

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

RAFAEL RODRIGUES DA SILVA

**ÁREAS EM QUE PYTHON VEM SENDO UTILIZADO NO APRENDIZADO DE
MÁQUINA: UM MAPEAMENTO SISTEMÁTICO**

DOIS VIZINHOS

2023

RAFAEL RODRIGUES DA SILVA

**ÁREAS EM QUE PYTHON VEM SENDO UTILIZADO NO APRENDIZADO DE
MÁQUINA: UM MAPEAMENTO SISTEMÁTICO**

**Areas where Python has been used in Machine Learning: a systematic
mapping**

Trabalho de Conclusão de Curso apresentado
como requisito para obtenção do título de
Especialista em Tecnologia Python para Negócios,
da Universidade Tecnológica Federal do Paraná.

Orientadora: Prof^a. Dr^a. Simone de Sousa Borges

DOIS VIZINHOS

2023



[4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

RAFAEL RODRIGUES DA SILVA

**ÁREAS EM QUE PYTHON FOI UTILIZADO NO APRENDIZADO DE MÁQUINA:
UM MAPEAMENTO SISTEMÁTICO**

Trabalho de Conclusão de Curso apresentado
como requisito para obtenção do título de
Especialista em Tecnologia Python para Negócios,
da Universidade Tecnológica Federal do Paraná.

Data de aprovação: 28/março/2023

Simone de Sousa Borges
Doutorado
Universidade Tecnológica Federal do Paraná

Helena Macedo Reis
Doutorado
Universidade Federal do Paraná

Marisângela Pacheco Brittes
Doutorado
Universidade Tecnológica Federal do Paraná

DOIS VIZINHOS

2023

Dedico este trabalho à minha orientadora
Profa. Dra. Simone de Sousa Borges, por
todo empenho em me ajudar a fazer o
melhor.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço primeiramente a Deus, por me conceder saúde, força e coragem para conseguir superar as dificuldades durante essa trajetória.

Aos meus pais Geraldo e Silvana que me deram força e me ensinaram a correr atrás dos meus sonhos. Ao meu irmão Felipe que sempre me motivou e incentivou a concluir mais essa jornada.

Em especial a minha noiva Viviane que a todo momento esteve comigo me encorajando e não mediu esforços para me ajudar em todos os momentos.

Agradeço a minha orientadora Profa. Dra. Simone de Sousa Borges, pela sabedoria, paciência e dedicação com que me guiou nesta trajetória.

Aos meus colegas de grupo Claudi Martins e José Eduardo pela parceria, tempo e esforço que empregaram durante o curso.

Enfim, agradeço a todos que direta ou indiretamente contribuíram para a realização desta formação e pesquisa.

Software é uma grande combinação de
arte e engenharia.

RESUMO

O aprendizado de máquina (*machine learning*) é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam e façam previsões sobre os dados. A combinação da simplicidade, versatilidade e o suporte da grande comunidade do *Python* o tornou uma escolha muito popular para inúmeras tarefas, como para avaliar grandes conjuntos de dados com a ajuda de algoritmos de aprendizado de máquina. Existe um crescente interesse pelo aprendizado de máquina, suas aplicações e implicações, e seus possíveis resultados, principalmente com o uso da linguagem de programação *Python*. Apesar deste interesse não se encontraram estudos sobre as diversas áreas de aplicações em que vem sendo conduzidos estudos e investigações. Assim, um mapeamento sistemático foi conduzido a fim de cooperar com uma visão geral das áreas e formas de aplicações. Inicialmente, 274 estudos sobre a utilização do *Python* na aplicabilidade de aprendizado de máquina foram analisados. Dentre estes, apenas 25 estudos satisfaziam os critérios de inclusão e exclusão de artigos definidos neste trabalho. Estes 25 artigos foram selecionados e categorizados de acordo com sua área de aplicação, bibliotecas do *Python*, algoritmos e os dados de insumo a serem utilizados. Como resultado, um mapa das pesquisas na área foi desenvolvido e os resultados de cada artigo foram identificados apontando que todos os 25 estudos obtiveram resultados positivos em suas propostas.

Palavras-chave: Aprendizado de máquina. Python. Algoritmo. Bibliotecas. Área de atuação. Mapeamento sistemático.

ABSTRACT

Machine learning is a subfield of artificial intelligence that focuses on developing algorithms and models that allow computers to learn and make predictions about data. The combination of Python's simplicity, versatility, and support from the large community has made it a very popular choice for numerous tasks, such as evaluating large datasets with the help of machine learning algorithms. There is a growing interest in machine learning, its applications and implications, and its possible results, mainly with the use of the Python programming language. Despite this interest, no studies were found on the various areas of applications in which studies and investigations have been conducted. Thus, a systematic mapping was conducted in order to cooperate with an overview of areas and forms of applications. Initially, 274 studies on the use of Python in machine learning applicability were analyzed. Among these, only 25 studies met the inclusion and exclusion criteria for articles defined in this work. These 25 articles were selected and categorized according to their application area, Python libraries, algorithms and input data to be used. As a result, a map of research in the area was developed and the results of each article were identified, indicating that all 25 studies obtained positive results in their proposals.

Keywords: Machine learning. Python. Algorithm. Libraries. Occupation area. Systematic mapping.

LISTA DE ILUSTRAÇÕES

Figura 1 - Bibliotecas utilizadas nos artigos	28
Figura 2 - Quantidade de diferentes algoritmos por área de atuação	30
Figura 3 - Quantidade de artigos com menções dos algoritmos por área de atuação	32
Quadro 1 - Artigos retornados por cada base de dados eletrônica e seleção final ...	21
Quadro 2 - Quantidade de artigos por área de atuação	22
Quadro 3 - Quantidade de artigos por categoria	24

LISTA DE ABREVIATURAS

QP	Questão de Pesquisa
IA	Inteligência Artificial

LISTA DE ACRÔNIMOS

IEEE	Institute of Electrical and Electronics Engineers
ACM	Association for Computing Machinery
DDoS	Distributed Denial of Service
ACCS	Australian Centre for Cyber Security
NOAA	National Oceanic and Atmospheric Administration
.CSV	Comma-separated values

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Questão de pesquisa	13
1.2	Objetivo geral	13
1.3	Justificativa	13
1.4	Estrutura do trabalho	13
2	ASPECTOS CONCEITUAIS	14
2.1	Python	14
2.2	Aprendizado de máquina	15
2.3	Uso de Python em aprendizado de máquina	17
3	METODOLOGIA	19
3.1	Metodologia utilizada	19
3.2	Questões de pesquisa	19
3.3	Processo de pesquisa	20
3.4	CrITÉrios de exclusão	21
3.5	Categorização dos artigos selecionados	23
4	DESENVOLVIMENTO	26
4.1	Uso de Python e suas bibliotecas	26
4.2	Algoritmos de aprendizado de máquina	28
4.3	Uso de dados	32
5	RESULTADOS	34
5.1	Resposta às questões de pesquisa	34
6	CONCLUSÕES	36
6.1	Trabalhos futuros	36
	REFERÊNCIAS	38

1 INTRODUÇÃO

O *Python* é uma linguagem de programação interpretada de alto nível que foi criada em 1991. Ela foi projetada para ser uma linguagem simples, fácil de aprender e acessível a uma ampla gama de programadores (ALURA, 2021). *Python* tornou-se uma das linguagens de programação mais populares do mundo em virtude a vários de seus pontos fortes, tendo resultados expressivos principalmente na área de aprendizado de máquina. O aprendizado de máquina (*machine learning*) é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam e façam previsões sobre os dados. Tem suas raízes no campo de reconhecimento de padrões e da teoria da aprendizagem computacional, das décadas de 1950 e 1960. A importância do uso de *Python* no aprendizado de máquina reside na sua capacidade de fornecer um ambiente de programação de alto nível para a análise e modelagem de dados. Motivados pelo crescente interesse e utilização dos conhecimentos e aplicações da área de inteligência artificial e seu subcampo aprendizado de máquina (ARBIT, 2018), foram levantados alguns questionamentos em relação às suas aplicações.

