UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JESAÍAS CARVALHO PEREIRA SILVA

UM MÓDULO INTELIGENTE BASEADO EM APRENDIZADO DE MÁQUINA PARA TREINAMENTO DE ESTUDANTES DE MEDICINA NO DOCTRAINING

Mossoró-RN

2020

JESAÍAS CARVALHO PEREIRA SILVA

UM MÓDULO INTELIGENTE BASEADO EM APRENDIZADO DE MÁQUINA PARA TREINAMENTO DE ESTUDANTES DE MEDICINA NO DOCTRAINING

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Linha de Pesquisa: Otimização e Inteligência computacional
Orientador: Profº Dr. Araken de Medeiros Santos,
Coorientador: Profº Dr. Francisco Milton Mendes Neto

Mossoró-RN

2020

S586m Silva, Jesaías Carvalho Pereira.
Um módulo inteligente baseado em aprendizado de máquina para treinamento de estudantes de medicina no DocTraining / Jesaías Carvalho Pereira Silva. - 2020.
111 f. : il.

Orientador: Araken de Medeiros Santos.
Coorientador: Francisco Milton Mendes Neto.
Dissertação (Mestrado) - Universidade Federal Rural do Semi-árido, Programa de Pós-graduação em Ciência da Computação, 2020.

1. Aprendizado de Máquina. 2. Medicina. 3. Jogo. 4. Múltiplos Classificadores. 5. Casos clínicos. I. Santos, Araken de Medeiros, orient. II. Mendes Neto, Francisco Milton, co-orient. III. Título.

JESAÍAS CARVALHO PEREIRA SILVA

UM MÓDULO INTELIGENTE BASEADO EM APRENDIZADO DE MÁQUINA PARA TREINAMENTO DE ESTUDANTES DE MEDICINA NO DOCTRAINING

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: 31 / 03 / 2020

Araken de Medeiros Santos

Prof. Dr. Araken de Medeiros Santos
Orientador e Presidente

Prof. Dr. Francisco Milton Mendes Neto
Coorientador - Universidade Federal Rural do Semi-Árido - UFERSA

Prof. Dr. Francisco Chagas de Lima Júnior
Membro Interno – Universidade do Estado do Rio Grande do Norte - UERN

Prof. Dr. Daniel Sabino Amorim de Araújo
Membro Externo - Universidade Federal do Rio Grande do Norte - UFRN

Dedico este trabalho especialmente a Deus, aos meus pais, Raimundo Nonato Pereira Silva e Isanete Carvalho Pereira Silva, e aos meus irmãos, Jesreel Carvalho Pereira Silva e Raisa Carvalho Pereira Silva.

Agradecimentos

Agradeço primeiramente a Deus, por ser sempre a luz que me guia em todos os momentos de minha vida, sem ele eu não poderia conseguir nada.

A minha família, pelo apoio diário e carinho que sempre tiveram comigo. Em especial, aos meus Pais, Raimundo Nonato Pereira Silva e Isanete Carvalho Pereira Silva, por nunca desistir de mim, sempre me incentivarem, acreditarem em meus sonhos e me instruírem por toda a minha vida. Aos meus irmãos, Raisa Carvalho Pereira Silva e Jesreel Carvalho Pereira Silva, pelo carinho, apoio e amizade, mesmo na distância.

Ao amigo, professor e orientador Araken de Medeiros Santos, pela orientação, apoio, dicas e por acreditar no meu trabalho para sempre ir mais além.

Sou grato ao amigo, professor e coorientador Francisco Milton Mendes Neto pela paciência, confiança, conselhos, orientações, acompanhamentos e todos os demais elementos que contribuíram para minha evolução como pessoa e cientista.

Também deixo meus agradecimentos, muito especial, aos meus amigos Glenda Muniz, Igor Melo, Arthur Domingues e Ademar Neto pela amizade, parceria, confiança e ajuda que foram fundamentais para o desenvolvimento de toda a pesquisa científica.

Além disso, também deixo meus agradecimento para os meus amigos do mestrado em Ciência da Computação (UERN/UFERSA), em especial, aos meus amigos do LES e LCC, que fazem parte de toda trajetória acadêmica juntos.

A todos os professores do Programa de Pós-Graduação em Ciências da Computação que contribuíram direta ou indiretamente para o meu crescimento.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio e por acreditarem no potencial dos seus pesquisadores.

Muito obrigado a todos que não tiveram nomes mencionados, mas contribuíram de uma forma ou outra em minha formação acadêmica.

"Grandes coisas fez o Senhor por nós, e, por isso, estamos alegres."

Salmos 126:3

Resumo

Nos últimos anos, estudantes de medicina do Brasil obtiveram baixo desempenho no exame do Conselho Regional de Medicina do Estado de São Paulo. Embora a avaliação mais recente mostre melhorias, eles ainda apresentam um desempenho ruim em especialidades como casos clínicos, diabetes e infarto do miocárdio. Nesse contexto, emergem várias estratégias para melhorar o ensino e a aprendizagem dos estudantes de medicina, uma delas são os jogos sérios. No entanto, existe um problema ao tentar treinar estudantes de medicina em casos clínicos por meio de um jogo, pois há um grande número de doenças, e o processamento de um banco de dados contendo essas doenças é trabalhoso, além de que é necessário sempre estar com essas informações atualizadas, visto que com o passar do tempo novas doenças vão surgindo e sintomas podem variar. Esta pesquisa propõe um módulo inteligente para auxiliar no gerenciamento de um jogo sério, realizando a classificação de dados médicos por meio de modelos de aprendizado de máquina e os disponibilizando a um jogo sério. Esse módulo inteligente faz parte do DocTraining, um projeto que conta com um jogo sério para treinamento dos estudantes de medicina. Este módulo possui um site para disseminação, gerenciamento e expansão de dados. As informações são disponibilizadas para o jogo sério por meio de um *web service*. Foram utilizados bancos de dados de casos clínicos, doença cardíaca e diabetes. Esses bancos de dados são rotulados por múltiplos classificadores de aprendizado de máquina que agregam as classificações por meio de confiança e voto majoritário. O módulo inteligente foi validado computacionalmente e com profissionais da área da medicina. Na validação computacional, os classificadores utilizaram a métrica de precisão e medida-F, por meio de divisão de porcentagem e validação cruzada. Em relação à validação com os profissionais, foi realizada uma entrevista semiestruturada com uma professora e um técnico em laboratório do curso de medicina da Universidade Federal Rural do Semi-Árido. Por meio dos resultados encontrados, foi possível observar que a combinação de classificadores melhorou os resultados iniciais. Destacando-se a combinação de *Naive Bayes*, *Máquina de vetores de suporte* e *Regressão logística* utilizando Confiança para classificação na base de doença cardíaca; *Máquina de vetores de suporte*, *Árvore de decisão* e *Regressão logística* usando Voto Majoritário para classificação de Diabetes; e *Máquina de vetores de suporte*, *K-Nearest Neighbor*, *Naive Bayes* e *Regressão logística* utilizando Confiança para classificação na base de Casos Clínicos. Os profissionais demonstraram que o módulo inteligente do DocTraining pode gerenciar os dados do jogo sério, tendo potencial para evoluir e se tornar uma ferramenta capaz de auxiliar no ensino-aprendizagem dos estudantes e professores de medicina.

Palavras-chave: Aprendizado de Máquina, Medicina, Jogo, Múltiplos Classificadores, Casos clínicos.

Abstract

In recent years, medical students from Brazil have obtained poor performance in the Regional Medical Council of São Paulo State exam. Although the most recent evaluation shows improvements, they still perform poorly in specialties such as clinical cases, diabetes and myocardial infarction. In this context, several strategies emerge to improve the teaching and learning of medical students, one of which is serious games. However, there is a problem when trying to train medical students in clinical cases through a game, as there are a large number of diseases, and processing a database containing these diseases is laborious, and it is always necessary to be with this updated information, as over time new diseases arise and symptoms may vary. It is necessary to update the data so that students have access to the most recent. This research proposes an intelligent module to assist in the management of a serious game, performing the classification of medical data through machine learning models and making them available to a serious game. In addition, it is also necessary that these data are easily administered by medical professionals who use the smart module. This smart module is part of DocTraining, a project that has a serious game for training medical students. This module has a website for disseminating, managing and expanding data. The information is made available for the serious game through a web service. Databases of clinical cases, heart disease, and diabetes were used. These databases are labeled using multiple machine learning classifiers that aggregate the classifications through trust and majority vote. The intelligent module was validated computationally and with medical professionals. In computational validation, the classifiers used the metric of Precision and F1 Score, using percentage split and cross-validation. In relation to validation with professionals, we conduct a semi-structured interview a professor and a laboratory technician from the medical course at the Federal Rural University of the Semi-Arid. With the results found, it was possible to observe that the combination of classifiers improved the initial results. Highlighting the combination of Naive Bayes, Support vector machine and Logistic regression using Confidence to classify the basis of heart disease; Support vector machine, Decision Tree and Logistic Regression using Majority Vote for classification of Diabetes; and Support vector machine, K-Nearest Neighbor, Naive Bayes and Logistic regression using Confidence for classification based on Clinical Cases. The professionals have shown that DocTraining intelligent module can manage serious game data, with the potential to evolve and become a tool capable of assisting the teaching-learning of medical students and teachers.

Keywords: Machine Learning, Medicine, Game, Ensemble Classifiers, Clinical Cases.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AD	*Árvore de Decisão*
AM	Aprendizado de Máquina
ANS	Aprendizado Não Supervisionada
APR	Aprendizado Por Reforço
AS	Aprendizado Supervisionado
ASS	Aprendizado Semissupervisionado
ANASEM	Avaliação Nacional Seriada dos Estudantes de Medicina
CE	Critérios para Exclusão
CI	Critérios para Inclusão
CQ	Critérios para Qualidade
CREMESP	Conselho Regional de Medicina do Estado de São Paulo
CSDA	Criador de Sintomas, Doenças e Amostras
DP	Divisão de Porcentagem
EGG	Eletroencefalografia
ENADE	Exame Nacional de Desempenho de Estudantes
EVP	*Emotive Virtual Patient*
FA	*Floresta aleatória*
IEC	Comissão Eletrotécnica Internacional
IEEE	*Institute of Electrical and Electronics Engineers*
IES	Instituições de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IMC	Índice de massa corporal
ISO	Organização Internacional de Normalização

JSON	*JavaScript Object Notation*
KDD	*Knowledge Discovery in Databases*
KNN	*K-Nearest Neighbor*
MD	Mineração de Dados
MEC	Ministério da Educação
MLC	Múltiplo Classificador
MLP	*Multilayer perceptron*
MVS	*Máquina de Vetores de Suporte*
NB	*Naive Bayes*
NC	Notas dos Concluintes
NPC	*Non Playable Characters*
PROMED	Projeto de Incentivo a Mudanças Curriculares para os Cursos de Medicina
QP	Questão Primária
RL	*Regressão Logística*
RSL	Revisão Sistemática de Literatura
SimDeCS	Simulação para Tomada de Decisão no Serviço de Saúde
SUS	Sistema Único de Saúde
TAM	*Tecnology Accept Model*
TDAH	Transtorno do Déficit de Atenção com Hiperatividade
UFERSA	Universidade Federal Rural do Semi-Árido
VC	Validação Cruzada
XP	Pontos de Experiência

Sumário

1 INTRODUÇÃO

Neste Capítulo, é apresentado todo o contexto da pesquisa que se baseia em uma análise de dados dos cursos de medicina do ensino superior brasileiro, motivação, problema de pesquisa, objetivos desta pesquisa, metodologia de pesquisa para alcançar os objetivos propostos e a estrutura deste documento.

1.1 CONTEXTO

Foram publicados em 2016, pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) resultados da Avaliação Nacional Seriada dos Estudantes de Medicina (ANASEM), aplicadas aos cursos de medicina de 256 instituições do Brasil. De acordo com os resultados obtidos 41,3% dos 22.086 participantes relataram desconhecimento do conteúdo como a maior dificuldade encontrada na prova e 36,9% relataram que estudaram alguns dos conteúdos da prova ou a maioria, mas não os aprendeu (BRASIL, 2016).

O INEP também realiza a avaliação e classificação dos cursos de graduação do país, utilizando os resultados do Exame Nacional de Desempenho de Estudantes (ENADE), este por sua vez, avalia a cada três anos o rendimento dos concluintes dos cursos de graduação, de acordo com os conteúdos programáticos de cada curso. O INEP deu início em 2004 aos ciclos de avaliação trienal para cada área do conhecimento dos cursos superiores, por intermédio do exame ENADE. Objetiva-se por meio do exame avaliar o desempenho dos estudantes com relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos de graduação, o desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional, e o nível de atualização dos estudantes com relação à realidade nacional e internacional (INEP, 2018b).

Assim, são elaborados indicadores de qualidade da Educação Superior. Após as avaliações é gerado o conceito ENADE, que é expresso em uma escala contínua por meio de cinco níveis, sendo o conceito 5 o maior, 1 o menor e 3 a média nacional (INEP, 2018a). No exame de 2016, foram aplicadas provas para 195.859 participantes, isso representa 90,7% dos 216.044 inscritos, com abstenção de 9,3% (INEP, 2018c). Em relação aos cursos de medicina, em 2016 foram analisados 177 cursos e contou com 16180 inscritos, que foram aplicadas provas a 15.865 estudantes, com abstenção de 1,9% (INEP, 2018d). Os cursos de Medicina foram avaliados nos anos de 2004, 2007, 2010, 2013 e 2016.

Sendo assim, objetivou-se nesta Seção analisar o desempenho dos cursos de nível superior de medicina do Brasil, levando em consideração os resultados das avaliações do

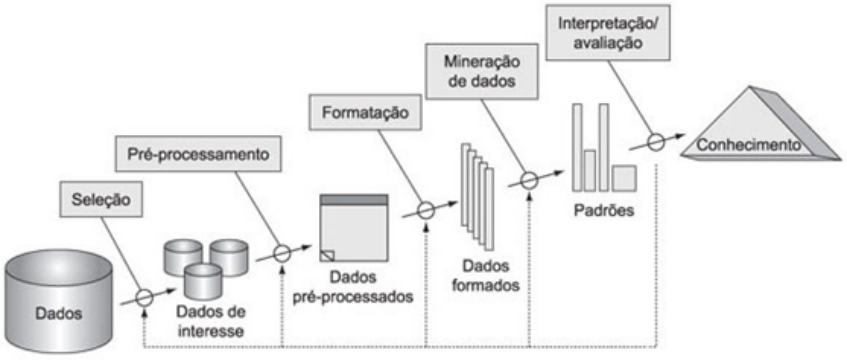
ENADE. Para alcançar este objetivo foi realizada uma análise estatística descritiva e aplicada Mineração de Dados (MD), com o intuito de descobrir quais categorias administrativas e regiões das Instituições de Ensino Superior (IES) possuem menores e maiores índices, e se houve uma redução de rendimento das instituições nos últimos anos.

1.1.1 Método de Análise

A metodologia desenvolvida nesta Seção foi dividida em duas etapas. A primeira etapa consistiu na realização da análise estatística descritiva das regiões e categorias administrativas das três últimas avaliações dos cursos de medicina. Por meio da análise estatística descritiva é possível organizar, sintetizar e descrever os dados (SANTOS, 2007).

Na segunda etapa, empregou o modelo *Knowledge Discovery in Databases* (KDD), apresentado por Fayyad, Piatetsky-Shapiro e Smyth (1996), vale ressaltar que esse modelo é bastante utilizado na ciência de dados. As fases do processo KDD estão representadas na Figura 1 e descritas em seguida. O processo do KDD envolve o uso do banco de dados, sendo que, os dados atravessam um tratamento programado e inteligente por intermédio de um encadeamento de atividades: seleção; pré-processamento; subamostragem e transformações necessárias; aplicação de métodos de MD para enumerar padrões; e avaliar os produtos de MD para identificar o subconjunto dos padrões enumerados considerados conhecimento.

Figura 1 – Modelo KDD



Fonte: Fayyad, Piatetsky-Shapiro e Smyth (1996)

Na fase de Seleção, as bases de dados usadas foram as do INEP sobre o ENADE[1], dos anos de 2010, 2013 e 2016. Os registros selecionados foram do curso de medicina e as variáveis selecionadas foram *região*, *categoria administrativa* e *conceito ENADE* da base de 2013 e *categoria administrativa* e *Conceito ENADE* das bases de 2010 e 2016.

Na fase de Pré-processamento, foram realizadas verificações de dados redundantes e remoção, caso houvesse, a fim de buscar eficiência do algoritmo de MD, e remoção de ruídos, se apropriado. Para resolver o problema de ausência de valores (*missing values*), há

[1] http://portal.inep.gov.br/conceito-enade/

basicamente três alternativas de solução, apresentadas por Prass (2004), que são técnica de imputação, substituição e exclusão dos registros. Salienta-se ainda que foi utilizada a última opção nos registros que não apresentavam conceito ENADE do curso. Também houve a inclusão das regiões em que os cursos estão localizados das bases dos anos de 2010 e 2016. Ao final dessa etapa os dados estavam pré-processados.

Na fase de Transformação, após os dados serem selecionados e pré-processados, foram armazenados e transformados em dados do tipo inteiro para que o algoritmo de MD fosse aplicado. Todos os dados foram agrupados e importados por meio da linguagem R[2].

Na fase de MD foram realizadas exploração e análise dos dados de forma automática, para descobrir padrões e regras, com intuito de fornecer informações relevantes. Nesta fase foi utilizado o método de *Clustering K-means* em cada uma das bases de dados selecionadas inicialmente (HAMERLY; ELKAN, 2004). O algoritmo foi aplicado para extração de padrões dos dados e para gerar regras que descrevam o comportamento da base de dados (BERRY; LINOFF, 1997).

Na fase de Interpretação/Avaliação dos dados minerados, foram apresentados modelos e padrões extraídos. Assim, é gerado o conhecimento que é documentado e relatado nesta Seção. Esta etapa foi realizada com auxílio de um especialista da área, como Prass (2004) afirma que deve ser feito. Os resultados e discussões são relatados na Subseção a seguir.

1.1.2 Resultados

Na análise de estatística é importante resumir e organizar os dados utilizando maneiras que facilitem sua interpretação e análise subsequente, por meio da representação visual das técnicas amostrais, permitindo investigar a simetria e estatística dos dados (GONÇALVES; VILELA; BEZERRA, 2018). Assim, nas Tabelas 1 e 2, são apresentados os cálculos totais de cursos com conceito ENADE, totais de cursos com conceito ENADE abaixo da média, porcentagem de cursos com conceito ENADE abaixo da média, média das Notas dos Concluintes (NC) e desvio padrão das NC.

Na Tabela 1, é apresentada a análise do desempenho das categorias administrativas dos cursos de medicina no ENADE 2010, 2013 e 2016. Em 2010, dos cursos de categoria pública avaliados, 2,94% possuíam conceito ENADE abaixo da média, com uma média da NC de 3,49 e desvio padrão de 0,8, os cursos de categoria privada, 32,89% apresentavam conceito ENADE abaixo da média, com média das NC de 2,40 e desvio padrão de 0,97.

Já em 2013, dos cursos públicos, 4,55% possuíam conceito ENADE abaixo da média, com média das NC de 3,32 e desvio padrão de 0,82, as públicas estaduais e federais obtiveram bons desempenhos com médias de 3,55 e 3,31 respectivamente. Dos cursos

[2] https://www.r-project.org/

Tabela 1 – Análise das Categorias Administrativas dos Cursos de Medicina no ENADE 2010, 2013 e 2016.

Ano		Pública				Privado						
		Pública Federal	Pública Estadual	Pública Municipal	Total Pública	Privada com fins lucrativos		Privada sem fins lucrativos				Total Privado
						Privada com fins lucrativos	Sociedade Civil	Privada sem fins lucrativos	Associação de utilidade Pública	Fundação	Sociedade	
Total	**2010**	68			**68**	76						**76**
	2013	43	18	5	**66**	30		70				**100**
	2016	47	23	5	**75**	26	6	59	1	8	1	**101**
Conceito Abaixo da Média	**2010**	2			**2**	25						**25**
	2013	1	1	1	**3**	17		19				**36**
	2016	1	3	2	**6**	9	2	19	0	1	0	**31**
Conceito Abaixo da Média (%)	**2010**	2,94			**2,94**	32,89						**32,89**
	2013	2,33	5,56	20	**4,55**	56,67		27,14				**36**
	2016	2,13	13,04	40	**8,00**	34,62	33,33	32,20	0	12,50	0	**30,69**
Média NC	**2010**	3,49			**3,49**	2,40						**2,4**
	2013	3,31	3,55	2,68	**3,32**	1,97		2,37				**2,25**
	2016	3,08	3,01	2,36	**3,01**	2,14	2,22	2,32	2,15	2,16	2,38	**2,26**
Desvio Padrão Média NC	**2010**	0,8			**0,80**	0,97						**0,97**
	2013	0,84	0,71	0,76	**0,82**	0,85		0,96				**0,95**
	2016	0,46	0,77	0,92	**0,62**	0,66	0,95	0,90	-	0,73	-	**0,82**

Fonte: Autoria Própria

privados, 36% apresentavam conceito ENADE abaixo da média, média das NC de 2,25 com desvio padrão de 0,95. O desempenho dos cursos de instituições privadas com fins lucrativos são piores, pois, 56,67% apresentavam conceito abaixo da média, média de NC de 1,97 e desvio padrão 0,85.

Na última avaliação dos cursos de Medicina no ENADE, os resultados foram relativamente satisfatórios em relação a 2013, mas apenas nos cursos privados. Os cursos públicos com conceito ENADE abaixo da média obtiveram uma baixa, e totalizaram 8%, a média das NC regrediu para 3,01 e o desvio padrão de 0,62. As públicas federais e estaduais regrediram, mas ainda obtiveram bons desempenhos com médias de 3,08 e 3,01 respectivamente. Já os cursos privados apresentaram uma melhora e houve diminuição dos cursos que estão com conceito ENADE abaixo da média para 30,69% e um aumento da média das NC para 2,26 com desvio padrão de 0,82. O desempenho dos cursos de instituições privadas com fins lucrativos ainda continuam baixos, mas obtiveram melhoras, pois, apresentam 34,62% com conceito abaixo da média e média de NC de 2,14 e desvio padrão 0,66.

Na Tabela 2, é apresentada a análise do desempenho das regiões dos cursos de medicina no ENADE 2010, 2013 e 2016. No âmbito geral houve um aumento dos cursos avaliados, o que pode ter ocorrido por conta do aumento dos cursos ao longo dos anos anteriores. Em 2010, 18,75% dos cursos estavam abaixo da média, dos 30 cursos do Nordeste apenas 1 estava abaixo da média, que correspondiam a 3,33% dessa região, seguida pela região Sul com 7,69%. Mais de 1/3 dos cursos da região Norte do país apresentavam desempenho inferior aos demais, e seguido pela região Sudeste com 28,57% dos cursos abaixo da média. Assim, o Nordeste e o Sul apresentavam os melhores resultados das NC

em 2010, com média de 3,37 e 3,30 e desvio padrão de 0,78 e 0,97 respectivamente.

Tabela 2 – Análise das Regiões dos Cursos de Medicina no ENADE 2010, 2013 e 2016.

Região	Total			Conceito Abaixo da Média			Conceito Abaixo da Média (%)			Média NC			Desvio Padrão NC		
	2010	2013	2016	2010	2013	2016	2010	2013	2016	2010	2013	2016	2010	2013	2016
CENTRO-OESTE	11	12	12	1	2	2	9,09	16,67	16,67	3,29	3,16	2,77	0,99	0,96	0,73
NORDESTE	30	34	39	1	8	7	3,33	23,53	17,95	3,37	2,77	2,69	0,78	0,87	0,68
NORTE	14	14	16	5	8	7	35,71	57,14	43,75	2,20	1,70	2,06	0,90	1,03	1,02
SUL	26	29	31	2	1	2	7,69	3,45	6,45	3,30	3,06	2,93	0,97	0,90	0,61
SUDESTE	63	77	78	18	20	19	28,57	25,97	24,36	2,63	2,59	2,46	1,07	1,06	0,88
GERAL	144	166	176	27	39	37	18,75	23,49	21,02	2,91	2,68	2,58	1,05	1,04	0,83

Fonte: Autoria Própria

Em 2013, 23,49% dos cursos estavam abaixo da média, a região Sul possuía apenas 1 curso abaixo da média, que correspondia a 3,45% dessa região, seguida pela região Centro-Oeste com 16,67%, o Nordeste e Sudeste apresentaram resultados em torno da média anual. O Norte teve um baixo desempenho, aumentando para 57,14% os seus cursos com conceito inferior. As regiões Centro-Oeste e Sul apresentaram melhores resultados das NC com médias de 3,16 e 3,06 e desvios padrões de 0,96 e 0,90 respectivamente.

Nas avaliações de 2016, os resultados foram relativamente satisfatórios em relação a 2013, pois 21,02% dos cursos estão abaixo da média, em 2013 era 23,49%. A região Sul possui 6,45% de seus cursos abaixo da média, o Norte continua com menor desempenho que corresponde a 43,75% dos seus cursos com conceito inferior, mas esses dados ainda são melhores que em relação a 2013. Entretanto, a média das NC continua regredindo, sendo 2,58 em 2016, o Sul e o Centro-Oeste apresentavam melhores resultados com médias de 3,16 e 3,06 e desvios padrões de 0,61 e 0,73, respectivamente. O que pode ter acontecido por conta das instituições privadas terem melhorado o ensino de acordo com as avaliações do ENADE, pois o Ministério da Educação (MEC) chegou a descredenciar algumas (MOTA *et al.*, 2014).

Tomando como base a metodologia KDD, após as fases de seleção, pré-processamento e transformação, os dados ficaram formatados, deixando-os úteis para análise e MD. A partir daí foram gerados três conjunto de dados (*dataset*). Na Figura 2, é apresentada a amostra para mineração da base de dados do ENADE de 2016.

O número de cursos das bases do ENADE de 2010, 2013 e 2016 foram 3.966, 3.519 e 4.300, respectivamente. Após as etapas de tratamento dos dados nos *datasets* de 2010, 2013 e 2016 restaram 144, 166 e 176 cursos de medicina respectivamente, com as colunas de categoria administrativa, região e conceito ENADE. Depois de se obter cada *dataset* foi aplicado o algoritmo *K-Means*, para identificar agrupamentos nos dados (HAMERLY; ELKAN, 2004). Segundo Li, Yang e Wang (2001), o *KMeans* possui a seguinte função objetivo, apresentada na Fórmula 1.1:

Figura 2 – Dados transformados, *dataset* ENADE 2016

	Categoria Administrativa	Região	Conceito Enade Faixa
1	1	4	2
2	1	1	4
3	1	4	3
4	1	2	3
5	1	3	1
6	1	2	3
7	2	5	3
8	3	4	3
9	3	5	3
10	3	5	3
11	3	4	3

Showing 1 to 11 of 176 entries

Fonte: Autoria Própria

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{1.1}$$

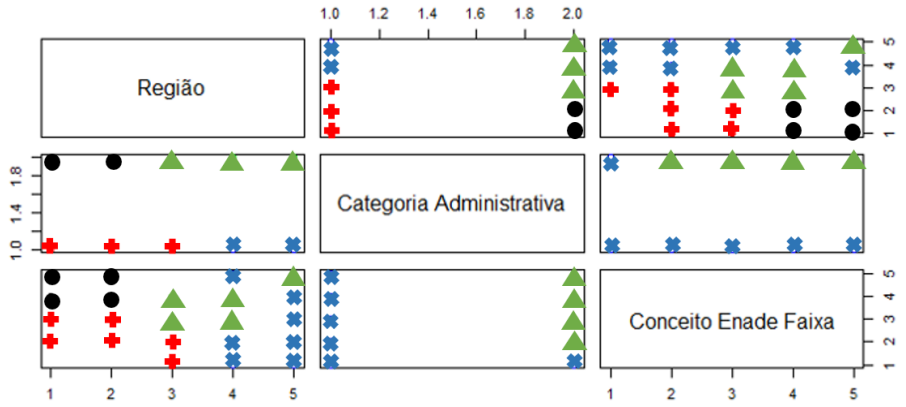
Na qual $\left\| x_i^{(j)} - c_j \right\|^2$ é uma medida da distância escolhida entre um ponto de dados, $x_i^{(j)}$ é o centro do *cluster* e c_j é um indicador da distância dos n pontos de dados de seus respectivos centros do *cluster*. O algoritmo apresenta o seguinte fluxo segundo JinHuaXu e HongLiu (2010):

1. Inicialmente são colocados K pontos no espaço representado pelos objetos que estão sendo agrupados. Esses pontos representam os centroides iniciais do grupo.
2. Cada objeto atribuído ao grupo que tem o centroide mais próximo.
3. Quando todos os objetos tiverem sido atribuídos, recalcule as posições dos centroides K.
4. Os passos 2 e 3 são repetidos até que os centroides não se movam mais.

Isso produz uma separação dos objetos em grupos dos quais a métrica a ser minimizada é calculada. Os resultados das três bases de dados são apresentados a seguir. Na Figura 3, é apresentado os *clusters* da base de dados de 2010 e a sua relação com cada atributo, na Figura 4 está a distribuição 2D dos *clusters* da base de dados de 2010. Foram gerados quatro *clusters*, com explicação de 85,11% dos dados.

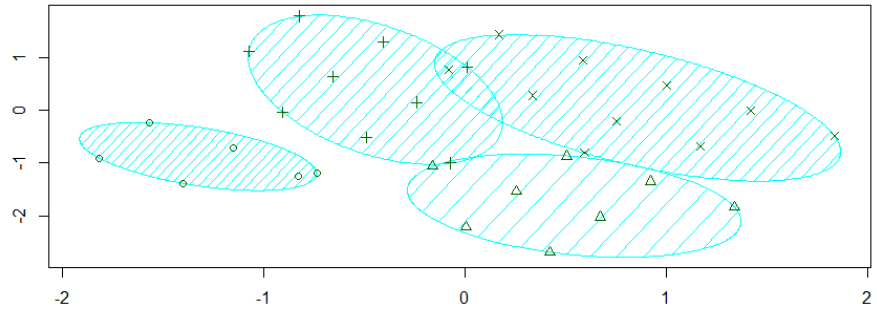
No *cluster* de formato "+" e cor Vermelha estão agrupados os cursos de instituições privadas (1.0), localizadas no Centro-Oeste (1), Nordeste (2) e Norte (3) com conceito ENADE 1, 2 e 3. O *cluster* no formato de "◯" e cor Preta agrupou os cursos de instituições

Figura 3 – Plotagem dos *Clusters*, *dataset* ENADE 2010



Fonte: Autoria Própria

Figura 4 – Distribuição dos *Clusters*, *dataset* ENADE 2010



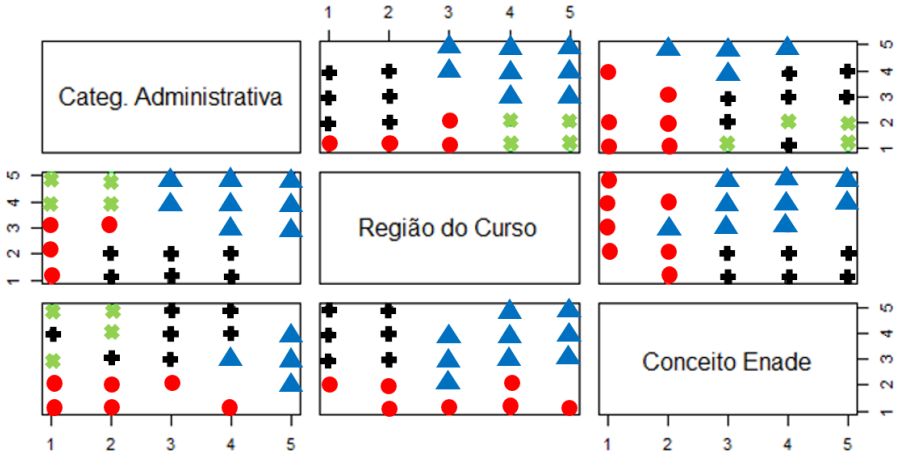
Fonte: Autoria Própria

públicas (2.0), localizadas no Centro-Oeste (1) e no Nordeste (2), de conceito ENADE 4 e 5, este *cluster* agrupou cursos com conceitos acima da média. O *cluster* no formato de "△" e cor Verde agrupou cursos de instituições públicas (2.0), localizadas no Norte (3), Sul (4) e Sudeste (5) com conceito ENADE 3, 4 e 5, este *cluster* agrupou cursos com conceito ENADE na média ou acima. O *cluster* de formato "×" e cor Azul agrupou os cursos de instituições privadas (1.0), localizadas no Sul (4) e Sudeste (5) com conceito ENADE 1, 2, 3, 4 e 5, mas em sua maioria 1 e 2.

