

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
BACHARELADO EM GESTÃO DA INFORMAÇÃO**

**TIAGO OLIVEIRA SPIRONELLO**

**GUIA COMPARATIVO DE CLASSIFICADORES DE TEXTO PARA OS ODS**

**UBERLÂNDIA  
2024**

TIAGO OLIVEIRA SPIRONELLO

**GUIA COMPARATIVO DE CLASSIFICADORES DE TEXTO PARA OS ODS**

Trabalho de Conclusão de Curso  
apresentado à Universidade Federal de  
Uberlândia como requisito à Conclusão de  
curso em Gestão da Informação.

Orientadora: Profa. Dra. Carla Bonato  
Marcolin

**UBERLÂNDIA**

**2024**

## RESUMO

Os algoritmos de classificação de texto são ferramentas computacionais que automatizam a tarefa de atribuir categorias ou rótulos a textos com base em seu conteúdo. O objetivo deste trabalho é realizar uma revisão sistemática da literatura com o propósito de identificar e analisar os principais algoritmos de classificação utilizados na categorização de materiais relacionados aos 17 Objetivos de Desenvolvimento Sustentável (ODS). A análise da base de dados revelou uma lacuna significativa na disponibilidade de recursos dedicados a essa temática, sugerindo que este é um campo ainda pouco explorado, especialmente considerando o prazo definido para a conclusão e avaliação dos resultados da Agenda 2030. Os principais algoritmos de classificação utilizados nesse contexto foram identificados, destacando-se a Regressão Logística, Naive Bayes, Máquina de Vetores de Suporte, Floresta Aleatória, Redes Neurais, K-ésimo Vizinho mais Próximo e Árvore de Decisão. Dentre eles, a Máquina de Vetores de Suporte demonstrou uma performance consistente e obteve excelentes resultados de classificação, sendo amplamente abordada nos estudos revisados. Esta pesquisa contribui para a compreensão dos principais algoritmos de classificação e para a identificação do mais adequado para análises desse tipo, visando potencializar o alcance das metas estabelecidas pelos ODS.

Palavras-chave: Classificação de texto; ODS; Inteligência Artificial; Agenda 2030.

## **ABSTRACT**

Text classification algorithms are computational tools that automate the task of assigning categories or labels to texts based on their content. The aim of this study is to conduct a systematic literature review to identify and analyze the main classification algorithms used in categorizing materials related to the 17 Sustainable Development Goals (SDGs). The analysis of the database revealed a significant gap in the availability of resources dedicated to this theme, suggesting that this is still a largely unexplored field, especially considering the deadline for the conclusion and evaluation of the Agenda 2030 results. The main classification algorithms used in this context have been identified, with Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, Neural Networks, K-Nearest Neighbor, and Decision Tree standing out. Among them, the Support Vector Machine demonstrated consistent performance and achieved excellent classification results, being widely addressed in the reviewed studies. This research contributes to understanding the main classification algorithms and identifying the most suitable one for such analyses, aiming to enhance the achievement of the goals established by the SDGs.

**Keywords:** Text classification; SDGs; Artificial Intelligence; Agenda 2030.

## LISTA DE ABRESVIATURAS E SIGLAS

ONU	Organização das Nações Unidas
ODM	Objetivos de Desenvolvimento do Milênio
ODS	Objetivos de Desenvolvimento Sustentável
IA	Inteligência Artificial
PLN	Processamento de Linguagem Natural
SDG	Sustainable Development Goals
SVM	Máquina de Vetores de Suporte
NB	Naive Bayes
DT	Árvore de Decisão
KNN	K-ésimo Vizinho mais Próximo
LR	Regressão logística
RF	Floresta aleatória

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	7
<b>2 MÉTODO DE PESQUISA</b> .....	11
<b>3 DISCUSSÃO E RESULTADOS</b> .....	13
<b>3.1 Análise descritiva da base</b> .....	13
<b>3.2 Principais modelos de classificação</b> .....	18
<b>3.2.1 Regressão logística (Logistic regression)</b> .....	19
<b>3.2.2 Naïve Bayes</b> .....	21
<b>3.2.3 Máquina de vetores de suporte (Support vector machine)</b> .....	22
<b>3.2.4 Árvore de decisão (Decision tree)</b> .....	23
<b>3.2.5 Floresta aleatória (Random forest)</b> .....	25
<b>3.2.6 Redes neurais (Neural networks)</b> .....	26
<b>3.2.7 K-ésimo vizinho mais próximo (K-nearest neighbors algorithm)</b> .....	27
<b>4 CONCLUSÃO</b> .....	30
<b>REFERÊNCIAS</b> .....	31

## 1 INTRODUÇÃO

No ano 2000, a Organização das Nações Unidas (ONU), com o apoio de 191 nações, estabeleceu um conjunto de metas com o propósito de enfrentar desafios globais, especialmente relacionados à pobreza, saúde, educação, igualdade de gênero e meio ambiente. Essas metas ficaram conhecidas como Objetivos de Desenvolvimento do Milênio (ODM), comprometendo as nações a unirem esforços em prol dessa causa, em um prazo de 15 anos para sua realização. Essas metas foram criadas para mobilizar esforços internacionais, promovendo a cooperação global em áreas cruciais de desenvolvimento humano, orientar políticas nacionais e internacionais, bem como para direcionar recursos para onde eram mais necessários, e estabelecido um horizonte temporal para orientar esses empreendimentos (ROMA, 2019).

Visando os resultados significativos que esses esforços conjuntos atingiram até o ano de 2015, os agora 193 Estados-membros da ONU adotaram um novo documento conhecido como Agenda 2030. Este plano cooperativo é ainda mais ambicioso do que seu precursor, com a finalidade de eliminar a extrema pobreza, enfrentar a desigualdade e a injustiça, bem como conter as mudanças climáticas até o ano de 2030 (ROMA, 2019). Para fornecer orientação, foram estabelecidos os 17 Objetivos de Desenvolvimento Sustentável (ODS), que se desdobram em 169 metas. Esses ODS têm o mesmo propósito de auxiliar os governos internacionais na elaboração, implementação e monitoramento de políticas públicas voltadas para o alcance dessas metas de maneira organizada e uniforme (ONU, 2015).

Os Objetivos de Desenvolvimento Sustentável, demonstrados na Figura 1, desempenham um papel relevante no contexto global atual. Sua concepção visa abranger uma ampla variedade de questões interconectadas, reconhecendo a impossibilidade de abordar os desafios globais de maneira isolada. Nesse sentido, eles fomentam parcerias estratégicas entre governos, setor privado, sociedade civil e comunidade científica, promovendo uma abordagem colaborativa como via principal para a resolução dessas questões globais. Em essência, esses objetivos concentram seus esforços na redução das desigualdades, tanto no interior das nações quanto entre elas e enfatizam que a promoção da justiça social e da inclusão assume um compromisso indispensável no alcance de um desenvolvimento sustentável verdadeiramente significativo e abrangente (ONU, 2015).

Figura 1 - 17 Objetivos de Desenvolvimento Sustentável



Fonte: Grupo de Trabalho da Sociedade Civil para a Agenda 2030.

O foco na sustentabilidade é um dos princípios fundamentais da Agenda 2030, enfatizando a importância de abordar as necessidades do presente sem comprometer a capacidade das gerações futuras de atender às suas próprias necessidades. Os ODS não buscam apenas soluções imediatas para problemas, mas visam criar um futuro mais resiliente e duradouro, implicando em adotar práticas que preservem e regenerem recursos naturais, reduzam a desigualdade e garantam que as melhorias na qualidade de vida sejam sustentáveis a longo prazo (SILVA, 2022). Também enfatizam a importância da educação e conscientização sobre a sustentabilidade, isso inclui a promoção de conhecimentos e habilidades que capacitem as pessoas a tomar decisões mais sustentáveis em suas vidas diárias.

Cada ODS é acompanhado por metas específicas e mensuráveis, desdobradas em indicadores quantificáveis. Esses indicadores têm uma importância fundamental, permitindo que governos, organizações e a sociedade avaliem o progresso de maneira objetiva, promovam a responsabilidade, tomem decisões informadas, comparem o desempenho internacionalmente e garantam uma abordagem abrangente, considerando aspectos econômicos, sociais e ambientais no caminho em direção ao desenvolvimento sustentável global. Por exemplo, uma das metas do ODS 2: Fome Zero e Agricultura Sustentável tem como objetivo "Até 2030, acabar com a fome e garantir o acesso de todas as pessoas, em particular os pobres e as pessoas



em situações vulneráveis, incluindo crianças, a alimentos seguros, nutritivos e suficientes todo o ano." (ONU, 2015) e se faz mensurável por meio da coleta e análise de dados sobre disponibilidade de alimentos, acesso a alimentos nutritivos e taxas de desnutrição, permitindo que os países e organizações ajustem suas estratégias para alcançar o objetivo de acabar com a fome até 2030.

