Lucas de Oliveira Batista Alves

USO DE ROTINAS DE APRENDIZADO DE MÁQUINA EM PRONTUÁRIO ELETRÔNICO PARA APOIO A DIAGNÓSTICOS DE PACIENTES OFTALMOLÓGICOS

Dissertação apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina, para obtenção do título de Mestre Profissional em Tecnologia, Gestão e Saúde Ocular.

São Paulo
2021

Lucas de Oliveira Batista Alves

USO DE ROTINAS DE APRENDIZADO DE MÁQUINA EM PRONTUÁRIO ELETRÔNICO PARA APOIO A DIAGNÓSTICOS DE PACIENTES OFTALMOLÓGICOS

Dissertação apresentada à Universidade Federal de São Paulo – Escola Paulista de Medicina, para obtenção do título de Mestre Profissional em Tecnologia, Gestão e Saúde Ocular.

Orientador:
Prof. Dr. Vagner Rogério dos Santos

São Paulo
2021

Ficha catalográfica elaborada pela Biblioteca Prof. Antonio Rubino de Azevedo,
Campus São Paulo da Universidade Federal de São Paulo, com os dados fornecidos pelo autor

Alves, Lucas de Oliveira Batista
Uso de rotinas de aprendizado de máquina em prontuário eletrônico para apoio a diagnósticos de pacientes oftalmológicos / Lucas de Oliveira Batista Alves. – São Paulo, 2021.
xiii, 41f.

Dissertação (mestrado) – Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós-Graduação em Tecnologia, Gestão e Saúde Ocular.

Título em inglês: Use of machine learning routines in electronic medical records to support ophthalmologic patient diagnoses.

1. Inteligência artificial. 2. Aprendizado de máquina. 3. Prontuário eletrônico. 4. Apoio diagnóstico.

UNIVERSIDADE FEDERAL DE SÃO PAULO

ESCOLA PAULISTA DE MEDICINA

PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA, GESTÃO E SAÚDE OCULAR

Chefe do Departamento:

Prof. Dr. Mauro Silveira de Queiroz Campos

Coordenador do Curso de Pós-graduação:

Prof. Dr. José Álvaro Pereira Gomes

Lucas de Oliveira Batista Alves

USO DE ROTINAS DE APRENDIZADO DE MÁQUINA EM PRONTUÁRIO ELETRÔNICO PARA APOIO A DIAGNÓSTICOS DE PACIENTES OFTALMOLÓGICOS

Presidente da banca:

Prof. Dr. Vagner Rogério dos Santos _____

Banca examinadora:

Dra. Rita Simone Lopes Moreira _____

Dr. Denys Emilio Campion Nicolosi _____

Dr. Caio Vinicius Saito Regatieri _____

Suplente:

Dr. Rossen Hazarbassanov _____

Data de aprovação: _____

Dedicatória

Dedico este trabalho aos meus pais e à minha esposa, com admiração e gratidão pelo apoio, carinho e presença ao longo da elaboração deste trabalho.

Agradecimentos

A Deus, por estar sempre comigo, me guiando, me abençoando e iluminando meus passos.

Minha gratidão especial ao Prof. Dr. Vagner Rogério dos Santos, meu orientador, pela pessoa e grande profissional que é.

À Unifesp, pela oportunidade de realização do curso.

À minha esposa e aos meus pais, pelo incentivo que me deram o tempo todo desta minha jornada.

Resumo

Objetivo: Implementar rotinas de inteligência artificial por meio do aprendizado de máquina para a construção de modelos de predição de diagnósticos com dados de prontuários eletrônicos dos pacientes do Departamento de Oftalmologia do Hospital São Paulo. **Método:** Elaboração de revisão bibliográfica sobre as principais técnicas e soluções de aprendizado de máquina, utilizados em prontuários eletrônicos, 1. extração, tratamento e análise dos dados de prontuários do Departamento; 2. construção e análise de modelos de vetorização de palavras relacionadas no contexto do banco de dados do Hospital São Paulo; 3. construção e validação dos modelos de predição de diagnósticos. **Resultados**: Os modelos de vetorização de palavras foram capazes de capturar a semântica de termos médicos e possibilitaram a construção de modelos de predição de diagnóstico, tornando o modelo de predição uma ótima ferramenta para auxiliar os profissionais de saúde. **Conclusão:** Os modelos de aprendizado de máquina mostraram resultados potenciais para auxiliar, como ferramentas de apoio, nos diagnósticos de pacientes oftalmológicos.

Abstract

Objective: To implement artificial intelligence routines through machine learning to construct diagnostic prediction models with data from electronic medical records of patients from the Department of Ophthalmology of Hospital São Paulo. **Method:** Preparation of a literature review of the main techniques and solutions of machine learning to use in electronic medical records, 1. extraction, treatment and analysis of data from medical records of the Department; 2. construction and analysis of vectorization models of related words in the context of the Database of Hospital São Paulo; 3. construction and validation of diagnostic prediction models. **Results:** The word vectorization models were able to capture the semantics of medical terms and enabled the construction of diagnostic prediction models, making the prediction model a great tool to assist health professionals. **Conclusion:** The machine learning models showed potential results to assist as diagnostic support tools of ophthalmologic patients.

Sumário

Lista de figuras

Lista de tabelas

Lista de abreviaturas, siglas e símbolos

AI	Inteligência Artificial
AM	Aprendizado de Máquina
API	*Application Programming Interface*
BD	Banco de Dados
CEP	Comitê de Ética em Pesquisa
CPU	*Central Processing Unit*
EPM	Escola Paulista de Medicina
GPU	*Graphics Processing Unit*
HSS	Hospital de São Paulo
IEEE	*Institute of Electrical and Electronic Engineers*
NLP	*Natural Language Processing*
PLN	Processamento de Linguagem Natural
SciELO	*Scientific Electronic Library Online*
Unifesp	Universidade Federal de São Paulo
V&V	Validação e Verificação
©	*Copyright*

1 INTRODUÇÃO

1.1 Introdução ao tema

Ao longo dos anos, a área da inteligência artificial desenvolveu várias técnicas de aprendizado de máquina (AM) para múltiplas aplicações (Kaur et al., 2016). A união das tecnologias computacionais empregadas no desenvolvimento de *softwares* utilizados na automação de prontuários médicos – também conhecidos como prontuário eletrônico –, e as tecnologias computacionais – conhecidas como aprendizado de máquina –, têm auxiliado na identificação de padrões de comportamento, relações causais de diferentes dados e grupos clínicos, até então despercebidas nos dados relacionados aos pacientes (Libralão et al., 2005).

Em sociedades orientadas pela tecnologia da informação, o conhecimento é um dos bens mais significativos de qualquer organização. Os papéis da tecnologia da informação nos cuidados de saúde estão bem estabelecidos. A gestão do conhecimento em cuidados de saúde oferece muitos desafios na criação, disseminação e preservação do conhecimento de saúde, usando tecnologias avançadas. O uso pragmático de sistemas de banco de dados, armazenamento de dados e tecnologias de gerenciamento de conhecimento pode contribuir muito para os sistemas de apoio à decisão nos cuidados de saúde (Ochoa Reyes et al., 2014).

1.2 Prontuário do paciente

O prontuário do paciente tem sido usado de forma constante desde 1.920, com o objetivo de armazenar todos os cuidados prestados ao paciente, gerando documentos para histórico clínico (Evans, 2016). Esses documentos contêm informações de resultados de exames, prescrições médicas, atendimentos, cirurgias e outras informações, servindo de memória para o médico e para todos os profissionais envolvidos (Carneiro, 2015).

O principal requisito para o prontuário do paciente é a obrigatoriedade de os registros médicos refletirem exatamente o curso da doença, indicar as suas possíveis causas e conter uma lista de atributos e valores, ordenados com a data do acontecimento de forma cronológica (Rada, 2008).

Os registros eram armazenados em papel e, somente na década de 60, surgiram os primeiros sistemas de informação hospitalares. Com o crescimento dos microcomputadores houve um impacto significativo nas aplicações de informática na área da saúde, as soluções de sistemas de informação em saúde estavam centradas em controles de estoque, prescrições e faturamentos. Basicamente, o uso de sistemas em saúde era para auxiliar a área administrativa dos hospitais, mas não as atividades clínicas.

Os registros médicos referentes ao paciente continuavam sendo armazenados em papel, ainda não existia estrutura que comportasse o registo médico de forma eletrônica. No final da década de 60, os sistemas hospitalares evoluíram e passaram a armazenar os dados do prontuário médico e, somente na década de 70 surgem os primeiros sistemas de prontuário eletrônico do paciente (Gubiani et al., 2003).

