



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS CAMPINA GRANDE
CENTRO PPGCTS
MESTRADO PROFISSIONAL EM CIÊNCIA E TECNOLOGIA EM SAÚDE**

ENNYO JOSÉ BARROS DE ARAÚJO

**APRENDIZADO DE MÁQUINA PARA DETECÇÃO DE COMPORTAMENTO
FRAUDULENTO EM DADOS SIMULADOS DO SISTEMA ODONTOLÓGICO DO
SUS**

CAMPINA GRANDE
2017

ENNYO JOSÉ BARROS DE ARAÚJO

**APRENDIZADO DE MÁQUINA PARA DETECÇÃO DE COMPORTAMENTO
FRAUDULENTO EM DADOS SIMULADOS DO SISTEMA ODONTOLÓGICO DO
SUS**

Dissertação apresentada ao Programa de Pós-Graduação Profissional em Ciência e Tecnologia em Saúde - PPGCTS da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de mestre em Tecnologia em Saúde.

Área de concentração: Medicina.

Orientador: Prof. Dr. Edmar Candeia Gurjão.

CAMPINA GRANDE
2017

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

A659a Araújo, Ênnyo José Barros de.

Aprendizado de máquina para detecção de comportamento fraudulento em dados simulados do Sistema odontológico do SUS [manuscrito] / Ênnyo José Barros de Araújo. - 2017.
79 p. : il. color.

Digitado.

Dissertação (Mestrado Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2017.

"Orientação: Prof. Dr. Edmar Candeia Gurjão, Pró-Reitoria de Pós-Graduação e Pesquisa".

1. Aprendizado de máquina. 2. Sistemas de saúde. 3. Sistemas de auditoria. 4. Detecção de fraude. I. Título.

21. ed. CDD 005.3

ENNYO JOSÉ BARROS DE ARAÚJO

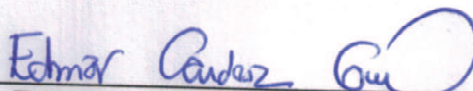
USO DE APRENDIZADO DE MÁQUINA PARA A DETECÇÃO DE FRAUDES NO SISTEMA ODONTOLÓGICO DO SUS

Dissertação apresentada ao Programa de Pós-Graduação Profissional em Ciência e Tecnologia em Saúde - PPGCTS da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de mestre em Tecnologia em Saúde.

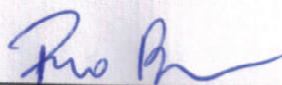
Área de concentração: Medicina.

Aprovada em: 06/04/2017.

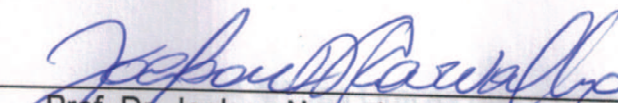
BANCA EXAMINADORA



Prof. Dr. Edmar Candeia Gurjão (Orientador)
Universidade Federal de Campina Grande (UFCG)



Prof. Dr. Paulo Eduardo Silva Barbosa
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Joelson Nogueira, de Carvalho
Universidade Federal da Paraíba (UFPB)

A M^a das Graças que a muito tempo se foi e não pôde acompanhar nenhum dos caminhos que trilhei, DEDICO.

AGRADECIMENTOS

Ao Professor Dr. Edmar Candeia Gurjão, meu orientador, por compartilhar comigo parte de sua experiência, competência e tranquilidade.

À professora Dra. Andréa de Farias Barros Andrade que com um sonho mudou o rumo da minha vida.

Ao meu pai José Alfrísio Alves de Araújo e a todos meus familiares por todo o apoio que me deram.

A minha esposa Selma por amarrar meu paraquedas todos os dias e a minha filha Sofia por suportar todas as minhas ausências.

Aos meus consultores que dedicaram seus conhecimentos e seu tempo para ajudar-me a construir este trabalho. Leonardo Henrique de Araújo Cavalcante, Silvio Rafael de Araújo Almeida e Sara Jane Xavier Gurjão. Agradeço pela ajuda inestimável.

A Allyson da Veiga Cavalcante, por fazer possível meus horários flexíveis para que eu pudesse estar neste mestrado.

A todo o corpo docente do mestrado que contribuíram com a minha formação e na concretização das ideias.

RESUMO

O Sistema Unificado de Saúde do Brasil (SUS) tem um dos maiores programas de assistência odontológica gratuita do mundo. O governo federal, via os seus entes federados, disponibiliza tratamento odontológico a 70% da população brasileira. Mesmo com números tão expressivos, esse sistema ainda possui muitas iniquidades e sofre de um problema endêmico no Brasil, a corrupção, neste caso, em forma de fraude. Estudos indicam que a fraude é responsável por até 40% de desperdícios de recursos em instituições de saúde, trazendo prejuízos financeiros enormes, sobretudo penalizando quem mais necessita da assistência médica, a população de baixa renda. Ferramentas de mineração de dados podem revelar padrões incognoscíveis para seres humanos, mas relativamente fáceis de encontrar para uma máquina treinada. Detectar e eliminar a fraude não é um problema trivial, requer tecnologias, profissionais especialistas e ferramentas. Neste trabalho utilizou-se aprendizado de máquina em dados simulados de atendimentos odontológicos seguindo o padrão o SUS, para separar os que tem indícios de fraude. Foram avaliados os algoritmos *Isolation Forest*, *K-Nearest Neighbors* e *Mahalanobis Distance*, e em todos os testes houve sucesso em separar os *clusters* de dados normais dos anômalos, demonstrando a possibilidade de utilização dessas ferramentas na detecção de fraudes de forma eficiente.

Palavras-Chave: Aprendizado de Máquina, SUS, Detecção, Fraude.

ABSTRACT

The Brazilian Unified Health System (Sistema Unificado de Saúde –SUS) has one of the largest free dental care programs in the world. The federal government, through its federated entities, provides dental treatment to approximately 70% of the Brazilian population. Even with such expressive numbers, this system still has many iniquities and suffers from an endemic problem in Brazil, corruption, in this case in forms of frauds. This is not a local problem, institutions around the world are studying and building technological solutions that can tackle one of the faces of frauds. Studies indicate that fraud is responsible for up to 40% of wasted resources in SUS, bringing huge financial losses, especially penalizes those who most need medical care, the low income population. Detecting and eliminating fraud is not a trivial problem, it requires technologies, experts and tools. Data mining techniques can reveal patterns difficult to be detected by humans. In this work using simulated data based on the SUS data standard machine learning is applied to detect frauds. Isolation Forest, *K*-Nearest Neighbors and Mahalanobis Distance algorithms was applied and successfully detected frauds.

Keywords: SUS. Machine Learning, Healthy System, Detection, Fraud.

LISTA DE ILUSTRAÇÕES

Figura 1	Recorte da Ficha Individual de Atendimento	51
Figura 2	Tela do software e-SUS AB	52
Figura 3	Histograma de Atendimento	54
Figura 4	Representação do modelo de aprendizagem	58
Figura 5	Submetendo dados ao modelo	58
Figura 6	Gráfico MLE (Distância de Mahalanobis)	57
Figura 7	Gráfico plotado do algoritmo KMeans	59
Figura 8	Gráfico plotado do algoritmo Isolation Forest	60

LISTA DE TABELAS

Tabela 1	Dados gerados em entrevistas.....	66
----------	-----------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

AIH	Autorização de Internação Hospitalar
ANA	Agência Nacional de Auditoria
CD	Cirurgião Dentista
CDS	Coleta de Dados Simplificada
CEP	Comitê de Ética em Pesquisa
CSV	Comma-separated values (Valores separados por virgula)
DATASUS	Departamento de informática do SUS
EHR	<i>Electronic Health Record</i>
HL7	<i>Health Level Seven International</i>
IaaS	<i>Infrastructure as a Service (Infraestrutura como Serviço)</i>
IoT	<i>Internet of Things (Internet das coisas)</i>
MS	Ministério da Saúde
PEC	Prontuário Eletrônico do Cidadão
PEP	Prontuário Eletrônico do Paciente
RES	Registro Eletrônico em Saúde
SISAB	Sistema de Informação em Saúde para Atenção Básica
SUS	Sistema Único de Saúde
UBS	Unidades Básicas de Saúde

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS.....	15
	2.1. OBJETIVO GERAL	15
	2.2. OBJETIVOS ESPECÍFICOS	15
3	JUSTIFICATIVA.....	16
4	PRESSUPOSTO	17
5	REFERENCIAL TEÓRICO	18
	5.1. DATASUS.....	19
	5.2. FRAUDES	24
	5.3. AUDITORIAS	27
	5.3.1. SISTEMA DE AUDITORIAS NO SUS.....	27
	5.3.2. SISTEMA DE AUDITORIAS ODONTOLÓGICAS.....	30
	5.4. INTELIGENCIA ARTIFICIAL	30
	5.4.1. APRENDIZADO DE MÁQUINA.....	32
	5.4.1.1. SUPERVISIONADO	33
	5.4.1.2. NÃO SUPERVISIONADO	34
	5.4.2. ESCOLHA DE ATRIBUTOS	34
	5.4.3. PRINCIPAIS ALGORITMOS DE APRENDIZADO DE MÁQUINA	36
	5.5. DETECÇÃO DE ANOMALIAS	39
	5.6. FERRAMENTAS	40
	5.6.1. SCIKIT-LEARN	40
	5.6.1.1. QUALIDADE E CONSISTÊNCIA.....	41
	5.6.1.2. TECNOLOGIAS COMPLEMENTARES	42
	5.6.1.3. FACILIDADES E EFICIÊNCIA.....	42
6	TRABALHOS RELACIONADOS.....	44
7	METODOLOGIA	47
8	EXPERIMENTO.....	49
	8.1. HIPOTHESES.....	49
	8.2. GERAÇÃO DOS DADOS	50
	8.2.1. SELEÇÃO DOS ATRIBUTOS	55
	8.2.2. O TREINAMENTO DOS MODELOS.....	55
9	RESULTADOS.....	57
	9.1.1. CONSIDERAÇÕES.....	61
10	CONCLUSÕES	63
11	BIBLIOGRAFIA	65
	ANEXO A – FICHA DE ATENDIMENTO INDIVIDUAL	74
	ANEXO B – ESTUDO SB BRASIL 2010.....	76
	ANEXO C – TABELAS SIA-SUS COM CÓDIGOS DE ATENDIMENTO	77

1 INTRODUÇÃO

O Brasil tem o maior sistema de saúde gratuito do mundo, mesmo assim, não consegue atingir de forma eficiente toda população. Existem sérios problemas a serem enfrentados, muitos deles podem ser resolvidos utilizando tecnologia da informação. Os softwares e técnicas de mineração de dados tradicionais não mais suprem as necessidades de análise e tomada de decisão. Os dados que o governo precisa estão espalhados entre vários municípios em formatos e arquitetura distintas. É necessário desenvolver novas estratégias de análise e controle de dados e informação. Cuidados com a saúde e governo geram enormes quantidades de dados, porém, e sem uma análise correta eles podem tornar-se inúteis (Archenaa & Anita, 2015).

A fraude é uma preocupação constante para empresas que utilizam sistemas de processamento de dados. Esse problema é responsável por consumir entre 7% e 17% de todos os recursos do sistema produtivo. Cartões de crédito, seguradoras, companhias telefônicas, planos de saúde e seguros saúde. Essa cifra chega a trilhões de dólares todos os anos. Por esse motivo, diversas instituições e centros de pesquisas estão empenhados em desenvolver soluções de softwares para mitigar os números resultantes dos casos de fraude.

Aprendizagem de máquina está entre as tecnologias mais promissoras para solucionar problemas de classificação e identificar padrões que podem gerar conhecimento. Conjuntamente com o avanço das técnicas de análises, houve avanço e barateamento das tecnologias de processamento e armazenamento de dados, gerando serviços como a Amazon AWS, onde se pode ter o processamento de dezenas de máquinas com custos relativamente baixo. Deste momento em diante começaram a surgir estudos sobre como utilizar recursos como a estatística e a ciência da computação para resolução de problemas em diversos domínios do conhecimento.

Um dos segmentos onde se aplica muito conhecimento e recursos são os que envolvem saúde e finanças. O mercado percebeu que poderia aplicar tais tecnologia na análise no grande volume de dados que são gerados todos os dias. Esse fenômeno não é novo, mas o nome popular dele é *Big Data*.

Entender e analisar dados pode melhorar os processos de decisão. O conhecimento não é mais adquirido somente analisando o passado dos dados, mas prevendo o futuro baseado em experiência adquirida com o processamento de diversas fontes de dados. Essas novas técnicas têm aplicações em praticamente todas as áreas. Uma das que mais se beneficiam são as finanças e cuidados com saúde. Para melhorar a qualidade no cuidado com a saúde e minimizar custos, é necessário analisar efetivamente grandes volumes de. Análises em *Big Data* ajudam a descobrir padrões e relacionamentos entre os dados com a ajuda de aprendizado de máquina (Archenaa & Anita, 2015).

Neste trabalho é proposto a simulação de dados de atendimento de unidades odontológica do SUS, submetê-los a um framework de inteligência artificial – Scikit-learning – e verificar se os resultados revelam anomalias que podem indicar fraudes. Aprendizado de máquina não supervisionado foi a solução utilizada nesta pesquisa e apontou resultados promissores. Pois não é tarefa apenas humana decidir se 1,4 bilhões de consultas são verdadeiras ou falsas. Necessita-se minerar dados. A mineração de dados é o processo de analisar os dados de diferentes perspectivas e resumi-los em informações úteis (CHAUHAN; SHUKLA, 2015).

Os resultados deste estudo podem ser úteis para auxiliar um dos muitos mecanismos utilizados para tentar controlar fraudes, um deles é a auditoria. Resultados de análises fornecerão meios para tornar o processo de auditoria mais eficiente. Pois é identificado na literatura que o sistema do SUS não utiliza ferramentas desta natureza para criar sua pauta de auditorias. A auditoria por si só não consegue acompanhar a dinâmica que os fraudadores utilizam.

O processo de auditoria tem uma clara missão social, garantir que recursos públicos estejam sendo usados de forma adequada e que procedimento clínicos estejam sendo executados com a qualidade necessária para garantir o bem-estar da população. O DENASUS desenvolve esse papel de gestor e auditor da qualidade dos serviços médicos. Neste trabalho propõe-se a explorar ferramentas tecnológicas afim de servir como uma ferramenta computacional complementar a tomada de decisão em indicação de auditorias.

2 OBJETIVOS

2.1.OBJETIVO GERAL

Usar aprendizado de máquina para encontrar anomalias (*Outliers*) em dados simulados do sistema odontológicos do SUS.

2.2.OBJETIVOS ESPECÍFICOS

- ✓ Encontrar números destoantes na quantidade de atendimento que um cirurgião dentista pode executar.
- ✓ Verificar a existência de percentual incongruente entre a quantidade de procedimentos executados em próteses com o percentual informado em estudos do governo.
- ✓ Testar a eficácia de três algoritmos de aprendizagem de máquina na detecção de anomalias.
- ✓ Utilizar a biblioteca de aprendizado de máquina *scikit-learn* para acelerar o processo de detecção de anomalias.
- ✓ Entender os processos de auditorias e detecção de fraudes no SUS.
- ✓ Gerar um software experimental utilizando *machine learning* para inferência de conhecimento.

3 JUSTIFICATIVA

A saúde brasileira vive já há algum tempo uma situação precária, segundo o economista André Médici: “O Brasil gasta pouco e gasta mal na saúde”. Gerir o maior sistema público de saúde do mundo tem muitos desafios. Utilizar tecnologias para auxiliar na gestão desse sistema é uma solução adequada. As iniquidades encontradas no SUS tem como parte do problema falta de recurso ou a má distribuição do mesmo. Então, se existem tecnologias que podem tornar mais eficiente a gestão e controle do SUS, ela merece ser investigada. Portanto, pesquisas direcionadas a esta área se justificam tanto pelo valor econômico como pelo valor social.

Vertentes de pesquisa em vários países apontam para o que é considerado o novo petróleo, os dados. Esse tema tem grande relevância no cenário mundial. Iniciativas privadas e governamentais têm projetos conectados às suas bases de dados. Aprendizado de máquina, *big data* e *biohacking* são temas presentes no Vale do Silício, local que serve de termômetro sobre temas emergentes em tecnológicas. Em seu livro sobre ética e *Big Data* (PATTERSON, 2012), afirma que estamos vivenciando uma das tecnologias mais importante da nossa era. Ela influenciará diversas áreas da nossa vida.

Estudos do National Transfer¹, cruzam a idade da população e suas necessidades de gastos com saúde e os resultados apontam para um cenário em que as contas com saúde são insustentáveis em longo prazo, logo, usar tecnologia para encontrar caminhos que levem a diminuição de custos, parece promissor.

¹ O objetivo do projeto *National Transfer Accounts* (NTA) é melhorar nossa compreensão da economia geracional. Como o crescimento populacional e a mudança da estrutura etária influenciam o crescimento econômico, a equidade de gênero e gerações, as finanças públicas e outras características importantes da macroeconomia

4 PRESSUPOSTO

O pressuposto deste trabalho é que as anomalias (*outlier*) encontradas possam servir de indícios para a averiguação de fraudes contra o sistema de saúde. Em um universo projetado de aproximadamente de 7% a 30% de fraudes, encontrar 1% ou 2% será considerado um avanço, já que o custo para chegar nessa solução é relativamente barato e os benefícios muitos. É impraticável do ponto de vista econômico ter uma equipe investigando fraude em cada município brasileiro, o custo seria equiparado ao valor gasto com a fraude. Por outro lado, é humanamente impossível poucas equipes lidarem com dados gerados por um sistema como o SUS, dessa forma, o uso de uma ferramenta de tomada de decisão que utilize aprendizado de máquina é uma possibilidade aceitável da ótica de custo benefícios. Delega-se as máquinas o que de melhor elas melhor sabem fazer: repetição, associação e cálculos.

Visto a escassez de trabalhos publicados sobre saúde e fraude no contexto aqui proposto, espera-se que esse trabalho traga luz a esse problema pouco explorado e outros pesquisadores possam aplica-los em outro domínio.

5 REFERENCIAL TEÓRICO

Buscar anomalias em bases de dados é um problema que vem ganhando visibilidades em meios acadêmicos. Encontrar distúrbios em dados é um problema importante que vem sendo pesquisado em diversas áreas e aplicada a diversos domínios. Detectar anomalias é definido como encontrar padrões em dados que não se comportam de forma esperada(CHANDOLA; BANERJEE; KUMAR, 2009). Diversas técnicas tem sido desenvolvidas para serem utilizadas em uma grande variedade de aplicações, tais como: detecção de fraudes em sistema financeiro, invasões em rede de computadores, sistemas de vigilância e em cuidados com a saúde(Yuen & Ortiz, 2016).

