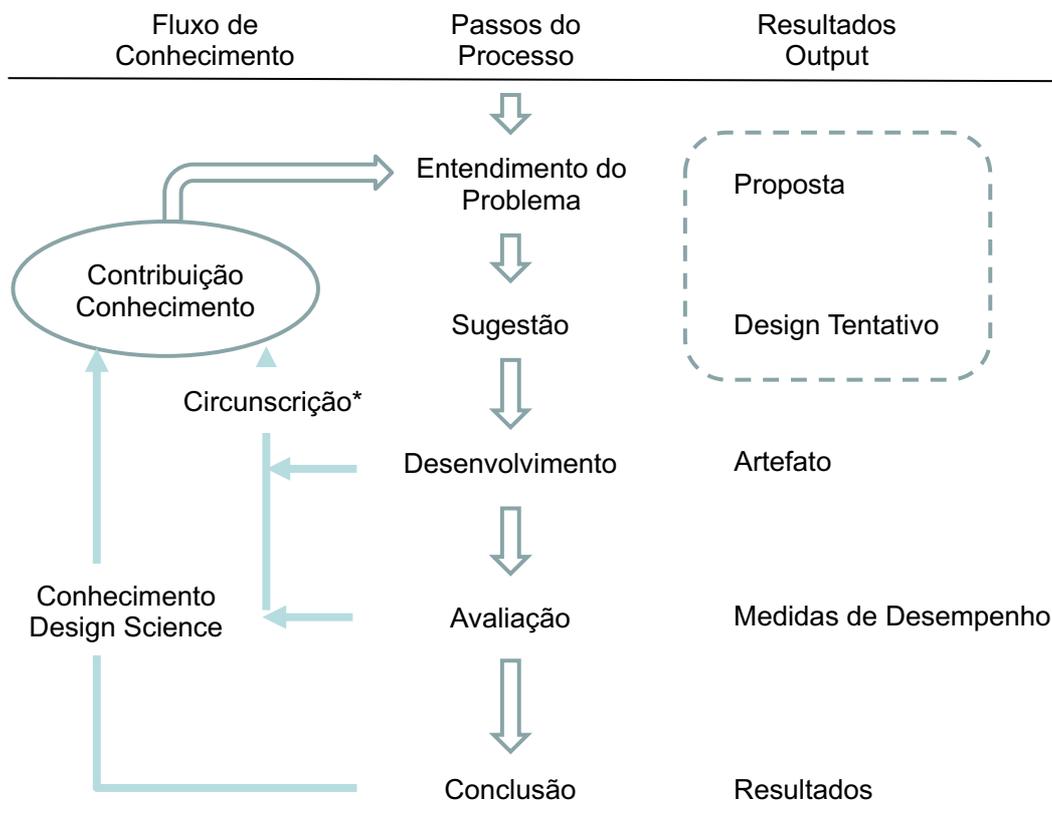
O principal objetivo do DSR na pesquisa de SI é produzir e avaliar um artefato de TI que suporte a solução para um problema organizacional identificado, de acordo com Hevner (2004). Os artefatos produzidos podem ser construtos (vocabulário e símbolos), modelos (abstrações e representações), métodos (algoritmos e práticas) e instanciações (sistemas implementados e protótipos). A pesquisa em DSR geralmente envolve a criação de um artefato e/ou a teoria do design como um meio para melhorar o estado atual da prática, bem como o conhecimento de pesquisa existente (BASKERVILLE et al., 2018).

A pesquisa em DSR se concentra no desenvolvimento de artefatos envolvendo duas atividades principais para melhorar e entender o comportamento dos aspectos dos sistemas de informações: (1) a criação de novos conhecimentos por meio do design de artefatos novos ou inovadores (coisas ou processos) e (2) análise do uso e/ou desempenho do artefato com reflexão e abstração.

Os artefatos criados no processo de pesquisa em DSR incluem, entre outros, algoritmos, interfaces humano/computador, metodologias de desenho de sistemas ou linguagens. Um modelo de processo geral é seguido por uma pesquisa em DSR, como se observa no modelo descrito na Figura 13. Esse modelo é uma adaptação de um modelo de design de processo computacional desenvolvido por Takeda et al. (1990). Embora as diferentes fases de um processo de design e de um processo de pesquisa em *design science* sejam semelhantes, as atividades realizadas nessas fases são consideravelmente diferentes.

Além disso, o que torna o modelo de processo de pesquisa em *design science* diferente do modelo de processo de design correspondente é o fato de que a contribuição do conhecimento novo (e verdadeiro) precisa ser um foco principal da pesquisa em *design science*. A Figura 13 demonstra um típico modelo de pesquisa utilizando DSR.

Figura 13 – Modelo de Processo de Pesquisa em Design Science (DSR-Design Science Research)



*Circunscrição é a descoberta do conhecimento restritivo sobre as teorias obtidas através da detecção e análise de contradições quando as coisas não funcionam de acordo com a teoria (MCCARTHY, 1980).

Fonte: Adaptado pelo autor (*Design Science in Information Systems*).

Entendimento do Problema: O entendimento do problema de uma pesquisa interessante pode vir de várias fontes, incluindo novos desenvolvimentos na indústria ou identificação de problemas em uma determinada disciplina. A leitura em uma disciplina aliada também pode oferecer a oportunidade de aplicação de novas descobertas no campo do pesquisador.

Os tipos de problemas relevantes para um esforço de pesquisa em DSR tendem a ser focados na solução de problemas em sua abordagem, em oposição a perguntas ou problemas que são respondidos por meio de explicação. Como parte da fase de entendimento do problema, o pesquisador considera critérios para avaliar o produto final do esforço de pesquisa. O resultado dessa fase é uma proposta, formal ou informal, para um novo esforço de pesquisa.

Sugestão: Imediatamente após o desenvolvimento de uma proposta com base no entendimento de um problema é a fase da sugestão. A sugestão é uma etapa criativa, em que

uma nova funcionalidade é prevista com base em uma nova configuração de elementos existentes ou novos. De fato, em qualquer proposta formal de pesquisa em DSR, como uma proposta para uma agência de subsídios do governo ou um patrocinador da indústria, um Design Tentativo e o desempenho de um protótipo baseado nesse design seriam parte integrante da Proposta.

Além disso, se, depois de investir um esforço considerável em um problema interessante, um Projeto Tentativo ou pelo menos o germe de uma ideia para solução de problemas não se apresentar ao pesquisador, a ideia (Proposta) será deixada de lado. É essa conexão íntima entre a Proposta e o Projeto Tentativo que serve como o motivo da linha pontilhada em torno das saídas da fase Entendimento de um Problema e Sugestão, mostrada na Figura 13.

Desenvolvimento: O Projeto Tentativo é desenvolvido e implementado nesta fase. Existem muitas formas de artefatos que podem ser desenvolvidos, desde teorias de design (GREGOR & JONES, 2007) a conceitos, modelos, processos ou instanciações (MARCH & SMITH, 1995; HEVNER et al., 2004).

As técnicas para implementação variam, é claro, dependendo do artefato a ser criado. Um algoritmo pode exigir a construção de uma prova formal para mostrar sua correção. Um sistema especialista que incorpore novas suposições sobre a cognição humana em uma área de interesse exigirá desenvolvimento de software, provavelmente usando um pacote ou ferramenta de alto nível.

A implementação em si pode ser muito tradicional e não precisa envolver novidades além do estado de prática para o artefato especificado; a novidade está principalmente no design, não na construção do artefato.

Avaliação: Uma vez construído, o artefato é avaliado de acordo com critérios sempre implícitos na Proposta (fase Entendimento do Problema). Desvios das expectativas, quantitativas e qualitativas, são cuidadosamente observados e devem ser tentativamente explicados.

A fase de avaliação contém uma subfase analítica, na qual são formuladas hipóteses sobre o comportamento esperado e os impactos do artefato, usando uma estratégia de avaliação consistente com as necessidades de avaliação (VENABLE et al. 2016).

Raramente, na pesquisa em DSR, hipóteses iniciais sobre comportamento são completamente confirmadas. Em vez disso, os resultados da fase de avaliação e as informações

adicionais obtidas na construção e execução do artefato são reunidos e devolvidos a outra rodada de Sugestões (consulte a seta de circunscrição da Figura 13).

Embora a pesquisa em *design science* frequentemente se concentre em examinar a utilidade de um artefato (por exemplo, HEVNER et al, 2004), outros sugeriram que a avaliação do artefato seja avaliada quanto à sua aptidão para se adaptar e sobreviver em um ambiente (GILL & HEVNER, 2013) ou considerando os impactos sociais do artefato (DELEOZ & PETTER, 2018).

As hipóteses explicativas, bastante amplas, raramente são descartadas, mas são modificadas para estar de acordo com as novas observações.

Os resultados da avaliação geralmente sugerem um novo design, frequentemente precedido por pesquisas adicionais para entender as razões pelas quais o comportamento e os impactos do artefato se desviaram do desempenho teórico esperado.

Conclusão: Essa fase pode ser apenas o fim de um ciclo de pesquisa ou é o final de um esforço de pesquisa específico. O final de um esforço de pesquisa geralmente é o resultado de satisfação, embora ainda haja desvios no comportamento do artefato em relação às (múltiplas) previsões hipotéticas revisadas; os resultados são considerados bons o suficiente.

Não são apenas os resultados do esforço, consolidados e redigidos nesta fase, mas o conhecimento obtido pelo esforço é frequentemente classificado como firme. São apresentados fatos que foram aprendidos e podem ser aplicado repetidamente; ou comportamento que possa ser invocado repetidamente, ou como pontas soltas; comportamento anômalo que desafia a explicação e pode muito bem servir como objeto de pesquisas adicionais.

A comunicação é muito importante na pesquisa (HEVNER et al., 2004). Portanto, essa fase, como conclusão de um esforço de pesquisa (indicado na Figura 13 pela pequena seta à esquerda que parte de “Contribuição do conhecimento”), precisa posicionar adequadamente a pesquisa que está sendo relatada e justificar sua contribuição ao conhecimento (GREGOR & HEVNER, 2013).

Dependendo do tipo de contribuição do conhecimento e do estado do conhecimento na área de pesquisa, as expectativas quanto à natureza e profundidade dos resultados da contribuição do conhecimento podem variar, como veremos a seguir sobre resultados (*outputs*) da pesquisa em DSR.

Resultados (*Outputs*) da Pesquisa em DSR: O resultado de um projeto de pesquisa em DSR deve ser o conhecimento em DSR. Para entender que forma essa contribuição de

conhecimento pode assumir, é recomendável começar com a compreensão dos possíveis tipos de contribuições de conhecimento da pesquisa em DSR.

O conhecimento do DSR se manifesta na forma de artefatos/constructos, modelos, estruturas, arquiteturas, princípios de design, métodos e/ou instanciações e teorias de design (ver a Tabela 1).

A instanciação é geralmente chamada de artefato material, enquanto os outros tipos de artefatos são chamados de artefatos abstratos.

Tabela 1 – Resultados (*outputs*) de Pesquisa Design Science (DSR)

1	Constructos	O vocabulário conceitual de um domínio
2	Modelos	Conjuntos de proposições ou declarações que expressam relacionamentos entre constructos
3	<i>Frameworks</i>	Guias reais ou conceituais para servir como suporte ou guia
4	Arquiteturas	Estruturas de alto nível de sistemas
5	Princípios de Design	Princípios e conceitos básicos para orientar o design
6	Métodos	Conjuntos de etapas usadas para executar tarefas - conhecimento prático
7	Instanciações	Implementações situadas em certos ambientes que operacionalizam ou não construções, modelos, métodos e outros artefatos abstratos; neste último caso, esse conhecimento permanece tácito.
8	Teorias de Design	Um conjunto prescritivo de instruções sobre como fazer algo para alcançar um determinado objetivo. Uma teoria geralmente inclui outros artefatos abstratos, como construções, modelos, estruturas, arquiteturas, princípios de design e métodos.

Fonte: adaptado pelo autor de March and Smith (1995).

De acordo com a metodologia de pesquisa *Design Science*, elaborou-se a Tabela 2, com os elementos propostos para essa dissertação de Mestrado.

Tabela 2 – Dados da Dissertação utilizando DSR

1	Entendimento do Problema e Motivação	Melhorar a gestão de Demanda e planejamento de inventários de equipamentos médico-hospitalares, com base em informações do mercado, <i>Market Share</i> e projeções de ganho e aumento de mercado.
2	Sugestão e definições dos objetivos para a solução	Propor o desenvolvimento de um protótipo de solução computacional que, por meio dos dados de mercado, <i>Market Share</i> obtido pelo processamento dos dados de equipamentos importados, demanda e níveis de inventários, possa calcular os níveis de inventários e de equipamentos que devem ser importados, derivado de uma estratégia de simulação de crescimento e ganho de mercado. O desenvolvimento terá como base os requerimentos obtidos por meio das entrevistas com os principais <i>stakeholders</i> da solução, descritos no Quadro 4.

3	Desenvolvimento e desenho da solução	Desenvolver um protótipo de solução computacional, utilizando linguagens de programação Java/HTML5 e algoritmos de <i>Machine Learning</i> (ML) de bibliotecas <i>Python</i> , capaz de importar os dados de equipamentos importados (disponibilizados mensalmente pela RFB), e por meio de palavras-chaves e NCM, possa identificar os itens desejados e calcular o % de mercado de cada fabricante. Em seguida, com os dados de demanda e inventário, e utilizando projeções de crescimento, bem como os algoritmos de ML de regressão linear simples, propor o nível ideal de inventário para esses produtos, para que não haja falta, tão pouco excessos.
4	Avaliação	Submeter a aplicação para validação dos dados de uma empresa de equipamentos médico-hospitalares e avaliar os resultados de acordo com as projeções propostas nos testes sobre <i>Market Share</i> e inventário. Posteriormente ao período de testes, efetuar uma entrevista semiestruturada com usuários chaves, para avaliar a usabilidade, facilidade e níveis de satisfação com a solução implementada, bem como identificar oportunidades de melhoria.
5	Conclusão e Comunicação	Protótipo de uma solução computacional para que possa ser introduzida no mercado, mesmo que ainda com oportunidades de melhorias. Elaboração de artigo científico a ser apresentado em Congressos da área e submetido para publicação em periódico.

