

Aplicação de algoritmos de machine learning na área farmacêutica: uma revisão

Appliance of machine learning algorithm in the pharmaceutical sector: a review

Aplicación de algoritmos de machine learning en el área farmacéutica: revisión

Recebido: 05/11/2021 | Revisado: 12/11/2021 | Aceito: 13/11/2021 | Publicado: 22/11/2021

Tarcio Correia de Campos

ORCID: <https://orcid.org/0000-0002-5565-6467>

Centro Universitário UNIFAVIP-Wyden, Brasil

E-mail: tarciocorreia7@gmail.com

Tibério Cesar Lima de Vasconcelos

ORCID: <https://orcid.org/0000-0001-7177-0561>

Centro Universitário UNIFAVIP-Wyden, Brasil

E-mail: tiberio.vasconcelos@professores.unifavip.edu.br

Resumo

A indústria farmacêutica com toda sua importância vem inovando e revolucionando no decorrer do tempo. A tecnologia da informação e seus segmentos, tem um papel imprescindível para que as mudanças ocorram, e este projeto mostrará o cenário de crescimento da indústria farmacêutica e a importância de sua tecnologia no Brasil, no mundo e o uso de algoritmos como ferramentas essenciais em diversas áreas do campo farmacêutico. No decorrer deste projeto, até o seu fim, será apresentado pontos mostrando como está o cenário industrial, tecnologias farmacêuticas, algoritmos sendo imprescindíveis na resolução de problemas, alianças entre indústrias, inovações para novos medicamentos, atendimentos e tratamentos. Trata-se de uma revisão de literatura integrativa utilizando as plataformas Google Acadêmico, PubMed, Scielo e Science Direct para buscar de artigos do período de 2003 a 2021 sobre aplicação de algoritmos de machine learning na área farmacêutica. O uso de algoritmos se mostrou eficaz facilitando no desenvolvimento de novas drogas e para resolver problemas existentes.

Palavras-chave: Árvore de decisão; Regressão linear de mínimos; Regressão logística; Indústria farmacêutica; Naive bayes; Support vector machine.

Abstract

The pharmaceutical industry with all of its importance has been innovating and revolutionizing in the course of the time. The information technology on its segments has a crucial role so the changes can happen, and this project will show the growth situation of the pharmaceutical industry and the importance of its technology in Brazil, in the world and also the use of algorithm as essentials tools in several areas in the pharmaceutical field. During the course of this project, to its end, will be described details showing how is the industry's outlook, pharmaceutical technologies, algorithms being important keys at problem solving, partnerships between industries, innovations for medications, medical services and treatments. This is an integrative literature review using the Google Scholar, PubMed, Scielo and Science Direct platforms to search for articles from 2003 to 2021 on the application of machine learning algorithms in the pharmaceutical area. The use of algorithms proved to be effective, facilitating the development of new drugs and in solving existing problems.

Keywords: Decision tree; Linear regression of minimums; Support vector machine; Logistic regression; Naive bayes; Pharmaceutical industry.

Resumen

La industria farmacéutica con toda su importancia viene innovando y revolucionando en el curso del tiempo. La tecnología de la información y sus seguimientos, tiene un papel indispensable para que las mudanzas ocurran, y este proyecto mostrará el escenario de crecimiento de la industria farmacéutica y la importancia de su tecnología en Brasil, en el mundo y el uso de algoritmos como herramientas esenciales en muchos ámbitos del campo farmacéutico. En el transcurso del proyecto, hasta su conclusión, se presentará puntos mostrando como se encuentra el escenario industrial, tecnologías farmacéuticas, algoritmos siendo imprescindibles en la resolución de problemas, alianzas entre industrias, innovaciones para nuevos medicamentos, atendimientos y tratamientos. Se trata de una revisión integradora de literatura utilizando las plataformas Google Scholar, PubMed, Scielo y Science Direct para la búsqueda de artículos desde 2003 hasta 2021 sobre la aplicación de algoritmos de aprendizaje automática en el sector farmacéutico. El uso de algoritmos demostró ser eficaz, facilitando el desarrollo de nuevos fármacos y para resolver problemas existentes.

Palabras clave: Árbol de decisión; Regresión lineal de mínimos; Máquina de vectores de apoyo; Regresión logística; Bayes ingenuos; Industria farmacéutica.

1. Introdução

A Inteligência Artificial (IA) na indústria farmacêutica se refere ao sistema de tecnologias interconectadas e automatizadas na indústria de biotecnologia que podem funcionar de forma autônoma, com pouca ou nenhuma intervenção humana. IA é uma tecnologia emergente que está encontrando seu caminho em muitas facetas do setor farmacêutico, desde o desenvolvimento de medicamentos, o diagnóstico e até mesmo o atendimento ao paciente (Bulgaru, 2020).