Porém não é apenas o governo responsável pelo alcance dessas metas, a sociedade como um todo tem sua parcela de responsabilidade, e as universidades assumem uma função indispensável na promoção e realização dos ODS, destacando-se especialmente por seu impacto significativo na pesquisa científica, com produção de conhecimento científico e tecnológico essenciais para a criação de soluções concretas para desafios globais. Além de contribuir com soluções para esses desafios, as pesquisas possuem o potencial de fornecer evidências empíricas para embasar políticas públicas e ações voltadas para o desenvolvimento sustentável.

As universidades também desempenham um papel crucial na capacitação de profissionais conscientes, colaborar com diferentes instituições e liderar a adoção de práticas sustentáveis em suas comunidades. No entanto, enfrentam desafios significativos, como métricas de produção científica que priorizam muitas vezes a quantidade em detrimento da qualidade e do impacto social, restrições financeiras, necessidade de parcerias e comunicação eficaz para públicos externos à academia. Estes desafios destacam a importância de um compromisso contínuo das universidades na busca pela realização dos ODS (SERAFIM e LEITE, 2021).

Neste contexto, a Inteligência Artificial (IA) exerce uma função essencial, principalmente na participação da comunidade científica no avanço desses objetivos. A IA, por meio de algoritmos avançados e análise de dados em larga escala, tem o potencial de impulsionar significativamente a consecução dos ODS de diversas maneiras. Ela oferece a capacidade de análise abrangente de dados, identificar padrões e tendências, prever eventos futuros e otimizar o uso de recursos em setores-chave, como energia, transporte e agricultura. Além disso, a IA facilita o monitoramento contínuo do progresso em direção aos ODS, tornando-a uma ferramenta aliada essencial em todas as áreas.

Atualmente, pode-se segmentar a atuação da Inteligência Artificial em duas vertentes distintas. A primeira delas é a capacidade de utilizar a tecnologia computacional para atingir efetivamente as metas estabelecidas pela Agenda 2030. Um exemplo notável é o projeto PlantVillage, que emprega a IA para auxiliar

agricultores no diagnóstico de doenças em suas plantações, desempenhando um papel essencial na promoção da segurança alimentar e na redução da pobreza, alinhando-se de maneira eficaz com o ODS 15 - Vida Terrestre (SANTOS DIVINO, 2021). O uso adequado dessa tecnologia pode resultar em uma ampla gama de soluções, abrangendo desde previsão de desastres naturais e cuidados de saúde acessíveis até a agricultura sustentável e educação personalizada. No entanto, é fundamental abordar questões éticas e legais para assegurar que os benefícios se estendam a todas as comunidades.

Por outro lado, a IA se faz bastante relevante no monitoramento contínuo do progresso em relação aos ODS. Ela capacita a análise abrangente de relatórios, documentos e dados provenientes de múltiplas fontes, permitindo a identificação de tendências, deficiências e áreas prioritárias que necessitam de atenção. Essa capacidade é essencial para uma compreensão aprofundada das metas e objetivos menos explorados, garantindo que nenhum aspecto seja negligenciado (DHIMAN, 2023).

Além disso, a IA pode auxiliar na disseminação e acessibilidade da produção científica e dos materiais mais relevantes e citados relacionados aos ODS. Isso inclui facilitar a tradução automática desses recursos para vários idiomas e regiões, o que, por sua vez, amplia consideravelmente o alcance global desses materiais (DHIMAN, 2023). Dessa forma, a IA desempenha um papel multifacetado na promoção e no cumprimento dos ODS, seja atuando no cumprimento efetivo das metas, ou impulsionando a compreensão, o monitoramento e a conscientização em nível global.

Utilizando como foco a atuação da IA no monitoramento da produção científica relacionada à Agenda 2030 e classificação desses materiais nos 17 Objetivos de Desenvolvimento Sustentável, o presente estudo tem por objetivo os classificadores, mais especificamente os classificadores de texto, algoritmos ou modelos utilizados em aprendizado de máquina e análise de dados para categorizar ou rotular materiais textuais em diferentes classes ou categorias com base no conteúdo do texto, sendo o caso, os ODS. A contribuição do estudo é compreender os principais algoritmos de classificação a fim de identificar o mais adequado para esse tipo de análise de forma a potencializar o alcance das metas estabelecidas pelos ODS.

Para alcançar esse objetivo é preciso entender a parte técnica por trás dos algoritmos de processamento de linguagem natural (PLN). Os classificadores de texto, operam através do processamento de dados textuais para identificar padrões e

características que os permitam categorizar os documentos em diferentes classes ou temas. Esses algoritmos utilizam uma variedade de técnicas, como a extração de características, tokenização, análise sintática e semântica, e modelagem de tópicos, muitas vezes combinando abordagens tradicionais de PLN com métodos mais avançados de aprendizado de máquina, como redes neurais (JURAFSKY, 2020). Compreender a funcionalidade e as nuances desses algoritmos é essencial para selecionar e adaptar as ferramentas mais adequadas para seu propósito específico.

## **2 MÉTODO DE PESQUISA**

Como o objetivo é identificar e analisar os principais algoritmos de classificação utilizados na categorização de materiais relacionados aos 17 Objetivos de Desenvolvimento Sustentável, a metodologia adotada neste estudo consiste na realização de uma revisão bibliométrica, que constitui uma análise sistemática e quantitativa da produção acadêmica em uma área específica de estudo (SOARES, 2018).

O termo "bibliometria" refere-se à aplicação de técnicas estatísticas e matemáticas para mensurar e analisar a produção, distribuição e utilização de informações presentes em documentos bibliográficos. Os resultados obtidos por meio desse processo conferem maior segurança, robustez e credibilidade às previsões e às decisões tomadas com base nessas análises (VANTI, 2002).

A revisão sistemática se destaca por sua abordagem detalhada e reproduzível no campo científico, diferenciando-se de outras revisões. Seu propósito é reduzir o viés presente em pesquisas não estruturadas e facilitar extensões para futuras investigações. Ao analisar a literatura de maneira sistemática, é possível avaliar a relevância de diversos autores em um determinado tema, fornecendo uma base sólida para o progresso do conhecimento a partir de estudos anteriores e, deste modo, identificando lacunas de pesquisa importantes.

Com o intuito de alcançar o objetivo estabelecido, foi conduzida uma busca por periódicos e materiais científicos acadêmicos nas bases de dados Web of Science e Scopus por meio do Periódicos Capes. A pesquisa utilizou a seguinte linha de busca: "(\"ODS\" OR \"SDG\" OR \"agenda2030\") AND (\"machine learning\" OR \"artificial intelligence\" OR \"NLP\" OR \"text classification\" OR \"aprendizado de máquina\" OR \"inteligência artificial\" OR \"processamento de linguagem natural\" OR \"classificador de

texto)”. O propósito é reunir uma significativa parcela da produção científica relacionada à inteligência artificial e à classificação de texto que aborde os Objetivos de Desenvolvimento Sustentável, a fim de não perder nenhum material relevante para este estudo. A pesquisa foi realizada em “Tópico” a qual engloba: Títulos, Resumos e Palavras-Chaves.

Um período temporal específico foi selecionado para a análise, abrangendo materiais publicados a partir de 2015, para se alinhar ao ano de publicação da Agenda 2030 pela ONU. Na base de dados Web of Science, identificaram-se 140 documentos, enquanto na Scopus, foram encontrados 284. Com a exclusão de 92 duplicatas presentes em ambas as bases, o total de documentos consolidados foi de 332. É importante ressaltar que o tipo dos documentos não foi limitado para essas pesquisas, dada a escassez de produção acadêmica específica sobre esse tema.

Após analisar os resumos dos materiais pré-selecionados, foram excluídos da avaliação aqueles que não trataram do tema central deste trabalho: a classificação de textos nos 17 Objetivos de Desenvolvimento Sustentável. A quantidade de documentos na base foi reduzida de 332 para 18, uma vez que a maioria deles abordou aplicações práticas de Inteligência Artificial alinhadas à agenda, tais como sistemas de monitoramento agrícola para previsão de safras, identificação de pragas e otimização de recursos (AMENYEDZI, 2023); diagnóstico médico assistido por IA para aprimorar a prevenção e tratamento de doenças, aumentando a eficiência dos sistemas de saúde (YOON, 2023); e o uso de sensores e algoritmos para monitorar a qualidade da água e detectar vazamentos em sistemas de abastecimento, contribuindo para a gestão sustentável dos recursos hídricos (MEHMOOD, 2020). Durante a leitura, identificou-se também materiais que atribuem outros significados às siglas "ODS" e "SDG", como "Oxide Dispersion Strengthened (ODS) steels" (SUMMERS, 2023), "Synthetic Data Generation (SDG) algorithms" (GALLONI, 2023) e "Obstacle Detection System (ODS)" (KOT, 2022), os quais também foram excluídos da base de estudo.