Adicionalmente, como parte da metodologia, foi realizada uma pesquisa bibliográfica para a fundamentação teórica envolvendo os temas: Sistemas Produtivos; Equipamentos Médico-Hospitalares; Dados Governamentais abertos; *Market Share*; *Machine Learning*; *Data Analytics*; *Robotic Process Automation* e Gestão de Demanda e Inventário.

Foi realizado um estudo de campo, por meio de entrevistas com os líderes das áreas estratégicas, comerciais e da cadeia de suprimentos, com o objetivo de levantar os requisitos necessários à solução computacional e identificar os aspectos chaves, para criar uma solução aplicável à realidade das partes interessadas. As entrevistas seguiram um roteiro pré-determinado, com questionários semiestruturados e todos os resultados registrados em forma de diagramas, mapas, fotos e quaisquer demais formas de registros possíveis (SAMPIERI; COLLADO; LUCIO, 2013). O questionário utilizado como base das entrevistas encontra-se na seção de Apêndices (ver Apêndice A). O resultado obtido em campo refletiu o mapeamento da cadeia de valor dos equipamentos importados no mercado brasileiro, incluindo o comportamento do mercado, planejamento de demanda e inventários, e níveis de inventários com foco no cliente.

3 RESULTADOS E DISCUSSÃO

O capítulo final desta dissertação descreve sobre a criação de um protótipo de solução computacional para solucionar um problema de negócio, encontrado principalmente nas empresas de equipamentos médico-hospitalares, mas não limitado apenas à essa indústria.

Posteriormente, o protótipo será convertido em um modelo de negócios, para ser comercializado inicialmente com a indústria foco dessa pesquisa, mas a comercialização poderá ser estendida a outras indústrias, dependendo da demanda.

Executou-se uma entrevista com duas empresas importadoras de equipamentos médico-hospitalares, para capturar os requisitos mais importantes para o protótipo da plataforma de software, dessa maneira possibilitando a criação de um protótipo. O roteiro da pesquisa, usado como base para as entrevistas, pode ser encontrado como Apêndice A.

O Quadro 4 mostra o perfil das pessoas que foram entrevistadas com o objetivo de captura dos requerimentos necessários para o desenvolvimento da solução computacional. O foco, a princípio, foi com pessoas que trabalham na indústria de equipamentos médico-hospitalares, pois é o principal foco dessa pesquisa.

Quadro 4 – Perfil das pessoas entrevistadas

Função	Formação	Tempo de Experiência (em anos)	Tempo de Empresa (em anos)	Indústria de Equipamentos Médicos?	Área de Trabalho
Sales Operations Manager	Gestão estratégica de vendas	+20	+15	Sim	Cromatografia e Espectrometria
Marketing Developer Manager	Biologia aplicada à saúde	+15	+8	Sim	Biotecnologia
Strategic & Innovation Manager	Adm. negócios	+10	+3	Sim	Estratégia e Inovação
CEO, Presidente	Engenharia industrial	+25	+5	Sim	Gerência Geral
CEO, Owner	Adm. empresas	+30	+20	Não	Gerência Geral

Fonte: Resultado da Pesquisa

O desenvolvimento da solução computacional foi realizado com o apoio da iniciativa privada, por meio de um acordo de cooperação tecnológica com a empresa Falcon Analytics.

Principais concorrentes, características e diferenciais do protótipo: Por meio de uma pesquisa efetuada na internet, buscou-se encontrar os três principais concorrentes do protótipo, bem como entender suas principais características e mercados de atuação, conforme o Quadro 5:

Quadro 5 – Principais concorrentes do protótipo e suas características

Nome da Solução	Empresa	Características	Atua somente no Brasil?	Mercado Específico
DataImport	Empresa com foco em geração de relatórios de matérias primas e produtos importados	a) Custos reais de produtos importados por diferentes empresas b) País de origem e de compra c) Descrições detalhadas da importação realizada d) Valores de fretes, seguro e quantidades e) Data e alfândega de desembarque	Sim	Não
FuncEx	Fundação Centro de estudos do comércio exterior	f) Acesso irrestrito e gratuito ao público para informações como: indicadores, câmbio, links e países g) Base de dados com informações de comércio exterior para que sejam produzidos relatórios, acesso pago	Sim	Não
Datagro	Consultoria agrícola independente, foco em: análises, dados primários e notícias sobre as principais commodities agrícolas mundiais	a) Relatórios diários b) Análises c) Preços de mercados físicos e futuros d) Dados de clima e) Fretes	Não	Sim Agrícola

Fonte: Resultado da Pesquisa

No anexo C, estão telas e endereço do website de cada uma das plataformas listadas no quadro 5.

Quadro 6 – Diferenciais do protótipo comparado com os principais concorrentes

Funcionalidade	Protótipo	Concorrentes
Relatórios por categoria de produtos	Sim	Sim
RPA – <i>Robotic Process Automation</i>	Sim	Não
<i>Machine Learning</i>	Sim	Não
<i>Big Data Analytics</i>	Sim	Não
Permite validar com os produtos internos da empresa?	Sim	Não
Calcula projeção de crescimento e níveis de inventários	Sim	Não

Fonte: Resultado da Pesquisa

3.1 Protótipo da solução computacional

Como descrito anteriormente, o protótipo da solução computacional será baseado nos dados de importação dos equipamentos médico-hospitalares, disponibilizados mensalmente no website da RFB. Os dados são disponibilizados em formato digital, mensalmente, e separados por NCM. Atualmente esses dados necessitam ser descarregados manualmente e será dessa forma, no protótipo, porém já estamos trabalhando em uma funcionalidade futura, que irá buscar os arquivos automaticamente assim que estiverem disponíveis, utilizando de recursos de RPA (*Robotic Process Automation*).

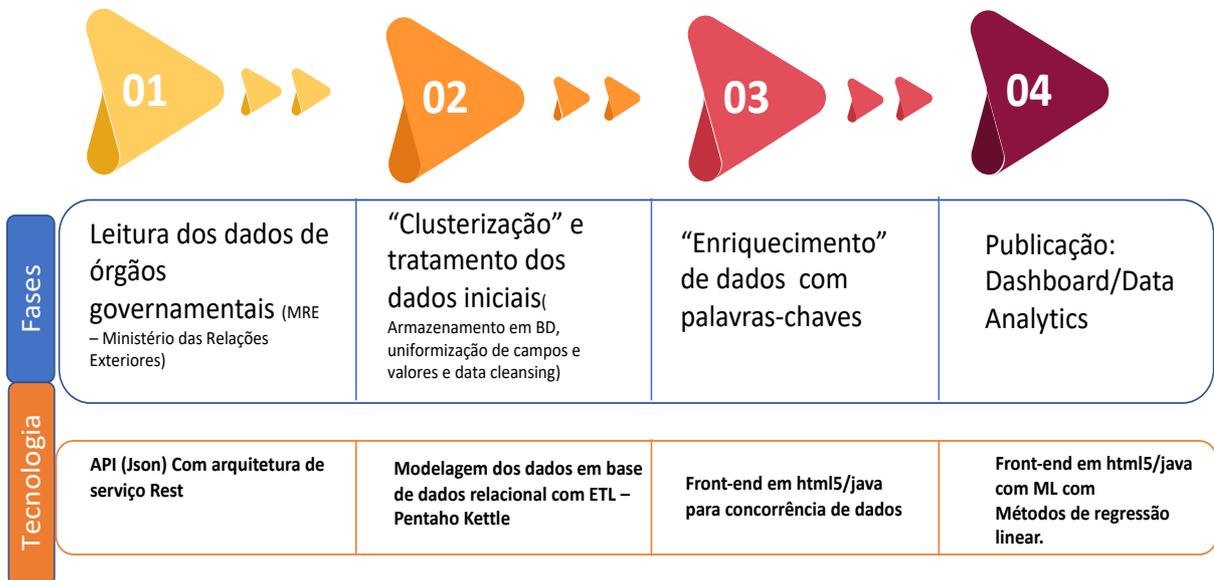
Depois que os arquivos são baixados e armazenados, são tratados, estruturados e transferidos para tabelas em um banco de dados para que possam ser processados. Esse processo é feito por meio de scripts e instruções de banco de dados SQL.

Em seguida, os dados são transformados através de palavras-chaves hierarquizadas, as quais serão constantemente enriquecidas através de algoritmos de *Machine Learning* (ML), que irão determinar informações cruciais para as análises de participação do mercado.

Serão aplicados conceitos de *Big Data Analytics*, também por meio de instruções SQL, para apresentação de *dashboards* executivos de posicionamento e participação de mercado, que auxiliarão nas tomadas de decisões.

Uma vez identificada a participação de mercado de um determinado produto, a plataforma oferecerá a possibilidade de simulação de crescimento de mercado, aumentando a demanda e irá sugerir o nível ideal de inventário utilizando o método de regressão linear, proporcionando ao time de *Supply Chain* insights para o planejamento dos produtos importados. Finalmente, todas as informações serão utilizadas e disponibilizadas através de ferramentas de *Data Analytics*. A Figura 14 mostra os passos, desde a obtenção dos dados primários, a transformação, processamento e enriquecimento dos dados, até a disponibilização para visualização.

Figura 14 – Fases dos Processos e Tecnologias da Solução Computacional



Fonte: Resultado da pesquisa

Passo 1: Nesse passo é efetuada a busca dos dados da RFB por meio de ferramentas de RPA, que serão detalhadas na seção de componentes da solução computacional. Uma vez que os dados são obtidos, são armazenados em uma pasta do servidor para que sejam processados pela aplicação.

Passo 2: Nesse passo, por meio de uma ferramenta *open-source* de ETL (Pentaho), que também será detalhada na próxima seção, os dados são transformados de acordo com as regras de negócios estabelecidas e armazenados nas tabelas de banco de dados da solução computacional.

Passo 3: Nesse passo, por meio de algoritmos de *Machine Learning* de classificação, as palavras-chaves são enriquecidas, para que possam ser identificados corretamente os equipamentos e seus fabricantes correspondentes.

Passo 4: Nesse passo, instruções SQL e regras de *Big Data Analytics*, serão apresentados *dashboards* executivos, referentes ao posicionamento de mercado de cada fabricante e equipamentos importados. Dessa maneira, nesse passo também é possível ao executivo simular um crescimento de mercado e, por meio de algoritmos de *Machine Learning*

de regressão linear simples, é possível determinar os níveis ideais de inventários de acordo com a demanda projetada.

Na próxima seção, serão detalhados os componentes da solução computacional.

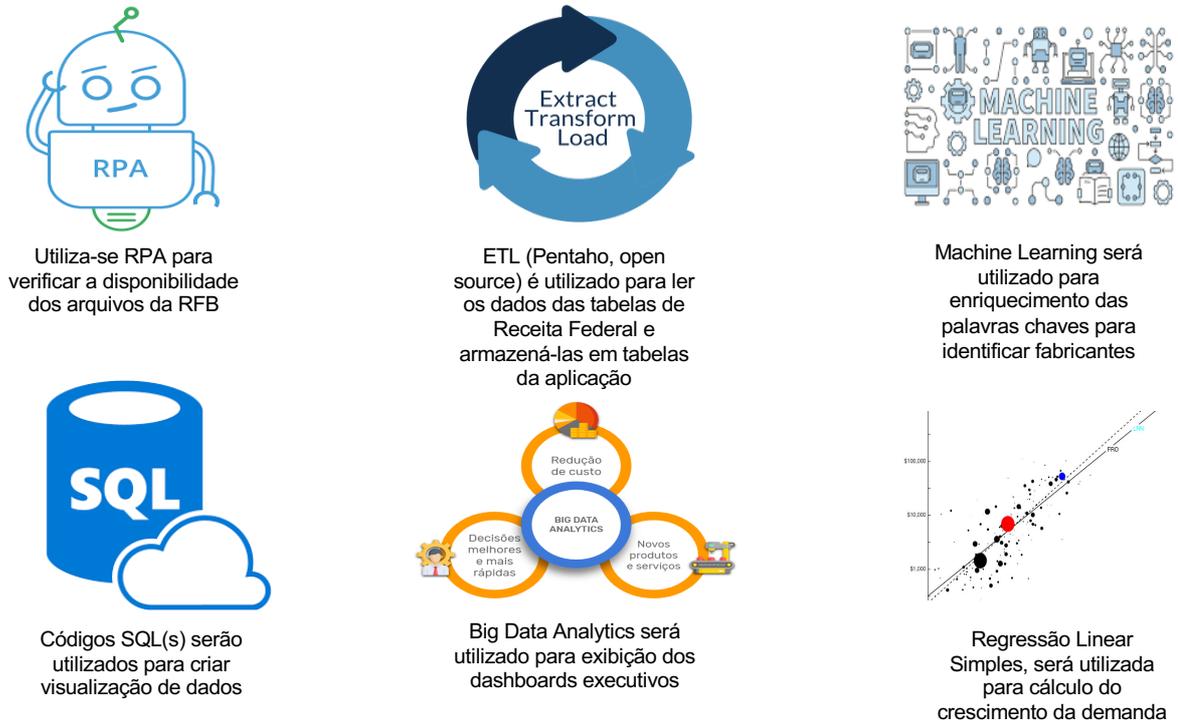
3.2 Componentes da Solução Computacional

Para a construção do protótipo da aplicação, foram utilizados alguns componentes de tecnologia, com foco principal em ferramentas *open-source* e ferramentas que são referências em seus segmentos, os quais serão descritos e documentados nessa seção. Foram utilizados componentes de RPA para monitorar a disponibilização do arquivo da RFB e posteriormente efetuar o download dos arquivos através dos capítulos de NCM referentes aos equipamentos médico-hospitalares. Posteriormente, foi utilizada uma ferramenta de ETL (*Extract/Transform/Load*), para extrair os dados, transformar e carregar nas tabelas de banco de dados da plataforma. A ferramenta utilizada para execução das funções de ETL foi o Pentaho, uma ferramenta *open-source*.