Na Figura 5, são apresentados os *clusters* da base de dados de 2013 e a sua relação com cada atributo. O *cluster* no formato de "△" e cor Azul agrupou cursos de instituições Públicas Estaduais (3), Públicas Federais (4) e Municipais (5), mas de maioria municipal, localizadas no Norte (3), Sudeste (4) e Sul (5), com conceito ENADE 2, 3, 4 e 5, mas, em maioria, de bom desempenho de conceitos 3 e 4.

O *cluster* no formato de "◯" e cor Vermelha agrupou cursos de baixo desempenho de instituições Privada com fins lucrativos (1), Privada sem fins lucrativos (2), Pública Estadual (3), Pública Federal (4), mas de sua maioria Privada, localizadas no Centro-Oeste (1), Nordeste (2) e Norte (3), com conceito ENADE 1 e 2. O *cluster* no formato de "+"

Figura 5 – Plotagem dos *Clusters*, *dataset* ENADE 2013

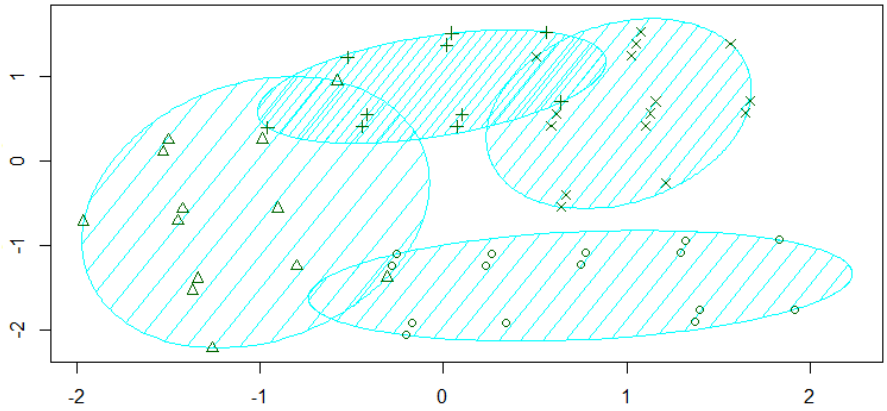


Fonte: Autoria Própria

e cor Preta agrupou cursos que obtiveram bom desempenho no ENADE de instituições Privada sem fins lucrativos (2), Pública Estadual (3) e Pública Federal (4), localizados no Centro-Oeste (1) e Nordeste (2), com conceito ENADE 3, 4 e 5. No *cluster* de formato "×" e cor Verde, estão agrupados poucos cursos, mas com bom desempenho de instituições privadas (1 e 2), localizados no Sudeste (4) e Sul (5) com conceito ENADE 3, 4 e 5.

A distribuição 2D dos *clusters* da base de dados de 2013, apresentada na Figura 6, obteve uma explicação de 80,83% dos dados, os *clusters* foram bem distribuídos, mas alguns pontos ainda apresentaram os mesmos domínios de alguns *clusters*.

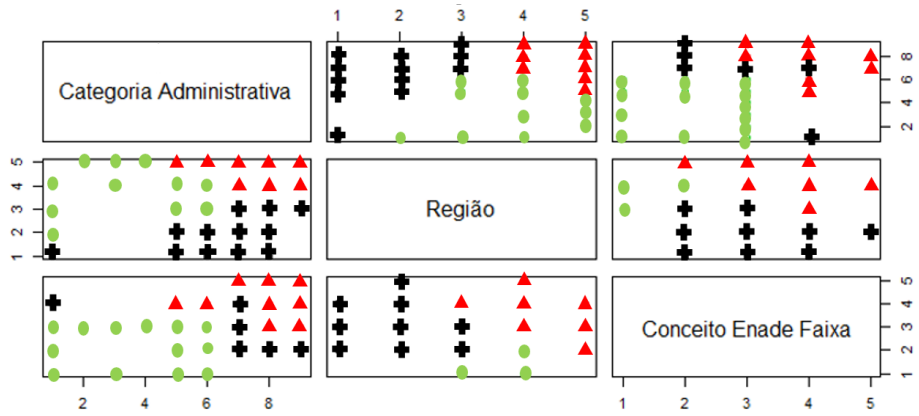
Figura 6 – Distribuição dos *Clusters*, *dataset* ENADE 2013



Fonte: Autoria Própria

Na última avaliação dos cursos de medicina do ENADE, apresentado na Figura 7, pode ser verificado a relação dos *clusters* com cada atributo. O *cluster* no formato de "△" e cor Vermelha agrupou os cursos das instituições Privada com fins lucrativos (5), Privada sem fins lucrativos (6), Pública Estadual (7) e Pública Federal (8) e Pública Municipal (9), sendo sua maioria Pública Federal e Municipal, localizados no Sudeste (4) e Sul (5) com conceito ENADE de 2, 3, 4 e 5, mas de maioria conceito 3 e 4.

Figura 7 – Plotagem dos *Clusters*, *dataset* ENADE 2016

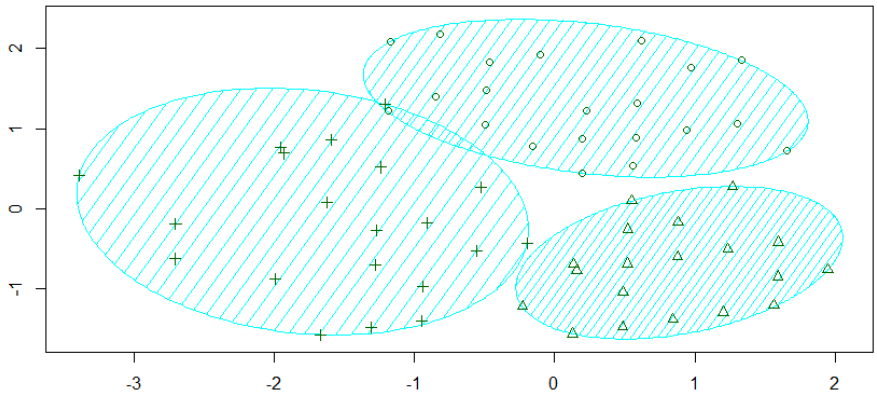


Fonte: Autoria Própria

No *cluster* de formato "+" e cor Preta, foram agrupados os cursos das instituições Públicas (7, 8 e 9), localizadas no Centro-Oeste (1), Nordeste (2) e Norte (3) com conceito ENADE 2, 3, 4 e 5, sendo a maior parte conceito 2 e 3. O *cluster* no formato de "◯" e cor Verde agrupou cursos das instituições Privadas (1, 2, 3, 4, 5, 6), localizados em sua maioria nas regiões Norte (3), Sudeste (4) e Sul (5) e com conceito ENADE 1 e 2, sendo a maioria conceito 1.

Inicialmente foram definidos quatro *clusters*, entretanto vários pontos pertenciam a mais de um domínio, com o intuito de uma melhor compreensão, esses dados foram agrupados em três *clusters*. Na Figura 8, é apresentada a distribuição 2D dos *clusters* da base de dados de 2016, obteve uma explicação de 78,78% dos dados, os *clusters* foram bem distribuídos, poucos pontos pertencem a mais de um domínio.

Figura 8 – Distribuição dos *Clusters*, *dataset* ENADE 2016



Fonte: Autoria Própria

Portanto, os *clusters* que agrupavam as instituições públicas tiveram uma baixa no rendimento ao final dessas três avaliações, entretanto as instituições públicas ainda apresentam uma ampla vantagem sobre as privadas. Novamente, isso pode ser justificado por conta do MEC ter descredenciado quatro instituições privadas em 2012, pois algumas

delas não possuíam prática para aulas (MOTA *et al.*, 2014), acarretando que outras instituições privadas melhorassem o ensino para não serem descredenciadas. Quando agrupados cursos da região Norte com instituições privadas, a possibilidade de se obter um conceito abaixo da média 3 foram maiores. Quando agrupados Sul ou Centro-Oeste com instituições públicas, as chances de se obter conceito ENADE na média ou acima foram maiores. Na Seção a seguir são apresentadas as discussões e soluções para o problema apresentado nesta Subseção.

1.1.3 Discussões

A análise da formação de médicos é importante pela relevância desse profissional para toda sociedade e o sistema de saúde. Assim, os ganhos da MD nesse contexto propõem iniciativas de incentivo às mudanças nessas instituições com baixo desempenho, para que busquem intervenções, a fim de melhorar o ensino-aprendizagem. Alguns autores trazem trabalhos que podem suprir essa carência, centralizados diretamente aos cursos de medicina, tais como o reforço das diretrizes nacionais dos cursos de medicina (Ministério da Educação, 2019), na qual cada curso está submetido, melhorando o ensino das instituições; métodos ativos de ensino, entre eles o *Problem Based Learning* (Aprendizagem baseada em problemas) e o *Team-Based Learning* (Aprendizagem Baseada em Equipes), em que o aluno fica autônomo e busca soluções para as questões do dia a dia, ou/e trabalho em grupo contínuo, que já estão sendo utilizados por algumas instituições de medicina (FAJARDO, 2019).

Outra forma de incentivar às mudanças no ensino nas escolas médicas, com baixo desempenho é por meio da aplicação do Projeto de Incentivo a Mudanças Curriculares para os Cursos de Medicina (PROMED) no qual oferece cooperação técnica e/ou operacional às escolas de graduação em Medicina, programa de estágios aos alunos e capacita o aluno com base na realidade e funcionamento do Sistema Único de Saúde (SUS). No trabalho de Oliveira *et al.* (2008), pôde ser percebido que a aplicação do PROMED nas escolas médicas encontrou-se em processo lento e abaixo das expectativas da saúde e população.

Nesta realidade, em que estão inseridos os estudantes e na qual vão atuar, outras avaliações e pesquisas apontam baixo conhecimento pelos estudantes de modo geral, como os relatórios da prova do Conselho Regional de Medicina do Estado de São Paulo (CREMESP), que indicam que os estudantes de medicina têm obtido baixo desempenho no exame do CREMESP, o qual avalia a formação médica de estudantes e egressos do 6º ano do curso. São aferidos os conteúdos básicos de áreas consideradas essenciais na Medicina, como Clínica Médica, Clínica Cirúrgica, Pediatria, Ginecologia e Obstetrícia, Saúde Pública, Epidemiologia, Saúde Mental, Bioética e Ciências Básicas. Em 2011, dentre os 418 participantes do exame, 191, que corresponde a 46%, foram reprovados.

Os estudantes apresentaram baixo rendimento em saúde pública com 49,0% de

acertos, obstetrícia com 54,1% de acertos, clínica médica com 56,5% de acertos e pediatria com 59,3% de acertos, especialidades que concentram a solução de muitos problemas de saúde da população (CFM, 2018).

Mais da metade dos estudantes de medicina, em 2016 foram reprovados nesse exame. Dos 2.677 participantes, 56,4% foram reprovados. Os estudantes não acertaram 60% da prova (CREMESP, 2019c). Em 2017, foi constatado que, dos 2.636 participantes, 78% erraram o diagnóstico laboratorial de diabetes mellitus e 60% demonstraram pouco conhecimento sobre doenças parasitárias, formas de veiculação e contaminação (CREMESP, 2019b). Estes dados são alarmantes, visto que os alunos, após o curso, lidarão com vidas humanas, havendo risco para a própria população. Enquanto isso, em 2018, os resultados foram melhores, mas ainda preocupam, pois dos 3.174 participantes, 38,2% foram reprovados no exame.

Os resultados continuam insatisfatórios, em clínica médica e ciências básicas, muitos dos recém-formados não sabem interpretar exames para diagnosticar e administrar a conduta terapêutica adequada em casos médicos básicos e problemas de saúde frequentes, tais como: 69% não souberam as diretrizes para aferição da pressão arterial; 68% não acertaram a conduta para paciente com infarto no miocárdio; 65% erraram o quadro laboratorial do diabetes mellitus descompensado; 59% não informaram corretamente o período de transmissão da gripe; e 44% não souberam identificar o agente causador e um dos principais transmissores da doença de chagas (CREMESP, 2019a).

Neste contexto, surgem várias estratégias que também podem auxiliar no ensino-aprendizagem, tais como o uso de dinâmicas de grupo, salas de aula invertidas, aprendizagem flexível, jogos, dentre outras (GOODE *et al.*, 2007). Algumas dessas técnicas também podem ser aplicados aos cursos de medicina, como no caso dos jogos sérios, que vem ganhando grande força no ensino-aprendizagem por possibilitar a reprodução de situações do mundo real em ambientes criados em computador (BOGONI; PINHO, 2014). Podendo ser usados como métodos de aprendizado ativo para incentivar alunos a revisar materiais (CASTRO; SIQUEIRA, 2017), pois este tipo de aprendizagem possui boa aceitação por parte dos professores e alunos de medicina.

Um exemplo é o trabalho de Vilagra (2012) que teve por objetivo apresentar uma revisão acerca das transformações que o ensino médico vem sofrendo nos últimos anos. A autora apresentou um manual de estratégias de ensino para ser utilizado pelos docentes de medicina. Foi percebido que os efeitos da inserção da prática médica precoce têm sido positivo para os alunos de medicina. Essa inserção pode ser dada pelo uso de estudo de caso e problematização, aula prática a beira do leito ou oficinas de prática. Portanto, essas estratégias apresentadas podem melhorar no ensino-aprendizagem desses cursos de medicina.

1.1.4 Conclusão

Apresentou-se nesta Subseção, uma análise do desempenho dos cursos de medicina do Brasil no ENADE 2010, 2013 e 2016. Foram analisados os desempenhos a níveis de regiões e categorias administrativas por meio de análise estatística usando média, desvio padrão, somatório e porcentagem. Além disso, foi aplicado MD nas bases de dados por meio de região, categoria administrativa e conceito ENADE. Foram apresentadas possíveis soluções encontradas na literatura sobre o problema abordado.

Com base na análise dos resultados encontrados foi possível afirmar que desde 2010 houve um crescimento de cursos de medicina abaixo na média das avaliações do ENADE, que as instituições públicas tiveram, no geral, avaliações melhores que as instituições privadas e principalmente as instituições Federais e Estaduais. Ainda se identificou, que a região Norte apresenta os cursos de menor desempenho e as regiões Sul, Centro-Oeste e Nordeste apresenta, na maioria, cursos de bom desempenho. Quando combinados cursos das regiões Sul e Centro-Oeste com instituições públicas federais e estaduais, os resultados são em quase na totalidade excelentes. Entretanto, quando combinado cursos da região Norte com instituições privadas (com fins lucrativos) quase sempre são abaixo da média.

Como descrito nas discussões desta Seção, o uso de dinâmicas de grupo e, principalmente métodos de aprendizado ativo, podem melhorar no ensino-aprendizado de estudantes de medicina. Esses resultados não indicam que os cursos das regiões e/ou categorias administrativas com conceito 1 ou 2 apresentam rendimento ruim, mas sim que um curso com conceito 3 agrega mais valor/conhecimento do que um curso com conceito 2 (G1, 2018).

O estudo descrito nesta Seção limita-se ao fato de não terem sido analisados todos os cursos de medicina do Brasil, pois uma parte não apresenta conceito ENADE e por isso não foram incluídos neste estudo; além do número limitado de variáveis para MD. Assim, no futuro esses cursos de medicina serão analisados mais a fundo, realizando análise de outros dados presentes nas bases de dados do INEP, como as notas dos estudantes.

1.2 MOTIVAÇÃO

Para que os médicos atuem em sua área é necessário inscrição no Conselho Regional de Medicina adquirida por meio de exames, ficando presumível o conhecimento da ciência médica pelo profissional. Além disso, desde abril de 2005, estão sujeitos à revalidação periódica do título, segundo a Resolução n. 1.755, de 12 de novembro de 2004, do Conselho Federal de Medicina. Tendo o mesmo que aprimorar continuamente seus conhecimentos e usar o melhor do progresso científico em benefício do paciente (CARVALHO *et al.*, 2012). Entretanto, mesmo com essas medidas ocorrem esses erros médicos pela falta de conhecimento prévio, que vem desde os estudos iniciais na graduação.

A má formação dos estudantes de medicina acarreta em erros médicos que afetam toda a sociedade. O erro pela deficiência de conhecimento técnico profissional é denominado de imperícia. Esses erros têm sido uma das grandes causas de mortes no mundo. Nos Estados Unidos, o departamento de Saúde e Serviços Humanos do escritório do Inspetor-Geral, examinando os registros de saúde de pacientes hospitalizados em 2008, relatou 180.000 mortes devido a erros médicos por ano, apenas entre os beneficiários do *Medicare* (MAKARY; DANIEL, 2016).

James (2013) realizou uma revisão da literatura que estimou danos evitáveis usando uma análise ponderada e descreveu uma faixa de incidência de 210.000 a 400.000 mortes por ano associada a erros médicos em pacientes hospitalares nos Estados Unidos, tomando como base estudos publicados de 2008 a 2011.

Balogh *et al.* (2015) identificaram que 10% dos casos de pacientes internados que evoluíram para óbito submetidos a estudos de necropsia nos Estados Unidos tiveram pelo menos um erro de diagnóstico. Dados de países membros da União Europeia, indicam consistentemente que falhas e eventos adversos relacionados à assistência à saúde ocorrem em 8% a 12% das hospitalizações. Tendo 23% dos cidadãos da União Europeia afirmado que foram diretamente afetados por alguma falha assistencial (Organização Mundial da Saúde, 2020).

No Brasil, o estudo de Mendes *et al.* (2009) verificou uma amostra randômica de 1.103 adultos de uma população de 27.350 internados em 2003, e identificou a incidência de eventos adversos de 7,6%, sendo 66,7% deles preveníveis. Em 2017, 54.076 pessoas morreram em decorrência da má conduta profissional. A cada hora, seis pessoas sofrem de erro no diagnóstico ou negligência médica no Brasil, havendo os erros clínicos uma relação com isso (COUTO *et al.*, 2018). A pratica clínica é cheia de vícios, condutas e conceitos errados, e determinadas condutas podem causar danos ao paciente, além de configurar má prática clínica (LOPES, 2020).

Diante desse quadro, surgiu a ideia de continuar o desenvolvimento de um jogo sério denominado DocTraining, apresentado inicialmente por Lima (2016), para auxiliar no processo de ensino-aprendizagem de estudantes de medicina, possibilitando o treinamento em casos clínicos, diabetes e doença cardíaca, na qual os alunos possuem deficiência de ensino, e, assim, mitigar os erros médicos. Entretanto, alguns problemas são enfrentados para desenvolver um jogo sério para alunos de medicina, esses problemas são descritos na próxima Seção.

1.3 PROBLEMA DE PESQUISA

Visto o cenário atual exposto na Seção anterior, e na Subseção 1.1.3, percebe-se a lacuna na formação dos estudantes de medicina, que pode ser suprida por meio

de simuladores de realidade virtual para medicina. Pois, seu uso pode possibilitar a reprodução de situações do mundo real em ambientes criados em computador, a fim de oferecer sensações e experiências presentes no mundo real (BOGONI; PINHO, 2014).

Uma das vantagens da simulação virtual é que o treinamento pode ser realizado com uma duração fixa e um determinado número de casos, como também pode ser controlada por níveis de competência do usuário (SATAVA; GALLAGHER; PELLEGRINI, 2003). Os simuladores para área da medicina apresentam a inexistência de riscos aos pacientes e estudantes, além de aumentar a precisão do aluno, e possibilitar repetição de treinamento e a transferência de experiência (MARAN; GLAVIN, 2003).

Entretanto, desenvolver um simulador ou um jogo para medicina requer cuidados, visto que os alunos irão lidar com vidas humanas, não podendo haver erros. Os dados apresentados para os estudantes em um simulador precisam ser verdadeiros e de confiança, assim, é necessário que os dados sejam analisados por especialistas.

Cada doença apresenta vários sintomas, entretanto algumas apresentam sintomas e taxas de exames semelhantes entre si. Portanto, processar grande quantidade de dados contido em base de amostras de doenças se torna um problema devido à complexidade e necessidade da análise dos dados pelo especialista da área. Com o passar do tempo novas doenças vão surgindo e sintomas podem variar. Sendo necessário atualizar os dados para que os estudantes tenham acesso ao que é mais recente. Assim, para simular um especialista que classifique os dados é possível utilizar técnicas de inteligência artificial, como o Aprendizado de Máquina (AM), que pode ser usado em jogos para medicina (SILVA *et al.*, 2019).

Além disso, tem-se a necessidade de possibilitar que profissionais da área realizem a inserção de novos dados de forma simples, fácil e de confiança, para que o simulador seja incrementado constantemente. Portanto, é necessário um ambiente on-line que possibilite o gerenciamento dos dados. Na próxima Seção estão os objetivos com base nos problemas de pesquisa aqui abordados.

1.4 OBJETIVOS

Considerando a problemática mencionada na Seção anterior, o presente trabalho propõe um módulo inteligente para auxiliar no gerenciamento de um jogo sério, realizando a classificação de dados médicos por meio de modelos de aprendizado de máquina e os disponibilizando a um jogo sério. Além disso, também é necessário que esses dados sejam facilmente administrados pelos profissionais da medicina que utilizarem o módulo inteligente.

Esse módulo inteligente faz parte do DocTraining, um projeto que conta com um

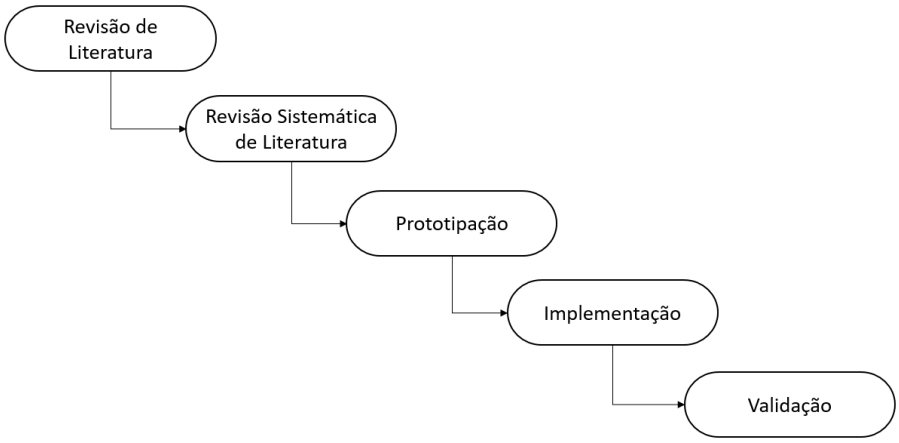
jogo sério para treinamento dos estudantes de medicina. Portanto, para o alcance do objetivo geral foram definidos os seguintes objetivos específicos:

- Identificar o estado da arte relacionado ao Aprendizado de Máquina nos Jogos para Medicina;
- Propor uma interface para utilização do módulo inteligente pelos professores/especialistas, na qual seja possível expandir os dados e gerenciar o jogo sério;
- Analisar a usabilidade e utilidade do módulo inteligente;
- Combinar múltiplos classificadores independentes na classificação de dados médicos, para gerar modelos de classificação que podem ser retreinados e atualizar os dados;
- Analisar a eficiência de classificação dos dados no módulo inteligente.

1.5 METODOLOGIA

Para propor esta nova abordagem, a metodologia de pesquisa utilizada neste trabalho seguiu um conjunto de fases, sendo dividida em: Revisão de Literatura; Revisão Sistemática de Literatura; Prototipação; Implementação; e Validação. Na Figura 9, pode-se observar o passo a passo da metodologia, que é descrita em seguida. Outros detalhes da metodologia também são apresentados ao longo do texto.

Figura 9 – Metodologia de Pesquisa.



Fonte: Autoria Própria

A Revisão de Literatura proporcionou a extração dos conceitos necessários para o desenvolvimento do módulo inteligente baseado em AM, além de identificar os principais trabalhos relacionados à essa temática.

A Revisão Sistemática de Literatura (RSL) tem o objetivo de caracterizar o estado da arte em uma determinada área do conhecimento, e identifica lacunas e oportunidades

de pesquisa. Por meio da RSL é possível realizar uma síntese do conhecimento existente sobre determinado assunto (BIOLCHINI *et al.*, 2007). Neste trabalho, a RSL determinou o estado da arte do aprendizado de máquina nos jogos para medicina, identificando as principais técnicas, necessidades e limitações apresentadas na literatura.

A Prototipação, segundo Pressman e Maxim (2016), tem o objetivo de identificar os requisitos do *software*. Sommerville (2007) define a prototipação como um processo que tem como propósito facilitar o entendimento dos requisitos e apresentar conceitos e funcionalidades do *software*. Nesta fase é apresentada com o protótipo, a arquitetura para gerar cada Múltiplo Classificador (MLC). Esta fase é responsável pela seleção de classificadores, na qual serão apresentados diversos MLCs.

Na Implementação do módulo inteligente, que é utilizado pelos professores e profissionais da medicina, foi utilizado o *framework* Django[3], para desenvolvimento *Web* de alto nível, com *design* limpo e pragmático, extremamente escalável. Foi utilizada a linguagem de programação *Python*[4], pois é uma das linguagens mais populares para computação científica. Por ser de natureza interativa de alto nível e pelo grande número de bibliotecas científicas, é uma boa opção para a análise exploratória de dados (PEDREGOSA *et al.*, 2011). Neste módulo inteligente também está o AM, que contêm os modelos dos classificadores que rotulam os dados médicos. A implementação ainda conta com o pré-processamento dos dados antes da aplicação das técnicas de AM.

A Validação foi feita por meio de duas etapas, sendo a primeira dos modelos de AM; e a segunda foi uma pré-validação com os especialistas que utilizaram o módulo inteligente.

- A etapa de validação dos modelos de AM foi subdividida em duas fases, na qual foram empregadas técnicas de Divisão de Porcentagem[5] (DP) e Validação Cruzada (VC) dos classificadores de AM e dos MLCs.

  Na primeira fase, a DP se baseou nas taxas de acertos e erros, na qual foi utilizada uma DP de 80% dos dados para treino e 20% para teste. E a VC empregou técnicas de *K-fold*, em que foram gerados cinco grupos com todas as instâncias do arquivo, sendo quatro para treinamento e uma para teste, alternando-os sucessivamente. Com isso, foram gerados cinco classificadores diferentes, em que o resultado final é a média dos cinco classificadores. Portanto, é visível a taxa real de acerto do classificador, dificultando que haja *overfitting*, na qual o modelo de AM se torna inútil com novos dados para classificação (SCIKIT-LEARN, 2019). A métrica de desempenho utilizada na DP e na VC foi a precisão.

[3] https://www.djangoproject.com/
[4] https://www.python.org/
[5] O termo em inglês para esse método é *percentage split*.

Na segunda fase, foi utilizada uma DP de 60% dos dados para treino e 40% para teste, de acordo com a pratica padrão (CATON *et al.*, 2018). A VC também empregou técnicas de *K-fold*, foram gerados oito grupos com todas as instâncias do arquivo, sendo sete para treinamento e uma para teste, sendo alternados sucessivamente. Assim, o resultado final foi a média dos oito classificadores. A métrica de desempenho utilizada nesta fase foi a *F-measure*, havendo variação para *Weighted F-measure* quando tratada de classificação multiclasse, como nos casos clínicos.

- Na etapa de pré-validação com os especialistas foram levados em conta as opiniões dos professores e profissionais da área em relação ao módulo inteligente e a confiança nos dados fornecidos para jogo. Assim, eles foram submetidos a uma entrevista semiestruturada, sendo possível coletar dados para validação do projeto.

1.6 ESTRUTURA DO DOCUMENTO

O presente trabalho está organizado conforme a seguinte estrutura: primeiro, no Capítulo 2, são apresentados conceitos sobre o aprendizado de máquina, e é apresentado o DocTraining; no Capítulo 3, é descrita uma Revisão Sistemática da Literatura que retrata o estado da arte referente ao uso do aprendizado de máquina nos jogos para medicina; no Capítulo 4, apresenta-se as bases de dados deste estudo, o pré-processamento aplicado, e a arquitetura de geração dos múltiplos classificadores; o Módulo Inteligente do DocTraining é apresentado no Capítulo 5; em seguida, no Capítulo 6, são apresentadas a validação computacional, a pré-validação realizada com os profissionais da medicina, e as discussões; e, por fim, no Capítulo 7, são apresentados as conclusões, contribuições científicas, limitações e trabalhos futuros desta pesquisa.

2 REFERENCIAL TEÓRICO

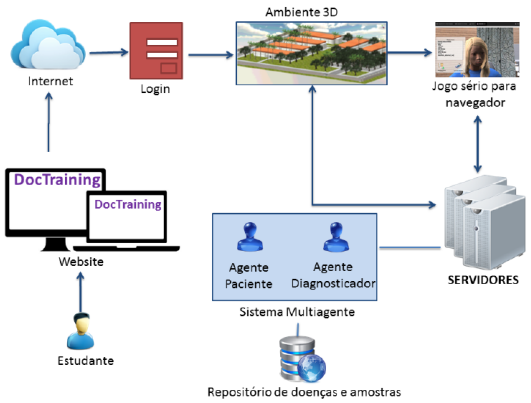
Neste Capítulo, são apresentados os conceitos e terminologias básicas ao entendimento deste trabalho e para a execução do desenvolvimento do módulo inteligente para auxiliar na simulação de casos clínicos, e na identificação de diabetes e doença cardíaca em pacientes. Primeiramente é apresentando o DocTraining. Na sequência é apresentado o Aprendizado de Máquina, sua definição, termos, hierarquia, classificação, processo de classificação, alguns classificadores. Por fim, é apresentada a combinação de classificadores, na qual são descritas as formas de se combinar classificadores.

2.1 DOCTRAINING

O DocTraining foi um ambiente 3D com o jogo sério para o treinamento de estudantes de medicina em casos clínicos, apresentado por Lima (2016). Foi desenvolvido com intuito de mitigar os erros médicos e apoiar o processo de aprendizagem dos estudantes.

A ferramenta de apoio é dividida em: ambiente 3D; sistema de gerenciamento de dados de usuário; jogo sério; e, CSDA (Criador de Sintomas, Doenças e Amostras). A simulação de casos clínicos foi feita por meio de ambiente 3D, aplicativo móvel usando sintetizador de voz e imersão através de óculos de realidade virtual. O sistema ainda conta com AM e Sistema multiagente para classificação de doenças. A visão geral para computadores de mesa do DocTraining é apresentada na Figura 10, o usuário entra no site e tem a opção de entrar no ambiente 3D.

Figura 10 – Visão geral para computadores de mesa e laptops.



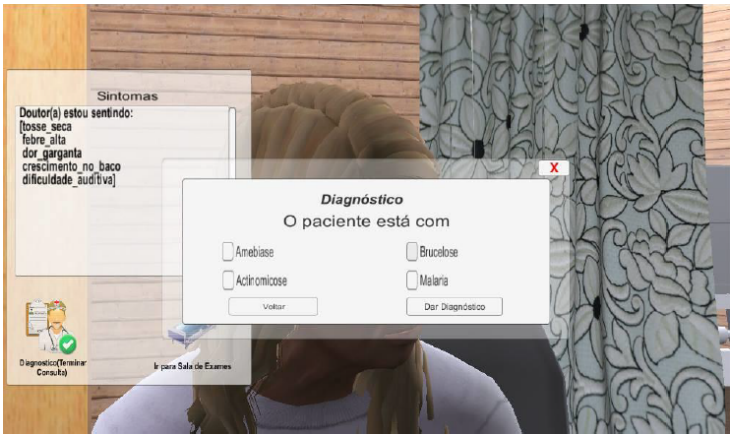
Fonte: Lima (2016)

Dentro do ambiente, o estudante de medicina pode entrar no hospital universitário e começar a consultar pacientes por meio do jogo sério. O jogo se comunica com o servidores,

onde o AM junto ao Sistema Multiagente fazem as classificações dos dados, a partir dos repositórios de doenças e amostras.