De acordo com Bulgaru (2020), os executivos da indústria farmacêutica estão inteirados na busca de alavancar a IA nas linhas de negócios. Grandes empresas da área fazem parcerias não apenas em busca de descobrimento e desenvolvimento de novos medicamentos, mas também para testes clínicos e atendimentos avançados ao paciente.

Usando algoritmos automatizados para fazer tarefas que normalmente dependem da inteligência humana, a IA se apropria disto na indústria farmacêutica e biotecnológica, e redefine a forma que os cientistas desenvolvem medicamentos, lidando com doenças entre outros fatores. Ele também apresenta dados da Narrative Science, mostrando que 61% das empresas que investem em estratégias inovadoras estão usando IA para identificar oportunidades que, de outra forma, teriam perdido. Para empresas farmacêuticas que prosperam em inovação, esta é uma estatística importante a ser entendida (Arsene, 2020).

A IA pode otimizar a indústria farmacêutica através de aprimoramento a P&D (Pesquisa e Desenvolvimento), seja no começo do projeto, identificação de novas moléculas, podendo ir até a validação e descobertas de medicamentos com base em alvos. No entanto, faz com que não só reduza o tempo do teste, mas também acelere a aprovação. Ao permitir que a fabricação seja otimizada, tornando-se mais rápida e eficiente, a indústria farmacêutica pode se beneficiar grandemente (Tierney, 2020). E são esses processos externos de tecnologia, envolvendo licenciamentos, alianças estratégicas, compra de tecnologias e o desenvolvimento de tecnologias contratando serviços de outras empresas, universidades e centros de pesquisa, em alguns casos podem substituir o esforço tecnológico próprio, mas frequentemente complementam os esforços interno de P&D (Santos, 2020).

Os autores Lane e Lubatkin (1998), estudiosos de como uma empresa aprende com outra, na relação de empresa estudante e empresa professor, em suas ideias, mostram que uma das formas de sobreviver em muitos mercados são com as alianças de P&D (Apud Koerich et al., 2014).

O Brasil se insere no contexto global da indústria farmacêutica como um dos mercados mais dinâmicos da última década, apresentando taxas de crescimento de dois dígitos. Impulsionado pelos medicamentos genéricos e pela ascensão de um grande número de novos consumidores, o mercado farmacêutico brasileiro alcançou a sexta posição mundial em 2013 (Gomes *et al.*, 2014).

Em nosso país, iniciou em 2008 o interesse maior sobre o assunto, através da política industrial e tecnológica criada pelo Ministério da Saúde, a partir deste momento, surgiram as Parcerias para o Desenvolvimento Produtivo (PDP), que incluía laboratórios públicos da rede oficial e laboratórios privados nacionais e estrangeiros, tendo como objetivo nacionalizar a produção e reduzir os preços de aquisição de medicamentos considerados estratégicos para o Sistema Único de Saúde (SUS). Essa mudança visava diminuir altos preços de medicamentos no Brasil, transferindo a tecnologia de produção de medicamentos com patentes perto da expiração para laboratórios nacionais (Torres, 2015).

Este trabalho objetiva revisar na literatura a aplicação na indústria farmacêutica em relação ao desenvolvimento, crescimento tecnológico e aprendizado de máquinas e como estes algoritmos são utilizados.

2. Metodologia

Este trabalho se mostra como uma revisão da literatura integrativa (RI), levando em consideração que é um tipo de estudo em que há integração de informações de um determinado tema presentes em artigos primários, sintetizando e com a

possibilidade de trazer a resposta de uma pergunta de pesquisa (Whittemore; Knafl, 2005). Realizado através de levantamento de outros trabalhos encontrados em bases de dados selecionadas, em investigação do tema abordado, trazendo informações específicas sobre aplicação de algoritmos de machine learning na área farmacêutica. O estudo foi realizado no Centro Universitário do Vale do Ipojuca – UNIFAVIP, localizado na Av. Adjar da Silva Casé, nº 800 – Indianópolis 55.024-740 Caruaru – PE.

A coleta de dados foi realizada em artigos, livros, revistas, material disponibilizado na internet, com os seguintes descritores: árvore de decisão, regressão linear de mínimos, regressão logística, algoritmo, indústria farmacêutica, tecnologia, naive bayes e support vector machine. Os critérios de inclusão foram: artigos com texto completo disponível, artigos originais em português, inglês que abordassem a aplicação de algoritmos de machine learning na área farmacêutica. Foram excluídas obras duplicatas, cartas ao editor e qualquer outro tema que não abordasse o assunto, onde, inicialmente foram selecionados 63 artigos, após os critérios de exclusão ficaram 34 artigos abordando o tema. A literatura utilizada foi entre os anos 2003 e 2021.

3. Resultados e discussão

3.1 indústria farmacêutica

A indústria farmacêutica e a ciência de desenvolvimento de medicamentos precisam de forma urgente de novas tecnologias de transformação na era atual da saúde digital e da inteligência artificial (IA). Os algoritmos e a IA são descritos por muitas vezes como tecnologias revolucionárias e esta revolução vem acontecendo na indústria farmacêutica nos últimos 5 anos (Koromina *et al.*, 2019).