Os dados primários disponibilizados pela RFB não identificam a empresa que está importando o produto, existe apenas uma descrição completa do produto importado. Através de uma inteligência de palavras-chaves é possível identificar os produtos que fazem parte do universo da pesquisa, bem como a empresa importadora, possibilitando análises detalhadas e o cálculo da participação de mercado. As palavras-chaves constantemente são enriquecidas através de algoritmos de *Machine Learning*.

Esses algoritmos são pré-definidos e disponíveis através de uma biblioteca da linguagem de programação *Python*. Esses algoritmos de *Machine Learning* supervisionados, também serão utilizados para a identificação de padrões/anomalias nas séries históricas dos produtos, identificando, por exemplo, aspectos de sazonalidade na demanda, juntamente com a regressão linear para cálculo dos níveis de inventários adequado, baseados em uma simulação de crescimento de mercado.

Figura 15 – Componentes da Solução Computacional



Fonte: Resultado da pesquisa

3.2.1 Exemplo das tabelas utilizadas na Solução Computacional

A seguir, serão apresentados exemplos de tabelas utilizadas na solução computacional, como mostram o Quadro 7 - *Dados da tabela de Importação RFB*, o Quadro 8 - *Dados da tabela de palavras-chaves*, e a Figura 16, com um exemplo da tabela de palavras preenchida.

Quadro 7 – Dados da tabela de Importações da RFB

#Id	Campo	Descrição
1	Número de Ordem	Número único para identificação dos processos de importação e exportação registrado pela RFB
2	Ano/Mês	Ano e mês que o processo foi registrado
3	Cod. NCM	NCM (Nomenclatura Comum do Mercosul) do Produto
4	Descrição Código NCM	Descrição da NCM
5	País	Código do país de origem do produto
6	País de Origem	Descrição do país de origem do produto
7	País	Código do país de aquisição do produto
8	País de Aquisição	Descrição do país de aquisição do produto
9	Und. Estat.	Código da unidade de medida

10	Unidade de Medida	Descrição da unidade de medida
11	Unidade Comerc.	Descrição da unidade de medida de comercialização do produto
12	Descrição do Produto	Descrição completa do produto
13	Qtde. Estatística	Quantidade importada
14	Peso Líquido	Peso líquido do produto
15	VMLE Dólar	Valor total do produto importado em dólar
16	VI. Frete Dólar	Valor do frete em dólar
17	VI. Seguro Dólar	Valor do seguro em dólar
18	Valor Un. Prod. Dólar	Valor da unidade do produto em dólar
19	Qtde. Comercial	Quantidade em que o produto é comercializado
20	Tot. Un. Prod. Dólar	Valor total do produto em dólar = Valor Un. Prod. Dólar * Qtde comercial
21	Unidade Desembarque	Unidade federativa de desembarque
22	Unidade Desembarço	Unidade federativa do desembarço do processo de importação
23	Incoterm	Termos internacionais de comércio. Exemplos: FOB, CIF, CIP, EXW, DAT, etc.
24	Nat. Informação	Natureza da informação: “efetiva” ou “especial”
25	Situação do Despacho	Situação atual do processo de importação

Fonte: Resultado da pesquisa

Quadro 8 – Dados da tabela de Palavras-chaves

#Id	Campo	Descrição
1	NCM	NCM (Nomenclatura Comum do Mercosul) do produto
2	Hierarquia	Hierarquia de relevância da palavra-chave, quanto menor, mais relevância possuirá
3	Palavra-Chave	Palavra que será utilizada para incluir ou excluir um registro de acordo com a hierarquia
4	Unid. Neg.	Relacionar o produto do registro selecionado à uma unidade de negócios da empresa
5	Sub. Unid. Neg	Relacionar o produto do registro selecionado à uma subunidade de negócios da empresa
6	Segm.	Relacionar o produto do registro selecionado ao segmento de comercialização
7	Palavra Exclusão	Caso o indicador seja de exclusão, incluir aqui a palavra para exclusão
8	Forncec	Relacionar o produto do registro selecionado Fornecedor que importa e comercializa o produto

Fonte: Resultado da pesquisa

Figura 16 – Exemplo da Tabela de Palavras-chaves preenchida

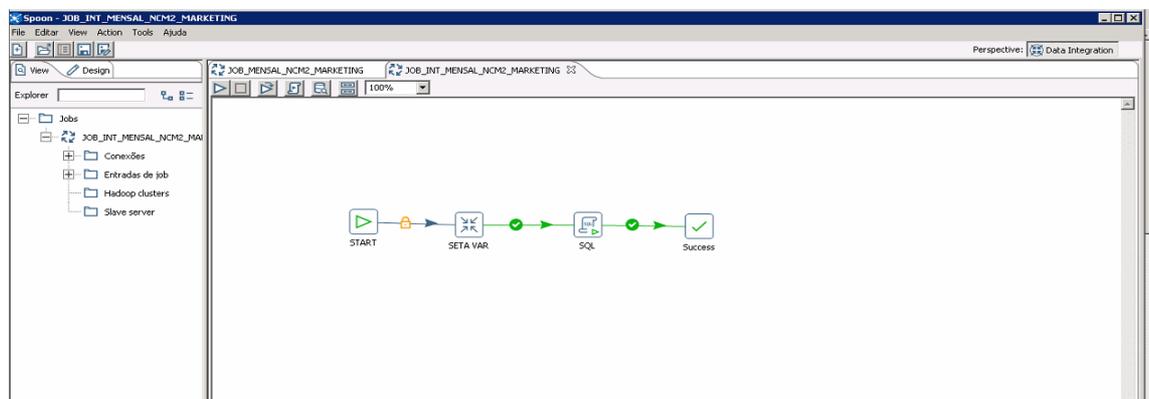
NCM	Hierarquia	PalavraChave	UnidNeg	SubUnidNeg	Segm	PalavraExclusao	Fornec
49019900	1	9938660	BID	Molecular	Termocicladores		Thermo Fisher
49019900	1	9938673	BID	Molecular	Termocicladores		Thermo Fisher
49019900	4	termociclador	BID	Molecular	Termocicladores		Others
49019900	4	THERMAL CYCLER	BID	Molecular	Termocicladores	BAX Q7	Others
84198940	1	4342718	BID	Molecular	Termocicladores		Thermo Fisher
84198940	1	4452300	BID	Molecular	Termocicladores		Thermo Fisher
84198940	1	4479071	BID	Molecular	Termocicladores		Thermo Fisher
84198940	1	Veriti DX	BID	Molecular	Termocicladores		Thermo Fisher
84198940	3	biorad	BID	Molecular	Termocicladores		biorad
84198940	3	Bio-Rad	BID	Molecular	Termocicladores		Bio-Rad
84198940	3	eppendorf	BID	Molecular	Termocicladores	qPCR	eppendorf
84198940	3	jena	BID	Molecular	Termocicladores		Analytik Jena
84198940	3	Thermo Fisher	BID	Molecular	Termocicladores		Thermo Fisher
84198940	3	Thermo Scientific	BID	Molecular	Termocicladores		Thermo Fisher
84198940	4	termociclador	BID	Molecular	Termocicladores		Others
84198940	4	THERMAL CYCLER	BID	Molecular	Termocicladores	BAX Q7	Others
84198999	1	T100	BID	Molecular	Termocicladores		Bio-Rad
84198999	1	185-1138	BID	Molecular	Termocicladores		Bio-Rad
84198999	1	186-1096	BID	Molecular	Termocicladores		Bio-Rad
84198999	1	3PrimeG Thermal Cycler	BID	Molecular	Termocicladores		Cole-Parmer Ltd
84198999	1	3PrimeX Thermal Cycler	BID	Molecular	Termocicladores		Cole-Parmer Ltd

Fonte: Resultado da pesquisa

3.2.2 Ferramenta de ETL – Pentaho Data Integration ou Kettle

O *Pentaho Data Integration*, ou *Kettle*, é uma ferramenta ETL (*Extraction, Transformation, and Loading*). Na Figura 17, é apresentado o *workflow* de uma geração simples, criada pelo *Pentaho Data Integration*.

Figura 17 – Exemplo da importação e transformação do arquivo da RFB com PDI



Fonte: Resultado da pesquisa

3.2.3 Biblioteca de Machine Learning da Linguagem Python

Existem dois tipos de algoritmos de *Machine Learning* supervisionados: regressão e classificação. O primeiro prevê saídas de valor contínuo, enquanto o segundo prevê saídas discretas. Por exemplo, prever o preço de uma casa em dólares é um problema de regressão, enquanto prever se um tumor é maligno ou benigno ou categorização de palavras é um problema de classificação.

Nesta subseção, serão apresentados os conceitos da biblioteca, tanto no que se refere à aplicação do algoritmo de regressão linear, como o de classificação das palavras-chaves, bem como a aplicação dos mesmos na solução computacional.

Primeiramente, será apresentada a biblioteca de regressão linear e como ela pode ser implementada para duas e várias variáveis usando o *Scikit-Learn*, que é uma das bibliotecas de *Machine Learning* mais populares e utilizada para *Python*. Essa biblioteca foi utilizada na solução computacional para calcular os níveis ideais de inventários baseado em uma demanda estimada de acordo com o crescimento de mercado estimado.

Será apresentado o código fonte da biblioteca e sua chamada, implementados na solução computacional, conforme etapas descritas a seguir.

Primeiramente, temos que carregar todas as bibliotecas:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as seabornInstance
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
%matplotlib inline
```

O seguinte comando importa o CSV com as variações de estoques para o *dataset pandas*:

```
dataset = pd.read_csv('/Documents/projects/ML/ML_
_LinearRegression/estoques.csv')
```

Definindo variáveis e preparando para execução do algoritmo de Regressão Linear:

```
X = dataset['MinStock'].values.reshape(-1,1)
y = dataset['MaxStock'].values.reshape(-1,1)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=0)

regressor = LinearRegression()
regressor.fit(X_train, y_train) #treinando o algoritmo

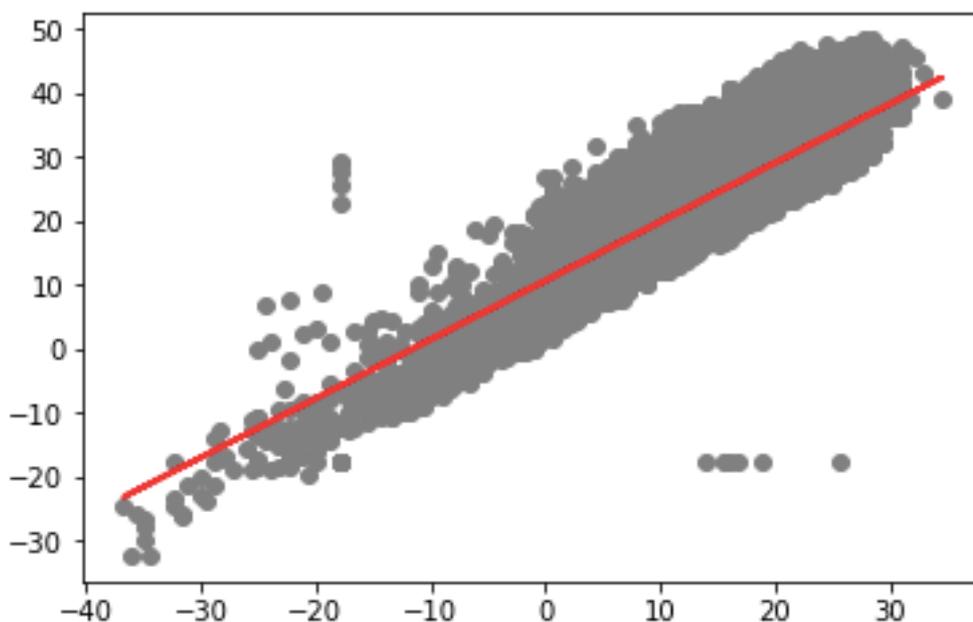
y_pred = regressor.predict(X_test)

plt.scatter(X_test, y_test, color='gray')
plt.plot(X_test, y_pred, color='red', linewidth=2)
plt.show()
```

Importante ressaltar: *Scikit-learn* é um pacote de bibliotecas *Python* que contém a implementação de muitos algoritmos de *Machine Learning*. *NumPy* é o pacote básico da linguagem *Python* que suporta vetores, matrizes e funções matemáticas. *Pandas* é uma biblioteca *Python* que fornece ferramentas de análise de dados e estruturas de dados de alta performance e fáceis de usar. *SciPy* é um pacote *open source* que implementa diversas técnicas de computação científica.

A Figura 18 mostra o resultado da execução do algoritmo de Regressão Linear, utilizando a tabela de quantidades mínimas e máximas de inventários em uma série histórica de um ano.

Figura 18 – Exemplo do Resultado do Algoritmo de *Machine Learning* de Regressão Linear



Será apresentado o algoritmo utilizado na classificação das palavras-chaves, nesse caso também foi utilizado um algoritmo da biblioteca *Scikit-Learn*, que, como comentado anteriormente, é uma das bibliotecas de *Machine Learning* mais populares e utilizada para *Python*, e foi adotada em nossa solução computacional. Foi utilizado no desenvolvimento o algoritmo de classificação *Naive Bayes*, bastante utilizado para classificação de textos e palavras. A linguagem *Python* e as bibliotecas podem ser obtidas pelo website: <https://www.python.org>.

Será apresentado o código fonte da biblioteca e sua chamada, implementados na solução computacional, conforme etapas descritas a seguir.