Muitas instituições estão empenhadas em encontrar e correlacionar informações aparentemente desconexas, encontrar padrões e anomalias. Entre as que mais se beneficiariam com o uso de tecnológicas são as que lidam com o sistema de saúde. Avanços em tecnologias de cuidados com a saúde, digitalização de registros médicos e telemedicina tem levado a um rápido crescimento de dados médicos(YAN et al., 2016). Os dados em saúde são gerados de formas diversas: imagens, tomográficas, ECG, resultado de exames, consultas, dados clínicos entre outros. Até recentemente era impraticável armazenar e analisar essa granularidade de informações, porém, o avanço das tecnologias e o barateamento do processamento e do armazenamento de dados estão possibilitando essas ações, o que leva a uma tendência mundial em informatização de registros de saúde.

As tecnologias aplicadas na saúde geram dados expressivos, principalmente num país como o Brasil que atende 71% da população², o que corresponde a mais de 100 milhões de habitantes em um sistema integrados de saúde. Segundo dados disponíveis no site do Portal Brasil, em 2014, foram registrados pelo SUS (Sistema Único de Saúde) 4,1 bilhões de tratamentos ambulatoriais, 1,4 bilhões de consultas médicas e 11,5 milhões de internações. Essas movimentações geram uma grande quantidade de dados associados: consultas, exames, anotações médicas, dados de sensores, prescrições e uso de insumos.

² PESQUISA NACIONAL DE SAÚDE - 71% dos brasileiros têm os serviços públicos de saúde como referência(“71% dos brasileiros têm os serviços públicos de saúde como referência”, [s.d.]).

Entre 2010 e 2015, o número de consultas na Atenção Básica se expandiu. Aumentou 47% no período, passando de 126,3 milhões para 186,1 milhões. Até hoje a população conta com 41.688 Unidades Básicas de Saúde (UBS) em funcionamento, 771 a mais do que em 2010 (40.917), que equivale a uma cobertura de 70% da população(BERALDO, 2016). Porém, o número de unidades utilizando o sistema ainda é baixo. Atualmente, 76% das unidades básicas de saúde ainda registram o histórico do paciente em papel(CAIXOTE, 2016). Recentemente o Ministério da Saúde adotou uma medida sob forma de portaria que solicita a informatização de todas as unidades de saúde, dessa forma, a coleta de dados da assistência básica será completamente informatizada. Essa medida significa na prática que as unidades de saúde que não obedecerem a portaria, ficaram com recurso retidos pela união, o que leva a crer que em pouco tempo existirá uma quantidade enorme de dados oriundos das UBS (Unidades Básicas de Saúde), gerando um cenário: armazenar, organizar, tratar, classificar e extrair informações úteis desse montante de dados.

Analisar manualmente as bases de dados tornou-se uma tarefa impraticável, por isso ferramentas para análise automática têm sido desenvolvidas(ZAMONER, 2013). Tecnologias de *Big Data* e Aprendizado de Máquina são promissoras nestes cenários que envolvem análise de grandes quantidades de dados de tipos e formatos distintos. Escondidos nestes dados podem estar contidos padrões que revelariam fraudes, erros médicos ou mesmo que auxiliariam na descoberta de tratamentos mais eficazes e baratos. Qualquer pesquisa que vise extrair conhecimento e melhorar o entendimento destes dados podem ajudar a mitigar custos que hoje são da ordem de 98,4 bilhões de reais.

5.1. DATASUS

O SUS tem como objetivo fornecer atendimento a toda população brasileira. Na prática, contudo, permanece favorecendo a política neoliberal de estado mínimo, já que aproximadamente 20% da população busca a assistência suplementar, por entender ser o SUS de difícil acesso e baixa qualidade(ANDRADE; ANDRADE, 2010). Mesmo deficitário o SUS ostenta números grandiosos de atendimento, segundo dados do Ministério da Saúde, o houveram 4,1 bilhões de atendimentos ambulatoriais, 1,4 bilhões de consultas médicas 11,5 milhões de internações, esses

são números de 2014, novas estatísticas devem elevar esses resultados (“População teve acesso a 1,4 bi de consultas médicas pelo SUS em um ano — Portal Brasil”, [s.d.]). O SUS acomoda outro programa bastante importante no tocante a saúde pública. O Programa Brasil Sorridente, ele é considerado o maior programa público de saúde bucal do mundo (SCHERER; SCHERER, 2015).

O SUS é descentralizado, ele possui uma estrutura distribuída de forma capilar, se estendendo até os municípios. Para administração dos sistemas do SUS, houve a criação de um órgão especial criado para gerir dados, criar soluções tecnológicas, treinamento, assistência e implantação de softwares e hardware. O DATASUS.

O DATASUS foi criado para gerir a quantidade de informação advinda do SUS, sua criação se deve ao decreto Nº 7.530 de 21 de junho de 2011, ficando encarregado de centralizar, controlar, integrar e criar sistemas e armazenar toda informação gerada do sistema de saúde. Vários sistemas integram o ecossistema de controle de todas entidades participantes do sistema único de saúde e de seus utilizadores. Ficou claro na época que sem um órgão responsável pelas informações, seria impossível gerir o SUS. A informação é fundamental para a democratização da saúde e o aprimoramento de sua gestão. A informatização das atividades do Sistema Único de Saúde (SUS), dentro de diretrizes tecnológicas adequadas, é essencial para a descentralização das atividades de saúde e viabilização do Controle Social sobre a utilização dos recursos disponíveis (“DATASUS”, [s.d.]).

O DATASUS possui dezenas de sistemas. Eles se caracterizam por serem de administração hospitalar, dos associados, de segurança e integração. Esses sistemas são de tecnológicas diferentes de bases de dados diferentes, integrados ou não em redes de computadores. É um conjunto heterogêneo de arquiteturas e tecnologias distintas, o que produz desafios tecnológicos no tangente a integração, normalização e captação destes dados.

Resulta deste processo um cenário de fragmentação que dificulta a visibilidade e o “cruzamento” das informações contidas nos diversos sistemas de informação, criando obstáculos ou mesmo inviabilizando prospecções e análises de questões relativas à saúde da população. Estas condições de trabalho subtraem subsídios dos tomadores de decisão (FERRAZ, 2009). Segundo Thiago Marzagão, em uma de suas palestras, os órgão do governo que usam algoritmos e mineração

de dados são: Receita Federal, Conselho Administrativo de Defesa Econômica(CADE), Polícia Federal e Controladoria Geral da União(THIAGO MARZAGÃO, [s.d.]). Na revisão bibliográfica realizada neste trabalho, não foi encontrada literatura associando sistemas informatizados de apoio a decisão aplicados ao sistema do SUS.

Os principais sistemas do DATASUS são:

- Sistema de Informações sobre Mortalidade (SIM);

O Sistema de Informações sobre Mortalidade (SIM) foi criado pelo DATASUS para a obtenção regular de dados sobre mortalidade no país. A partir da criação do SIM foi possível a captação de dados sobre mortalidade, de forma abrangente, para subsidiar as diversas esferas de gestão na saúde pública. Com base nessas informações é possível realizar análises de situação, planejamento e avaliação das ações e programas na área.

- Sistema de Informações sobre Nascidos Vivos (Sinasc)

O DATASUS desenvolveu o Sistema de Informações sobre Nascidos Vivos (SINASC) visando reunir informações epidemiológicas referentes aos nascimentos informados em todo território nacional. Sua implantação ocorreu de forma lenta e gradual em todas as Unidades da Federação.

- Sistema de Informações de Agravos de Notificação (Sinan)

O Sistema de Informação de Agravos de Notificação - Sinan é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória.

- Sistema de Informações Hospitalares do SUS (SIH/SUS)

A finalidade do AIH (Sistema SIHSUS) é registrar todos os atendimentos provenientes de internações hospitalares que foram financiadas pelo SUS, e a partir deste processamento, gerar relatórios para que os gestores possam fazer os pagamentos dos estabelecimentos de saúde. Além disso, o nível Federal recebe mensalmente uma base de dados de todas as internações autorizadas (aprovadas ou não para pagamento) para que possam ser repassados às Secretarias de Saúde os valores de Produção de Média e Alta complexidade, além dos valores de CNRAC, FAEC e de Hospitais Universitários – em suas variadas formas de contrato de gestão.

- Cadastro Nacional de Estabelecimentos de Saúde (CNES)

Disponibiliza informações relacionadas aos estabelecimentos, profissionais de saúde, mantenedoras e cadastro de equipes. Apresenta no site , toda a infraestrutura de serviços de saúde bem como a capacidade instalada disponível no país.

- Sistema de Informações do Programa Nacional de Imunização (SI-PNI)

O SI-PNI é um sistema desenvolvido para possibilitar aos gestores envolvidos no Programa Nacional de Imunização, a avaliação dinâmica do risco quanto à ocorrência de surtos ou epidemias, a partir do registro dos imunobiológicos aplicados e do quantitativo populacional vacinado, agregados por faixa etária, período de tempo e área geográfica. Possibilita também o controle do estoque de imunobiológicos necessário aos administradores que têm a incumbência de programar sua aquisição e distribuição.

- CADSUS - Sistema de Cadastramento de usuários do SUS

Sistema de cadastramento de Usuários do Sistema Único de Saúde permite a geração do Cartão Nacional de Saúde, que facilita a gestão do Sistema Único de Saúde e contribui para o aumento da eficiência no atendimento direto ao usuário.

- SIAB - Sistema de Informação de Atenção Básica

O Sistema de Informação da Atenção Básica foi implantado para o acompanhamento das ações e dos resultados das atividades realizadas pelas equipes do Programa Saúde da Família - PSF. O SIAB foi desenvolvido como instrumento gerencial dos Sistemas Locais de Saúde e incorporou em sua formulação conceitos como território, problema e responsabilidade sanitária. Através dele obtêm-se informações sobre cadastros de famílias, condições de moradia e saneamento, situação de saúde, produção e composição das equipes de saúde. Principal instrumento de monitoramento das ações do Programa Saúde da Família, tem sua gestão na Coordenação de Acompanhamento e Avaliação do Departamento de Atenção Básica / SAS.

- SISCOLO/SISMAMA - Sistema de Informação do câncer do colo do útero e Sistema de Informação do câncer e mama

Sistema informatizado de entrada de dados desenvolvido pelo DATASUS em parceria com o INCA, para auxiliar a estruturação do Viva Mulher (Programa

Nacional de Controle do Câncer do Colo do Útero e de Mama). Coleta e processa informações sobre identificação de pacientes e laudos de exames citopatológicos e histopatológicos, fornecendo dados para o monitoramento externo da qualidade dos exames, e assim orientando os gerentes estaduais do Programa sobre a qualidade dos laboratórios responsáveis pela leitura dos exames no município.

- HIPERDIA - Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos

O Hiperdia destina-se ao cadastramento e acompanhamento de portadores de hipertensão arterial e/ou diabetes mellitus atendidos na rede ambulatorial do Sistema Único de Saúde – SUS, permitindo gerar informação para aquisição, dispensação e distribuição de medicamentos de forma regular e sistemática a todos os pacientes cadastrados. O sistema envia dados para o Cartão Nacional de Saúde, funcionalidade que garante a identificação única do usuário do Sistema Único de Saúde – SUS.

- e-SUS-SAMU

O sistema utilizado para captura de dados do SAMU (Serviço de Atendimento Móvel de Urgência) foi desenvolvido para funcionar de forma autônoma, será realizada a integração do sistema desenvolvido para o SAMU, e também os registros dos procedimentos que tiverem sido realizados nos atendimentos de urgência e que deverão ser incorporados nos boletins de produção gerenciados pela central de regulação.

- GIL - Gerenciamento de Informações Locais

Tem como finalidade sistematizar as ações no atendimento dos estabelecimentos de saúde, contribuindo para a melhoria da gestão em saúde. Esse sistema pode ser instalado em qualquer EAS da rede ambulatorial básica do SUS, independentemente de seu porte ou grau de complexidade. Sua manutenção é garantida pelo DATASUS, atendendo as atualizações legais definidas pelo próprio Ministério da Saúde.

- e-SUS Hospitalar

O e-SUS Hospitalar tem seu foco no HIS (Hospital Information System) e PEP (Prontuário Eletrônico do Paciente), além de oferecer uma ferramenta para geração de relatórios, formulários dinâmicos baseados em metadados, painel de indicadores de gestão, sistema de laboratório e ERP. Ele pode atender as demandas de Hospitais Públicos na esfera Federal, Estadual e Municipal, bem como o Projeto

S.O.S. Emergências, UPA's e todas as demandas no âmbito do Ministério da Saúde.

- SIHSUS - Sistema de Informações Hospitalares do SUS

A finalidade do AIH (Sistema SIHSUS) é registrar todos os atendimentos provenientes de internações hospitalares que foram financiadas pelo SUS, e a partir deste processamento, gerar relatórios para que os gestores possam fazer os pagamentos dos estabelecimentos de saúde. Além disso, o nível Federal recebe mensalmente uma base de dados de todas as internações autorizadas (aprovadas ou não para pagamento) para que possam ser repassados às Secretarias de Saúde os valores de Produção de Média e Alta complexidade, além dos valores de CNRAC, FAEC e de Hospitais Universitários – em suas variadas formas de contrato de gestão.

- INTEGRADOR - Sistema de Integração dos Sistemas de Informação em Saúde

Aplicação desenvolvida para a esfera municipal com tecnologia de software livre, sempre que possível. É responsável pela constituição do banco de dados municipal, composto pelo Cadastro de Estabelecimento de Saúde – CNES, Cadastro de Usuários do SUS – CADSUS e informações dos eventos de saúde realizados na rede, pelos diversos sistemas de informatização locais. Prevê a criação de padrões de comunicação entre os sistemas, em consonância com as deliberações da XII Conferência Nacional de Saúde, para a Política Nacional de Informação e Informática do SUS, e a exportação dos dados para os demais sistemas de informação de saúde, integrando-os para melhor gerenciamento do SUS. (“Sistemas - DATASUS”, [s.d.]

5.2.FRAUDES

Fraude é um ato de má-fé que tem por objetivo fraudar ou ludibriar alguém. Ato de falsificar documentos, marcas e produtos. A etimologia da palavra não consegue dar o real significado, proporções e consequência do ato de fraudar impetrados dentro do serviço público de saúde. Fraudes existem onde e quando envolvem transações de dinheiro. Cuidados de saúde é especialmente um alvo tentador para os ladrões (GANGOPADHYAY; CHEN, 2016). Hoje o Brasil tem que lidar com pessoas de má índole, organizações criminosas infiltradas no sistema

público e com possíveis erros de operação por servidores honestos. Como separar esses casos e gerar um relatório analítico que descreva com maior precisão a realidade?

A fraude está aumentando com a expansão da tecnologia moderna e da comunicação global, resultando em perdas substanciais para os negócios. Conseqüentemente, a detecção de fraude tornou-se uma questão importante a ser explorada (YUFENG KOU et al., [s.d.]).

As fraudes impetradas no SUS, como em uma fraude comum, necessita que informações fraudulentas de pacientes e procedimentos sejam fornecidas como verdade. Gerar registros médicos falsos, agravar com um laudo uma condição de saúde que o paciente não está acometido. Informar que um procedimento foi executado, quando na verdade ele só existe de forma digital e/ou manuscrita. Tudo isso é feito para que o sistema do SUS efetue pagamentos inexistentes.

A fraude não acomete somente os cofres, ela pode acarretar problemas ao fraudado, quando suas os fraudadores falsificam documentos, realizam procedimentos indevidos e modificam informações sensíveis dos dados dos pacientes, trocando por dados falsos. O que se segue pode ser desastroso para a saúde do paciente, pois procedimentos podem ser adotados em cascata aos constatados.

Desvio de dinheiro, falsificação de prontuários, realização de procedimentos fantasmas, roubo de material, recebimento de propina para adiantar pessoas em filas de exames. Esses são alguns dos muitos tipos de fraudes a que é acometido o sistema do SUS. Não é raro ver nas mídias denúncias relacionadas ao sistema de saúde. Esse problema não está limitado a países em desenvolvimento, países desenvolvidos enfrentam o mesmo problema. Para tanto, existe uma associação mundial, a Global Health Care Anti-Fraud Network (GHCAN), uma entidade internacional estruturada com outras associações no mundo todos voltadas para o combate de fraudes em sistemas de seguro de saúde.

De acordo com estimativas do governo federal dos Estados Unidos e de grupos baseados em questões como a Associação Nacional Anti-Fraude de Cuidados de Saúde (NHCAA), até 10% de todos os gastos com saúde nos Estados Unidos Podem ser perdidos anualmente por fraude, abuso e desperdício(“Combating healthcare provider fraud and abuse”, [s.d.]). Se levarmos em conta que em 2011, US \$ 2,27 trilhões foram gastos em cuidados de saúde nos EUA, significa dezenas

de bilhões foram gastos indevidamente. Somente em um caso na Flórida(EUA), duas pessoas conseguiram receber 31 milhões de dólares falsificando documentos(“Global Health Care Anti-Fraud Network » U.S. — 2 Individuals Charged in \$31-Million Health Care Fraud Scheme”, [s.d.]). Os números brasileiros são diferentes, enquanto o índice de fraude de forma global é medido entre 10 e 15%, no Brasil eles começam em 20%, podendo chegar a 50% em determinadas fraudes. Todos esses números são estimados, pois como o SUS é público, não existe interesse em divulgar abertamente. É possível que isso seja refletido no pequeno número de pesquisas ligadas a fraudes publicadas no Brasil. Os números são absurdamente altos, mas eles são apenas parte do problema. O lado social desse problema é a parte mais grave.

As cidades mais pobres são as que mais sofrem, muitos países utilizam sistemas para a detecção de fraudes, mas é um problema de difícil solução. A cada dia as fraudes ficam mais sofisticadas e é fácil escondê-la num amontoado de dados gerados pelo sistema de saúde. Este problema decorre, em parte, da complexidade da detecção de fraudes, enquanto que apenas especialistas em medicina podem julgar se uma receita médica é uma fraude ou não. Não sendo economicamente viável contratar especialistas para verificar cada prescrição manualmente(ZHU; WANG; WU, 2011).