O material intitulado “Analytics of machine learning-based algorithms for text classification”, de autoria de Sayar Ul Hassana, Jameel Ahamed, Khaleel Ahmada, foi incluído manualmente na base de dados devido à sua potencial contribuição positiva para este projeto.

O procedimento técnico adotado nesta pesquisa emprega o Bibliometrix, um pacote do software R-Studio, que possibilita análises tanto quantitativas quanto

qualitativas, enquanto realiza simultaneamente o mapeamento de dados bibliográficos. Essa ferramenta apresenta um notável potencial devido à sua natureza de código aberto, continuamente aprimorada por uma extensa comunidade de desenvolvedores. Além disso, pode ser adaptado e aprimorado pelos próprios usuários, permitindo alinhar a pesquisa de maneira mais eficaz às necessidades específicas (DERVIS, 2019). O Bibliometrix possibilita a análise de desempenho e a avaliação de dados científicos presentes em bases de dados. Seus resultados indicam os grupos de pesquisadores que abordam uma temática específica, incluindo informações sobre países, periódicos relevantes e colaboração internacional. Além disso, a ferramenta permite o mapeamento de estudos científicos, visando revelar a estrutura dinâmica dos dados e representar os arranjos cognitivos em uma área de pesquisa (COBO MARTIN, 2011).

### **3 DISCUSSÃO E RESULTADOS**

A estrutura desta seção está subdividida em duas partes, sendo a primeira: a análise quantitativa e qualitativa da base de documentos, realizada por meio do Bibliometrix. Esta abordagem inclui a distribuição anual das produções relacionadas ao tema, os periódicos de maior relevância, os autores mais influentes e uma representação visual por meio de uma nuvem de palavras.

Na segunda metade desta seção, será realizada uma análise aprofundada dos classificadores mais abordados nos documentos, explorando as perspectivas dos autores sobre cada um, destacando pontos positivos e negativos. Além disso, será discutido o contexto e a importância dessa classificação na produção acadêmica em relação aos 17 Objetivos de Desenvolvimento Sustentável.

#### **3.1 Análise descritiva da base**

O Quadro 1 a seguir apresenta os documentos utilizados nesta análise bibliométrica, selecionados com base na leitura de seus resumos e na interpretação pessoal de sua potencial contribuição positiva para o contexto em questão.

Quadro 1 - Documentos da base

Ano	Título	Autores	Journal/ Periódico	Qualis/ Capes	Tipo do documento
2023	A STUDY ON HIGHER EDUCATION STUDENTS ATTITUDE TOWARD ARTIFICIAL INTELLIGENCE IN ACHIEVING SUSTAINABLE SOCIAL EMPOWERMENT IN ASSAM	GUPTA S;BHATTACHARJEE S	STUDIES IN BIG DATA	-	CAPÍTULO DE LIVRO
2023	DISCOVERING NEW PATHWAYS TOWARD INTEGRATION BETWEEN HEALTH AND SUSTAINABLE DEVELOPMENT GOALS WITH NATURAL LANGUAGE PROCESSING AND NETWORK SCIENCE	SMITH T;VACCA R;MANTEGAZZA L;CAPUA I	GLOBALIZATION AND HEALTH	A2	ARTIGO
2023	THE SUSTAINABLE DEVELOPMENT GOALS AND AEROSPACE ENGINEERING A CRITICAL NOTE THROUGH ARTIFICIAL INTELLIGENCE	SÁNCHEZ-RONCERO A;GARIBO-I-ORTS ;CONEJERO J;EIVAZI H;MALLOR F;ROSENBERG E;FUSO-NERINI F;GARCÍA-MARTÍNEZ J;VINUESA R;HOYAS S	RESULTS IN ENGINEERING	B4	ARTIGO
2023	SAILING THE DATA SEA TO ADVANCE RESEARCH ON THE SUSTAINABLE DEVELOPMENT GOALS	SPEZZATTI A;KHERADMAND E;GUPTA K;PERAS M;ZAMINPEYMA R	PHILOSOPHICAL STUDIES SERIES	-	CAPÍTULO DE LIVRO
2023	ARTIFICIAL INTELLIGENCE FOR SUSTAINABLE DEVELOPMENT GOALS BIBLIOMETRIC PATTERNS AND CONCEPT EVOLUTION TRAJECTORIES	SINGH A;KANAUJIA A;SINGH V;VINUESA R	SUSTAINABLE DEVELOPMENT	A1	ARTIGO
2023	SEMISUPERVISED EVENT PREDICTIONS WITH GRAPH NETWORKS	NEMEC P;LEBAN G;ROZANEC J;GROBELNIK M	ICPE 2023 - COMPANION OF THE 2023 ACM/SPEC INTERNATIONAL CONFERENCE ON PERFORMANCE ENGINEERING	-	ARTIGO DE CONFERÊNCIA
2023	ARTIFICIAL INTELLIGENCE AND SUSTAINABLE DEVELOPMENT GOALS NEXUS VIA FOUR VANTAGE POINTS	NASIR O;JAVED R;GUPTA S;VINUESA R;QADIR J	TECHNOLOGY IN SOCIETY	A1	ARTIGO
2023	UNPACKING THE SUSTAINABLE DEVELOPMENT GOALS SDGS INTERLINKAGES A SEMANTIC NETWORK ANALYSIS OF THE SDGS TARGETS	SONG J;JANG C	SUSTAINABLE DEVELOPMENT	A1	ARTIGO
2022	A ROBERTA APPROACH FOR AUTOMATED PROCESSING OF SUSTAINABILITY REPORTS	ANGIN M;TAŞDEMİR B;YILMAZ C;DEMIRALP G;ATAY M;ANGIN P;DIKMENER G	SUSTAINABILITY (SWITZERLAND)	A2	ARTIGO
2023	OPERATIONALIZING DIGITAINABILITY ENCOURAGING MINDFULNESS TO HARNESS THE POWER OF DIGITALIZATION FOR SUSTAINABLE DEVELOPMENT	GUPTA S;CAMPOS Z J;DEL R C G;TOMIČIĆ A;ANDRÉS M S;MAHFOUZ M;OSEMWEIGIE I;PHEMIA C S V;SCHMITZ M;MAHMOUD N;NYAREGH M	SUSTAINABILITY (SWITZERLAND)	A2	ARTIGO
2022	MAPPING THE SUSTAINABLE DEVELOPMENT GOALS SDGS IN SCIENCE TECHNOLOGY AND INNOVATION APPLICATION OF MACHINE LEARNING IN SDGORIENTED ARTEFACT DETECTION	HAJIKHANI A;SUOMINEN A	SCIENTOMETRICS	A1	ARTIGO
2022	MASSIVE OPEN ONLINE COURSES ROLE IN PROMOTING UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS	WANG Y;LI Y;HU X;XU Y;LIANG X;LEI C	PROCEEDINGS OF 2022 IEEE LEARNING WITH MOOCS, LWMOOCS 2022	-	ARTIGO DE CONFERÊNCIA
2021	AUTOMATED MAPPING OF ENVIRONMENTAL HIGHER EDUCATION RANKING SYSTEMS INDICATORS TO SDGS INDICATORS USING NATURAL LANGUAGE PROCESSING AND DOCUMENT SIMILARITY	BUZABOON A;ALBOFLASA H;ALNASER W;SHATNAWI S;ALBINALI K	2021 INTERNATIONAL CONFERENCE ON INNOVATION AND INTELLIGENCE FOR INFORMATICS, COMPUTING, AND TECHNOLOGIES, 3ICT 2021	-	ARTIGO DE CONFERÊNCIA
2021	SDGS A RESPONSIBLE RESEARCH ASSESSMENT TOOL TOWARD IMPACTFUL BUSINESS RESEARCH	RODENBURG K;DE S V;HUGHES J	SUSTAINABILITY (SWITZERLAND)	A2	ARTIGO
2020	AI FOR MONITORING THE SUSTAINABLE DEVELOPMENT GOALS AND SUPPORTING AND PROMOTING ACTION AND POLICY DEVELOPMENT	MILLER L;BOLTON M;BOULTON J;MINTROM M;NICHOLSON A;RUDIGER C;SKINNER R;RAVEN R;WEBB G	2020 IEEE / ITU INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE FOR GOOD, AI4G 2020	-	ARTIGO DE CONFERÊNCIA
2020	SUSTAINABLE DEVELOPMENT GOALS AND CITIES	ALVAREZ-RISCO A;DEL-AGUILA-ARCENTALES S;ROSEN M	BUILDING SUSTAINABLE CITIES: SOCIAL, ECONOMIC AND ENVIRONMENTAL FACTORS	-	CAPÍTULO DE LIVRO
2021	HIDDEN IN PLAIN SIGHT BUILDING A GLOBAL SUSTAINABLE DEVELOPMENT DATA CATALOGUE	HODSON J;SPEZZATTI A	LECTURE NOTES IN NETWORKS AND SYSTEMS	-	ARTIGO DE CONFERÊNCIA
2021	THE INTERRELATION OF SUSTAINABLE DEVELOPMENT GOALS IN PUBLICATIONS AND PATENTS A MACHINE LEARNING APPROACH	HAJIKHANI A;SUOMINEN A	CEUR WORKSHOP PROCEEDINGS	-	ARTIGO DE CONFERÊNCIA
2022	ANALYTICS OF MACHINE LEARNING-BASED ALGORITHMS FOR TEXT CLASSIFICATION	HASSAN SU;AHAMED J;AHMAD K	SUSTAINABLE OPERATIONS AND COMPUTERS	-	ARTIGO

Fonte: Elaboração própria.