```
from sklearn.naive_bayes import MultinomialNB
```

Essa linha de comando importa o algoritmo *Naive Bayes* para que possa ser utilizado em nosso protótipo.

Esse algoritmo tem dois parâmetros de matrizes, um chamado *features*, que são as características em inglês, onde colocamos as palavras dos produtos (equipamentos). E tem o parâmetro *labels*, que são as etiquetas ou resultados em inglês, onde definimos se pertence ou não a um importador.

Os próximos comandos são para treinar o algoritmo e depois avaliar o resultado se a palavra/produto pertence ou não ao importador:

```
Clf = MultinomialNB() */ Essa linha cria um objeto
classificador
Clf.Fit(Features, labels) */ Essa linha treina o algoritmo,
passando as características e os resultados obtidos nos dados.
Pred = clf.predict (palavras_produto, importador) */ aqui
define se as palavras-chaves pertencem ao fabricante importador.
```

Dessa maneira, cada vez mais teremos palavras-chaves específicas determinando o fabricante importador, aperfeiçoando cada vez mais a acurácia do cálculo do *Market Share*.

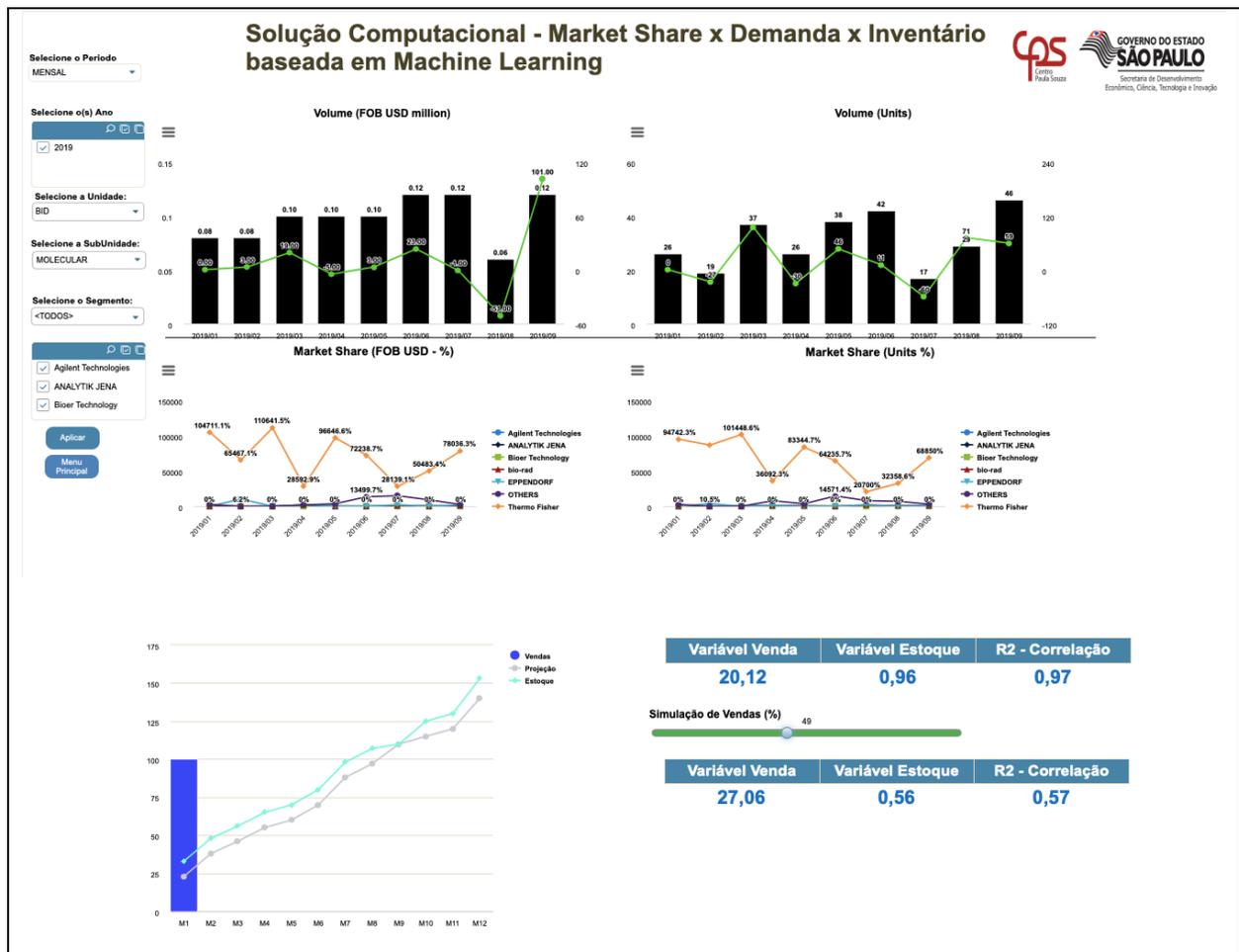
O nível de hierarquia das palavras-chaves varia em uma escala de 5 a 1, onde o 5 é mais genérico e o 1 mais específico, e que já determina o fabricante/importador. O objetivo do *Machine Learning* é que cada vez mais as palavras sejam de hierarquia 1, determinando de imediato o fabricante.

3.2.4 Solução Computacional em perspectiva

O protótipo permite que o usuário selecione o período que deseja analisar, em seguida a linha de produtos, bem como a comparação entre os fabricantes/concorrentes. Com os parâmetros definidos, a solução exibirá graficamente a participação de mercado de cada fornecedor em quantidades e unidade monetária (dólar).

Uma vez que esses dados sejam apresentados, será solicitado que o usuário da plataforma indique um percentual de crescimento de mercado, e, baseado nesse parâmetro, na base inferior da tela será exibida uma série temporal com a demanda e inventários sugeridos para esse produto com mais um ano de visibilidade. E no quadrante ao lado serão exibidos os padrões/anomalias identificadas na série temporal através do algoritmo de *Machine Learning*, como mostra a Figura 19.

Figura 19 – Demonstração da Solução Computacional



Fonte: Resultado da pesquisa

A solução computacional foi desenvolvida utilizando códigos em JAVA, HTML5 e instruções SQL, *Structured Query Language*, Linguagem de Consulta Estruturada ou SQL, que é a linguagem de pesquisa declarativa padrão para banco de dados relacional (base de dados relacional). Os dados brutos foram obtidos através do website da Receita Federal Brasileira, onde são publicados mensalmente. Para acessar os dados, basta acessar o seguinte endereço do website da RFB: <http://receita.economia.gov.br/dados/resultados/comercio-exterior>. O período utilizado foi de janeiro a outubro de 2019.

Posteriormente, esses dados foram enriquecidos através algoritmos de *Machine Learning* e apresentados na forma de *dashboards*, como mostra a Figura 19.

Para o cálculo de regressão linear simples, para obter os níveis ideais de inventário, baseado em uma série histórica, vamos utilizar a biblioteca de *Machine Learning* da linguagem de programação *Python*, conhecida como *Scikit-learn*. Por meio dela, pode-se implementar várias funções de regressão.

A solução demonstrou uma acurácia de 90% no cálculo do *Market Share*. Esse resultado foi obtido por meio dos registros de importações da empresa fabricante, que testou a solução, e, comparados com os dados disponibilizados pelos dados da RFB, chegou-se a 100% de acurácia.

O protótipo foi avaliado de duas formas: quanto à acurácia dos resultados e quanto à usabilidade e utilidade.

Com relação aos resultados, o protótipo demonstrou uma acurácia de 90% no cálculo do *Market Share*. Esse resultado foi obtido por meio dos registros de importações da empresa fabricante, que testou a solução, e foram comparados com os dados disponibilizados pelos dados da RFB.

No Quadro 9, são apresentadas algumas colunas de um arquivo real, obtido por meio do site da RFB, onde uma das colunas chaves é o número da ordem ou número do processo de importação. Com os registros de importações da empresa é feita uma validação dos processos realizados no período, e executa-se uma comparação com os dados do arquivo da receita, garantindo que 100% das informações e valores de importações estão corretos.

Quadro 9 – Exemplo: algumas colunas do arquivo da RFB

NUMERO DE ORDEM	ANOMES	COD.NCM	DESCRICAÇÃO DO CODIGO NCM	PAIS	PAIS DE ORIGEM	PAIS	PAIS DE AQUISICAO
29830000100001	201812	27011200	HULHA BETUMINOSA, NVO AGLOMERA	249	ESTADOS UNIDOS	249	ESTADOS UNIDOS
29830000200001	201812	27011200	HULHA BETUMINOSA, NVO AGLOMERA	249	ESTADOS UNIDOS	249	ESTADOS UNIDOS
24140000100001	201812	27011200	HULHA BETUMINOSA, NVO AGLOMERA	249	ESTADOS UNIDOS	249	ESTADOS UNIDOS

Fonte: Resultado da pesquisa

Essa validação somente é possível para a empresa usuária do protótipo em virtude das informações disponíveis para as validações.

Para os demais fabricantes, que são concorrentes da empresa usuária do protótipo, a acurácia é estimada entre 75% e 80%, podendo melhorar de acordo com o nível de aperfeiçoamento das palavras-chaves. Esses dados são validados pelos dados públicos de vendas de equipamentos desses fabricantes, disponibilizados na internet.

Com relação à usabilidade e utilidade, foi realizada uma avaliação com 3 usuários chaves do protótipo, buscando avaliar as funcionalidades/características principais do protótipo. O Quadro 10 exibe um resumo das avaliações efetuadas e os resultados obtidos:

Quadro 10 – Avaliação das usuárias chaves do protótipo

Função	Formação	Tempo de Experiência (em anos)	Tempo de Empresa (em anos)	Funcionalidade/Características	1 - Atende Totalmente 2 – Atende Parcialmente 3 – Não Atende
Sales Operations Manager	Gestão estratégica de vendas	+20	+15	Cálculo do <i>Market Share</i>	1
				Projeção de crescimento e demanda	1
				Níveis de inventários	2
				Usabilidade	1
				Facilidade	2
Marketing Developer Manager	Biologia aplicada a saúde	+15	+8	Cálculo do <i>Market Share</i>	1
				Projeção de crescimento e demanda	1
				Níveis de inventários	2
				Usabilidade	1
				Facilidade	1
Strategic & Innovation Manager	Adm. negócios	+10	+3	Cálculo do <i>Market Share</i>	1
				Projeção de crescimento e demanda	1
				Níveis de inventários	2
				Usabilidade	1

				Facilidade	2
--	--	--	--	------------	---

Fonte: Resultado da Pesquisa

Foi elaborado também um resumo com principais comentários de dois usuários chaves do protótipo, que se encontra no Anexo C. Nesses comentários, são apresentados aspectos referentes à situação anterior da divisão e posteriormente à implementação do protótipo.

O Apêndice D apresenta como cada componente foi desenvolvido, incluindo a utilização das sentenças SQL, bem como a biblioteca de *Machine Learning*.

CONCLUSÃO

Essa pesquisa teve como objetivo principal o desenvolvimento de um protótipo de solução computacional que seria responsável pelo cálculo do *Market Share* de empresas importadoras de equipamentos médico-hospitalares, bem como calcular a demanda de equipamentos com base em simulações de crescimento de mercado.

O mercado de saúde foi o escolhido em virtude da relevância devido ao crescimento exponencial nos últimos anos e sua representatividade no PIB (Produto Interno Bruto) do Brasil. Também pelo situação atual de uma pandemia mundial.

Segundo a ANAHP (2020), as despesas com saúde no Brasil em 2018 foram da ordem de R\$ 637 bilhões, correspondendo a 9,3% do PIB, e esses gastos estão distribuídos nos diversos setores da saúde, da seguinte forma:

- a) Gastos Públicos – R\$ 273,30 bilhões – 43%
 - a. Governo Federal – R\$ 118,10 bilhões – 19%
 - b. Governo Estadual – R\$ 70,30 bilhões – 11%
 - c. Governo Municipal – R\$ 84,90 bilhões – 13%

- b) Gastos Privados – R\$ 363,70 bilhões – 57%
 - a. Gastos Particulares (*out of pocket*) – R\$ 171,60 bilhões – 27%
 - b. Saúde Complementar – R\$ 192,10 bilhões – 30%

Essa estimativa foi calculada com base nos dados da Secretaria do Tesouro Nacional, OMS e ANS. Estão disponíveis no website da ANAHP e podem ser acessados pelo endereço: <https://www.anahp.com.br/dados-do-setor/>.

Nesse momento de pandemia, causado pelo coronavírus (COVID-19), muitas instituições da área de saúde estão fazendo importações de vários equipamentos médico-hospitalares, além de outros tipos de EPI (s) (Equipamentos de Proteção Individual). Dessa maneira, o protótipo ajudaria as empresas a mapear o mercado e criar estratégias para atender a demanda do mercado. Outra forma de ajudar seria permitindo às empresas que necessitam de equipamentos e EPI (s) encontrarem importadores desses equipamentos.

O desenvolvimento do protótipo foi efetuado por meio de um acordo de cooperação tecnológica com a empresa Falcon Analytics.

Foram implementados no protótipo componentes que o diferenciam dos principais concorrentes de mercado, como, por exemplo: conceitos de RPA, algoritmos de *Machine Learning*, de regressão e classificação e, para exibição dos resultados, conceitos de *Big Data Analytics*.

O protótipo foi implementado em uma das unidades importadoras de uma empresa de equipamentos médico-hospitalares e o resultado foi considerado satisfatório, por meio de uma pesquisa com o usuário chave, com respeito às funcionalidades, facilidade de utilização e interpretação dos dados que foram validados pela área de marketing. Ver depoimento na seção de apêndices, Apêndice C.

Limitações:

O foco da pesquisa para construção do protótipo foi baseado nas informações de equipamentos importados para o mercado de saúde brasileiro. Existe a possibilidade de estender a pesquisa para outros tipos de produtos importados, bem como para outros países que mantêm dados governamentais abertos.

Trabalho Futuro:

O protótipo da solução computacional está baseado, como foi descrito na dissertação, nos dados da Receita Federal do Brasil (RFB), e poderá ser base para estudos futuros, onde a solução poderá expandida para outros países da América Latina, como: Chile, México, entre outros que mantêm dados governamentais abertos.

REFERÊNCIAS

- AGUIRRE, S.; & RODRIGUEZ, A. Automation of a business process using robotic process automation (RPA): A case study. In: FIGUEROA-GARCÍA, J.; E. LÓPEZ-SANTANA, E.; VILLA-RAMÍREZ, J.; & FERRO- ESCOBAR, R. (Eds.). **Applied computer sciences in engineering**. Berlin: Springer, pp. 65-71, 2017.
- AGUNE, R. M.; GREGORIO FILHO, A. S.; BOLLIGER, S. P. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. In: **Congresso Consad de Gestão Pública**, III, Brasília, 2010.
- ANAHP. **Observatório ANAHP**. São Paulo, SP, 2015. Disponível em: http://www.anahp.com.br/files/observatorio_anahp_2015_miolo_210x297_web.pdf . Acesso em 14 de fevereiro de 2015.
- ANAHP. **Observatório ANAHP**. São Paulo, SP, 2020. Disponível em: <https://www.anahp.com.br/dados-do-setor/>. Acesso em 14 de maio de 2020.
- ATAMAN, M. B; HEERDE, H. J. van, & MELA, C. F. The long-term effect of marketing on brand sales. **Journal of Marketing Research**, 47(5), pp. 866-882, 2010.
- BADEN-FULLER, Charles; HAEFLIGER, Stefan. Business models and technological innovation. **Long range planning**, v. 46, n. 6, pp. 419-426, 2013.
- BAIJU, N.T. **Machine Learning: The complete history in a timeline**. December 7, 2019. Disponível em: <https://roboticsbiz.com/machine-learning-the-complete-history-in-a-timeline/>.
- BISHOP, C. M. **Pattern recognition and Machine Learning**. Berlin: Springer, First Edition, Fevereiro, 2006.
- BASKERVILLE, R.; BAIYERE, A.; GREGOR, S.; HEVNER, A. and ROSSI, M. Design Science Research Contributions: Finding a Balance between Artifact and Theory. **Journal of the Association for Information Systems**, 19(5), pp. 358-376, 2018.
- BRASIL. Agência Nacional de Vigilância Sanitária – ANVISA. **Portaria no 2.043, de 12 de dezembro de 1994 e Portaria SVS no 686, de 27 de agosto de 1998**. Conceitos técnicos, produtos para área da saúde. Inciso IV do Art. 3o, Decreto no 79.094, de 5 de janeiro de 1997. Disponível em: <http://portal.anvisa.gov.br/wps/content/Anvisa+Portal/Anvisa/Inicio/Produtos+para+Saude>. Acesso em 11 de fevereiro de 2016.
- BREIMAN, Leo; FRIEDMAN, J.H.; OLSHEN, R.A. e STONE, C.J. **Classification of regression trees**. Wadsworth & Books /Cole Advanced Books & Software, Monterey, CA, 1984.
- BRONNENBERG, B. J.; MAHAJAN, V. & VANHONACKER, W. R. 2000. The emergence of market structure in new repeat-purchase categories: the in-terplay of Market Share and retailer distribution. **Journal of Marketing Research**, 37(1), pp. 16-31, 2000.

BUZZELL, R. D; GALE, B. T. & SULTAN, R. G. M. Market Share – a key to profitability. **Harvard Business Review**, 53(1), pp. 97-106, 1975.

CHANG, V.; & WILLS, G. A model to compare cloud and non-cloud storage of Big Data. **Future Generation Computer Systems**, 57, pp. 56–76, 2016.

CHESBROUGH, Henry. Business model innovation: opportunities and barriers. **Long range planning**, v. 43, n. 2, pp. 354-363, 2010.

CHOPRA, S.; MEINDL, P. **Supply chain management: strategy, planning, and operation**. New Jersey: Prentice Hall, 2001.

CROWDFLOWER. **CrowdFlower 2015 Data Scientist Report**. Online. 2016. Disponível em: <https://visit.figure-eight.com/2016-data-scientist-report>. Acesso em: 20/05/2016.

DAVIDOV, D.; TSUR, O.; RAPPOPORT, A. **Enhanced sentiment learning using twitter hashtags and smileys**. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. Proceedings of the 23rd international conference on computational linguistics: posters [S. l.], pp. 241–249, 2010.

DELOITTE. Global Healthcare Outlook, 2020. **Laying a foundation for the future**. Dr. Stephanie Allen. Disponível em: <https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html>. Acesso em 05 de março de 2020.

DI MARTINO, B.; AVERSA, R.; CRETILLA, G.; ESPOSITO, A. & KOŁODZIEJ, J. Big data (lost) in the cloud. **International Journal of Big Data Intelligence**, 1(1–2), pp. 3–17, 2014.

DINIZ, V. Como conseguir dados governamentais abertos. In: CONGRESSO CONSAD DE GESTÃO PÚBLICA, III, Brasília, 2010.

EAVES, D. **The three laws of open government**. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Acesso em 30 set. 2009.

FARRIS, P. W; OLVER, J. & Kluyver, C. The relationship between distribution and Market Share. **Marketing Science**, 8(2), pp. 107-128, 1989.

FARRIS, P. W, BENDLE, N. T, PFEIFER, P. E, & REIBSTEIN, D. J. **Marketing metrics 50+ metrics every executive should master**. Upper Saddle River, NJ: Wharton School. 2006.

FU, Y.; JIANG, H. & XIAO, N. A scalable inline cluster deduplication framework for big data protection. In: **Middleware**, Springer, pp. 354–373, 2012.

GALEGALE, N. V.; FONTES, E. L. G.; GALEGALE, B. P. Uma contribuição para a segurança da informação: um estudo de casos múltiplos com organizações brasileiras. **Perspectivas em Ciência da Informação**, vol. 22, no. 3, p. 75–97, Belo Horizonte, Julho/Setembro, 2017.

GILL, T. G. & HEVNER, A. R. **A Fitness-Utility Model for Design Science Research**. ACM Transactions on Management Information Systems 4(2), p. 5, 2013.

GREGOR, S., & JONES, D. The Anatomy of a Design Theory. **Journal of the Association for Information Systems**, 8(5): Article 19, 2007.

GREGOR, S., & HEVNER, A. Positioning and Presenting Design Science Research for Maximum Impact. **MIS Quarterly** 37(2), pp. 337–355, 2013.

GUJARATI, D.N.M. **Econometria Básica. 5.ed.**, São Paulo: McGraw-Hill, 2011.

HENDERSON, B. The product portfolio. **BCG Perspectives**, pp. 66, 1-2, 1970.

HENDERSON, B. D. **The experience curve revisited: IV.** The growth share matrix of the product portfolio. Boston: Boston Consulting Group, 1973.

HEVNER, A. R.; MARCH, S. T.; PARK, J. and RAM, S. Design Science in Information Systems Research. **MIS Q.**, vol. 28, no. 1, pp. 75–105, 2004.

HILL, C, GRIFFITHS, W, JUDGE, G. **Econometria, 3a edição.** Saraiva 2010.

HURWITZ, J., & KIRSCH, D. **Machine Learning For Dummies®, IBM Limited Edition.** NJ, Published by John Wiley & Sons, Inc, pp. 3-20, 2018.