1.3 Aprendizado de máquina

O aprendizado de máquina é uma tecnologia ligada à área de Inteligência Artificial (AI) que torna possível a construção de *softwares*, com capacidade de identificar padrões em conjuntos de dados, permitindo que os *softwares* tomem decisões com base nas experiências adquiridas de forma automatizada (Monard, Baranauskas, 2017).

Assumindo que um *software* nada mais é que um sequenciamento de instruções, seguindo regras claras de avaliações de dados (algoritmos) e com o objetivo de resolver um problema, com base em entradas de dados, processamento dos dados pelos algoritmos e saídas com os resultados do processamento. Normalmente, existem vários algoritmos para resolver um mesmo problema, o desenvolvedor de *software* seleciona um determinado algoritmo, segundo as necessidades, desempenho e custo (Dietterich et al., 2010).

Comparando o aprendizado de máquina com as atividades de um desenvolvedor de *softwares* tradicional, é possível afirmar que os algoritmos utilizados na tecnologia de aprendizado de máquina permitem gerar sequências de avaliações de dados, onde a tecnologia de AM assume que, para algumas atividades, não há um algoritmo para resolver o problema. Assim, para resolução de um problema como AM

devemos utilizar um conjunto de exemplos de dados classificados para podermos ensinar a AM o que é esperado como resultado dos dados. Por exemplo: para dizer que um *e-mail* é *spam* usamos um conjunto de *e-mails* classificados, informando quais são *spam* e quais não são, ou seja, compensamos a falta do conhecimento dos algoritmos, pelos dados de exemplo para entrada e saída (Dietterich et al., 2010).

Nos últimos anos, o aprendizado de máquina vem sendo um tópico em alta quando se fala em ciência de dados e projetos de alta tecnologia. As primeiras aplicações, usando os conceitos de aprendizado de máquina, surgiram no início da década de 90, quando começaram os processos de mineração de dados e construção de *softwares* adaptativos (Sheth, 2017).

Existem dois tipos principais de modelos de aprendizado de máquina: os modelos de aprendizado supervisionados e os modelos não supervisionados (McCrea, 2017).

1.3.1 Aprendizado de máquina supervisionado

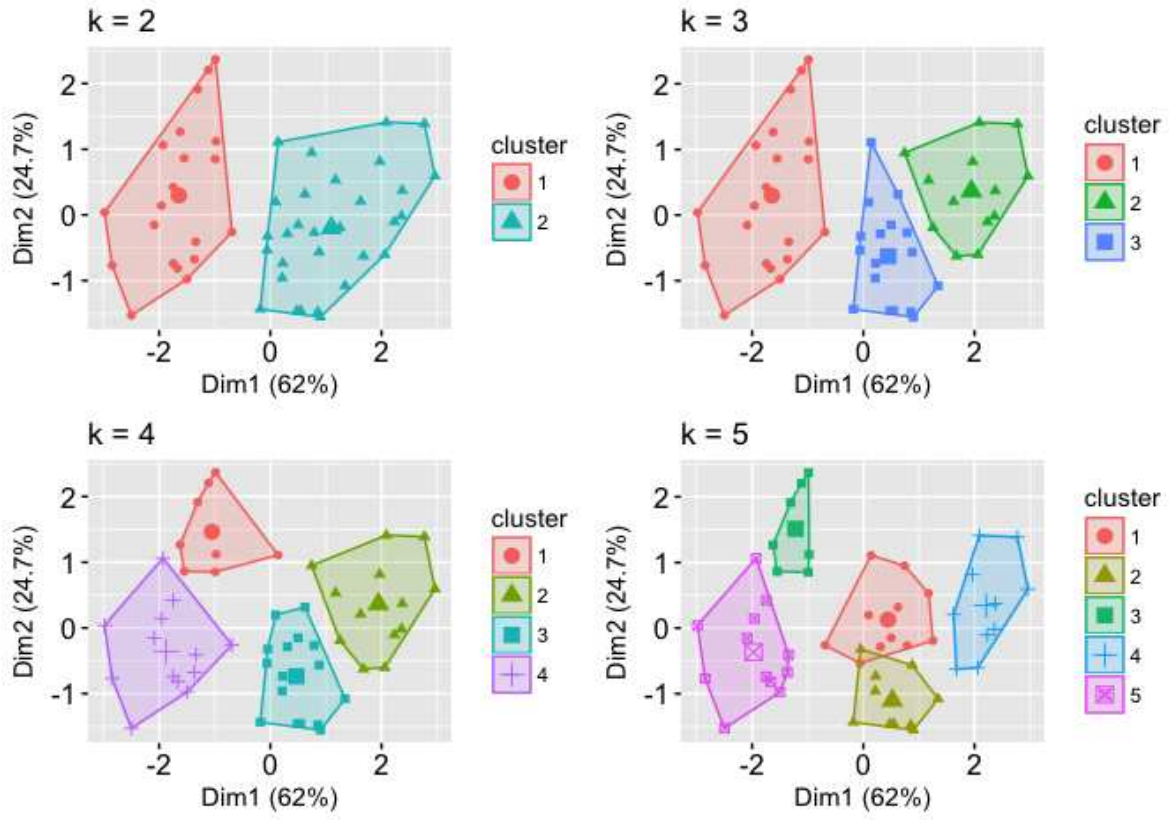
Na maioria dos modelos de aprendizado supervisionado, o objetivo é desenvolver uma função preditora de criação de hipóteses. Basicamente, para a criação de um modelo supervisionado são usadas múltiplas entradas de dados que já foram pré-determinadas (Camilo, da Silva, 2009).

No estudo feito sobre “Determinação de vícios refrativos oculares utilizando Support Vector Machines” foi usado um modelo supervisionado para fazer classificação de imagens oftalmológicas de pacientes com refrativos oculares (miopia, hipermetropia e astigmatismo), com o objetivo de auxiliar na identificação das lentes corretivas mais adequadas ao problema do paciente (Libralão, 2005).

O aprendizado supervisionado concentra-se na classificação, que envolve a escolha de subgrupos para melhor descrever uma nova característica de dados e hipóteses.

1.3.2 Aprendizado de máquina não supervisionado

O objetivo do aprendizado de máquina não supervisionado é de encontrar os relacionamentos entre os dados, identificar padrões e agrupamentos das informações. Para este modelo, não usamos exemplos de dados pré-determinados para fazer as classificações. Este modelo acaba se tornando uma tarefa mais difícil de ser desenvolvida, e o resultado do modelo não supervisionado acaba tendo seu desempenho avaliado por um modelo supervisionado subsequente (Chen et al., 2017). Entre as técnicas existentes, temos a de agrupamentos (*clusters*) (Figura 1). Os algoritmos usados por essa técnica para a classificação de um conjunto de dados têm como objetivo mostrar uma série de grupos com base nas semelhanças e diferenças entre os dados, permitindo a obtenção de grupos naturais de dados, de modo que os dados dentro de um grupo sejam muito semelhantes um com o outro e, ao mesmo tempo, sejam muitos diferentes de outros grupos de dados (Vieira, 2017).



Fonte: Disponível em: https://uc-r.github.io/kmeansclustering. Acesso em: 28 jun. 2020.
Figura 1. Exemplo de *cluster* de AM não-Supervisionado utilizando o algoritmo K-Means.

1.3.3 Principais ferramentas abertas para uso de aprendizado de máquina

As ferramentas para implementação de modelos de aprendizado de máquina em sua grande maioria, são feitas sobre as licenças de código aberto, permitindo a contribuição de qualquer pessoa para o projeto, assim, facilitando o crescimento dos projetos, melhoras em seus códigos fontes e criação de comunidades para discussões e resoluções de problemas.

Algumas empresas de *software* tradicionais vinham guardando os códigos de suas ferramentas de aprendizado de máquina como se fossem um segredo do qual era um elemento essencial do seu modelo comercial. O *software* de código aberto, ao contrário, está disponível publicamente, e os aplicativos gerados são distribuídos livremente. Esta mudança na filosofia permite que os usuários colaborem para o desenvolvimento e na criação de *softwares* de código aberto, como o navegador de internet Mozilla Firefox e o sistema operacional Linux, que são *softwares* de código aberto amplamente populares.