A despeito deste interesse, o autor deste trabalho não possui conhecimento da existência de estudos que apresentem uma visão geral sobre o uso da linguagem *Python* aplicado no aprendizado de máquina. Com o propósito de preencher essa lacuna, foi realizado um mapeamento sistemático das pesquisas existentes envolvendo *Python* e aprendizado de máquina. O mapeamento sistemático é uma metodologia que abrange a busca por literatura a fim de verificar a extensão e a quantidade de estudos publicados (que são chamados de estudos primários) na área de interesse (PETERSEN *et al.*, 2008). Neste trabalho, são descritos os resultados de um mapeamento sistemático juntamente com os elementos fundamentais do protocolo de pesquisa utilizado e a maneira como o procedimento foi realizado. A seguir serão apresentados em maiores detalhes a questão de pesquisa, os objetivos e a justificativa deste trabalho.

1.1 Questão de pesquisa

Quais áreas vêm utilizando mais soluções baseadas em aprendizagem de máquina desenvolvidas com *Python*?

1.2 Objetivo geral

Investigar o estado da arte do uso de *Python* no desenvolvimento de aplicações de aprendizagem de máquina.

1.3 Justificativa

Conhecer em quais áreas estão concentradas as aplicações baseadas em aprendizado de máquina e desenvolvidas com *Python* é importante porque destaca a versatilidade e o poder dessa linguagem nesse campo. Desta forma, compreender as várias maneiras pelas quais o *Python* vêm sendo usado no aprendizado de máquina pode contribuir para a tomada de decisões mais confiáveis em relação a ferramentas e técnicas utilizadas em projetos específicos, possibilitando o aproveitamento eficaz não somente dos recursos da linguagem, bem como dos recursos disponíveis na comunidade *Python*.

1.4 Estrutura do trabalho

Este trabalho está organizado da seguinte maneira: na Seção 2 são apresentados os principais conceitos estudados neste trabalho. Na Seção 3, além de descrever os principais elementos do protocolo de pesquisa utilizado, são apresentados detalhes sobre como o mapeamento sistemático foi conduzido, como os estudos foram selecionados, avaliados e categorizados, e quais foram os critérios utilizados para análise dos resultados. O desenvolvimento do trabalho está inserido na Seção 4, onde as informações relevantes coletadas e as descobertas realizadas foram registradas. Na seção 5, os principais resultados do mapeamento são comentados. E, finalmente, as conclusões são apresentadas na seção 6.

2 ASPECTOS CONCEITUAIS

Nesta seção são introduzidos conceitos fundamentais para o entendimento deste estudo, abordando *Python*, aprendizado de máquina e o uso de *Python* em aprendizado de máquina.

2.1 Python

O *Python* é uma linguagem de programação interpretada de alto nível que foi criada em 1991 por Guido van Rossum. Ela foi projetada para ser uma linguagem simples, fácil de aprender e acessível a uma ampla gama de programadores (ALURA, 2021). *Python* tornou-se uma das linguagens de programação mais populares do mundo em virtude a vários de seus pontos fortes, alguns deles são:

- **Legibilidade:** o código *Python* é conhecido por sua alta legibilidade, o que facilita a compreensão e a manutenção de outros programadores.
- **Versatilidade:** *Python* pode ser usado para uma ampla gama de tarefas, incluindo: desenvolvimento web, desenvolvimento de jogos, computação científica, análise de dados, estatística e inteligência artificial.
- **Comunidade grande e ativa:** o *Python* possui uma comunidade grande e ativa de usuários e desenvolvedores, o que levou à criação de um grande número de bibliotecas e ferramentas para inúmeras tarefas.
- **Fácil de aprender:** o *Python* possui uma sintaxe simples e intuitiva, tornando-o uma linguagem ideal para iniciantes e amadores.

No entanto, o *Python* também tem alguns pontos fracos, incluindo:

- **Desempenho:** *Python* é uma linguagem interpretada, o que pode torná-la mais lenta do que as linguagens compiladas como o C ou C++.

- **Tipagem dinâmica:** o *Python* usa tipagem dinâmica, o que pode torná-lo mais propenso a erros de tempo de execução e também pode dificultar a detecção de erros durante o desenvolvimento.

Algumas das principais peculiaridades do *Python* incluem:

- **Uso de espaço em branco:** o *Python* usa espaços em branco (recuo) para delimitar blocos de código, o que é diferente de diversas outras linguagens de programação que geralmente usam chaves ou palavras-chave.
- **Tipagem dinâmica:** conforme mencionado anteriormente, o *Python* é uma linguagem de tipagem dinâmica, o que significa que as variáveis não possuem um tipo predeterminado em sua definição e podem ser alteradas em tempo de execução.
- **Grande biblioteca padrão:** o *Python* tem uma grande biblioteca padrão que inclui diversos módulos para uma ampla variedade de tarefas, incluindo entrada e saída de arquivos, expressões regulares e serviços da web.
- **Linguagem interpretada:** *Python* é uma linguagem interpretada, o que significa que o código é executado linha por linha, em vez de ser compilado em código de máquina.

No geral, a combinação da simplicidade, versatilidade e o suporte da grande comunidade do *Python* o tornou uma escolha muito popular para inúmeras tarefas, desde o desenvolvimento da web e análise de dados até a computação científica e inteligência artificial (ALTITUDE TVM, [s.d.]; L TECNOLOGIA, 2021).

2.2 Aprendizado de máquina

O aprendizado de máquina (*machine learning*) é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam e façam previsões sobre os dados.

Tem suas raízes no campo de reconhecimento de padrões e da teoria da aprendizagem computacional, das décadas de 1950 e 1960.

O desenvolvimento do aprendizado de máquina foi facilitado pelos grandes avanços no poder de computação e de armazenamento, bem como pela disponibilidade de grandes e diversos conjuntos de dados (*big data*). Com esses recursos, os pesquisadores foram capazes de criar algoritmos e modelos que conseguem aprender com os dados e fazer previsões sobre novos dados com um nível de precisão cada vez maior. O aprendizado de máquina tem vários pontos fortes que o tornam uma ferramenta poderosa para resolver uma ampla gama de problemas. Por exemplo:

- **Automação:** os algoritmos de aprendizado de máquina podem automatizar tarefas que, de outra forma, exigiriam grande intervenção humana, como por exemplo: classificação de imagens ou processamento de linguagem natural.
- **Escalabilidade:** os algoritmos são capazes de processar grandes quantidades de dados e fazer previsões para milhões de entradas em questão de segundos ou minutos.
- **Precisão aprimorada:** muitos algoritmos de aprendizado de máquina superam os métodos tradicionais quando se fala em precisão, especialmente quando grandes quantidades de dados estão envolvidas.

Entretanto, o aprendizado de máquina também possui vários pontos fracos, incluindo:

- **Overfitting:** o *overfitting* acontece quando um modelo é muito complexo e acaba por aprender muito bem os dados de treinamento, levando a um desempenho ruim em dados novos e não vistos.
- **Viés:** os algoritmos de aprendizado de máquina podem aprender e reproduzir vieses presentes nos dados de treinamento, o que pode ter consequências negativas se esses vieses forem prejudiciais ou discriminatórios.

- **Falta de transparência:** muitos algoritmos de aprendizado de máquina são considerados "caixas pretas" porque é difícil entender realmente como eles chegaram a suas previsões. Isso acaba por dificultar a confiança nos resultados, o diagnóstico e a correção de erros.
- **Requer grandes quantidades de dados:** os algoritmos de aprendizado de máquina geralmente exigem grandes quantidades de dados para que possam ser treinados a fim de produzir resultados precisos, isso pode limitar sua aplicabilidade em alguns domínios ou áreas de aplicação.