Vem ocorrendo uma rápida introdução de IA na área da saúde e principalmente em 2016 e 2017, várias empresas farmacêuticas realizaram investimentos em empreendimentos conjuntos com empresas de IA. O propósito é melhorar diagnósticos ou biomarcadores e a identificação de alvos de medicamentos e o desenvolvimento de novos medicamentos (Mak; Pichika, 2019).

Mak e Pichika (2019), mostraram que foram realizadas inúmeras parcerias entre indústrias farmacêuticas e empresas de IA a nível global, um exemplo foi a parceria entre a DeepMind Technologies, uma subsidiária do Google que colaborou com a Royal Free London NHS Foundation Trust para auxiliar no tratamento de lesão renal aguda. A transição da medicina geral para a saúde IA moderna se concentra na base dos dados. As análises desses dados subjacentes, juntamente com o aprendizado de máquina, são posteriormente formuladas em algoritmos - contribuindo fortemente para a saúde moderna progressiva que incorpora IA.

A indústria farmacêutica possui um setor com forte dinamismo, pois, se relacionam diretamente com processos tecnológicos e produtivos, caracterizadas pela constante procura de inovações e melhorias nos processos. Estados Unidos, União Europeia e Japão compõem a tríade dos lugares onde encontramos os principais produtores de inovação e possuem os maiores faturamentos, tendo maior parte do mercado. O setor farmacêutico é marcado por uma grande quantidade de patentes registradas, com o propósito de proteção dos laboratórios que fazem o lançamento dos medicamentos inovadores e querem exclusividade (Jesus, 2018).

Lobato e Ortiz (2019), mostram que a análise da quantidade de pedidos dos depósitos de patentes de tecnologia ou produção de tecnologia inovadora é um forte indicador internacional do desenvolvimento tecnológico mundial. Em 2017 os cinco países que mais fizeram depósitos de patentes no mundo foram: Estados Unidos, China, Japão, Alemanha e República da Coreia.

Quando se identifica as partes estratégicas para a indústria farmacêutica aqui no Brasil é preciso saber as especificidades do sistema de inovação nacional, tendo em vista elementos da base produtiva junto com a tecnológica em

saúde, o modelo de atenção à saúde – universalidade, integralidade e equidade – e a regulação da produção e inovação em saúde (Gadelha, 2012).

Para orientar os nichos de estratégias da indústria nacional, é necessário levar em consideração não apenas o que é tendência tecnológica que está marcando a evolução da indústria farmacêutica no mundo, mas também, as partes específicas em associação ao sistema nacional de inovação em saúde junto da articulação de sua base produtiva (Vargas, 2017).

Olhando para as macrotendências associadas a política de atenção à saúde, aos perfis sociodemográficos e epidemiológicos, fica evidente o quão importante é a inclusão de novas plataformas tecnológicas tanto de biotecnologia, como de nanotecnologia para assistir a saúde e quão grande é o desafio para base produtiva de saúde nacional. Analisando a indústria farmacêutica no país, encontramos dificuldades se tratando de como funciona a base produtiva, da estrutura regulatória e da infraestrutura de ciência, tecnologia e inovação (CT&I) que são ligadas à área da saúde (Vargas, 2017).

O Brasil é um grande atrativo para empresas mundiais no ramo farmacêutico, principalmente no segmento de genéricos e de medicamentos similares. Com a regulação dos medicamentos genéricos no Brasil, facilitou a entrada de produtores de medicamentos, aumentando a concorrência no setor, levando em consideração que o país não é um realizador de gastos com P&D (Nishijima et al., 2014).

3.2 Algoritmos utilizados na indústria farmacêutica

3.2.1 Árvore de decisão

Quando se utiliza a árvore de decisão, teremos não somente uma visão geral sobre o problema, mas também um caminho para agir. Tendo como resultado a análise de decisão, uma forma racional para a compreensão do problema estudado por completo. Friedland DJ et al (1998, p. 38). É preciso avaliar criticamente toda a estrutura que se montou da árvore de decisão, probabilidades, estimativas e pressupostos, deste ponto, avaliar se os resultados serão plausíveis para o propósito final (Balbinotto Neto, 2017).

Para haver uma liberação de medicamentos não esteroidais com qualidade, segurança e eficácia, é preciso excluir presença de microrganismos, aqueles que podem degradar o produto e que seja um indicador de higiene durante a fabricação, podendo causar efeito adverso no paciente. A árvore de decisão foi usada para estabelecer quais microrganismos eram questionáveis. Na árvore de decisão, foi usada vários cenários com agentes biológicos diferentes, seja microrganismos inofensivos ou oportunistas patógenos, em combinação com 9 produtos diferentes, representando cada tipo de via de administração para drogas estéreis. Os resultados mostraram que o uso desta abordagem tornou a avaliação dos microrganismos fácil e verificável (Burgalassi et al., 2021).