1.3.3.1 Apache Mahout

O Apache Mahout (TM) foi construído por membros da equipe de desenvolvimento do Apache Lucene (ferramenta de pesquisa). O desenvolvimento teve início em 2008 com o objetivo de manter implementações robustas, documentadas e escaláveis. O Mahout se resume em uma estrutura de álgebra linear distribuída, matemática expressiva, projetada para permitir que matemáticos, estatísticos e cientistas de dados implementem rapidamente seus próprios algoritmos de aprendizado de máquina.

Normalmente, o seu uso vem em conjunto com outras tecnologias como *Apache Spark* ou *Apache Hadoop* que são ferramentas de alta *performance* para armazenamento e processamento de dados (Ingersoll, 2017).

1.3.3.2 *TensorFlow*

TensorFlow é umas das bibliotecas de aprendizado de máquina que mais vem crescendo em uso e conteúdo (Github, 2017), exemplificando as rotinas de implementações. O *TensorFlow* foi originalmente desenvolvido pelo *Google*, com arquitetura flexível que permite a implantação em *Central Processing Unit* (CPU) e *Graphics Processing Unit* (GPU), em computadores pessoais, servidor ou dispositivo móvel com uma única *Application Programming Interface* (API) de uso. O *TensorFlow* foi projetado para ser uma biblioteca poderosa de rede neural, mas também é possível criar outros algoritmos de aprendizado de máquina, como árvores de decisão e outros (Shaikh, 2016).

As grandes vantagens no uso do *TensorFlow* são: destacadas por possuir:

a) Implementação intuitiva, e permitir facilmente a visualização dos modelos gráficos por meio de sua ferramenta TensorBoard;
b) Treinamentos com uso de CPU/GPU em computação distribuída;
c) Flexibilidade das plataformas, como servidores e aplicações para dispositivos móveis.

1.3.4 Uso de aprendizado de máquina na área da saúde

O crescimento na geração de dados relacionados à saúde apresentou grandes oportunidades para melhorar o acompanhamento da saúde dos pacientes. O aprendizado de máquina tem desempenhado um papel essencial na área da saúde e está sendo cada vez mais aplicado aos cuidados de saúde, incluindo a segmentação de imagens médicas, apoio a diagnóstico e comparação de dados históricos de prontuários (McCrea, 2017). O aprendizado de máquina está sendo usado para a análise da importância dos parâmetros clínicos e das suas combinações para o prognóstico, previsão da progressão da doença, extração de conhecimento médico, planejamento clínico e como ferramenta de suporte para terapias (Kaur et al., 2016).

Alguns estudos mostram que o aprendizado de máquina tem ajudado na monitoração de pacientes em casos delicados, de acordo com Guide et al. (2014), no

artigo intitulado “A machine learning system to improve heart failure patient assistance”. Ele trata sobre assistência a pacientes com insuficiência cardíaca. No artigo é apresentado um sistema de apoio à decisão clínica, que combina o uso de um aparelho portátil para coleta de informação remotas de parâmetros clínicos, possibilitando o monitoramento. As informações extraídas do aparelho portátil permitiram a construção de sistemas de aprendizado de máquina que permitem uma melhor análise dos dados capturados. Depois que os modelos de aprendizado de máquina estavam suficientemente treinados, a ferramenta foi usada para exibir saídas inteligentes para avaliação da gravidade do caso do paciente, identificação do tipo de insuficiência cardíaca, predições, comparações cronológicas e prognóstico baseado em pontuações (Guidi et al., 2014).

Os algoritmos de aprendizado de máquina são eficazes em reconhecer padrões em conjuntos de dados complexos e acaba se categorizando com uma capacidade adequada para aplicações médicas, especialmente as que dependem de medições complexas. Como resultado, o aprendizado de máquina é frequentemente usado em diagnósticos e detecção de doenças, podendo contribuir com melhores decisões sobre os planos de tratamento para os pacientes, por meio de sistemas eficazes de saúde (Kaur et al., 2016).

1.3.4.1 Uso de aprendizado de máquina em Oftalmologia

A oftalmologia é uns dos melhores campos para uso das técnicas de inteligência artificial na área da saúde. Com o grande volume de dados de alta qualidade e os avançados ~~nos~~ métodos de exames por imagens, tornam a oftalmologia um grande candidato para aplicações inteligentes (Khare et al., 2017).

Um estudo feito pela *Google Research* demonstra a precisão alcançada no uso desta tecnologia. Essa pesquisa mostra a interpretação de 128.175 imagens de fundo de olho, classificadas por oftalmologistas, representadas por vários estágios da retinopatia diabética, são eles: sem retinopatia diabética, leve, moderada, grave, proliferativa, edema macular diabético recomendável e retinopatia diabética recomendável. Nas configurações do algoritmo, foi usada uma alta sensibilidade para permitir descartar, de forma confiável, os casos negativos (tendo um baixo nível de

falsos positivos). Nesta configuração, o algoritmo alcançou 96-97% de sensibilidade e 93% de especificidade (Qi, 2017).

Em sua grande maioria, o uso de aprendizado de máquina na área oftalmológica costuma usar resultados de exames de imagem, modelo supervisionado de aprendizado de máquina com pré-classificação das informações para apoio diagnóstico (Ohsugi et al., 2017).

2 OBJETIVOS

2.1 Objetivo geral

Proposta de aplicar Rotinas de Aprendizado de Máquina em Prontuário Eletrônico para apoio a diagnósticos de pacientes em oftalmologia.

2.2 Objetivos específicos

- Analisar e aplicar transformação dos dados;
- Aplicar rotinas de Aprendizado Máquina na base de dados;
- Elaborar métodos para processamento de dados.

3 REVISÃO DA LITERATURA

3.1 Revisão da literatura e seleção das técnicas, segundo as palavras-chave

Este estudo é baseado em revisão de escopo, avaliando o uso de técnicas de aprendizado de máquina com aplicação em saúde. A pesquisa por artigos foi realizada nas bases de dados PubMed[1], Scientific Electronic Library Online (SciELO)[2] e Institute of Electrical and Electronic Engineers (IEEE)[3]. Como critério de inclusão, decidiu-se por artigos publicados no período de 2009 a 2019. As buscas foram realizadas entre 13 de maio de 2018 até 13 de maio de 2020.

Esta metodologia apresenta similaridades com uma revisão sistemática, dados pelos seguintes tópicos: (1) identificação da pergunta norteadora, (2) identificação de estudos relevantes, (3) seleção dos estudos e (4) mapeamento dos dados.

3.2 Descritores (fontes: MeSH e DeCS):

Os termos utilizados foram: “Electronic Health Records” + “Machine Learning”, “Clinical Text” + “Machine Learning”, “Medical Records” + “Machine Learning”, “Electronic Health Records” + “Natural Language Process“, “Clinical Text” + “Natural Language Process“, “Medical Records” + “Natural Language Process“, “Electronic Health Records” + “Neural Network“, “Clinical Text” + “Neural Network“ e “Medical Records” + “Neural Network“.

3.3 Seleção das técnicas, segundo a revisão da literatura

No retorno da busca pelos termos descritos na etapa anterior, com base na leitura do resumo e similaridade com o objetivo deste estudo, foram selecionados cinco artigos para leitura completa. Essa seleção está demonstrada na tabela 1, com as respectivas técnicas utilizadas em cada artigo.

[1] www.ncbi.nlm.nih.gov/pubmed
[2] www.scielo.org
[3] ieeexplore.ieee.org

3.4 Análise das informações de prontuário eletrônico contidos no banco de dados do Hospital São Paulo

O Departamento de Tecnologia da Informação do Hospital São Paulo disponibilizou um usuário de banco de dados com acesso às tabelas de Anotações Clínicas do Prontuário Eletrônico da instituição, mediante a autorização de uso (Anexo 1), para a análise e criação das rotinas de exportação dos dados em uma nova estrutura de banco de dados.

3.5 Tratamento e processamento dos dados do Departamento de Oftalmologia e Ciências Visuais existentes no banco de dados do Hospital São Paulo

O Departamento de Tecnologia da Informação do Hospital São Paulo disponibilizou acesso a nuvem *Google Cloud*® da instituição, para uso do banco de dados *Google Big Query*® para armazenar as Anotações Clínicas reprocessadas, com acesso a ferramenta de visualização *Google Data Studio*®. Para a etapa de processamento e transporte dos dados para a nuvem, o departamento também liberou uma Máquina Virtual no Data Center interno da Universidade Federal de São Paulo (Unifesp), com acesso ao banco de dados *Oracle*® para a leitura dos dados.