Apesar dessas deficiências, o aprendizado de máquina teve um impacto gigantesco em uma vasta gama de setores e continua a ser uma área cada vez mais ativa de pesquisa e desenvolvimento (ENGENHARIA HÍBRIDA, 2022; ROCHA, [s.d.]; LEARNING DATA, 2021).

2.3 Uso de Python em aprendizado de máquina

Do ponto de vista técnico, a importância do *Python* no aprendizado de máquina reside na sua capacidade de fornecer um ambiente de programação de alto nível para a análise e modelagem de dados. Algumas das maneiras específicas pelas quais o *Python* ajuda a dar suporte ao aprendizado de máquina são:

- **Tipagem dinâmica:** as variáveis tipadas dinamicamente do *Python* facilitam e muito o trabalho com conjuntos de dados grandes e complexos e a execução de operações que requerem ou utilizam vários tipos de dados.
- **Bibliotecas de alto nível:** as bibliotecas do *Python* para aprendizado de máquina, como por exemplo: *NumPy*, *Pandas*, *Scikit-Learn*, *TensorFlow* e *PyTorch*, fornecem poderosos algoritmos e funções pré-construídos que podem ser facilmente aplicados a várias tarefas de aprendizado de máquina. Essas e outras muitas bibliotecas ajudam a simplificar e facilitar a implementação de modelos

complexos de aprendizado de máquina e fornecem uma vasta gama de funcionalidades para manipulação e análise de dados.

- **Visualização de aprendizado de máquina:** o *Python* possui diversas bibliotecas para visualizar dados e resultados de modelos de aprendizado de máquina, como a *Matplotlib* e a *Seaborn*, que fornecem ferramentas para explorar e entender as saídas de aprendizado de máquina.
- **Extensibilidade:** o *Python* é altamente extensível, isso significa que os usuários e programadores podem escrever suas próprias funções, classes e bibliotecas personalizadas para atender às suas necessidades específicas. Isso possibilita a construção de modelos de aprendizado de máquina complexos e altamente personalizados.
- **Grande comunidade:** há uma comunidade grande e ativa de desenvolvedores, cientistas de dados e especialistas em aprendizado de máquina que contribuem para o desenvolvimento e manutenção de bibliotecas e ferramentas em *Python*. Essa comunidade oferece diversos recursos, inúmeros tutoriais e suporte para quem trabalha ou está estudando aprendizado de máquina em *Python*.

No geral, a facilidade de uso, as bibliotecas de alto nível, os recursos de visualização, a extensibilidade e a grande comunidade do *Python* o tornam uma ferramenta valiosa e poderosa para aprendizado de máquina. Esses recursos possibilitam que os usuários implementem modelos complexos e personalizados de maneira rápida e eficaz, realizem análises de dados e visualizem resultados, tornando-o uma escolha robusta para qualquer projeto de aprendizado de máquina (IMPACTA, 2020; DATA SCIENCE ACADEMY, 2021).

3 METODOLOGIA

Nesta seção são apresentadas informações sobre a metodologia utilizada e os meios utilizados para a realização do trabalho.

3.1 Metodologia utilizada

O tipo de pesquisa realizada foi a pesquisa exploratória (WAZLAWICK, 2014), e a metodologia utilizada para viabilizá-la foi a realização de um mapeamento sistemático (PETERSEN *et al.*, 2008). A escolha do mapeamento se justifica principalmente por esta ser uma metodologia de pesquisa usada para identificar, analisar e categorizar a literatura existente sobre um tópico ou assunto específico, de forma rápida e confiável, ainda que mais simples quando comparada a outros métodos de pesquisa (PETERSEN *et al.*, 2008). O processo de mapeamento sistemático utilizado é o proposto por Petersen *et al.*, (2008) e envolve as cinco etapas principais, são elas:

- definição de questões de pesquisa;
- realização da pesquisa de estudos primários relevantes;
- triagem dos documentos;
- *keywording* dos resumos;
- a extração de dados e mapeamento.

3.2 Questões de pesquisa

Levando em conta que as questões de pesquisa devem reger os objetivos do mapeamento, as seguintes questões foram definidas:

QP1: Em quais áreas há o maior número de aplicações?

QP2: Quais são as principais bibliotecas utilizadas nos trabalhos encontrados?

QP3: Os resultados, na maioria dos estudos, são positivos ou negativos?

QP4: Por quais motivos o *Python* é a linguagem mais utilizada para aprendizado de máquina?

3.3 Processo de pesquisa

No primeiro momento foi definido o tema do mapeamento realizando uma pesquisa de artigos através do Google Acadêmico, a primeira busca foi feita no contexto de “aprendizado através de jogos digitais”, onde não foi obtido sucesso em encontrar artigos relevantes para a utilização neste trabalho de mapeamento, após isso foi realizada mudanças no tema algumas vezes, para por exemplo: “aprendizado de *Python* através de jogos digitais” e “uso de *Python* para desenvolvimento de jogos de aprendizagem”, onde também não foi encontrados artigos relevantes. Os resultados se mostraram promissores quando foi produzida uma mudança no foco de pesquisa para “aplicação de *Python* com *Machine Learning*” e após a leitura do resumo de diversos artigos, houve então a definição do tema: “Quais são, atualmente, as diferentes aplicações de *Python* em *Machine Learning*”.

Após definir o tema, houve a possibilidade então, através da mesma pesquisa e estudo, de estabelecer o objetivo principal do trabalho de mapeamento sistemático: “Investigar o estado da arte do uso de *Python* no desenvolvimento de aplicações de aprendizagem de máquina”.

Para a definição da *string de pesquisa*, foram utilizadas duas *keywords* para representar melhor o contexto do mapeamento sistemático: “*Python*” e “*Machine Learning*”, foram encontrados resultados promissores em três sites, são eles: IEEE, ACM e Elsevier conforme apresentado na Quadro 1. Esta pesquisa priorizou artigos em inglês que estão em “*open access*” e que foram aplicados entre 2020 e 2023, pois o foco da mesma é ter estudos recentes como insumo do mapeamento. Após realizar alguns testes definiu-se a seguinte *string de pesquisa*: *abstract* possui as palavras “*Python*” e “*Machine Learning*” ou *document title* possui as palavras “*Python*” e “*Machine Learning*” ou *author keywords* possui as palavras “*Python*” e “*Machine Learning*”, ou seja, tecnicamente a *string* de pesquisa utilizada foi a seguinte:

- ("Abstract": "Python" AND "Abstract": "Machine Learning") OR ("Document Title": "Python" AND "Document Title": "Machine Learning") OR ("Author Keywords": "Python" AND "Author Keywords": "Machine Learning")

Quadro 1 - Artigos retornados por cada base de dados eletrônica e seleção final

Base de Dados	Quantidade
ACM Digital Library	42
Elsevier (Science Direct)	207
IEEE Xplore	25
Total	274
Seleção final	25

Fonte: Autoria própria (2023)

Os demais sites de artigos não foram utilizados por um ou vários dos seguintes motivos:

- Não possui filtro de “open access”;
- Não possui filtro de um ou vários dos campos de busca (*abstract*, *document title* e *author keywords*);
- O site está desatualizado ou não retornou nenhum artigo.

É importante ressaltar que a pesquisa foi realizada no mês de janeiro do ano de 2023, utilizando o filtro de data e que no site da ACM foi necessário contar manualmente os artigos *open access*, por conta de o mesmo não fornecer um filtro para esta finalidade. No primeiro momento, todos os artigos encontrados foram analisados, a fim de constatar a relevância para a utilização neste estudo e responder às questões de pesquisa.