Desde a divulgação da Agenda 2030 em 2015 pela Organização das Nações Unidas, observa-se uma notável expansão nas produções acadêmicas relacionadas ao desenvolvimento sustentável, abrangendo suas diversas áreas de estudo. Nos últimos anos, com a crescente popularização da Inteligência Artificial, esses dois temas têm convergido no meio científico, tornando o uso da tecnologia para a promoção da sustentabilidade um tópico atual e de grande relevância.

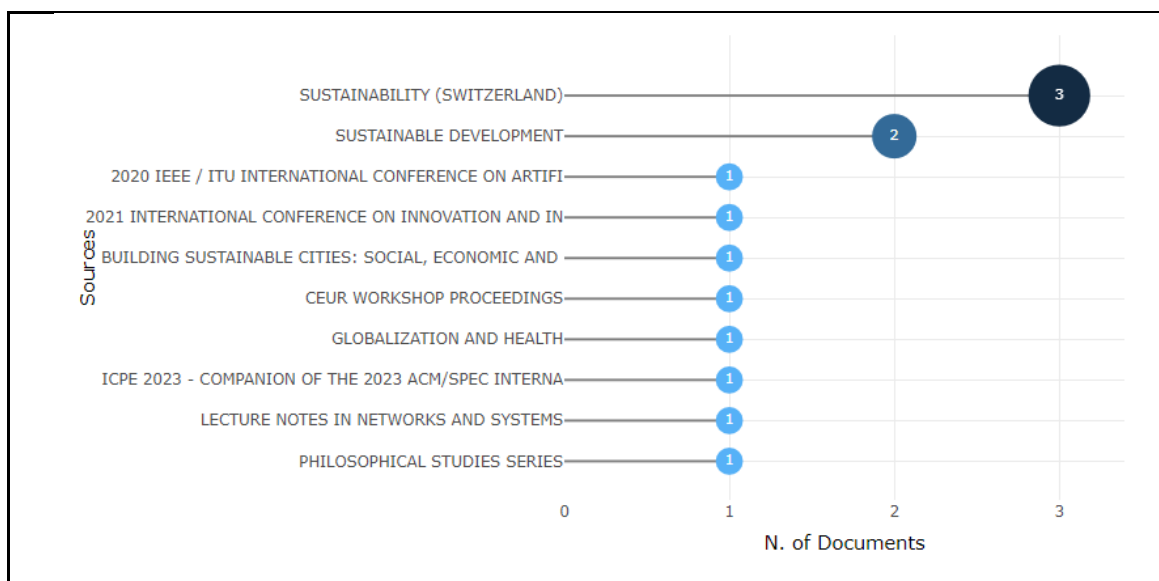
Os primeiros estudos específicos identificados na pesquisa datam do ano de 2020, totalizando 2 produções. A partir dessa data, observa-se uma tendência ascendente, com 4 produções registradas nos anos de 2021 e 2022. Destaca-se, no entanto, que a relevância atingiu seu ponto máximo no último ano, 2023, com mais que o dobro de produções, totalizando 9.

Esta análise revela uma perspectiva otimista em relação ao tema, alimentando a expectativa de que até 2030 haja uma exploração mais profunda e abrangente, visando preencher lacunas e contribuir significativamente para a classificação das produções mundiais nos 17 ODS.

Na Figura 2, são apresentados os 10 periódicos mais significativos da base de dados, evidenciando que o periódico "Sustainability" desempenha um papel proeminente na contribuição para o campo da classificação de texto relacionado aos Objetivos de Desenvolvimento Sustentável. Em seguida, o "Sustainable Development" se destaca com duas publicações, enquanto os demais periódicos contam com apenas uma.

Destaca-se a escassez de publicações sobre o tema nos periódicos analisados, reforçando que esta é uma área de estudo ainda pouco explorada, com uma distribuição limitada das produções entre os principais periódicos.

**Figura 2 - Periódicos mais presentes**

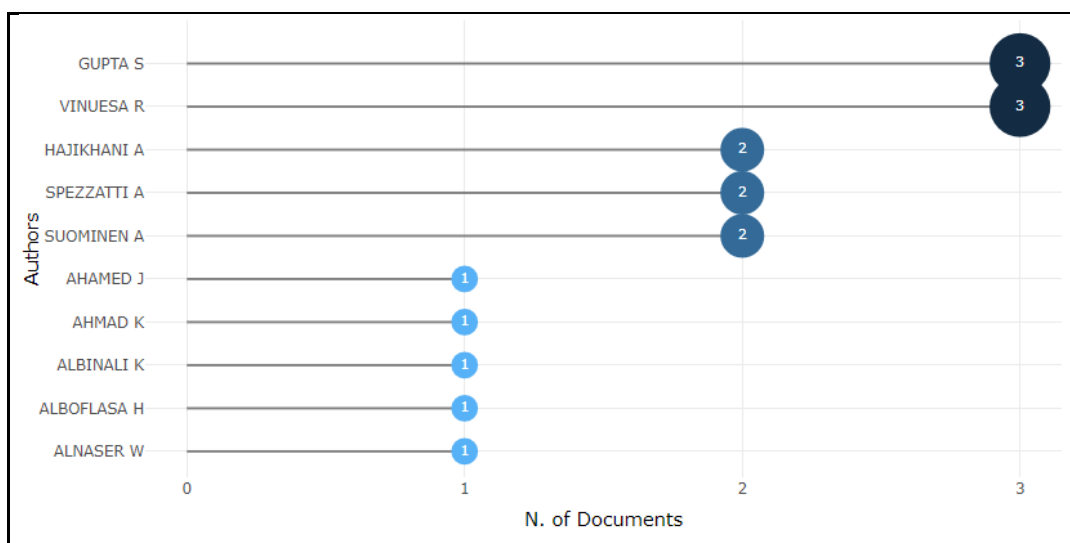


Fonte: Bibliometrix (2024).

Em seguida, procedeu-se à análise dos autores mais relevantes, apresentando um ranking de suas produções na Figura 3, juntamente com uma representação gráfica das colaborações na Figura 4, cujas cores estão normalizadas de acordo com as afiliações dos autores.

Observa-se que os dois autores mais produtivos no âmbito do tema, "Gupta S" e "Vinuesa R", não apenas lideram em termos de produções, mas também se destacam por sua significativa colaboração. Notavelmente, eles são os únicos a engajar em trabalhos colaborativos com autores afiliados a diferentes instituições.