3.6 Criação e análise dos Modelos de Linguagem Natural

Para a construção do modelo de palavras correlacionadas foram utilizadas 1.789.584 anotações clínicas do prontuário do HSP, com a biblioteca *python gensim* (Gensim, 2009).

Foram utilizados oito diagnósticos localizados na Revista Brasileira de Oftalmologia (Vargas, Rodrigues, 2010) para a avaliação dos modelos de *Natural Language Processing* (NLP), tendo como critério de análise as cinco primeiras palavras com maior nível de acurácia relacionadas a cada diagnóstico.

3.7 Seleção de diagnósticos na base de dados do Departamento de Oftalmologia e Ciências Visuais

Foram selecionados 15 diagnósticos da tabela de Classificação Internacional de Doenças (CID-10), com base na frequência de registros de Anotações Clínicas que apresentaram os diagnósticos.

3.8 Criação e análise dos modelos de predição de diagnósticos

Os modelos de predição de diagnóstico foram construídos, utilizando a técnica de validação cruzada com divisão da base de dados em três conjuntos, os algoritmos utilizados foram *K-Nearest Neighbors*, *Random Forest*, *Naive Baye*, *Neural Network* e *Gradient Boosting*, e para cada algoritmo foram criados dois modelos, um utilizando a técnica de *One Hot Encoder* e outro com *Label Encoder*.

Foram criadas tabelas de comparação dos modelos, as análises foram agrupadas por técnicas e algoritmos, aplicando a métrica de acurácia de cada modelo, método e algoritmo. Também foram utilizados os relatórios das validações cruzadas, apresentando cada um dos três conjuntos de dados e o valor médio entre as validações de cada modelo final.

4 MÉTODOS

4.1 Método de abordagem

O presente trabalho de pesquisa foi realizado no período de 13 maio de 2018 até 13 maio de 2020, e validado no Departamento de Oftalmologia e Ciências Visuais da Escola Paulista de Medicina (EPM) da Unifesp, de abril de 2019 até novembro de 2019.

4.2 Aspectos éticos

Este trabalho foi submetido e aprovado pelo Comitê de Ética em Pesquisa (CEP) da EPM/Unifesp, na reunião de 17 de abril de 2018, sob o número 8520080418 (Anexo 1).

4.3 Fluxograma

5 RESULTADOS

5.1 Artigos selecionados

Com base na leitura dos resumos dos artigos encontrados nas bases de dados acadêmicas, foram selecionados cinco artigos com maior relevância com o estudo para leitura completa e tabelamento das ferramentas e técnicas utilizadas (Tabela 1).

Tabela 1. Artigos localizados e selecionados de bases de dados acadêmicas

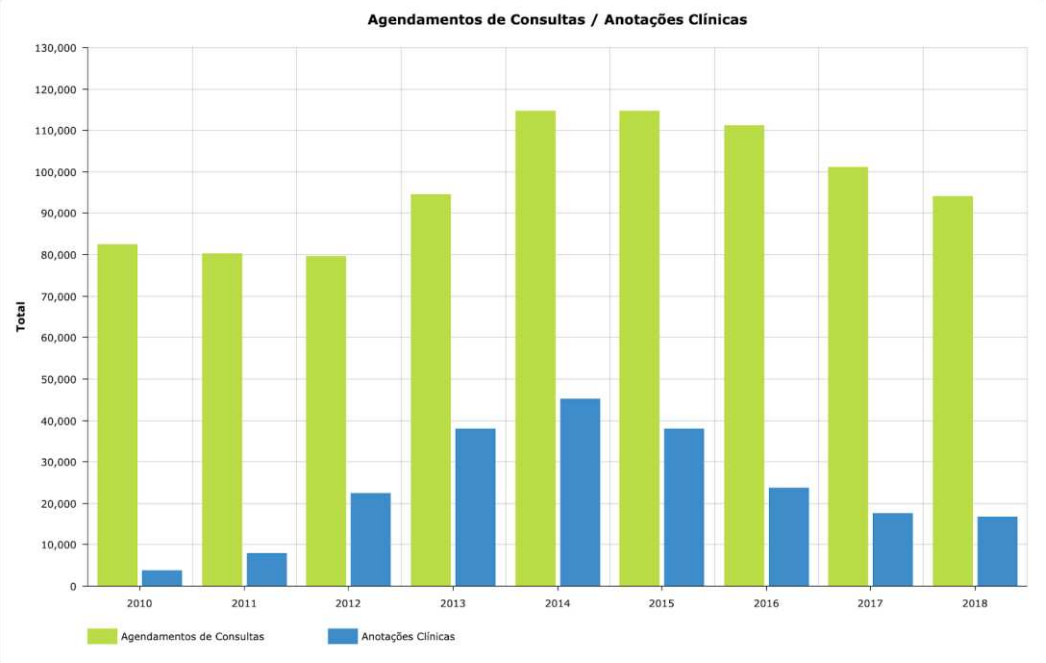
Título do Artigo	Técnicas / Ferramentas Utilizadas
A comparison of word embedding for the biomedical natural language processing (Wang Y et al., 2018)	Processamento de Linguagem Natural, Vetorização de Palavras, Rede Neural.
Embedding methods for EHR data and the utility of clinical text (Corbin, Machiraju, 2017)	Processamento de Linguagem Natural, Vetorização de Palavras, Rede Neural.
Predicting age by mining electronic medical records with deep learning characterizes differences between chronological and physiological age (Wang Z et al., 2017)	Processamento de Linguagem Natural, Vetorização de Palavras, Rede Neural.
Learning effective embeddings from medical notes (Dubois, Romano, 2017)	Processamento de Linguagem Natural, Vetorização de Palavras, Rede Neural.
What can natural language processing do for clinical decision support? (Demner-Fushman et al., 2009)	Processamento de Linguagem Natural

Fonte: Próprio autor.

5.2 Análise das Informações de Prontuário Eletrônico contidos no banco de dados do Hospital São Paulo

O período utilizado na filtragem dos dados foi de 31/12/2018 até 01/01/2010. Neste período, o departamento de Oftalmologia do HSP registrou 872.024 agendamentos de consultas. Desses agendamentos, foram registradas 212.588 anotações clínicas em prontuário eletrônico. Os dados contam com registros de anotações de 62.641 pacientes, contendo 94.069 diagnósticos e 13.498 medicamentos registrados nas anotações clínicas (Figura 2).

Os dados coletados incluem atendimentos de seis especialidades, 43 subespecialidades, com total de 212.588 registros de atendimentos registrados em prontuário eletrônico.



Fonte: Próprio autor.

Figura 2. Total de anotações clínicas e agendamentos de consultas registrados no Departamento de Oftalmologia do Hospital São Paulo.

Tabela 2. Resumo dos volumes de dados utilizados para o estudo

Agendamento de Consultas	872.024
Anotações Clínicas em Prontuário Eletrônico	212.588
Pacientes	62.641
Especialidades	6
Subespecialidades	43
Diagnósticos	94.069
Medicamentos	13.498

Fonte: Próprio autor.

Tabela 3. Registros de atendimentos por Especialidades

Especialidade	**Registros de Atendimentos**
Oftalmologia Geral	133.812
Pré-Anestésico	6.275

Fonte: Próprio autor.