3.4 Critérios de exclusão

Realizada a triagem com o registro de informações relevantes e uma divisão dos estudos encontrados, através da leitura do título, resumo, conclusão e *keywords* o número de artigos foi reduzido para 41. Após isso foram aplicadas regras de

exclusão para cada um dos 41 artigos pré-selecionados com o objetivo de validar as informações catalogadas. As regras de exclusão aplicadas foram as seguintes:

- Artigos sem resumo, conclusão ou *Keywords*;
- Artigos que são tutoriais de aplicação de biblioteca *Python*;
- Artigos com resultado inconclusivo (onde o(s) autor(es) não expressaram claramente na conclusão que o que foi realizado é melhor ou teve melhores resultados em comparação com outros estudos/pesquisas/métodos);
- Artigos onde o tema ou a aplicação está relacionada a TI;
- Artigos que não possuem “Python” e “*Machine Learning*” em seu resumo descritivo;
- Artigo com resumo inconclusivo, onde o autor não foi capaz de expor o conteúdo de forma clara.

Após a aplicação dessas regras de exclusão o resultado obtido foi de 25 estudos. É importante destacar que durante esse processo foi obtida a área de aplicação dos artigos resultantes e logo abaixo está a relação da quantidade de artigos por área de aplicação. É possível notar que a Medicina é a área que mais se tem utilizado da aplicação de aprendizado máquina em seus estudos e pesquisas com um total de 8 artigos, áreas como Cibersegurança, Geologia e Plantas aparecem em segundo lugar com 2 artigos cada. É interessante notar a grande diversidade das áreas de atuação, sendo 16 áreas distintas.

Quadro 2 - Quantidade de artigos por área de atuação

Área de atuação	Quantidade
Aeroespacial	1
Agricultura	1
Cibersegurança	2
Circuito de software	1
Clima	1
Energia/eletricidade	1
Geologia	2

Imagens de microscópio	1
Indústria	1
Ligas metálicas	1
Medicina	8
Pecuária leiteira	1
Plantas	2
Reciclagem/reutilização de resíduos	1
Redes de celulares	1
Sistema de refrigeração	1
Total	25

Fonte: Autoria própria (2023)

3.5 Categorização dos artigos selecionados

Depois de realizar a leitura na íntegra dos artigos foi possível encontrar conceitos, assuntos e palavras chaves que permitiram efetuar a classificação em algumas categorias relevantes levando em consideração as questões de pesquisa a serem respondidas e o tema deste trabalho. As categorias foram definidas com base no objetivo de cada artigo, abaixo estão listadas as categorias definidas:

- **Análise comparativa:** Esta categoria inclui os artigos que propuseram uma análise comparativa de determinada metodologia de aprendizado de máquina com outras, propostas por outros artigos;
- **Aplicação e experimentos com modelo proposto:** Esta categoria abrange os artigos que demonstraram a aplicação e os experimentos realizados com um modelo de aprendizado de máquina proposto;
- **Estratégia de otimização:** Os artigos desta categoria, demonstram estratégias de otimização de resultados na aplicação de aprendizado de máquina;
- **Nova arquitetura de aprendizado profundo:** Estudos classificados nessa categoria demonstram uma nova arquitetura de aprendizado de máquina, com foco em aprendizado profundo;

- **Nova biblioteca:** Artigos que apresentam as funcionalidades de uma nova biblioteca *Python* para implementação em aprendizado de máquina;
- **Novo algoritmo:** Esta categoria inclui os estudos desenvolvidos com o propósito de apresentar um novo algoritmo;
- **Novo framework:** Os artigos desta categoria exibem as funcionalidades presentes no novo *framework* desenvolvido para aprendizado de máquina;
- **Novo pipeline:** Demonstra o uso e as aplicações de um novo pipeline.
- **Novo simulador:** Os estudos classificados nesta categoria demonstram a implementação de um simulador baseado em aprendizado de máquina;
- **Revisão sistemática:** Estudos que aplicaram a metodologia de revisão sistemática.

Importante frisar que essas categorias não definem completamente os artigos estudados, visto que vários deles possuem uma grande quantidade de informações referentes a sua área de aplicação e problema proposto, artigos esses que expressam perfeitamente a justificativa para sua realização e a necessidade de se criar, por exemplo uma nova biblioteca. É através destes estudos que as pesquisas podem ser testadas, validadas e conseqüentemente resultam na evolução da área de aplicação e a resolução de novos problemas são criadas. É possível observar claramente no quadro 3 que a categoria mais encontrada foi “Aplicação e experimentos com modelo proposto” com 15 artigos sendo categorizados dessa forma. Isso mostra o grande interesse no desenvolvimento de novos modelos e mostrar os resultados obtidos.

Quadro 3 - Quantidade de artigos por categoria

Categorias	Quantidade
Análise comparativa	1
Aplicação e experimentos com modelo proposto	15
Estratégia de otimização	1

Nova arquitetura de aprendizado profundo	1
Nova biblioteca	2
Novo algoritmo	1
Novo framework	1
Novo pipeline	1
Novo simulador	1
Revisão sistemática	1
Total	25

Fonte: Autoria própria (2023)

4 DESENVOLVIMENTO

Nesta seção são apresentadas as descobertas relevantes realizadas durante todo o processo de aplicação da metodologia, desde a pesquisa até a leitura dos artigos. Aqui estão registradas as informações sobre o uso do *Python* e suas bibliotecas, os algoritmos de aprendizado de máquina utilizados nos artigos e a importância do uso dos dados em cada trabalho.

4.1 Uso de Python e suas bibliotecas

Chang, Bhavani, Xu, 2022 destacaram em seu artigo que *Python* é amplamente considerada como a linguagem de programação mais eficaz e útil. Ele contém diversas bibliotecas que foram utilizadas no projeto proposto de aprendizado de máquina. Como visto na literatura, nesta abordagem o aprendizado de máquina é considerado um subconjunto do modelo de Inteligência Artificial. As bibliotecas de *Python* foram usadas para fazer previsões poderosas com *Scikit Learn*, que é uma ferramenta de previsão e aplicação prática de aprendizado de máquina.

A flexibilidade de executar diversos algoritmos é composta por um grande poder de aplicação e uma sintaxe distinta de outras linguagens de programação. Além disso, o *Python* provou ser uma linguagem apropriada para avaliar grandes conjuntos de dados com a ajuda de algoritmos de aprendizado de máquina para receber *insights* significativos, conforme Copeland (2019 apud CHANG 2022). A linguagem *Python* é considerada uma linguagem de programação muito adequada tanto para simulações, quanto para programação do mundo real. Ela é considerada a linguagem de alto nível mais poderosa para aplicação em aprendizado de máquina (MEDIUM, 2021).