Na luta contra a fraude, as ações abrangem duas grandes categorias: prevenção da fraude e detecção de fraude(BOLTON et al., 2001). Este trabalho leva em consideração a detecção e não a prevenção. A prevenção requer planejamento e muito investimento em PIN de segurança, identificação por retina, rastreamento de transações, certificados digitais, software especialista, entre outros.

As principais formas de fraudes na saúde, entre outros, são:

- Faturamento por serviços não prestados;
- Desvio de verbas destinadas à saúde;
- Falsificação de documentos para adiantar consultas em detrimento de outrem;
- Faturamento para um serviço não coberto como um serviço coberto;
- Realizar procedimento ou usar órtese/próteses com qualidade inferior a declarada;

- Cobrança por implantação de equipe que não existe no local indicado;
- Relatórios incorretos de diagnósticos ou procedimentos;
- Corrupção (propinas e subornos);
- Emissão falsa ou desnecessária de medicamentos prescritos;
- Desvio de medicamentos;
- Desvio de insumos;
- Falsificação do ponto de presença em unidades de saúde;

Uma das hipóteses da causa de fraude, pode ser devido ao baixo índice de informatização das Unidades Básicas de Saúde, tendo como resultado, facilitação de processos fraudulentos que são difíceis de detectar. Outro ponto de falha que deve ser considerado é o processo de inserção de informações no sistema, em uma análise preliminar foi detectado que as informações são inseridas de forma muito genérica, levando em consideração apenas a cobrança de procedimentos mais comuns, deixando muito espaço para inserção manual de códigos e descritivos de operações, o que pode causar discrepância entre o descrito e o encontrado. É possível que as auditorias têm muitas dificuldades em gerar documentos com a comprovação das não conformidades, pois as condições ficam precárias, o que promove baixa fiscalização e possivelmente muita impunidade.

5.3.AUDITORIAS

5.3.1. SISTEMA DE AUDITORIAS NO SUS

Gerir um sistema complexo como o do SUS requer abordagens sistemáticas afim de detectar, prevenir e monitorar ações que gerem erros, fraudes e práticas abusivas. Como também, entender processos, atuar e apontar soluções que direcione esforços na resolução de problemas. A auditoria preenche bem esses requisitos servindo também com propósitos não punitivos e educadores. Auditoria é uma das ferramentas de qualidade mais eficientes para monitorar um sistema de gestão, pois, quando bem aplicada, diagnostica não conformidades no serviço avaliado, sendo executada por profissionais capacitados que apresentam, além do

conhecimento técnico-científico, atributos pessoais como imparcialidade, prudência e diplomacia, por meio de pareceres embasados por leis, portarias e resoluções (Noronha e Salles, 2004).

Esse mecanismo gerencial pode ser aplicado para averiguar o uso de recursos públicos, desempenho de equipes, cumprimento de normas e/ou especificações. É possível também, mensurar a qualidade de um produto ou serviço oferecido. A implantação de uma auditoria pode seguir diferentes tipos de metodologias, sendo facultado ao auditor a melhor forma de verificação. O sistema de saúde é uma atividade complexa que envolve vários setores e ações, como prevenção, cuidado, tratamento e promoção, em que atuam e interagem diferentes atores, entre os quais a população, corporações, gestores, dirigentes políticos, profissionais e segmentos empresariais (AYACH, 2013). O trabalho de auditoria no SUS é extremamente complexo, pois necessita de grande quantidade de informações que precisam ser cuidadosamente extraídas, trabalhadas e interpretadas, pois muitos interesses e responsabilidades estão em foco quando se audita a saúde (MINISTÉRIO DA SAÚDE, 2011). Para executar esta tarefa, existe um órgão especial do governo, DENASUS (Departamento Nacional de Auditoria do SUS), órgão integrante da estrutura da Secretaria de Gestão Estratégica e Participativa do Ministério da Saúde e componente federal do Sistema Nacional de Auditoria (SNA).

Esse órgão propõem uma metodologia em forma de cartilha que explica e informa como deve se proceder uma auditoria, existem muito pontos importantes nela, mas para ilustrar esse trabalho, será feito um recorte de pontos frágeis ou que aparentam fragilidade. Conforme explica o documento, as finalidades das auditorias são:

- Aferir a preservação dos padrões estabelecidos e proceder ao levantamento de dados que permitam conhecer a qualidade, a quantidade, os custos e os gastos da atenção à saúde;
- Avaliar os elementos componentes dos processos da instituição, serviço ou sistema auditado, objetivando a melhoria dos procedimentos, por meio da detecção de desvios dos padrões estabelecidos;
- Avaliar a qualidade, a propriedade e a efetividade dos serviços de saúde prestados à população, visando à melhoria progressiva da assistência à saúde;

- Produzir informações para subsidiar o planejamento das ações que contribuam para o aperfeiçoamento do SUS e para a satisfação do usuário.

Sobre os tipos de auditoria, eles definem como:

- Conformidade – Examina a legalidade dos atos de gestão dos responsáveis sujeitos a sua jurisdição, quanto ao aspecto assistencial, contábil, financeiro, orçamentário e patrimonial;
- Operacional – Avalia os sistemas de saúde, observados aspectos de eficiência, eficácia e efetividade.

E sobre a natureza das auditorias elas podem ser:

- Regular ou Ordinária – Ações inseridas no planejamento anual de atividades dos componentes de auditoria;
- Especial ou Extraordinária – Ações não inseridas no planejamento, realizadas para apurar denúncias ou para atender alguma demanda específica.

Porém, após esse ponto não se sabe se por omissão proposital ou por falta de metodologia, fica implícito a forma como se escolhe os processos a serem auditados. É usada a frase planejamento anual, denúncias ou demandas específicas. Não foi encontrado nesta pesquisa os meios pelos quais são escolhidas as auditorias, principalmente as que tratam das auditorias de conformidade. Um pressuposto que se levanta é que pela natureza política que norteia os três entes federados, é que com exceção das denúncias, a escolha seja arbitrária. Não técnica.

Avaliando a grande quantidade de registros gerados pelo SUS, fica ressaltado que os profissionais encarregados de realizar auditorias necessitam de ferramentas tecnológicas que os auxiliem a sistematizar a escolha de unidades a serem auditadas. Tomar esse tipo de decisão é humanamente impossível, levando em consideração o tamanho e diversidade da estrutura do sistema de saúde brasileira. Acredito que uma ferramenta informatizada, construída com máquinas de aprendizado, executando um sistema de classificação pode ajudar a melhorar a escolha de processos a ser auditados. Gerar resultados por intermédio de sistemas automáticos estaria de acordo com a resolução CFC nº 828/1998. E comporia a fase analítica da auditoria. O Tribunal de Contas da União considera como papéis de trabalho, entre outros, planilhas, formulários, questionários preenchidos, fotografias,

vídeos, áudios, arquivos magnéticos, ofícios, memorandos, portarias, cópias de contratos e outros documentos, matrizes de planejamento e de procedimentos, de achados de auditoria e de responsabilização(MINISTÉRIO DA SAÚDE, 2011).

5.3.2. SISTEMA DE AUDITORIAS ODONTOLÓGICAS

A auditoria odontológica é uma área de conhecimento na qual não existem estudos consistentes para a elucidação dos principais pontos de divergência entre auditores e cirurgiões-dentistas clínicos (EDUARDO DARUGE et al., 2011). Em dois trabalhos encontrados sobre auditoria e glosa dos procedimentos não fica claro o motivo principal que levou a rejeição dos procedimentos, em (EDUARDO DARUGE et al., 2011) fica claro quais os procedimento que são mais glosados, mas não fica claro o que leva a rejeição do procedimento. No trabalho de (AYACH; MOIMAZ; GARBIN, 2013) o autor afirma que 42,4% da glosa vem do mesmo serviço no mesmo dente no mesmo paciente, mas novamente existe uma incerteza se foi erro ou fraude. Observando a escassez de literatura sobre o assunto de fraude no Brasil, é notório a dificuldade de estabelecer pesquisas neste segmento, seja por falta de interesse do poder público em conhecer os números, seja por falta de interesse na pesquisa ou por inacessibilidade aos dados – como foi o caso deste trabalho.

5.4. INTELIGENCIA ARTIFICIAL

Turing propôs a pergunta: "As máquinas podem pensar?". Isto deve começar com as definições do significado dos termos "máquina" e "pensar"(TURING, 1950). Abrindo seu artigo mais famoso o Sr. Turing levanta esse questionamento e propõe um jogo. O jogo da imitação, o que viria a se chamar 'Teste de Turing'. Esse jogo vai completar 67 anos sem um vencedor, dada a complexidade de ações que um programa de computador deve tomar para 'enganar' o jogador humano.

Por definição, inteligência artificial pode ser definida coma ciência e a engenharia de fazer máquinas inteligentes, especialmente programas de computador inteligentes. Relaciona-se com a tarefa semelhante de usar

computadores para entender a inteligência humana, mas a AI não precisa se limitar a métodos biologicamente observáveis (“Basic Questions”, [s.d.]).

O nome inteligência artificial é um simbolismo complexo, definir inteligência humana é por si um conceito difícil de explicar e classificar, pois inteligência é um conjunto de habilidades humanas, distribuídas em diversas áreas cognitivas. Várias áreas da filosofia tentam entender sobre o processo que faz um ser inteligente. Posto isso, se torna ainda mais difícil definir o que significa inteligência artificial. John McCarthy, em um dos seus artigos questiona como um programa de computador perceberia o mundo, quais seriam as interpretações de suas entradas. Neste ponto ele afirma: Assim, alguns dos principais problemas tradicionais da filosofia surgem na inteligência artificial (MCCARTHY; HAYES, 1969). Em parte, esses questionamentos de 1969 continuam sem respostas. Muito se avançou nessa ciência, mas produzir inteligência em máquinas não é um trabalho trivial, já sabia o pesquisador John McCarthy, considerado um dos pais da Inteligência Artificial, juntamente com Marvin Minsky. Treinar algoritmos para procurar grupos específicos e padrões não é uma tarefa fácil, envolve estudos multidisciplinares complexos. É a junção de várias ciências estabelecidas, com novos campos de pesquisas.

Existem várias pesquisas que se utilizam de IA para resolver problemas diversos, mas nos escopo de saúde uma das promissoras são:

- Medicina de precisão, o que pode gerar tratamentos individualizados;
- Medicina cognitiva, onde softwares leem milhares de artigos médicos, cruza com os sintomas do paciente e informa ao médico caminhos que o ajudarão a descobrir a enfermidade do paciente;
- Ajudar pesquisadores a formular drogas mais eficientes;
- Detectar doenças genéticas com maior precisão;
- Disseminar pesquisas que melhorariam o conhecimento científico;
- Classificar e usar modelos preditivos para identificar pacientes que se beneficiariam com a medicina preventiva;
- Usar marcadores para diagnosticar doenças antes que os sintomas apareçam;
- Uso de análise em linguagem natural para criar uma nova geração de *chat boots* para ‘conversar’ com pacientes sobre sua condição de saúde e indicar profissionais que poderiam lhe auxiliar em algum tipo de tratamento;

- Detectar e minimizar fraudes e desperdício;
- Prever epidemias.

5.4.1. APRENDIZADO DE MÁQUINA

De forma simplista, podemos dizer que aprendizagem de máquina estuda algoritmos de computador para aprender a fazer coisas. Ela é muito eficiente em realizar análises em dados que humanos perdem a compreensão do tamanho ou tecnologias similares não conseguem processar. Os algoritmos de aprendizagem de máquina aprendem utilizando exemplos de dados no presente, para criar um modelo e resolver problemas similares submetidos posteriormente.

O Youtube em 2015 gerou 70 milhões de horas de visualização (“Estatísticas - YouTube”, [s.d.]). O Google indexa mais de 130 trilhões de páginas (“How Search Works - The Story – Inside Search – Google”, [s.d.]). O Large Synoptic Survey Telescope no Chile gera 140 TB de informação a cada 5 dias (“Data, data everywhere | The Economist”, [s.d.]). Este dilúvio de dados exige métodos automatizados de análise de dados, que é o que a aprendizagem de máquina fornece (MURPHY, 2012).

Máquinas de aprendizagem eram vistas com desconfiança e sempre a sombra de outras disciplinas. Décadas atrás, nenhuma grande empresa de software tinha interesses em desenvolver projetos nesta área. A abordagem era muito complexa, desenvolver regras para domínios simples poderiam consumir milhares de testes lógicos, porém, em meados dos anos 70, houve uma proposição se era possível ensinar aos algoritmos utilizando exemplos, essa visão mudou o rumo desta ciência.

Em 1983 iniciou a revolução, dezenas de artigos, teses e livros começaram a abordar o tema. Depois de um período de encantamento, as linhas de pesquisa concretas cristalizaram: indução das regras if-then para sistemas baseados no conhecimento; Indução de classificadores, programas capazes de melhorar suas habilidades com base na experiência; Ajuste automático de programas Prolog(KUBAT, 2015). Atualmente elas ocupam lugar de destaque em pesquisas. Sendo considerado um dos ramos mais promissores para um futuro próximo. A maioria das pessoas se utiliza dessa tecnologia sem perceber. Classificadores de spam, seleção e recomendações de filmes, músicas, produtos e amigos.

Aprendizado de máquina é uma subcategoria da inteligência artificial. Caracteriza-se pela junção da ciência da computação com a matemática. Um dos desafios dessa união é a construção de algoritmos eficientes que podem lidar com grandes bases de dados. A aprendizagem de máquina é a disciplina científica que focaliza em como os computadores aprendem dos dados (CLEOPHAS; ZWINDERMAN, 2014). O conceito entre autores não varia muito, todos concordam que o conhecimento e eficiência dos algoritmos vem por intermédio de exemplos contidos em dados. A aprendizagem de máquina é o estudo sistemático de algoritmos e sistemas que melhoram o seu conhecimento ou desempenho com a experiência (FLACH, 2012). É muito improvável que possamos construir qualquer tipo de sistema inteligente capaz de qualquer das facilidades que associamos à inteligência, como linguagem ou visão, sem usar o aprendizado para chegar lá (“COS 511: Theoretical Machine Learning”, 2008).

Os modelos estatísticos em geral são de dois tipos: supervisionados e não supervisionados.

5.4.2. SUPERVISIONADO

A aprendizagem supervisionada funciona com métodos de classificação. Suas ações envolvem escolher subgrupos que melhor descreva uma instancia de dados. E gerar previsões, que envolve estimar um parâmetro desconhecido.

No aprendizado supervisionado, as entradas e saídas são mapeadas, tendo como objetivos gerar um conjunto de pares entrada-saída. Divide-se os dados em dois conjuntos, o de treinamento e o de teste. Cada entrada de treinamento é um vetor D-dimensional, representando, informações de um domínio: batimentos cardíacos, temperatura, pressão, idade entre outros. Esses são os chamados atributos, porém, a entrada poderia ser um objeto estruturado complexo, como uma imagem, uma frase, uma mensagem de e-mail, uma série temporal, uma forma molecular, um gráfico, etc. Da mesma forma, a forma da variável de saída ou de resposta pode, em princípio, ser qualquer coisa, mas a maioria dos métodos assume que a entrada é uma variável categórica ou nominal de algum conjunto finito, (como mamífero ou não mamífero) ou uma escalar real (como o valor de venda de uma casa). Quando a entrada é categórica, o problema é conhecido como classificação

ou reconhecimento de padrões, quando entrada é valor real, o problema é conhecido como regressão (MURPHY, 2012).

5.4.2.1. NÃO SUPERVISIONADO

O segundo tipo principal de aprendizagem de máquina é a aprendizagem não supervisionada. Aqui temos apenas entradas **Erro! Fonte de referência não encontrada.** e o objetivo é encontrar padrões nos dados. Este é um problema muito menos bem definido, pois não existe conhecimento prévio de quais padrões devem ser encontrados, e não há nenhuma métrica de erro óbvio para usar (ao contrário da aprendizagem supervisionada, onde podemos comparar nossa previsão de y para um dado X ao valor observado) (MURPHY, 2012).

Aprendizagem não supervisionada pode ser exemplificada com algoritmos que encontra padrões de doenças coronarianas num exame de ECG. Ao contrário do aprendizado supervisionado os dados não todos os dados têm rótulos identificadores.

5.4.3. ESCOLHA DE ATRIBUTOS

Máquinas de aprendizado geram conhecimento automaticamente dos dados, no entanto o sucesso desta tarefa depende da qualidade dos dados que estão sendo processados, caso os dados não tenham correlação ou sejam irrelevantes, o conhecimento resultante pode ter sua veracidade diminuída ou mesmo não resultar em resultados úteis. Portanto, a seleção de atributos é considerada como uma etapa de pré-processamento para melhorar a tarefa de aprendizado da máquina. Seleção de atributos é um pré-processamento, técnica que localiza um subconjunto mínimo de atributos que captura as propriedades relevantes de um conjunto de dados para permitir uma classificação adequada (PHYU; OO, [s.d.]).

A escolha de atributos demanda um estudo separado, já que pode ser considerada uma das tarefas mais difíceis na mineração de dados. Essa atividade requer um conhecimento profundo sobre a natureza e os tipos de dados com os quais se está lidando. A escolha dos atributos errados pode confundir o algoritmo

preditivo, como também desperdiçar recursos computacionais processando informações irrelevantes e/ou duplicadas. Dessa forma deve-se criar um subconjunto dos dados originais e remover dados irrelevante e/ou duplicados.

Alguns estudos apontam resolução de problemas com 40 atributos (AMALDI; KANN, 1998). Nos últimos anos a maioria dos artigos explora domínios com centenas ou dezenas de milhares de variáveis ou características(GUYON; ELISSEFF; DE, 2003). Então, a escolha de atributos tem destaque na construção de modelos de aprendizagem de máquina. Novas técnicas estão sendo utilizadas para mitigar os problemas de escolhas de atributos. A natureza do aprendizado se divide em duas tarefas, definir quais atributos são relevantes e como combinar esses atributos. Nessa perspectiva, a seleção de características relevantes e a eliminação de irrelevantes, é um dos problemas centrais na aprendizagem de máquina (BLUM; LANGLEY, [s.d.]). A literatura no traz uma série de definições diferentes sobre o que é relevante em aprendizagem de máquina, o que torna uma variável relevantes? Uma importante pergunta é: importante para o quê? As importâncias dos atributos estão intimamente conectadas ao objetivo os dados se propões.