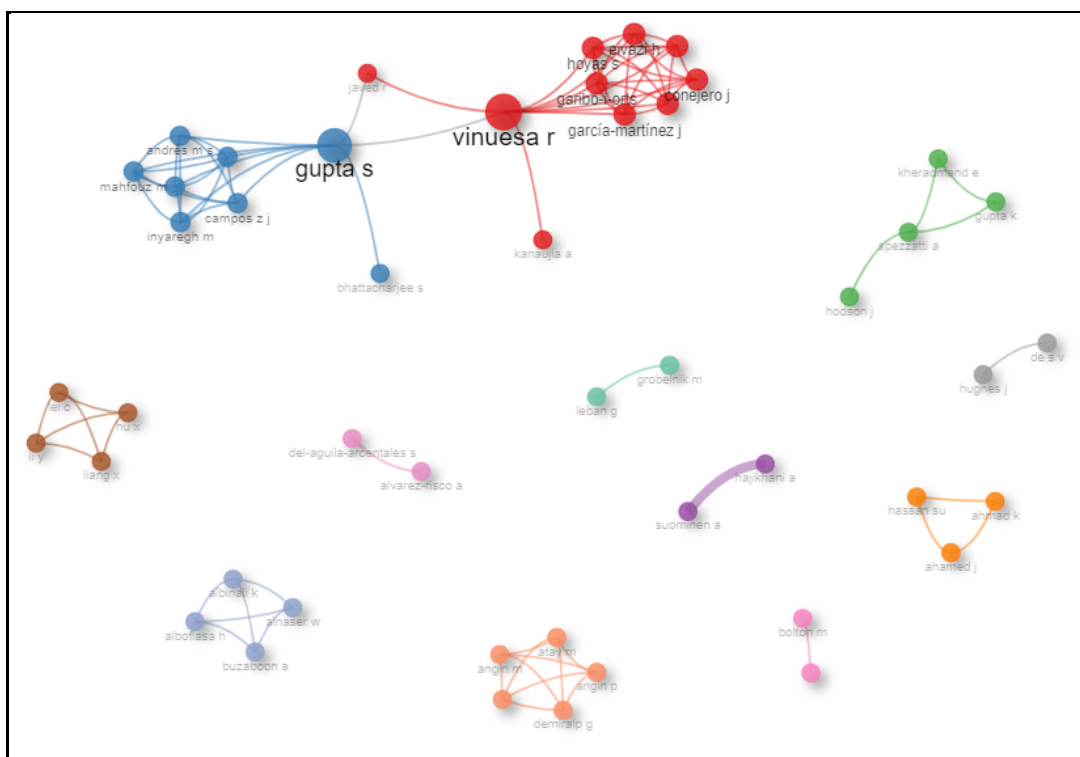
**Figura 3 - Autores mais presentes**



Fonte: Bibliometrix (2024).



**Figura 4 - Rede de colaboração**



**Fonte:** Bibliometrix (2024).

A Figura 5 apresenta uma nuvem de palavras, destacando as 50 palavras-chave mais frequentes nos materiais analisados neste estudo. É notável que as expressões "Sustainable Development Goal" (Objetivo de Desenvolvimento Sustentável) e "Sustainable Development" (Desenvolvimento Sustentável) se destacam como as mais relevantes, refletindo o foco central deste trabalho nos ODS e na Agenda2030. Em seguida, foi observado a ênfase em termos relacionados ao segundo tema central de nossa análise, como "Artificial Intelligence" (Inteligência Artificial), "Machine Learning" (Aprendizado de Máquina), "Learning Systems" (Sistemas de Aprendizagem) e "Natural Language Processing" (Processamento de Linguagem Natural). Como já mencionado, a classificação de texto por inteligência artificial faz parte de uma vertente do aprendizado de máquinas conhecido como processamento de linguagem natural, que se concentra na interação entre computadores e linguagem humana.



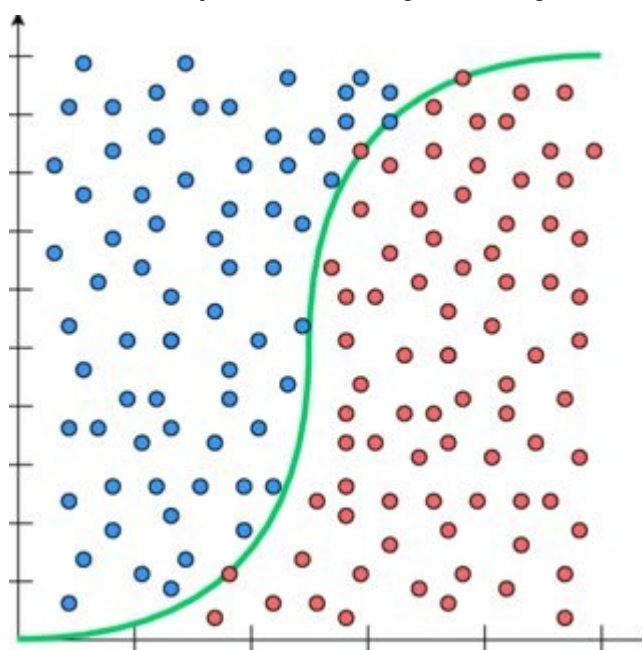
Dessa forma, o trabalho busca preencher lacunas e oferecer uma visão abrangente sobre a aplicação e eficácia dos algoritmos de classificação em contextos relacionados aos Objetivos de Desenvolvimento Sustentável, considerando a diversidade de abordagens encontradas na literatura revisada.

O Quadro 2, apresentado ao término desta seção, sintetiza os principais algoritmos de classificação discutidos nos materiais, acompanhados por um contador de ocorrências e as respectivas performances analisadas. Além destes, outros algoritmos, tais como Otimização Sequencial Mínima (SINGH, 2024), Fatoração de Matriz Não Negativa (SÁNCHEZ-RONCERO, 2023), Representações Distribuídas de Tópicos (SÁNCHEZ-RONCERO, 2023) e Word2Vec (SONG, 2023), também foram mencionados. No entanto, sua presença não foi suficientemente expressiva para permitir comparações significativas.

### **3.2.1 Regressão logística (Logistic regression)**

A Regressão Logística é uma técnica popular e eficaz utilizada em problemas de classificação binária. Ao contrário do nome sugere, a Regressão Logística não é uma técnica de regressão, mas sim uma abordagem probabilística que estima a probabilidade de uma instância pertencer a uma classe específica. O modelo emprega a função logística para transformar a soma ponderada das características de uma instância em uma probabilidade no intervalo de 0 a 1 (MOHI, 2020). Esta probabilidade é então comparada a um limiar, geralmente 0,5, para determinar a classificação final. O treinamento do modelo envolve a otimização dos parâmetros, ajustando-os de modo a maximizar a verossimilhança dos dados observados. A Regressão Logística é particularmente útil em situações onde se busca entender a relação entre variáveis independentes e a probabilidade de ocorrência de um evento. A Figura 6 exemplifica uma função sigmóide em verde, dividindo o espaço amostral.

**Figura 6** - Classificação utilizando Regressão Logística



**Fonte:** Adaptado de "A Systematic Review on Airlines Industries based on Sentiment Analysis and Topic Modeling" de Hasib, Khan (2023).

Ao analisar o desempenho do algoritmo de Regressão Logística nos materiais que embasam este estudo, inicialmente no trabalho de Angin (2022), foi conduzida uma comparação com os algoritmos Máquina de Vetores de Suporte (SVM), Naive Bayes (NB) e Árvore de Decisão (DT) para os 15 primeiros Objetivos de Desenvolvimento Sustentável (ODS), utilizando a métrica de qualidade F1-Score. Em apenas uma ocasião, ele superou todos os outros, alcançando um F1-Score de 0.93 para o ODS 9. Adicionalmente, o algoritmo também apresentou os melhores resultados para os ODS 2, 4, 6, 8, 11, 12, 13, 14, porém nestes com o resultado igualado ao SVM. Nesta comparação, os tipos de texto classificados foram relatórios de sustentabilidade, publicações governamentais, notas políticas e literatura acadêmica relacionada à sustentabilidade.

No estudo conduzido por Hajikhani (2022), o desempenho do algoritmo foi comparado com o SVM e o NB para os 16 primeiros ODS, utilizando métricas como F1-Score, Precision e Recall, e categorizando textos como publicações científicas e documentos de patentes. Nessa avaliação, o algoritmo se destacou especialmente para o ODS 15, emergindo como o melhor algoritmo em todas as três métricas, e para o ODS 3, liderando em Precision e F1-Score. É relevante notar que essa análise foi realizada em conjunto com outros dois algoritmos, Word2Vec com Regressão

Logística e Doc2Vec com Regressão Logística, os quais foram excluídos dessa comparação de desempenho por não estarem entre os algoritmos selecionados.

### **3.2.2 Naïve Bayes**

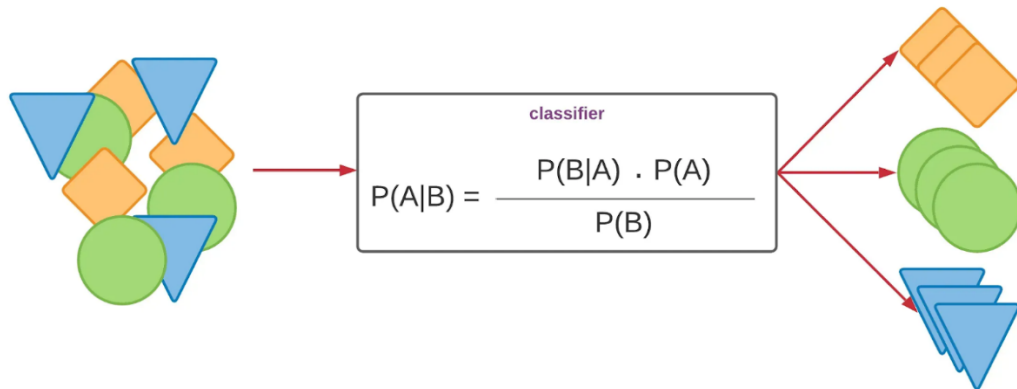
O algoritmo de classificação Naïve Bayes é fundamentado no teorema de Bayes e destaca-se pela simplicidade e eficiência. A abordagem "Naïve" deriva da suposição de independência condicional entre as características, o que simplifica os cálculos. O modelo calcula as probabilidades condicionais das características dadas às classes durante o treinamento (BILAL, 2015). Existem diferentes variantes do Naïve Bayes, sendo as duas mais comuns: Gaussiana e Multinomial.