O DocTraining ficou complexo e algumas funcionalidades não ficaram bem elaboradas, como a complexidade das consultas e o AM. Além dos elementos de gamificação, qualidade gráfica do jogo, variedades de conteúdo e o principal, estrutura do código desenvolvido. A sua versão não é mais compatível com a versão atual da *Unity*, na qual o *software* foi desenvolvido. Dessa maneira, o sistema encontra-se obsoleto em relação a tecnologia atual, e fora de funcionamento. Na Figura 11, pode-se observar uma consulta e diagnóstico de um paciente pelo usuário/estudante.

Figura 11 – Tela de seleção de diagnóstico.



Fonte: Lima (2016)

O CSDA tomava como base valores *booleanos*, ou seja, valores de Verdadeiro ou Falso, que eram utilizados pelo jogo nos sintomas informados pelo paciente. Isso também limitou o jogo, pois os pacientes já informavam sobre dados que deveriam ser descobertos com exames. Na Figura 12, é apresentado o CSDA, na qual os professores eram capazes de gerenciar os dados do AM, entretanto esse acesso era local, o que dificultava esse incremento.

Figura 12 – Componente criado para uso do especialista de saúde.

Fonte: Lima (2016)

Lima (2016) e SOUSA NETO *et al.* (2018), definiram como trabalhos futuros, utilizar comitês de classificadores para verificação de melhoria na acurácia do cálculo de diagnóstico de amostras não rotuladas. Combinando múltiplos classificadores independentes para verificar melhorias significativas de veracidade das respostas encontradas, tem-se o objetivo de sempre rotular amostras com o melhor grau de certeza, pois as amostras de Diabetes classificadas por Lima (2016), conseguiram 75,39% como o melhor resultado usando VC com 10-*folds* por meio da métrica de precisão. Além da inserção de classificadores multirrótulo para classificação de doenças que compartilham sintomas exatamente iguais.

2.2 APRENDIZADO DE MÁQUINA

Anos atrás a inteligência artificial era vista como uma área teórica, com poucas aplicações e com pouco valor prático. Com o crescimento dos dados nas últimas décadas, houve a necessidade de ferramentas computacionais mais sofisticadas e autônomas para redução de intervenção humana e dependência de especialistas. Assim, por meio de algumas técnicas, seriam capazes de criar por si próprias, a partir da experiência passada, uma hipótese, ou função, capaz de resolver o problema que se deseja tratar. Este processo de indução de uma hipótese, por meio de experiência passada, dá-se o nome AM (LUGER, 2014).

2.2.1 Definição

O AM é importante para aplicações práticas de inteligência artificial por conta da indução (LUGER, 2014). A indução é a forma de inferência lógica que permite que conclusões gerais sejam obtidas de exemplos ou observações particulares. Assim, o AM é programado para aprender com experiência passada, usando o princípio da inferência chamado de indução, onde se obtém conclusões genéricas a partir de um conjunto de exemplos (FACELI *et al.*, 2011).

Monard e Prati (2005), definem AM como uma subárea de pesquisa muito importante da Inteligência Artificial, com uma importante característica essencial para o comportamento de inteligência, que é a capacidade de aprender. Segundo Rossi (2016), os objetivos dos algoritmos de AM é aprender, generalizar, ou ainda extrair padrões ou características das classes com base nos rótulos informados pelo especialista.

Assim, o uso de algoritmos de aprendizado de máquina exige o menor esforço humano, pois os modelos de classificação são mais fáceis de serem construídos (BENBRAHIM; BRAMER, 2009). Para Faceli *et al.* (2011), existem várias aplicações bem-sucedidas de técnicas de AM na solução de problemas reais, tais como:

- Reconhecimento de palavras faladas;
- Predição de taxas de cura de pacientes com diferentes doenças;
- Detecção do uso fraudulento de cartões de crédito;
- Condução de automóveis de forma autônoma em rodovias;
- Ferramentas que jogam gamão e xadrez de forma semelhante a campeões;
- Diagnóstico de câncer por meio da análise de dados de expressão gênica.

2.2.2 Conceitos Básicos do Aprendizado de Máquina

São encontradas na literatura algumas definições sobre conceitos de aprendizado de máquina, nesta Subseção são apresentados alguns conceitos, tais como: Indutor; Exemplo; Rótulo; Atributo; Conjunto de Treinamento; Conjunto de Testes; Precisão Média; *F-measure*; e, Validação Cruzada. A seguir, são apresentadas tais definições.

- **Indutor**: Segundo Monard e Baranauskas (2003), o objetivo do indutor (ou programa ou algoritmo) consiste em extrair um bom classificador a partir de um conjunto de exemplos rotulados. O indutor pode ser usada para classificar novos dados (ainda não rotulados) com a meta de predizer corretamente. Assim, o classificador pode ser avaliado considerando sua precisão, *recall* (revocação), *F1-score* (também conhecido por *F-score* ou *F-measure)* e velocidade de aprendizado.

- **Exemplo**: Um exemplo, também conhecido como registro ou dado, é uma tupla (ou também um vetor ou linha da tabela) de valores de atributos. Um exemplo descreve o objeto de interesse ou central, tal como um paciente, dados médicos sobre uma determinada doença, exames ou histórico de clientes (MONARD; BARANAUSKAS, 2003).

- **Rótulo ou Classe**: Rótulo também chamado de atributo-alvo ou classe, descreve o conceito de algo que se deseja aprender para tornar viável a realização de predições a seu respeito (FACELI *et al.*, 2011). No uso de aprendizado supervisionado, todo exemplo possui pelo menos um atributo denominado rótulo, mas podem haver mais (SILVA, 2017). Quando na classificação são possíveis apenas dois resultados, como sim ou não, é chamada de classificação binária, mas caso esse número seja variado é chamada de classificação multiclasse.

- **Atributo**: Os atributos também chamados campos ou variáveis, são utilizados para descrever características ou aspectos de um exemplo. Geralmente, os atributos se dividem em dois modelos: discretos ou contínuos. Atributos discretos contêm um número limitado ou finito de valores. Um tipo especial de atributo discreto é o atributo binário, também chamado de booleano, que possui apenas dois valores de saída, como: 0 ou 1; sim ou não; ausência ou presença; verdadeiro ou falso; ou outros dois valores. Os atributos contínuos, por sua vez, são representados por números infinitos de valores. Esses atributos contínuos são resultados de medidas e representados por números reais ou inteiros, como peso, tamanho e distância (FACELI *et al.*, 2011; SILVA, 2017).

- **Conjunto de Treinamento**: O conjunto de treinamento é utilizado como entrada pelos algoritmos de aprendizado para construção do classificador. Assim, esse conjunto deve ser representativo na distribuição da população dos dados do domínio, contendo todos os exemplos das classes (SILVA, 2017).

- **Conjunto de Teste**: O conjunto de teste tem a função de avaliar o modelo construído. Esse conjunto não deve ser apresentado ao algoritmo de aprendizado durante a elaboração o treinamento. O ideal é que o conjunto de testes não tenha exemplos em comum com os exemplos do conjunto de treinamento, pois o algoritmo deve ser capaz de classificar exemplos que ainda não conhece (SILVA, 2017).

- **Precisão Média**: A precisão média se resume em uma curva de precisão de recuperação como a média ponderada de precisões alcançadas em cada limite, com o aumento na recuperação do limite anterior usado como peso, de acordo com a Fórmula 2.1, apresentada a seguir.

$$precisão = \sum_{X}^{n} \left(R_n - R_{n-1}\right) P_n \tag{2.1}$$

Na qual P_n e R_n são precisão e a recuperação no enésimo limiar (SCIKIT-LEARN, 2019). Caso essa métrica calculada pelos pesos, são calculadas métricas para cada rótulo e encontra a média ponderada pelo número de instâncias verdadeiras para cada rótulo.

- ***F-measure* ou medida-F:** Apesar de precisão ser uma métrica mais popular, que descreve o percentual de acertos do modelo em comparação com o total de previsões realizadas, ela não fornece detalhes levando em considerações os acertos pelo número de classes. Assim, o *F-measure* é uma média harmônica entre precisão e *recall* (revocação ou cobertura). A métrica de desempenho *F-measure* apresenta a Fórmula 2.2 para tarefas de classificação binária. Quando a classificação é de *multiclasses* ou *multirrótulos* a média do *F-measure* de cada classe deve ser ponderada dependendo do parâmetro, como no caso do *Weighted F-measure*, que calcula métricas para cada rótulo e encontra a média ponderada pelo suporte (número de instâncias verdadeiras para cada rótulo).

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{2.2}$$

- **Validação Cruzada**: Realizando as validações dos classificadores, é possível ajustar as configurações do modelo manualmente para que ele se torne ideal. Havendo a possibilidade do conhecimento dos dados "vazar" para o modelo, ocorrendo o chamado *overfitting*, onde o modelo aprendeu os dados de treino e se ajustou a eles, podendo se tornar inútil com novas amostras. Para que isso não se torne um problema é possível utilizar a Validação Cruzada (*Cross-validation*), onde os dados são divididos em K conjuntos menores e são usados K - *1* para treinamento e 1 para teste. Após avaliada a performance com o conjunto de teste que não foi submetido o processo é repetido alternando os conjuntos, assim é usado um dos conjuntos para teste e os demais para treino. O desempenho será a média dos K classificadores durante o processo (SCIKIT-LEARN, 2019).

2.2.3 Hierarquia do Aprendizado

Faceli *et al.* (2011) em seu livro[6] mostram que os algoritmos de AM podem ser organizados de diferentes maneiras. Uma dessas maneiras é em relação ao paradigma de aprendizado, esses critérios de AM podem ser divididos em preditivas e descritivas, e são apresentados a seguir:

[6] Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina - ISBN 978-85-216-1880-5

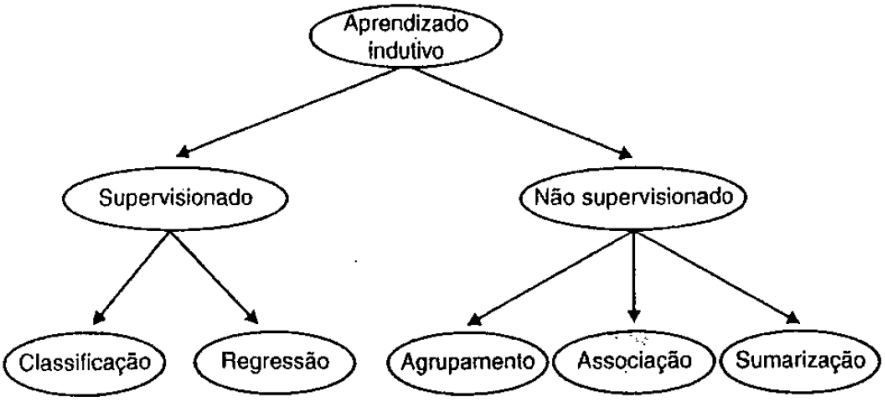
- **Preditivas**: Tem o objetivo de encontrar padrões, a partir dos dados de treinamento que possam ser utilizados na previsão de rótulos de novos exemplos. Algoritmos de AM, utilizados nesta tarefa, seguem o paradigma de aprendizado supervisionado. O termo supervisionado se refere ao fato da presença de um supervisor externo, que conhece previamente saída, ou os rótulos. Assim, ele pode avaliar a capacidade da hipótese induzida e predizer o valor de saída para novos exemplos.

- **Descritivos**: Tem o objetivo de explorar ou descrever um conjunto de dados. Os algoritmos utilizados para descrição, não fazem uso de atributos de saída (ou rótulos), seguindo um paradigma de aprendizado não supervisionado, pois não há presença de um especialista que conhece os dados. Uma aplicação descritiva é no agrupamento de dados, onde a meta é encontrar grupos de objetos semelhantes no conjunto de dados. A aplicação da descrição também pode ser para encontrar regras de associação que relacionam um grupo de atributos a outro grupo de atributos.

De acordo com esses paradigmas, para Faceli *et al.* (2011), tem-se uma hierarquia de AM, com base no uso de indução. O aprendizado por meio da indução realiza generalizações a partir dos dados. Tem-se então o aprendizado supervisionado (preditivo) e o não supervisionado (descritivo). Assim, nas tarefas supervisionadas os rótulos dos dados são: discreto, no caso de classificação; e, contínuo, no caso de regressão.

No paradigma descritivo, são genericamente divididos em: agrupamento, em que os dados são agrupados de acordo com sua similaridade; sumarização, cujo objetivo é encontrar uma descrição simples e compacta para um conjunto de dados; e associação, que consiste em encontrar padrões frequentes de associações entre os atributos de um conjunto de dados. Na Figura 13, pode-se observar a hierarquia de aprendizado segundo Faceli *et al.* (2011). Mais desses conceitos são apresentados na Subseção 2.2.5.

Figura 13 – Hierarquia de aprendizado.



Fonte: Faceli *et al.* (2011)

2.2.4 Classificação de Dados

Segundo Monard e Prati (2005), os sistemas de AM podem ser classificação segundo várias dimensões, como: Modo; Paradigma; Forma de aprendizado; e, Linguagem de descrição utilizada para descrever exemplos e conhecimento. Na Tabela 3 estão as classificações do AM.

Tabela 3 – Classificação de Aprendizado de Máquina.

Modos	**Paradigmas**	**Formas**	**Linguagens de Descrição**
Supervisionado	Simbólico	Incremental	Exemplos ou Objetos
	Estatístico		
Não supervisionado	Baseado em exemplos	Não-Incremental	Hipóteses
Semissupervisionado	Conexionista		Conhecimento de Domínio
	Evolutivo		

Fonte: Monard e Prati (2005)

Os Paradigmas são definidos por Monard e Baranauskas (2003) da seguinte forma:

- **Simbólico**: Os sistemas de aprendizado simbólico buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contra exemplos desse conceito. As representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras ou rede semântica.

- **Estatístico**: Consiste em utilizar modelos estatísticos para encontrar uma boa aproximação do conceito induzido. Por exemplo, um classificador linear assume que as classes podem ser expressas como combinações lineares dos valores dos atributos, procurando uma combinação linear que forneça a melhor aproximação sobre o conjunto de dados. Dentre os métodos estatísticos podem ser citados os baseados em *redes bayesianas*.

- **Baseado em Exemplos**: Tem como objetivo classificar exemplos nunca vistos, através de exemplos similares vistos anteriormente. São exemplos desse paradigma os algoritmos do tipo Raciocínio Baseado em Casos ou *Nearest Neighbor* (vizinho mais próximo).

- **Conexionista**: Redes Neurais são construções matemáticas simplificadas inspiradas no modelo biológico do sistema nervoso. A representação de uma *Rede Neural* envolve unidades altamente interconectadas e, por esse motivo, o nome conexionismo é utilizado para descrever a área de estudo. A metáfora biológica com as conexões neurais do sistema nervoso tem interessado muitos pesquisadores, e tem fornecido muitas discussões sobre os méritos e as limitações dessa abordagem de aprendizado. Em particular, as analogias com a biologia tem levado muitos pesquisadores a acreditar que as *Redes Neurais* possuem um grande potencial na resolução de

problemas que requerem intenso processamento sensorial humano, tal como visão e reconhecimento de voz.

- **Genético**: Um classificador genético consiste de uma população de elementos de classificação que competem para fazer a predição. Elementos que possuem uma performance fraca são descartados, enquanto os elementos mais fortes proliferam, produzindo variações de si mesmos. Este paradigma possui uma analogia direta com a teoria de Darwin, na qual sobrevivem os mais bem adaptados ao ambiente.

Já em relação a forma de aprendizado, os algoritmos de AM são classificados em não-incremental e incremental por Monard e Prati (2005):

- **Não-incremental**: Quando necessita que todos os exemplos utilizados pelo algoritmo estejam simultaneamente disponíveis.
- **Incremental**: Não necessita que todos os exemplos utilizados pelo algoritmo estejam simultaneamente disponíveis, pois tenta atualizar a hipótese corrente sempre que novos exemplos são adicionados ao conjunto de treinamento.

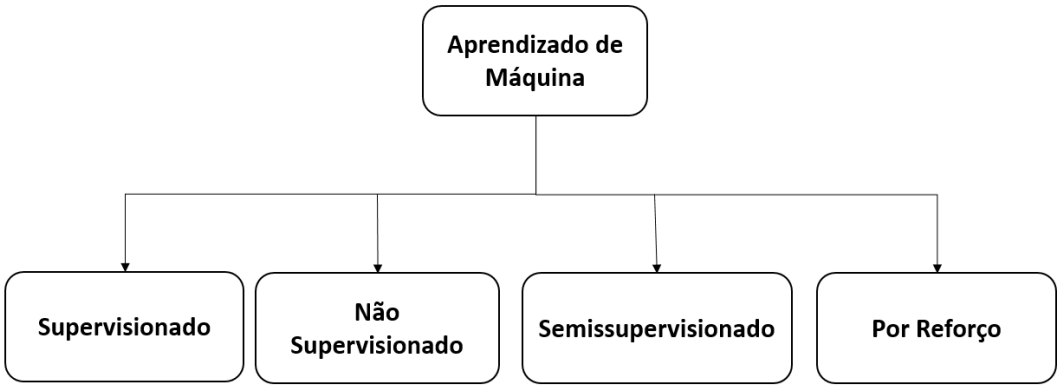
Qualquer que seja o tipo de aprendizado, são necessárias linguagens para descrever exemplos, hipóteses e conhecimento do domínio. Essas linguagens dependem do paradigma. Os Modos de AM possuem outras divisões além das apresentadas por Monard e Prati (2005). Na Subseção a seguir são apresentadas as classificações dos modos ou tipos de AM.

2.2.5 Tipos de Aprendizado de Máquina

Como apresentadas nas Subseções 2.2.3 e 2.2.4, o AM possui alguns tipos ou modos. Diversos autores na literatura apresentam os tipos de AM. Lorena e Carvalho (2007), os classifica em Aprendizado Supervisionado e Aprendizado Não Supervisionado. Já Faceli *et al.* (2011), os classifica em Aprendizado Supervisionado, Aprendizado Não Supervisionado, e Aprendizado Por Reforço. Matsubara (2004) apresenta outra definição diferente dos demais, onde ele cita o Aprendizado Semissupervisionado.

Russell e Norvig (2014) apresentam os AM citados anteriormente, ou seja: Aprendizado Supervisionado; Aprendizado Não Supervisionado; Aprendizado Por Reforço e Aprendizado Semissupervisionado. Na Figura 14 estão os tipos de aprendizado de máquina, e em seguida são apresentadas as suas definições.

Figura 14 – Tipos de Aprendizado de Máquina.



Fonte: Autoria Própria

Aprendizado Por Reforço

O Aprendizado Por Reforço (APR), é apresentado por Faceli *et al.* (2011) como uma tarefa na qual a meta é reforçar ou recompensar uma ação considerada positiva e punir uma ação considerada negativa. Um exemplo de tarefa de reforço é a de ensinar um robô a encontrar a melhor trajetória entre dois pontos.

Algoritmos de aprendizado utilizados nessa tarefa, em geral, punem a passagem por trechos pouco promissores e recompensam a passagem por trechos promissores. A aprendizagem por reforço é o único caminho possível para treinar um programa com desempenho de alto nível (RUSSELL; NORVIG, 2014).

Aprendizado Não Supervisionado

O Aprendizado Não Supervisionado (ANS), tem como objetivo explorar ou descrever um conjunto de dados (FACELI *et al.*, 2011). No ANS, segundo Russell e Norvig (2014), o agente aprende padrões na entrada, embora não seja fornecido nenhum *feedback* explícito. A tarefa mais comum de aprendizagem não supervisionada é o agrupamento: a detecção de grupos de exemplos de entrada potencialmente úteis.

Segundo Lorena e Carvalho (2007), no ANS não há a presença de um professor, ou seja, não existem exemplos rotulados. O algoritmo de ANS aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade. Essas técnicas são utilizadas principalmente quando o objetivo for encontrar padrões ou tendências que auxiliem no entendimento dos dados.

Aprendizado Supervisionado

No Aprendizado Supervisionado (AS), é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada

é conhecido (MONARD; BARANAUSKAS, 2003). Esse AM tem como objetivo encontrar um modelo ou hipótese, a partir de dados de treinamento, que podem ser utilizados para prever um rótulo ou valor que caracterize um exemplo novo, com base nos atributos de entrada (FACELI *et al.*, 2011).

Russell e Norvig (2014) afirmam que nesse tipo de aprendizado o agente observa alguns exemplos de pares de entrada e saída, e aprende uma função que faz o mapeamento da entrada para a saída. Quando a saída do conjunto de treinamento são valores finito (como ensolarado, nublado ou chuvoso), o problema da aprendizagem será chamado de classificação, e se esses dados apresentarem apenas dois valores, será chamado de classificação *booleana* ou binária.

Quando a saída do conjunto de treinamento for um número (como temperatura de amanhã), o problema de aprendizagem é chamado de regressão. Tecnicamente, a solução de um problema de regressão é encontrar uma expectativa condicional ou valor médio do treinamento porque a probabilidade de acharmos exatamente o número de valor real certo para o treinamento é 0 (RUSSELL; NORVIG, 2014).

Aprendizado Semissupervisionado

No Aprendizado Semissupervisionado (ASS) são usados exemplos rotulados e não rotulados, pois no AS às vezes é trabalhoso, lento e custoso conseguir dados rotulados e usar o ANS pode não ser trivial para descobrir o conceito embutido nos nós dos *clusters* encontrados. Como é uma tarefa difícil solicitar a um especialista que classifique todos esses dados, ainda é possível solicitar ao especialista que classifique alguns desses dados com alta certeza. Assim, é possível utilizar algoritmos de ASS (MATSUBARA, 2004).

Segundo Russell e Norvig (2014), no ASS são dados alguns poucos exemplos rotulados e deve-se fazer o que puder na coleção de exemplos não rotulados, mesmo os rótulos em si podem não ser as verdades que esperamos.

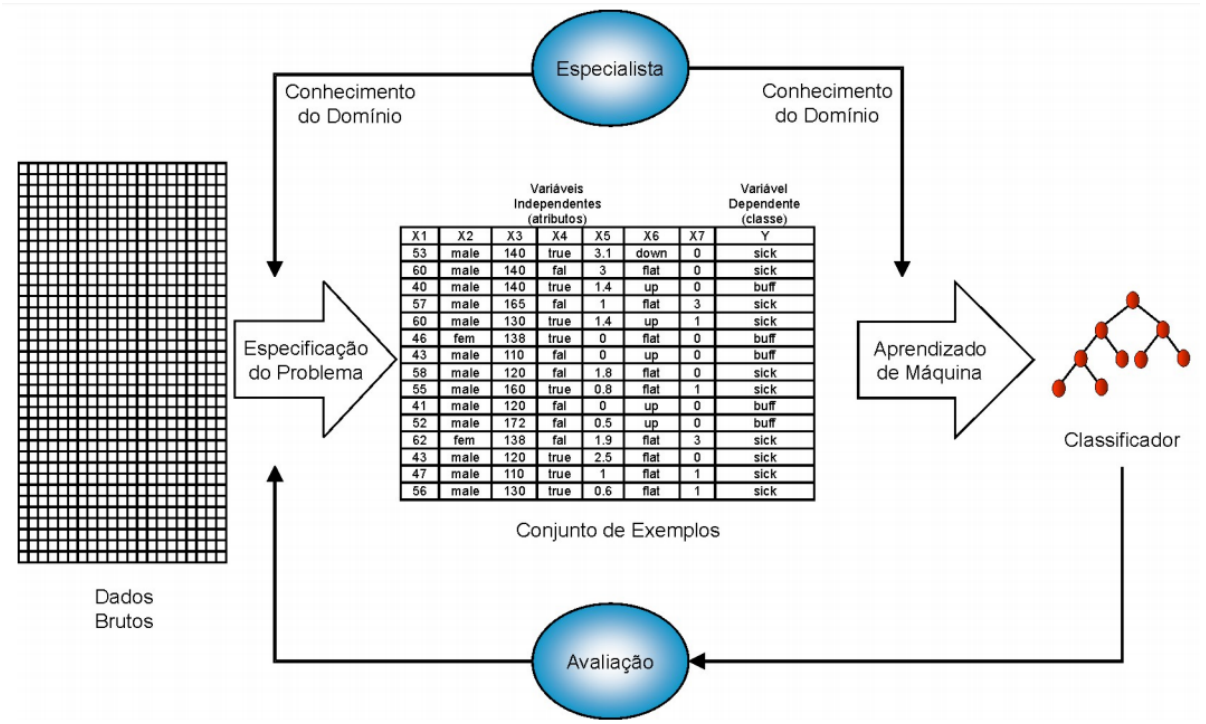
O ASS emprega estratégia que estendem o AS ou ANS para incluir informações adicionais do outro paradigma de aprendizagem. Por exemplo, a classificação semissupervisionada tem como objetivo treinar um classificador com dados rotulados e não rotulados, para obter um melhor classificador do que se fosse treinado só com dados rotulados (MANCHEGO, 2013).

2.2.6 Processo de Classificação

O processo de classificação, segundo Monard e Baranauskas (2003), pode ser ilustrado de maneira geral pela Figura 15, com o conhecimento sobre o domínio é possível escolher os dados ou fornecer alguma informação previamente conhecida como entrada ao indutor. Após a indução, o classificador é geralmente avaliado e o processo de classificação

pode ser repetido, se necessário, por exemplo, adicionando outros atributos, exemplos ou mesmo ajustando alguns parâmetros no processo de indicação.

Figura 15 – Processo de classificação de dados.

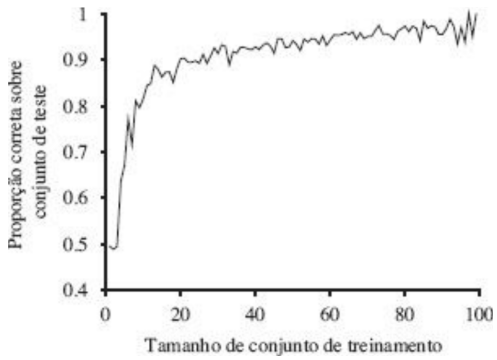


X1	X2	X3	X4	X5	X6	X7	Y
53	male	140	true	3.1	down	0	sick
60	male	140	fal	3	flat	0	sick
40	male	140	true	1.4	up	0	buff
57	male	165	fal	1	flat	3	sick
60	male	130	true	1.4	up	1	sick
46	fem	138	true	0	flat	0	buff
43	male	110	fal	0	up	0	buff
58	male	120	fal	1.8	flat	0	sick
55	male	160	true	0.8	flat	1	sick
41	male	120	fal	0	up	0	buff
52	male	172	fal	0.5	up	0	buff
62	fem	138	fal	1.9	flat	3	sick
43	male	120	true	2.5	flat	0	sick
47	male	110	true	1	flat	1	sick
56	male	130	true	0.6	flat	1	sick

Fonte: Monard e Baranauskas (2003)

Em seu livro[7], Russell e Norvig (2014) indagam sobre a quantidade de dados necessários para o aprendizagem de máquina, e em seguida afirma que, quanto mais dados de treinamento, há um melhor treinamento, conforme apresentado na Figura 16, onde é possível visualizar uma curva de aprendizagem em algoritmo de árvore de decisão.

Figura 16 – Exemplo de curva de aprendizagem em árvore de decisão em domínio de restaurante.



Fonte: Russell e Norvig (2014)

Pode-se observar que à medida que o tamanho de conjunto de treinamento cresce, também aumenta a proporção de classificações corretas dos testes, chegando próximo a 1

[7] Inteligência Artificial – Tradução da Terceira Edição - ISBN 978-85-352-3701-6

ou 100%. Pois com mais conjunto de treinamento o indutor terá mais conhecimento sobre aquele domínio, simulando um especialista real. Para realizar esse processo é necessário o classificador, assim são apresentadas na Subseção a seguir algumas técnicas para gerar classificadores.

2.2.7 Classificadores

Nesta Subseção são apresentadas algumas técnicas para gerar classificadores de AM.

Máquina de Vetores de Suporte

A *Máquina de Vetores de Suporte* (MVS) é um técnica de AS usado para classificação de dados binários linearmente não probabilísticos. A MVS tem a capacidade de ajustar um hiperplano que separa duas ou mais classes diferentes. Quando os dados são linearmente não separáveis, é usado o truque do *kernel* que mapeia as entradas de dados para espaços dimensionais mais altos de recursos nos quais eles são facilmente separáveis, ou seja, esses dados passam para mais de duas dimensões onde será possível separá-los por meio de uma curva linear (ALCHALABI *et al.*, 2017).

Rede Neural Artificial

A *Rede Neural Artificial* é uma técnica de AM amplamente utilizada na área. Uma de suas muitas aplicações é no planejamento de caminhos, pois ela tem uma grande capacidade de aprender caminhos não lineares (AHMAD *et al.*, 2013).

Nela é definido um conjunto de neurônios de entrada que são ativados para realizar o processamento. Os dados adquiridos por essa ativação dos neurônios são então ponderados e transformados por uma função determinada na rede, sendo repassados para outros neurônios. O processo é finalizado quando um neurônio de saída é ativado. Finalizando com a classificação daquela exemplo.

Redes Bayesianas

As *Redes Bayesianas* ganharam importância no mundo científico graças à sua utilidade na modelagem e tratamento em incerteza, principalmente na medicina (KINCAID *et al.*, 2003). As *Redes Bayesianas* podem ser adequadas no uso de simuladores de casos clínicos. Pearl (2014) sugeriu que o raciocínio humano deve adotar uma estratégia diferente, que desvie o foco da faceta quantitativa da representação das probabilidades, para dar mais atenção às relações de dependência entre variáveis.

Com base nisso, a estrutura de conhecimento utilizada para avaliação humana é do tipo de gráficos de dependência e percorrer as conexões entre seus nós consiste nos

processos básicos de pesquisa, o que fazem as Redes Bayesianas (FLORES *et al.*, 2013). As redes Bayesianas são adequadas para modelar o conhecimento e apoiar o raciocínio sob a modelagem do conhecimento de incerteza (FLORES *et al.*, 2013).

Árvore de decisão

A *Árvore de decisão* é normalmente usada em problemas de classificação. Nesta técnica, o conjunto de dados é aprendido e modelado. O objetivo é subdividir o problema complexo e grande em subproblemas menores que são mais simples de serem resolvidos.

Assim, cada subproblema ainda pode ser subdividido em menores, com intuito de simplificar o problema (CHAVES, 2012). Por fim, a estrutura se torna semelhante à de uma árvore, na qual cada nó é responsável por realizar uma decisão até chegar em uma folha, que representa o rótulo do exemplo.

Regressão linear

A *regressão linear* é uma equação para se estimar a condicional de uma variável, utilizando os valores de algumas outras variáveis. Assim, é possível prever um valor esperado utilizando outros dados. Entretanto diferente de outras técnicas, na *regressão linear* o resultado é algo continuo, como a temperatura.

Q-learning

O *Q-learning* é uma técnica de APR que faz com que agentes computacionais aprendam a agir de maneira ideal em domínios markovianos. Produzindo efeito incremental para programação dinâmica que impõe demandas computacionais limitadas. Melhorando sucessivamente suas avaliações da qualidade de ações específicas em estados específicos (WATKINS; DAYAN, 1992).

Portanto o *Q-learning* aprende uma política, e diz a um agente que ação tomar em uma circunstância. O *Q-learning* encontra uma política ideal no sentido de maximizar o valor esperado da recompensa total em todas e quaisquer etapas sucessivas, a partir do estado atual (MELO, 2001).

Regressão logística

A *Regressão logística* é uma técnica de AM baseada em estatística. Na qual a partir de um conjunto de observações, um modelo é capaz de realizar a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias (JR; LEMESHOW; STURDIVANT, 2013).

Quando aplicada ela gera um resultado que varia entre 0 e 1 por meio de regressão, mas ao fim com base nesse resultado é feita uma classificação dos dados. Sendo assim

diferente da *regressão linear*, que produz uma saída de valor contínuo e não discreto, como na *Regressão logística*.

K-Vizinhos mais próximos

A técnica de K-Vizinhos mais próximos, também conhecida por *K-Nearest Neighbor* é uma técnica de AM baseado em estatística, na qual os exemplos de teste são adicionados em um espaço de recursos. A saída no caso de uma classificação é uma associação de classe utilizando os K vizinhos mais próximos sendo o objeto atribuído à classe mais comum entre esses vizinhos mais próximos (COVER; HART, 1967).

Se K for 1, por exemplo, então o rótulo será atribuído ao vizinho que estiver mais próximo. Nesta técnica ainda pode ser adaptado o peso dos vizinhos com base no problema e/ou o número de vizinhos, na qual é tomada a decisão final.