JIANG, P.; WINKLEY, J.; ZHAO, C.; MUNNOCH, R.; MIN, G. & YANG, L.T. **An intelligent information forwarder for healthcare big data systems with distributed wearable sensors.** in *IEEE Systems Journal*, vol. 10, no. 3, pp. 1147-1159, 2014.

KERREMANS, M. **Gartner market guide for process mining.** Report G00353970. Gartner, 2018.

KOHAVI, Ron e PROVOST, Foster. **Glossary of terms: Machine Learning 30.** Disponível em: <http://bit.ly/2CIYKNC>, 1998.

KOHONEN, Teuvo. **Self-Organization and Associative Memory.** Springer-Verlag Series in Information Science, 1987.

KONSTAN, J. & RIEDL, J. Deconstructing recommender systems: how Amazon and Netflix predict your preferences and prod you to purchase. **IEEE Spectrum**, 2012.

KOTLER, P, KELLER, K. **Administração de marketing.** 14a Edição. São Paulo: Pearson Education do Brasil, 2012.

Lacity, M. C., & Willcocks, L. P. Robotic process automation at Telefonica O2. **MIS Quarterly Executive**, 15(1), pp. 21-35, 2016.

LAPORTE, T.; DEMCHAK, C.; FRIIS, C. Webbing governance: national differences in constructing the face of public organizations. In: GARSON, G. **Handbook of public information systems.** New York: Marcel Dekker Publishers, 2000.

LE CLAIR, C. **The Forrester wave, robotic process automation,** the 12 providers that matter most and how they stack up. Forrester, Cambridge, 2017.

- MARCH, S., & SMITH, G. Design and Natural Science Research on Information Technology. **Decision Support Systems** **15**, pp. 251-266, 1995.
- MCCARTHY, J. **Circumscription—A Form of Non-Monotonic Reasoning**. **Artificial Intelligence** **13**(1-2), pp. 27-39, 1980.
- MENGER, V. Deduce: A pattern matching method for automatic de-identification of dutch medical text. **Telematics and Informatics**, Elsevier, v. 35, n. 4, pp. 727-736, 2018.
- MITCHEL, Tom. Machine Learning. **McGraw-Hill Science/Engineering/Math**. Disponível em: <http://www.cs.ubbcluj.ro/~gabis/ml/ml-books/McGrawHill%20-%20Machine%20Learning%20- Tom%20Mitchell.pdf>. 1997.
- OPEN GOV DATA. **Eight principles of open government data**. Disponível em: https://public.resource.org/8_principles.html. Acesso em: 8 dez. 2007.
- OSORIO, Fernando Santos. **Um Estudo sobre Reconhecimento Visual de Caracteres através de Redes Neurais**. Dissertação de Mestrado, CPGCC, UFRGS, Porto Alegre - Brasil. Outubro 1991.
- PAK, A., PAROUBEK, P. Twitter based system: Using twitter for disambiguating sentiment ambiguous adjectives. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 5th International Workshop on Semantic Evaluation**. [S.l.], p.436-439, 2010.
- PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A. and CHATTERJEE, S. A Design Science Research Methodology for Information Systems Research, **J. Manag. Inf. Syst.**, vol. 24, no. 3, pp. 45-77, 2007.
- PRESSMAN, Roger S. Engenharia de Software. **ArtMed**, 7. ed. VitalSource Bookshelf Online, 09/2010.
- RECEITA FEDERAL. Disponível em: < <http://idg.receita.fazenda.gov.br/>>. Acessado em 04 de outubro de 2017.
- REHMAN, M. H.; KHAN, A. & BATOOL, A. Big Data Analytics in mobile and cloud computing environments. In: Hussain, Q. (Ed.). **Handbook of research on next-Generation high performance computing**, Vol. 1, IGI Global, 2016.
- RIES, Eric. **A startup enxuta**. São Paulo: Leya, 2014.
- RUST, R. T. & ZAHORIK, A. J. Customer satisfaction, customer retention, and Market Share. **Journal of Retailing**, **69**(2), pp. 193-215, 1993.
- SAKAKI, T.; OKAZAKI, M.; MATSUO, Y. **Earthquake shakes twitter users: real-time event detection by social sensors**. In: ACM. Proceedings of the 19th international conference on Worldwideweb [S.l.], pp.851-860, 2010.

- SALES, Alessandra Simoni Ferraz. **Logística na Cadeia de Suprimentos da Indústria Automobilística**. Belo Horizonte: Editora C/Arte, 2005.
- SAMPIERI, Roberto H.; COLLADO, Carlos F.; LUCIO, María del Pilar B. **Metodologia de pesquisa**. Porto Alegre: McGraw Hill, 2013.
- SIM, I. (2016). Two ways of knowing: Big data and evidence-based medicine. **Annals of Internal Medicine**, 164(8), pp. 562-563, 2016.
- SOUZA, Dalva I. de; MULLER, Deise M.; FRACASSI, Maria A. T.; ROMEIRO, Solange B. B. **Manual de orientações para projetos de pesquisa**. Novo Hamburgo: FESLSVC, 2013.
- SRINIVASAN, S; LESZCZYC, P. T. L. & Bass, F. M.. Market Share re- sponse and competitive interaction: the impact of temporary, evolving and structural changes in prices. **International Journal of Research in Marketing**, 17(4), pp. 281-305, 2000.
- SUTTON, Richard S. & BARTO, Andrew G. **Reinforcement Learning: An Introduction**. MIT Press (A Bradford Book), Cambridge, MA, 1998.
- TAKEDA, H.; VEERKAMP, P.; TOMIYAMA, T. and YOSHIKAWAM, H. Modeling Design Processes. **AI Magazine Winter**, pp. 37–48, 1990.
- TORNBOHM, C. **Gartner market guide for robotic process automation software**. Report G00319864. Gartner, 2017.
- TROVATI, M. Reduced topologically real-world networks: a big-data approach. **International Journal of Distributed Systems and Technologies (IJ DST)**, 6(2), pp. 13–27, 2015.
- VAN DER AALST, W.M.P. **Process mining: data science in action**. Springer, Heidelberg, 2016.
- VENABLE, J.; PRIES-HEJE, J. and BASKERVILLE, R. FEDS: A Framework for Evaluation in Design Science Research. **European Journal of Information Systems** 25(1), pp. 77-89, 2016.
- WILBUR, K. C. & FARRIS, P. Distribution and Market Share. **Journal of Retailing**, 90 (2), pp. 154-167, 2014.
- WITTEN, Ian H. e FRANK, Eibe. **Data mining: practical machine learning tools and techniques**. Elsevier, pp. 29, 2005.
- WU, C.; BUYYA, R. & RAMAMOHANARAO, K. **Big Data Analytics = Machine Learning + Cloud Computing**. 2016.
- YANG, C.; ZHANG, X.; ZHONG, C.; LIU, C.; PEI, J.; RAMAMOHANARAO, K., et al. A spatiotemporal compression-based approach for efficient big data processing on cloud. **Journal of Computer and System Sciences**, 80(8), pp. 1563–1583, 2014.
- ZHAI, Y.; ONG, Y.-S. & TSANG, I. W. The emerging big dimensionality. **Computational Intelligence Magazine, IEEE**, 9(3), pp. 14–26, 2014.

APÊNDICES

APÊNDICE A – Roteiro de entrevistas para determinar a usabilidade e aplicação do protótipo

1. Em quais mercados a empresa atua? Existe algum mercado foco?
2. Quais são os principais clientes (cite três)? E quais são os segmentos de mercado?
3. Quem são os principais concorrentes?
4. A empresa atualmente utiliza alguma ferramenta (de mercado ou desenvolvida internamente) para medição de posicionamento de mercado *Market Share* dos equipamentos importados? Se sim, qual?
5. Existe uma área de inteligência de mercado ou marketing que avalia essas informações?
6. Essas informações são consideradas relevantes para o desenvolvimento de ações de mercado? De que forma?
7. Seria útil uma ferramenta que buscasse e calculasse essas informações automaticamente? E que pudesse sugerir insights de ações?
8. Que informações/análises considera importante que estivessem disponíveis em uma ferramenta com essas características?
9. Qual a posição de mercado da empresa? Se não for a líder? Sabe que é a empresa líder? E por que acredita que está na liderança?
10. Com que frequência a empresa faz análises internas/externas para entender o desempenho dos negócios e trabalhar no desenvolvimento de vantagens competitivas?
11. Acredita que conhecer o mercado e os concorrentes seja importante?
12. Tem conhecimento de alguma ferramenta de mercado que proporcione essas informações?
13. Se a empresa decide investir para obter essas informações processadas mensalmente, quanto acredita que seria um custo factível?
 - a. 500 – 1.000 reais/mês
 - b. 1.001 – 2.000 reais/mês
 - c. 2.001 – 3.000 reais/mês

APÊNDICE B – Outras informações dos principais concorrentes do Protótipo

Concorrente 1: A Figura 20 mostra o primeiro concorrente da solução computacional, que é o FuncEx. O FuncexData é uma ferramenta de apoio a profissionais e instituições envolvidas na promoção e no desenvolvimento do comércio exterior brasileiro.

Endereço da página na web: <http://www.funcexdata.com.br>.

Figura 20 – FuncEx, concorrente da Solução Computacional

FUNCEX data | **FUNCEX** fundação centro de estudos do comércio exterior | Ajudando o Brasil a expandir fronteiras

HOME | BASE DE DADOS | INDICADORES | CÂMBIO | LINKS | PAÍSES

FuncexData

Torne-se um assinante e tenha acesso à sua consulta

O **FuncexData** é uma ferramenta de inteligência comercial desenvolvida para instituições cuja missão contemple a promoção do comércio exterior, a internacionalização das empresas brasileiras ou a venda de serviços a empresas exportadoras ou importadoras. Para tanto, o **FuncexData** disponibiliza, de forma ágil e interativa, dados e informações que possibilitam a elaboração de análises e diagnósticos, o planejamento de ações e tarefas, a seleção e identificação de públicos-alvo, além de outras iniciativas que demandem embasamento estatístico e uso de inteligência comercial. O **FuncexData** contempla dois ambientes diferenciados:

- O primeiro, de acesso irrestrito ao público, é composto por informações, dados e séries estatísticas que podem ser consultados de forma gratuita através dos canais: **INDICADORES, CÂMBIO, LINKS e PAÍSES.**
- O segundo, de acesso restrito a assinantes, o canal **BASE DE DADOS** é estruturado sob a forma de três Bancos de Dados que podem ser consultados através do uso de ferramentas de busca que permitem o cruzamento de dados selecionados a partir de um amplo cardápio de categorias pré-definidas.

Os canais **Indicadores** e **Câmbio** disponibilizam todo o acervo de indicadores elaborados pela Funcex, com destaque para os índices de comércio exterior (preço e quantum). O usuário tem acesso à totalidade das séries históricas disponíveis, que são atualizadas com periodicidade mensal, trimestral ou anual, de acordo com a sua metodologia específica de elaboração.

O canal **Base de Dados** é dividido em três bancos de dados

- Estatísticas de Comércio Exterior** fornece informações de valores exportados e importados segundo a origem e destino dos fluxos (país ou bloco), em diversos níveis (total brasileiro, UF's, unidades da receitas e vias), com base em número elevado e diversificado de categorias ou tipologias dos produtos. O banco se presta, particularmente, para a elaboração de análises conjunturais ou estruturais da evolução dos fluxos de comércio exterior, assim como para o exame da composição e grau de diversificação das pautas exportadora e importadora, tanto em termos setoriais ou de produtos quanto de mercados.
- Estatísticas Gerais das Empresas Exportadoras e Importadoras:** disponibiliza valores segundo origem e destino dos fluxos e em diversos níveis para algumas categorias e tipologias de empresas exportadoras ou importadoras, inclusive segundo o tamanho da empresa.
- Cadastro Geral das Empresas Exportadoras e Importadoras** que é baseado nas atualizações periódicas realizadas pela Funcex para a elaboração dos catálogos de empresas que participam diretamente da atividade de comércio exterior. Atualmente a Funcex dispõe de dados cadastrais completos e atualizados (razão social, endereço e CEP, telefone, fax, e-mail, website e contato/ principal executivo na área de comércio exterior) de mais de 17.000 empresas exportadoras e 11.000 empresas importadoras.

Fonte: Resultado da Pesquisa

Concorrente 2: A Figura 21 mostra o segundo concorrente da solução computacional, que é o Datagro. A Datagro, mais do que uma plataforma que apresenta análises e dados primários e secundários de importação e exportação, também fornece consultoria aos seus clientes, onde o foco principal é o setor de commodities agrícolas.

Endereço da página na web: <https://www.datagro.com>

Figura 21 – Datagro, concorrente da Solução Computacional

DATAGRO Sobre a DATAGRO Portal Plataforma Contato

CONSULTORIA AGRÍCOLA INDEPENDENTE

Análises, dados primários e notícias sobre as principais commodities agrícolas mundiais

Conteúdo
Análises, notícias e dados

Portal DATAGRO

Consultoria
Agrícola, Industrial, Gestão Assistida, Governos

DATAGRO Consultoria

Educação e Eventos
Conferências, Eventos, Seminários e Cursos

DATAGRO Conferences

+35
ANOS

+40
CLIENTES EM
PAÍSES

+10
ANÁLISES EM
CULTURAS E
MERCADOS

+50K
SÉRIES DE DADOS

+130
RELATÓRIOS/MÊS

+13K
PESSOAS PRESENTES
NOS EVENTOS

+100K
ACOMPANHAM OS
EVENTOS ONLINE

Fonte: Resultado da Pesquisa

Concorrente 3: A Figura 22 mostra o terceiro concorrente da solução computacional, que é o DataImport. O DataImport gera apenas relatórios para equipes de importação, para ajudá-los a entender melhor o comportamento do mercado interno com respeito à aquisição de produtos e matérias primas importadas.

Endereço da página na web: <https://dataimport.com.br>

Figura 22 – DataImport, concorrente da Solução Computacional



DATAIMPORT
data for competitiveness

[HOME](#) [SOBRE NÓS](#) [CONTATO](#) [ACESSAR SISTEMA](#)

DATAIMPORT

VOCÊ QUER SABER QUANTO SEU CONCORRENTE ESTÁ PAGANDO POR UM PRODUTO OU MATÉRIA PRIMA? **NÓS SABEMOS!**

SEGURANÇA E CONFIABILIDADE NA SUA TOMADA DE DECISÃO

CONSULTE-NOS AGORA MESMO E SAIBA COMO UM DE NOSSOS RELATÓRIOS PODE SE TORNAR DECISIVO PARA O SEU NEGÓCIO

Fonte: Resultado da Pesquisa

APÊNDICE C – Avaliação das áreas usuárias do protótipo

Área: *Chromatography and Mass Spectrometer Division | Latin America*

Situação anterior ao uso do protótipo

O departamento não possuía gestão de inteligência de mercado e visibilidade do negócio, além do prisma de suas próprias vendas (análises de pedidos e faturamentos) e aspectos empíricos de seus próprios gestores e *stakeholders* (experiência e *feeling* de mercado). A gente poderia pedir estudos de mercados de empresas especializadas e terceiras, entretanto isso demandava um custo dispendioso e agredia nossos orçamentos, sem sabermos se o estudo era preciso de verdade. Entender esse aspecto do nosso segmento era uma dificuldade real que enfrentávamos.

Avaliação da utilização do protótipo

Visibilidade real e baseada em algoritmos matemáticos do mercado/departamento. Tamanho do mercado, participações do mercado da empresa e de seus competidores e subsídios para a elaboração de planejamentos estratégicos.

Depoimento da usuária chave do protótipo

"O projeto nos trouxe dois grandes benefícios: compartilhar os bons resultados com os gestores, mantendo as equipes motivadas e engajadas, e sinalizar a tempo os desvios frente ao planejado, possibilitando agilidade nas análises para o processo de recuperação". Sales Operation Chromatography and Mass Spectrometer Division.

Área: *Life Science Group Division | Latin America*

Situação anterior ao uso do protótipo

As informações das divisões eram restritas e alcançavam apenas dados que eram produzidos internamente, como dados de notas fiscais e pedidos de compras.

Aproximadamente 90% dos produtos são oriundos de outros países e passam por processos de importações. Necessitavam de dimensionamento e acuracidade apurada, tanto

para não haver rupturas quanto para não causar estoques altos. O grande desafio era obter do empirismo as informações de mercado associado às próprias vendas. Ter o conceito de inteligência de mercado de forma processual e sistêmica era o objetivo a perseguir.

Avaliação da utilização do protótipo

Visibilidade real e baseada em algoritmos matemáticos do mercado/departamento. Tamanho do mercado, participações do mercado da empresa e de seus competidores e subsídios para a elaboração de planejamentos estratégico. E outros benefícios, como:

- a) Visão clara para o time global;
- b) Melhor acuracidade nos tamanhos de estoque e pedidos;
- c) Projeto de preço de venda baseado em estudo e sensibilidade através deste.

Depoimento da usuária chave do protótipo

"Ter a possibilidade de mostrar informações da divisão com segurança, captar recursos de investimento embasado em informações, e entender o mercado com números, trouxe o respeito ao time local e permitiu melhorarmos os processos e aumentar equipe e verbas do time global". Market Development Manager, BioScience.

APÊNDICE D – Propriedades e código fonte dos componentes do protótipo

Seleção do Período: Essa propriedade permite ao usuário selecionar o período que deseja analisar da base de dados. As informações serão apresentadas para o usuário de três maneiras: anual, trimestral e mensal.

Abaixo, o código SQL para selecionar o período:

```
select 'ANUAL' AS PERIODO, 'ANUAL' AS PERIODO2
UNION ALL
select 'TRIMESTRAL' AS PERIODO, 'TRIMESTRAL' AS PERIODO2
UNION ALL
SELECT 'MENSAL' AS PERIODO, 'MENSAL' AS PERIODO2
```

Seleção dos anos disponíveis na base de dados: Por meio da seguinte instrução SQL, é possível efetuar uma varredura no banco de dados para verificar os dados de todos os anos disponíveis, para que sejam apresentados no quadro e o usuário possa selecionar os anos para análise.

```
SELECT DISTINCT YEAR (DATA) ANO, YEAR (DATA) ANO2
FROM FATO11_NCM2
ORDER BY 2 DESC
```

Seleção das Unidades de Negócios: Por meio da seguinte instrução SQL, é possível efetuar uma varredura no banco de dados para verificar os dados de todas as unidades de negócios disponíveis, para que sejam apresentadas no quadro e o usuário possa selecionar as unidades de negócios que deseja analisar.

Existe a possibilidade de assinalar a opção “TODOS”.

```
SELECT '<TODOS>', '<TODOS>'
UNION ALL
SELECT distinct CONVERT (VARCHAR (10), FATO6) AS COD,
CONVERT (VARCHAR (50), DESCRICAO) AS UNIDADE
FROM CAD_FATO6_NCM2
WHERE DESCRICAO IS NOT NULL AND DESCRICAO <> ''
AND FATO6 NOT IN (19, 22, 16, 25, 26)
ORDER BY 2
```

Seleção das Subunidades de negócios: Por meio da seguinte instrução SQL, é possível efetuar uma varredura no banco de dados para verificar os dados de todas as subunidades de

negócios disponíveis, para que sejam apresentadas no quadro e o usuário possa selecionar as subunidades de negócios que deseja analisar.

Existe a possibilidade de assinalar a opção “TODOS”.

```
SELECT '<TODOS>', '<TODOS>'
UNION ALL
SELECT DISTINCT CONVERT(VARCHAR(10),CF7.FATO7) AS COD,
      CONVERT(VARCHAR(50),CF7.DESCRICAO) AS UNIDADE
  FROM CAD_FATO7_NCM2 CF7 INNER JOIN FATO11_NCM2 F7 ON
(CF7.FATO7 = F7.FATO7)
  WHERE DESCRICAO IS NOT NULL AND DESCRICAO <> ''
  AND (F7.FATO6 = '{cboUnidade}' or '{cboUnidade}' =
'<TODOS>')
ORDER BY 2
```

Seleção dos Segmentos de Negócios: Por meio da seguinte instrução SQL, é possível efetuar uma varredura no banco de dados para verificar os dados de todos os segmentos disponíveis no banco de dados, para que sejam apresentados no quadro e o usuário possa selecionar os segmentos que deseja analisar. Existe a possibilidade de assinalar a opção “TODOS”.