Na variante Gaussiana, assume-se que as características seguem uma distribuição normal, ou gaussiana. Este modelo é adequado para dados contínuos e quantitativos, sendo frequentemente aplicado em problemas nos quais as características podem ser representadas por valores reais, como em problemas de reconhecimento de padrões (BAFJAISH, 2020).

Já o Naïve Bayes Multinomial é aplicado a dados discretos, frequentemente associados a contagens ou frequências, como em representações de documentos por palavras. Esse modelo é especialmente eficaz em tarefas como classificação de textos, análise de sentimentos e filtragem de spam, onde as características são representadas por contagens de palavras (BAFJAISH, 2020).

Em ambas as vertentes, o processo de classificação envolve calcular a probabilidade posterior de cada classe para uma instância de teste, levando em conta as características presentes. A classe com a probabilidade mais alta é então atribuída à instância. Apesar da simplificação "Naïve", o Naïve Bayes continua sendo uma ferramenta poderosa e eficaz em uma variedade de aplicações de aprendizado de máquina.

**Figura 7** - Classificação utilizando Naive Bayes



**Fonte:** Adaptado de “Prediction of myocardial infarction complications” de Conca, Alessandro (2022).

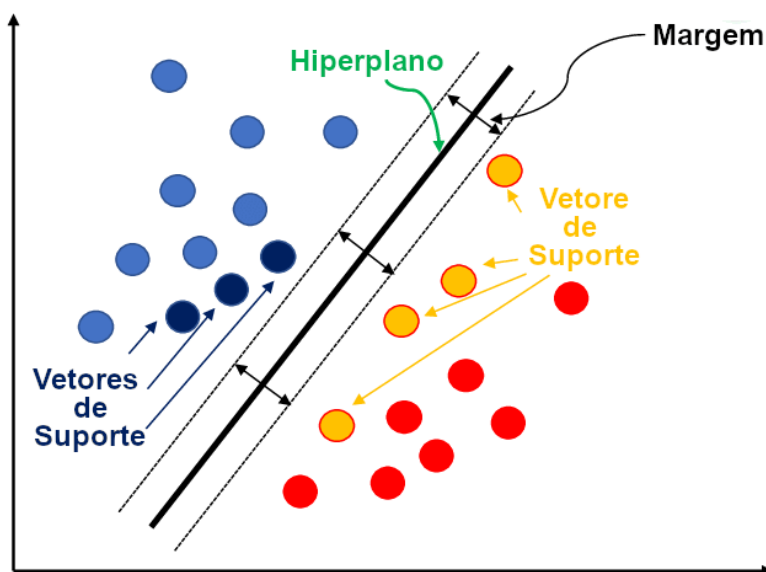
Em relação ao desempenho do algoritmo Naive Bayes, destacou-se como o segundo melhor classificador em termos de Accuracy, Precision, Recall e F1-Score, conforme evidenciado por Hassan (2022), em comparação com o Máquina de vetores de suporte (SVM), K-ésimo Vizinho mais Próximo (KNN), Regressão logística (LR) e Floresta aleatória (RF), ficando apenas atrás do KNN.

Além disso, vale ressaltar sua performance excepcional na classificação de materiais relacionados ao ODS 14, superando SVM e LR e liderando em todas as três métricas empregadas: Precision, Recall e F1-Score, conforme destacado por Hajikhani (2022).

### 3.2.3 Máquina de vetores de suporte (Support vector machine)

A abordagem principal do SVM é encontrar um hiperplano de decisão que maximize a margem entre as distintas classes no espaço de características (KURNIA, 2020). Em outras palavras, o SVM busca encontrar a superfície de separação ótima que maximiza a distância entre os exemplos de treinamento das classes, considerando os chamados vetores de suporte. Como ilustra na Figura 8, estes são os pontos mais próximos da fronteira de decisão, e a otimização da margem é realizada através da minimização de uma função de custo, que penaliza a classificação incorreta e favorece a generalização do modelo (HASSAN, 2022). O SVM é particularmente eficaz em espaços de alta dimensão e é capaz de lidar com conjuntos de dados complexos, além de oferecer versatilidade ao permitir a utilização de diferentes funções de kernel para lidar com dados não linearmente separáveis.

**Figura 8** - Classificação utilizando Máquina de vetores de suporte



**Fonte:** Adaptado de MEDIUM. Disponível em:

<https://medium.com/@bsinghrathore32/support-vector-machines-svm-day-31-6714f6534a35>.

Acesso em: 18 mar. 2024.

Na comparação entre classificadores nos materiais estudados, o algoritmo SVM se destaca como o mais abrangente. Mostra-se sólido e consistente, liderando em grande parte das comparações, tanto em termos de métricas de qualidade quanto para os Objetivos de Desenvolvimento Sustentável (ODS). Conforme destacado por Hajikhani (2022), em comparação com Naive Bayes e Regressão Logística, o SVM alcança os melhores resultados de classificação para os ODS 2, 5, 9 e 11 em termos de Precision, Recall e F1-Score, liderando em todas as três métricas de qualidade. Além disso, para os ODS 1, 4, 6, 8, 10 e 13, é o melhor em pelo menos duas métricas.

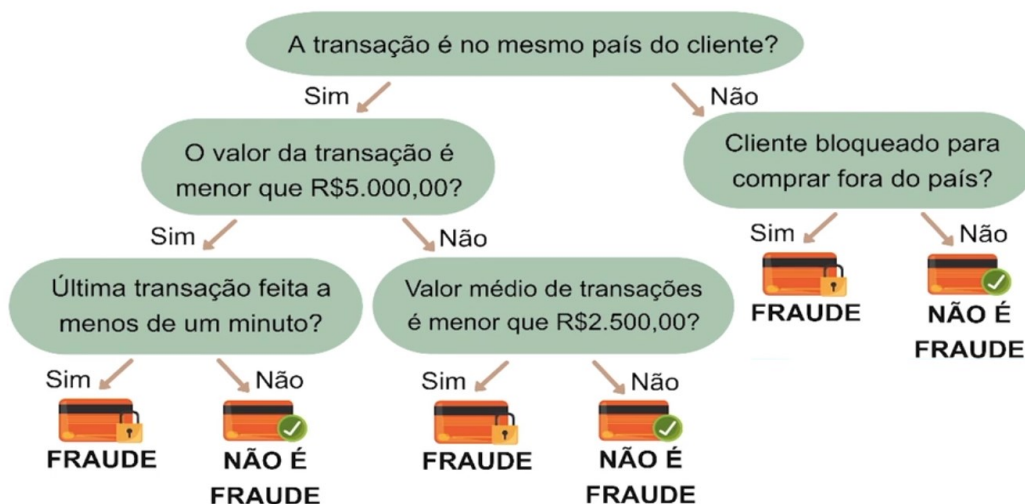
Também conforme mencionado em Angin (2022), o algoritmo SVM se mostrou como o melhor classificador para todos os 15 primeiros ODS analisados, exceto no caso do ODS 9, onde a Regressão Logística teve um desempenho superior. A métrica utilizada neste estudo foi o F1-Score.

### **3.2.4 Árvore de decisão (Decision tree)**

O algoritmo de classificação Árvore de Decisão é uma abordagem intuitiva e interpretável que visa tomar decisões sequenciais com base nas características dos dados. Durante o treinamento, o modelo constrói uma estrutura de árvore hierárquica, na qual cada nó representa uma decisão baseada em uma determinada característica.

Os dados são particionados recursivamente de acordo com essas decisões até que seja alcançada uma folha, que corresponde à classe final atribuída à instância (CHARBUTY, 2021). O processo de construção da árvore envolve escolher, em cada nó, a característica que melhor separa os dados, muitas vezes utilizando critérios como ganho de informação ou índice Gini. A árvore resultante é uma representação visual e interpretável das decisões tomadas pelo modelo. Árvores de Decisão podem lidar com dados de diferentes tipos e são capazes de capturar relações não lineares. No entanto, podem ser propensas a *overfitting*, especialmente em árvores profundas. Estratégias como poda da árvore e limitação da profundidade são comumente empregadas para mitigar esse problema. Devido à sua simplicidade e versatilidade, as Árvores de Decisão são amplamente utilizadas em diversas aplicações, desde diagnósticos médicos até sistemas de recomendação.