Tabela 4. Subespecialidades com o maior volume de Registros de Atendimentos

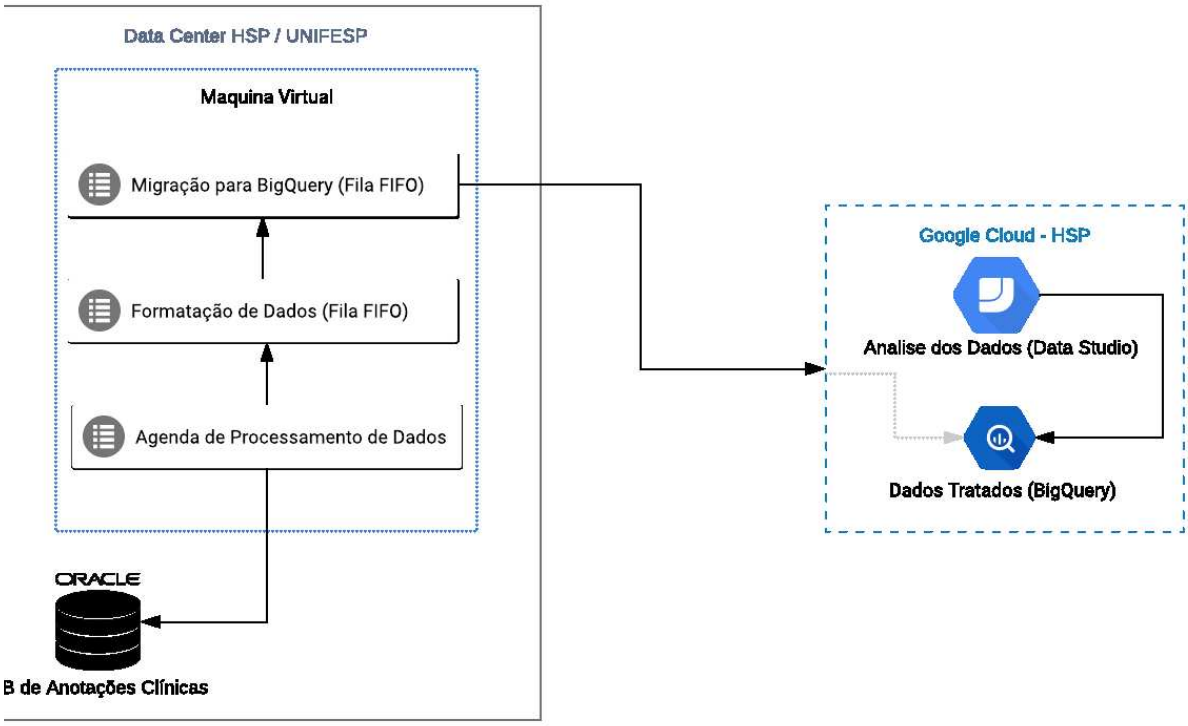
Subespecialidade	Registros de Atendimentos
Catarata	31.386
Externa Córnea	26.258
Geral	22.893
Uvea	18.914
Nominal	14.427
Catarata Congênita	1.385
Glaucoma	40.037
Refração	2.871

Fonte: Próprio autor.

5.3 Ajustes dos dados do Departamento de Oftalmologia e Ciências Visuais existentes no banco de dados do Hospital São Paulo

Os atendimentos médicos registrados em prontuário eletrônico utilizados para o estudo estavam armazenados em banco de dados em formato particularizado pelo *software* utilizado para gerenciar as informações. Para uso dos dados, foi necessária a construção de uma ferramenta para tratamento, consistência, formatação dos dados e transformação para um novo formato que permitisse o andamento do projeto.

Os dados tratados foram armazenados no banco de dados *BigQuery*® guardando as informações: Idade, Sexo, Especialidade, Subespecialidade, Perfil de atendimento, Medicamentos, Exames realizados, Procedimentos realizados, Diagnósticos e Anotações clínicas. As Anotações clínicas foram armazenadas na nova estrutura em dois formatos; um formato tabular com as perguntas e respostas e no formato de documento, contendo toda a Anotação Clínica do atendimento.



Fonte: Próprio autor.
Figura 3. Diagrama de arquitetura da rotina de extração, tratamento, consistência, formatação e migração dos dados.

O processamento foi divido em três etapas: (1) A etapa inicial ficou responsável por agendar as rotinas de leitura dos dados do Prontuário Eletrônico em lotes de registros; (2) Na segunda etapa foi aplicada a transformação dos dados para uma nova estrutura de modelo de dados; (3) A terceira etapa ficou responsável por transferir e armazenar os dados na nuvem do HSP, utilizando a base de dados *Google Big Query®*.

5.4 Criação e análise dos Processamentos de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é necessário para criar a representação contextual das informações utilizadas para o processo de aprendizado. O método utilizado no projeto foi o *Word2Vec* (criação de vetores de palavras relacionadas), que permitiu a representação das ocorrências das palavras em um mesmo contexto, tornando as palavras em agrupamentos com significado semântico.

O *Word2Vec* é utilizado para representação vetorial de palavras, permitindo a detecção de propriedades semânticas e relações linguísticas entre palavras por meio de uma rede neural profunda (*Deep Neural Networks*).

Para o projeto foi criado uma tabela de comparação de desempenho de modelos de *Word2Vec* já treinados, com a versão do modelo que foi criado a partir do Banco de Dados do HSP para, posteriormente, serem utilizados na criação dos modelos de classificação, do Departamento de Oftalmologia.

5.4.1 Seleção dos critérios de análise

Para a análise dos modelos, como critério de desempenho, foram selecionados oito diagnósticos, comparando qual modelo apresenta as 5 melhores palavras relacionadas. Os diagnósticos foram selecionados a partir de um artigo da Revista Brasileira de Oftalmologia (Vargas, Rodrigues, 2010), em que aparecem como principais diagnósticos em um estudo de perfil de atendimento de pacientes oftalmológicos de uma Unidade Mista de Saúde, na cidade de São Paulo.

Tabela 5. Diagnósticos utilizados para avaliação dos modelos/dicionários de vetores de palavras

Diagnósticos
Blefarite
Pterígio
Catarata
Conjuntivite
Ambliopia
Glaucoma
Estrabismo
Hordéolo

Fonte: Próprio autor.

5.4.2 Criação do Modelo/Dicionário com base de dados do HSP

Foram utilizadas 1.789.584 Anotações Clínicas do prontuário do HSP, com unidades e especialidades variadas, incluindo as Anotações Clínicas de Oftalmologia.

Para a criação do modelo, foi utilizada a biblioteca *python gensim* (Gensim,

2009), com os parâmetros: tamanho do vetor 300, tamanho da janela como 10 e mínimo de ocorrência de palavras como 2. Como resultado, foi gerado um dicionário com 305.812 palavras.

5.4.3 Comparação dos modelos do *Word2Vec*

Tabela 6. Comparação entre os modelos Wikipédia© e Prontuário HSP de *Word2Vec*

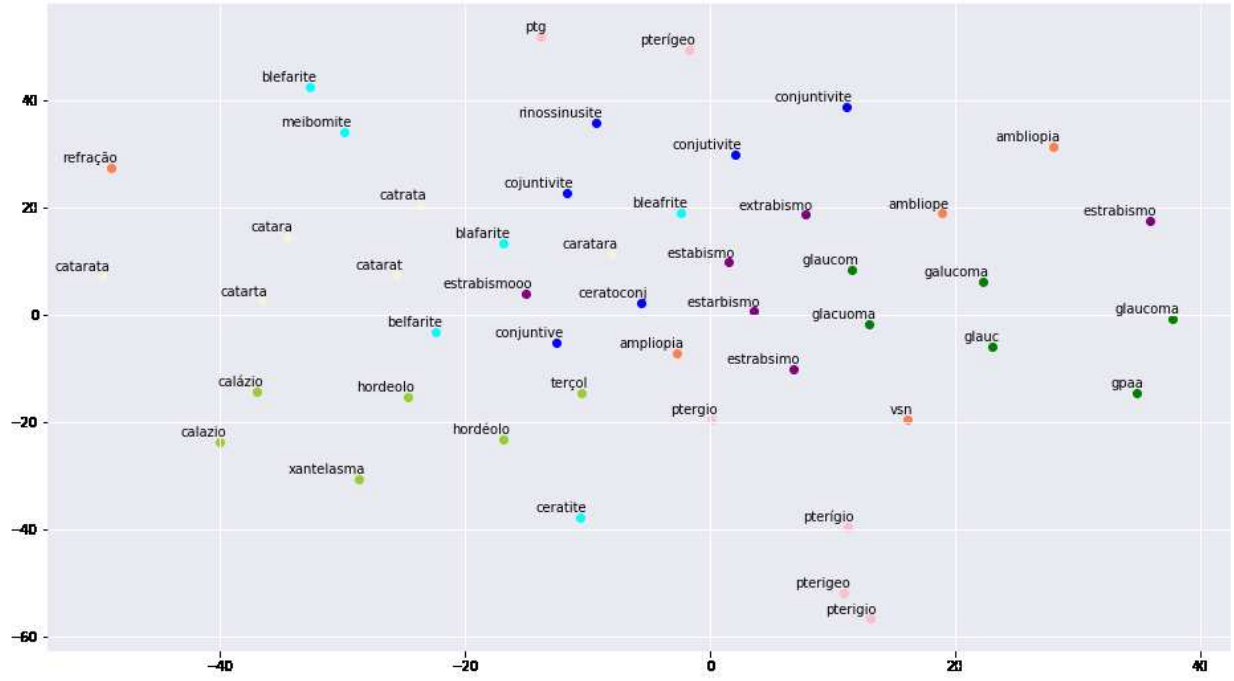
(continua)