Isso mostra quão poderosa e abrangente se tornou a utilização de *Python* em aprendizado de máquina, visto que as aplicações podem ser realizadas em diferentes áreas e com múltiplas finalidades, visando sempre a facilidade de implementação e o poder de processamento, tornando as execuções das tarefas mais rápidas e confiáveis. Foi observada a utilização e desenvolvimento das ferramentas Python para a resolução de problemas desde a agricultura até a área espacial, de diversas aplicações em medicina até a pecuária leiteira, sempre

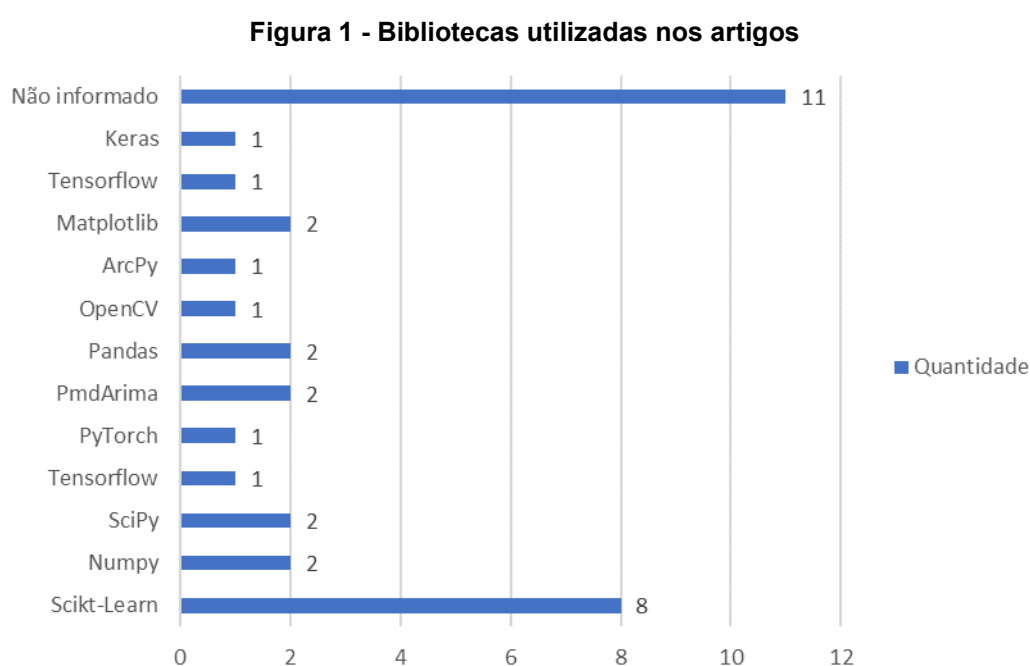
utilizando diversos recursos disponíveis pela própria linguagem de programação e por estudos e ferramentas de outros autores.

Os serviços de desenvolvimento do *Python* são uma opção adequada para esta que é uma linguagem forte e poderosa usada para incentivar habilidades computacionais na obtenção de *insights* valiosos das informações, por exemplo, dos pacientes que sofrem de doenças cardíacas, que, por sua vez, ajudarão no suporte a aplicativos baseados em saúde. É oportuno caso alguém tenha que entregar-se à diversidade de desenvolver algo com a ajuda de conexão com a Internet ou tenha trabalhado de forma autônoma sem nenhuma conexão com a Internet (SRINATH, 2017 apud ISMAIL, 2022).

Além do *Python*, outras linguagens de programação usadas para aprendizado de máquina incluem R, *Java*, C++ e *Julia*. R é uma linguagem muito popular para computação estatística e análise de dados, enquanto que *Java* e C++ são conhecidos por sua velocidade e eficiência, tornando-os adequados para aplicativos de aprendizado de máquina em larga escala. *Julia* é uma linguagem de programação relativamente nova que foi projetada especificamente para computação numérica e computação científica, incluindo aprendizado de máquina. Apesar dos recursos dessas outras linguagens de programação, o Python ganhou força significativa na comunidade de aprendizado de máquina devido à sua simplicidade, flexibilidade e disponibilidade de muitas bibliotecas e estruturas poderosas, incluindo *TensorFlow*, *Keras*, *PyTorch* e *Scikit Learn* (CANALTECH, 2021; TERRA, 2021; DATA SCIENCE ACADEMY, 2021; ACERVO LIMA, 2020).

A linguagem também é preferida pelos cientistas de dados devido à disponibilidade de extensas bibliotecas, como *SciPy*, *Pandas*, *Numpy* e muitas outras (CHANG, BHAVANI, XU, 2022). As bibliotecas *Python* são coleções de código pré-escrito que fornecem um conjunto de ferramentas e funcionalidades aos desenvolvedores, facilitando o desenvolvimento de aplicativos e sistemas de software. O *Python* possui um vasto número de bibliotecas, muitas projetadas especificamente para aprendizado de máquina, análise de dados, desenvolvimento web e computação científica. No geral, as bibliotecas *Python* desempenham um papel importante no desenvolvimento de aplicações e fornecem aos desenvolvedores uma variedade de ferramentas e funcionalidades que podem acelerar de forma significativa o processo de desenvolvimento (AWARI, 2022).

Ao realizar o levantamento destes dados, foi observado que nem todo artigo possuía a declaração explícita de qual biblioteca *Python* foi utilizada, desta forma é possível que esteja enviesado. No entanto é possível observar que a quantidade de artigos que não informaram a biblioteca utilizada é bem expressiva, sendo no total 11 artigos, a biblioteca mais utilizada foi a *Scikit Learn*, sendo mencionada em 8 artigos, as bibliotecas *Keras*, *Tensorflow*, *ArcPy*, *OpenCV*, *PyTorch* e *Tensorflow* foram utilizadas em um artigo (podendo ser no mesmo artigo), as demais bibliotecas foram utilizadas em 2 artigos cada.



Fonte: Autoria própria (2023)

4.2 Algoritmos de aprendizado de máquina

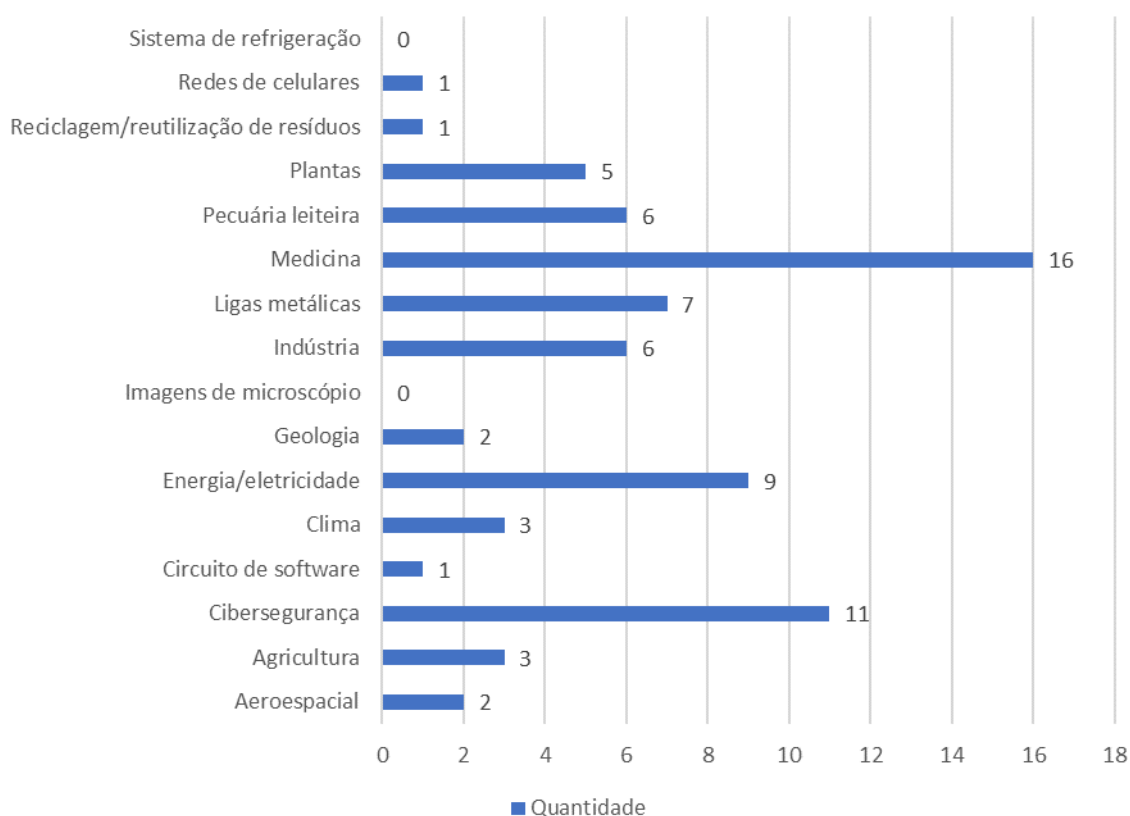
Não só a utilização das bibliotecas é importante no *Python* como também os algoritmos. Um algoritmo é uma sequência de raciocínios, instruções ou operações para alcançar um objetivo, sendo necessário que os passos sejam finitos e operados sistematicamente para que se possa obter o resultado esperado (ROCK CONTENT, 2019). O *Python* possui uma ampla variedade de algoritmos de aprendizado de máquina que são frequentemente utilizados para várias tarefas, como classificação, regressão, agrupamento e detecção de anomalias. Alguns dos tipos mais comuns de algoritmos de aprendizado de máquina em *Python* incluem:

- **Regressão Linear:** é um algoritmo de regressão que prevê um valor contínuo, como a previsão de preços de imóveis com base em seu tamanho e localização;
- **Regressão Logística:** um tipo de algoritmo de classificação que prevê um resultado binário, como por exemplo: se um cliente vai comprar um produto ou não;
- **Árvores de Decisão:** é um tipo de algoritmo que usa um modelo de árvore de decisões e suas possíveis consequências para que possa fazer previsões;
- **Random Forest:** algoritmo de conjunto que combina várias árvores de decisão para melhorar a precisão das previsões;
- **K-Nearest Neighbors:** este é um tipo algoritmo de classificação que prevê a classe de um ponto de um grupo de dados observando as classes de seus k vizinhos mais próximos;
- **Support Vector Machines:** um tipo de algoritmo de classificação que tenta encontrar um hiperplano no espaço de alta dimensão e que separa diferentes classes de dados;
- **Redes Neurais:** é uma família de algoritmos inspirados na estrutura e na função do cérebro humano, são usados para uma vasta gama de tarefas, incluindo o reconhecimento de imagens e o processamento de linguagem natural (CRONAPP, 2019).

Estes são apenas alguns exemplos dos tipos mais comuns de algoritmos de aprendizado de máquina em *Python*. Dependendo da tarefa e do conjunto de dados específicos utilizados, outros algoritmos podem ser mais apropriados e eficazes, podendo gerar resultados melhores e mais assertivos, por isso é importante observar a exploração e os experimentos realizados com diferentes algoritmos para encontrar o melhor ajuste para cada aplicação executada nos artigos. Levando em conta estas informações, foi realizado o levantamento dos diferentes algoritmos citados nas pesquisas por área de atuação. É importante esclarecer que alguns algoritmos não foram contabilizados por conta da forma com que foram nomeados nos artigos.

De acordo com a figura 2 é possível observar que a área da Medicina possui, com grande diferença, a maior quantidade de algoritmos distintos, contando com 16, seguido por Cibersegurança e Energia/Eletricidade com 11 e 9 tipos de algoritmos, respectivamente. Os artigos de Sistema de refrigeração e Imagens de microscópio não possuíam nenhuma declaração de uso de algoritmos, enquanto que Redes de celulares, Reciclagem/Reutilização de resíduos e Circuito de software declararam apenas um algoritmo, os artigos das demais áreas de atuação declararam entre 1 e 9.

Figura 2 - Quantidade de diferentes algoritmos por área de atuação

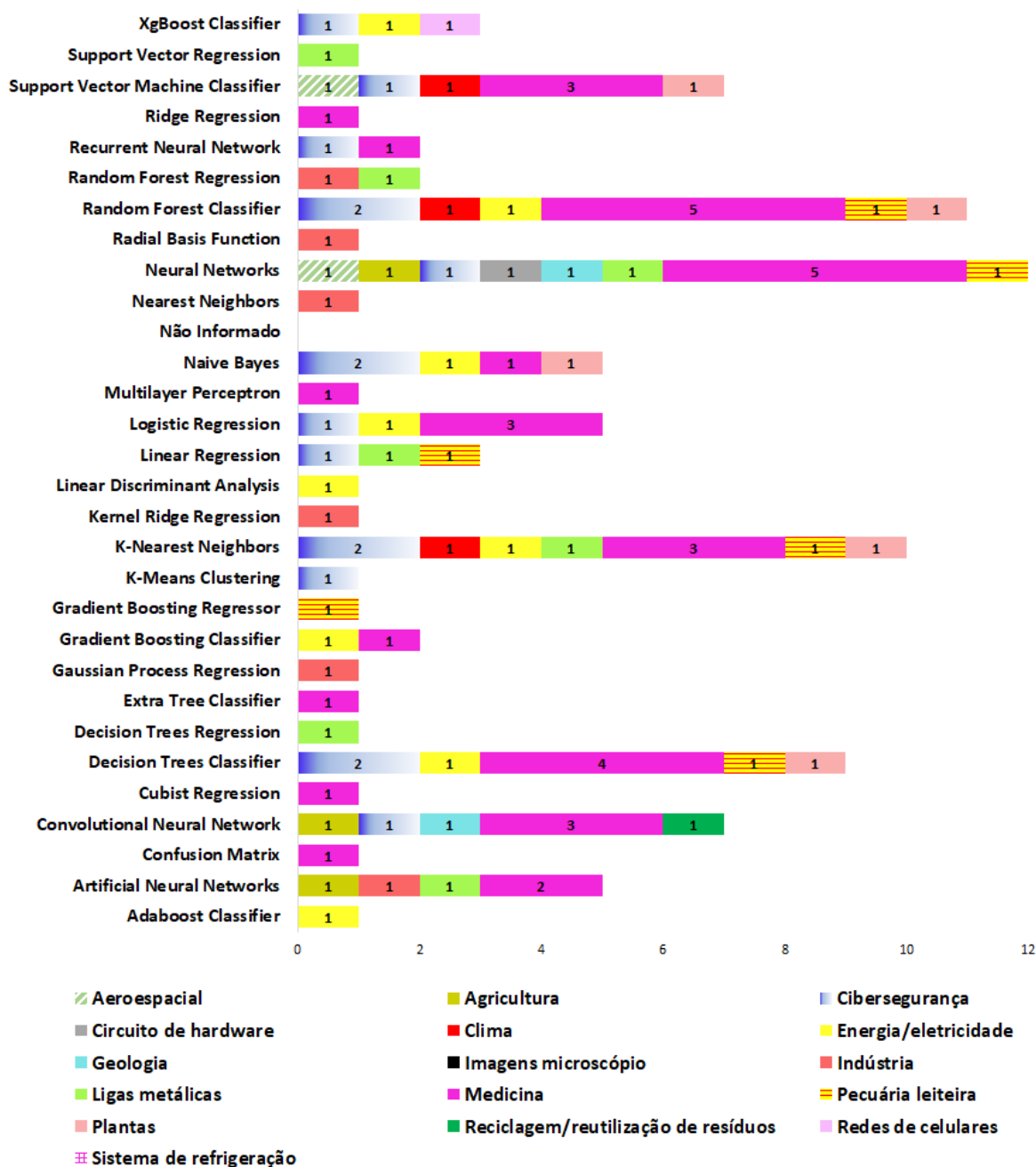


Fonte: Autoria própria (2023)

Ao compilar todas as informações coletadas em relação às áreas de atuação e aos algoritmos utilizados, foi realizada a montagem de um gráfico para possibilitar a visualização da quantidade de artigos com menções dos algoritmos por área de atuação, conforme exibido na figura 3. É possível observar que os algoritmos *Random Forest Classifier* e *Neural Networks* foram os mais utilizados na área da Medicina, onde 5 artigos possuem declaração de utilização para cada um destes algoritmos (podendo ser os mesmos artigos entre diferentes algoritmos). Outro dado

importante é que o algoritmo *Neural Networks* está presente em 12 artigos, representando o algoritmo que mais foi utilizado, seguido de perto por *Random Forest Classifier* e *K-Nearest Neighbors*, presentes em 11 e 10 artigos respectivamente. Energia/Eletricidade é a área de atuação com menos artigos e com mais algoritmos utilizados, tendo apenas um artigo, declararam a utilização de 9 algoritmos diferentes, seguida pela área de atuação Ligas metálicas, que também possui apenas um artigo e utiliza 7 algoritmos diferentes.