**Figura 9** - Classificação utilizando Árvore de Decisão



**Fonte:** Adaptado de ALURA. Disponível em: <<https://www.alura.com.br/conteudo/arvores-decisao-aprofundando-modelos-machine-learning>>. Acesso em: 18 mar. 2024.

Em relação ao desempenho, a Árvore de Decisão apresentou resultados satisfatórios. No entanto, conforme observado por Angin (2022), quando comparada a outros algoritmos, ficou evidente que existem algoritmos com desempenho superior.

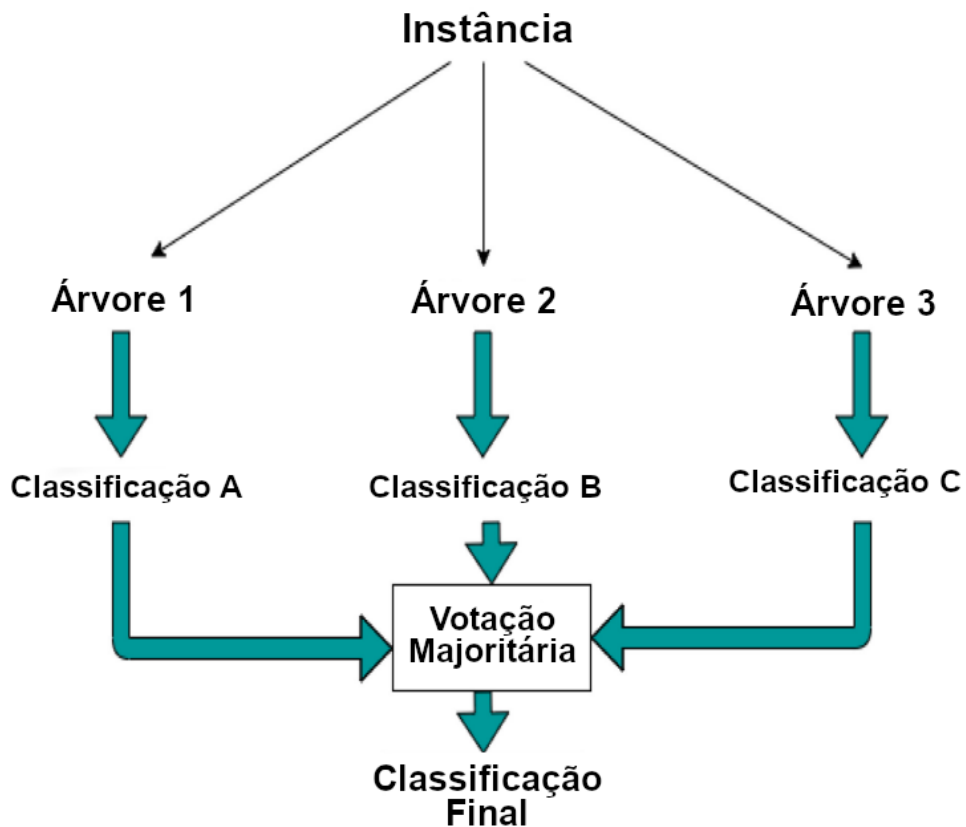
Em outra análise realizada por Singh (2024), embora o desempenho da Árvore de Decisão não tenha sido comparado com outros classificadores, uma revisão da literatura revelou sua associação frequente com os ODS 4 e 16.



### 3.2.5 Floresta aleatória (Random forest)

A Floresta Aleatória é um algoritmo de classificação que se destaca por sua eficácia ao combinar múltiplas árvores de decisão. Cada árvore é treinada em uma amostra aleatória do conjunto de dados, e a previsão final é determinada pela votação majoritária das árvores individuais. Esse método de *ensemble learning* reduz a variância e o *overfitting*, melhorando a capacidade de generalização do modelo (CHARBUTY, 2021). Durante o treinamento, a Floresta Aleatória introduz aleatoriedade na escolha de subconjuntos de dados e nas características consideradas em cada divisão da árvore. Essa abordagem diversificada e adaptável torna o modelo robusto em relação a diferentes padrões no conjunto de treinamento (ZHENG, 2019). A Floresta Aleatória é particularmente eficaz em lidar com dados de alta dimensionalidade, essa versatilidade e a capacidade de lidar com conjuntos de dados heterogêneos fazem da Floresta Aleatória uma ferramenta valiosa em aprendizado de máquina.

Figura 10 - Classificação utilizando Floresta Aleatória



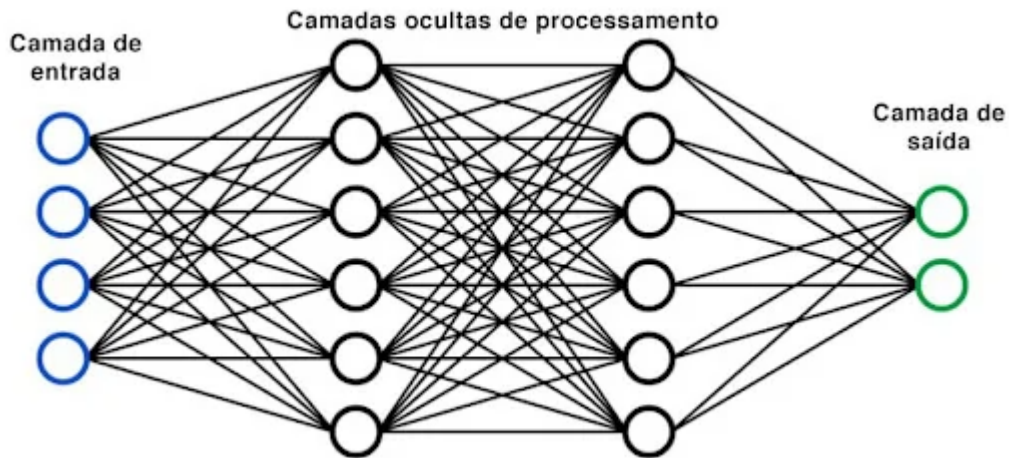
Fonte: Adaptado de “Analytics of machine learning-based algorithms for text classification” de Hassan, Sayar (2022).

O algoritmo exibe uma performance satisfatória ao analisar os materiais que avaliam sua capacidade de classificação. No entanto, ao ser confrontado com outros algoritmos, torna-se evidente que existem classificadores mais eficazes para os cenários em questão (HASSAN, 2022). De acordo com uma revisão de literatura conduzida por Singh (2024), a Floresta Aleatória é frequentemente associada ao ODS 16 - Paz, Justiça e Instituições Eficazes.

### **3.2.6 Redes neurais (Neural networks)**

As Redes Neurais são algoritmos de classificação inspirados no funcionamento do cérebro humano, com a capacidade de aprender padrões complexos a partir dos dados. Elas consistem em camadas de neurônios interconectados, divididas em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada conexão entre os neurônios possui um peso que é ajustado durante o treinamento. Durante a fase de *forward propagation*, os dados de entrada são propagados através da rede, passando por transformações lineares e não lineares nas camadas ocultas, até a camada de saída, que fornece as previsões do modelo (LAI, 2015). O treinamento da rede neural ocorre através do algoritmo de *backpropagation*, que utiliza o gradiente descendente para ajustar os pesos das conexões, minimizando a diferença entre as previsões da rede e os rótulos reais dos dados de treinamento. Redes Neurais podem capturar relações não lineares e aprender representações abstratas dos dados, tornando-as poderosas em tarefas complexas (LAI, 2015). No entanto, seu treinamento pode ser computacionalmente intensivo e requer um conjunto de dados significativo para evitar o *overfitting*. Essa flexibilidade e capacidade de lidar com dados complexos fazem das Redes Neurais uma escolha valiosa em diversas aplicações, como reconhecimento de imagem, processamento de linguagem natural e reconhecimento de padrões.