Diagnóstico	Modelo com Dados do Wikipédia©	Modelo com Dados do HSP
Blefarite	Blefarites	Meibomite
	Blefarite	Belfarite
	Blefarospasmo	Blafarite
	Blefaroespasmo	Bleafrite
	Ceratite	Ceratite
Pterígio	Pterigóide	Pterigeo
	Pterigomaxilar	PTG
	Pterigóidea	Pterígio
	Pterigopalatino	Ptergio
	Pterigoide	Pterígeo
Catarata	Catarata	Catara
	Cataratas	Catarta
	Cataratas	Catarat
	Glaucoma	Caratara
	Córnea	Catrata
Conjuntivite	Conjuntivite	Conjutivite
	Conjutivite	Conjuntive
	Rinoconjuntivite	Cojuntivite
	Ceratoconjuntivite	Rinossinusite
	Ceratoconjuntivite	Ceratoconj
Ambliopia	Ambliopia	Estrabismo
	Estrabismo	VSN
	Presbiopia	Refração
	Hipermetropia	Ampliopia
	Retinopatia	Ambliope

(continuação)

Glaucoma	Glaucoma	Galucoma
	Estrabismo	Glaucom
	Ceratocone	Glacuoma
	Retinopatia	GPAA
	Oftalmológico	Glauc
Estrabismo	Estrabismo	Extrabismo
	Estrabismos	Estarbismo
	Astigmatismo	Estrabismooo
	Ambliopia	Estrabsimo
	Glaucoma	Estabismo
Hordéolo	hordéolo	Calazio
	Terçol	Hordéolo
	Maléolo	Calázio
	Lagoftalmo	Terçol
	Blefarite	Xantelasma

Fonte: Próprio autor.



Fonte: Próprio autor.

Figura 4. Gráfico de correlação de palavras *Word2Vec* com eixo X e Y normalizados de -60 a 60.

5.5 Seleção de diagnósticos na base de dados do Departamento de Oftalmologia e Ciências Visuais

O Departamento de Oftalmologia possui 94.069 diagnósticos em seus registros de Prontuário Eletrônico, com uma variedade de 852 diagnósticos únicos, utilizando a tabela CID-10.

Foram selecionados 15 diagnósticos com o maior número ocorrências em anotações clínicas e agrupados pelo CID-10 principal, para minimizar a variabilidade dos registros.

Tabela 7. Grupos de diagnósticos utilizados como variável target na criação e análise dos modelos de aprendizado de máquina

Catarata	Transtornos da Refração e da Acomodação	Outros Transtornos da Pálpebra
Outros Transtornos do Cristalino	Conjuntivite	Outros Transtornos da Retina
Outros Transtornos do Nervo Óptico e das Vias Ópticas	Transtornos da Retina em Doenças Classificadas em Outra Parte	Glaucoma
Estrabismo	Outras Inflamações da Pálpebra	Outros Transtornos da Conjuntiva
Distúrbios Visuais	Ceratite	Órgãos e Tecidos Transplantados

Fonte: Próprio autor

5.6 Criação e análise dos modelos de predição de diagnósticos

Para construção dos modelos de predição foram utilizadas as variáveis: gênero, idade do paciente na consulta, especialidade, subespecialidade e um vetor de tamanho 300 com o valor médio dos índices de palavras correlacionadas das anotações clínicas e como variável *target*, o diagnóstico.

Na construção dos modelos foram utilizados cinco algoritmos: *K-Nearest Neighbors*, *Random Forest Classifier*, *Naive Bayes*, *Neural Network Multi-layer Perceptron Classifier* e *Gradient Boosting*.

5.6.1 Transformação dos dados

Para a transformação dos dados foram feitos os tratamentos seguindo duas técnicas: *Label Encoder* e *One Hot Encoder*. Ambos possuem o objetivo de transformar as variáveis categóricas (sexo, especialidade, subespecialidade) em numéricas, para conseguirmos utilizar na construção do modelo, tendo em mente que os algoritmos de aprendizado de máquina aceitam apenas números com valores de variáveis explicativas.

Tabela 8. Exemplo de transformação de dados de gênero do paciente, dados antes da transformação

Gênero
M
F
F
F

Fonte: Próprio autor

Tabela 9. Exemplo de transformação de dados de gênero do paciente com *Label Encoder*

Gênero
0
1
1
1

Fonte: Próprio autor

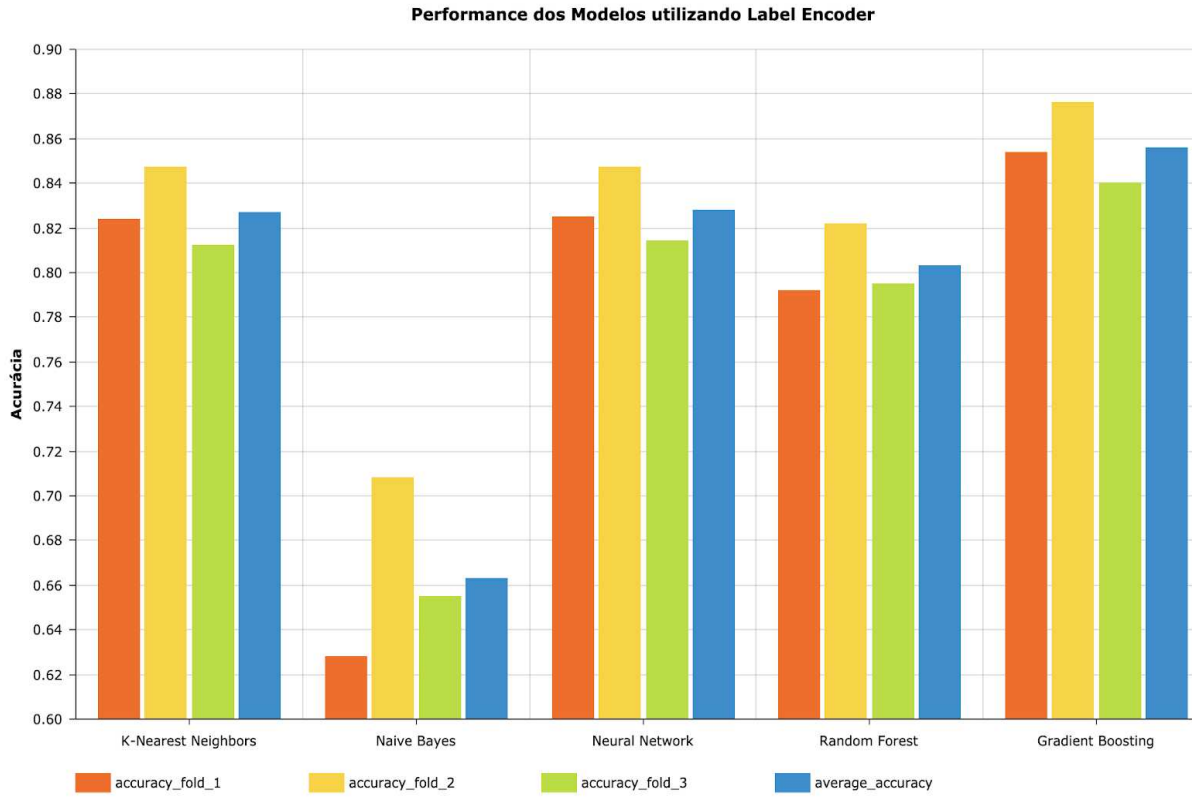
Tabela 10. Exemplo de transformação de dados de gênero do paciente com *One Hot Encoder*

M	F
1	0
0	1
0	1
0	1

Fonte: Próprio autor

Foram gerados dois conjuntos de dados, um com *Label Encoder* e outro com *One Hot Encoder* para validação e verificação da *performance* dos modelos na etapa subsequente.