Existem muitas técnicas que são utilizadas para a preparação dos dados que serão treinados por algoritmos, destacamos:

- Seleção Univariada – A *Univariate feature selection* realiza testes estatísticos e podem ser usados para selecionar os atributos que têm o relacionamento mais forte com a variável de saída.
- Removendo atributo com baixa variação – *Variance Threshold* é uma abordagem simples para a seleção de atributos. Remove todos os atributos cuja variância não atinja algum limite. Por padrão, ele remove todos os atributo de variância zero, ou seja, os que têm o mesmo valor em todas as amostras(“1.13. Feature selection — scikit-learn 0.18.1 documentation”, [s.d.]).
- Eliminação de atributos Recursivo - A *Recursive Feature Elimination* (RFE) funciona eliminando recursivamente atributos e construindo um modelo sobre os atributos que permanecem.
- Análise do componente principal - A *Principal Component Analysis* (PCA) usa álgebra linear para transformar o conjunto de dados em uma forma compactada. Geralmente isso é chamado de técnica de redução de dimensão de dados. Uma propriedade de PCA é que você pode

escolher o número de dimensões ou componente principal no resultado transformado.

- Importância do atributo – Utiliza-se algoritmos de árvores de decisão para estimar a importância dos atributos.

5.4.4. PRINCIPAIS ALGORITMOS DE APRENDIZADO DE MÁQUINA

Cada problema necessita de uma abordagem diferente para sua resolução. É comum utilizar uma parcela significativa de tempo para determinar qual algoritmo e técnica é mais adequada para a realização de análises em dados, posto isso, neste estudo serão abordados apenas os algoritmos que tem alguma relevância para o resultado final deste trabalho.

ISOLATION FOREST

É utilizada para detectar anomalias em um conjunto de dados. O algoritmo parte do pressuposto que anomalia de dados são escassos e diferentes do conjunto normal de dados. Assim, esses pontos são mais facilmente encontrados por um mecanismo chamado de isolamento. Este método é altamente útil e é fundamentalmente diferente de todos os métodos existentes. Ele introduz o uso do isolamento como um meio mais eficaz e eficiente para detectar anomalias do que as medidas básicas de distância e densidade comumente usadas. O uso do isolamento permite que o método proposto, iForest, explore uma sub amostragem em uma extensão que não seja viável nos métodos existentes, criando um algoritmo que tenha uma complexidade de tempo linear com uma baixa constante e uma baixa exigência de memória(LIU; TING, [s.d.]).

SUPPORT VECTOR MACHINE

SVM (*Support Vector Machine*) são técnicas úteis para a classificação de dados. A rede de vetores de suporte é uma nova aprendizado de máquina para

problemas de classificação de dois grupos. A máquina conceitualmente implementa a seguinte ideia: os vetores de entrada são mapeados não linearmente para um espaço de grande dimensão. Neste espaço de característica é construída uma superfície de decisão linear. Propriedades especiais da superfície de decisão garantem a alta capacidade de generalização da máquina de aprendizagem (CORTES; VAPNIK, 1995). As vantagens das máquinas de vetor de suporte são:

- Eficaz em espaços dimensionais elevados;
- Ainda eficaz nos casos em que o número de dimensões é maior do que o número de amostras;
- Usa um subconjunto de pontos de treinamento na função de decisão (chamados vetores de suporte), portanto também é eficiente na memória;
- Versátil: diferentes funções do Kernel podem ser especificadas para a função de decisão. Os kernels comuns são fornecidos, mas também é possível especificar kernels personalizados (“1.4. Support Vector Machines — scikit-learn 0.18.1 documentation”, [s.d.]).

K-NEAREST NEIGHBORS

O princípio por trás dos métodos de vizinho mais próximo é encontrar um número predefinido de amostras de treinamento mais próximo da distância até o novo ponto e prever o rótulo a partir delas. O número de amostras pode ser uma constante definida pelo usuário (k-nearest mais próximo de aprendizagem), ou variar com base na densidade local de pontos (raio de base de aprendizagem vizinho). A distância pode, em geral, ser qualquer medida métrica: a distância euclidiana padrão é a escolha mais comum. (“1.6. Nearest Neighbors — scikit-learn 0.18.1 documentation”, [s.d.]

Apesar de sua simplicidade, k-nearest tem sido bem-sucedido em um grande número de problemas de classificação e regressão, incluindo dígitos manuscritos ou cenas de imagem de satélite. Sendo um método não-paramétrico, muitas vezes é bem-sucedido em situações de classificação onde o limite de decisão é muito irregular.

A aproximação k- nearest mais próximo tem sido bem investigada na literatura e mostrou ser uma poderosa técnica não paramétrica para estimativa de densidade

e classificação. No entanto, encontrar os vizinhos k -mais próximos de uma amostra de teste entre N amostras de projeto é um processo demorado, particularmente para grandes N (FUKUNAGE; NARENDRA, 1975).

PCA

A análise de componentes principais (PCA) provou ser uma técnica extremamente popular para a redução da dimensionalidade e é discutido extensamente na maioria dos textos sobre análise multivariada. Suas muitas áreas de aplicação incluem compressão de dados, análise de imagem, visualização, reconhecimento de padrões, regressão e previsão de séries temporais (TIPPING; BISHOP, [s.d.]).

A Análise de Componentes Principais tem sido considerada um dos resultados mais valioso da álgebra. PCA é usado abundantemente em todas as formas de análise - da neurociência à computação gráfica - porque é um método simples, não paramétrico de extrair informações relevantes de conjuntos de dados confusos. É possível com PCA reduzir um conjunto complexo de dados em uma dimensão inferior e extrair informações uteis.

A análise de componentes principais, identifica atributos correlacionados na distribuição de dados e projeta os dados para este subespaço de menor dimensão (HODGE; AUSTIN, 2004a). PCA assume que os subespaços determinados pelos componentes principais são compactos, o que limita a sua aplicabilidade, sobretudo para distribuições esparsas. Porém, é um pré-processador muito bom para selecionar um subconjunto de atributos para métodos que sofrem com os problemas de grandes dimensões. A PCA identifica a componente principal da maior variância, uma vez que cada componente tem um autovalor associado cuja magnitude corresponde à variância dos pontos do vector componente. PCA retém os principais componentes k com maior variação e descarta todos os outros para preservar a máxima informação e manter redundância mínima (HODGE; AUSTIN, 2004a).

DISTÂNCIA DE MAHALANOBIS

A distância de Mahalanobis é a distância entre um ponto de dados e um centroide de espaço multivariado. No contexto de detecção de outliers, a distância

de Mahalanobis é utilizada para calcular a distância de cada observação para o centro de seu grupo baseado na sua matriz de covariância (MARQUES, 2015). Ela é um dos métodos multivariados mais poderosos para detecção de outliers.

Mahalanobis é computacionalmente caro para calcular grandes conjuntos de dados dimensionais elevado, em comparação com a distância Euclidiana, pois requer percorrer todo o conjunto de dados para identificar as correlações de atributo (HODGE; AUSTIN, 2004a).

5.5. DETECÇÃO DE ANOMALIAS

A detecção de valores atípicos ou *outlier* é questão fundamental na mineração de dados, especificamente na detecção de fraudes. *Outliers* foram informalmente definidos como observações em um conjunto de dados que parecem ser inconsistentes com o restante desse conjunto (FERDOUSI; MAEDA, 2006). As anomalias são basicamente dados gerados de forma não-padrão, que destoa dos restantes dos dados, por essas características ele pode ser usado para detectar fraudes. Detectar anomalias pode ser uma tarefa crucial em determinados sistemas, detectar vibrações em uma aeronave, detectar intrusão em redes protegidas, verificar anomalias em radiografias e/ou ressonância. Em geral se busca estabelecer os padrões derivados dos dados. No entanto, encontrar anomalias é a busca por padrões raros e importantes que são gerados por uma parte dos dados analisados (HA; SEOK; LEE, 2015). Estudar esses eventos podem revelar características diferentes nos mecanismos que geram os dados, conhecer essas características geram informações uteis sobre anormalidades e a forma como são adquiridos. Um exemplo poderia ser analisar todos os dados dos sensores conectados a um paciente antes de um evento de um acidente vascular cerebral ou uma parada cardíaca e verificar anomalias nos dados e estabelecer uma relação entre as anomalias e o desencadear do evento.

Anomalias surgem devido a falhas mecânicas, alterações no comportamento do sistema, comportamento fraudulento, erro humano, erro de instrumento ou simplesmente através de desvios naturais nas populações (HODGE; AUSTIN, 2004b).

Os métodos de detecção de valores anômalos originais eram arbitrários, mas agora são utilizadas técnicas de princípio e sistemáticas, extraídas da gama completa de ciências da computação e estatística (HODGE; AUSTIN, 2004b).

Dois cenários, supervisionados e não supervisionados, são utilizados no desenvolvimento e aplicação de métodos de detecção de valores atípicos (HA; SEOK; LEE, 2015). É possível utilizar diversas formas na detecção de ruídos, mas a detecção de valores anômalos é geralmente considerada como um problema de aprendizagem não supervisionado devido à falta de conhecimento prévio sobre a natureza de vários objetos atípicos (SURI; MURTY; ATHITHAN, 2012).

5.6.FERRAMENTAS

5.6.1. SCIKIT-LEARN

Este projeto foi iniciado em 2007 como um projeto Google Summer of Code de David Cournapeau. Mais tarde naquele ano, Matthieu Brucher começou a trabalhar neste projeto como parte de sua tese. Em 2010 Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort e Vincent Michel do INRIA assumiram a liderança do projeto e fizeram o primeiro lançamento público, em 01 de fevereiro de 2010. Desde então, vários lançamentos surgiram após um ciclo de cerca de 3 meses e um próspero programa internacional. A comunidade tem liderado o desenvolvimento (“About us — scikit-learn 0.18.1 documentation”, [s.d.]). É uma biblioteca muito utilizada pela comunidade de analista de dados e vem arrecadando adeptos e sempre presentes em eventos e congressos de Python, AI, ML e neurociência.

O Python uma das principais, senão a mais popular linguagem de programação para aplicações científicas, todo esse ambiente e estrutura é utilizado pelo scikit-learn. Entre outras coisas ela inclui vários algoritmos de aprendizado supervisionado e não supervisionado, ferramentas de avaliação e avaliação de modelos, bem como ferramentas para pré-processamento e engenharia de recursos (“mloss | Project details:scikitlearn”, [s.d.]). Sendo considerada uma biblioteca de aprendizado de máquina de uso geral construída em Python, ela foi pensada para que pesquisadores não especialistas possam utilizar os algoritmos mais clássicos de

aprendizado de máquina para resolução vários dos seus problemas. Essa característica da biblioteca é muito importante, porque abre a possibilidade de área do conhecimento não especialista, possam utilizar o potencial das máquinas de aprendizado em seus dados. Entre elas podemos incluir: biologia, genética, negócios físicos e web, astronomia, física e muitos outros.

A utilização do ecossistema Python, juntamente com uma licença BSD flexível, favorece a incorporação dessa scikit-learn em muitos produtos finais, diminuindo a distância entre o tempo que se leva da concepção, teste e desenvolvimento, até a construção da aplicação como produto final. Essas características se refletem na grande demanda de utilização da biblioteca por empresas privadas. Em depoimento de Jeremy Achin CEO DataRobot diz: Nós estamos construindo a próxima geração de software de análise preditiva para tornar os cientistas de dados mais produtivos, scikit-learn é parte integrante do nosso sistema. A variedade de técnicas de aprendizagem de máquinas em combinação com as implementações sólidas que o scikit-learn oferece torna uma biblioteca imprescindível para aprendizado de máquina em Python. Jeremy Achin, CEO & Co-founder DataRobot Inc (“Who is using scikit-learn? — scikit-learn 0.18.1 documentation”, [s.d.]).

Existem muitas bibliotecas que se propõe a fazer o mesmo que o Scikit-learn, porém, existem características nela que a destaca de outras bibliotecas em Python. Algumas razões são:

- É distribuída sob a licença BSD;
- Incorpora código compilado para uma maior eficiência(Cython), diferente de MDP(ZITO et al., 2008) e pybrain (SCHAUL et al., 2010);
- É dependente apenas de numpy e scipy para facilitar a distribuição, diferente de pymvpa (HANKE et al., 2009) que tem dependências em R e shogun;
- Usa programação imperativa, diferente de pybrain que usa um data-flow framework. (SOFIYANTI; FITMAWATI; ROZA, 2015).

5.6.1.1. QUALIDADE E CONSISTÊNCIA

A ferramenta tem seu desenvolvimento focado na qualidade do código. Muitos recursos de engenharia de software são empregados para fazer dessa biblioteca mais fácil e confiável. As interfaces são bem similares não importa que implementação de algoritmos será utilizada, essa característica promove uma experiência de similaridades e consistência aos códigos.

Uma excelente documentação é disponibilizada para os usuários, com manuais, exemplos e explicações que facilitam a utilização. Incluindo documentação narrativa, referências de classe, um tutorial, instruções de instalação (SOFIYANTI; FITMAWATI; ROZA, 2015). A linguagem utilizada em manuais e explicações faz jus a sua característica de ser acessível a todos tipos de pesquisador, pois evita os jargões de aprendizado de máquina sempre que possível.

5.6.1.2. TECNOLOGIAS COMPLEMENTARES

Algumas tecnologias complementares são integradas a biblioteca para facilitar e fornecer estruturas consistentes que facilitem a análise e processamento dos dados, Numpy é uma delas. Ela é usada para dar a estrutura base usada para dados e parâmetros do modelo. Os dados de entrada são apresentados como matrizes numpy, integrando-se perfeitamente com outras bibliotecas Python científicas. Scipy conecta pacotes numéricos baseados em FORTRAN, sendo muito importante para portabilidade e instalação. Para aumentar a velocidade o Python foi combinado com o C na forma da linguagem Cython, fornecendo um desempenho melhor do que apenas o uso de linguagens de alto nível.

5.6.1.3. FACILIDADES E EFICIÊNCIA

A interface é um ponto forte da biblioteca, em sua arquitetura foi preferível utilizar interfaces do que herança, essa convenção de código deixa tudo muito simples de usar. Aliado a facilidade se tem eficiência, seu desenvolvimento foi focado em maximizar a eficiência computacional. O tempo de processamento do scikit-learn comparado a outras bibliotecas desenvolvidas em Python é bastante

vantajoso. Ela evita cópias de memória, o que fornece até 40% menos sobrecarga do que as conexões Python originais. O libsvm foi escrito de forma a melhorar a eficiência em dados densos, e utilizar as capacidades de pipeline dos processadores modernos.

6 TRABALHOS RELACIONADOS

Tradicionalmente o sistema financeiro é o segmento que faz mais uso de inteligência artificial, porém, muitas iniciativas em outras áreas estão adaptando e desenvolvendo novas técnicas de classificação para ampliar o domínio de aplicação. No caso da saúde o uso da inteligência artificial é tão necessário quanto no sistema financeiro, visto a quantidade de dados gerados, vide o tamanho do SUS no Brasil, e a diversidade de informações nesses dados, que torna as análises inviáveis para o ser humano.

Aprendizagem de máquina é uma técnica interdisciplinar que combina informática, estatística, ciência dos dados e ciência da computação, esse conjunto de ferramentas é bastante útil quando aplicados em dados voltados para saúde (BOONCHIENG; DUANGCHAEMKARN, 2016). Esse novo ramo da ciência da computação tem se revelado bem flexível e cabe como solução de muitos problemas em vários domínios. Um dos segmentos que fornece muito aporte para essa tecnológica é a saúde. Aprendizagem de máquina nos ajuda a extrair recursos úteis de uma grande quantidade de dados para resolver ou prever eventos relacionados à saúde, incluindo suporte à decisão médica, previsão, classificação, classificação, agrupamento, detecção de anomalias. Por exemplo, máquinas de aprendizado são utilizadas para classificar doenças cardíacas e definir suas severidades (HIJAZI et al., 2016) e muitas pesquisas estão apontando para diagnósticos utilizando sistemas preditivos de diagnósticos (CHEN et al., 2016).

No domínio biomédico, quase todos os conjuntos de dados são dados não estruturados e são geralmente multidimensionais (BOONCHIENG; DUANGCHAEMKARN, 2016). Somente com auxílio de Big Data e aprendizagem de máquina será possível construir modelos úteis de análises de dados. É possível unir dados de vigilância com dados meteorológicos, análise de tendências em redes sociais, informações de buscadores de internet e desenvolver algoritmos que pode processar essas informações e inferir epidemias em populações específicas e como elas se espalham. De posse destes dados medidas podem ser tomadas para melhor distribuir recursos humanos e financeiros para atender essa demanda.

Há pesquisas para a detecção de anomalias no sistema financeiro de indústrias farmacêuticas, com o objetivo de dar segurança a investidores potenciais

nesse segmento industrial (SHEN; YANG, 2015), utilização de aprendizado de máquina para verificação da taxa de mortalidade de pacientes mal nutridos após serem submetidos a cirurgias, podendo indicar cuidados especiais para pacientes oriundos de determinada cirurgia (WATCHARAPASORN, 2016).

A fundação Michael J. Fox, investiga a utilização da técnica de *Deep Learning* em dispositivos vestíveis no acompanhamento de pacientes com doença de Parkinson (ESKOFIER et al., 2016). Aprendizado de máquina tem sido usados para detectar anormalidade na retina de pacientes acometidos de diabetes (RAMAN; THEN, 2016). Há ainda a detecção de queda e desmaios em pessoas com epilepsia, utilizando sensores inerciais com Arduino (RIBEIRO et al., 2016).

Existe pesquisa na área de inteligência artificial na detecção de fraude no sistema odontológico em planos de saúde particular (MÁRIO; FONTENELLE, 2014). Muito se investe em detecção de fraude no sistema de saúde, planos de seguro saúde investem muito em novas técnicas que se adaptem a diferentes tipos de fraudes, é o caso de um sistema interativo de detecção de fraude e abuso eletrônico baseado em aprendizado de máquina em seguro de saúde (KOSE; GOKTURK; KILIC, 2015).

Muitas dessas pesquisas têm seu foco em *outlier*, o uso dessa técnica pode descobrir padrões ocultos e executando tarefas que foram consideradas impossíveis para uma máquina realizar.

Este trabalho, igual outros anteriormente citados, tem intuito de utilizar tecnologia para auxiliar profissionais a melhorar suas capacidades de promover o bem-estar e a saúde das pessoas acometidas de alguma enfermidade e necessitam de auxílio especializado, no caso deste trabalho, usuários do sistema público de assistência odontológica. Apesar de não ser escopo da pesquisa detectar dados de enfermidades diretamente é possível inferir algumas causas através de padrões gerados nos dados.