Naive Bayes

O método de Bayes Ingênuo, popularmente conhecido por *Naive Bayes* é uma técnica de aprendizado supervisionado baseado na aplicação do teorema de Bayes com a suposição de independência condicional entre os pares de recursos, dado a classificação da classe (LEWIS, 1998). O *Naive Bayes* é um método rápido, pois se baseia apenas em estatística, diferente dos métodos mais sofisticados (ZHANG, 2004).

2.3 COMBINAÇÃO DE CLASSIFICADORES

Combinar múltiplas entradas para inferir informações sobre o ambiente é muito natural, pois é feita diariamente pelos humanos, nós combinamos diariamente informações acústicas, visuais, táteis, olfatórias e térmicas para gerir sobre o mundo a volta. Um exemplo muito frequente é o de sistemas biológicos que adotam tais esquemas (ACHERMANN; BUNKE, 1996 apud SALVADEO, 2009).

Neste contexto, fusão de sensores e combinação de classificadores têm sido explorados, sendo baseados nessas inspirações biológicas, como uma alternativa para obter melhores desempenhos em tarefas de reconhecimento de padrões sem precisar aumentar a complexidade. Apesar dessa área que está crescendo rapidamente, ainda há muito empirismo nessa área (SALVADEO, 2009).

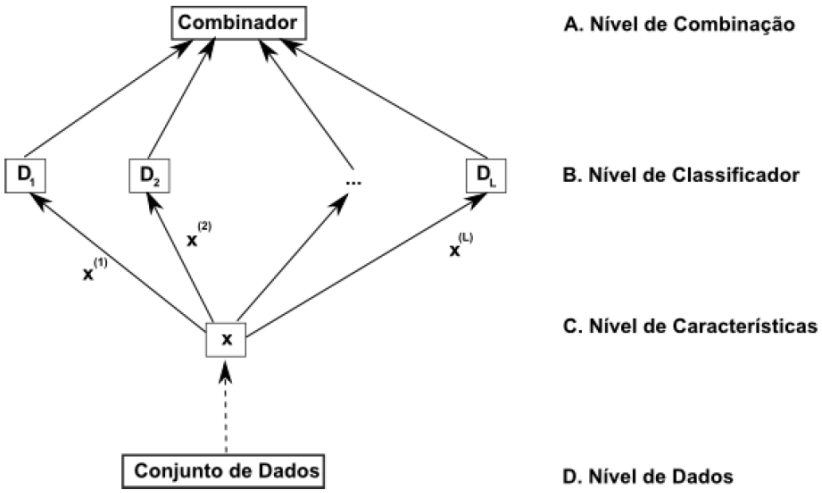
Várias aplicações forçam seus esforços em combinar múltiplos classificadores, resolvendo problemas complexos de modo mais robusto e eficiente (YANG; AULT; PIERCE, 2000); pois nenhum classificador é capaz de resolver todos os problemas em qualquer base de dados (WOLPERT, 2002). A combinação de classificadores baseia-se no objetivo de unir classificadores com características diferentes, e melhor ainda se complementares. A união

de diferentes classificadores traz as vantagens destes e dispensa as fraquezas particulares de cada um. Assim, é esperado que uma combinação melhore o desempenho (BI *et al.*, 2004).

As saídas dos classificadores individuais devem ser fortemente não correlacionados nos erros de classificação, ou seja, os classificadores não devem errar na classificação de uma mesma amostra, ou ao menos não devem associar uma amostra à mesma classe incorreta (SALVADEO, 2009). Com essa complementaridade entre os classificadores espera-se que havendo o erro de um, os outros acertem, ou, pelo menos, forneçam uma classificação diferente dos outros erros, fazendo com que uma fusão dos classificadores possa obter bons desempenhos. A fim de obter essa complementariedade um *ensemble* (conjunto de classificadores) pode ser estabelecido.

Para isso, segundo Salvadeo (2009), podemos variar os combinadores (nível de combinação), os classificadores individuais ou especialistas (nível de classificador), os atributos determinados por técnicas de seleção ou extração de atributos diferentes (nível de características) e os subconjuntos de dados (nível de dados). Na Figura 17 pode-se observar essa ilustração.

Figura 17 – Abordagens para a construção de *ensembles* em combinação de classificadores.



Fonte: Salvadeo (2009)

A combinação de classificadores é um novo classificador, que toma como entrada a resposta dos classificadores e como saída uma decisão final sobre eles. As respostas dos classificadores podem se apresentar em 3 níveis segundo Xu, Krzyzak e Suen (1992 apud SALVADEO, 2009):

- **Abstração** (menos informação para a tomada de decisão): apenas a informação do rótulo da classe associada ao padrão é fornecida por cada classificador.

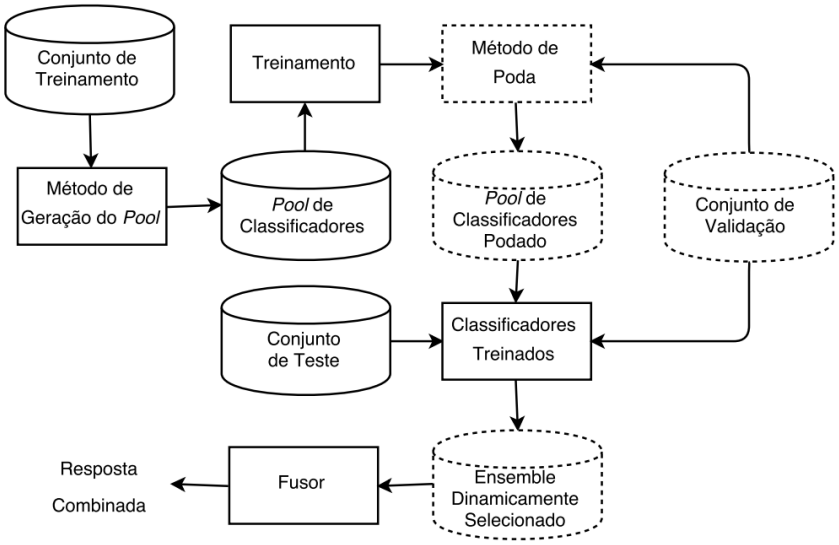
- ***Ranking***: uma lista contendo a ordem das classes é fornecida pelos classificadores individuais, onde a classe mais provável está no topo da lista e recebe a 1^a posição no *rank*.

- **Medida** (mais informação disponível para a tomada de decisão): uma medida que pode ser interpretada como o grau de confiança de determinada classe ser a correta é fornecida por cada classificador.

Assim, antes de combinar a saída de diversos classificadores, as saídas precisam ser colocadas num mesmo domínio e normalizadas. Isso implica que o tipo de informação gerada pelos classificadores individuais influenciam a forma de combinar, como o domínio de probabilidades que é normalmente considerado (SALVADEO, 2009).

Já Pinheiro (2017), expõe a arquitetura de um sistema de múltiplos classificadores, conforme apresentado na Figura 18. Segundo Pinheiro (2017), esse sistema inicia com a geração do *pool* de classificadores, isto é, alguma metodologia para criar classificadores diferentes para serem combinados; em seguida todo o *pool* (poda) de classificadores é treinado; alguns desses classificadores podem ser removidos, com base e métodos de poda, que utilizam variadas abordagens, seja com base na diversidade, performance ou algum outro pré-requisito. Para realizar a remoção utiliza-se um conjunto de validação; com os classificadores treinados é possível iniciar a fase de teste, isto é, apresentar exemplos ainda não classificados para verificar o desempenho do sistema como um todo.

Figura 18 – Arquitetura de um sistema de múltiplos classificadores.



Fonte: Pinheiro (2017)

Durante a fase de teste é possível existir uma seleção dinâmica do *ensemble*, para utilizar apenas os melhores classificadores naquele exemplo que será classificado, para isso é utilizado um conjunto de validação; assim, no final, o *fusor* que irá gerar uma única

resposta ou classificação a partir das respostas dadas por todos os classificadores. As etapas de linha tracejada não são de uso obrigatórias, pois podem variar dependendo da necessidade da aplicação.

Uma definição mais atual, em relação a anterior citada por Xu, Krzyzak e Suen (1992 apud SALVADEO, 2009), é que a forma mais comum de combinar as respostas dos classificadores é utilizando voto majoritário, essas combinações podem funcionar de três modos segundo Woźniak, Graña e Corchado (2014):

- A classe será aquela que todos os classificadores votaram nela;
- A classe será aquela que conseguir uma quantidade de votos maior que a metade dos classificadores após a poda;
- A classe será aquela que conseguir a maior quantidade de votos em comparação com as outras.

Portanto, por meio dessas combinações apresentadas anteriormente, são possíveis construir MLCs para aumentar o desempenho inicial dos classificadores, tomando como base para decisão das combinações os tipos de dados, as precisões e as saídas dos classificadores. A seguir são apresentados alguns meta-classificadores que também fazem a combinação dos indutores para formar um MLC.

2.3.1 Meta-classificadores

Nesta Subseção são definidos conceitos sobre alguns meta-classificadores que utilizam algumas das técnicas base e gera diversos classificadores alternando os parâmetros e agregando os resultados de cada um para impulsiona-los. Também é apresentado a *Floresta aleatória*, um método de classificadores *ensemble*.

Bagging

O meta-classificador de *Bootstrap aggregating* mais conhecido por *Bagging* foi publicado por Breiman (1996) no ano de 1996. Ele utiliza vários classificadores base de um mesmo algoritmo, e reduz a variância na classificação. Para isso são gerados conjuntos sucessivos e independentes das amostras de dados no treinamento. Cada uma dessas amostras é gerada escolhendo de forma aleatória n exemplos, por meio de substituição, para que todas apresentem a mesma quantidade de exemplos. O resultado final é agregado por meio da confiança (média) desses classificadores, ou pelo voto majoritário.

Boosting

O *Boosting* é um meta-classificador utilizado para aprimorar o desempenho de qualquer algoritmo, como *Rede Neural Artificial* e *Árvore de decisão*. Assim, ele transforma classificadores "fracos" em "fortes". Neste meta-classificador os parâmetros do classificador base são ajustados de acordo com os erros anteriores cometidos.

No final é gerado um classificador "forte" que possui a composição de outros classificadores ajustados, em que o resultado final é agregado também pela média ou pelo voto majoritário (FREUND; SCHAPIRE *et al.*, 1996). A diferença em relação ao *Bagging* e que o *Boosting* gera sequencialmente o conjunto de treino e os classificadores de acordo com o resultado da iteração anterior. Enquanto o *Bagging* gera o conjunto de treino de forma aleatória e pode gerar os classificadores de forma paralela.

Floresta aleatória

Floresta aleatória é um método de classificadores *ensemble* usado para classificação ou regressão e dentre outras tarefas. Ela utiliza a *árvore de decisão* como classificador base, e constrói diversas *árvores de decisão* no momento do treinamento (BREIMAN, 2001).

A *floresta aleatória* corrige o hábito das *árvores de decisão* e tenta se ajustar ao seu conjunto de treinamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A previsão do conjunto é dada utilizando a confiança dos classificadores individuais. Inicialmente a *floresta aleatória* utilizava o *Bagging* para gerar a combinação, entretanto há pesquisas que utilizam *Boosting* (ZHANG; XIE, 2010).

3 REVISÃO SISTEMÁTICA DA LITERATURA

Neste Capítulo, é apresentado todo o processo de condução da Revisão Sistemática da Literatura realizada e seus resultados, nela estão todos os trabalhos relacionados encontrados no âmbito internacional de acordo com o protocolo desta RSL.

3.1 ESCOPO DA REVISÃO SISTEMÁTICA DA LITERATURA

Na medicina, o AM pode ser aplicado a um conjunto de dados com o propósito de desenvolver modelos de riscos robustos (DEO, 2015), o que não é surpreendente, pois podem ser aplicados a uma ampla gama de campos como financeiro, astronômico e biológico (DEO *et al.*, 2014), facilmente reduzindo a tarefa de prever resultados de diversos recursos ou encontrando padrões recorrentes em conjunto de dados multidimensionais. A aplicação do AM na medicina enfrenta alguns obstáculos como da precisão realizada por um especialista humano em relação à máquina (DEO, 2015).

No desenvolvimento dos jogos sérios, a inteligência artificial, por meio do AM, oferece um potencial significativo, aprimorando a experiência do jogador em todas as etapas do jogo, ajudando a melhorar vários estágios dentro do desenvolvimento de jogos sérios (FRUTOS-PASCUAL; ZAPIRAIN, 2017). Os jogos sérios apresentam características explícitas e cuidadosamente pensadas com propósitos educacionais e não se destinam a serem jogados somente para diversão (ABT, 1987).

Segundo Clua (2014), o uso de jogos sérios para fins relacionados à saúde tem se tornado uma forte tendência e o seu impacto pode abrir formas de tratamento sem precedentes. Dentro deste contexto, várias pesquisas tem sido realizadas buscando explorar e mapear as mais diversas áreas da medicina e da saúde para o universo dos jogos digitais (CLUA, 2014).

O AM pode ser usado de diversas maneiras na medicina, como na previsão de crises e doenças. Os jogos para medicina são de grande valia e também podem auxiliar de diversas maneiras, tais como: treinamento de médicos; identificação de doenças; e, tratamento de doenças. Para desenvolver um jogo para medicina é possível utilizar o AM. Entretanto existem várias tipos, técnicas e tarefas de AM existentes. Assim, surgiu a necessidade de se descobrir quais as principais formas de se aplicar o AM para desenvolver jogos para medicina.

Diante disso, nesta RSL, são apresentados os jogos, técnicas, tarefas e áreas médicas

explorados pelo AM. As RSL são consideradas estudos secundários, que têm sua fonte de dados nos estudos primários (GALVÃO; PEREIRA, 2014). Por meio da RSL é possível identificar, avaliar e interpretar toda a pesquisa disponível relevante para uma questão específica ou área temática, ou fenômeno de interesse de forma repetitiva e imparcial (KITCHENHAM, 2004).

3.2 PROCESSO SISTEMÁTICO

Nesta Seção, são apresentados os detalhes da metodologia utilizada na RSL, que consiste em: definir os objetivo da revisão; definir as questões de pesquisa; definir a *string* de busca, definir as bases de dados; definir os critérios de inclusão, exclusão e qualidade; definir os procedimentos de seleção dos estudos; e, definir a extração dos dados. O processo sistemático dessa RSL foi realizado baseado no modelo de protocolo apresentado por Biolchini *et al.* (2005).

3.2.1 Objetivos e Escopo

O objetivo geral desta RSL é identificar o estado da arte da utilização do AM nos jogos para medicina, por meio da identificação e análise de jogos que utilizam AM com aplicações na medicina. Assim, para o alcance do objetivo geral foram definidos os seguintes objetivos específicos:

- **Objetivo 1:** Identificar os principais tipos e técnicas de aprendizagem de máquina utilizado nos jogos para medicina.
- **Objetivo 2:** Identificar os principais jogos sérios que utilizam aprendizagem de máquina na área da medicina.
- **Objetivo 3:** Identificar as principais áreas da medicina em que os jogos encontrados estão sendo aplicados.

3.2.2 Questões de Pesquisa

A questão primária de pesquisa a ser tratada neste trabalho é apresentada a seguir e foi decomposta em outras questões de pesquisa, que buscam alcançar o objetivo proposto.

- **Questão primária (QP):** Como o aprendizado de máquina está sendo utilizado nos jogos para medicina?
- **Questão secundária (QS1):** Qual o principal tipo de aprendizagem de máquina utilizado nos jogos para medicina?

- **Questão secundária (QS2):** Quais as principais técnicas de aprendizagem de máquina nos jogos para medicina?
- **Questão secundária (QS3):** Quais os principais jogos sérios que utilizam aprendizagem de máquina na área da medicina?
- **Questão secundária (QS4):** Quais as principais áreas da medicina em que os jogos encontrados estão sendo aplicados?

3.2.3 Definição da *String* de Busca

Para definir a *string* de busca, é importante encontrar as palavras-chave relacionadas à pesquisa. As palavras-chave nas revisões sistemáticas são definidas por meio das questões de pesquisa sob análise. Inicialmente, foram definidas as palavras-chave que seriam utilizadas na *string* de busca, e posteriormente realizou-se uma busca-piloto. Concluiu-se que as palavras deveriam ser todas no idioma inglês e que alguns sinônimos fossem adicionados. Na Tabela 4 estão as palavras-chave e sinônimos da estratégia de busca em Inglês.

Tabela 4 – Palavras-chave e Sinônimos.

Palavra-chave	**Sinônimos**
Machine learning	*Automatic learning, Automatically learning*
Medicine	*Medic, Health*
Game	*Games, Simulator, Simulation*

Fonte: Autoria Própria

Após realizar as combinações das palavras-chave e sinônimos foi definida a seguinte *string* geral de busca, conforme apresentada na Tabela 5.

Tabela 5 – *String* Geral de Busca.

String Geral
(“Machine learning” OR “Learning automatically” OR “Automatic learning”) AND (Games OR Game OR Simulator OR Simulation) AND (Medicine OR Medical OR Health)

Fonte: Autoria Própria

3.2.4 Bases de Dados

Nesta pesquisa foram utilizadas bases de dados eletrônicas indexadas e máquinas de busca eletrônica, conhecidas internacionalmente que indexam trabalhos da grande área da computação ou interdisciplinar. Com base nesses critérios, as bases de dados escolhidas são *Institute of Electrical and Electronics Engineers* (IEEE), *ACM Digital Library*, *Science Direct* e *Scopus*, conforme apresentado na Tabela 6.

Tabela 6 – Bases de Dados.

Base de Dados	**Endereço na Web**
IEEE	ieeexplore.ieee.org
ACM Digital Library	dl.acm.org
Science Direct	www.sciencedirect.com
Scopus	www.scopus.com

Fonte: Autoria Própria

3.2.5 Critérios de Inclusão, Exclusão e Qualidade

Os critérios de inclusão e exclusão são definidos para garantir a imparcialidade na condução da RSL. Trabalhos aceitos para análise devem atender a, pelo menos, um critério de inclusão. Foram definidos 2 critérios para inclusão (CI) de trabalhos, conforme apresentados a seguir.

- CI1: Trabalhos que apresentam técnicas de aprendizagem de máquina para jogos eletrônicos na área da medicina.
- CI2: Trabalhos que apresentam jogos eletrônicos que utilizam aprendizagem de máquina na área da medicina.

Trabalhos excluídos da análise devem atender a, pelo menos, um critério de exclusão. Foram definidos 3 critérios para exclusão (CE) de trabalhos, conforme apresentados a seguir.

- CE1: Trabalhos que não apresentam técnicas de aprendizagem de máquina;
- CE2: Trabalhos que não apresentam jogos eletrônicos;
- CE3: Trabalhos que não são relacionados à medicina.

Trabalhos com qualidade devem atender a todos os critérios de qualidade. Foram definidos 3 critérios para qualidade (CQ) de trabalhos, conforme apresentados a seguir.

- CQ1: Trabalhos publicados em jornais, revistas ou conferências com *qualis* igual ou acima de B5;
- CQ2: Trabalhos que apresentem 4 páginas ou mais;
- CQ3: Trabalhos escritos no idioma Inglês ou Português.

3.2.6 Procedimento para Seleção dos Estudos

O procedimento para seleção de estudos foi definido em 3 fases conforme detalha-se a seguir:

- **Fase 1:** Corresponde à busca e coleta de trabalhos. Nesta fase, as *strings* de busca formadas pela combinação dos sinônimos das palavras-chave identificadas são submetidas às máquinas de busca selecionadas e os trabalhos encontrados foram coletados.
- **Fase 2:** Corresponde à leitura dos resumos. Nesta fase, entre os trabalhos retornados, são lidos os resumos e títulos, e constatando-se a relevância de um trabalho, com base em algum dos critérios de inclusão, o trabalho é pré-selecionado para ser lido na íntegra. Os trabalhos que se encaixaram em algum dos critérios de exclusão são excluídos, e caso não se encaixem em nenhum, também são pré-selecionados.
- **Fase 3:** Corresponde à leitura dos trabalhos completos. Os trabalhos que foram aprovados na etapa anterior são lidos integralmente, novamente aplicados os CI e CE, e também aplicados os CQ. Os trabalhos aceitos nesta fase são utilizados para responder as questões de pesquisa.

3.2.7 Extração de dados

Nesta RSL foi realizada a extração dos dados considerados importantes neste estudo. A seguir, na Tabela 7, estão os dados extraídos dos trabalhos selecionados. Esses dados auxiliaram na condução e nas respostas das questões de pesquisa dessa RSL.

Tabela 7 – Formulário de Extração de dados.

Item	**Descrição**
Título	Título do estudo primário
Ano	Ano em que o estudo primário foi publicado
Base	Base de Dados do estudo
Fonte	A conferência, revista ou livro em que o estudo primário foi publicado.
Qualis	*Qualis* do trabalho
Idioma	Idioma do trabalho
Páginas	Número de Páginas
Objetivo	Objetivo do trabalho
Jogo	Nome do Jogo
Área do Jogo	Temática abordada da saúde
Visualização do Jogo	Forma do jogo ser visualizado
AM	Tipo de AM
Técnica	Técnica de AM
Tarefa	Tarefa de AM
Uso do AM	Forma que o AM foi usado
Ferramenta do AM	Ferramenta que foi usada para o AM
Eficiência	Eficiência da solução
Vantagens	Benefícios do uso da solução proposta
Impacto	Impacto da solução proposta
Limitações	Limitações apresentadas pelo autor
Validação	Como o trabalho foi validado.
Observações	Observações

Fonte: Autoria Própria

3.3 CONDUÇÃO DA REVISÃO SISTEMÁTICA

Nesta Subseção, são apresentados os resultados obtidos com as buscas dos estudos primários. As buscas nas bases de dados foram realizadas entre agosto e outubro de 2018 e foram limitadas aos trabalhos publicados entre os anos de 2008 e 2018. Na base de dados da IEEE, foi realizado uma *Advanced Search* por meio de *Command Search* com a *string* geral.

Na base de dados da *ACM Digital Library*, foi realizada uma *Advanced Search* por meio de *Query syntax*, a *string* de busca foi adaptada por conta dos operadores lógicos e foi usado o "+". Na base de dados da *Scopus*, foi realizada uma *Advanced Search*, por meio de *Query string*, esta busca foi realizada no título, resumo e palavras-chave dos trabalhos. Na base de dados da *Science Direct*, foi realizada uma *Advanced Search*, por meio do *Title, abstract or keywords* dos trabalhos. Na Tabela 8 estão as *strings* de buscas submetidas a cada base de dados.

Tabela 8 – *Strings* de busca.

Base	**String**
IEEE	*("Machine learning" OR "Learning automatically" OR "Automatic learning") AND (Games OR Game OR Simulator OR Simulation) AND (Medicine OR Medical OR Health)*
ACM Digital Library	*+("Machine learning" "Learning automatically" "Automatic learning") +(Games Game Simulator Simulation) +(Medicine Medical Health)*
Science Direct	*("Machine learning" OR "Learning automatically" OR "Automatic learning") AND (Games OR Game OR Simulator OR Simulation) AND (Medicine OR Medical OR Health)*
Scopus	*TITLE-ABS-KEY (("Machine learning" OR "Learning automatically" OR "Automatic learning") AND (games OR game OR simulator OR simulation) AND (medicine OR medical OR health))*

Fonte: Autoria Própria

Na Fase 1, foram encontrados 1.040 trabalhos, entre os quais 146 eram trabalhos repetidos. Na fase 2, foram selecionados 40 trabalhos após leitura dos resumos. Na Fase 3, após a leitura completa e baseado nos critérios de inclusão e qualidade, foram selecionados 12 trabalhos. A seguir, pode-se observar na Tabela 9 os detalhes em relação ao quantitativo dos trabalhos em cada fase.

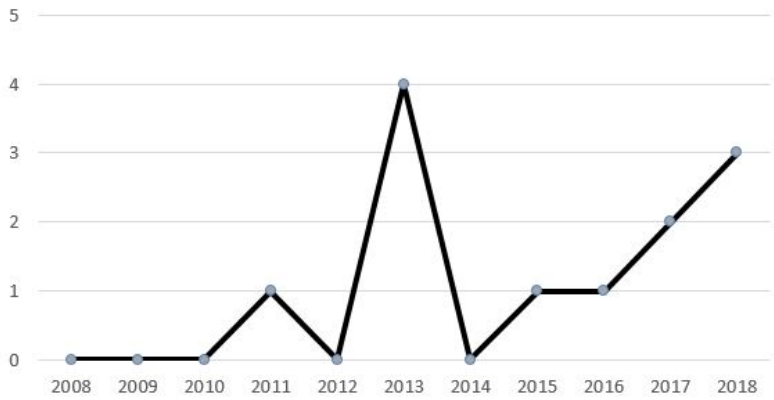
Tabela 9 – Resultados das buscas.

Base	**Fase 1**	**Fase 2**	**Fase 3**
IEEE	362	18	6
ACM Digital Library	313	9	2
Science Direct	25	1	0
Scopus	340	12	4
Total	1.040	40	12

Fonte: Silva *et al.* (2019)

Observa-se que a base de dados da IEEE registrou o maior número de trabalhos selecionados nesta pesquisa, com 6 trabalhos. Os trabalhos selecionados foram publicados entre os anos de 2011 e 2018, em 2013 foram encontrados 4 trabalhos, sendo o maior número de trabalhos encontrados por ano nessa revisão sistemática, seguido por 2018, com 3 trabalhos selecionados. Na Figura 19, pode-se observar os detalhes em relação a distribuição dos trabalhos por anos.

Figura 19 – Distribuição de estudos primários por ano.



Fonte: Autoria Própria

3.4 RESULTADOS E DISCUSSÕES

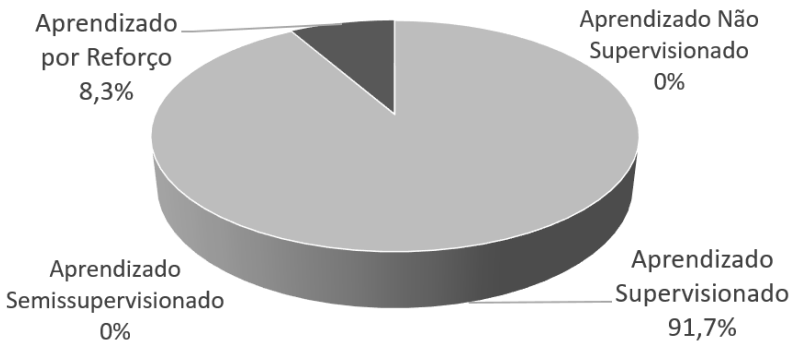
Os trabalhos que foram selecionados têm o potencial de auxiliar a responder aos questionamentos que nortearam essa pesquisa. A sua principal questão é: Como o AM está sendo utilizado nos jogos para medicina? E a partir dessa questão surgiram as questões secundárias que são apresentadas e respondidas nas subseções a seguir. Ao final tem-se a resposta da questão primária de pesquisa.

3.4.1 Qual o Principal Tipo de Aprendizagem de Máquina Utilizado nos Jogos para Medicina?

Na Seção 2.2.5, foram apresentados alguns dos tipos de AM existentes, sendo dividido em Aprendizado Supervisionado, Aprendizado Não Supervisionada, Aprendizado Semissupervisionado e o Aprendizado por Reforço. Os trabalhos encontrados por meio da *string* de busca foram divididos entres esses AM. Na Figura 20, pode-se observar a distribuição do uso dos tipos de AM nos trabalhos aceitos.

O trabalho apresentado por Chouhan *et al.* (2015), utilizou o AS, com o objetivo de aprender e prever os níveis de atenção dos jogadores para o próximo instante. Esse aprendizado se mostrou eficaz para realizar esta previsão. Entretanto, o trabalho possivelmente apresenta um erro no método de Entropia, pois houve leituras mais baixas dos níveis de

Figura 20 – Tipo de Aprendizado de Máquina.



Fonte: Silva *et al.* (2019)

atenção devido aos movimentos oculares e, portanto, um estágio de pré-processamento deve ser usado para descartar artefatos oculares ou algum outro algoritmo é necessário para estimar níveis de atenção quando movimentos oculares, principalmente a mudança no olhar estiver envolvida.

No trabalho apresentado por Heller *et al.* (2013), foi utilizado o AS para identificar o Transtorno do Déficit de Atenção com Hiperatividade (TDAH) dos jogadores. O aprendizado se mostrou eficaz com 70% de acerto, entretanto outros jogos apresentam maior nível de acerto como *Brief Rating Scale* descrito pelos autores.

Ahmad *et al.* (2013) utilizaram o AS para tentar aprender e encontrar o caminho ideal em microcirurgias realizadas por especialistas no simulador. O aprendizado se mostrou relativamente médio em eficiência. Seu desempenho não foi muito encorajador para o exercício de agarrar em microcirurgias, provavelmente pelos poucos dados de treinamento, o que pode ser melhorado com a inclusão de mais dados de treinamento.

No trabalho apresentado por Alchalabi *et al.* (2017), o AS foi usado na representação dos modelos de níveis de atenção dos usuários. O aprendizado se mostrou eficiente na classificação dos dados de Eletroencefalografia (EEG) dos usuários, entretanto, foram usados poucos números de amostras e apenas de usuários saudáveis e não de pessoas diagnosticadas com TDAH, o que era objetivo do trabalho.

Alchalabi *et al.* (2018) apresentaram o mesmo aprendizado do trabalho mencionado anteriormente, mas foram apresentados novos resultados e validações, o que demonstrou uma taxa de 96% de acerto na classificação dos dados do EEG para detectar o estado de atenção correto durante o jogo em indivíduos saudáveis, e de 98% na classificação dos dados do EEG para detectar o estado de atenção correto durante o jogo em indivíduos com TDAH.

No trabalho apresentado por Hervás *et al.* (2016), os autores usaram o AS para detecção da interação empática e não empática, possibilitando o diagnóstico de Distúrbios da Comunicação Social. O seu uso teve bom desempenho somente com dados sintéticos, e

está longe de representar uma solução definitiva para fornecer suporte no diagnóstico de Distúrbios da Comunicação Social, o que pode ter acontecido pelo pequeno conjunto de treinamento.

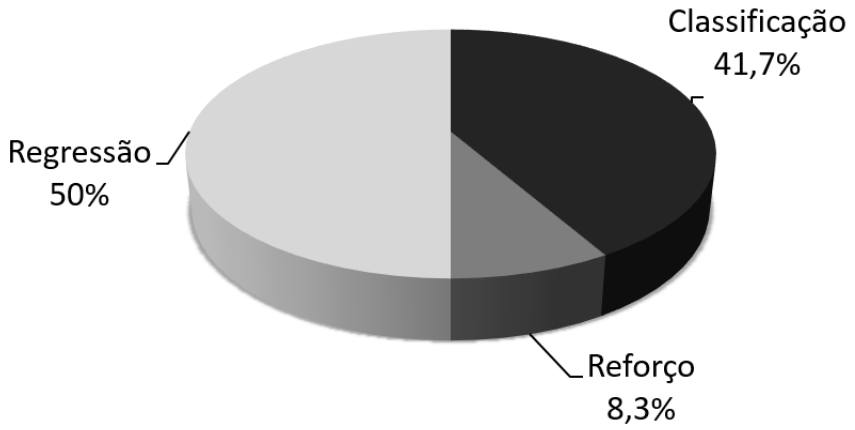
O APR foi utilizado em um trabalho apresentado por Perdiz *et al.* (2018), a fim de intermediar possíveis ações do usuário e instruções de atuação do jogo, e tentar adaptar sua resposta para maximizar os resultados e assim chegar ao melhor resultado no jogo. Apesar dos jogadores não apresentarem um efeito perceptível nos *scores*, o uso da técnica de aprendizado por reforço pode ser tomada ao projetar jogos controlados por *biosignal* com pessoas com deficiência e *Assistive Living* em mente.

Os trabalhos apresentados por Lima *et al.* (2016), Zielke *et al.* (2018), Flores *et al.* (2013), George *et al.* (2011) e Kim *et al.* (2013) não informaram o tipo de aprendizado utilizado. Entretanto, acredita-se que esses trabalhos usam o AS, pois descrevem que o aprendizado pode aprender com experiências passadas, ou que foi utilizado dados de treinamento, o que se trata de uma característica do AS.