```
SELECT '<TODOS>', '<TODOS>'
UNION ALL

SELECT DISTINCT CONVERT(VARCHAR(50),CF11.FATO11) AS COD,
      CONVERT(VARCHAR(50),CF11.DESCRICAO) AS SEGMENTO
  FROM CAD_FATO11_NCM2 CF11 INNER JOIN FATO11_NCM2 F11 ON
(CF11.FATO11 = F11.FATO11)
  WHERE DESCRICAO IS NOT NULL AND DESCRICAO <> ''
  AND CF11.FATO11 NOT IN ('exclusao')
  AND (F11.FATO6 = '{cboUnidade}' or '{cboUnidade}' =
'<TODOS>')
  AND (F11.FATO7 = '{cboSubUnidade}' or '{cboSubUnidade}'
= '<TODOS>')
ORDER BY 2
```

Seleção das Empresas com participação de mercado: Por meio da seguinte instrução SQL, é possível efetuar uma varredura no banco de dados para verificar os dados de todos os fabricantes importadores disponíveis no banco de dados, respeitando as seleções anteriores, para que sejam apresentados no quadro e o usuário possa selecionar os fabricantes que deseja analisar.

Por *default*, todos já estarão selecionados, mas existe a possibilidade de comparar fabricantes específicos, basta o usuário desmarcar o fabricante que não deseja que faça parte da análise.

```

SELECT DISTINCT CF11.FATO3      AS COD
                ,CF11.DESCRICAO AS SEGMENTO
FROM CAD_FATO3_NCM2 CF11
JOIN FATO11_NCM2     F11 ON (CF11.FATO3 = F11.FATO3)
WHERE CF11.DESCRICAO IS NOT NULL
      AND CF11.DESCRICAO <> ''
      AND F11.FATO3 <> '107'
      AND YEAR(F11.DATA) IN ({LSTANO})
      AND (F11.FATO6 = '{cboUnidade}' or '{cboUnidade}' =
'<TODOS>')
      AND (F11.FATO7 = '{cboSubUnidade}' or
'{cboSubUnidade}' = '<TODOS>')
      AND (F11.FATO11 = '{CBOSEG}' OR '{CBOSEG}' = '<TODOS>')
ORDER BY 2

```

Gráfico de volume de importações, de acordo com o período e parâmetros selecionados, será apresentado em: % volume em dólares. O conjunto de instruções SQLs abaixo tem como objetivo gerar o gráfico com o volume de importações dos produtos e outros parâmetros selecionados.

```

DECLARE @PERIODO VARCHAR(30)
SET @PERIODO = '{CBOPERIODO}'

IF @PERIODO = 'ANUAL'
BEGIN

SELECT YEAR(DATA) AS ANO
      ,ROUND(SUM(CONVERT(MONEY,A.MTR3)) / 1000000.00,2) AS
VALOR
      , (ROUND(SUM(CONVERT(MONEY,A.MTR3)) /
(SELECT ROUND(SUM(CONVERT(MONEY,A1.MTR3)),2)
FROM FATO11_NCM2 A1
WHERE (A1.FATO6 IN ('{cboUnidade}') OR
'{cboUnidade}' = '<TODOS>')
AND (A1.FATO7 IN ('{cboSubUnidade}') OR
'{cboSubUnidade}' = '<TODOS>')
AND (A1.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' =
'<TODOS>')
AND A1.FATO11 NOT IN ('exclusao')
AND A1.FATO6 NOT IN (19,22,16,25,26)
AND A1.FATO3 <> 107

```

```

        AND YEAR(A1.DATA) = YEAR(A.DATA) - 1),2) - 1) *
100.00 AS VARIACAO
FROM FATO11_NCM2 A
WHERE (A.FATO6 IN ('{cboUnidade}') OR '{cboUnidade}' =
'<TODOS>')
AND (A.FATO7 IN ('{cboSubUnidade}') OR '{cboSubUnidade}' =
'<TODOS>')
AND (A.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' = '<TODOS>')
AND A.FATO11 NOT IN ('exclusao')
AND A.FATO6 NOT IN (19,22,16,25,26)
AND A.FATO3 <> 107
AND YEAR(DATA) in ({LSTANO})
GROUP BY YEAR(DATA)
ORDER BY 1 ASC

END

IF @PERIODO = 'TRIMESTRAL'
BEGIN

SELECT CONVERT (VARCHAR(04), YEAR (DATA)) + '/' +
CONVERT (VARCHAR(01), DATEPART (QQ, DATA)) AS ANO
        , ROUND (SUM (CONVERT (MONEY, A.MTR3)) / 1000000.00, 2) AS
VALOR
        , (ROUND (SUM (CONVERT (MONEY, A.MTR3)) /
        (SELECT ROUND (SUM (CONVERT (MONEY, A1.MTR3)), 2)
        FROM FATO11_NCM2 A1
        WHERE (A1.FATO6 IN ('{cboUnidade}') OR
'{cboUnidade}' = '<TODOS>')
        AND (A1.FATO7 IN ('{cboSubUnidade}') OR
'{cboSubUnidade}' = '<TODOS>')
        AND (A1.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' =
'<TODOS>')
        AND A1.FATO11 NOT IN ('exclusao')
        AND A1.FATO6 NOT IN (19,22,16,25,26)
        AND A1.FATO3 <> 107
        AND DATEPART (QQ, A1.DATA) = CASE WHEN
DATEPART (QQ, A.DATA) <> 1 THEN DATEPART (QQ, A.DATA) - 1
        ELSE 4
        END
        AND YEAR (A1.DATA) = CASE WHEN DATEPART (QQ, A.DATA)
<> 1 THEN YEAR (A.DATA)
        ELSE YEAR (A.DATA) - 1
        END ), 2) - 1) * 100.00
AS VARIACAO
FROM FATO11_NCM2 A
WHERE (A.FATO6 IN ('{cboUnidade}') OR '{cboUnidade}' =
'<TODOS>')
AND (A.FATO7 IN ('{cboSubUnidade}') OR '{cboSubUnidade}' =
'<TODOS>')
AND (A.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' = '<TODOS>')

```

```

AND YEAR(DATA) in ({LSTANO})
AND A.FATO11 NOT IN ('exclusao')
AND A.FATO6 NOT IN (19,22,16,25,26)
AND A.FATO3 <> 107
GROUP BY CONVERT(VARCHAR(04),YEAR(DATA)) + '/' +
CONVERT(VARCHAR(01),DATEPART(QQ,DATA)),DATEPART(QQ,A.DATA),Y
EAR(A.DATA)
ORDER BY 1 ASC

```

```
END
```

```
IF @PERIODO = 'MENSAL'
BEGIN
```

```

SELECT CONVERT(VARCHAR(04),YEAR(DATA)) + '/' + RIGHT('0' +
CONVERT(VARCHAR(02),MONTH(DATA)),2) AS ANO
      ,ROUND(SUM(CONVERT(MONEY,A.MTR3)) / 1000000.00,2) AS
VALOR
      , (ROUND(SUM(CONVERT(MONEY,A.MTR3)) /
(SELECT ROUND(SUM(CONVERT(MONEY,A1.MTR3)),2)
FROM FATO11_NCM2 A1
WHERE (A1.FATO6 IN ('{cboUnidade}') OR
'{cboUnidade}' = '<TODOS>')
AND (A1.FATO7 IN ('{cboSubUnidade}') OR
'{cboSubUnidade}' = '<TODOS>')
AND (A1.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' =
'<TODOS>')
AND A1.FATO11 NOT IN ('exclusao')
AND A1.FATO6 NOT IN (19,22,16,25,26)
AND A1.FATO3 <> 107
AND MONTH(A1.DATA) = CASE WHEN MONTH(A.DATA) <>
1 THEN MONTH(A.DATA) - 1
ELSE 12
END
AND YEAR(A1.DATA) = CASE WHEN MONTH(A.DATA) <> 1
THEN YEAR(A.DATA)
ELSE YEAR(A.DATA) - 1
END ),2) - 1) * 100.00

```

```

AS VARIACAO
FROM FATO11_NCM2 A
WHERE (A.FATO6 IN ('{cboUnidade}') OR '{cboUnidade}' =
'<TODOS>')
AND (A.FATO7 IN ('{cboSubUnidade}') OR '{cboSubUnidade}' =
'<TODOS>')
AND (A.FATO11 IN ('{CBOSEG}') OR '{CBOSEG}' = '<TODOS>')
--AND A.FATO3 IN ({LSTFORNEC})
AND YEAR(DATA) in ({LSTANO})
AND A.FATO11 NOT IN ('exclusao')
AND A.FATO6 NOT IN (19,22,16,25,26)
AND A.FATO3 <> 107

```

```

GROUP BY CONVERT (VARCHAR (04) , YEAR (DATA) ) + '/' + RIGHT ('0' +
CONVERT (VARCHAR (02) , MONTH (DATA) ) , 2) , MONTH (A . DATA) , YEAR (A . DAT
A)
ORDER BY 1 ASC

END

```

Gráfico de *Market Share*, apresentando os importadores participantes, de acordo com o período e parâmetros selecionados, será apresentado em: % volume em dólares. Esse conjunto de instruções SQLs tem como objetivo gerar o gráfico com o *Market Share* dos importadores participantes dos produtos e outros parâmetros selecionados.

```

DECLARE @ANOS VARCHAR (MAX)
DECLARE @COLUNAS VARCHAR (MAX)
DECLARE @ANO VARCHAR (10)
DECLARE @PERIODO VARCHAR (30)
--DECLARE @FILTROANO VARCHAR (04)

SET @PERIODO = '{CBOPERIODO}'
--SET @FILTROANO = YEAR (GETDATE ()) - 1

IF @PERIODO = 'ANUAL'
BEGIN

SET @ANOS = ''
SET @COLUNAS = ''
DECLARE cursor_objects CURSOR FOR

SELECT DISTINCT YEAR (DATA) AS ANO
FROM FAT011_NCM2
WHERE (FATO6 = ('{cboUnidade}')) OR '{cboUnidade}' =
'<TODOS>')
AND (FATO7 = ('{cboSubUnidade}')) OR '{cboSubUnidade}' =
'<TODOS>')
AND (FATO11 IN ('{CBOSEG}')) OR '{CBOSEG}' = '<TODOS>')
AND FATO3 IN ({LSTFORNEC})
AND YEAR (DATA) in ({LSTANO})
ORDER BY 1

OPEN cursor_objects
FETCH NEXT FROM cursor_objects INTO @ANO
WHILE @@FETCH_STATUS = 0
BEGIN
SET @ANOS = @ANOS + '[' + @ANO + '], '
SET @COLUNAS = @COLUNAS + 'ROUND (ISNULL ([ ' + @ANO + '], 0), 1)
AS [ ' + @ANO + '], '
FETCH NEXT FROM cursor_objects INTO @ANO
END

```

```

CLOSE cursor_objects
DEALLOCATE cursor_objects

SET @ANOS = (SELECT SUBSTRING(@ANOS, 0, LEN(@ANOS)))
SET @COLUNAS = (SELECT SUBSTRING(@COLUNAS, 0,
LEN(@COLUNAS)))

EXEC (
'SELECT FORNEC2,'+@COLUNAS+
' FROM (SELECT YEAR(A.DATA) AS ANO,
CASE WHEN (SUM(A.MTR3)/(select sum(x.MTR3)
from FATO11_NCM2 x
WHERE YEAR(X.DATA) in ({LSTANO})
AND (x.FATO6 IN (' + '''' + '{cboUnidade}' +
'''' + ') OR ' + '''' + '{cboUnidade}' + '''' + ' =
''<TODOS>''))
and (x.FATO7 IN (' + '''' +
'{cboSubUnidade}' + '''' + ') OR ' + '''' +
'{cboSubUnidade}' + '''' + ' = ''<TODOS>''))
and (x.FATO11 IN (' + '''' + '{CBOSEG}' +
'''' + ') OR ' + '''' + '{CBOSEG}' + '''' + ' = ''<TODOS>''))
AND x.FATO3 IN ({LSTFORNEC})
AND x.FATO11 NOT IN ('exclusao')
AND x.FATO6 NOT IN (19,22,16,25,26)
AND x.FATO3 <> 107
AND YEAR(A.DATA) = YEAR(X.DATA) )) * 100 >
10.00 THEN CF3.DESCRICAO ELSE CF3.DESCRICAO END AS FORNEC2
, ROUND((SUM(A.MTR3) /
(select sum(x.MTR3)
from FATO11_NCM2 x
WHERE YEAR(X.DATA) in ({LSTANO})
AND (x.FATO6 IN (' + '''' +
'{cboUnidade}' + '''' + ') OR ' + '''' + '{cboUnidade}' +
'''' + ' = ''<TODOS>''))
and (x.FATO7 IN (' + '''' +
'{cboSubUnidade}' + '''' + ') OR ' + '''' +
'{cboSubUnidade}' + '''' + ' = ''<TODOS>''))
and (x.FATO11 IN (' + '''' + '{CBOSEG}' +
'''' + ') OR ' + '''' + '{CBOSEG}' + '''' + ' = ''<TODOS>''))
AND x.FATO11 NOT IN ('exclusao')
AND x.FATO6 NOT IN (19,22,16,25,26)
AND x.FATO3 <> 107
AND YEAR(A.DATA) = YEAR(X.DATA) )) *
100,2) AS QTDE'+
' FROM FATO11_NCM2 A INNER JOIN CAD_FATO3_NCM2 CF3 ON
(A.FATO3 = CF3.FATO3) '+
' WHERE YEAR(A.DATA) in ({LSTANO})
AND (A.FATO6 IN (' + '''' + '{cboUnidade}' + '''' + ')
OR ' + '''' + '{cboUnidade}' + '''' + ' = ''<TODOS>''))
and (A.FATO7 IN (' + '''' + '{cboSubUnidade}' + '''' +
') OR ' + '''' + '{cboSubUnidade}' + '''' + ' = ''<TODOS>''))

```

```

        and (A.FATO11 IN (' + '''' + '{CBOSEG}' + '''' + ') OR
' + '''' + '{CBOSEG}' + '''' + ' = '<TODOS>')
        AND A.FATO3 IN ({LSTFORNEC})
        AND A.FATO11 NOT IN ('exclusao')
        AND A.FATO6 NOT IN (19,22,16,25,26)
        AND A.FATO3 <> 107 ' +
' GROUP BY YEAR(A.DATA),CF3.DESCRICAO) T '+
'PIVOT ( SUM(QTDE)'+
'          FOR ANO IN ('+@ANOS+')'+
' ) AS P')

```

END