Diferente dos trabalhos citados, esse, busca promover a saúde de forma indireta, atuando para encontrar fraudes que tem como consequência enfermidades resultantes deste processo. A fraude pode se manifestar de várias formas, uma delas seria a baixa qualidade de atendimento, podendo gerar: perda de elementos dentários, dores, infecções e outras enfermidades desencadeadas de um atendimento não executado de forma adequada. A intenção deste trabalho é apontar anomalias em dados, mas que a investigação da sua origem traga benefícios à

população que utiliza destes serviços na forma de atendimentos mais abundantes e eficientes.

7 METODOLOGIA

A pesquisa foi de natureza tecnológica. Houve uma busca por utilizar de tecnologias que pudessem responder as perguntas do estudo e que fossem importantes para enriquecer a pesquisa e gerar um experimento. Utilizar tecnologia é um importante recurso de busca por progresso da ciência e para aplicabilidades na resolução de problemas reais, "...tornando-se a chave para a competitividade estratégica e para o desenvolvimento social e econômico de uma região"(SILVEIRA; BAZZO, 2009).

A pesquisa pode ser considerada com objetivo exploratório. O caráter multidisciplinar demandou muita pesquisa em assuntos de pouco domínio do pesquisador sobre o problema, eles tiveram que receber atenção para ajudar a constituir algumas hipóteses e enriquecer a pesquisa. Grande parte do trabalho utilizou a pesquisa bibliográfica, principalmente em artigos científicos, pois a maioria dos trabalhos relacionados estavam disponíveis neste tipo de publicação. Além de ser parte da investigação científica, a pesquisa bibliográfica é um importante instrumento na educação continuada do profissional (VOLPATO, 2000). Durante a pesquisa foram gerados vários experimentos para testar os dados colhidos e gerados durante a pesquisa.

O interesse em fraudes no sistema odontológico tem sido baixo, foi realizado pesquisa no sistema de busca Scielo usando os descritores; Fraude no SUS, Odontologia, SUS e fraude. Esses descritores não trouxeram resultados relevantes, quando esse estudo for publicado vai se somar a não mais que cinco estudos publicados especificamente sobre junção de inteligência artificial em busca de fraude no sistema odontológico. Em outras áreas estudos de aprendizado de máquina são bem abundantes.

Foram realizadas entrevistas com três especialistas, esses informaram sobre a quantidade de atendimento realizada em suas unidades de saúde e responderam livremente a perguntas feitas sobre a dinâmica dos serviços executados e sobre sua relação com as auditorias realizadas nas unidades de saúde.

Utilizando dados simulados desenvolveu-se um experimento para testar as hipóteses, esse procedimento foi considerado como a melhor forma de provar as hipóteses.

Para realizar os experimentos foram utilizados alguns softwares e equipamentos. Os itens para realizar os experimentos foram bem simples:

- Notebook Lenovo – G400s. Core i7 com 16GB de RAM e 500 GB de armazenamento em disco rígido.
- Monitor externo SAMSUNG de 22“.
- Editor de código PYCHARM (*communit edition*) para editar e testar códigos Python;
- Editor de código Atom – uso geral;
- LibreOffice Calc - para formatação dos arquivos CSV;
- Linux Mint 18 – sistema operacional;
- PHP 7.0;
- Python 2.7 e 3.5 – as duas versões estavam disponíveis para eventuais incompatibilidades nos elementos que compõe a biblioteca scikit-learn.
- Anaconda – esse componente foi bem importante para a pesquisa. Quando instalado o mesmo providencia todo ecossistema para a computação científica, instalando componentes, servidores embutidos, editores – Ipython -, scikit-learn e bases de dados de teste;
- Acesso à internet – realização de pesquisas e instalação de pacotes quando necessários.
- Scikit-learn – Essa é a biblioteca que foi utilizada para prover os algoritmos de aprendizado de máquina, instalando essa ferramenta foram economizados meses de estudo.

8 EXPERIMENTO

8.1. HIPOTÉSES

A literatura mostra que as fraudes são recorrentes em diversos segmentos do setor público e do setor produtivo, elas são sofisticadas e envolvem inúmeras formas de burla à sistemas que pode envolver indivíduo e/ou um grupo deles. Do ponto de vista econômico; é inviável localizar fraudes com analistas humanos em sistemas que transacionam centenas de milhares de dados. É possível que novas tecnológicas possam auxiliar neste processo.

O sistema odontológico do SUS possui as características citadas: pessoas e processos informatizados. Profissionais executando tarefas e informando através de sistemas os procedimentos que foram executados e suas quantidades. Segundo a literatura pesquisada, esse é um cenário onde a fraude pode se desenvolver. Assim surgem as hipóteses:

- Técnicas de Aprendizado de Máquina podem encontrar agrupamentos de dados que indiquem comportamentos anômalos em distribuição de dados oriundos de simulações de atendimentos odontológico?
- É possível utilizar o estudo governamental SBBrasil que revela estatísticas atuais das necessidades da população quanto a utilização dos serviços odontológico para estabelecer parâmetros de discrepâncias entre o encontrado por algoritmos em dados simulados e os percentuais indicados no estudo?

Ao investigar essas hipóteses, pode-se lançar luz sobre outros pontos derivados das mesmas questões. Como identificar problemas relacionados a:

- Baixa qualidade do atendimento;
- Não cumprimento da carga horária efetiva;
- Erros de digitação de diversos tipos:
 - Inserção do código do profissional no registro do e-SUS;
 - Inserção do procedimento errado;

- Indicação de tratamento no dente errado ou repetição do mesmo;
- A unidade de saúde pode estar assinalando serviços que não foram realizados, já que a inserção dos dados se dá a posteriori do atendimento;
- A unidade está assinalando serviços mais caros dos que os que realmente foram prestados;

8.2. GERAÇÃO DOS DADOS

Os dados foram baseados em entrevistas com três especialistas que trabalham no serviço público em unidades de saúde básica. Esses profissionais foram escolhidos por refletir a intenção de restringir o domínio de atuação deste trabalho. Para coletar os dados foi utilizado a 'Ficha de Atendimento Individual', que é bem comum aos CD entrevistados. A ficha de atendimento se assemelha ao software de preenchimento on-line realizados no software e-SUS AB, portanto, existe equivalência entre eles. Todos os campos presente da ficha possuem correspondente no software de preenchimento on-line, dessa forma o estudo considera que não existe diferença para a geração dos dados simulados tendo como modelo o software ou a ficha.

OS CD anotavam as quantidades de atendimentos que os mesmos realizavam de cada procedimento. Quando algum item ficava sem resposta lhes era perguntado a ausência da informação. Em geral a resposta está ligada a falta de infraestrutura na unidade de atendimento ou a falta de especialidade do profissional. A ficha de atendimento tem muitas variáveis, inicialmente foram suprimidos 22 dados relacionados a identificação dos profissionais e dos pacientes, restando 60 variáveis que foram objeto de estudo. Parte dos dados inseridos é referente a encaminhamento a outros profissionais e a entrega de material de higiene, essas 19 variáveis foram consideradas sem relevância para o estudo e foram descartadas.

Algumas técnicas foram utilizadas para verificar a relevância dos dados restantes para a aplicação dos algoritmos, boa parte do trabalho de mineração de dados está no pré-processamento, entender à natureza dos dados é essencial para o sucesso da eficácia dos resultados.

A quantidade de atendimentos por CD já era esperada, mas as entrevistas forneceram a realidade da distribuição, quais os procedimentos mais realizados por

paciente. A quantidade de atendimentos já havia sido pesquisada, então houve confirmação do encontrado na literatura e o entrevistado. Em geral um profissional pode atender um paciente a cada meia hora (VASCONCELLOS, 1980).

SAÚDE MÃO PERTO DE VOCE		SAÚDE SISTEMA S-SUS		FICHA DE ATENDIMENTO ODONTOLÓGICO INDIVIDUAL				DIGITADO POR: _____ DATA: / /																																																																																																																																												
				CONFERIDO POR: _____				FOLHA Nº: _____																																																																																																																																												
Nº DO CARTÃO SUS DO PROFISSIONAL		CBO		Cód. CHES UNIDADE		Cód. CHES EQUIPE		TURNO																																																																																																																																												
Nº DO CARTÃO SUS DO PROFISSIONAL		CBO		Nº DO CARTÃO SUS DO PROFISSIONAL		CBO		DATA: / /																																																																																																																																												
Local de Atendimento: 01 - UBS 02 - Unidade Móvel 03 - Rua 04 - Domicílio 05 - Escola/Creche 06 - Outros																																																																																																																																																				
<table border="1"> <thead> <tr> <th>Nº</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> <th>7</th> <th>8</th> <th>9</th> <th>10</th> </tr> </thead> <tbody> <tr> <td>Nº PRONTUÁRIO</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Nº CARTÃO SUS</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Data de nascimento</td> <td>Diá/mês</td> <td>/</td><td>/</td><td>/</td><td>/</td><td>/</td><td>/</td><td>/</td><td>/</td><td>/</td> </tr> <tr> <td></td> <td>Ano</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td>Local de atendimento</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Paciente com Necessidades Especiais</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Gestante</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td rowspan="5">Tipo Atendimento</td> <td>Consulta Agendada</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Emergência</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Atendimento de Urgência</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Primeira Consulta Odontológica Programática</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> <tr> <td>Consulta de Retorno em Odontologia</td> <td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td><td><input type="checkbox"/></td> </tr> </tbody> </table>										Nº	1	2	3	4	5	6	7	8	9	10	Nº PRONTUÁRIO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Nº CARTÃO SUS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Data de nascimento	Diá/mês	/	/	/	/	/	/	/	/	/		Ano										Local de atendimento	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Paciente com Necessidades Especiais	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Gestante	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Tipo Atendimento	Consulta Agendada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Emergência	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Atendimento de Urgência	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Primeira Consulta Odontológica Programática	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Consulta de Retorno em Odontologia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nº	1	2	3	4	5	6	7	8	9	10																																																																																																																																										
Nº PRONTUÁRIO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
Nº CARTÃO SUS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
Data de nascimento	Diá/mês	/	/	/	/	/	/	/	/	/																																																																																																																																										
	Ano																																																																																																																																																			
Local de atendimento	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
Paciente com Necessidades Especiais	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
Gestante	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
Tipo Atendimento	Consulta Agendada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
	Emergência	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
	Atendimento de Urgência	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
	Primeira Consulta Odontológica Programática	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										
	Consulta de Retorno em Odontologia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>																																																																																																																																										

Figura 1- Recorte da Ficha Individual de Atendimento

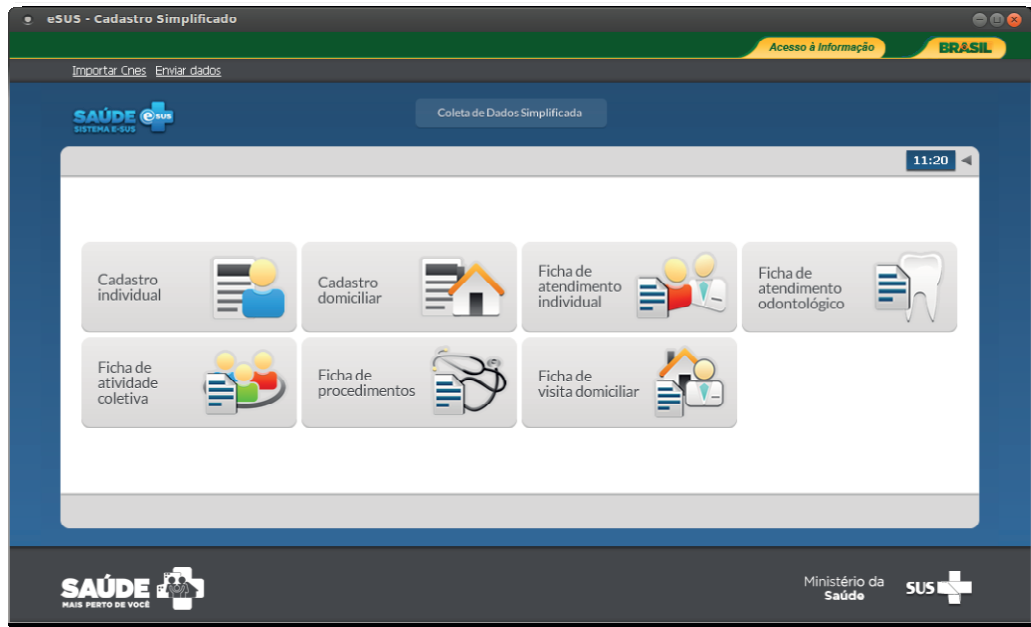


Figura 2- Tela do software e-SUS AB

Após obter as quantidades de atendimento dos CD, foi extraído a média e gerada populações compatíveis com esses números. Os dados refletem algumas disparidades entre técnicas individuais, inexistência de equipamentos e métodos de atuação e conduta de cada CD que trabalha na unidade. Fica claro que os mesmos serviços oferecidos em unidades diferentes têm diferenças ligadas a estrutura necessária para que cada profissional exerça suas atividades.

Alguns scripts em linguagem PHP foram utilizados para gerar as populações de dados que foram objeto do estudo. Essas rotinas consistiam em gerar arquivos em formato CSV com dados fictícios de atendimento.

A tabela contém os dezessete procedimentos mais relevantes executados na unidade de atendimento e seus respectivos valores, todos são atendimentos clínicos de baixa complexidade. O código é a anotação adotada pela tabela SIA-SUS de códigos de procedimentos que pode ser vista no anexo C. Intervenção é o procedimento executado e CD-001 o CD-004 são os cirurgiões que executaram os procedimentos e suas respectivas quantidades mensais. Esses dados são baseados numa carga horária de oito horas por dia, cinco dias por semana.

Tendo um conjunto próximo da realidade, novos conjuntos de dados foram criados para simular as anomalias e testar as hipóteses do trabalho proposto.

Tabela 1 – Dados gerados em entrevistas

Código	Intervenção	CD-001	CD-002	CD-003	CD-004
307020010	ACESSO A POLPA DENTÁRIA E MEDICAÇÃO (POR DENTE)	10	8	6	12
101020074	APLICAÇÃO TOPICA DE FLUOR (INDIVIDUAL POR SESSAO)	10	7	9	16
307010015	CAPEAMENTO PULPAR	3	5	0	0
307020029	CURATIVO DE DEMORA C/ OU S/ PREPARO BIOMECÂNICO	10	9	5	8
401010031	DRENAGEM DE ABSCESSO	1	0	3	2
404020097	EXCISÃO E SUTURA DE LESÃO NA BOCA	50	40	45	42
414020120	EXODONTIA DE DENTE DECÍDUO	10	8	5	3
414020138	EXODONTIA DE DENTE PERMANENTE	50	40	35	45
101020082	EVIDENCIAÇÃO DE PLACA BACTERIANA	50	35	50	43
301080011	ABORDAGEM COGNITIVA COMPORTAMENTAL DO FUMANTE	50	40	45	50
204010187	RADIOGRAFIA PERI-APICAL, INTER-PROXIMAL (BITE-WING)	20	0	30	5
307030016	RASPAGEM ALISAMENTO E POLIMENTO SUPRAGENGIVAIS (POR SEXTANTE)	20	18	16	20
307010023	RESTAURAÇÃO DE DENTE DECIDUO	20	16	25	18
307010031	RESTAURAÇÃO DE DENTE PERMANENTE ANTERIOR	50	40	50	40
307010040	RESTAURAÇÃO DE DENTE PERMANENTE POSTERIOR	60	30	25	35
301100152	RETIRADA DE PONTOS DE CIRURGIAS BÁSICAS (POR PACIENTE)	50	40	40	40
101020090	SELAMENTO PROVISORIO DE CAVIDADE DENTÁRIA	10	5	4	3
	Total	474	341	393	382

Quando plotado, os dados já revelam padrões que geram especulações:

1. Por que o dentista CD-001 é mais produtivo que os demais?
2. Por que os valores de exodontia de dente permanente é diferente dos valores de sutura, já que são procedimentos consecutivos?
3. Por que o CD-002 não realiza radiografias peri-apical?
4. Por que todos os CD têm índices tão altos de exodontia?

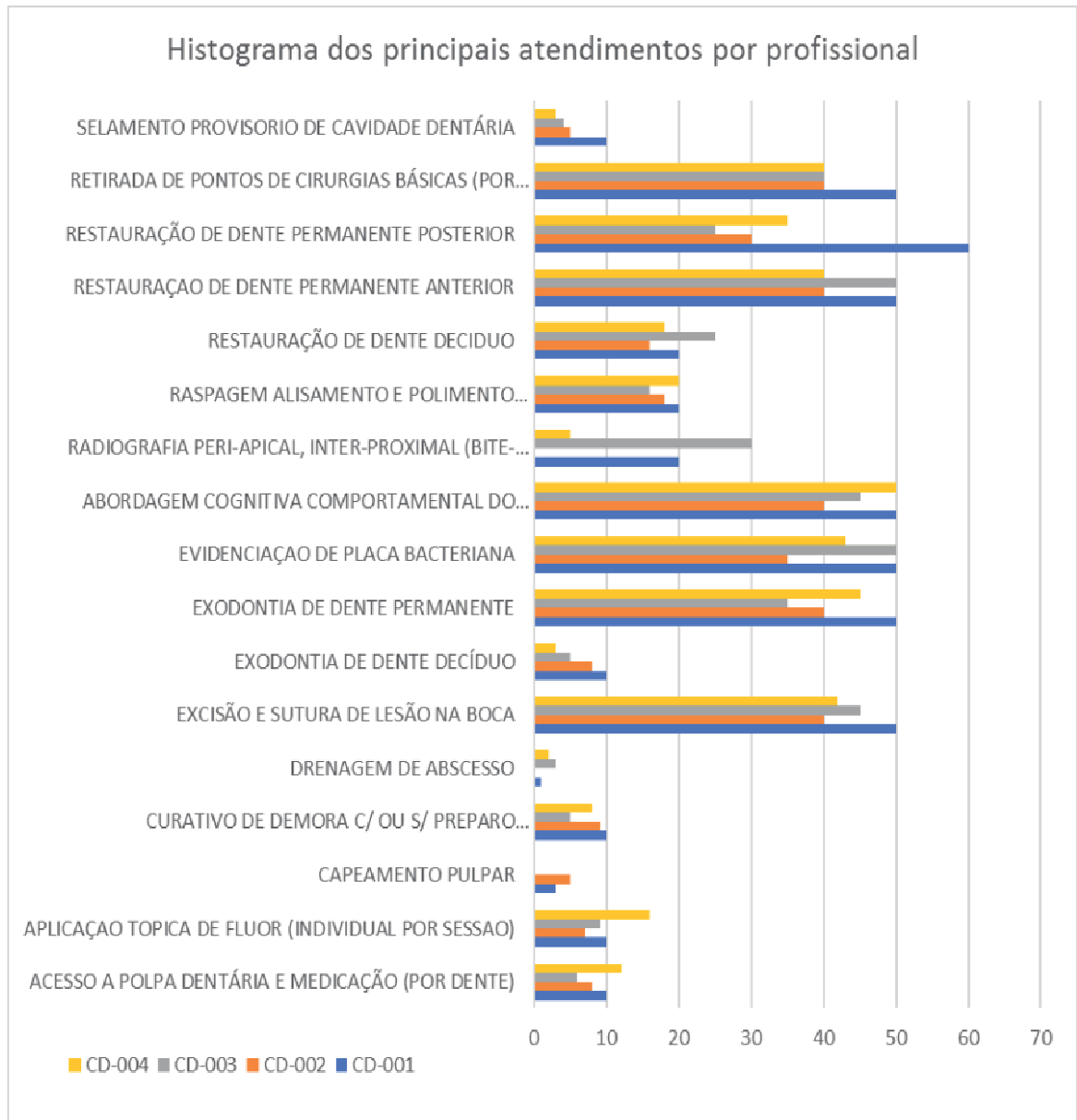


Figura 3- Histograma de Atendimento

É fácil para o operador humano perceber anomalias em gráficos e resumos quando eles tratam de um pequeno conjunto de dados, resumo de atendimento de quatro CD neste caso. É visualmente perceptível as anomalias numa distribuição normal de dados. Observamos os maiores e os menores e instintivamente os classificamos como anômalos. Essa situação muda se tivéssemos que plotar os 60 mil CD brasileiros e os seus atendimentos.