Dessa forma, pode-se observar que o Aprendizado Supervisionado é o mais usado nos jogos para medicina, pois foi utilizado por 91,7% dos trabalhos. O AS tem uma precisão que é necessária ser utilizada nesses jogos a fim do benefício de indivíduos e/ou jogadores e/ou pacientes, que se caracteriza por meio do treinamento dos dados no AM. Não foi identificado o uso de ANS e ASS nestes trabalhos. Apesar do Aprendizado Por Reforço ter representado 8,3% do total de trabalhos selecionados, ele pode se tornar uma alternativa viável nos jogos para medicina.

Foi apresentado na Subseção 2.2.3 as tarefas de AM, sendo subdivisões dos tipos de AM, assim são divididas em classificação e regressão no AS; e agrupamento, sumarização e associação no ANS. Foi incluído o reforço como tarefa no APR. Assim, foi adaptada em relação a pesquisa anterior (SILVA *et al.*, 2019). Na Figura 21 pode-se observar a distribuição das tarefas de AM nos trabalhos selecionados. Dentre as tarefas de AM, seis, que corresponde à 50% dos trabalhos, utilizaram regressão, apresentada nos trabalhos de Chouhan *et al.* (2015), Heller *et al.* (2013), Alchalabi *et al.* (2017), Alchalabi *et al.* (2018), George *et al.* (2011) e Kim *et al.* (2013).

Figura 21 – Tarefa do Aprendizado de Máquina nos Trabalhos.



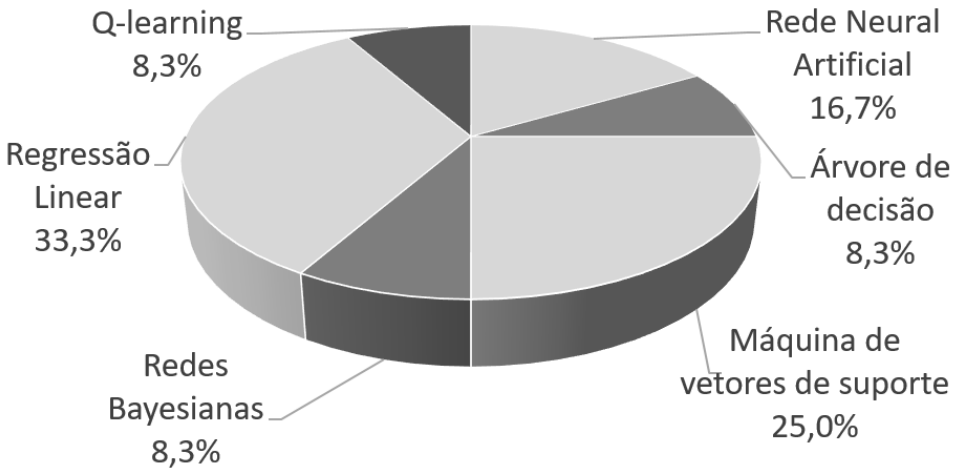
Fonte: Autoria Própria

Cinco, que corresponde a 41,7%, utilizaram a tarefa de classificação dos dados, sendo identificada nos trabalhos de Lima *et al.* (2016), Ahmad *et al.* (2013), Zielke *et al.* (2018), Flores *et al.* (2013) e Hervás *et al.* (2016). E um, que corresponde a 8,3%, utilizou a tarefa de reforço, apresentada no trabalho Perdiz *et al.* (2018).

3.4.2 Quais as Principais Técnicas de Aprendizagem de Máquina nos Jogos para Medicina?

Em relação as técnicas de AM usadas, foram divididas em relação aos principais classificadores de AM utilizadas, essas e outras foram descritas anteriormente na Subseção 2.2.7. Na Figura 22, é apresentada a porcentagem de uso das técnicas nos trabalhos selecionados.

Figura 22 – Técnicas de Aprendizado de Máquina.



Fonte: Autoria Própria

A MVS foi usada em 25% dos trabalhos analisados. Por Alchalabi *et al.* (2017) e Alchalabi *et al.* (2018) ela foi usada com a tarefa de *Regressão Linear*, enquanto Hervás

et al. (2016) utilizaram a MVS com *Regressão logística*, que é outra forma de regressão, entretanto ela assume valores categóricos ou binários, diferente da regressão linear. Isso demonstra a boa combinação da MVS com a Regressão.

Dois dos trabalhos selecionados apresentam o uso de *Rede Neural Artificial*, o que corresponde à 16,6% do total. Essa *Rede Neural Artificial* foi usada por meio da *Rede Elman* por Ahmad *et al.* (2013), e Processamento de Linguagem Natural por Zielke *et al.* (2018).

Apenas um dos trabalhos encontrados apresentou de forma clara o uso de *Redes Bayesianas*, que corresponde à 8,3% do total de trabalhos selecionados. O trabalho apresentado por Flores *et al.* (2013) utilizou as *Redes Bayesianas*, para realizar classificações. O trabalho usou o aprendizado para modelar conhecimento e apoiar o raciocínio sob a modelagem do conhecimento de incerteza, o que se alinha a um dos objetivos das Redes Bayesianas. O trabalho obteve uma boa validação e seguiu o processo da Organização Internacional de Normalização (ISO) / Comissão Eletrotécnica Internacional (IEC) 14598-6.

Apenas um trabalho utilizou a *árvore de decisão*, que corresponde à 8,3% do total de trabalhos selecionados. Lima *et al.* (2016) utilizou o a árvore de decisão por meio do algoritmo *J48* para classificar os sintomas da doença transmitidos pelo paciente virtual em um jogo sério.

Quatro trabalhos, que corresponde à 33,3%, utilizaram *regressão linear* como técnica principal para classificar os dados. Outros dois trabalhos também utilizaram a *regressão linear*, entretanto a técnica utilizada foi a MVS, com intuito de gerar uma regressão linear, por isso não foram consideradas como a técnica utilizada, mas apenas como a tarefa de AM.

O trabalho de Perdiz *et al.* (2018) utilizou o *Q-learning* para intermediar entre possíveis ações do usuário e instruções de atuação do jogo, tentando adaptar sua resposta a do usuário para maximizar os resultados do jogo.

3.4.3 Quais os Principais Jogos Sérios que Utilizam Aprendizagem de Máquina na Área da Medicina?

Para identificação dos principais jogos que usaram aprendizagem de máquina na área da medicina, foram levados em conta alguns critérios de qualidade como *qualis* e número de páginas, além do progresso da pesquisa por meio de outras publicações, validação do jogo, limitações da pesquisa, vantagens e outros pontos importantes da pesquisa. Portanto, chegou-se a conclusão de que os principais jogos que utilizam AM na área da medicina são o FOCUS e SimDeCS (Simulação para Tomada de Decisão no Serviço de Saúde). Na Tabela 10 estão alguns dos dados analisados.

Tabela 10 – Dados dos Jogos.

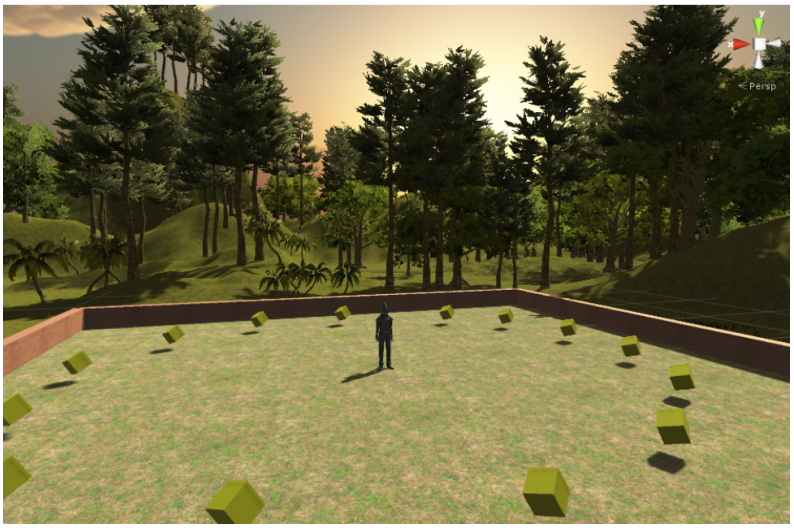
Nome do Jogo	Qualis	Páginas	Limitações	Validação	Vantagens	Publicações
DocTraining	B3	9	Jogo sério está em teste alfa	Não Validado	Treinamento de avaliações clinicas e de diabetes, Visualização das manchas dos pacientes, Sistema multi-agente	1
Não Informado pelo Autor	A2	6	Possível erro no método de Entropia	Usuários	Mostrou deficiências do algoritmo Sample entropy, Melhoria dos níveis de atenção	1
Groundskeeper	B5	9	Brief Rating Scale, versão pai e professor do Conners recrutam um nível mais alto de cuidado e eram mais preditivos.	26 Crianças com e 26 sem TDAH	Identificação de déficits de desempenho executivo e diferenciação baseada em padrões de comportamento entre TDAH, tipo combinado, TDAH, tipo desatento, ansiedade e depressão com boa precisão.	1
Não Informado pelo Autor	B2	6	O desempenho do algoritmo não foi muito encorajador para o exercício de agarrar	4 Cirurgiões Especialistas	Simulador para Treinamento de Cirurgias	1
EVP - Emotive Virtual Patient	B3	8	Estreito campo de visão dos Hardwares utilizados. Qualquer falha no sistema quebra a imersão dos estados real e virtual.	36 Alunos	Disponibiliza de pacientes, mentor, e professor virtual.Permitindo que os professores visualizem e avaliem remotamente o desempenho dos alunos durante a entrevista, Utiliza a rede US Ignite GENI de alta velocidade e baixa latência médica em tempo real	1
FOCUS	A1 e B3	9 e 6	Necessidade de cenários mais interativos. Poucas número de amostras	4 Usuários com TDAH e 5 Usuários Saudáveis	Sistema de detecção de pacientes com TDAH. Precisão de até 96% e 98% na classificação dos dados do EEG para detectar o estado de atenção correto durante o jogo em indivíduos saudáveis e com TDAH respectivamente.	2
SimDeCS	B3	6	Baixa Satisfação média com a confiabilidade, de 45%	24 Pessoas (13 médicos, 05 professores, 05 alunos de graduação e 01 pós-graduando) e ISO / IEC 14598-6	Feedback e orientações sobre as decisões clínicas tomadas. Sistema multi-agente	1
Não Informado pelo Autor	B3	8	Não teve efeito perceptível nos escores dos jogadores nos primeiros testes	5 Usuários	Melhor pontuação em evitar colisão e melhoria das sessões de não-RL para RL entre todos os participantes, maioria dos participantes expressou a capacidade de se adaptar ao paradigma do jogo rapidamente. Abordagem provavelmente pode ser tomada ao projetar jogos controlados por biosignal com pessoas com deficiência e Assistive Living em mente.	1
Não Informado pelo Autor	B1	4	-	10 Participantes	Melhora de atenção e relaxamento com o jogo. Configurações adequadas para interação com videogames simples.	1
ViziCal	A1	8	Não foi encontrada diferença no desempenho entre a aceleração e a posição das articulações	9 Homens	Pode ser usado para projetar jogos de exercícios que geram maiores benefícios à saúde, identificaram um conjunto de características que permitem prever o gasto de energia. Provaram que os acelerômetros preveem gasto energético.	1
Não Informado pelo Autor	B4	8	Pequeno conjunto de treinamento e solução está longe de representar uma solução definitiva para fornecer suporte no diagnóstico de MSC	30 sujeitos e conjunto de dados sintético	Em geral, a classificação das dificuldades de empatia e socialização teve médio desempenho. Desempenho médio na avaliação do algoritmo com dados sintéticos. Participação de psicólogos, terapeutas ocupacionais no projeto	1

Fonte: Autoria Própria

O FOCUS, apresentado por Alchalabi *et al.* (2017) e posteriormente por Alchalabi

et al. (2018), teve o objetivo de treinar e fortalecer a capacidade de atenção dos pacientes com TDAH e detectar seu nível de atenção. Este trabalho teve duas publicações inclusas nessa RSL com *qualis* A1 e B3, e 9 e 6 páginas. Foi validado usando 9 usuários, 4 com TDAH e 5 Saudáveis, com uma precisão de até 96% e 98% na classificação dos dados do EEG para detectar o estado de atenção correto durante o jogo em indivíduos saudáveis e com TDAH, respectivamente. Assim, esse trabalho apresenta um novo método para classificação de dados de EEG e possibilita a identificação de pessoas com TDAH. Na Figura 23, é apresentado o ambiente do jogo FOCUS.

Figura 23 – Ambiente do jogo FOCUS.



Fonte: Alchalcabi, Eddin e Shirmohammadi (2017)

É apresentado, segundo os autores, como o primeiro trabalho que estabeleceu as bases para a integração de um classificador de AM com um jogo sério para detectar o TDAH e como primeiro trabalho na literatura de instrumentação e medição que tentou medir a atenção para detectar o TDAH. Este Jogo é controlado por meio de *EMOTIV EPOC+*[8] . Os dados EEG brutos foram extraídos e registrados durante as sessões de teste usando *scripts Python* executados em segundo plano e os modelos de classificação foram construídos também usando *Python*.

O SimDeCS, apresentado por Flores *et al.* (2013), teve o objetivo de realizar o monitoramento do usuário/estudante de medicina durante o processo de simulação, fornecendo *feedback* e orientações sobre as decisões clínicas tomadas. Esse trabalho tem *qualis* B3, e 6 páginas. Foi validado com 24 pessoas, sendo 13 médicos, 05 professores, 05 alunos de graduação e 01 aluno de pós-graduação, usando o *software* e preenchendo questionário. Também foi avaliado em termos de qualidade técnica e usabilidade, obedecendo à norma brasileira ISO / IEC 14598-6, que recomenda um mínimo de oito avaliadores (ABNT, 2004). O simulador trabalha o desenvolvimento das capacidades técnicas e competência no

[8] https://www.emotiv.com/epoc/

diagnóstico formulado, seguindo o próprio ritmo de aprendizagem do usuário/jogador. Na Figura 24 pode-se observar a interface do jogo SimDeCS.

Figura 24 – Interface do SimDeCS.

Fonte: Flores *et al.* (2013)

O processo de formulação do diagnóstico médico pode ser visto como um conjunto de etapas como: entrevista médica, exame físico, formulação de hipóteses diagnósticas e requisição (ou não) de exames complementares. Várias pessoas se dedicaram a esse propósito, contando com profissionais da área de saúde para modelar o conhecimento específico em Redes Bayesianas, o qual inclui também especialistas em modelagem de área computacional e especialistas na educação, todos trabalhando com táticas pedagógicas para avançar junto com profissionais no desenvolvimento. Quanto ao próprio sistema, está em fase final de desenvolvimento com três redes (dor de cabeça, dispepsia e parasitoses), possibilitando moldar cerca de 80 casos clínicos por professores que se preocupam em delinear cada caso pessoal. Apenas dez casos clínicos foram preparados para os estudantes se exercitarem.

O ViziCal, apresentado por Kim *et al.* (2013), teve o objetivo de prever com precisão o gasto energético de jogar um *exergame*, este trabalho tem *qualis* A1 e 8 páginas. Foi validado com 9 homens (idade média 20,7 (Desvio Padrão = 2,24), peso 74,2 kg (Desvio Padrão = 9,81), IMC (Índice de massa corporal) 23,70 (Desvio Padrão = 1,14),% de gordura 14,41 (Desvio Padrão = 1,93)). O jogo por meio do sensor *Kinect* mapeava a profundidade e extraia a posição 3D de 20 articulações do esqueleto, a 200 quadros por segundo. As articulações incluem centro do quadril, coluna, centro do ombro, cabeça, ombro, cotovelo, punho, mão, quadril, joelho, tornozelo e pé. Na Figura 25, apresenta-se

as posições das articulações do esqueleto de um usuário que está jogando um exergame.

Figura 25 – Posições das articulações do esqueleto de um usuário que está jogando um exergame.

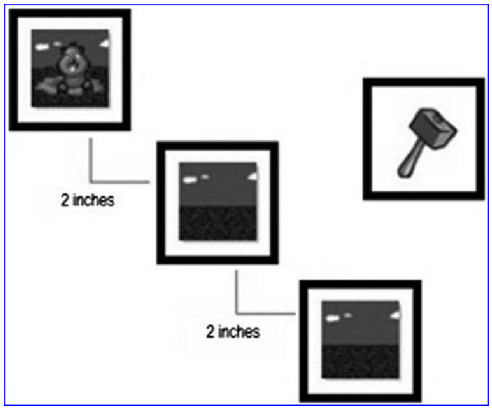
Fonte: Kim *et al.* (2013)

Nesse jogo, os jogadores pontuam destruindo alvos usando gestos na parte superior do corpo, como socos, mas também usando cabeçadas. Eles incluíram gestos com a cabeça, por esse tipo de movimento ser difícil de medir usando acelerômetros, já que eles são tipicamente unidos a cada membro, também foram incluídos saltos para mais gastos energéticos. Um alvo é renderizado pela primeira vez usando um círculo verde, com um raio de 50 *pixels*. O alvo fica verde por 1 segundo antes de ficar amarelo e depois desaparece após 1 segundo. O jogador marca 5 pontos se o alvo for destruído quando verde e 1 quando amarelo, com intuito de motivar os jogadores a destruir os alvos o mais rápido possível. Um alvo é renderizado como uma linha verde, quando cada alvo é destruído com sucesso um som é tocado. Os autores conseguiram identificar um conjunto de características que permitem prever o gasto de energia, além de provar que os acelerômetros preveem gasto energéticos.

O Groundskeeper, apresentado por Heller *et al.* (2013), tem *qualis* B5 e 9 páginas. Foi validado com 52 crianças (50% delas com TDAH). O jogo usa quatro *Sifteo Cubes* e uma placa de posicionamento. Um cubo é sempre usado como um "martelo" para atingir os alvos do jogo. Inicialmente, três cubos são colocados em uma linha reta vertical, cada um desses três cubos tem uma imagem de grama verde e o azul do céu como pano de fundo. Imagens de um coelho, um jardineiro (um homem com um cortador de grama), um *Gopher*[9] ou alguns pássaros pequenos que aparecem em cada tela por 1, 1,5 ou 3 segundos aleatoriamente. Na Figura 26 pode-se observar o funcionamento do Groundskeeper.

[9] *Geomyidae* é uma família da ordem dos roedores, aparentada com os esquilos.

Figura 26 – Funcionamento do Groundskeeper.

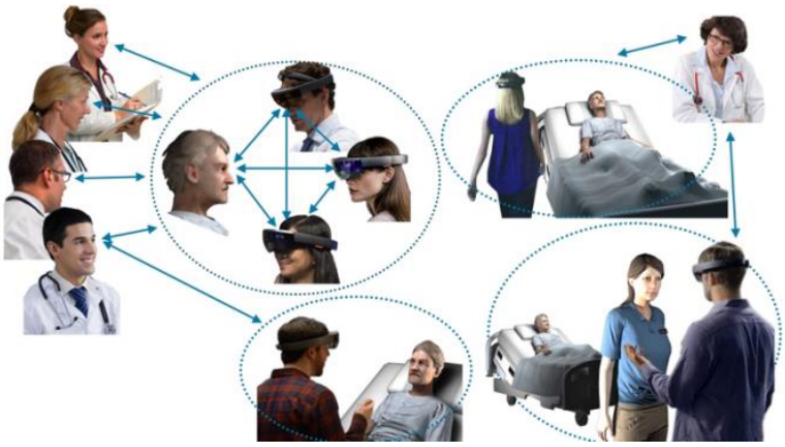


Fonte: Heller *et al.* (2013)

O objetivo do jogo é evitar distratores e esperar que apareça apenas uma imagem de um *Gopher*, momento em que o jogador deve tocar o cubo no *Gopher*, que marca um sucesso com um ruído. Cada uma das 17 sessões de jogo tem 90 segundos de duração, com um intervalo de 20 segundos entre cada sessão. Por meio do jogo foi possível identificar déficits de desempenho executivo como TDAH. O jogo coleta vários dados que poderão ser usados para fornecer pistas adicionais para o diagnóstico de distúrbios cognitivos.

O *Emotive Virtual Patient* (EVP), apresentado por Zielke *et al.* (2018), teve o objetivo de descrever a plataforma de sistema humano emotivo virtual de realidade mista, este trabalho tem *qualis* B3 e 8 páginas. Foi validado com 36 alunos. Esse simulador permite que os estudantes de medicina pratiquem entrevistas com um paciente virtual em uma experiência de realidade aumentada, usando atualmente o *Microsoft HoloLens*. São utilizadas conversas de linguagem natural baseadas em redes neurais complexas que representam as emoções, culturas únicas e padrões de comportamento geral dos pacientes da vida real. Na Figura 27 estão exemplos de entrevistas com pacientes virtuais no EVP.

Figura 27 – Exemplos de entrevistas com pacientes virtuais e reais no EVP.

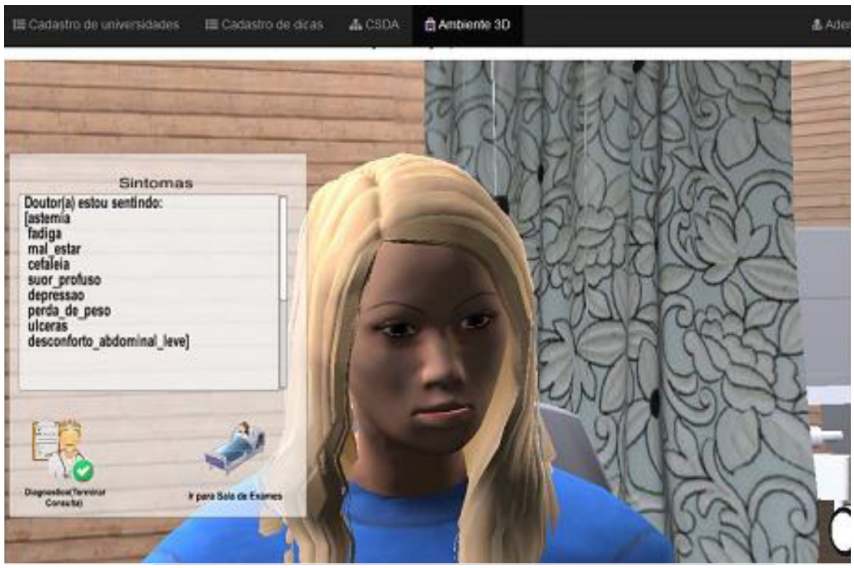


Fonte: Zielke *et al.* (2018)

No processo de aprendizagem e avaliação, é usada a participação do docente e crítica do desempenho clínico virtual do estudante de medicina. Os estudantes de medicina experimentam um cenário de paciente virtual e os professores podem ver remotamente a mesma simulação de realidade aumentada e a linguagem corporal do aluno em tempo real. O EVP também fornece um *feedback* para o aluno, pois além de um professor virtual, há um mentor.

O DocTraining, apresentado por Lima *et al.* (2016), teve o objetivo de apresentar uma ferramenta para auxiliar alunos e professores de medicina como forma de mitigar o crescente número de erros médicos em casos clínicos, este trabalho tem *qualis* B3, 9 páginas e não foi validado, pois se encontrava em teste alfa. Neste jogo os usuários podem interagir com outros usuários por meio de mensagens, além de NPC (*Non Playable Characters*) que oferecem dicas. Ao entrar no hospital universitário o usuário possui um consultório virtual, e nele aparece o paciente que informa o que está sentindo para ser dado uma avaliação clínica geral do paciente. Na Figura 28, é apresentada uma consulta com um paciente virtual no DocTraining.

Figura 28 – Paciente virtual informando os sintomas.



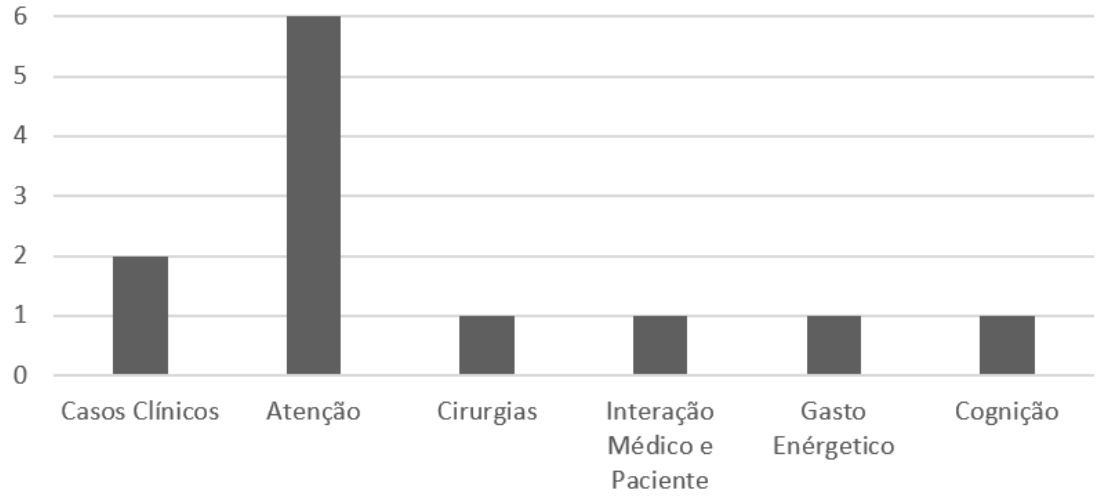
Fonte: Lima (2016)

Após o estudante entrar com o diagnóstico, o jogo verifica se está correto. Se o aluno tiver dificuldades de dar o diagnóstico ele pode verificar o corpo do paciente na sala de exames, onde é possível verificar manchas no corpo e lesões, além de aproximar a câmera para ver melhor as partes do corpo do paciente. Os professores podem inserir novas doenças e dicas no jogo, que são analisadas antes de ficarem disponíveis. O jogo conta com um sistema multiagente.

3.4.4 Quais as Principais Áreas da Medicina em que os Jogos Estão Sendo Aplicados?

Os trabalhos foram classificados quanto as temáticas da medicina mais abordadas, sendo divididos em: Casos clínicos; Atenção; Cirurgias; Interação Médico e Paciente; Gastos Energéticos Corporal; e, Cognição. Na Figura 29 pode-se observar a distribuição dos trabalhos nas temáticas da saúde.

Figura 29 – Áreas da saúde dos jogos.



Fonte: Silva *et al.* (2019)

Seis trabalhos estão relacionados à Atenção. Chouhan *et al.* (2015) trabalharam a atenção e memória por meio de um jogo com intuito de memorizar uma sequência de objetos e, em seguida, selecioná-los corretamente na ordem de sua aparência para as diferentes modalidades de estímulos. O jogo apresentou melhoria nos níveis de atenção dos jogadores. Perdiz *et al.* (2018) focaram em aumentar o nível de atenção dos jogadores por meio do uso de movimentos oculares para controlar um *Card*, apesar de apresentar uma melhora relativa na atenção, não houve efeito perceptivo nos *scores* dos jogadores.

George *et al.* (2011) tiveram o objetivo de melhorar a atenção e relaxamento. O trabalho foi validado com dez usuários e usou dados de EEG para medir a atenção e relaxamento do usuário, apresentando melhora de atenção e relaxamento dos usuários. Esse trabalho ainda apresentou configurações adequadas para interação com *videogames* simples e utilizou o *software* de aprendizado *Matlab* para uso da Regressão Linear.

Desses seis trabalhos, três estão ligados ao TDAH. O jogo Groundskeeper, apresentado por Heller *et al.* (2013), objetivou construir modelos para prever o funcionamento executivo de desordens, com dados coletados no jogo dos sujeitos de teste. Foram abordados TDAH, transtorno depressivo, transtornos de ansiedade, transtorno desafiador de oposição, transtorno do pânico ou transtornos alimentares. Esse jogo usa *Sifteo Cubes* e coleta dados de respostas dos usuários durante o jogo. Usou o AS, através da Regressão Linear, e usou o

software Weka para o aprendizado. O trabalho identificou déficits de desempenho executivo e diferenciação baseada em padrões de comportamento entre TDAH, tipo combinado, TDAH, tipo desatento, ansiedade e depressão. O jogo FOCUS também apresentou em seus dois trabalhos identificação de atenção e TDAH.

Dois dos trabalhos selecionados apresentaram casos clínicos. O trabalho apresentado por Lima *et al.* (2016) apresentou o DocTraining, um jogo para auxiliar alunos e professores de medicina como forma de mitigar o crescente número de erros médicos em casos clínicos. Esse trabalho usou aprendizado na identificação das doenças de casos clínicos e diabetes no jogo, e usou o *software Weka* além de usar sistema multiagente. Apesar da variedade de utilidades abordadas no trabalho, o jogo encontra-se em teste alfa. O trabalho apresentado por Flores *et al.* (2013) focou em Decisões Clínicas e trabalhou as capacidades técnicas do usuário.

Baseado nesses resultados, pode-se identificar que as principais áreas da saúde com jogos que utilizam AM é a área de Atenção, por ser tratado em metade dos trabalhos encontrados com foco em Transtorno do Déficit de Atenção e Hiperatividade; e, os Casos Clínicos, apresentados em dois trabalhos. Além das áreas de Cirurgias, Interação Médico e Paciente, Gasto Energético Corporal e Cognição.

3.4.5 Como o Aprendizado de Máquina está Sendo Utilizado nos Jogos para Medicina?

Por meio das respostas das questões de pesquisa anteriores, foi possível verificar o estado da arte do Aprendizado de Máquina em Jogos para Medicina. Foi possível notar que a maioria dos trabalhos abordam principalmente o Aprendizado Supervisionado, além do Aprendizado Por Reforço, que também pode ser uma tecnologia viável nos jogos para medicina.

Isso pode estar diretamente relacionado a confiança dos dados, pois um dos desafios do AM se trata da responsabilidade, e os médicos precisam se sentir confortáveis com os riscos dos erros médicos, que não devem ser maior do que os que já ocorrem (DEO, 2015). Como nesses casos os dados necessitam de acompanhamento médico o uso do treinamento por meio do AS se torna ideal, pois é simulado um especialista.

O AS nestes trabalhos utilizam principalmente a *Regressão Linear*, *Máquina de Vetores de Suporte*, *Rede Neural Artificial*, *Q-learning* e as *Redes Bayesianas*, que são bastante usadas na área da medicina, e que pode crescer com os anos. Essa técnicas estão sendo utilizadas com as tarefas de regressão, e também a classificação. Isso pode proporcionar uma boa acurácia das pesquisas (SILVA *et al.*, 2019).

Esses tipos de AM, técnicas e tarefas estão sendo bastantes aplicados a Memória, pela facilidade encontrada atualmente em se monitorar os dados de EEG dos indivíduos

(SILVA *et al.*, 2019). Assim, os jogos podem trabalhar os transtornos, concentração, dentre outras coisas relacionadas a memória. Também estão sendo aplicadas a casos clínicos, onde uma boa proposta para isso são as *Redes Bayesianas* por proporcionarem a modelagem de conhecimento e raciocínio, que podem ser aplicadas nas definições de doenças e classificação do AM (FLORES *et al.*, 2013).

3.5 CONCLUSÕES E LIMITAÇÕES

Neste Capítulo foram apresentados o planejamento, a condução e os resultados de uma RSL sobre o AM nos jogos para medicina, publicados nos últimos 10 anos, no âmbito internacional. Os trabalhos foram pesquisados nas bases de dados da IEEE, *ACM Digital Library*, *Science Direct* e *Scopus*. A busca pelos trabalhos resultou na pré-seleção de 40 trabalhos, dentre os quais 12 foram incluídos para a extração de dados.

Portanto, foi possível observar que o AM nos jogos para medicina possui aplicação no AS, através do uso de Máquina de Vetores de Suporte, *Rede Neural Artificial*, *Redes Bayesianas*, *Árvore de Decisão* e *Regressão Linear*, com foco na tarefa do AM de Regressão e Classificação. Esses jogos focaram principalmente na área da Memória e Casos clínicos. Foi perceptível que mesmo com o avanço dessa tecnologia, ainda foi pouco explorada, pois com base na leitura dos trabalhos foi observado que ainda existem várias áreas da medicina que foram pouco exploradas e poucas técnicas de AM utilizadas em relação a diversidade existente.

Por meio dos resultados encontrados, foi possível mapear as técnicas utilizadas, os principais jogos, o principal tipo de AM utilizado e as principais áreas abordadas. Porém, apenas uma pequena parcela desses trabalhos fez o uso do AM nos jogos para medicina, o que se tornou uma limitação deste trabalho. Além disso, outras limitações foram: o não uso da base de dados da PubMed[10]; os poucos sinônimos utilizados nas *strings* de busca; e um dos critérios de qualidade ter utilizado o *qualis* ao invés de fator de impacto.