A Organização Mundial da Saúde (OMS 2016), afirmou que a árvore de decisão é frequentemente usada em análises de riscos. E o uso de uma árvore de decisão recentemente foi sugerido pela Parenteral Drug Association (PDA, 2014) para avaliar microrganismos, por exemplo, em valores de atividade de água que evitam o crescimento de microrganismos (Apud Burgalassi *et al.*, 2021).

Foi idealizado um estudo para melhorar ferramentas para prever a absorção de drogas em humanos, acelerando o desenvolvimento da formulação e processo dos medicamentos usando a árvore de decisão, assim, auxiliando o investigador na escolha dos métodos *in vitro* mais adequados para otimizar o projeto de formulação oral e processo de desenvolvimento. A árvore de decisão é considerada como um documento vivo, orientando os investigadores durante todo o processo de decisão para selecionar o teste *in vitro* que mais se adequará durante o desenvolvimento da formulação e prevendo se haverá outros métodos que precisarão ser adicionados ao longo do desenvolvimento (Andreas *et al.*, 2018).

Separando a estrutura geral da árvore de decisão em estágios, onde foram vistas formulações de liberação retardada, a identificação de que tipo de doença e equipamento de solução estaria disponível e qual o melhor para o teste de liberação, o

método de dissolução adequado, o tipo de formulação e também o ativo farmacêutico, que são grandes pontos que influenciam na seleção e onde os protocolos experimentais são disponibilizados e a apresentação da estrutura geral destes documentos. A perspectiva é que a com o feedback da indústria farmacêutica em sua comunidade científica, essa árvore de decisão seja revisada e ampliada, para que no futuro seja ampliada, cobrindo uma cadeia maior de métodos experimentais (Andreas *et al.*, 2018).

Ilyas *et al.* (2021), viram que existe um crescimento em pesquisa relacionada à remoção de Contaminantes Orgânicos Emergentes (EOCs), como os fármacos, de águas residuais usando áreas úmidas construídas, porém, há ausência de pesquisas sobre a tradução de ciências disponíveis e evidências científicas em ferramentas de apoio à decisão, vendo isto, foi desenvolvido uma nova estrutura de árvore de decisão, propondo uma demonstração de alto potencial para melhorar o conhecimento e apoiar aplicativos para realização da remoção. Esta ferramenta pode ser útil para o ramo existente de tomadores de decisão, sejam eles formuladores de políticas, engenheiros de design, operadores, cientistas pesquisadores, educadores e cidadãos.

Na árvore de decisão projetou-se as águas residuais focando na remoção de parâmetros convencionais, demandas químicas e bioquímicas de oxigênio, total sólidos em suspensão, nitrogênio total e fósforo total. Neste estudo, alguns produtos farmacêuticos foram encontrados nas águas residuais, como o paracetamol, diclofenaco, ibuprofeno, tonalide e triclosan (Ilyas *et al.*, 2021).

A árvore de decisão foi uma ferramenta escolhida para identificar indivíduos que não aderiram ao trabalho anti-hipertensivo, através dela se conseguiu elaborar regras lógicas de decisão com informações obtidas e com capacidade de prever variáveis independentes que guiam a determinado fim. Utilizou-se da árvore de decisão junto com um banco de dados, a árvore mostrou um resultado capaz de classificar de forma correta 87,28% dos indivíduos, assim, contribuindo com as equipes de saúde para poder realizar as abordagens. Neste estudo, a árvore de decisão mostrou que os pacientes que não aderem ao tratamento correto são por conta da dificuldade em comparecer ao posto de saúde, outro ponto importante levantado foi que os indivíduos que tem pouco conhecimento sobre a Hipertensão Arterial Sistêmica, também tem dificuldades para ser aderente ao tratamento (Medeiros *et al.*, 2014).

3.2.2 Regressão linear de mínimos

Um trabalho foi realizado na tentativa de estabelecer um modelo do comportamento de retenção para compostos farmacêuticos de forma gradiente em Cromatografia de Fluido Supercrítico (SFC). Separando propositalmente uma vasta gama de propriedades físico-químicas conhecidas e usadas como sondas de retenção durante os experimentos, sendo todos compostos originados do campo farmacêutico e com variedades de moléculas interessantes para a SFC. Para realização deste trabalho usou a Regressão Linear Mínima Quadrada Múltipla (MLR), que foi primeiro aplicada individualmente para avaliar a possibilidade de estabelecer modelos de retenção para cada sonda. Esses modelos tiveram que descrever a ligação entre as condições experimentais e a retenção observada e permitir apreciar o impacto dos diferentes parâmetros experimentais na retenção das sondas (Andri *et al.*, 2017).