8.2.1. SELEÇÃO DOS ATRIBUTOS

Inicialmente foi gerado uma população de cem mil registros, e após algumas análises, notou-se a uniformidade dos dados, o que confundiu o modelo e não foram gerados resultados relevantes. Neste momento verificou-se a necessidade de traçar uma estratégia para modelar os dados na obtenção de resultados adequados.

A primeira ação foi eliminar atributos com baixa relevância para o estudo. Aplicou-se então a técnica de eliminação recursiva de atributos (RFE), o que diminui de 41 para apenas 10 o número de variáveis relevantes. Complementar a essa técnica utilizamos a redução de dimensão, para melhor correlacionar os dados (PCA) e facilitar a plotagem dos dados, reduzir o espaço dimensional e aumentar a correlação entre as variáveis melhora o rendimento dos algoritmos. Para a redução de atributos foi feito:

- Primeiro treina-se o modelo com todos os atributos e observa-se o desempenho.
- Depois exclui-se os 10% dos atributos mais fracos, repete-se o treinamento do modelo e observa a performance e o grau de certeza que está sendo gerados.
- Faz-se isso para mais atributos e verifica-se qual a combinação gerou os melhores resultados;
- Diminui-se o número de amostras.

8.2.2. O TREINAMENTO DOS MODELOS

O treinamento dos modelos se deu submetendo as populações de dados simulados aos algoritmos do framework. Nesse instante scikit-learning cria um modelo de classificação baseado no algoritmo selecionado.

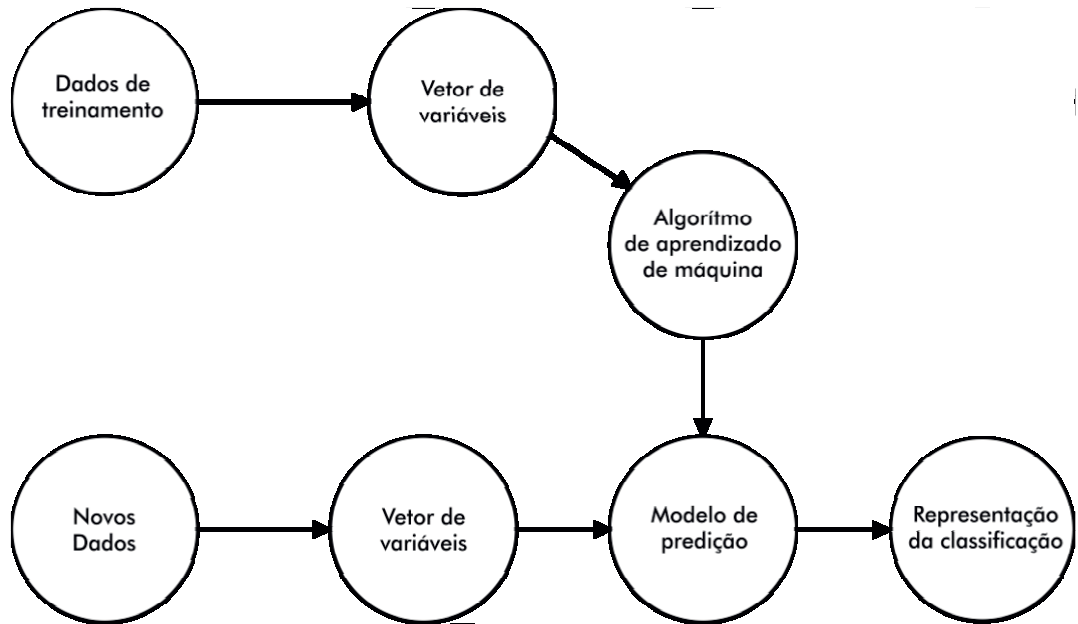


Figura 4- Representação do modelo de aprendizagem

Quando o modelo está treinado, novas populações com anomalias foram submetidas ao modelo. O resultado é plotado em forma de gráfico para verificar se houve a classificação de *clusters*.

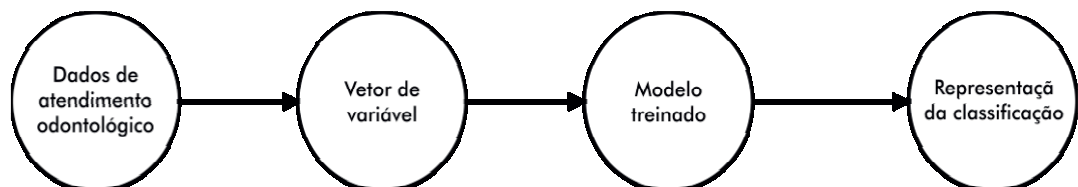


Figura 5- Submetendo dados ao modelo

9 RESULTADOS

Os dois primeiros testes foram desenvolvidos para responder se o classificador conseguiria selecionar atendimentos de complexidades distintas sendo executados numa mesma unidade de saúde. Uma unidade básica não deveria executar atendimentos de responsabilidades de outros centros de atendimento, visto que não possui o equipamento nem o especialista necessário para tal operação.

Submetendo essas populações de dados no classificador de densidade que usa a técnica conhecida com Distância de Mahalanobis. O resultado apresentado na Figura 6 mostra uma clara representação de como o classificador separa essas populações em *cluster*. O algoritmo conseguiu detectar densidades diferentes na geração de dados, pois a população anômala tinha um pequeno número de indivíduos com características distintas. As disparidades estão representadas em vermelho, enquanto a população normal está descrita em azul. O *box plot* deixa claro a diferença de densidade da população normal para a população anormal.

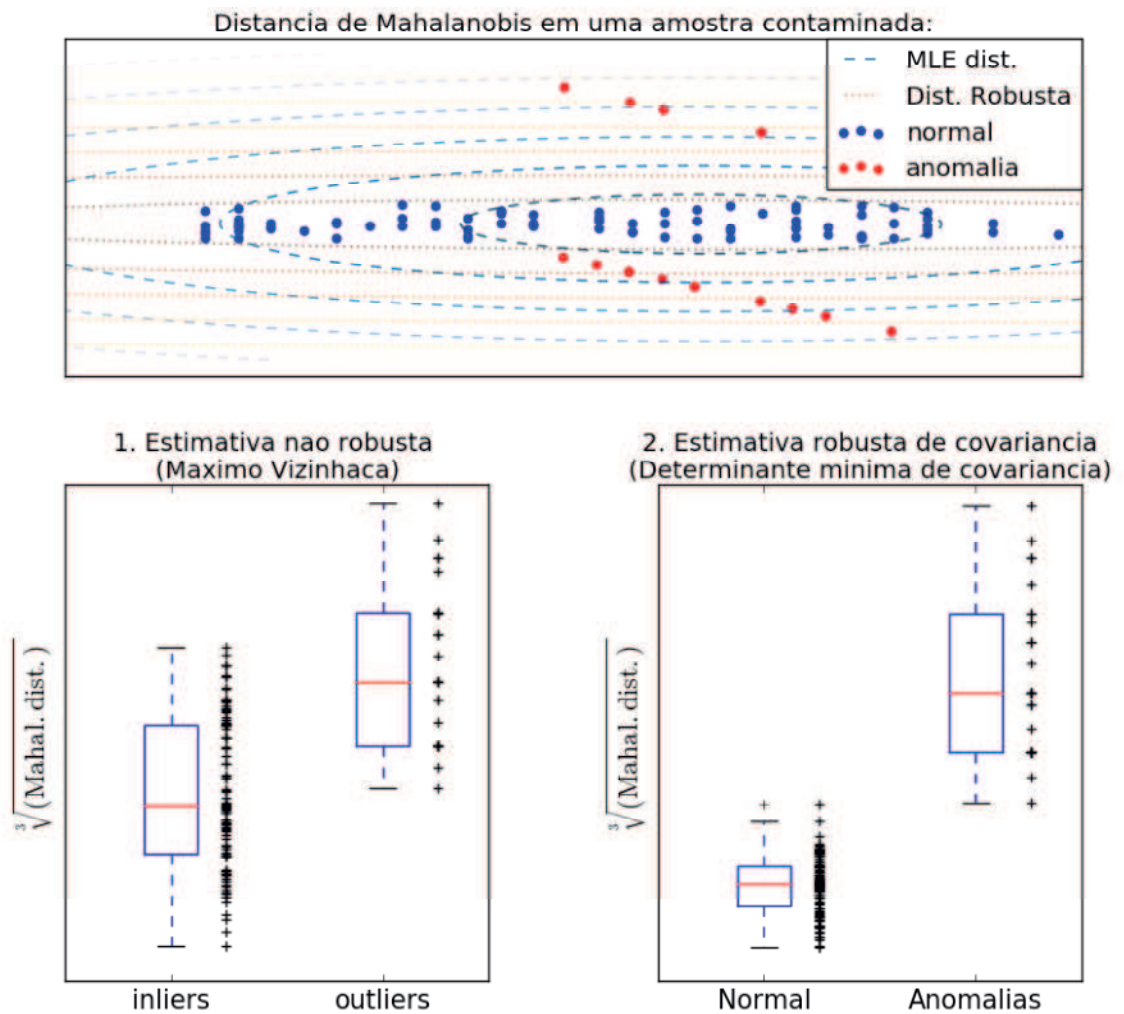


Figura 6- Gráfico MLE (Distância de Mahalanobis)

No segundo experimento aplicou-se o mesmo conjunto de dados do experimento anterior para verificar se técnicas distintas resultariam em resultados diferentes uns dos outros. No segundo experimento utilizamos o K-means, este algoritmo tenta separar os dados em *cluster*. Foi informado para o teste, dois *clusters*, sabendo que nosso teste consistia em um domínio discreto.

Os resultados estão apresentados na Figura 7, e foram semelhantes. Foi possível para o algoritmo perceber que existiam representações de dados que poderia ser dividida em dois segmentos.

É possível perceber que existe uma disparidade entre a geração normal de dados e as anomalias geradas, fato que facilita a identificação pelo algoritmo.

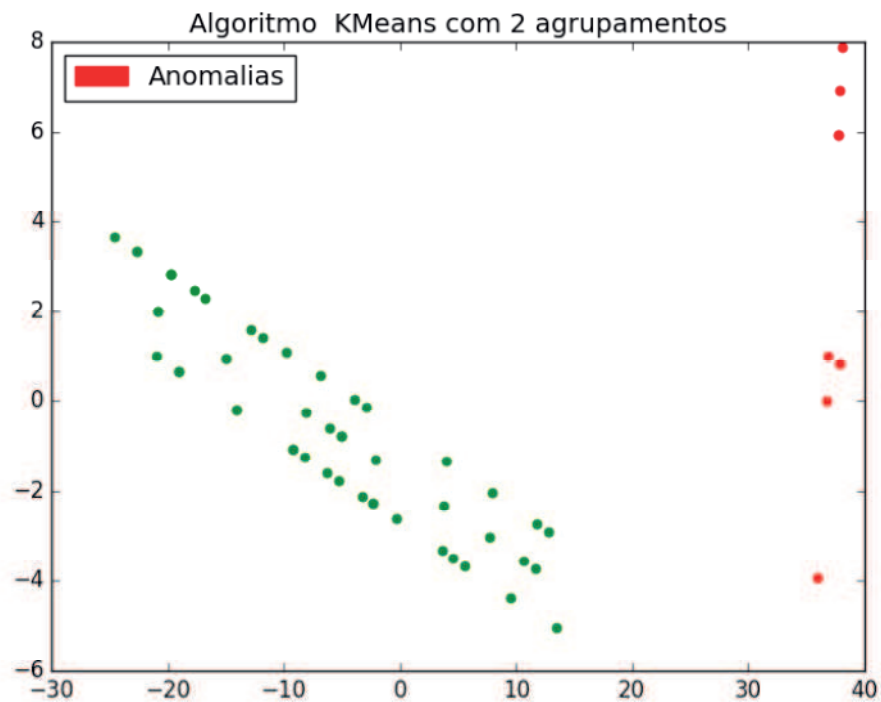


Figura 7- Gráfico plotado do algoritmo KMeans

No último experimento foi estabelecido o percentual de atendimento normal em adolescentes no Brasil e criadas duas populações, uma que estivesse dentro do percentual outra com dados anômalos. Foi usado o algoritmo de Floresta Isolada (*Isolation Forest*) para descobrir essas anomalias. Mesmo não existindo grande disparidade entre a população normal e as anomalias geradas o algoritmo foi capaz de identifica-los com precisão.

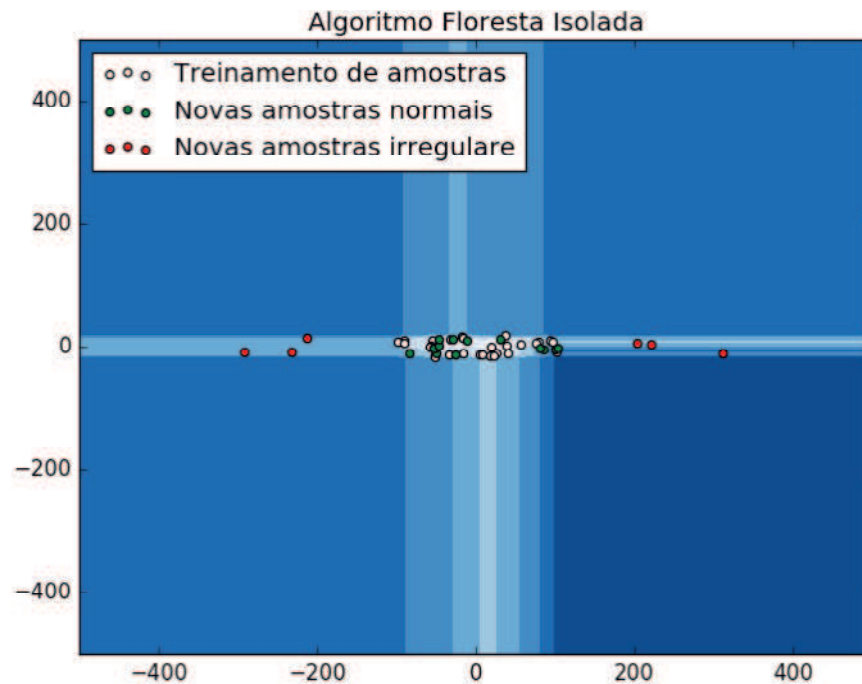


Figura 8- Gráfico plotado do algoritmo Isolation Forest

É possível utilizar aprendizado de máquina como ferramenta auxiliar na classificação de dados fraudulentos em dados de atendimento odontológico. Usando técnicas de classificação foi possível separar dados normais de atendimento de anomalias que podem ser consideradas fraudulentas.

Neste estudo preliminar, foi observado que os resultados plotados refletem os ruídos inserido nos dados analisados. Os três algoritmos estabeleceram uma classificação em clusters distintos; os dados normais e os anômalos.

Avaliar a performance de algoritmos de clusterização não é uma ação trivial. Muitos trabalhos usam técnicas diferentes de verificação, mas tudo depende do tipo de algoritmos que se está utilizando, *cross validation procedure*, métodos estatísticos e avaliação do tamanho do raio e da densidade do cluster, são alguns dos exemplos de verificação de performance. Neste trabalho houve apenas a avaliação visual da plotagem dos dados em clusters divididos por cores, estes estavam bem separados e com bordas bem definidas entre os clusters. São necessários mais estudos para estabelecer as melhores métricas de definição de performance para os resultados deste trabalho.

Existem uma preocupação em se estabelecer os índices de confiança dos algoritmos. Inicialmente todos os dados são gerados por um script em linguagem de programação, portanto, não temos dados reais do sistema odontológico do SUS. Implicando que qualquer índice de confiança que se estabelece pode ser questionado, visto que, a própria origem dos dados gera um alto grau de incerteza. Não existe forma de verificar a veracidade dos resultados gerados, pois os mesmos não têm paralelo com dados reais para poderem ser verificados

9.1.1. CONSIDERAÇÕES

Ao analisarmos a ficha de inserção dos dados de atendimento (ANEXO A), houve um entendimento que os procedimentos realizados são assinalados - para fins de recebimento do pagamento do SUS - mas a forma como tratamento é executado é desprezada, fato esse que torna praticamente não rastreáveis os procedimentos e técnicas utilizados pelo CD, dificultando uma auditoria de qualidade nos serviços prestados.

O presente trabalho apresenta uma abordagem, porém, outras podem ser utilizadas para conseguir alternativa dentre centenas de outras. É preciso entender a natureza dos dados, a quantidade de processamento, limitações financeiras, limitações técnicas, acurácia dos resultados e uma dezena de outras variáveis que norteiam a pesquisa em dados. "...incluindo: conjuntos de dados maciços, precisão, privacidade, medidas de desempenho e complexidade" (GOLMOHAMMADI; ZAIANE, 2012).