Em resumo, os resultados apresentados nesta RSL apresentam as seguintes contribuições: fornecem uma visão geral do AM; fornecem características (requisitos) usadas nos trabalhos selecionados; e, identificam técnicas que podem ser utilizadas em novos jogos.

[10] https://www.ncbi.nlm.nih.gov/pubmed/

4 ÁREA ESCOLHIDA PARA O ESTUDO E GERAÇÃO DE MÚLTIPLOS CLASSIFICADORES

Neste Capítulo, são apresentados detalhes dos dados utilizados pelo aprendizado de máquina no módulo inteligente e da geração dos MLCs, sendo dividido em base de dados, na qual são descritos alguns detalhes dos atributos; em seguida, o pré-processamento aplicado nos dados, em que foram utilizadas técnicas de limpeza, verificação de atributos repetidos e incremento da base de dados; e, por fim, o protótipo de geração dos múltiplos classificadores, em que são vistos os classificadores aplicados neste estudo e os meta-classificadores.

4.1 BASE DE DADOS

As bases de dados deste trabalho contêm informações relevantes na área da medicina/saúde sobre doenças, esses dados são relevantes, pois serão classificados pelo AM e disponibilizados no jogo sério. Foram utilizadas bases de dados sobre casos clínicos, diabetes e presença de doença cardíaca em pacientes, para ser focado em assuntos que alunos de medicina mais possuem deficiência de ensino de acordo com o CREMESP.

A base de dados de casos clínicos foi desenvolvida inicialmente por Lima (2016) baseado no guia de bolso sobre Doenças Infecciosas e Parasitárias do Ministério da Saúde (2010) e do *site Minha Vida*[11]. Vale ressaltar que esses dados foram validados por SOUSA NETO *et al.* (2018) usando os próprios estudantes e professores especialistas, alcançando boa concordância na veracidade dos dados entre eles.

Esses dados são organizados no formato de *DataFrame*, na qual todas as colunas são sintomas, e assumem valores booleanos que representam verdadeiro e false, os sintomas são ardência ao urinar, febre alta, olhos vermelhos, vômitos, dor de cabeça, fraqueza muscular, tontura, dor no corpo, cansaço, náusea, perda de apetite, dentre outros. Entretanto a última coluna é classe, e assume valores que indicam o nome da doença como por exemplo, Actinomicose, Amebíase, Anemia, Cancro mole, Catapora, Conjuntivite, Câncer de mama, Litíase urinária, Malária, Mão-pé-boca, Uretrite, dentre outras. Assim, totalizando 196 sintomas, 41 doenças e 201 casos clínicos, na qual uma mesma doença pode apresentar sintomas distintos.

A base de dados de diabetes é do Instituto Nacional de Diabetes e Doenças

[11] http://www.minhavida.com.br/

Digestivas e Renais, que contém dados de pacientes reais com os atributos de: número de vezes em que ficou grávida, concentração de glicose no plasma, pressão arterial diastólica, tríceps espessura da dobra da pele, insulina sérica de 2 horas, índice de massa corporal, função pedigree do diabetes e idade. Esses pacientes são classificados em saudáveis ou doentes em relação à diabetes, totalizando oito atributos e 768 pacientes.

A base de dados de doença cardíaca é do Instituto Húngaro de Cardiologia, Hospitais Universitários Zurique e Basileia da Suíça, Centro Médico *Long Beach* e Fundação da Clínica *Cleveland*. Ela contém os seguintes atributos dos pacientes: idade, sexo, tipo de dor no peito, pressão arterial em repouso, níveis séricos de colesterol (mg/dl), açúcar no sangue em jejum acima de 120 mg, resultados eletrocardiográficos em repouso, frequência cardíaca máxima alcançada, angina induzida por exercício, depressão do ST no eletrocardiograma induzido pelo exercício relativo ao descanso, a inclinação do segmento ST do pico do exercício, número de vasos principais (0-3) coloridos por *flourosopy* e thal (normal, defeito fixo, defeito reversível). Sendo os pacientes classificados em saudáveis ou doentes em relação à presença de doença cardíaca. Totalizando 13 atributos e 303 pacientes. As bases de dados de diabetes e doença cardíaca se encontram disponíveis no *site* da *Kaggle*[12].

4.2 PRÉ-PROCESSAMENTO

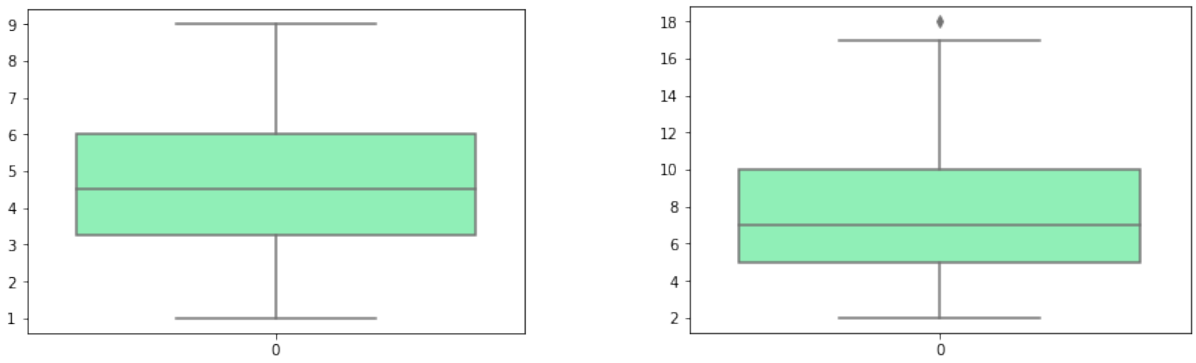
O pré-processamento dos dados foi feito por meio de uma análise estatística básica, verificações de redundância, e valores ausentes nos dados. Isso é importante pois antes de serem aplicadas as técnicas de AM, é necessário que os dados estivessem pré-processados, com intuito de haver uma melhor eficácia dos modelos de classificação.

A base de dados de casos clínicos foi ampliada sendo adicionadas 34 novos casos clínicos das mesmas fontes de dados anterior. Estes dados foram normalizados de acordo com a 1ª, 2ª e 3ª formas normais (MACHADO, 2018) e deixaram de ser organizados em *DataFrame* e passaram a ser organizado no formato de entidade e relacionamento. Assim, uma doença possui vários sintomas e um sintoma está em várias doenças.

Nestes dados não havia valores ausentes, mas foram verificadas redundância de sintomas que apresentavam nomes distintos para mesma representação. Como solução um dos sintomas foi excluído e em seu lugar foi adicionado o relacionamento com o sintoma não excluído. No fim, essa base de dados possui um total de 213 sintomas, 50 doenças e 235 casos clínicos, na qual uma mesma doença pode apresentar sintomas distintos como acontece na vida real. Em média, cada caso clínico apresenta 4,5 exemplos, entretanto há um que apresenta um exemplo e outros que apresentam nove exemplos. Cada caso clínico apresenta em média 7,6 sintomas, sendo a maioria deles na distribuição de cinco a dez. Na Figura 30 pode ser observado o *boxplot* desses dados.

[12] https://www.kaggle.com

Figura 30 – Detalhes base de dados sobre Casos Clínicos



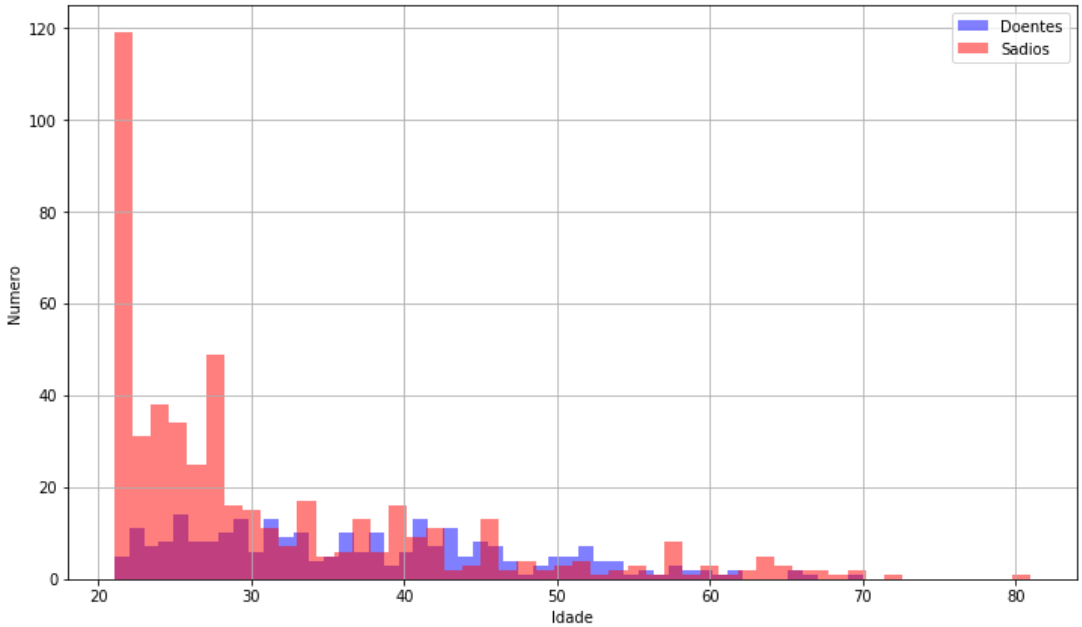
(a) Quantidade Exemplos

(b) Quantidade de Sintomas

Fonte: Autoria Própria

Na base de dados de diabetes não foi encontrado nenhum atributo repetido e nenhum valor ausente. Ela possui 268 pacientes doentes e 500 saudáveis. É possível observar que levando em consideração idade em relação ao número de doentes, quanto maior a idade mais fácil estar doente, pois os jovens são saudáveis. Na figura 31 é apresentada essa distribuição. Assim, essa boa distribuição auxilia no aprendizado de máquina.

Figura 31 – Detalhes Base de dados sobre diabetes

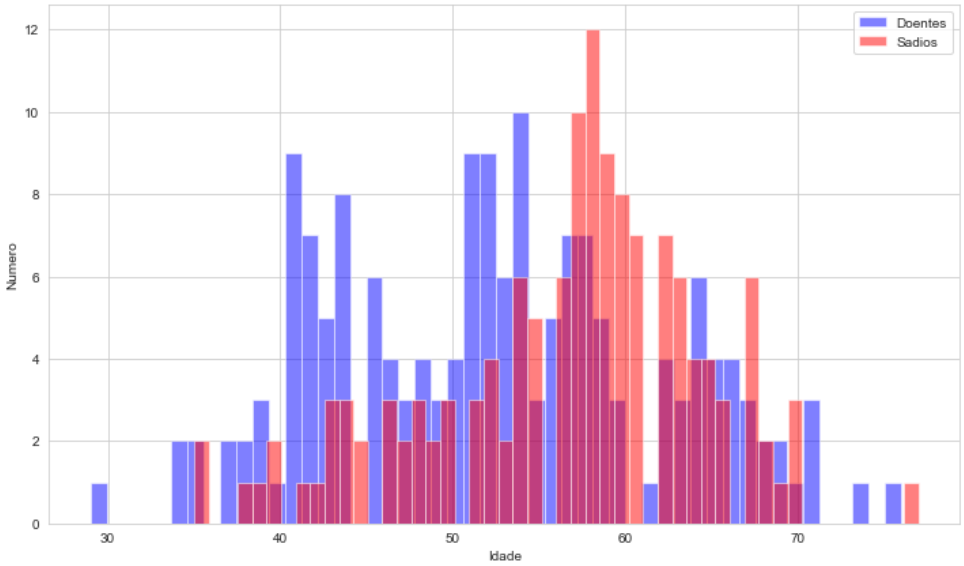


Fonte: Autoria Própria

Na base de dados de doença cardíaca também não foi encontrado nenhum atributo repetido e nenhum valor ausente. Ela possui 138 pacientes doentes e 165 saudáveis, 96 mulheres e 207 homens distribuídos entre idades de 29 à 77 anos. É possível observar que a maioria dos casos de doença cardíaca acontece em jovens, o que faz com que um

diagnóstico precoce salve uma vida inteira. Na Figura 32 é apresentada essa distribuição. Assim, essa boa distribuição auxilia no aprendizado de máquina.

Figura 32 – Detalhes Base de dados sobre doença cardíaca



Fonte: Autoria Própria

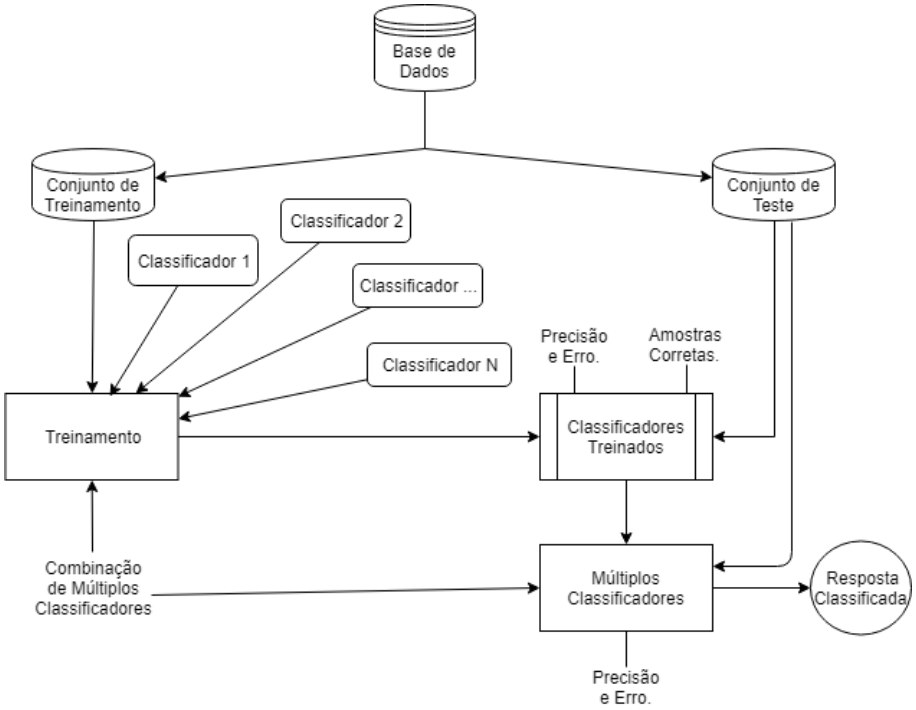
4.3 GERAÇÃO DE MÚLTIPLOS CLASSIFICADORES

O protótipo da arquitetura de geração de MLCs do AM é dividido em quatro partes. A Base de Dados, o Treinamento, os Classificadores Treinados e os MLCs. Esse protótipo pode ser observado na Figura 33.

Inicialmente, a Base de Dados foi dividida em conjuntos de treinamento e de teste, com 20% e 80% respectivamente. O conjunto de treinamento serviu para realizar inicialmente o treinamento dos classificadores. Enquanto o conjunto de teste serviu para realizar a validação e descobrir a medida e erro dos classificadores testados e dos MLCs. Vale ressaltar que essa divisão de dados foi feita inicialmente para validar utilizando a métrica de precisão e que novamente foi aplicada usando a métrica de medida-F usando a DP de 60% para treino e 40% para teste, ambas calculadas pela ferramenta *scikit-learn*, sendo a precisão *precision weighted avg*, e a medida-F *f1-score weighted avg*.

No Treinamento, foi utilizado o conjunto de treinamento em cada um dos classificadores utilizados. Esses classificadores podem variar e, com essa variação, podem chegar ao bom desempenho, dependendo dos dados e da sua combinação. Os classificadores de AM utilizadas foram *Regressão Logística* (RL), *K-Nearest Neighbor* (KNN), MVS, *Árvore de Decisão* (AD), *Rede Neural Artificial* usando *Multilayer Perceptron* (MLP) (PAL; MITRA, 1992) e *Naive Bayes* (NB), todas disponíveis no Scikit-learn (2019). Essas técnicas são bastante utilizadas na literatura, além de que, na RSL mencionada no Capítulo 3, foram

Figura 33 – Arquitetura da Geração de Múltiplos Classificadores



Fonte: Autoria Própria

encontradas algumas dessas técnicas, que estão sendo utilizadas atualmente nos jogos sérios para medicina.

Nos Classificadores Treinados, o conjunto de teste foi submetido a cada classificador e foram verificados seus resultados. Ao término, haviam a precisão, erro e medida-F de cada classificador, além de cada amostra de dados do conjunto de teste que foi classificada de forma correta ou errada por eles.

Nos MLCs, foram realizadas as combinações dos melhores classificadores, tomando como base a taxa de acerto de cada um e as amostras de teste classificadas corretamente para que fossem selecionados modelos que se complementassem o máximo possível. A resposta da classificação dependeu dos classificadores que foram selecionados e do método mais eficiente para a combinação.

A combinação de cada MLC se deu por meio: do Voto Majoritário, na qual a resposta final foi aquela que teve maior quantidade de votos; e da Confiança, em que foram realizadas as medidas previstas de cada classe ser a correta pelos classificadores e a classe escolhida foi a que tinha a maior média de probabilidade. Foram realizadas diversas combinações dos classificadores para se obter uma boa medida-F. Portanto, as etapas de Treinamento, Classificadores Treinados e MLCs foram repetidas. Em cada MLC gerado foi verificada a medida-F e erro, por meio do conjunto de teste.

Além de serem formados MLCs seguindo a arquitetura de geração de MLCs apresen-

tada anteriormente, também foram aplicados meta-classificadores capazes de realizar impulso para estabilizar classificadores e até melhorar os resultados. Esses meta-classificadores utilizam alguns indutores base e gera diversos classificadores alternando os parâmetros do classificador base e agregando os resultados de cada um para impulsiona-los. Os meta-classificadores *ensemble* utilizados foram *Floresta Aleatória* (FA) por meio do algoritmo *RandomForestClassifier* (SCIKIT-LEARN, 2019); o *AdaBoost-SAMME* (HASTIE *et al.*, 2009) que utiliza a técnica de *Boosting*; e o *Bagging* por meio do algoritmo de *BaggingClassifier* (SCIKIT-LEARN, 2019).

Por meio disso foi possível identificar a melhor forma de classificar os dados apresentados neste Capítulo, ou seja, quais os MLCs gerados que vão atuar no módulo inteligente classificando cada uma das suas respectivas bases de dados. Os resultados dos testes desses classificadores e MLCs são descritos no capítulo 6. A seguir é detalhado o módulo inteligente do DocTraining, que contem esse sistema de aprendizado de máquina.

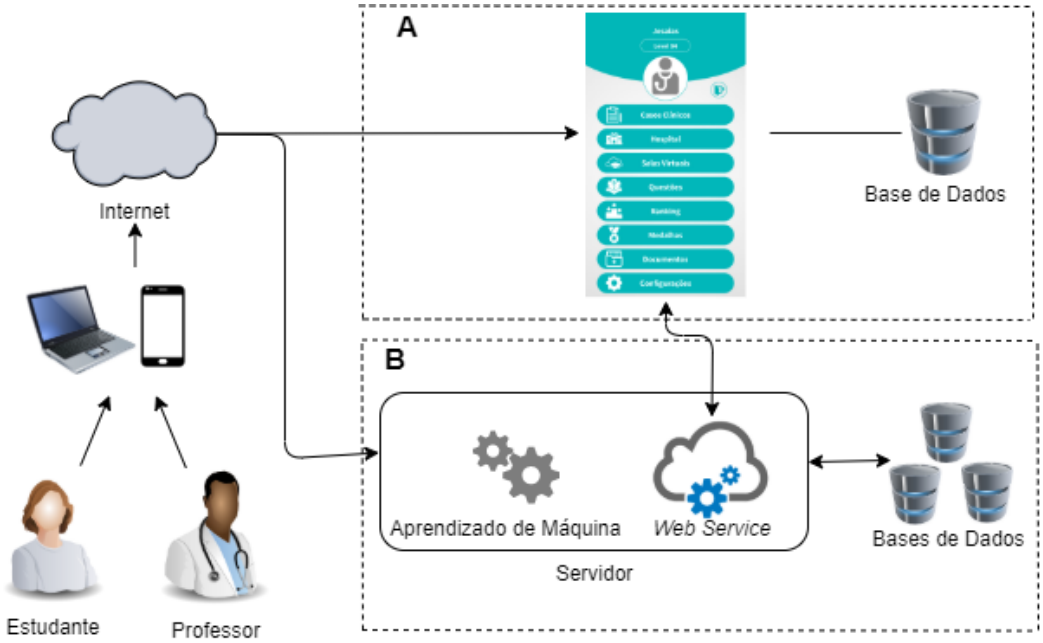
5 MÓDULO INTELIGENTE DO DOCTRAINING

Neste Capítulo, é apresentado o módulo inteligente do DocTraining, sendo dividido em visão geral do DocTraining; Sistema on-line para gerenciar os dados; *Web Service*; Sistema de Aprendizado de Máquina; e, em seguida, é apresentado brevemente o Jogo *mobile*.

5.1 VISÃO GERAL

O desenvolvimento do DocTraining é dividido em dois projetos separados, sendo eles projeto **A** e **B**, como apresentado na Figura 34. O projeto **A** corresponde ao jogo *mobile* 2D com toda a estrutura necessária para interação do jogador com os casos clínicos e outros elementos de aprendizagem e jogabilidade, na qual o estudante de medicina pode acessar via *smartphones* com acesso à internet e consultar os pacientes. Os dados do jogo *mobile* são das amostras de doenças reais, que são classificados por meio dos modelos de AM.

Figura 34 – Arquitetura Geral do DocTraining



Fonte: Autoria Própria

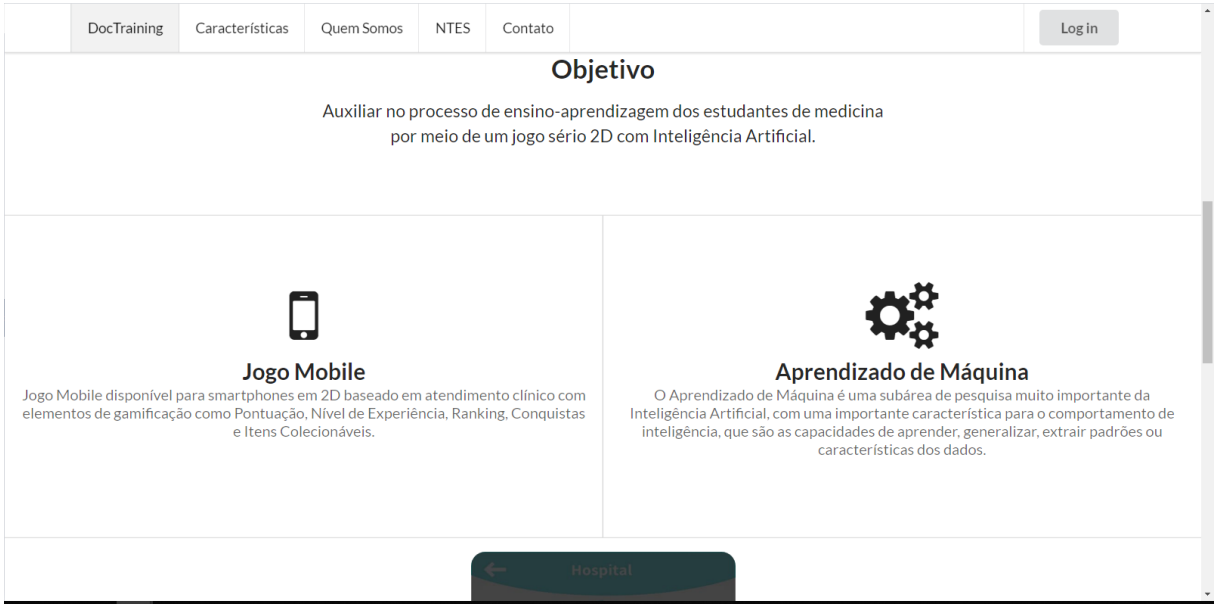
O jogo *mobile* possui funcionalidades como: casos clínicos por níveis, sistema de *ranking*, conquistas, curiosidades na área médica e salas customizadas. O projeto **B** corresponde à classificação de dados por meio dos algoritmos de AM e de MLCs. Estes

projetos são agregados utilizando um *web service*, que fornece a integração dos dois projetos, sem considerar a plataforma de desenvolvimento ou o ambiente operacional, tornando-os independentes (Fu; Peng; Hu, 2015). Também há um *web site* para divulgação e gerenciamento dos dados para o jogo *mobile*. Neste trabalho são apresentados os resultados do projeto **B**.

5.2 SISTEMA ON-LINE PARA GERENCIAMENTO DE DADOS

O projeto conta com um *site*[13] que pode ser acessado por qualquer computador com acesso à internet, assim não será limitado a nenhuma rede. O *site* tem o objetivo de divulgar quem é a equipe desenvolvedora, os serviços prestados pelo sistema, formas de entrar em contato com os pesquisadores e *link* de *download* do jogo *mobile*. Vale ressaltar que todo o *site* possui uma *interface responsiva* que se adapta de acordo com o computador ou dispositivo móvel do usuário. Na Figura 35 é visto o *site* de divulgação.

Figura 35 – *Site* de divulgação do projeto



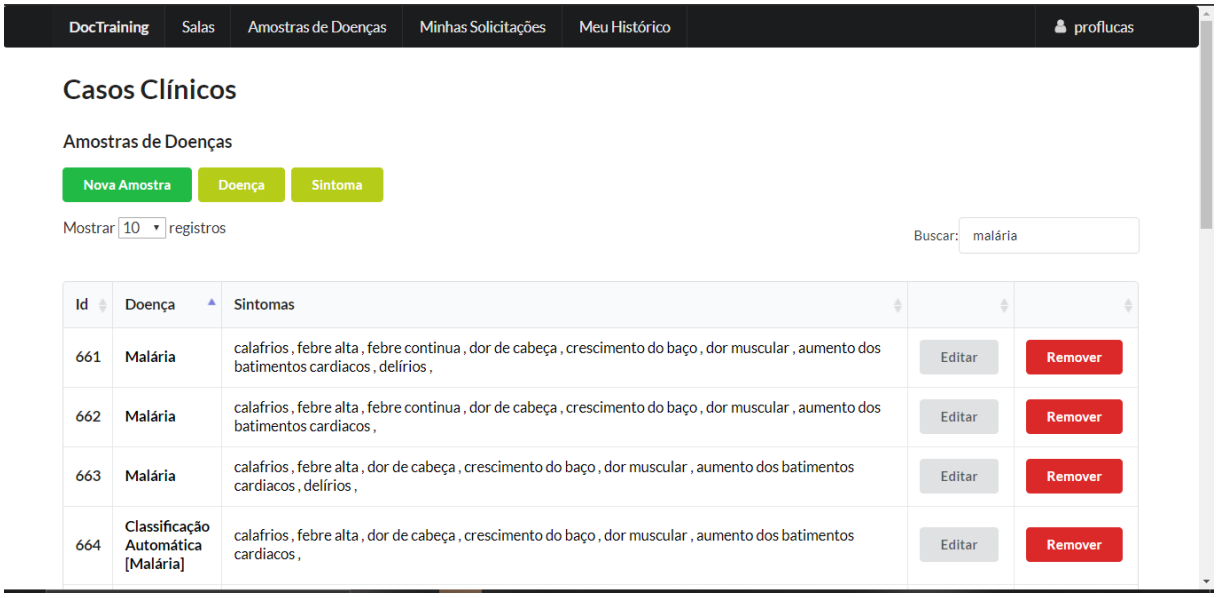
Fonte: Autoria Própria

Ao realizar *login* é possível que os professores ou especialistas da área gerenciem as amostras de doenças e pacientes virtuais. Eles podem cadastrar novas doenças e novos sintomas no módulo. Também é possível inserir novas amostras, sendo obrigatório informar pelo menos os sintomas. Caso a amostra já seja classificada, deve também inserir a doença. As amostras que não possuem rótulos não servem de entrada/conhecimento para o AM, mas serão classificadas por meio do AM, tomando como base todas as amostras rotuladas. Assim, será possível manter o módulo inteligente sempre atualizado e escalável, pois há

[13] https://doctraining.herokuapp.com/

possibilidade de crescer, o que é desejável a todo sistema/aplicação. Na Figura 36, pode ser observada a *interface* das amostras de dados na visão dos professores.

Figura 36 – Tela de gerenciamento de dados do DocTraining



Fonte: Autoria Própria

Vale ressaltar que o sistema conta com dois tipos de visões, sendo uma para administradores, na qual é possível visualizar todas as solicitações de alterações das amostras de doenças, todos os *logs* e os usuários cadastrados no *site* (podendo desativar o *login* dos usuários); e a visão dos professores/especialistas, na qual não podem ver os usuários cadastrados. As visualizações das solicitações de alterações de amostras de doenças e *logs* são visíveis quando relacionadas ao professor/especialista que estiver conectado, isso pode ser observado na Figura 37.

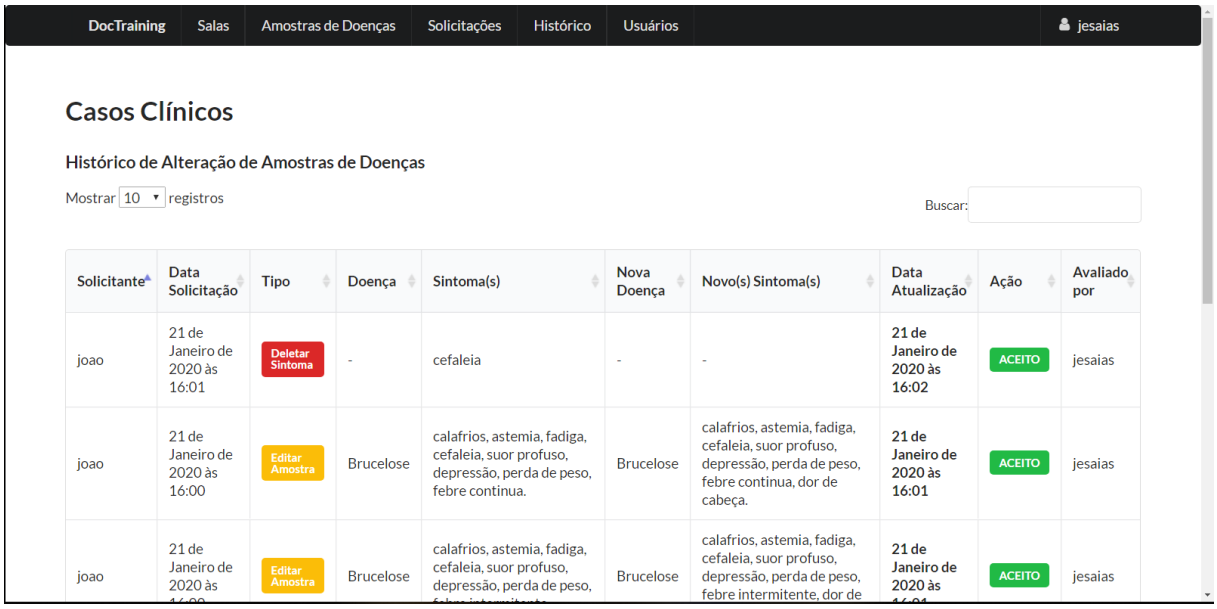
Apesar de apenas os professores e especialistas serem capazes de gerenciar os dados, eles não poderão inserir, editar ou deletar os dados diretamente, pois esses dados servem de conhecimento para o AM. Assim, inserir dados repetidos e/ou editar, e/ou apagar informações que se tornem errôneas não se torna proveitoso para aplicação. Dessa forma, esses dados antes de serem alterados passam pela verificação e permissão de um dos administradores do DocTraining, que podem aceitar ou recusar cada umas das solicitações feitas pelos professores ou especialistas.

Após avaliação das solicitações de alterações dos dados, todo histórico dos pedidos é armazenado em uma tabela de *log*. Assim, será possível saber qual usuário solicitou, qual administrador avaliou o pedido, quando uma informação foi alterada ou solicitada para alterar e qual era a informação da época, tornando as informações úteis para auditoria ou relatórios para obter o cenário da época. Isso pode ser observado na Figura 38, na qual é visto esse histórico na visão de um administrador.