A regressão linear de mínimos, junto com mínimos quadrados parciais, foram usados para analisar a Dapoxetina (DAP) e a Tadalafila (TAD), que associados formam o Erectafil, usado para o tratamento da ejaculação precoce. Foi realizada a análise de seus binários e formulações farmacêuticas, fazendo comparações entre eles, também discutindo suas vantagens e desvantagens. A proposta foi analisar misturas binárias e dosagem farmacêutica e obtiveram sucesso detectando as concentrações de ambas as drogas juntas, tanto na mistura do conjunto de treinamento, como no conjunto de teste. Os resultados obtidos são um incentivo para um desempenho inteligente modelos quimiométricos para a determinação quantitativa de vários medicamentos com equipamentos e dispositivos simples e de baixo custo (Anwar *et al.*, 2020).

Um levantamento foi feito para analisar o quantitativo de produtos farmacêuticos vendidos durante os anos de 2003 a 20015 no Brasil, e foi usado a regressão linear simples para realizar uma estimativa das vendas para o ano de 2016, por fim, exibir as vendas reais de 2016 e o que foi estimado pela técnica de regressão linear. E a dúvida neste estudo era se a regressão linear simples seria capaz de mostrar uma boa exatidão em relação a previsão de vendas para o setor farmacêutico. Durante a estimativa, o maior índice de eficiência foi de 93,63% e o menor foi de 78,36%, pode-se achar que os resultados apresentados pela regressão linear simples não sejam satisfatórios, mas pode ser explicado devido ao mercado farmacêutico ser sazonal, onde, tem meses que teremos picos de vendas, e uma precisão exata do valor vendido fica improvável. O estudo concluiu que o objetivo ao usar a regressão linear foi alcançado e serviu o realizar a estimativa do valor de vendas no mercado farmacêutico (Oliveira; Neto, 2017).

Foi feita uma aplicação do modelo de regressão linear para avaliar os dados de estabilidade de medicamento. Foram usadas três amostras de três lotes diferentes do produto, compondo dois princípios ativos em dois tipos de embalagem (PVCD e Flow Pack - FP), tendo o controle da temperatura e umidade durante o armazenamento, onde, as amostras de forma individual eram coletadas por um tempo pré-determinado para ter informações das características físicas, químicas, biológicas e microbiológicas. Pelo modelo de regressão linear foi observado que o produto tende a degradar nas diferentes embalagens. Em determinado produto, se viu que a embalagem PVCD + FP é mais eficiente do que usar a PVCD, mas o uso da FP torna o produto mais caro. Comumente, o modelo de regressão linear se aplica ao estudo de estabilidade estimando taxas como de formação de impurezas dos produtos, mas, este estudo foi realizado para usar a regressão linear nos estudos de dados de estabilidade, verificando a eficiência das duas embalagens na proteção do produto (Souza; Raminelli, 2013).

3.2.3 Support Vector Machine

Burbridge *et al.* (2001), usaram um algoritmo da comunidade de aprendizagem e aplicaram em um problema que é conhecido na descoberta dos medicamentos. Utilizaram Vetor de Suporte de Máquina (Support Vector Machine - SVM), um novo algoritmo de aprendizagem para classificação. O SVM prova seu potencial para a análise da relação estrutura-atividade. O problema para qual houve sua aplicação, foi a previsão da inibição da dihidrofolato redutase por pirimidinas. Existe a Minimização de Risco Estrutural (SRM), onde em uma sequência de espaços de hipóteses de complexidade em crescimento, o SRM escolhe qual erro de treinamento será minimizado, onde, após esta etapa se escolhe o que minimiza um superior ligado ao erro que está acontecendo a generalização, e o SVM consegue realizar essas etapas ao mesmo tempo, fazendo o controle do tamanho do peso dos recursos.

Quando se utiliza os parâmetros SVM, não precisa utilizar-se de outros treinamentos, como a árvore de decisão, por exemplo. A máquina de vetores de suporte entra em uma classificação inteligente e robusta, apresentando precisão técnica, um sistema de aprendizado de competência automatizada, eficiente algoritmo de processamento, com capacidade de reproduzir resultados verificáveis (Burbridge *et al.*, 2001).

Dados financeiros de várias empresas farmacêuticas foram capturados para elaborar um sistema que avaliasse o risco de crédito e para comparar o risco efeito de avaliação do modelo logístico e modelo de Máquina de Vetor de Suporte (SVM) baseado em análise fatorial. Os resultados mostraram a superioridade de 3,7% do modelo SVM em cima do modelo logístico, com erro inferior a 12,3%, indicando sua superioridade e eficácia quando este se aplica à avaliação de risco de crédito da indústria farmacêutica (Liu; Zeng, 2020).