```

IF @PERIODO = 'TRIMESTRAL'
BEGIN

```

```

SET @ANOS = ''
SET @COLUNAS = ''
DECLARE cursor_objects CURSOR FOR

```

```

SELECT DISTINCT CONVERT(VARCHAR(04),YEAR(DATA)) + '/' +
CONVERT(VARCHAR(01),DATEPART(QQ,DATA)) AS TRIMESTRE
FROM FATO11_NCM2
WHERE (FATO6 = ('{cboUnidade}')) OR '{cboUnidade}' =
'<TODOS>')
AND (FATO7 = ('{cboSubUnidade}')) OR '{cboSubUnidade}' =
'<TODOS>')
AND (FATO11 IN ('{CBOSEG}')) OR '{CBOSEG}' = '<TODOS>')
AND FATO3 IN ({LSTFORNEC})
AND YEAR(DATA) in ({LSTANO})
ORDER BY 1

```

```

OPEN cursor_objects
FETCH NEXT FROM cursor_objects INTO @ANO
WHILE @@FETCH_STATUS = 0
BEGIN
SET @ANOS = @ANOS + '['+@ANO+'],'
SET @COLUNAS = @COLUNAS + 'ROUND(ISNULL(['+@ANO+'], 0),1)
AS '['+@ANO+'],'
FETCH NEXT FROM cursor_objects INTO @ANO
END
CLOSE cursor_objects
DEALLOCATE cursor_objects

```

```

SET @ANOS = (SELECT SUBSTRING(@ANOS, 0, LEN(@ANOS)))
SET @COLUNAS = (SELECT SUBSTRING(@COLUNAS, 0,
LEN(@COLUNAS)))

```

```

EXEC (
'SELECT FORNEC2,'+@COLUNAS+

```

```

' FROM (SELECT CONVERT (VARCHAR (04), YEAR (A.DATA)) + ' +
'''' + '/' + '''' + ' +
CONVERT (VARCHAR (01), DATEPART (QQ, A.DATA)) AS ANO,
CASE WHEN (SUM (A.MTR3) / (select sum (x.MTR3)
from FATO11_NCM2 x
WHERE YEAR (DATA)
IN ( ' + '{LSTANO}' +
' ) AND
(x.FATO6 IN ( ' + '''' + '{cboUnidade}' + '''' + ' ) OR ' +
'''' + '{cboUnidade}' + '''' + ' = '<TODOS>'))
and
(x.FATO7 IN ( ' + '''' + '{cboSubUnidade}' + '''' + ' ) OR ' +
'''' + '{cboSubUnidade}' + '''' + ' = '<TODOS>'))
and
(x.FATO11 IN ( ' + '''' + '{CBOSEG}' + '''' + ' ) OR ' + ''''
+ '{CBOSEG}' + '''' + ' = '<TODOS>'))
AND
x.FATO3 IN ({LSTFORNEC})
AND x.FATO11
NOT IN ('exclusao'))
AND x.FATO6
NOT IN (19,22,16,25,26)
AND x.FATO3
<> 107
AND
CONVERT (VARCHAR (04), YEAR (A.DATA)) + ' + '''' + '/' + '''' +
' + CONVERT (VARCHAR (01), DATEPART (QQ, A.DATA)) =
CONVERT (VARCHAR (04), YEAR (X.DATA)) + ' + '''' + '/' + '''' +
' + CONVERT (VARCHAR (01), DATEPART (QQ, X.DATA)) ) ) * 100 >
10.00 THEN CF3.DESCRICAO
ELSE CF3.DESCRICAO
END AS FORNEC2
, ROUND ((SUM (A.MTR3) / (select sum (x.MTR3) from
FATO11_NCM2 x
WHERE YEAR (DATA)
IN ( ' + '{LSTANO}' +
' ) AND
(x.FATO6 IN ( ' + '''' + '{cboUnidade}' + '''' + ' ) OR ' +
'''' + '{cboUnidade}' + '''' + ' = '<TODOS>'))
and
(x.FATO7 IN ( ' + '''' + '{cboSubUnidade}' + '''' + ' ) OR ' +
'''' + '{cboSubUnidade}' + '''' + ' = '<TODOS>'))
and
(x.FATO11 IN ( ' + '''' + '{CBOSEG}' + '''' + ' ) OR ' + ''''
+ '{CBOSEG}' + '''' + ' = '<TODOS>'))
AND x.FATO11 NOT
IN ('exclusao'))
AND x.FATO6 NOT
IN (19,22,16,25,26)
AND x.FATO3 <>
107

```

```

AND
CONVERT (VARCHAR(04), YEAR (A. DATA)) + ' + '''' + '/' + '''' +
' + CONVERT (VARCHAR (01), DATEPART (QQ, A. DATA)) =
CONVERT (VARCHAR (04), YEAR (X. DATA)) + ' + '''' + '/' + '''' +
' + CONVERT (VARCHAR (01), DATEPART (QQ, X. DATA)) )) * 100,2)
AS QTDE'+
'
FROM FATO11_NCM2 A INNER JOIN CAD_FATO3_NCM2 CF3
ON (A.FATO3 = CF3.FATO3) '+
'
WHERE YEAR(DATA) IN( ' + '{LSTANO}' +
') AND (A.FATO6 IN ( ' + '''' + '{cboUnidade}' +
'''' + ') OR ' + '''' + '{cboUnidade}' + '''' + ' =
''<TODOS>'' )
and (A.FATO7 IN ( ' + '''' + '{cboSubUnidade}'
+ '''' + ') OR ' + '''' + '{cboSubUnidade}' + '''' + ' =
''<TODOS>'' )
and (A.FATO11 IN ( ' + '''' + '{CBOSEG}' +
'''' + ') OR ' + '''' + '{CBOSEG}' + '''' + ' = ''<TODOS>'' )
AND A.FATO3 IN ({LSTFORNEC})
AND A.FATO11 NOT IN (''exclusao'')
AND A.FATO6 NOT IN (19,22,16,25,26)
AND A.FATO3 <> 107
'
GROUP BY CONVERT (VARCHAR (04), YEAR (A. DATA)) + ' +
'''' + '/' + '''' + ' +
CONVERT (VARCHAR (01), DATEPART (QQ, A. DATA)), CF3.DESCRICAO) T '+
'PIVOT ( SUM(QTDE)'+
'
FOR ANO IN ('+@ANOS+')'+
' ) AS P')

```

END

```

IF @PERIODO = 'MENSAL'
BEGIN

```

```

SET @ANOS = ''
SET @COLUNAS = ''
DECLARE cursor_objects CURSOR FOR

```

```

SELECT DISTINCT CONVERT (VARCHAR (04), YEAR (DATA)) + '/' +
RIGHT ('0' + CONVERT (VARCHAR (02), MONTH (DATA)), 2) AS MES
FROM FATO11_NCM2
WHERE (FATO6 = ('{cboUnidade}')) OR '{cboUnidade}' =
''<TODOS>'' )
AND (FATO7 = ('{cboSubUnidade}')) OR '{cboSubUnidade}' =
''<TODOS>'' )
AND (FATO11 IN ('{CBOSEG}')) OR '{CBOSEG}' = ''<TODOS>'' )
AND FATO3 IN ({LSTFORNEC})
AND YEAR (DATA) in ({LSTANO})
ORDER BY 1

```

```

OPEN cursor_objects

```



```

,ROUND((SUM(A.MTR3)/(select sum(x.MTR3) from
FATO11_NCM2 x
WHERE YEAR(DATA)
IN( ' + '{LSTANO}' +
') AND
(x.FATO6 IN ( ' + '''' + '{cboUnidade}' + '''' + ') OR ' +
'''' + '{cboUnidade}' + '''' + ' = ''<TODOS>'')
and
(x.FATO7 IN ( ' + '''' + '{cboSubUnidade}' + '''' + ') OR ' +
'''' + '{cboSubUnidade}' + '''' + ' = ''<TODOS>'')
and
(x.FATO11 IN ( ' + '''' + '{CBOSEG}' + '''' + ') OR ' + ''''
+ '{CBOSEG}' + '''' + ' = ''<TODOS>'')
AND x.FATO11 NOT
IN (''exclusao'')
AND x.FATO6 NOT
IN (19,22,16,25,26)
AND x.FATO3 <>
107
AND
CONVERT (VARCHAR(04),YEAR (A.DATA)) + ' + '''' + '/' + '''' +
' + RIGHT( ' + '''' + '0' + '''' + ' +
CONVERT (VARCHAR(02),MONTH (A.DATA)),2) =
CONVERT (VARCHAR(04),YEAR (X.DATA)) + ' + '''' + '/' + '''' +
' + RIGHT( ' + '''' + '0' + '''' + ' +
CONVERT (VARCHAR(02),MONTH (X.DATA)),2) )) * 100,2) AS
QTDE'+
' FROM FATO11_NCM2 A INNER JOIN CAD_FATO3_NCM2 CF3
ON (A.FATO3 = CF3.FATO3) '+
' WHERE YEAR(DATA) IN( ' + '{LSTANO}' +
') AND (A.FATO6 IN ( ' + '''' + '{cboUnidade}' +
'''' + ') OR ' + '''' + '{cboUnidade}' + '''' + ' =
''<TODOS>'')
and (A.FATO7 IN ( ' + '''' + '{cboSubUnidade}'
+ '''' + ') OR ' + '''' + '{cboSubUnidade}' + '''' + ' =
''<TODOS>'')
and (A.FATO11 IN ( ' + '''' + '{CBOSEG}' +
'''' + ') OR ' + '''' + '{CBOSEG}' + '''' + ' = ''<TODOS>'')
AND A.FATO3 IN ({LSTFORNEC})
AND A.FATO11 NOT IN (''exclusao'')
AND A.FATO6 NOT IN (19,22,16,25,26)
AND A.FATO3 <> 107 ' +
' GROUP BY CONVERT (VARCHAR(04),YEAR (A.DATA)) + ' +
'''' + '/' + '''' + ' + RIGHT( ' + '''' + '0' + '''' + ' +
CONVERT (VARCHAR(02),MONTH (A.DATA)),2),CF3.DESCRICAO) T '+
'PIVOT ( SUM(QTDE)'+
' FOR ANO IN (''+@ANOS+'')'+
' ) AS P')
END

```

APÊNDICE E – Tipos de Algoritmos de *Machine Learning*

Quadro 11 – Tipos de Algoritmos de *Machine Learning*

10 Principais Algoritmos de Machine Learning	Propósito	Utilizado no Protótipo
1 Árvores de Decisão	Uma árvore de decisão é uma ferramenta de apoio que utiliza um gráfico ou modelo de decisões e suas possíveis consequências, incluindo resultados de eventos fortuitos, custos de recursos e utilidade.	Não
2 Classificação Naïve Bayes	Os classificadores Naïve Bayes são uma família de classificadores probabilísticos simples com base na aplicação Bayes 'teorema com forte independência entre as características. A imagem em destaque é a equação – em que $P(A B)$ é a probabilidade posterior, $P(B A)$ é a probabilidade, $P(A)$ é a probabilidade prévia e $P(B)$ é preditor de probabilidade prévia.	Sim
3 Regressão Linear Simples	Em estatística, regressão linear simples é o quadrado mínimo estimador de um modelo de regressão linear com uma única variável explicativa. Em outras palavras, regressão linear simples se encaixa uma linha reta através do conjunto de pontos n de tal forma que faça a soma dos quadrados <i>residuais</i> do modelo (isto é, as distâncias entre os pontos verticais do conjunto de dados e a linha reta) tão pequena quanto possível.	Sim
4 Regressão logística	A regressão logística é uma poderosa forma estatística de modelar um resultado binomial com uma ou mais variáveis explicativas. Ela mede a relação entre a variável dependente categórica e uma ou mais variáveis independentes, estimando as probabilidades usando uma função logística, que é a distribuição logística cumulativa.	Não
5 Support Vector Machine	SVM é um algoritmo binário da classificação. Dado um conjunto de pontos de 2 tipos em lugar N dimensional, SVM gera um hiperplano ($N - 1$) dimensional para separar esses pontos em 2 grupos. Digamos que você tem alguns pontos de 2 tipos em um papel que são linearmente separáveis. SVM encontrará uma linha reta que separa esses pontos em 2 tipos e situados o mais longe possível de todos esses pontos.	Não
6 Ensemble Methods	São algoritmos de aprendizagem que constroem um conjunto de classificadores e, em seguida, classificam novos pontos de dados, tendo um ponderado voto de suas previsões. O método de conjunto original é a média bayesiana, mas os algoritmos mais recentes incluem codificação de saída, correção de erros, <i>bagging</i> e reforço.	Não
7 Algoritmos de Agrupamento (Clustering)	É a tarefa de agrupar um conjunto de objetos de tal forma que os do mesmo grupo (cluster) são mais semelhantes uns aos outros do que aqueles em outros grupos.	Não
8 Decomposição em valores singulares	Na álgebra linear, a SVD é uma fatorização de uma matriz complexa real. Para uma dada matriz $m \times n$ M , existe uma decomposição tal que $M = U\Sigma V$, onde U e V são matrizes unitárias e Σ é uma matriz diagonal.	Não
9 Análise de Componentes Principais	PCA é um procedimento estatístico que usa uma transformação ortogonal para converter um conjunto de observações de variáveis, possivelmente correlacionadas em um conjunto de valores de variáveis linearmente não correlacionadas, chamadas componentes principais.	Não
10 Análise de componentes independentes	ICA é uma técnica estatística para revelar fatores ocultos que estão subjacentes a conjuntos de variáveis aleatórias, medições ou sinais. A ICA define um modelo generativo para os dados multivariados observados, que normalmente é dado como um grande banco de dados de amostras. No modelo, as variáveis de dados são assumidas como misturas lineares de algumas variáveis latentes desconhecidas, e o sistema de mistura também é desconhecido. As variáveis latentes são consideradas não-gaussianas e mutuamente independentes, e são chamadas de componentes independentes dos dados observados.	Não

Fonte: Resultado da Pesquisa