Figura 37 – Tela de solicitações de alteração de amostras

Fonte: Autoria Própria

Figura 38 – Histórico de alterações



Fonte: Autoria Própria

Estão disponíveis também salas virtuais, na qual o professor consegue adicionar novas salas e novas perguntas em cada uma das salas, como observado na Figura 39. Assim, no jogo *mobile*, os alunos poderão adentrar nas salas, responder as perguntas e adquirir conhecimento de algo específico. Os conteúdos ficam a critério do professor. Os administradores do DocTraining são capazes de entrar nessas salas, por meio do módulo inteligente, e o ver o conteúdo. Por fim, os administradores têm o poder de deletar uma sala, caso observem que o conteúdo apresentado não seja relevante.

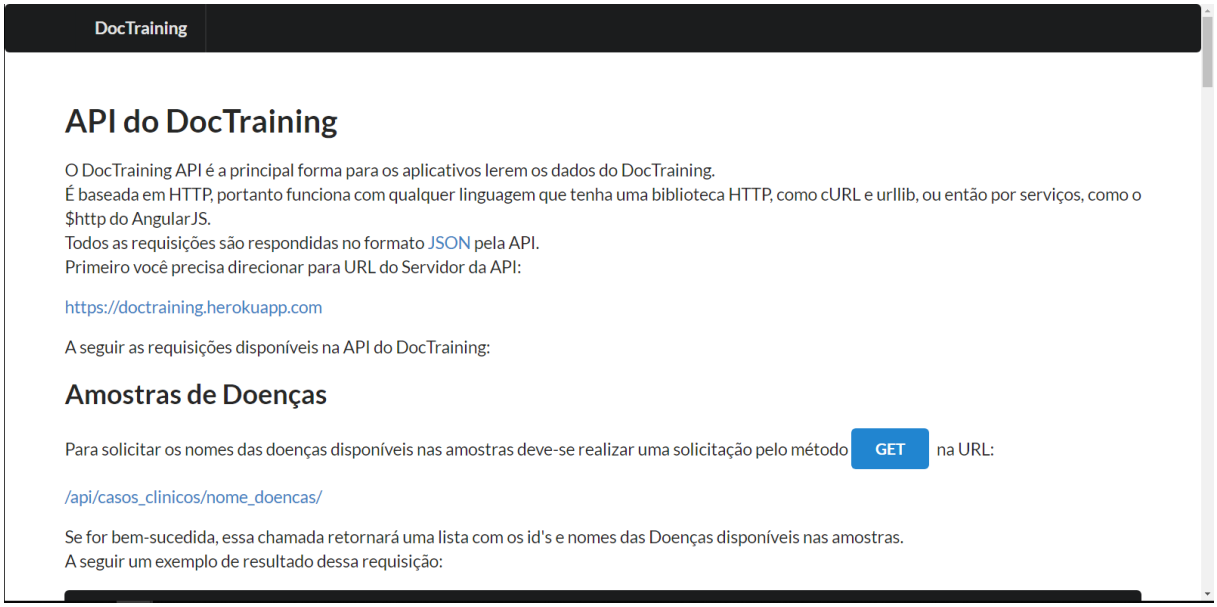
Figura 39 – Salas Virtuais

Fonte: Autoria Própria

5.3 *WEB SERVICE* DO DOCTRAINING

Foi desenvolvido um *Web Service* do DocTraining para que fosse feita a comunicação entres os dois projetos. Esse *Web Service* foi documentado[14] para facilitar o desenvolvimento do jogo *mobile*. Assim, quando for desenvolvido uma nova aplicação para o DocTraining será facilitado por conta do *Web Service* e da documentação disponível. Na Figura 40 é possível visualizar parte dessa documentação.

Figura 40 – *Web Service*



Fonte: Autoria Própria

[14] https://doctraining.herokuapp.com/api/

O *Web Service* do DocTraining conta atualmente com 10 requisições, sendo elas: todas as doenças, todos os sintomas, todos os casos clínicos, um caso clínico, todos os casos de diabetes, um caso de diabete, todos os casos de doença cardíaca, um caso de doença cardíaca, todas as salas virtuais e perguntas de uma determinada sala virtual. Todos eles se tratam de métodos *GET*, ou seja, o jogo *mobile* faz uma solicitação por meio do método *GET* e o *Web Service* retorna os dados correspondentes que estão presentes na bases de dados e classificados pelo AM. Na Figura 41, é apresentada uma requisição de um caso clínico que, após processamento pelo AM, envia as informações para o jogo *mobile* no formato de *JavaScript Object Notation* (JSON).

Figura 41 – Exemplo de resposta em JSON do *Web Service*


```
{
      "id":163,
      "sintomas": [
          "ardência nos olhos",
          "sensação de algo nos olhos",
          "lacrimação",
          "secreção transparente nos olhos",
          "olhos vermelhos"
      ],
      "doenca": "CONJUNTIVITE"
  }
```


Fonte: Autoria Própria

É possível observar que os dados são apresentados por meio de *id*, *doença* e lista de *sintomas*. Essa requisição sempre gera uma amostra aleatória, pois no jogo não é interessante apresentar um mesmo problema para o estudante de medicina, mas sim selecionar aleatoriamente, como acontece em casos reais. Entretanto, como no jogo *mobile* o professor pode gerenciar o ambiente de aprendizado dos seus alunos, as demais requisições de casos clínicos se tornam interessantes, pois é possível gerenciar o jogo *mobile* para selecionar quais doenças serão apresentadas aos alunos como motivo de aprendizado.

5.4 SISTEMA DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina atua internamente no módulo inteligente. Ele é treinado e em seguida são feitos testes automaticamente, na qual os administradores têm acesso a quantidade de acertos e erros que os modelos estão tendo nos testes. Em seguida, eles estão prontos para realizar a classificação dos dados de cada uma das bases. Cada base de dados possui um modelo de MLC responsável por isso. Os MLCs foram escolhidos de acordo com os testes realizados e descritos na Seção 6.1.

Assim, esses modelos de AM gerados podem ser retreinados automaticamente e atualizar a base, caso necessário. A classificação dos dados no módulo inteligente é feita

entre as 02h e 04h, e entre 12h e 12:30h, sendo horários que, possivelmente, não haverá pico de uso do jogo *mobile* que consome o *web service* e tão pouco de vários usuários conectados no módulo inteligente.

5.5 JOGO *MOBILE*

Apesar de não ser o foco deste trabalho apresentamos aqui alguns breves resultados do projeto **A** para entendimento desta pesquisa. O jogo sério *mobile* foi desenvolvido ao mesmo tempo e coincidindo com o projeto **B**, assim a equipe de desenvolvimento definia e implementava as várias versões e protótipos funcionais.

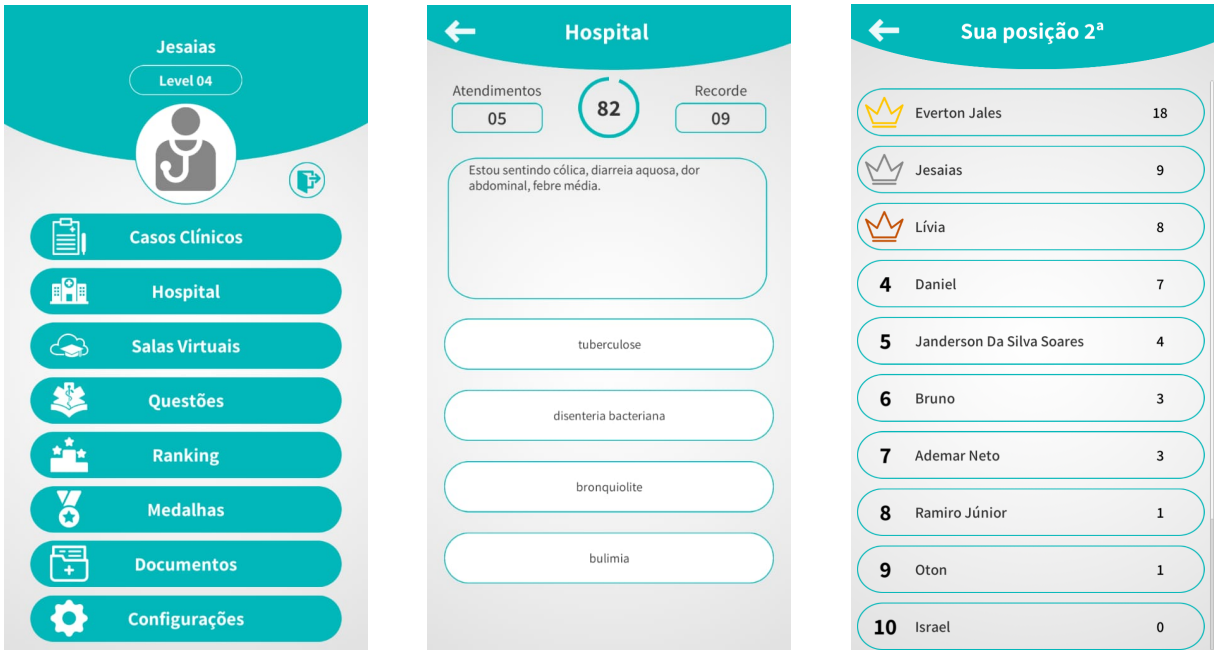
Dessa forma, foi possível testar na prática a funcionalidade de alguns requisitos com a equipe responsável pelos testes e validação para obter melhores refinamentos tanto para o jogo *mobile* como para o módulo inteligente do DocTraining. Portanto, foram incorporados elementos que estimulassem os aspectos motivacionais, sendo algo bem aplicado quando se trata de educação. Seguem os detalhes dos Cinco elementos do jogo *mobile* presentes na versão final.

- Pontuação: pontuação por precisão, que o jogador recebe ao dar um diagnóstico corretamente a um paciente. A pontuação possui um valor tabelado para cada ação do jogador.
- Nível de experiência: serve para identificar jogadores veteranos e jogadores novatos.
- Ranking: Serve para atribuir a cada jogador uma classificação quanto à sua dedicação ao jogo. Inicialmente o *mobile* foi desenvolvido com 2 tipos de ranking, sendo eles, o ranking dos pontos de precisão e ranking de pontos de conhecimento.
- Conquistas: Esse tipo de recurso impulsiona o jogador a completar tarefas dentro do jogo *mobile* para poder ganhar determinada conquista.
- Documentos: na medida que o jogador vai passando de nível, respondendo e acertando algumas questões, documentos contendo algumas informações sobre determinada área da medicina ou curiosidades vão sendo ganhos.

Por fim, tem-se os resultados da implementação final do jogo *mobile* com tudo que foi apresentado anteriormente e já conectado com o modulo inteligente. Na Figura 42 é visto algumas telas da versão final do jogo *mobile*.

Na Figura 42 (a) é apresentado o menu principal do jogo *mobile*, onde o jogador pode navegar para outras seções do jogo*mobile*, como as configurações, visualizar os documentos adquiridos, responder questões, entrar no hospital médico, casos clínicos específicos, salas virtuais, entre outras. Ao entrar no hospital médico o estudante tem pode

Figura 42 – DocTraining Mobile



(a) Menu Principal (b) Hospital Médico (c) Ranking de Atendimentos

Fonte: Autoria Própria

atender diversos pacientes específicos de acordo com os sintomas apresentados pelo mesmo e indicar qual o provável diagnostico de acordo com os sintomas apresentados, isso pode ser observado na Figura 42 (b), em que na medida que o jogador vai resolvendo um caso, outros vão surgindo, caso ele acerte, o tempo é acrescido em 10 segundos, caso ele erre, perde 3 segundos. Dessa forma, o jogador deve tentar realizar o máximo de atendimentos sem errar, demonstrando assim agilidade em realizar diagnósticos. Vale ressaltar que esses dados são disponibilizados por meio do *web service* no módulo inteligente e que são classificados por meio do AM.

No fim dos atendimentos são gerados Pontos de Experiência (XP) e de Atendimentos, como pode ser observado na Figura 42 (c), para motivar os jogadores a continuar realizando atendimentos. Outro recurso disponível no jogo *mobile* que ressaltamos são as salas virtuais, quem conectam os professes aos alunos e os aproximam. Esses salas são customizadas de acordo com o desejo dos professores. No módulo inteligente os professores criam as salas e adicionam as perguntas. No jogo *mobile* os alunos podem entrar nas salas e responder as perguntas disponível, também adquirindo XP. O jogo *mobile* foi validado com 14 estudantes de medicina, que responderam pré-teste, pós-teste e questionário de satisfação.

6 RESULTADOS

A validação do trabalho se deu por meio de duas etapas. A primeira etapa consistiu em testes acerca dos resultados de acertos dos algoritmos e múltiplos classificadores de aprendizado de máquina em cada uma das três bases de dados. A segunda etapa foi por meio da pré-validação realizada com profissionais do curso de medicina da cidade de Mossoró, no estado do Rio Grande do Norte. A seguir são apresentados cada uma dessas etapas de validação. E, por fim, as discussões dos resultados aqui apresentados.

6.1 VALIDAÇÃO COMPUTACIONAL

Nesta Seção são apresentados os resultados dos teste da validação computacional, na qual foram utilizadas diferentes técnicas de AM nas bases de dados. Foram usadas três bases de dados no AM. Detalhes, das três, são apresentados na Tabela 11. A base de casos clínicos possui 213 atributos, 50 classes e 235 instâncias. A base de Diabetes possui oito atributos, duas classes e 768 instâncias. Enquanto a base de Doença Cardíaca possui 13 atributos, duas classes e 303 instâncias.

Tabela 11 – Detalhes das bases de dados utilizadas para o aprendizado.

Base de Dados	**Nº de Atributos**	**Nº de Classes**	**Nº de Instâncias**
Casos Clínicos	213	50	235
Diabetes	8	2	767
Doença Cardíaca	13	2	303

Fonte: Autoria Própria

As técnicas de AM utilizadas foram RL, KNN, MVS, AD, MLP e NB. Assim, foram aplicadas essas técnicas nas bases de dados e gerada a Precisão (DP de 20% e VC com *5-fold*) e medida-F (DP de 40% e VC com *8-fold*) dos modelos de AM construídos. Na Tabela 12 são mostrados os resultados obtidos para as duas formas de validação nas bases de dados de doença cardíaca.

Tabela 12 – Resultados das Classificações na Base de Doença Cardíaca

Classificador	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
RL	86,0%	83,5%	85,0%	82,9%
KNN	75,4%	64,3%	69,9%	65,1%
MVS	88,2%	82,2%	81,7%	82,1%
AD	77,8%	76,2%	79,5%	75,1%
MLP	87,0%	82,5%	81,0%	80,8%
NB	88,8%	80,8%	79,9%	79,4%

Fonte: Autoria Própria

Os melhores resultados foram da RL, pois obteve precisão na DP de 86%, na VC de 83,5%, e medida-F na DP de 85% e da VC de 82,9%; a MVS também obteve bom desempenho, chegando a precisão de 82,2% na VC e medida-F de 82,1% na VC. Enquanto que o KNN não obteve bons desempenhos em relação aos demais. Apesar disso, esses classificadores apresentaram bons resultados. Após verificação desses resultados, foram gerados os múltiplos classificadores. Na Tabela 13 estão as combinações dos classificadores.

Tabela 13 – Resultados dos MLCs na Base de Doença Cardíaca

Combinação	**Classificadores**	**Precisão**		**Medida-F**	
		DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
Voto Majoritário	**NB, MVS, RL**	88,2%	83,8%	82,5%	82,7%
Voto Majoritário	**NB, RL**	88,5%	83,8%	85,2%	83,0%
Voto Majoritário	**NB, MVS**	90,3%	83,5%	82,7%	82,7%
Confiança	**NB,RL**	90,8%	82,8%	85,0%	83,3%
Confiança	**NB,MVS,RL**	90,8%	84,1%	84,1%	83,5%
Confiança	**NB,MVS,MLP**	89,5%	83,5%	81,6%	83,5%

Fonte: Autoria Própria

Na combinação por Voto Majoritário o melhor MLC foi combinando NB e RL, com precisão na DP de 88,5% e VC de 83,8%, e medida-F na DP de 86% e de 84% na VC. Enquanto, na combinação usando Confiança, o melhor resultado foi da NB, MVS e RL juntas, com precisão de 90,8% na DP, 84,1% na VC, e medida-F na DP de 94,1% e na VC de 83,5%. Assim, ficou visível que a combinação melhorou os resultados iniciais das técnicas, pois houve melhora tanto na precisão quanto na medida-F dos resultados. Nessa base, as melhores combinações por Voto Majoritário e Confiança apresentaram resultados quase iguais. Os resultados dos meta-classificadores *ensemble* estão detalhados na Tabela 14, esses resultados foram comparados aos dos MLCs.

Tabela 14 – Resultados dos meta-classificadores *ensembles* na Base de Doença Cardíaca

Classificadores	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
FA	88,8%	84,2%	84,1%	80,8%
Bagging	86,0%	83,8%	85,1%	83,1%
AdaBoost	90,8%	84,5%	85,1%	81,9%

Fonte: Autoria Própria

A FA teve como classificador base a AD nesta base de dados e nas demais também, com 30 estimadores (indutores). O *Bagging* teve 10 estimadores e o *AdaBoost* teve 100 estimadores, para eles foi escolhido como classificador base a RL, pois ela apresentou os melhores resultados dos classificadores iniciais. Os resultados desses meta-classificadores foram agregados por meio da confiança. O *Bagging* e o *AdaBoost* apresentaram resultados semelhantes e conseguiram melhorar os classificadores iniciais. Entretanto esses resultados ainda são menores que os dos MLCs formados. Assim, foi escolhida para classificação nesta base de dados o NB e RL juntos, por meio do Voto Majoritário. Pois conseguiram uma melhor classificação, além de estabilizar os classificadores iniciais. Na Tabela 15 estão os resultados das classificações na base de dados de casos clínicos.

Tabela 15 – Resultados das Classificações na Base de Casos Clínicos

Classificador	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
RL	94,7%	99,7%	92,6%	99,7%
KNN	95,7%	99,2%	91,6%	99,2%
MVS	94,3%	99,2%	91,5%	99,2%
AD	75,5%	76,2%	74,8%	78,5%
MLP	95,7%	99,2%	91,6%	99,2%
NB	93,3%	98,9%	88,1%	98,9%

Fonte: Autoria Própria

Nesta base de dados, os melhores resultados foram da RL, pois obteve precisão na DP de 94,7%, na VC de 99,7%, e medida-F na DP de 92,6% e da VC de 99,7%; o KNN também obteve bom desempenho, chegando a precisão e medida-F de 99,2% na VC. Enquanto que a AD não obteve bons desempenhos em relação aos demais. Os demais classificadores apresentaram bons resultados. Após verificação desses resultados, foram gerados os múltiplos classificadores. Na Tabela 16 estão as combinações dos classificadores.

Tabela 16 – Resultados dos MLCs na Base de Casos Clínicos

Combinação	**Classificadores**	**Precisão**		**Medida-F**	
		DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
Voto Majoritário	**MVS, KNN, RL**	94,7%	99,2%	91,4%	99,2%
Voto Majoritário	**MVS, MLP, NB**	95,7%	99,2%	91,5%	99,2%
Voto Majoritário	**MVS, KNN, RL, MLP, NB**	95,7%	99,2%	92,6%	99,7%
Confiança	**MVS, KNN, NB, RL**	95,7%	99,7%	92,8%	99,7%
Confiança	**MVS, KNN, NB**	95,7%	99,2%	91,6%	99,2%
Confiança	**RL, KNN, MVS**	95,7%	99,2%	91,6%	99,2%

Fonte: Autoria Própria

Na combinação por Voto Majoritário, o melhor MLC foi combinando MVS, KNN, RL, MLP e NB, com precisão na DP de 95,7% e VC de 99,2%, e medida-F na DP de 92,6% e de 99,7% na VC. Enquanto na combinação usando Confiança, o melhor resultado foi da MVS, KNN, NB e RL juntas, com precisão de 95,7% na DP, 99,7% na VC, e medida-F na DP de 92,8% e na VC de 99,7%. Novamente, ficou visível que a combinação melhorou os resultados iniciais, pois houve melhora tanto na precisão quanto na medida-F dos resultados. As melhores combinações por Voto Majoritário e Confiança também apresentaram resultados quase iguais. Os resultados dos meta-classificadores *ensemble* na base de Casos Clínicos estão detalhados na Tabela 17.

Tabela 17 – Resultados dos meta-classificadores *ensembles* na Base de Casos Clínicos

Classificadores	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
FA	90,8%	98,1%	83,2%	97,2%
Bagging	94,7%	99,7%	91,0%	99,7%
AdaBoost	94,7%	99,7%	92,6%	99,4%

Fonte: Autoria Própria

A FA teve 10 estimadores. O *Bagging* e o *AdaBoost* tiveram 15 estimadores, para eles foi escolhido como classificador base a RL, pois ela apresentou novamente, os melhores resultados. Os resultados foram agregados por meio da confiança. O *Bagging* apresentou os melhores resultados dos meta-classificadores, com resultados iguais ao do melhor MLC, exceto na DP com resultados poucos menores. Assim, foi escolhida para classificação nesta base de dados o MVS, KNN, NB e RL juntos, por meio da Confiança, pois conseguiram uma melhor classificação e estabilidade dos classificadores. Na Tabela 18 estão os resultados das classificações na base de dados de Diabetes.

Tabela 18 – Resultados das Classificações na Base de Diabetes

Classificador	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
RL	76,5%	77,0%	75,9%	76,7%
KNN	74,0%	74,4%	73,7%	73,0%
MVS	77,8%	76,8%	77,2%	75,8%
AD	75,6%	72,1%	65,4%	70,2%
MLP	74,2%	69,0%	70,8%	69,3%
NB	76,6%	64,3%	53,6%	55,4%

Fonte: Autoria Própria

Os melhores resultados foram da RL, com precisão na DP de 76,5%, na VC de 77%, e medida-F na DP de 75,9% e da VC de 76,7%; e da MVS, que obteve a precisão de 76,8% na VC e medida-F de 75,8% na VC. Enquanto o NB não obteve bons desempenhos em relação aos demais. Apesar disso, esses classificadores apresentaram bons resultados. Após verificação desses resultados, foram gerados os múltiplos classificadores. Na Tabela 19 estão os resultados dos MLCs.

Tabela 19 – Resultados dos MLCs na Base de Diabetes

Combinação	**Classificadores**	**Precisão**		**Medida-F**	
		DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
Voto Majoritário	**RL, KNN, MVS**	77,9%	76,9%	76,0%	75,7%
Voto Majoritário	**MVS, AD, RL**	77,3%	77,7%	74,8%	76,6%
Voto Majoritário	**MVS, RL, MLP**	77,7%	77,3%	74,8%	75,8%
Confiança	**MVS, RL**	75,3%	77,0%	76,9%	76,3%
Confiança	**MVS, AD, RL**	78,4%	76,1%	68,9%	74,2%
Confiança	**KNN, RL, MLP**	74,7%	76,4%	76,3%	75,3%

Fonte: Autoria Própria

Na combinação por Voto Majoritário, o melhor MLC foi combinando MVS, AD e RL, com precisão na DP de 77,3% e VC de 77,7%, e medida-F na DP de 74,8% e de 76,6% na VC. Enquanto na combinação usando Confiança, o melhor resultado foi da MVS e RL juntas, com precisão de 75,3% na DP, 77% na VC, e medida-F na DP de 76,9% e na VC de 76,3%. Nesta base, os resultados dos MLCs não foram superiores aos dos classificadores iniciais, mas apresentaram medidas semelhantes, entretanto, um pouco menor. A seguir, são descritos os resultados dos meta-classificadores *ensemble*, detalhados na Tabela 14.

Tabela 20 – Resultados dos meta-classificadores *ensembles* na Base de Diabetes

Classificadores	**Precisão**		**Medida-F**	
	DP 20%	**VC 5-*fold***	**DP 40%**	**VC 8-*fold***
FA	74,3%	77,1%	76,6%	74,4%
Bagging	75,7%	77,6%	75,9%	76,5%
AdaBoost	77,1%	77,0%	76,3%	76,3%

Fonte: Autoria Própria

A FA teve 450 estimadores, já o *Bagging* e o *AdaBoost* tiveram 10 e 5 estimadores, sucessivamente, com a RL como classificador base. Novamente, os resultados desses meta-classificadores foram agregados por meio da confiança. Os resultados do *Bagging* foram os melhores, considerando apenas a VC com precisão de 77,6% e medida-F de 76,5%. O *AdaBoost* apresentou resultados pouco menores. Esses resultados ainda são menores que os dos MLCs formados. Portanto, foi escolhida para classificação nesta base de dados a MVS, AD e RL juntos, por meio do Voto Majoritário, que obtiveram melhor classificação que os classificadores iniciais.

Na validação dos classificadores e MLCs não foi levado em conta o tempo de processamento, pois o mesmo apresentou baixo tempo de execução (entre 0 segundo e 8 segundos) em um computador que possui *Windows* 10 com as seguintes configurações: Processador: Intel Core i5 750 CPU 2.67 GHz; Memória: 4 Gigabytes. Na próxima Seção estão os resultados da validação com os profissionais, que tiveram a oportunidade de visualizar esses dados que foram classificados pelo AM e apresentados nesta Seção.

6.2 PRÉ-VALIDAÇÃO COM PROFISSIONAIS

Nesta Seção são apresentados os resultados da entrevista semiestruturada para pré-validação do trabalho. Vale ressaltar que esta pesquisa tem a aprovação do Comitê de Ética em Pesquisa da Universidade do Estado do Rio Grande do Norte, sob o seguinte número de parecer: 3.787.735 de acordo com o apêndice A.

A entrevista semiestruturada foi realizada com dois profissionais da área da medicina que tiveram a oportunidade de utilizar o módulo *web* do DocTraining por uma semana, essa entrevista foi feita separadamente.

Desses dois profissionais uma é professora dos cursos de medicina da Universidade Federal Rural do Semi-Árido (UFERSA), com especialização. Vale ressaltar que o Projeto **A**, mencionado no Capítulo 5 que está diretamente ligado a esta pesquisa, foi validado com os alunos desta professora; e um técnico de laboratório dos cursos de medicina da UFERSA, que possui mestrado. A seguir são discutidas as respostas para as respectivas perguntas, além dessas perguntas, também foram realizadas outras, sendo também realizadas pontuações pelo profissional sobre detalhes importantes do projeto.

1. **O jogo *mobile* sério influência de alguma forma em suas práticas de ensino?**

   A professora informou que sim, que influencia e que este pode ser utilizado como método ativo de ensino, pois, segundo ela, estão sendo bastante utilizados no curso de medicina da UFERSA. Já o Técnico informou que ele pode ser utilizado como ferramenta auxiliar, tipo *quiz*.

2. **O que você mais gostou no Módulo *Web* do DocTraining?**

   Quando indagada sobre isto, em um primeiro momento a professora disse que não sabia opinar. Já o técnico disse que gostou da praticidade de uso. Com isso foram surgindo novas perguntas para que eles pudessem se expressar melhor. Destacamos aqui alguns pontos informados por eles para uma melhor análise.

   Segundo os entrevistados, o módulo *web* do DocTraining apresenta conteúdos interessantes e é capaz de gerenciar os dados do jogo *mobile*, sendo possível adicionar valor ao ensino de medicina para os estudantes obterem com mais clareza os conteúdos das disciplinas ou áreas de interesse.

   Eles declararam interesse em continuar utilizando o projeto. Com isso, é possível observar que, por meio dos conteúdos e funcionalidades, os profissionais acreditam no potencial do módulo *web* do DocTraining e na evolução para se tornar uma ferramenta capaz de auxiliar no ensino-aprendizagem dos estudantes. Por último, a professora realizou o convite para que o projeto seja apresentado aos demais alunos e professores do curso de medicina.

3. **O que você menos gostou ou sentiu falta no módulo *web* do DocTraining?**

   Novamente, a professora disse que não sabia opinar, mas em seguida disse que sentiu ausência de imagens médicas. Entretanto, o técnico disse que faltou informações no módulo *web* acerca de onde foram retiradas as informações das bases de dados deste trabalho, pois, segundo ele, isto é importante até mesmo para os alunos saberem de onde devem estudar. Ao serem indagados sobre se conseguem utilizar módulo *web* do DocTraining sem a ajuda de um instrutor, ambos demostraram que ainda precisam de auxílio para utiliza-lo.

   Mas eles afirmaram que o jogo apresenta uma interface agradável e compreensível. O técnico de laboratório indagou sobre a fonte dos dados apresentados no módulo *web*, e, ao ser informado sobre a fonte, este afirmou que confia nos dados e na classificação deles.

   Por meio dessas respostas foi possível identificar que o módulo *web* do DocTraining ainda carece de ajustes, como melhorar a clareza de suas informações, dentre elas as fontes dos dados, que ainda não é apresentada no o módulo *web*. E onde o usuário

está, pois aparentemente ao editar algumas informações foi constatado que pode acabar confundindo o usuário.

4. **O que você mais gostou no jogo *mobile*?**

   Segundo a professora, o que ela mais gostou foi a variedade de coisas apresentadas (Hospital, Salas, Casos Clínicos). Já o técnico informou que gostou da facilidade de uso e linguagem simples. Destacamos aqui que estes dados são repassados para o Jogo *mobile* por meio do *web service* do DocTraining. Isso demonstra novamente, que, ao fim, juntos, Projetos **A** (Jogo *Mobile*) e **B** (Módulo Inteligente) apresentam simplicidade e facilidade de uso, algo importante para um projeto deste tipo crescer e evoluir.

5. **O que você menos gostou no jogo *mobile*?**

   Assim como no módulo *web* do DocTraining, a professora disse que sentiu falta das imagens médicas. O técnico informou que sentiu ausência de dados de tratamentos. Esses dados também são limitados, pois ainda não consta informações presentes no módulo inteligente sobre imagens médicas ou de tratamento médico para doenças, algo que limita também o jogo *mobile*.

6.3 DISCUSSÕES

A etapa de validação computacional apresentou resultados positivos sobre o uso das técnicas de AM na classificação dos dados médicos, que foram ampliados em seguida com a combinação dos classificadores. Assim, pode-se afirmar que a classificação dos dados pelo AM foi bem-sucedida, em especial na base de dados de casos clínicos e doença cardíaca. Na entrevista com os profissionais, os resultados mostraram que eles concordaram nas classificações que o módulo inteligente realiza, e que este é capaz de gerenciar os dados do jogo sério, sendo classificados pelo AM, como mencionado anteriormente.

Ao longo desta pesquisa, foi descrito que o problema abordado em AM é de classificação. Portanto, os modelos aqui apresentados são treinados, e, em seguida, estão prontos para receber como entrada novos exemplos e apresentar como saída o rotulo referente ao exemplo. Portanto, ao utilizar o modulo os profissionais da área medica poderão observar a descrição de que o exemplo foi classificado automaticamente.

Entretanto esses resultados não são explicáveis, ou seja, não é possível que tenham acesso à explicação do porquê aquele exemplo foi classificado daquela forma. Sendo o conhecimento adquirido pela máquina, neste caso, não explicado, como no casos de algoritmos de *Rede Neural*. Dessa forma podem surgir questionamentos pelos professores e profissionais que utilizarem o modulo inteligente quanto aquela classificação, possibilitando que discordem da classificação, e que mesmo eles discordando os modelos de classificação

podem estar certos e não os profissionais, ou vice-versa. Mas não havendo explicação para distinguir isso. Isto não invalida os resultados da pesquisa aqui apresentados, mas tornam-se um ponto de discussão nesta pesquisa, pois está relacionada a confiabilidade. Mesmo que esses dados sejam classificados automaticamente pelo AM é possível que os profissionais os rotulem manualmente, seguindo o processo de solicitação de edição de amostras.

Por fim, quanto a usabilidade e importância do módulo inteligente a pré-validação com os especialistas obteve um resultado satisfatório, já que a maioria das respostas foram promissoras. Demonstrando que a ferramenta computacional desenvolvida pode ser confiável para ser usada e gerenciar os dados presentes no jogo sério, como as salas virtuais e os dados médicos, trazendo benefícios ao processo de ensino-aprendizagem. Entretanto ainda consta a necessidade de melhorias referentes a facilidade de uso do módulo inteligente.

De modo geral, o DocTraining, quando comparado os resultados desta pesquisa em relação com as similares na versão anterior, desenvolvido por (LIMA, 2016) é visto que houve grandes avanços. Dentre eles estão na arquitetura, que agora é dividida em dois projetos, utilizando um *web service*, facilitando que seja desenvolvido outro jogo sério, o que era inviável na versão anterior, pois os dados estavam em arquivos *.arff* do *Weka* e copiados para o banco de dados, assim houve problemas quanto ao crescimento do projeto, que acessava diretamente o banco de dados. Assim, agora o jogo sério interage com o *web service* consumindo os dados presentes no módulo inteligente.