Houve uma investigação de máquina de vetores de suporte (SVM) algoritmos, para ver seu desempenho e capacidade preventivos na área de designer de medicamento. Onde se viu de forma clara quão importante é a otimização dos parâmetros de SVM em forma de seleção de variáveis para poder desenvolver modelos estatísticos. No SVM foi analisado valores de parâmetros e quais variáveis incluir no momento da caracterização dos compostos que estavam sob investigação. Concluiu-se

que a tecnologia de máquinas de vetores de suporte pode ter sua aplicação com sucesso na derivação de modelos estatísticos, apresentando boas estatísticas e boas capacidades preditivas dentro de áreas, sejam elas tradicionais ou derivadas relações quantitativas de estrutura-atividade (Norinder, 2003).

3.2.4 Regressão logística

Os modelos de regressão fazem parte de uma das ferramentas estatísticas de mais importância para a área de estatística de dados quando a pretensão é modelar relações entre as variáveis. Encontramos na regressão logística uma técnica estatística com o objetivo de modelar, fazendo relação logística entre uma variável resposta dicotômica e entre variáveis explicativas numéricas (Cabral, 2013).

Na última década acelerou o desenvolvimento de nano formulações farmacêuticas, porém, os nano portadores de drogas ainda enfrentam desafios. Alguns modelos matemáticos preditivos foram adotados para formulação de microemulsão, manufatura e aumento de escala, entre eles, a regressão logística. Pela regressão logística consegue prever probabilidades de uma microemulsão passar no teste de controle de qualidade, suas variáveis de saída são binárias, podendo então ser usada na modelagem de uma resposta binária, vendo se uma formulação de microemulsão atende ou não a alguma especificidade de identificação de atributos de qualidade. Sendo assim, usou a regressão logística para modelar a probabilidade de formulações de microemulsão que atendem às especificações de identificação de atributos de qualidade (Herneisey *et al.*, 2019).

Al-Tahat *et al.* (2013) usaram um dos seguimentos da Modelagem de Regressão Logística para fazer prevenção e investigação de falhas encontradas em ferramentas de comprimidos na indústria farmacêutica. Usando esta ferramenta, ajudaria a diminuir erros futuros de punção. Foi investigado três atributos de punção, cinco atributos do produto, teor de umidade e tamanho da peneira. Aplicando a regressão logística aos tipos de falha de previsão, de acordo com os aspectos citados, investigando a contribuição de cada um para diferentes tipos de falha. A regressão ligou e avaliou de forma quantitativa o que cada fator tem de efeito e contribuição para que as falhas ocorressem, mostrando ser uma ferramenta forte.

3.2.5 Naive Bayes

Naive Bayes é um dos modelos mais conhecidos na área de aprendizado de máquina e na aplicação de probabilidade, usando a teoria de Bayes (estatístico, filósofo e ministro presbiteriano Thomas Bayes) como princípio fundamental. Tendo como ideias iniciais a suposição de independência entre as variáveis do problema, realizando uma classificação de probabilidades com observações, as caracterizando em classes pré-definidas. Com seguimentos sendo aplicados em análises de créditos, busca por falhas em sistemas mecânicos e diagnósticos médicos (Casali, 2021).

Para classificar os medicamentos desconhecidos são utilizadas seis propriedades químicas: massa molecular relativa, número de doadores de ligações de hidrogênio, número de aceitadores de ligações de hidrogênio, área de superfície polar, constante hidrofóbica e número de chaves de rotação flexíveis. Aplicando o algoritmo Naive Bayes nos dados que foram coletados, a classe alvo dos novos medicamentos pôde ser prevista precisamente. E este estudo classificou os medicamentos usados para febre e febre tifoide, baseando nas propriedades químicas citadas acima e sendo atribuída a cada doença utilizando o classificador Naive Bayes (Indraja; K, 2018).

Indraja e K (2018) usaram o algoritmo Naive Bayes o aplicando no banco de dados para prever se a nova droga pertencia à classe da febre ou da tifoide, recuperando os valores dos atributos do medicamento contra a febre e calculando a soma dos valores desses atributos da massa molecular relativa para os medicamentos contra a febre.

No estágio inicial do projeto de drogas tem uma necessidade que seja criado métodos alternativos que tenham confiança e eficiência in vivo para avaliar os produtos químicos que possuam potencial de neurotoxicidade. Neste estudo, foram desenvolvidos modelos computacionais de predição para neurotoxicidade induzida por drogas usando o Naive Bayes.

Uma vantagem deste algoritmo é saber lidar com vários dados de forma simultânea, com tolerância a ruídos aleatórios durante o processo de construção do modelo e foi aplicado para diferenciar neurotóxicos de não neurotóxicos, baseado em oito propriedades físico-químicas moleculares e dez tipos de impressões digitais moleculares com onze diâmetros diferentes. O Naive Bayes mostrou uma ferramenta robusta e desenvolveu com sucesso modelos de predição computacional induzidos por drogas de neurotoxicidade (Zhang *et al.*, 2020).