Figura 3 - Quantidade de artigos com menções dos algoritmos por área de atuação



Fonte: Autoria própria (2023)

4.3 Uso de dados

Tão importante quanto a escolha e utilização dos algoritmos corretos é a utilização de dados corretos para a realização dos treinamentos e testes dos algoritmos de aprendizado de máquina. Alguns artigos utilizaram a geração de dados virtuais para serem utilizados em seus procedimentos, enquanto que outros

baixaram os dados de repositórios públicos ou dos sites oficiais de instituições públicas, como citamos nos exemplos abaixo:

- Após a calibração dos modelos, um conjunto de experimentos virtuais foram realizados para gerar dados virtuais para cada *chiller*. As entradas deste modelo foram variadas de forma aleatória em suas faixas válidas e com as frequências esperadas em um sistema real, seguindo o hipercubo latino, que é um método de amostragem. Isso tornou possível caracterizar o sistema em todo o espaço de entrada, o que não seria possível se fossem dados reais a serem usados diretamente (ZABALA *et al.*, 2020).
- A seleção do conjunto de dados UNSW-nb15 foi realizada do *GitHub*, que possui dados sobre os ataques *DDoS*. Este conjunto de dados é disponibilizado pelo *Australian Centre for Cyber Security (ACCS)* (SRINATH, 2017 apud ISMAIL *et al.*, 2022). O conjunto de dados utilizado consiste em diferentes informações sobre os ataques *DDoS*, incluindo: número de ID, representação do meio da rede, rótulo dos ataques e a gravidade dos ataques (ISMAIL *et al.*, 2022).
- Foi baixado manualmente dados meteorológicos diários de *Dublin Phoenix Park* pelo período de um ano entre 01/06/2019 e 01/06/2020 do NOAA (*National Oceanic and Atmospheric Administration*), com informações de temperatura diária máxima e mínima, a quantidade diária de precipitação, temperatura média diária e a profundidade diária da neve. Esses dados são compilados em um arquivo *.CSV* e foram utilizados como a entrada bruta para os modelos de aprendizado de máquina. O valor de cada variável é registrado em relação à data, desta forma, cada variável é uma série temporal com um intervalo de tempo básico de um dia (WU *et al.*, 2022)

5 RESULTADOS

Esta seção apresenta os resultados consolidados da pesquisa realizada. Além disso, as questões de pesquisa são revisitadas e respondidas

5.1 Resposta às questões de pesquisa

Para tal, são utilizadas as quatro questões de pesquisa definidas no início do processo de mapeamento sistemático: “**QP1**: Em quais áreas há o maior número de aplicações?”; “**QP2**: Quais são as principais bibliotecas utilizadas nos trabalhos encontrados?”; “**QP3**: Os resultados, na maioria dos estudos são positivos ou negativos?”; “**QP4**: Por quais motivos o *Python* é a linguagem mais utilizada para aprendizado de máquina?”.

Ao fim da pesquisa e após ser feito o levantamento das informações pertinentes ao tema do trabalho, como: esclarecimentos sobre a linguagem de programação *Python*, sobre o campo de aprendizado de máquina, às áreas de atuação dos artigos, bibliotecas e algoritmos utilizados, obtenção e utilização dos dados, entre outras os seguintes resultados foram obtidos:

- Nesta pesquisa, a Medicina é área que possui mais aplicações do aprendizado de máquina com o uso do *Python*, sendo 8 os artigos encontrados nesta área (**QP1** – ver Quadro 2);
- Declarada em 8 artigos, a biblioteca *Scikit Learn* foi a mais utilizada nos trabalhos selecionados, seguida por *Matplotlib*, *Pandas*, *PmdArima*, *Scipy* e *Numpy*, que foram utilizadas em 2 artigos cada, sendo que diferentes bibliotecas podem ter sido usadas num mesmo artigo (**QP2** – ver Figura 1);
- Todos os 25 artigos da seleção final possuem declarações por parte de seus autores que obtiveram resultados positivos na implementação do aprendizado de máquina com a linguagem *Python* (**QP3**);
- São vários os motivos que fazem o *Python* ser a linguagem mais utilizada em aprendizado de máquina, alguns deles incluem: é considerada a linguagem de programação mais eficaz e útil, ela

contém diversas bibliotecas, possui flexibilidade de executar diversos algoritmos, é composta por um grande poder de aplicação e uma sintaxe distinta de outras linguagens de programação, além disso, o *Python* provou ser uma linguagem apropriada para avaliar grandes conjuntos de dados com a ajuda de algoritmos de aprendizado de máquina, é considerada uma linguagem de programação muito adequada tanto para simulações, quanto para programação do mundo real, ela é considerada a linguagem de alto nível mais poderosa para aplicação em aprendizado de máquina (**QP4** – ver Seções 2.3 e 4.1).

Em relação a questão de pesquisa principal deste trabalho “Quais áreas vêm utilizando mais soluções baseadas em aprendizagem de máquina desenvolvidas com *Python*?” foi constatado que são mais amplas e variadas possíveis, pois incluem desde as aplicações que foram encontradas nos artigos estudados, como por exemplo: Cibersegurança, Aeroespacial e Medicina, e também em outras áreas e outras formas de aplicações que foram observadas durante as pesquisas, alguns exemplos incluem: processamento de linguagem natural, reconhecimento de fala, robótica, classificação e reconhecimento de padrões em imagens e vídeos, dentre muitos outros. O mais importante a se notar é que novas aplicações estão sendo desenvolvidas constantemente à medida que a tecnologia avança e as empresas descobrem novas maneiras de aproveitar o poder do aprendizado de máquina, desde a resolução de problemas até o mapeamento de novas aplicações.

6 CONCLUSÕES

Este trabalho teve como principal objetivo realizar uma pesquisa exploratória utilizando como metodologia o mapeamento sistemático a fim de investigar o estado da arte do uso de *Python* no desenvolvimento de aplicações de aprendizagem de máquina, e dessa forma proporcionar uma visão geral do que tem sido aplicado e esclarecer em quais cenários e por quais motivos estão sendo feitas essas aplicações.

Para cumprir com esse objetivo, realizou-se a definição de quatro questões de pesquisa, a sondagem de assuntos relevantes para a utilização na pesquisa genérica e *string* de pesquisa e foi estabelecido os filtros que seriam utilizados nos buscadores acadêmicos, com isso foi possível obter a seleção final de 25 artigos. Assim sendo, houveram diversas descobertas relevantes durante todo o processo de aplicação da metodologia, como por exemplo: as informações sobre o uso do *Python* e suas bibliotecas, os algoritmos de aprendizado de máquina utilizados nos artigos e a importância da utilização dos dados em cada trabalho.

Como contribuição, esta pesquisa apresenta um mapeamento sistemático recente e que abrange a investigação do estado da arte da aplicação de *Python* em aprendizado de máquina, tendo sido possível responder às questões de pesquisa e várias percepções (*insights*) foram coletados, contribuindo assim para uma visão, ainda que geral, do domínio investigado abrangendo seu entendimento. Como observado, as áreas de aplicação são mais amplas e variadas possíveis, pois incluem desde as aplicações que foram encontradas nos artigos estudados, como também em outras áreas que foram observadas durante as pesquisas, alguns exemplos incluem: processamento de linguagem natural, reconhecimento de fala, robótica, classificação e reconhecimento de padrões em imagens e vídeos, dentre muitos outros. O mais interessante é que novas aplicações estão sendo desenvolvidas constantemente à medida que a tecnologia avança.