Detectar fraudes é uma tarefa difícil, mesmo tendo sido delimitado um domínio de baixa complexidade com é o caso da odontologia. Usar algoritmos de aprendizado de máquina para encontrar anomalias requer entendimento dos dados, ajustes e o uso de muitos especialistas. Não sendo possível deixar essa tarefa de forma automática. "... por isso esta tarefa é considerada como semiautomática, os médicos especialistas são obrigados a verificar as prescrições consideradas como suspeitas de fraude com algoritmos automáticos(ZHU; WANG; WU, 2011).

É plausível o uso de aprendizado de máquina aplicado aos dados gerados pelo Sistema Único de Saúde para obtenção de relatório indicativos de fraudes. Os sistemas de aprendizado de máquina não supervisionados - como os testados neste estudo - estão aptos a se adequar as mudanças dos métodos e modos de

fraudar, pois estará treinado para encontrar as anomalias geradas por mecanismos diferentes da geração normal de dados.

10 CONCLUSÕES

Mesmo o algoritmo indicado que existe um conjunto de dados destoante, outras possibilidades devem ser consideradas. O Sistema de introdução de dados nas unidades de atendimento é precário. Os registros das atividades não são inseridos diretamente no sistema, fichas são preenchidas e depois digitadas por um auxiliar ou digitador do ente federado. Essas mudanças sucessivas dos atores podem levar a erros – factíveis – que não significam necessariamente fraude, mas pode indicar ingerência, desorganização erros operacionais, baixa qualidade de atendimento, negligencia, descaso e outras formas não previstas e/ou ainda desconhecidas. Muitas dessas falhas podem ser refletidas para o sistema em forma de anomalias. Mesmo não sendo fraude no sentido principal, esses comportamentos são deletérios e penaliza os mais fragilizados que são os usuários do SUS.

Esse é uma pesquisa preliminar, muitos outros pontos têm que ser considerados, mas é plenamente possível utilizar tecnologias inteligentes para analisar dados do SUS e gerar informações uteis, não somente a busca de fraudes, mas respostas a pesquisas clinicas, controle e antecipação de epidemias, controle logístico e melhor distribuição de recursos. É possível utilizar aprendizado de máquina pode ajudar a gerar processos mais eficientes em unidades básicas de saúde, melhorando o acerto em diagnósticos e encaminhamentos, diminuindo a fila de espera e acelerando tratamentos.

É esperado para o ano de 2017 a informatização de mais de 90% das unidades de saúde, dessa forma, a união, estados e municípios necessitaram de estratégias para gerir os dados gerados em suas unidades, melhorar seu processo de decisão e estabelecer um controle mais efetivo da utilização de mão de obra e recursos. Centros de pesquisa como o NUTES podem propor diversos estudos neste sentido ao SUS, visto que possui infraestrutura e quadro funcional adequado.

Ficou compreendido no estudo que aprofundamentos devem ser realizados nessa linha de pesquisa, entender melhor a natureza dos dados, verificar a tenacidade e adequação dos algoritmos ao tamanho e a forma dos dados. Acredito que esse trabalho lança uma luz sobre como é possível contribuir para saúde pública utilizando computação, estatística, paciência e um computador simples. A

pesquisa em aprendizado de máquina é democrática e acessível a qualquer domínio de problema, bastando ao pesquisador utilizar das muitas ferramentas disponíveis gratuitamente, incluindo bibliotecas de aprendizado, cursos e dados experimentais.

É perceptível que para fornecer um aporte dessa magnitude o estado deve se equipar com ferramentas que o auxiliem nessa administração. Pois paralelamente a esse crescimento assistencial, também cresceu o número de fraudes, gerando uma pergunta: de que forma será possível fiscalizar a lisura dos dados oriundos dos entes federados? Do ponto de vista monetário é impossível verificar se todas as solicitações de pagamento por serviço prestado são lícitas.

Portanto, o objetivo desse estudo se alinha a necessidade do estado de ter ferramentas administrativas que forneçam informações de apoio a decisão no que tange as auditorias. Ter tal ferramenta implica em diminuição do prejuízo do estado com fraudes e melhoria na qualidade dos serviços prestados à população.

11 BIBLIOGRAFIA

1.13. Feature selection — scikit-learn 0.18.1 documentation. Disponível em: <http://scikit-learn.org/stable/modules/feature_selection.html>. Acesso em: 12 fev. 2017.

1.4. Support Vector Machines — scikit-learn 0.18.1 documentation. Disponível em: <<http://scikit-learn.org/stable/modules/svm.html>>. Acesso em: 12 fev. 2017.

1.6. Nearest Neighbors — scikit-learn 0.18.1 documentation. Disponível em: <<http://scikit-learn.org/stable/modules/neighbors.html>>. Acesso em: 13 fev. 2017.

1.9. Naive Bayes — scikit-learn 0.18.1 documentation. Disponível em: <http://scikit-learn.org/stable/modules/naive_bayes.html>. Acesso em: 13 fev. 2017.

6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python). Disponível em: <<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/>>. Acesso em: 13 fev. 2017.

71% dos brasileiros têm os serviços públicos de saúde como referência. Disponível em: <<http://portalsaude.saude.gov.br/index.php/cidadao/principal/agencia-saude/17961-71-dos-brasileiros-tem-os-servicos-publicos-de-saude-como-referencia>>. Acesso em: 12 mar. 2017.

About us — scikit-learn 0.18.1 documentation. Disponível em: <<http://scikit-learn.org/stable/about.html>>. Acesso em: 10 fev. 2017.

AMALDI, E.; KANN, V. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. **Theoretical Computer Science**, v. 209, n. 1–2, p. 237–260, dez. 1998.

ANDRADE, E. N. DE; ANDRADE, E. D. O. O SUS e o direito à saúde do brasileiro: leitura de seus princípios, com ênfase na universalidade da cobertura: [revisão]. **Rev. bio,t. (Impr.)**, v. 18, p. 61–74, 2010.

AYACH, C. Auditoria no Sistema Único de Saúde : o papel do auditor no serviço

odontológico Auditing in the Brazilian National Health System : the auditor ' s role in oral health services. **Saúde e Sociedade**, v. 22, n. 1, p. 237–248, 2013.

AYACH, C.; MOIMAZ, S. A. S.; GARBIN, C. A. S. Auditoria no Sistema Único de Saúde: o papel do auditor no serviço odontológico. **Saúde e Sociedade**, v. 22, n. 1, p. 237–248, mar. 2013.

Basic Questions. Disponível em: <<http://www-formal.stanford.edu/jmc/whatisai/node1.html>>. Acesso em: 13 mar. 2017.

BERALDO, N. **Ministro da Saúde reforça prazo para municípios adotarem o prontuário eletrônico**. Disponível em: <<http://portalsaude.saude.gov.br/index.php/cidadao/principal/agencia-saude/26106-ministro-da-saude-reforca-prazo-para-municipios-adotarem-o-prontuario-eletronico>>. Acesso em: 3 jan. 2017.

BLUM, A. L.; LANGLEY, P. Selection of Relevant Features and Examples in Machine Learning. [s.d.].

BOLTON, R. J. et al. Unsupervised Profiling Methods for Fraud Detection. **PROC. CREDIT SCORING AND CREDIT CONTROL VII**, p. 5--7, 2001.

BOONCHIENG, E.; DUANGCHAEMKARN, K. **Digital Disease Detection: Application of Machine Learning in Community Health Informatics**. 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). **Anais...IEEE**, jul. 2016 Disponível em: <<http://ieeexplore.ieee.org/document/7748841/>>. Acesso em: 14 dez. 2016

CAIXOTE, D. **Em 60 dias, posto de saúde sem prontuário eletrônico perderá recurso federal**. Disponível em: <<http://portalsaude.saude.gov.br/index.php/cidadao/principal/agencia-saude/25937-em-60-dias-posto-de-saude-sem-prontuario-eletronico-perdera-recurso-federal>>. Acesso em: 3 jan. 2017.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. **ACM Computing Surveys (CSUR)**, v. 41, n. September, p. 1–58, 2009.

CHAUHAN, P.; SHUKLA, M. A Review on Outlier Detection Techniques on Data Stream by Using Different Approaches of K-Means Algorithm. **2015 International Conference on Advances in Computer Engineering and Applications**, p. 580–585, mar. 2015.

CHEN, D. et al. **Diagnostic visualization for non-expert machine learning practitioners: A design study**. 2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). **Anais...IEEE**, set. 2016Disponível em: <<http://ieeexplore.ieee.org/document/7739669/>>. Acesso em: 14 dez. 2016

CLEOPHAS, T. J.; ZWINDERMAN, A. H. **Machine Learning in Medicine - Cookbook Three**. [s.l: s.n.].

Combating healthcare provider fraud and abuse. [s.d.].

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, set. 1995.

COS 511: Theoretical Machine Learning. 2008.

Covariance Matrix. Disponível em: <<http://stattrek.com/matrix-algebra/covariance-matrix.aspx>>. Acesso em: 13 fev. 2017.

Data, data everywhere | The Economist. Disponível em: <<http://www.economist.com/node/15557443>>. Acesso em: 7 fev. 2017.

DATASUS. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=01>>. Acesso em: 14 jan. 2017.

EDUARDO DARUGE, J. et al. Avaliação dos procedimentos clínicos mais glosados nos convênios odontológicos. **RFO**, v. 16, n. 2, p. 136–139, 2011.

ESKOFIER, B. M. et al. **Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson’s disease assessment**. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). **Anais...IEEE**, ago. 2016Disponível em: <<http://ieeexplore.ieee.org/document/7590787/>>. Acesso em: 14 dez. 2016

Estatísticas - YouTube. Disponível em: <<https://www.youtube.com/yt/press/pt-BR/statistics.html>>. Acesso em: 7 fev. 2017.

FERDOUSI, Z.; MAEDA, A. **Unsupervised Outlier Detection in Time Series Data.** 22nd International Conference on Data Engineering Workshops (ICDEW'06). **Anais...IEEE**, 2006Disponível em: <<http://ieeexplore.ieee.org/document/1623916/>>. Acesso em: 22 jan. 2017

FERRAZ, L. H. V. D. C. O SUS, o DATASUS e a informação em saúde: uma proposta de gestão participativa. **Cadernos de saude publica**, p. 109, 2009.

FLACH, P. **MACHINE LEARNING The Art and Science of Algorithms that Make Sense of Data.** [s.l.] Cambridge University Press, 2012.

FUKUNAGE, K.; NARENDRA, P. M. A Branch and Bound Algorithm for Computing k-Nearest Neighbors. **IEEE Transactions on Computers**, v. C-24, n. 7, p. 750–753, 1975.

GANGOPADHYAY, A.; CHEN, S. **Health Care Fraud Detection with Community Detection Algorithms.** 2016 IEEE International Conference on Smart Computing (SMARTCOMP). **Anais...IEEE**, maio 2016Disponível em: <<http://ieeexplore.ieee.org/document/7501694/>>. Acesso em: 20 jan. 2017

Global Health Care Anti-Fraud Network » U.S. — 2 Individuals Charged in \$31-Million Health Care Fraud Scheme. Disponível em: <<http://www.ghcan.org/2016/11/08/u-s-2-individuals-charged-in-31-million-health-care-fraud-scheme/#>>. Acesso em: 21 jan. 2017.

GOLMOHAMMADI, K.; ZAIANE, O. R. **Data Mining Applications for Fraud Detection in Securities Market.** 2012 European Intelligence and Security Informatics Conference. **Anais...IEEE**, ago. 2012Disponível em: <<http://ieeexplore.ieee.org/document/6298820/>>. Acesso em: 20 jan. 2017

GUYON, I.; ELISSEEFF, A.; DE, A. M. An Introduction to Variable and Feature Selection. **Journal of Machine Learning Research**, v. 3, p. 1157–1182, 2003.

HA, J.; SEOK, S.; LEE, J. S. A precise ranking method for outlier detection.

Information Sciences, v. 324, p. 88–107, 2015.

HANKE, M. et al. PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. **Neuroinformatics**, v. 7, n. 1, p. 37–53, 2009.

HIJAZI, S. et al. Machine Learning in Cardiac Health Monitoring and Decision Support. **Computer**, v. 49, n. 11, p. 38–48, nov. 2016.

HODGE, V. J.; AUSTIN, J. A survey of outlier detection methodologies. **Artificial Intelligence Review**, v. 22, n. 2, p. 85–126, out. 2004a.

HODGE, V. J.; AUSTIN, J. A Survey of Outlier Detection Methodologies. **Artificial Intelligence Review**, v. 22, n. 2, p. 85–126, out. 2004b.

How Search Works - The Story – Inside Search – Google. Disponível em: <<https://www.google.com/insidesearch/howsearchworks/thestory/>>. Acesso em: 7 fev. 2017.

KOSE, I.; GOKTURK, M.; KILIC, K. An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. **Applied Soft Computing Journal**, v. 36, n. November 2016, p. 283–299, 2015.

KUBAT, M. **An Introduction to Machine Learning**. [s.l: s.n.].

LIU, F. T.; TING, K. M. Isolation Forest. [s.d.].

MÁRIO, J.; FONTENELLE, B. O USO DA INTELIGÊNCIA ARTIFICIAL NA DETECÇÃO DE FRAUDES E ABUSOS EM TRATAMENTO ODONTOLÓGICOS. 2014.

MARQUES, H. O. Avaliação e seleção de modelos em detecção não supervisionada de outliers. p. 64, 2015.

MCCARTHY, J.; HAYES, P. J. Some philosophical problems from the standpoint of artificial intelligence. **Machine Intelligence**, v. 4, n. 463–502, p. 463–502, 1969.

MCKINNEY, W. **Python for Data Analysis**. First ed. [s.l.] O'REILLY, 2013.

MINISTÉRIO DA SAÚDE. **Auditoria do SUS**. [s.l: s.n.].

mloss | Project details:scikitlearn. Disponível em:

<<http://mloss.org/software/view/240/>>. Acesso em: 10 fev. 2017.

MURPHY, K. P. Machine Learning: A Probabilistic Perspective. **MIT Press**, p. 25, 2012.

PATTERSON, K. D. D. **Ethics of Big Data.** [s.l.] O'REILLY, 2012.

PHYU, T. Z.; OO, N. N. Performance Comparison of Attributes Selection for Machine Learning Task. [s.d.].

População teve acesso a 1,4 bi de consultas médicas pelo SUS em um ano — Portal Brasil. Disponível em: <<http://www.brasil.gov.br/saude/2015/06/populacao-teve-acesso-a-1-4-bi-de-consultas-medicas-pelo-sus-em-um-ano>>. Acesso em: 12 mar. 2016.

RAMAN, V.; THEN, P. Proposed Retinal Abnormality Detection and Classification Approach. p. 636–641, 2016.

RIBEIRO, E. et al. **On the use of inertial sensors and machine learning for automatic recognition of fainting and epileptic seizure.** 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom). **Anais...IEEE**, set. 2016Disponível em:

<<http://ieeexplore.ieee.org/document/7749420/>>. Acesso em: 14 dez. 2016

SCHAUL, T. et al. PyBrain. **Journal of Machine Learning Research**, v. 11, p. 743–746, 2010.

SCHERER, C. I.; SCHERER, M. D. DOS A. Advances and challenges in oral health after a decade of the “Smiling Brazil” Program. **Revista de Saude Publica**, v. 49, 2015.

SHEN, Q.; YANG, R. **Thompson-Tau Outlier Detection Method for Detecting Abnormal Data of Listed Pharmaceutical Companies in China.** 2015 8th International Symposium on Computational Intelligence and Design (ISCID). **Anais...IEEE**, dez. 2015Disponível em:

<<http://ieeexplore.ieee.org/document/7468973/>>. Acesso em: 14 dez. 2016

SILVEIRA, R. M. C. F.; BAZZO, W. Ciência, tecnologia e suas relações sociais: a percepção de geradores de tecnologia e suas implicações na educação tecnológica. **Ciência & Educação (Bauru)**, v. 15, n. 3, p. 681–694, 2009.

Sistemas - DATASUS. Disponível em: <<http://datasus.saude.gov.br/sistemas-e-aplicativos>>. Acesso em: 14 jan. 2017.

SOFIYANTI, N.; FITMAWATI, D. I.; ROZA, A. A. *Stenochlaena Riauensis* (Blechnaceae), A new fern species from riau, Indonesia. **Bangladesh Journal of Plant Taxonomy**, v. 22, n. 2, p. 137–141, 2015.

SURI, N. N. R. R.; MURTY, M. N.; ATHITHAN, G. Unsupervised feature selection for outlier detection in categorical data using mutual information. **Hybrid Intelligent Systems (HIS), 2012 12th International Conference on**, p. 253–258, 2012.

THIAGO MARZAGÃO. **Big Data no Planalto: como o governo minera datasets gigantescos para reprimir crimes**. Disponível em: <<https://www.infoq.com/br/presentations/big-data-no-planalto-como-o-governo-minera-datasets>>. Acesso em: 24 abr. 2017.

TIPPING, M. E.; BISHOP, C. M. Mixtures of Probabilistic Principal Component Analysers. **Neural Computation**, v. 11, n. 2, p. 443–482, [s.d.].

TURING, A. M. COMPUTING MACHINERY AND INTELLIGENCE. **Mind**, v. 49, p. 433–460, 1950.

VAN CAPELLEVEEN, G. et al. Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. **International Journal of Accounting Information Systems**, v. 21, p. 18–31, 2016.

VASCONCELLOS, M. DO C. C. O assistente odontológico na equipe de prestação de serviços odontológicos. **Revista de Saúde Pública**, v. 14, n. 1, p. 123–136, mar. 1980.

VOLPATO, E. D. S. N. Pesquisa bibliográfica em ciências biomédicas. **Jornal de Pneumologia**, v. 26, n. 2, p. 77–80, abr. 2000.

WATCHARAPASORN, M. P. The Surgical Patient Mortality Rate Prediction by

machine learning algorithms. 2016.

WEBB, G. I.; BOUGHTON, J. R.; WANG, Z. Not So Naive Bayes: Aggregating One-Dependence Estimators. **Machine Learning**, v. 58, n. 1, p. 5–24, jan. 2005.

Who is using scikit-learn? — scikit-learn 0.18.1 documentation. Disponível em: <<http://scikit-learn.org/stable/testimonials/testimonials.html>>. Acesso em: 10 fev. 2017.

YAN, K. et al. **A Hybrid Outlier Detection Method for Health Care Big Data.** 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom). **Anais...IEEE**, out. 2016 Disponível em: <<http://ieeexplore.ieee.org/document/7723688/>>. Acesso em: 14 dez. 2016

YUEN, K.-V.; ORTIZ, G. A. Outlier detection and robust regression for correlated data. **Computer Methods in Applied Mechanics and Engineering**, v. 313, p. 632–646, 2017.