4. Considerações Finais

Existem várias tecnologias ativas no mercado e ainda assim é necessário o surgimento de novas que auxiliem no processo de inovação. Nesta revisão foram identificados algoritmos fazendo a indústria inovar e alcançar agilidade, sendo inseridos cada vez mais na área da saúde. É visto que a área farmacêutica é um dos seguimentos que mais se utilizam de tecnologias da informação e se mostra disposta a buscar investimentos. Buscar parcerias entre empresas é a solução para aumentar o alcance, ter soluções mais ágeis e obter mais lucros.

Quando se utiliza algoritmos no desenvolvimento de um medicamento, é possível prever problemas que podem surgir durante o processo e saber como agir, como é o caso da árvore de decisão. Com a utilização da regressão linear de mínimos se consegue fazer quantitativos e estimativas de estabilidade de forma segura. Usando o support vector machine, se tem um algoritmo de aprendizagem de classificação, sendo uma ferramenta inteligente e com robustez. Para dados estatísticos com objetivo de modelagem, o uso da regressão logística é um dos mais importantes algoritmos para resolução de problemas nestas áreas. Foi visto também o Naive Bayes, que no aprendizado de máquina é bem conhecido e forte na aplicação de probabilidade, e este algoritmo se destaca em sua importância, pois, quando se inicia uma nova droga é necessário criar métodos seguros e eficiência in vivo para avaliar seu nível de toxicidade, e é possível com o Naive Bayes, ele tem a capacidade de lidar com vários dados de forma simultânea.

A indústria farmacêutica ganha com o uso desses algoritmos, impactando no aperfeiçoamento em cada etapa do setor farmacêutico. Aplicando os mesmos dentro da indústria, ela ganha ao adquirir agilidade no desempenho, nas capacidades preventivas na área de designer de medicamento, na qualidade, segurança, estabilidade, eficácia, capacidades preventivas de microrganismos, probabilidades ou investigando falhas. Sendo um tema interessante, novo na área de saúde e que abre espaço para estudos futuros.

Agradecimentos

Agradecemos ao nosso Deus por nos guiar em cada passo, a nossa família e amigos que nos impulsionam a irmos mais longe e a todos que de alguma forma contribuíram para construção deste artigo.

Referências

- Al-Tahat, M. D., Jawwad, A. K. M. A., & Nahleh, Y. L. A. (2013). Ordinal logistic regression model of failure mode and effects analysis (FMEA) in pharmaceutical tableting tools. *Engineering Failure Analysis*, 27, 322-332.
- Andreas, C. J., Rosenberger, J., Butler, J., Augustijns, P., McAllister, M., Abrahamsson, B., & Dressman, J. (2018). Introduction to the OrBiTo decision tree to select the most appropriate in vitro methodology for release testing of solid oral dosage forms during development. *European Journal of Pharmaceutics and Biopharmaceutics*, 130, 207-213.
- Andri, B., Dispas, A., Marini, R. D. E. A., Hubert, P., Sassi, P., Al Bakain, R., ... & Vial, J. (2017). Combination of partial least squares regression and design of experiments to model the retention of pharmaceutical compounds in supercritical fluid chromatography. *Journal of Chromatography A*, 1491, 182-194.
- Anwar, B. H., Abdelhamid, N. S., Magdy, M. A., & Naguib, I. A. (2020). Linear support vector regression and partial least-squares for determination of dapoxetine hydrochloride and tadalafil in binary pharmaceutical mixtures. *Journal of AOAC International*, 103(1), 132-139.
- Arsene, C. *Artificial Intelligence & Pharma: o que vem a seguir?* Digital Authority Partners, 04 August 2020.