6.1 Trabalhos futuros

Como trabalhos futuros, é sugerido a criação de novas questões de pesquisa para abranger e alcançar novas temáticas visando melhores resultados,

ampliar os filtros de pesquisa com o intuito de deixar o mapeamento menos tendencioso amplificando assim os resultados obtidos. A partir deste estudo é possível realizar a aplicação das questões de pesquisa (com as devidas adaptações) para as demais subáreas da inteligência artificial como: busca (*pathfinding*), planejamento automatizado, processamento de linguagem natural, representação de conhecimento, raciocínio e raciocínio probabilístico, robótica e percepção, aplicação da IA (SANTOS, 2015).

Outra sugestão é implementar os mesmos procedimentos deste trabalho de mapeamento para as demais linguagens de programação que podem ser utilizadas em aprendizado de máquina, como as já citadas: R, *Java*, C++ e *Julia* com a finalidade de observar se os resultados e a abrangência das áreas de aplicação são as mesmas, se há diferenças nos resultados obtidos e comparar todas essas informações.

REFERÊNCIAS

ACERVO LIMA. 2020. **As 5 principais linguagens de programação e suas bibliotecas para aprendizado de máquina em 2020**. Disponível em: <https://acervolima.com/as-5-principais-linguagens-de-programacao-e-suas-bibliotecas-para-aprendizado-de-maquina-em-2020/>. Acesso em: 02 mar. 2023.

ALTITUDE TVM. [s.d.]. **Entendendo o python com os pontos fortes e fracos do python, você sabe?**. Disponível em: <https://altitudetvm.com/pt/komputer/1232-pengertian-python-beserta-kelebihan-dan-kekurangan-python-sudah-tahu.html>. Acesso em: 02 mar. 2023.

ALURA. 2021. **Python: origem do nome**. Disponível em: <https://www.alura.com.br/artigos/python-origem-do-nome>. Acesso em: 02 mar. 2023.

PORTAL ARBIT. 2018, **Machine learning: utilize na sua empresa**. Disponível em: <https://blog.arbit.com.br/machine-learning-utilize-na-sua-empresa/#:~:text=O%20conceito%20de%20Machine%20Learning,para%20ganhar%20diferencia%C3%A7%C3%A3o%20e%20escala>. Acesso em: 02 mar. 2023.

AWARI. 2022. **Conheça as principais bibliotecas python para data science**. Disponível em: https://awari.com.br/bibliotecas-python-para-data-science/?utm_source=blog. Acesso em: 02 mar. 2023.

CANALTECH. 2021. **5 melhores linguagens de programação para aprendizado de máquina**. Disponível em: <https://canaltech.com.br/mercado/5-melhores-linguagens-de-programacao-para-aprendizado-de-maquina-220750/>. Acesso em: 02 mar. 2023.

CHANG, V. *et al.*, **An artificial intelligence model for heart disease detection using machine learning algorithms**, Healthcare Analytics, Volume 2, 2022, 100016, ISSN 2772-4425. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772442522000016>. Acesso em: 19 fev. 2023.

CRONAPP. 2019. **Algoritmos de machine learning: conheça os 6 principais e descubra os impactos na sua empresa**. Disponível em: <https://blog.cronapp.io/algoritmos-de-machine-learning/>. Acesso em: 02 mar. 2023.

DATA SCIENCE ACADEMY. 2021. **Por que a linguagem python é tão popular em machine learning e inteligência artificial?**. Disponível em: <https://blog.dsacademy.com.br/por-que-a-linguagem-python-e-tao-popular-em-machine-learning-e-inteligencia-artificial/>. Acesso em: 02 mar. 2023.

ENGENHARIA HÍBRIDA. 2022. **Machine learning: como funciona, tipos, aplicações**. Disponível em: <https://www.engenhariahibrida.com.br/post/machine-learning-como-funciona-tipos-aplicacoes>. Acesso em: 02 mar. 2023.

IMPACTA. 2020. **Python: melhor linguagem para machine learning, entenda!**. Disponível em: <https://www.impacta.com.br/blog/python-melhor-linguagem-para-machine-learning-entenda/>. Acesso em: 02 mar. 2023.

ISMAIL *et al.* **A machine learning** - based classification and prediction technique for ddos attacks, in IEEE Access, vol. 10, pp. 21443-21454, 2022, doi: 10.1109/ACCESS.2022.3152577. Disponível em: <https://ieeexplore.ieee.org/document/9716094>. Acesso em: 19 fev 2023.

L TECNOLOGIA. 2021. **Vantagens e desvantagens da linguagem python**. Disponível em: <https://ltecnologia.com.br/blog/vantagens-e-desvantagens-da-linguagem-python/>. Acesso em: 02 mar. 2023.

LEARNING DATA. 2021. **Fundamentos básicos do machine learning que você precisa saber**. Disponível em: <https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber>. Acesso em: 02 mar. 2023.

MEDIUM, 2021, **Heart disease detection using machine learning in python**. Disponível em: <https://randerson112358.medium.com/heart-diseasedetection-using-machine-learning-python-a701f39396cb>, Acesso em: 19 fev. 2023.

PETERSEN, K. *et al.* **Systematic Mapping Studies in Software Engineering**. Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering. 17, 2008.

BARANAUSKAS, J. A. DCM FFCLRP USP, [s.d.]. **AM I - Aprendizado de máquina conceitos e definições**. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Conceitos-Definicoes.pdf>. Acesso em: 02 mar. 2023.

ROCK CONTENT. 2019. **Saiba como funciona um algoritmo e conheça os principais exemplos existentes no mercado**. Disponível em: <https://rockcontent.com/br/blog/algoritmo/#:~:text=Um%20algoritmo%20%C3%A9%20uma%20sequ%C3%Aancia%20de%20racioc%C3%ADnios%2C%20instru%C3%A7%C3%B5es%20ou%20opera%C3%A7%C3%B5es,de%20informa%C3%A7%C3%B5es%20mediadas%20pelas%20instru%C3%A7%C3%B5es..> Acesso em: 02 mar. 2023.

SANTOS, F. M. IASC – INTELIGÊNCIA ARTIFICIAL SOB CONTROLE, 2015, **Subáreas da inteligência artificial do ponto de vista computacional**. Disponível em: <https://iascblog.wordpress.com/2015/10/23/subareas-da-inteligencia-artificial-do-ponto-de-vista-computacional/>. Acesso em: 14 mar. 2023.

TERRA. 2021. **5 melhores linguagens de programação para aprendizado de máquina**. Disponível em: <https://www.terra.com.br/byte/5-melhores-linguagens-de-programacao-para-aprendizado-de-maquina,420320c50210d3fbc8fd95a079b69d64g1htdi93.html>. Acesso em: 02 mar. 2023.

WAZLAWICK, R. S. 2014, **Metodologia de pesquisa para ciência da computação**. 2a ed. Ro de Janeiro. Elsevier.

WU, J. *et al.* **Boosting climate analysis with semantically uplifted knowledge graphs**, in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4708-4718, 2022, doi: 10.1109/JSTARS.2022.3177463. Disponível em: <https://ieeexplore.ieee.org/document/9780580>. Acesso em: 19 fev 2023

ZABALA, L. *et al.* **Virtual testbed for model predictive control development in district cooling systems**, Renewable and Sustainable Energy Reviews, Volume 129, 2020, 109920, ISSN 1364-0321, Disponível em: <https://www.sciencedirect.com/science/article/pii/S2772442522000016>. Acesso em: 19 fev 2023