YUFENG KOU et al. **Survey of fraud detection techniques.** IEEE International Conference on Networking, Sensing and Control, 2004. **Anais...IEEE**, [s.d.] Disponível em: <<http://ieeexplore.ieee.org/document/1297040/>>. Acesso em: 20 jan. 2017

ZAMONER, F. W. **Técnica de aprendizado semissupervisionado para detecção de outliers.** [s.l: s.n.].

ZHANG, H. The Optimality of Naive Bayes Naive Bayes and Augmented Naive Bayes. [s.d.].

ZHU, S.; WANG, Y.; WU, Y. **Health care fraud detection using nonnegative matrix factorization.** 2011 6th International Conference on Computer Science & Education (ICCSE). **Anais...IEEE**, ago. 2011 Disponível em: <<http://ieeexplore.ieee.org/document/6028688/>>. Acesso em: 20 jan. 2017

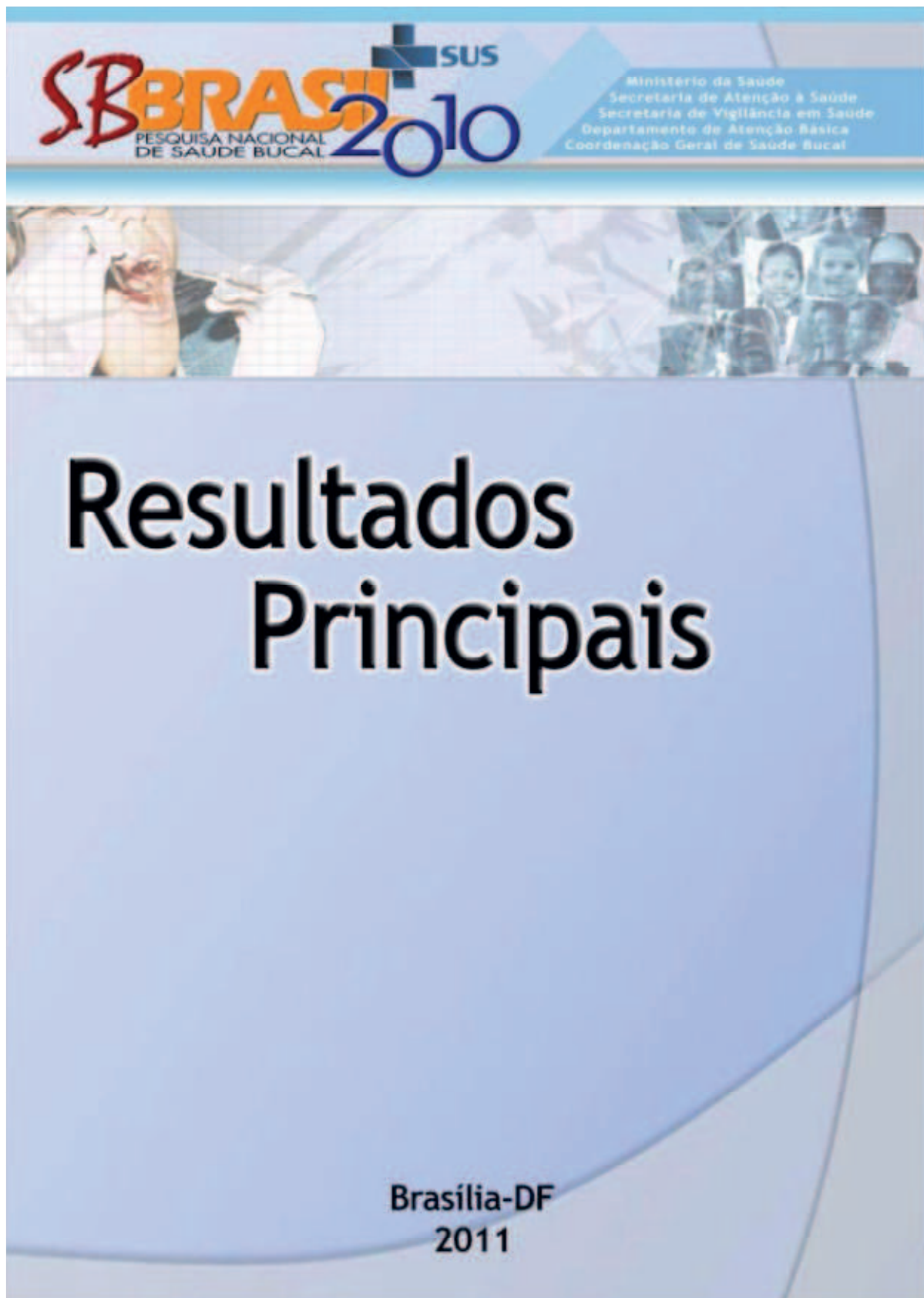
ZITO, T. et al. Modular toolkit for Data Processing (MDP): a Python data processing

framework. **Frontiers in Neuroinformatics**, v. 2, p. 8, 2008.

Nº		1	2	3	4	5	6	7	8	9	10	11	12	13	
PROCEDIMENTOS (Quantidade realizada)	Profilaxia / remoção da placa bacteriana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Pulpotomia dentária	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Radiografia Periapical / Interoaxial	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Raspagem alívio e polimento supragengivais (por sextante)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Raspagem alívio e polimento subgengivais (por sextante)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Restauração de dente decíduo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Restauração de dente permanente anterior	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Restauração de dente permanente posterior	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Retirada de pontos de cirurgias básicas (por paciente)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Selamento provisório de cavidade dentária	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Tratamento de alveólite	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Urticária / Urticomas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Outros (DA)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
FORNECIMENTO	Escova Dental	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Creme Dental	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Fio Dental	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
CONDUZIR*	Retorno para consulta agendada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Agendamento p/ Outros Profissionais/AB	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Agendamento p/ NASF	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Agendamento p/ Grupos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Tratamento Concluído	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	ENCAMINHAMENTO	Atendimento à Pacientes c/ necessidades especiais	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Cirurgia BMF	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Endodontia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Estomatologia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Implantodontia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Odontopediatria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Ortodontia / Ortognatia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Periodontia		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Prótese Dentária		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Radiologia		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Outros	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		

Legenda: Opção Múltipla de Escolha Opção Única de Escolha (Marcar X na opção desejada)
 Local de Atendimento: 01 - UBS 02 - Unidade Móvel 03 - Rua 04 - Domicílio 05 - Escola/Creche 06 - Outros 07 - Polo (Academia da Saúde) 08 - Instituição / Abrigo
 09 - Unidade prisional ou anexo 10 - Unidade socioeducativa
 * Campo Obrigatório
 ** Este campo não é obrigatório caso o tipo de atendimento for de demanda espontânea

ANEXO B – ESTUDO SB BRASIL 2010



ANEXO C – TABELAS SIA-SUS COM CÓDIGOS DE ATENDIMENTO

ANEXO 42 - TABELA SIA-SUS SIH-SUS DAS CONSULTAS - PROCEDIMENTOS – ATENDIMENTOS – ACOMPANHAMENTOS DO CIRURGIÃO DENTISTA CLÍNICO <small>(nome descrito para este CBO na Tabela Unificada => CIRURGIÃO DENTISTA CLÍNICO GERAL - CBO 223208)</small> E CIRURGIÃO DENTISTA DE SAÚDE DA FAMÍLIA <small>(nome descrito para este CBO na Tabela Unificada => CIRURGIÃO DENTISTA DE SAÚDE DA FAMÍLIA - CBO 2232B1)</small> 6ª VERSÃO 2008 Conforme a Tabela Unificada SIA-SIH SUS – SIGTAP 3.0.1.3 (de 06/11/07)	
<small>27/12/2007 – Última atualização SIGTAP 3.0.1.3 (SIGTAP Desktop – competência 08-2008)</small>	
<small>Estão disponíveis no Site do SIA/S, no endereço www.pmf.sc.gov.br/saude as Planilhas para consulta: Relat6ria Procedimento X CBO 223208 Relat6ria Procedimento X CBO 2232B1 Relat6ria Relat6rio Prod X CID Principal do Anexo 42</small>	
<small>Sugerimos consultar as demais Planilhas que s6o disponibilizadas no Site do SIA/S. Para sempre que a SigTap Desktop, em sua 6rea de compet6ncia, apresenta situa66es como os Procedimentos utilizados na Produ666o do SIA-SUS, estas Planilhas s6o disponibilizadas no endereço: www.pmf.sc.gov.br/saude em Informa66es em Sa6de, Sistema RAU RAAC.</small>	
CÓDIGO	DESCRIÇÃO
0301080011	ABORDAGEM COGNITIVA COMPORTAMENTAL DO FUMANTE (POR ATENDIMENTO /PACIENTE)
0307020010	ACESSO A POLPA DENTÁRIA E MEDICAÇÃO (POR DENTE)
0101020058	APLICAÇÃO DE CARIOSTÁTICO (POR DENTE)
0101020066	APLICAÇÃO DE SELANTE (POR DENTE)
0101020074	APLICAÇÃO TÓPICA DE FLUOR (INDIVIDUAL POR SESSÃO)
0301060037	ATENDIMENTO DE URGÊNCIA EM ATENÇÃO BÁSICA
90078	ATUALIZAÇÃO DE PRONTUÁRIO – INFOSAÚDE
0307010015	CAPEAMENTO PULPAR
0307040011	COLOCAÇÃO DE PLACA DE MORDIDA
0414010019	CONTENÇÃO DE DENTES POR SPLINTAGEM <small>Exceto para o Cirurgião Dentista de Saúde da Família</small>
0301010137	CONSULTA /ATENDIMENTO DOMICILIAR NA ATENÇÃO BÁSICA
0301010030	CONSULTA DE PROFISSIONAIS DE NÍVEL SUPERIOR NA ATENÇÃO BÁSICA (EXCETO MÉDICO) <small>Solicita66es ao Profissional, preferencialmente utilizar este c6digo, quando 6rbita atuando no Unidade de Sa6de (US)</small>
0301010048	CONSULTA DE PROFISSIONAIS DE NÍVEL SUPERIOR NA ATENÇÃO ESPECIALIZADA (EXCETO MÉDICO) <small>Solicita66es ao Profissional, preferencialmente utilizar este c6digo, quando 6rbita atuando em Pol6clinica e nas Unidades de Pronto Atendimento (UPA) Exceto para o Cirurgião Dentista de Saúde da Família</small>
0307020029	CURATIVO DE DEMORA C/ OU S/ PREPARO BIOMECÂNICO
0401010031	DRENAGEM DE ABSCESSO <small>EXIGE REGISTRO DO CID-10 PRINCIPAL (campo CID-10 ? folha RAAC)</small>
0101020082	EVIDENCIAÇÃO DE PLACA BACTERIANA
0404020097	EXCISÃO E SUTURA DE LESÃO NA BOCA

ANEXO 58 - TABELA SIA-SUS SIH-SUS DAS CONSULTAS – PROCEDIMENTOS ATENDIMENTOS – ACOMPANHAMENTOS DO CIRURGIÃO DENTISTA ENDODONTISTA CEO II – Centro de Especialidades Odontológicas Tipo II (nome descrito para este CBO na Tabela Unificada => CIRURGIÃO DENTISTA-ENDODONTISTA - CBO 223212) 6ª VERSÃO 2008 Conforme a Tabela Unificada SIA-SIH SUS – SIGTAP 3.0.1.3 (de 06/11/07)	
FF122007 – Última atualização 02/04/2008 (SIGTAP Dentlog – competência 03-2008)	
O Anexo 52 – TABELA SIA-SUS SIH-SUS DAS CONSULTAS – PROCEDIMENTOS – ATENDIMENTOS – ACOMPANHAMENTOS DO CIRURGIÃO DENTISTA CLÍNICO - CBO 223068 e DO DIFUNDIDO DENTISTA DE SAÚDE DA FAMÍLIA- CBO 223201 Agrupado às atividades do Cirurgião Dentista em Clínica Geral e Integrã o XE do CEO II	
Tabela disponível no Site do SUS, no endereço www.pmf.sc.gov.br/saude as Relações para consulta: Relação Procedimentos X CBO 223212 Relação Relação Proc X CBO Principal do Anexo 52	
Sugere-se consultar as demais Relações que são disponibilizadas no Site do SUS. Para sempre que a DigTaj Dentlog, em sua nova competência, apresenta atualizações para os Procedimentos utilizados na Produção de SIA-SUS, estas Relações são disponibilizadas no endereço: www.pmf.sc.gov.br/saude as informações em Saúde, Sistema RAJ/ RAAC.	
CÓDIGO	DESCRIÇÃO
0307020010	ACESSO A POLPA DENTÁRIA E MEDICAÇÃO (POR DENTE)
0414020022	APICECTOMIA C/ OU S/ OBTURAÇÃO RETROGRADA
0301060061	ATENDIMENTO DE URGÊNCIA EM ATENÇÃO ESPECIALIZADA
90078	ATUALIZAÇÃO DE PRONTUÁRIO – INFOSAUDE
0307010015	CAPEAMENTO PULPAR
0307040011	COLOCAÇÃO DE PLACA DE MORDIDA
0414010019	CONTENÇÃO DE DENTES POR SPLINTAGEM
0301010048	CONSULTA DE PROFISSIONAIS DE NÍVEL SUPERIOR NA ATENÇÃO ESPECIALIZADA (EXCETO MÉDICO)
0307020029	CURATIVO DE DEMORA C/ OU S/ PREPARO BIOMECÂNICO
0307020037	OBTURAÇÃO DE DENTE DECIDUO
0307020045	OBTURAÇÃO EM DENTE PERMANENTE BIRRADICULAR
0307020053	OBTURAÇÃO EM DENTE PERMANENTE C/ TRÊS OU MAIS RAÍZES
0307020061	OBTURAÇÃO EM DENTE PERMANENTE UNIRRADICULAR
0307020070	PULPOTOMIA DENTÁRIA
0204010160	RADIOGRAFIA OCLUSAL
0204010187	RADIOGRAFIA PERI-APICAL, INTERPROXIMAL (BITE-WING)
0307030032	RASPAGEM CORONO-RADICULAR (POR SEXTANTE)
0307020088	RETRAMENTO ENDODONTICO EM DENTE PERMANENTE BI-RADICULAR
0307020096	RETRAMENTO ENDODONTICO EM DENTE PERMANENTE C/ 3 OU MAIS RAÍZES
0307020100	RETRAMENTO ENDODONTICO EM DENTE PERMANENTE UNI-RADICULAR

ANEXO 60 - TABELA SIA-SUS SIH-SUS DAS CONSULTAS – PROCEDIMENTOS ATENDIMENTOS – ACOMPANHAMENTOS DO CIRURGIÃO DENTISTA TRAUMATOLOGISTA BUCO MAXILO FACIAL CEO II - Centro de Especialidades Odontológicas Tipo II (nome descrito para este CBO na Tabela Unificada => CIRURGIÃO DENTISTA-TRAUMATOLOGISTA BUCOMAXILOFACIAL - CBO 223268) 6ª VERSÃO 2008 Conforme a Tabela Unificada SIA-SIH SUS – SIGTAP 3.0.1.3 (de 06/11/07)	
(P/12/2007 – Última atualização 05/04/2008 (SIGTAP Dentist) – competência 05-0908)	
O Anexo 60 – TABELA SIA-SUS SIH-SUS DAS CONSULTAS – PROCEDIMENTOS – ATENDIMENTOS – ACOMPANHAMENTOS DO CIRURGIÃO DENTISTA CLÍNICO - CBO 223268 e DO CIRURGIÃO DENTISTA DE SAÚDE DA FAMILIA - CBO 223261 apresenta as atividades de Cirurgião Dentista em Clínica Geral e integra a RE do CEO II	
Este(s) dependent(es) no Site do SIA/S, no endereço www.pmf.sc.gov.br/saude ou Relatório para consulta: Relatório Procedimento X CBO 223268 Relatório Resumo Plan. X CID Principal do Anexo 60 Relatório Resumo Plan. X CID Principal do Anexo 61	
Superintende consultar os serviços Relatórios que são disponibilizados no Site do SIA/S. Para sempre que a DUT/OD Dentista, em sua nova competência, apresenta atualizações para os Procedimentos utilizados na Produção do SIA-SUS, estes Relatórios são disponibilizados no endereço: www.pmf.sc.gov.br/saude em informações em Saúde, Sistema RAH/RAC.	
CÓDIGO	DESCRIÇÃO
0307020010	ACESSO A PÓLPA DENTÁRIA E MEDICAÇÃO (POR DENTE)
0414020014	ALVEOLOTOMIA /ALVEOLECTOMIA (POR ARCO DENTARIO)
0414020022	APICETOMIA C/ OU S/ OBTURAÇÃO RETROGRADA
0414020030	APROFUNDAMENTO DE VESTIBULO ORAL (POR SEXTANTE)
0301060061	ATENDIMENTO DE URGÊNCIA EM ATENÇÃO ESPECIALIZADA
90078	ATUALIZAÇÃO DE PRONTUÁRIO – INFOSAÚDE
0201010232	BIÓPSIA DE GLÂNDULA SALIVAR
0201010526	BIÓPSIA DOS TECIDOS MOLES DA BOCA EXIGE REGISTRO DO CID-10 PRINCIPAL (campo CID-10 1 linha RAH)
0307010015	CAPEAMENTO PULPAR
0307040011	COLOCAÇÃO DE PLACA DE MORDIDA
0301010048	CONSULTA DE PROFISSIONAIS DE NÍVEL SUPERIOR NA ATENÇÃO ESPECIALIZADA (EXCETO MÉDICO)
0414010019	CONTENÇÃO DE DENTES POR SPLINTAGEM
0414020049	CORREÇÃO DE BRIDAS MUSCULARES
0414020057	CORREÇÃO DE IRREGULARIDADES DE REBORDO ALVEOLAR
0414020073	CURETAGEM PERIAPICAL
0401010031	DRENAGEM DE ABSCESSO EXIGE REGISTRO DO CID-10 PRINCIPAL (campo CID-10 1 linha RAH)
0414020081	ENXERTO GENGIVAL
0401010066	EXCISÃO E/ OU SUTURA SIMPLES DE PEQUENAS LESÕES /FERIMENTO DE PELE, ANEXOS E MUCOSA
0414020120	EXODONTIA DE DENTE DECIDUO
0414020138	EXODONTIA DE DENTE PERMANENTE