- Balbinotto Neto, Giacomo. *Análise de decisão aplicada à farmacoeconomia*. In Instituto de Avaliação de Tecnologia em Saúde. P. 1-303, 23. Porto Alegre.
- Bulgaru, I. Indústria farmacêutica na era da inteligência artificial: o futuro é brilhante. 12 May 2020.
- Burbidge, R., Trotter, M., Buxton, B., & Holden, S. (2001). *Drug design by machine learning: support vector machines for pharmaceutical data analysis*. *Computers & chemistry*, 26(1), 5-14.
- Burgalassi, S., Ceccanti, S., Vecchiani, S., Leonangeli, G., Federigi, I., Carducci, A., & Verani, M. (2021). Objectionable microorganisms in pharmaceutical production: Validation of a decision tree. *European Journal of Pharmaceutical Sciences*, 166, 105984.
- Cabral, C. I. S. (2013). *Aplicação do modelo de regressão logística num estudo de mercado* (Doctoral dissertation).
- Camargo, A. C. M. *Inovações farmacêuticas no Brasil*. In Uniemp Inovação. ano I, n. 1, september 2004.
- Gadelha, C. A. G. (2012). *A dinâmica do sistema produtivo da saúde: inovação e complexo econômico-industrial*. Editora Fiocruz.
- Gomes, R. D. P., Pimentel, V. P., Cardoso, M. L., & Pieroni, J. P. (2014). O novo cenário de concorrência na indústria farmacêutica brasileira.
- Herneisey, M., Lambert, E., Kachel, A., Shychuck, E., Drennen, J. K., & Janjic, J. M. (2019). *Quality by design approach using multiple linear and logistic regression modeling enables microemulsion scale up*. *Molecules*, 24(11), 2066.
- Ilyas, H., Masih, I., & van Hullebusch, E. D. (2021). A decision tree framework to support design, operation, and performance assessment of constructed wetlands for the removal of emerging organic contaminants. *Science of the Total Environment*, 760, 143334.
- Indraja, B., & Annapurani, K. (2018). Classification of medicines using naive bayes classifier. *Research Journal of Pharmacy and Technology*, 11(5), 1940-1944.
- Rojas Jesus, A. D. R. (2018). *A indústria farmacêutica: um estudo focado nos tratados internacionais e no comércio exterior do Peru*.
- Koerich, G. V., de Lorenzi Cancellier, É. L. P., & Tezza, R. (2014). Capacidade de absorção em pequenas empresas: proposição de um instrumento de medição. VII Encontro de estudos em empreendedorismo e gestão de pequenas empresas (EGEPE). Goiânia, 01-18.
- Koromina, M., Pandi, M. T., & Patrinos, G. P. (2019). *Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics*. *Omics: a journal of integrative biology*, 23(11), 539-548.
- Liu, P., & Zeng, Z. (2021). *Evaluation of financial credit risk on pharmaceutical supply chain based on support vector machine*. In *Journal of Physics: Conference Series* (Vol. 1774, No. 1, p. 012030). IOP Publishing.
- Lobato, A. O. C., & Ortiz, R. M. (2019). A inovação e a proteção da propriedade intelectual no Brasil: análise da dependência nacional da tecnologia farmacêutica estrangeira. *Revista GEINTEC*, 9(1), 4809-4824.
- Mak, K. K., & Pichika, M. R. (2019). *Artificial intelligence in drug development: present status and future prospects*. *Drug discovery today*, 24(3), 773-780.
- Medeiros, A. R. C., Araújo, Y. B. D., Vianna, R. P. D. T., & Moraes, R. M. D. (2014). *Modelo de suporte à decisão aplicado à identificação de indivíduos não aderentes ao tratamento anti-hipertensivo*. *Saúde em debate*, 38, 104-118.
- Mesquita, M. A. D., & Santoro, M. C. (2004). *Análise de modelos e práticas de planejamento e controle da produção na indústria farmacêutica*. *Production*, 14, 64-77.
- Nishijima, M., Biasoto Jr, G., & Lagroteria, E. (2014). *A competição no mercado farmacêutico brasileiro após uma década de medicamentos genéricos: uma análise de rivalidade em um mercado regulado*. *Economia e sociedade*, 23, 155-186.
- Norinder, U. (2003). *Support vector machine models in drug design: applications to drug transport processes and QSAR using simplex optimisations and variable selection*. *Neurocomputing*, 55(1-2), 337-346.
- Paula, G. N. D. (2001). Os desafios da indústria farmacêutica global e o redesenho do negócio pela Merck. *Revista de Administração de Empresas*, 41, 76-87.
- Santos, M. C. B. G., & Pinho, M. (2012). *Estratégias tecnológicas em transformação: um estudo da indústria farmacêutica brasileira*. *Gestão & Produção*, 19, 405-418.
- Souza, R. C., & Raminelli, J. *Aplicação do modelo linear na avaliação de dados de estabilidade de medicamento*. *Semina: Ciências Exatas e Tecnológicas*, 34(1), 57-66.
- Tierney, S. *Artificial intelligence in pharma: utilising a valuable resource*. In *European Pharmaceutical Review*. P. 1-1, 10 March. 2020.
- Torres, R. L. (2015). *Capacitação tecnológica na indústria farmacêutica brasileira*. Rio de Janeiro: UFRJ (Tese de Doutorado em Economia da Indústria e da Tecnologia).
- Vargas, M. A. (2017). Indústria de base química no Brasil: potencialidades, desafios e nichos estratégicos. Gadelha P, Gadelha CAG, Noronha JC, Pereira TR, organizadores. *Brasil Saúde Amanhã: complexo econômico industrial da saúde*. Rio de Janeiro: Editora Fiocruz.
- Whittemore, R.; Knafel, K. The integrative review: updated methodology. *Journal of advanced nursing*, 52(5), 546-553, 2000.
- Zhang, H., Mao, J., Qi, H. Z., Xie, H. Z., Shen, C., Liu, C. T., & Ding, L. (2020). *Developing novel computational prediction models for assessing chemical-induced neurotoxicity using naïve Bayes classifier technique*. *Food and Chemical Toxicology*, 143, 111513.