5.6.2 *Performance* dos modelos, utilizando *Label Encoder*



Fonte: Próprio autor

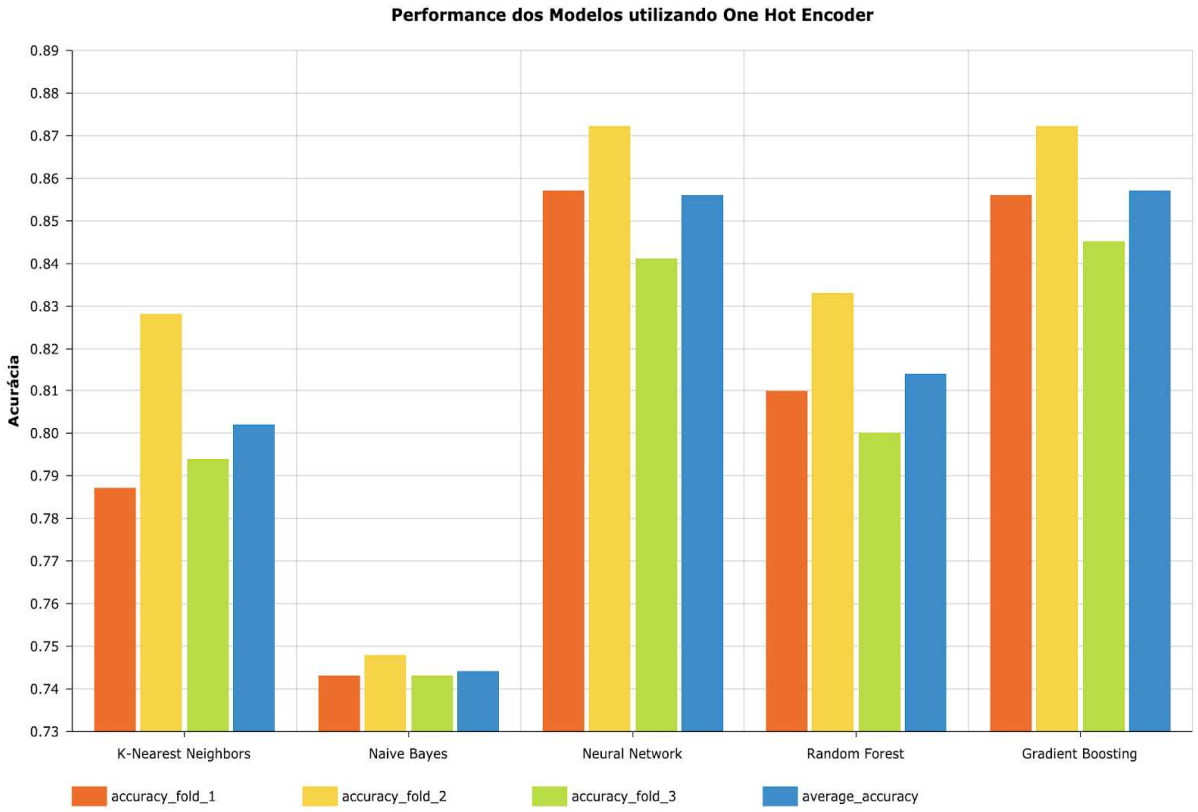
Figura 5. Gráfico de acurácia de validação cruzada dos modelos de diagnósticos, utilizando *Label Encoder.*

Tabela 11. Análise da *performance* dos modelos utilizando *Label Encoder*

Algoritmos	Acurácia Média
K-Nearest Neighbors	0.82
Naive Bayes	0.66
Neural Network	0.83
Random Forest	0.80
Gradient Boosting	0.85

Fonte: Próprio Autor

5.6.3 *Performance* dos modelos, utilizando *One Hot Encoder*



Fonte: Próprio Autor

Figura 6. Gráfico de acurácia de validação cruzada dos modelos de diagnósticos, utilizando *One Hot Encoder.*

Tabela 12. Análise da *performance* dos modelos utilizando *One Hot Encoder*

Algoritmos	Acurácia Média
K-Nearest Neighbors	0.80
Naive Bayes	0.74
Neural Network	0.85
Random Forest	0.81
Gradient Boosting	0.85

Fonte: Próprio Autor

6 DISCUSSÃO

Este estudo teve como objetivo a aplicação dos métodos de aprendizado de máquina na base de dados de Anotações Clínicas de Prontuário Eletrônico do Departamento de Oftalmologia do Hospital São Paulo, para serem avaliadas a compatibilidade e a convergência das informações com os métodos selecionados.

6.1 Extração, tratamento e análise dos dados

O formato de armazenamento dos dados pelo *software* de Prontuário Eletrônico tornou o processo de extração custoso para implementação do estudo. Foi necessária a criação de rotinas de extração e consistências das informações. Em alguns casos, os dados tiveram que ser agrupados em uma mesma especialidade ou subespecialidade como por exemplo as especialidades: Oftalmologia Geral, Geral e Oftalmologia e suas subespecialidades como: Oftalmología, Geral, R1, R2, R3 e outras, sendo possível notar que algumas informações eram apenas para gestão administrativa da instituição.

Com a análise dos dados extraídos é possível afirmar que menos de 25% das Anotações Clínicas de atendimentos de consultas são registradas em sistema eletrônico, gerando uma limitação dos dados para uso no estudo.

6.2 Construção e análise de modelo de vetorização de palavras relacionadas em contexto do banco de dados do Hospital São Paulo

O modelo de vetores de palavras relacionadas gerado a partir dos dados de Anotações Clínicas do Hospital São Paulo, apresentou a capacidade de identificação dos diagnósticos, abreviações e também as relações entre as palavras com erros de digitação e ortografia.

Em comparação com o modelo de vetores de palavras gerado com dados do *Wikipedia*, o modelo do Hospital São Paulo apresentou uma relação de diagnósticos que melhor descreve os dados clínicos.

6.3 Construção e análise dos modelos de aprendizado de máquina para predição de diagnósticos

Para a construção dos modelos de predição de diagnóstico foi necessário criar agrupamentos de diagnóstico, utilizando a tabela de CID-10, para reduzir a variabilidade. A base de dados estava com muitos diagnósticos, tinham um volume muito pequeno de anotações clínicas, e gerava muita variação nos modelos, produzindo uma redução na acurácia. Em alguns casos, a acurácia chegou a ser menos da metade da acurácia final.

Apesar do modelo de predição de diagnóstico ter um nível de acurácia de até 0,85 na predição dos diagnósticos selecionados, ele apenas tem o objetivo de apresentar o diagnóstico com maior relação com a anotação clínica, utilizando como referências os dados históricos do departamento. O modelo não tem o propósito de diagnosticar o paciente. Logo, o modelo de predição de diagnóstico é uma ferramenta complementar de apoio ao diagnóstico para o profissional médico.

7 CONCLUSÕES

7.1 Geral

Os modelos de aprendizado de máquina demonstraram ser um conjunto de ferramentas em potencial, quando aplicado em base de dados de anotações clínicas de oftalmologia, podendo auxiliar na hipótese diagnóstica e de análises contínuas no cuidado clínico do paciente.

7.2 Modelo de vetores de palavras relacionadas

A comparação entre os modelos de vetores de palavras (W2V) de textos de anotações clínicas e textos do Wikipédia© demonstrou que os modelos com textos de anotações clínicas podem capturar a semântica de termos médicos melhor que aqueles treinados com dados do Wikipédia©, com base na avaliação dos diagnósticos selecionados.

7.3 Modelo de predição de diagnósticos

O modelo de predição de diagnóstico por anotações clínicas mostrou resultados potenciais para ser uma ferramenta de apoio de diagnósticos de pacientes oftalmológicos, com uma média de acurácia de 0,85 para o algoritmo Gradient Boosting, utilizando o método de *One Hot Encoding* nos diagnósticos pré-selecionados.

8 REFERÊNCIAS

Camilo CO, da Silva JC. Mineração de dados: conceitos, tarefas, métodos e ferramentas [relatório técnico na internet]. Goiânia: Universidade Federal de Goiás; 2009 [citado 28 nov. 2020]. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf

Carneiro SD. Prontuário eletrônico do paciente: percepção de aceitação e facilidade de uso de profissionais da área de saúde [dissertação na internet]. Belo Horizonte: Fundação Mineira de Educação e Cultura; 2015 [citado 12 fev. 2021]. Disponível em: https://repositorio.fumec.br/bitstream/handle/123456789/466/severino_carneiro_mes_sigc_2016.pdf?sequence=1&isAllowed=y

Chen CC, Juan HH, Tsai MY, Lu HH. Unsupervised learning and pattern recognition of biological data structures with density functional theory and machine learning. Sci Rep. 2018 Jan 11;8(1):557. doi: 10.1038/s41598-017-18931-5.

Corbin CK, Machiraju GB. Embedding methods for EHR data and the utility of clinical text [Internet]. Stanford, CA; 2017 [cited 2020 Nov 12]. Available from: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/custom/15846252.pdf

Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform. 2009 Out;42(5):760-72. doi: 10.1016/j.jbi.2009.08.007.