Outro ponto foi o Criador de Sintomas, Doenças e Amostras da versão anterior, que era acessado apenas localmente, no Computador que o tivesse, e que com algumas adaptações poderia ser utilizado na mesma rede em que estivessem os dados que ela edita. Portanto, isso também dificultava que houvesse incremento pelos profissionais, pois necessitaria do deslocamento deles para realizar isso, além de não ser possível saber quem fez as alterações. Nesta nova versão o sistema para gerenciamento de dados é on-line, sendo acessado por qualquer computador que possua internet, possuindo histórico de atualização dos dados, que indicam data, hora, usuário que atualizou, dados anteriores e os novos.

Ainda constam nesta nova versão as salas virtuais, que é um novo recurso que os professores podem utilizar, adicionando conteúdos relevantes para os alunos responderem. O sistema de aprendizado de máquina também apresentou resultados positivos, pois com o uso dos múltiplos classificares os modelos apresentados tornaram-se estáveis, evitando grandes variações nas classificações.

7 CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado um módulo inteligente para auxiliar no gerenciamento de um jogo sério que auxilia no ensino-aprendizagem dos estudantes e professores de medicina. Por meio dos resultados neste Capítulo são apresentadas as conclusões obtidas na realização desta pesquisa, as limitações existentes no trabalho e os trabalhos futuros.

7.1 CONCLUSÕES

O módulo inteligente do DocTraining conta com um *web* site para divulgação, gerenciamento e expansão de dados pelos professores e administradores. As informações são disponibilizadas para o jogo *mobile*, por meio de um *web service*, no formato JSON. Na validação foram utilizados modelos de aprendizado de máquina para classificação de doença cardíaca, diabetes e casos clínicos, por meio de divisão de porcentagem (40% e 20%) e validação cruzada (5 e 8 *fold*).

Também foi utilizada a combinação de classificadores para melhorar a precisão e medida-F nos resultados dos indutores iniciais. Outra combinação de classificadores foi utilizando meta-classificadores. Assim, foi possível observar que a melhor forma de rotular os dados foi por meio dos múltiplos classificadores, sejam eles formados por meio de Voto Majoritário e Confiança ou por Meta-classificadores para melhorar os resultados dos classificadores iniciais.

Neste estudo, para classificação de doença cardíaca a combinação por *Naive Bayes*, *Máquina de vetores de suporte* e *Regressão logística* usando Confiança mostrou ser a mais eficiente. Para base de dados de Diabetes, a combinação por *Máquina de vetores de suporte*, *Árvore de decisão* e *Regressão logística* usando Voto Majoritário mostrou ser a mais eficiente.

Entretanto, na base de dados de Casos Clínicos, os resultados ainda apresentam medidas semelhantes tanto para os múltiplos classificadores como para os meta-classificadores, com bons resultados para ambos, mas o múltiplo classificador formado por *Máquina de vetores de suporte*, *K-Nearest Neighbor*, *Naive Bayes* e *Regressão logística* usando Confiança apresentou os melhores resultados. O uso dos múltiplos classificadores e dos meta-classificadores melhorou os resultados iniciais, como era esperado. A fusão dos classificadores se deu por meio da função de Votação Majoritária e pela Confiança (probabilidade de uma classe ser a prevista).

Por meio da entrevista os profissionais da medicina, foi possível observar que o módulo inteligente do DocTraining tem a capacidade de continuar evoluindo e se tornar

uma ferramenta poderosa para auxiliar no desenvolvimento de outros jogos sérios que utilizem o *web service*. O trabalho foi feito para que seja uma base de evolução de ambos os projetos, com propósito de continuar seu desenvolvimento.

Sendo assim, é possível notar que os projetos encontram-se robustos e com proposta de melhorias, por meio dos trabalhos futuros descritos na Seção 7.3. Com os resultados encontrados, ficou comprovado que o módulo inteligente pode ser capaz de gerenciar os dados do jogo *mobile*, pois possui toda uma estrutura para este objetivo.

Esta pesquisa apresenta as seguintes contribuições científicas: uma visão geral do aprendizado de máquina e a identificação do estado da arte relacionado ao aprendizado de máquina nos jogos para medicina; apresenta o uso dos algoritmos de aprendizado de máquina na classificação dos dados para um módulo inteligente. Assim, construindo modelos de classificação que se tornem estáveis e com melhores resultados, servindo como ponto de partida para novas pesquisas. Outras contribuições estão relacionadas à computação aplicada, na qual, de um modo especifico, será possível que novos jogos sérios sejam desenvolvidos, consumindo os dados presentes no *web service* e, de modo geral, auxiliando no ensino-aprendizagem de estudantes de medicina, por intermédio desses jogos.

7.2 LIMITAÇÕES

Algumas limitações na solução proposta foram identificadas após a execução do experimento. Dentre elas incluem:

- O número pequeno de funcionalidades, visto que há uma quantidade maior de funcionalidades que podem ser exploradas no projeto e no *web service*.
- Foram realizadas apenas seleções de classificadores por meio de método estático, pois este único subconjunto de classificadores rotula todas as instâncias de teste.
- A pré-validação foi feita apenas com dois especialistas, que utilizaram o módulo inteligente por pouco tempo, assim, devido a pequena amostra os dados, os resultados da entrevista não podem ser generalizados para todos os profissionais de saúde.
- As bases de doença cardíaca e diabetes não foram testadas no jogo *mobile*.

7.3 TRABALHOS FUTUROS

Por meio das conclusões e limitações descritas anteriormente, em trabalhos futuros serão realizadas modificações no módulo inteligente por meio das conclusões obtidas mediante pré-validação com os profissionais. Assim serão realizadas modificações nos dados dos casos clínicos, sendo os sintomas divididos em aqueles que o paciente pode informar

no jogo e aqueles que são descobertos apenas por meio de exames. Outro ponto será a modificação do *web service*, que contará com esses dados de exames e de doenças com sintomas semelhantes.

O jogo será então capaz de apresentar outra doença que apresenta sintomas semelhantes, incentivando a investigação do estudante para dar um diagnostico preciso sobre os sintomas iniciais apresentados. Também será adicionada a combinação de classificadores usando o método dinâmico, assim, para cada nova instância serão escolhidos os classificadores que podem ter maiores chances de acertar (CRUZ; SABOURIN; CAVALCANTI, 2018).

Por fim, será realizada uma validação completa com mais profissionais da medicina usando o módulo inteligente por, no mínimo, um semestre, para que seja possível ver o desenvolvimento dos alunos, ao fim serão realizadas entrevistas e aplicados questionários de aceitação de tecnologia, seguindo o modelo TAM (*Tecnology Accept Model*), proposto por Davis (1989) ou alguma de suas variações.

Referências

ABNT. *ABNT NBR ISO/IEC 14598-6: Engenharia de Software-avaliação de produto-parte 6: documentação de módulos de avaliação*. [S.l.]: ABNT, 2004. Citado na página 65.

ABT, C. C. *Serious games*. [S.l.]: University press of America, 1987. Citado na página 53.

ACHERMANN, B.; BUNKE, H. Combination of classifiers on the decision level for face recognition. *Universitat Bern. Bern*, (IAM-96-2002), 1996. Citado na página 48.

AHMAD, M. A. *et al.* Benchmarking expert surgeons' path for evaluating a trainee surgeon's performance. In: ACM. *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. [S.l.], 2013. p. 57–62. Citado 4 vezes nas páginas 46, 60, 62 e 63.

ALCHALABI, A. E. *et al.* Feasibility of detecting adhd patients' attention levels by classifying their eeg signals. In: IEEE. *Medical Measurements and Applications (MeMeA), 2017 IEEE International Symposium on*. [S.l.], 2017. p. 314–319. Citado 5 vezes nas páginas 46, 60, 61, 62 e 64.

ALCHALABI, A. E. *et al.* Focus: Detecting adhd patients by an eeg-based serious game. *IEEE Transactions on Instrumentation and Measurement*, IEEE, 2018. Citado 4 vezes nas páginas 60, 61, 62 e 65.

Alchalcabi, A. E.; Eddin, A. N.; Shirmohammadi, S. More attention, less deficit: Wearable eeg-based serious game for focus improvement. In: *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. [S.l.: s.n.], 2017. p. 1–8. Citado na página 65.

BALOGH, E. P. *et al.* The path to improve diagnosis and reduce diagnostic error. In: *Improving Diagnosis in Health Care*. [S.l.]: National Academies Press (US), 2015. Citado na página 29.

BENBRAHIM, H.; BRAMER, M. Text and hypertext categorization. In: SPRINGER, BERLIN, HEIDELBERG. *Bramer M. (eds) Artificial Intelligence An International Perspective, Lecture Notes in Computer Science*. [S.l.], 2009. v. 5640, p. 11–38. Citado na página 37.

BERRY, M. J.; LINOFF, G. *Data mining techniques: for marketing, sales, and customer support*. [S.l.]: John Wiley & Sons, Inc., 1997. Citado na página 19.

BI, Y. *et al.* Combining multiple classifiers using dempster's rule of combination for text categorization. In: SPRINGER. *International Conference on Modeling Decisions for Artificial Intelligence*. [S.l.], 2004. p. 127–138. Citado na página 49.

BIOLCHINI, J. *et al.* Systematic review in software engineering. *System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES*, v. 679, n. 05, p. 45, 2005. Citado na página 54.

BIOLCHINI, J. C. de A. *et al.* Scientific research ontology to support systematic review in software engineering. *Advanced Engineering Informatics*, Elsevier, v. 21, n. 2, p. 133–151, 2007. Citado na página 32.

BOGONI, T. N.; PINHO, M. S. Avaliação de um simulador háptico de realidade virtual para treinamento de endodontia. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2014. v. 25, n. 1, p. 259. Citado 2 vezes nas páginas 27 e 30.

BRASIL. *Relatório Síntese dos Resultados, Avaliação Nacional Seriada dos Estudantes de Medicina - 2016*. [S.l.]: Inep/MEC, 2016. Citado na página 17.

BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996. Citado na página 51.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 52.

CARVALHO, B. R. de *et al.* Erro médico: implicações éticos, jurídicas e perante o código de defesa do consumidor. *Revista de Ciências Médicas*, v. 15, n. 6, 2012. Citado na página 28.

CASTRO, R.; SIQUEIRA, S. Aprendizagem ativa em sistemas de informação: Novas técnicas propostas e reflexões sobre as experiências. In: SBC. *Anais do XIII Simpósio Brasileiro de Sistemas de Informação*. [S.l.], 2017. p. 535–542. Citado na página 27.

CATON, S. *et al.* Dynamic model evaluation to accelerate distributed machine learning. In: IEEE. *2018 IEEE International Congress on Big Data (BigData Congress)*. [S.l.], 2018. p. 150–157. Citado na página 33.

CFM. *Cremesp divulga resultado de exame dirigido a estudantes de Medicina e recém-formados*. 2018. Disponível em: <http://www.portal.cfm.org.br>. Citado na página 27.

CHAVES, B. B. *Estudo do algoritmo AdaBoost de aprendizagem de máquina aplicado a sensores e sistemas embarcados.* Tese (Doutorado) — Universidade de São Paulo, 2012. Citado na página 47.

CHOUHAN, T. *et al.* A comparative study on the effect of audio and visual stimuli for enhancing attention and memory in brain computer interface system. In: IEEE. *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. [S.l.], 2015. p. 3104–3109. Citado 3 vezes nas páginas 59, 61 e 70.

CLUA, E. W. G. Jogos sérios aplicados a saúde. *Journal of Health Informatics*, v. 6, 2014. Citado na página 53.

COUTO, R. C. *et al.* Ii anuário da segurança assistencial hospitalar no brasil. Instituto de Estudos de Saúde Suplementar, 2018. Citado na página 29.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE transactions on information theory*, IEEE, v. 13, n. 1, p. 21–27, 1967. Citado na página 48.

CREMESP. *Exame do CREMESP 2018 aprova 61% dos médicos recém-formados*. 2019. Disponível em: <https://www.cremesp.org.br/pdfs/Relatorio-Exame-Cremesp_certo.pdf>. Citado na página 27.

CREMESP. *Exame do Cremesp aprova mais da metade dos médicos recém-formados*. 2019. Disponível em: <https://www.cremesp.org.br>. Citado na página 27.

CREMESP. *Exame do CREMESP reprova mais da metade dos médicos recém-formados*. 2019. Disponível em: <https://www.cremesp.org.br/pdfs/releasefinal2examecremesp2016.pdf>. Citado na página 27.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, Elsevier, v. 41, p. 195–216, 2018. Citado na página 98.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, JSTOR, p. 319–340, 1989. Citado na página 98.

DEO, R. C. Machine learning in medicine. *Circulation*, Am Heart Assoc, v. 132, n. 20, p. 1920–1930, 2015. Citado 2 vezes nas páginas 53 e 71.

DEO, R. C. *et al.* Prioritizing causal disease genes using unbiased genomic features. *Genome biology*, BioMed Central, v. 15, n. 12, p. 534, 2014. Citado na página 53.

FACELI, K. *et al.* Inteligência artificial–uma abordagem de aprendizado de máquina. *Rio de Janeiro: LTC*, 2011. Citado 7 vezes nas páginas 37, 38, 39, 40, 42, 43 e 44.

FAJARDO, V. *Faculdades de medicina adotam novos modelos de aula; saiba o que são os métodos TBL e PBL*. 2019. Disponível em: <https://g1.globo.com/educacao/guia-de-carreiras/noticia/2018/08/08/faculdades-de-medicina-adotam-novos-modelos-de-aula-saiba-o-que-sao-os-metodos-tbl-e-pbl.ghtml>. Citado na página 26.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado na página 18.

FLORES, C. D. *et al.* Leveraging the learning process in health through clinical cases simulator. In: IEEE. *Serious Games and Applications for Health (SeGAH), 2013 IEEE 2nd International Conference on*. [S.l.], 2013. p. 1–6. Citado 8 vezes nas páginas 47, 61, 62, 63, 65, 66, 71 e 72.

FREUND, Y.; SCHAPIRE, R. E. *et al.* Experiments with a new boosting algorithm. In: CITESEER. *icml*. [S.l.], 1996. v. 96, p. 148–156. Citado na página 52.

FRUTOS-PASCUAL, M.; ZAPIRAIN, B. G. Review of the use of ai techniques in serious games: Decision making and machine learning. *IEEE Transactions on Computational Intelligence and AI in Games*, v. 9, n. 2, p. 133–152, June 2017. ISSN 1943-068X. Citado na página 53.

Fu, P.; Peng, Q.; Hu, X. A web service composition system based on semantic parsing. In: *2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. [S.l.: s.n.], 2015. p. 561–569. Citado na página 80.

G1. *Enade não diz se curso é bom, só se é melhor ou pior que outro; entenda*. 2018. Disponível em: <http://g1.globo.com/educacao/noticia/2013/10/enade-nao-diz-se-curso-e-bom-so-se-e-melhor-ou-pior-que-outro-entenda.html>. Citado na página 28.

GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. *Epidemiologia e Serviços de Saúde*, SciELO Public Health, v. 23, p. 183–184, 2014. Citado na página 54.

GEORGE, L. *et al.* Using scalp electrical biosignals to control an object by concentration and relaxation tasks: design and evaluation. *arXiv preprint arXiv:1111.5285*, 2011. Citado 2 vezes nas páginas 61 e 70.

GONÇALVES, E.; VILELA, J.; BEZERRA, J. Análise estatística de notas e interações em cursos a distância. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2018. v. 29, n. 1, p. 71. Citado na página 19.

GOODE, S. *et al.* Enhancing is education with flexible teaching and learning. Journal of Information Systems Education, 2007. Citado na página 27.

HAMERLY, G.; ELKAN, C. Learning the k in k-means. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 281–288. Citado 2 vezes nas páginas 19 e 21.

HASTIE, T. *et al.* Multi-class adaboost. *Statistics and its Interface*, International Press of Boston, v. 2, n. 3, p. 349–360, 2009. Citado na página 78.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer Science & Business Media, 2009. Citado na página 52.

HELLER, M. D. *et al.* A machine learning-based analysis of game data for attention deficit hyperactivity disorder assessment. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 2, n. 5, p. 291–298, 2013. Citado 5 vezes nas páginas 60, 61, 67, 68 e 70.

HERVÁS, R. *et al.* A learning system to support social and empathy disorders diagnosis through affective avatars. In: IEEE. *Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), International Conference on*. [S.l.], 2016. p. 93–100. Citado 3 vezes nas páginas 60, 62 e 63.

INEP. *Definidos procedimentos de cálculo e divulgação dos Indicadores de Qualidade da Educação Superior de 2017*. 2018. Disponível em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/definidos-procedimentos-de-calculo-e-divulgacao$\$-dos-indicadores-de-qualidade-da-educacao-superior-de-2017/21206>. Citado na página 17.

INEP. *Enade*. 2018. Disponível em: <http://portal.inep.gov.br/relatorios>. Citado na página 17.

INEP. *Enade 2016 tem participação de mais de 90% dos inscritos e menor abstenção da série histórica*. 2018. Disponível em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/enade-2016-tem-participacao-de-mais-de-90-dos-inscritos-e-menor-abstencao-da-serie-historica/21206>. Citado na página 17.

INEP. *RESUMO ENADE 2016*. 2018. Disponível em: <http://download.inep.gov.br/educacao_superior/enade/documentos/2016/resumo_enade_2016_abstencao.pdf>. Citado na página 17.

JAMES, J. T. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety*, LWW, v. 9, n. 3, p. 122–128, 2013. Citado na página 29.

JINHUAXU; HONGLIU. Web user clustering analysis based on kmeans algorithm. In: *2010 International Conference on Information, Networking and Automation (ICINA)*. [S.l.: s.n.], 2010. v. 2, p. V2–6–V2–9. ISSN 2162-5476. Citado na página 22.

JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado na página 47.

KIM, M. *et al.* Vizical: Accurate energy expenditure prediction for playing exergames. In: ACM. *Proceedings of the 26th annual ACM symposium on User interface software and technology*. [S.l.], 2013. p. 397–404. Citado 3 vezes nas páginas 61, 66 e 67.

KINCAID, J. P. *et al.* Simulation in education and training. In: *Applied system simulation*. [S.l.]: Springer, 2003. p. 437–456. Citado na página 46.

KITCHENHAM, B. *Procedures for Performing Systematic Reviews*. Department of Computer Science, Keele University, UK, 2004. Citado na página 54.

LEWIS, D. D. Naive (bayes) at forty: The independence assumption in information retrieval. In: SPRINGER. *European conference on machine learning*. [S.l.], 1998. p. 4–15. Citado na página 48.

LI, I. T. Y.; YANG, Q.; WANG, K. Classification pruning for web-request prediction. In: *WWW Posters*. [S.l.: s.n.], 2001. Citado na página 21.

LIMA, R. M. *Doctraining: Um Ambiente 3D Com Jogo Sério para o Treinamento de Estudantes de Medicina em Casos Clínicos*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal Rural do Semi-Árido - UFERSA e Universidade do Estado do Rio Grande do Norte - UERN, Programa de Pós-Graduação em Ciência da Computação, 2016. Citado 7 vezes nas páginas 29, 34, 35, 36, 69, 73 e 95.

LIMA, R. M. *et al.* A 3d serious game for medical students training in clinical cases. In: *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*. [S.l.: s.n.], 2016. p. 1–9. Citado 5 vezes nas páginas 61, 62, 63, 69 e 71.

LOPES, G. Q. *Vinte erros comuns na prática clínica*. 2020. Disponível em: <https://pebmed.com.br/vinte-erros-comuns-na-pratica-clinica/>. Citado na página 29.

LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado 2 vezes nas páginas 42 e 43.

LUGER, G. F. *Artificial Intelligence*. [S.l.]: Pearson Education, 2014. Citado 2 vezes nas páginas 36 e 37.

MACHADO, F. N. R. *Banco de Dados-Projeto e Implementação*. [S.l.]: Editora Saraiva, 2018. Citado na página 74.

MAKARY, M. A.; DANIEL, M. Medical error—the third leading cause of death in the us. *Bmj*, British Medical Journal Publishing Group, v. 353, p. i2139, 2016. Citado na página 29.

MANCHEGO, F. E. A. *Anotação automática semissupervisionada de papéis semânticos para o português do Brasil*. Dissertação (Mestrado em Ciências Matemáticas e de Computação) — Universidade de São Paulo, 2013. Citado na página 44.

MARAN, N. J.; GLAVIN, R. J. Low-to high-fidelity simulation–a continuum of medical education? *Medical education*, Wiley Online Library, v. 37, p. 22–28, 2003. Citado na página 30.

MATSUBARA, E. T. *O algoritmo de aprendizado semi-supervisionado co-training e sua aplicação na rotulação de documentos*. Dissertação (Mestrado em Ciências Matemáticas e de Computação) — Universidade de São Paulo, 2004. Citado 2 vezes nas páginas 42 e 44.

MELO, F. S. Convergence of q-learning: A simple proof. *Institute Of Systems and Robotics, Tech. Rep*, p. 1–4, 2001. Citado na página 47.

MENDES, W. *et al.* The assessment of adverse events in hospitals in brazil. *International Journal for Quality in Health Care*, Oxford University Press, v. 21, n. 4, p. 279–284, 2009. Citado na página 29.

Ministério da Educação. *Entidades prestam apoio a propostas de melhoria de programas de residência*. 2019. Disponível em: <http://portal.mec.gov.br/ultimas-noticias/212-educacao-superior-1690610854/18939-entidades-medicas-apoiam-proposta-de-especialistas-para-aprimorar-o-ensino>. Citado na página 26.

Ministério da Saúde. Doenças infecciosas e parasitárias: guia de bolso. *Brasília: Ministério da Saúde*, p. 1–444, 2010. Citado na página 73.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, v. 1, n. 1, p. 39–56, 2003. Citado 5 vezes nas páginas 37, 38, 41, 44 e 45.

MONARD, M. C.; PRATI, R. C. Aprendizado de máquina simbólico para mineração de dados. XIII Escola Regional de Informática, p. 1–26, 2005. Citado 3 vezes nas páginas 37, 41 e 42.

MOTA, A. *et al.* Exame do cremesp como indicador da qualidade do ensino médico. *Revista Brasileira de Educação Médica*, v. 38, p. 150–159, 2014. Citado 2 vezes nas páginas 21 e 26.

OLIVEIRA, N. A. d. *et al.* Mudanças curriculares no ensino médico brasileiro: um debate crucial no contexto do promed. SciELO Brasil, 2008. Citado na página 26.

Organização Mundial da Saúde. *10 facts on patient safety*. 2020. Disponível em: <https://www.who.int/features/factfiles/patient_safety/en/>. Citado na página 29.

PAL, S.; MITRA, S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks*, IEEE Press Piscataway, NJ, USA, v. 3, n. 5, p. 683–697, 1992. Citado na página 76.

PEARL, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. [S.l.]: Elsevier, 2014. Citado na página 46.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 32.

PERDIZ, J. *et al.* Measuring the impact of reinforcement learning on an electrooculography-only computer game. In: *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. [S.l.: s.n.], 2018. p. 1–8. ISSN 2573-3060. Citado 4 vezes nas páginas 61, 62, 63 e 70.

PINHEIRO, R. H. W. *Combinação de Classificadores em Diferentes Espaços de Características para Classificação de Documentos*. Tese (Doutorado em Ciência da Computação) — Programa de Pós-Graduação em Ciência da Universidade Federal de Pernambuco - UFPE, 2017. Citado na página 50.

PRASS, F. S. *Estudo Comparativo Entre Algoritmos de Análise de Agrupamentos em Data Mining*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Ciência da Computação, nov 2004. Citado na página 19.

PRESSMAN, R.; MAXIM, B. *Engenharia de Software-8ª Edição*. [S.l.]: McGraw Hill Brasil, 2016. Citado na página 32.

ROSSI, R. G. *Classificação automática de textos por meio de aprendizado de máquina baseado em redes*. Tese (Doutorado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, 2016. Citado na página 37.

RUSSELL, S.; NORVIG, P. *Inteligência Artificial: Tradução da 3a Edição*. [S.l.]: Elsevier Brasil, 2014. v. 1. ISBN 978-85-352-3701-6. Citado 4 vezes nas páginas 42, 43, 44 e 45.

SALVADEO, D. H. P. *Combinação de Múltiplos Classificadores para Reconhecimento de Face Humana*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal São Carlos - UFSCar, Programa de Pós-Graduação em Ciência da Computação, 2009. Citado 4 vezes nas páginas 48, 49, 50 e 51.

SANTOS, C. M. L. S. A. Estatística descritiva. *Manual de Auto-aprendizagem, Lisboa, Edições Sílabo*, 2007. Citado na página 18.

SATAVA, R. M.; GALLAGHER, A. G.; PELLEGRINI, C. A. Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. *Journal of the American College of Surgeons*, Elsevier, v. 196, n. 6, p. 933–937, 2003. Citado na página 30.

SCIKIT-LEARN. *Scikit-learn user guide*. Release 0.21.2. https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf, 2019. Citado 4 vezes nas páginas 32, 39, 76 e 78.

SILVA, J. C. P. *et al.* Aprendizado de máquina nos jogos para medicina: Uma revisão sistemática. In: *Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2019. p. 70–81. Disponível em: <https://doi.org/10.5753/sbcas.2019.6243>. Citado 7 vezes nas páginas 30, 58, 60, 61, 70, 71 e 72.

SILVA, W. K. N. da. *Construções de Comitês de Classificadores Multirrótulos no Aprendizado Semissupervisionado Multidescrição*. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal Rural do Semi-Árido - UFERSA e Universidade do Estado do Rio Grande do Norte - UERN, Programa de Pós-Graduação em Ciência da Computação, 2017. Citado na página 38.

SOMMERVILLE, I. Engenharia de software, 8 edição. *Pearson, Addison Wesley*, v. 8, n. 9, p. 10, 2007. Citado na página 32.

SOUSA NETO, A. *et al.* Avaliação de um ambiente virtual gamificado para auxiliar o ensino-aprendizagem de estudantes de medicina. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2018. v. 29, n. 1, p. 496. Citado 2 vezes nas páginas 36 e 73.

VILAGRA, S. M. B. W. *ELABORAÇÃO DE MANUAL DE ESTRATÉGIAS DE ENSINO PARA DOCENTES DE MEDICINA*. Dissertação ((Mestrado Profissional em Ensino em Ciências da Saúde e do Meio Ambiente)) — Centro Universitário de Volta Redonda – UniFOA, 2012. Citado na página 27.

WATKINS, C. J.; DAYAN, P. Q-learning. *Machine learning*, Springer, v. 8, n. 3-4, p. 279–292, 1992. Citado na página 47.

WOLPERT, D. H. The supervised learning no-free-lunch theorems. In: *Soft computing and industry*. [S.l.]: Springer, 2002. p. 25–42. Citado na página 48.

WOŹNIAK, M.; GRAÑA, M.; CORCHADO, E. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, Elsevier, v. 16, p. 3–17, 2014. Citado na página 51.

XU, L.; KRZYZAK, A.; SUEN, C. Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 22, n. 3, p. 418–435, 1992. Citado 2 vezes nas páginas 49 e 51.

YANG, Y.; AULT, T.; PIERCE, T. Combining multiple learning strategies for effective cross validation. In: *ICML*. [S.l.: s.n.], 2000. p. 1167–1174. Citado na página 48.

ZHANG, H. The optimality of naive bayes. *AA*, v. 1, n. 2, p. 3, 2004. Citado na página 48.

ZHANG, Z.; XIE, X. Research on adaboost. m1 with random forest. In: IEEE. *2010 2nd International Conference on Computer Engineering and Technology*. [S.l.], 2010. v. 1, p. V1–647. Citado na página 52.

ZIELKE, M. A. *et al.* Exploring medical cyberlearning for work at the human/technology frontier with the mixed-reality emotive virtual human system platform. In: IEEE. *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. [S.l.], 2018. p. 1–8. Citado 4 vezes nas páginas 61, 62, 63 e 68.

Apêndices

APÊNDICE A – COMITÊ DE ÉTICA EM PESQUISA

UERN - UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: DocTraining Mobile: Um Jogo Sério para o Treinamento de Estudantes de Medicina em Casos Clínicos

Pesquisador: Everton Jales de Oliveira

Assistentes: Jesaías Carvalho Pereira Silva e Francisco Milton Mendes Neto

Área Temática:

Versão: 2

CAAE: 23653019.0.0000.5294

Instituição Proponente: UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO - UFERSA

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.787.735

Apresentação do Projeto:

O projeto de pesquisa é uma atividade mestrado que tem como objetivo desenvolver e avaliar um jogo sério, que proporcione uma experiência teórica, com finalidade de maximizar o auto-aprendizado de estudantes de medicina em casos clínicos. O jogo será desenvolvido para smartphones e contará com vários elementos como ranking, pontos de experiência, conquistas, itens colecionáveis, de forma a estimular os estudantes no uso do jogo e através de uma experiência lúdica que oferece conhecimento que facilitem a identificação de doenças e assuntos da grade curricular de medicina de maneira geral. O jogo conta com uma base de dados de doenças cardíacas, diabetes e casos clínicos, que são classificados por meio de múltiplos classificadores de aprendizado de máquina. O projeto foi dividido em 3 fases, a primeira fase do projeto foi o levantamento de requisitos por meio de uma apresentação oral da proposta e aplicação de questionário com 25 professores do curso de medicina, que não foi descrito na metodologia do projeto em anexo; a segunda foi o desenvolvimento do simulador e, por fim, a terceira fase será o experimento prático com os alunos. Nessa fase, será feito um estudo experimental com os alunos para avaliar a questão do auto aprendizado proporcionado pelo jogo, em seguida será aplicado um questionário de aceitação de tecnologia, seguindo o modelo TAM (Tecnology Accept Model), para fazer uma análise qualitativa e quantitativa da utilidade e facilidade de uso do jogo. Para finalizar todo o processo de validação e coleta de dados, também será realizado um grupo focal, que é uma discussão estruturada para obter

informações relevantes de um grupo de pessoas sobre um tópico específico. E a realização de uma entrevista semiestruturada com alguns professores ou profissionais da área. Serão analisados os pontos negativos e positivos do jogo com base na visão dos alunos e professores, sugestões de melhorias e levantamento de novos requisitos para futuras versões.

Objetivo da Pesquisa:

A finalidade do projeto DocTraining Mobile é validar um jogo sério de casos clínicos na medicina, com o intuito de maximizar o auto-aprendizado dos alunos nesse quesito.

Avaliação dos Riscos e Benefícios:

Os riscos e benefícios foram avaliados

Comentários e Considerações sobre a Pesquisa:

A pesquisa é relevante

Considerações sobre os Termos de apresentação obrigatória:

Todos os termos obrigatórios foram anexados

Conclusões ou Pendências e Lista de Inadequações:

Todas as pendências foram sanadas

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1451012.pdf	07/12/2019 11:41:34		Aceito
Outros	carta_de_anuencia.pdf	07/12/2019 11:36:30	Everton Jales de Oliveira	Aceito
Outros	Roteiro_entrevista_semiestruturada_para_professores_DocTraining.pdf	07/12/2019 11:32:58	Everton Jales de Oliveira	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Professores_Doctraining.pdf	07/12/2019 11:31:47	Everton Jales de Oliveira	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Alunos_Doctraining.pdf	07/12/2019 11:30:44	Everton Jales de Oliveira	Aceito

UERN - UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE

Continuação do Parecer: 3.787.735

Projeto Detalhado / Brochura Investigador	Projeto_de_pesquisa_DocTraining.pdf	07/12/2019 11:29:16	Everton Jales de Oliveira	Aceito
Folha de Rosto	folha_de_rosto.pdf	10/10/2019 16:36:52	Everton Jales de Oliveira	Aceito
Outros	Questionario_Satisfacao_DoctrainingM.pdf	09/10/2019 17:54:18	Everton Jales de Oliveira	Aceito
Declaração de Pesquisadores	Declaracao_compromisso_pesquisador.Pdf	09/10/2019 17:32:21	Everton Jales de Oliveira	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

MOSSORO, 20 de Dezembro de 2019

Assinado por:
Ana Clara Soares Paiva Tôrres
(Coordenador(a))

Endereço: Rua Miguel Antonio da Silva Neto, s/n
Bairro: Aeroporto **CEP:** 59.607-360
UF: RN **Município:** MOSSORO
Telefone: (84)3312-7032 **E-mail:** cep@uern.br