Dietterich T, Bishop C, Heckerman D, Jordan M, Kearns M, editors. Introduction to machine learning. 2nd. ed. Cambridge, MA: Massachusetts Institute of Technology**;** 2010. Available from: https://dl.matlabyar.com/siavash/ML/Book/ML%20Book_Alpaydin.pdf

Dubois S, Romano N. Learning effective embeddings from medical notes [Internet]. Stanford, CA; 2017 [cited 2020 Jul 22]. Available from: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2744372.pdf

Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform. 2016;9Suppl 1):S48-S61.

Gensim. Topic modelling for humans. [Internet]. Valleta, MT; 2009. [cited 2020 Jul 22]. Available from: https://radimrehurek.com/gensim/.

GitHub. Most active repositories. [Internet]. Cincinati, OH: GitHub; 2017 [cited 2020 Jan 12]. Available from: https://octoverse.github.com/.

Guidi G, Pettenati MC, Melillo P, Iadanza E. A machine mearning system to improve heart failure patient assistance. IEEE J Biomed Health Inform. 2014 Nov;18(6):1750-6. doi:10.1109/JBHI.2014.2337752.

Ingersoll G. Introducing Apache Mahout Scalable, commercial-friendly machine learning for building intelligent applications [Internet]. San Francisco, CA; 2017 [cited 2017 Dez 12]. Available from: https://www.ibm.com/developerworks/library/j-mahout/#artrelatedtopics.

Gubiani JS, da Rocha RP, D'Ornellas MC. Interoperabilidade semântica do prontuário eletrônico do Paciente. [Anais n internet]. II Simpósio de Informática da Região Centro/RS. Santa Maria, RS: 2003 [citado 21 abr. 2020]. Disponível em: https://www.lume.ufrgs.br/bitstream/handle/10183/554/000415301.pdf?sequence=1

Kaur H, Wasan S. Empirical study on applications of data mining techniques in healthcare. J Comput Sci. 2006;2(2):194-200. doi:10.3844/jcssp.2006.194.200

Khare A, Jeon M, Sethi IK, Xu B. Machine learning theory and applications for healthcare. J Healthc Eng. 2017;2017:5263570.doi:10.1155/2017/5263570

Libralão LG, Netto AV, de Carvalho APLF, de Oliveira MCF. Determinação de vícios refrativos oculares utilizando Support Vector Machines. Sba: Controle & Automação. 2005 Jun;16(2):146-58. doi: 10.1590/S0103-17592005000200004.

McCrea N. An introduction to machine learning theory and its applications: a visual tutorial with examples [Internet]. 2017. [cited 2020 Jan 12]. Available from: https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer.

Monard MC, Baranauskas JA. Conceitos sobre aprendizado de máquina. In: Rezende SO, organizadora. Sistemas inteligentes: fundamentos e aplicações. Barueri: Manole; 2003. p. 89-114.

Ochoa Reyes AJ, Orellana Garcia A, Sanchez Corales Y, Davila Hernandez F. Web component for the analysis of clinical information using the technique of clustering data mining. RCIM [Internet] 2014 [cited 2021 Mar 12];6(1):5-16. Available from: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1684-18592014000100002&lng=es&nrm=iso&tlng=en

Ohsugi H, Tabuchi H, Enno H, Ishitobi N. Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep. 2017 Aug 25;7(1):9425. doi: 10.1038/s41598-017-09891-x.

Qi SR. Deep learning in ophthalmology - How Google did it [Internet]. Health AI; 2017 [cited 22 Mar 2020]. Available from: https://medium.com/health-ai/deep-learning-in-ophthalmology-using-128-175-retinal-images-59814e8a3f68

Rada R. Information systems and healthcare enterprises. Hershey (NY): IGI Publishing; 2008.

Shaikh F. An introduction to implementing neural networks using TensorFlow in Python. [Internet]. 2016 Out 3 [cited 2017 Dec 12] Available from: https://www.analyticsvidhya.com/blog/2016/10/an-introduction-to-implementing-neural-networks-using-tensorflow/.

Sheth A. History of machine learning [Internet]. 2017 Ago 25 [cited 2018 Dec 12]. Available from: https://medium.com/bloombench/history-of-machine-learning-7c9dc67857a5.

Vargas MA, Rodrigues M de LV. Perfil da demanda em um serviço de oftalmologia de atenção primária**.** Rev Bras Oftalmol. 2010 Abr; 69(2):77-83 doi:10.1590 /S0034-72802010000200002

Vieira AFG. Técnicas de aprendizaje de máquina utilizadas para la minería de texto. Investigácion Bibliotecologica. 2017;31(71):103-26.

Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, et al. A comparison of word embeddings for the biomedical natural language processing. J Biomed Inform. 2018 Nov;87:12-20. doi: 10.1016/j.jbi.2018.09.008.

Wang Z, Li L, Glicksberg BS, Israel A, Dudley JT, Ma'ayan A. Predicting age by mining electronic medical records with deep learning characterizes differences between chronological and physiological age. J Biomed Inform. 2017 Dez;76:59-68. doi: 10.1016/j.jbi.2017.11.003.

ANEXOS

Anexo 1 – Comitê de Ética em Pesquisa

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO

COMITÊ DE ÉTICA EM PESQUISA

SP

São Paulo, 17 de abril de 2018
CEP N 8520080418

Ilmo(a). Sr(a).
Pesquisador(a): Lucas De Oliveira Batista Alves
Depto/Disc: Oftalmologia E Ciencias Visuais
Prof. Dr. Vagner Rogério Dos Santos (orientador)

Título do projeto: "Uso de Rotinas de Aprendizado de Máquina em Prontuário Eletrônico para Apoio a Diagnósticos de Pacientes Oftalmológicos".

Parecer Consubstanciado do Comitê de Ética em Pesquisa UNIFESP/HSP

Proposta de aplicação de Aprendizado Máquina em Prontuário Eletrônico para Apoio a Diagnósticos de Pacientes em Oftalmologia. Objetivos Específicos: – Elaborar avaliação sobre as principais técnicas e soluções para aprendizado de máquina (AM) em prontuários eletrônicos. Definição e especificações dos ambientes de desenvolvimento de AM e compatibilidade com o banco de dados (BD) do Hospital São Paulo (HSP) e Departamento de Oftalmologia. Estabelecer interface homem computador (IHM) da aplicação segundo BD do HSP e Departamento de Oftalmologia.Elaborar, rotinas, métodos de aplicação de AM.

O Comitê de Ética em Pesquisa da Universidade Federal de São Paulo/Hospital São Paulo, na reunião de 10/04/2018, **ANALISOU** e **APROVOU** o protocolo de estudo acima referenciado. A partir desta data, é dever do pesquisador:
1. Comunicar toda e qualquer alteração do protocolo.
2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do protocolo.
3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.
4. **Relatórios parciais** de andamento deverão ser enviados **anualmente** ao CEP até a conclusão do protocolo.

Atenciosamente,

Prof. Dr. Miguel Roberto Jorge

Coordenador do Comitê de Ética em Pesquisa da
Universidade Federal de São Paulo/Hospital São Paulo

Rua Botucatu, 572, 1º andar conjunto 14 CEP 04023-061 São Paulo - tel: 55 (11) 5571-1062 / fax: 55 (11) 5539-7162
Horário de atendimento: Segundas, Terças, Quintas e Sextas, das 9:00 as 13:00hs : e-mail:cep@unifesp.edu.br
CEP Nº 8520080418

Bibliografia consultada

Academia Brasileira de Letras. Vocabulário Ortográfico da Língua Portuguesa, Volp. Busca no vocabulário. [Internet]. 2016 Set [cited 2021 Feb 01]. Available from: http://www.academia.org.br/nossa-lingua/busca-no-vocabulario

Associação Brasileira de Normas Técnicas. NBR 14724: Informação e documentação: trabalhos acadêmicos: apresentação. Rio de Janeiro; 2011.

Fundação Instituto Brasileiro de Geografia e Estatística. Normas de apresentação tabular. 3a ed. Rio de Janeiro: IBGE; 1993. 62 p.

Normas para teses e dissertações [Internet]. 3a ed. São Paulo: Universidade Federal de São Paulo, Biblioteca Antônio Rubino de Azevedo, Coordenação de Cursos; 2021 [cited 2021 Jun 30]. Available from: http://www.bibliotecacsp.unifesp.br/Documentos-Apostila/normas-para-teses-e-dissertacoes

Pereira TA, Montero EFS. Terminologia DeCS e as novas regras ortográficas da língua portuguesa: orientações para uma atualização[Internet]. Acta Cir Bras [Internet]. 2016 [cited 2021 Jan 10];27(7):509-14. Available from: http://www.scielo.br/pdf/acb/v27n7/a14v27n7.pdf