**Figura 11** - Classificação utilizando Redes Neurais



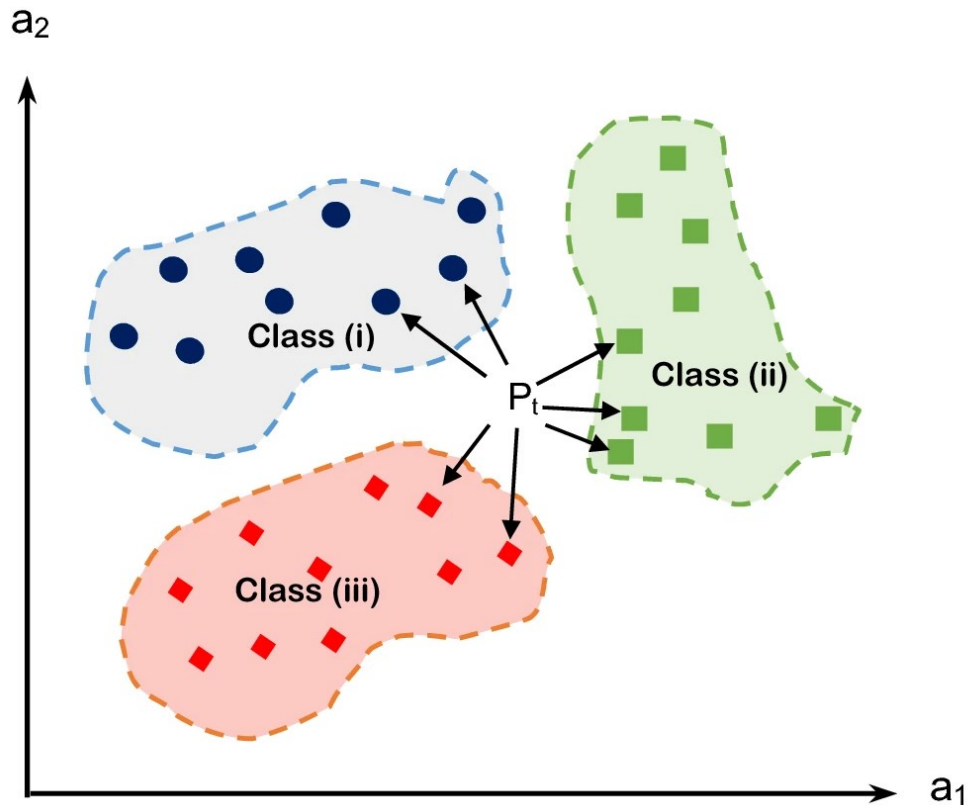
**Fonte:** Retirado de OPENCADD. Disponível em: <<https://www.opencadd.com.br/blog/o-que-sao-redes-neurais>>. Acesso em: 18 mar. 2024.

Nenhum estudo da base de dados apresentou resultados de classificação para algoritmos de Redes Neurais. No entanto, na revisão da literatura, esses algoritmos emergiram como componentes significativos, especialmente relacionados aos ODS 3, 7, 11 e 13 (SINGH, 2024).

### 3.2.7 K-ésimo vizinho mais próximo (K-nearest neighbors algorithm)

O algoritmo de classificação K-NN é uma abordagem intuitiva e simples no campo de aprendizado de máquina. Funciona com base na ideia de que instâncias semelhantes tendem a se agrupar no espaço de características (ZHENG, 2019). O processo de classificação ocorre identificando a classe predominante entre os k vizinhos mais próximos de uma instância de teste (JINDAL, 2015), como mostra a Figura 12. A escolha do valor de k é crucial, pois determina a quantidade de vizinhos considerados para a decisão de classificação. A métrica de distância, frequentemente a distância euclidiana, é utilizada para avaliar a proximidade entre instâncias. O K-NN é robusto para dados ruidosos, adaptando-se bem a diferentes formas de distribuição dos dados. No entanto, sua eficiência pode ser impactada por conjuntos de dados de grandes dimensões, pois o cálculo da distância pode se tornar computacionalmente custoso (BAOLI, 2003). Apesar dessas considerações, o K-NN é uma ferramenta versátil e fácil de compreender, sendo útil em diversas aplicações de classificação.

**Figura 12** - Classificação utilizando K-ésimo vizinho mais próximo



**Fonte:** Retirado de MEDIUM. Disponível em: <<https://medium.com/data-hackers/knn-k-nearest-neighbor-o-que-%C3%A9-aebe0f833eb>>. Acesso em: 18 mar. 2024.

No estudo conduzido por Hassan (2022), o desempenho do algoritmo de classificação K-ésimo Vizinho mais Próximo foi avaliado em comparação com o Máquina de vetores de suporte, Naïve Bayes Multinomial, Floresta aleatória e Regressão logística na classificação de duas bases de dados distintas. Os resultados revelaram sua superioridade em todas as quatro métricas de qualidade utilizadas: Acurácia, Precisão, Sensibilidade e F1-Score. Esses achados destacam a solidez desse algoritmo em tal contexto de aplicação.

O Quadro 2 a seguir sintetiza as informações desta seção em um guia comparativo dos algoritmos de classificação mais prevalentes na literatura revisada, voltados para classificação de textos em relação aos 17 Objetivos de Desenvolvimento Sustentável (ODS). É importante notar que nem todos os estudos realizaram uma comparação direta dos algoritmos, e aqueles que o fizeram apresentaram um contexto específico, como foco em um ou alguns ODS, uma seleção limitada de classificadores e seus propósitos individuais. Portanto, este resumo reflete as informações analisadas neste estudo, podendo haver discrepâncias em relação a outras análises.

**Quadro 2 - Comparativo dos algoritmos**

Algoritmos de classificação	Contagem de utilização nos materiais da base	ODS mais citados	Melhor performance	Performance (%)
Regressão logística (Logistic regression)	6	3, 4, 5, 9, 14, 15	21 de 67	31,34%
Naïve Bayes	6	4, 8, 14	11 de 67	16,41%
Máquina de vetores de suporte (Support vector machine)	5	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 16	42 de 67	62,68%
Floresta aleatória (Random forest)	3	16	0 de 04	0,00%
Redes Neurais (Neural Networks)	3	3, 7, 11, 13	01 de 01	100,00%
K-ésimo Vizinho mais Próximo (K-nearest neighbors algorithm)	2	-	04 de 04	100%
Árvore de Decisão (Decision Tree)	2	4, 16	0 de 15	0,00%

Fonte: Elaboração própria.

## 4 CONCLUSÃO

O presente estudo se propôs a realizar uma revisão sistemática da literatura com o intuito de identificar e analisar os principais algoritmos de classificação utilizados na categorização de materiais relacionados aos 17 Objetivos de Desenvolvimento Sustentável, delineados pela Organização das Nações Unidas e divulgados através da Agenda 2030.

A análise descritiva da base revelou uma lacuna significativa na disponibilidade de materiais dedicados a essa temática, indicando um campo ainda pouco explorado, especialmente considerando que estamos a apenas seis anos do prazo estipulado para a conclusão e avaliação dos resultados da Agenda 2030.

Através da revisão da literatura, foi possível identificar os principais algoritmos de classificação empregados nesse contexto. Dentre eles destacam-se a Regressão Logística, Naive Bayes, Máquina de Vetores de Suporte, Floresta Aleatória, Redes Neurais, K-ésimo Vizinho mais Próximo e Árvore de Decisão. Notavelmente, o algoritmo de Máquina de Vetores de Suporte apresentou uma performance consistente, sendo amplamente abordado nos estudos revisados e obtendo excelentes resultados de classificação. Em muitos casos, destacou-se como o algoritmo mais eficaz na análise ou como o segundo melhor, além de estar associado a uma ampla gama de ODS.

Entretanto, entre os poucos trabalhos encontrados sobre o tema, nem todos compararam os algoritmos de classificação ou forneceram resultados explícitos. Alguns desses trabalhos apenas discutiram a necessidade e a importância do tema, ou conduziram uma análise sobre a distribuição da produção entre os 17 ODS, identificando quais são mais promissores e quais ainda carecem de exploração científica. Por outro lado, os estudos que realizaram comparações entre algoritmos tinham um escopo específico, concentrando-se apenas em alguns classificadores selecionados e áreas de atuação.

Para pesquisas futuras, sugere-se colocar essa revisão da literatura em prática, avaliando os principais algoritmos de classificação de texto em uma base diversificada que contenha todos os ODS. Seria importante avaliar os pressupostos de cada algoritmo e otimizá-los para alcançar os melhores resultados, contribuindo assim significativamente para a rotulação da produção científica conforme a Agenda 2030. Isso ajudaria a mapear a evolução e identificar lacunas com mais facilidade.

## REFERÊNCIAS

- AMENYEDZI, D. K. et al. Signal Preprocessing Towards IoT Acoustic Data for Farm Pest Detection. In: IEEE EUROCON 2023 - 20th International Conference on Smart Technologies, 2023, Torino, Italy. Proceedings... Torino, Italy, 2023. p. 78-83. DOI: 10.1109/EUROCON56442.2023.10198956.
- ANGIN, M. et al. A RoBERTa Approach for Automated Processing of Sustainability Reports. Sustainability, v. 14, p. 16139, 2022. DOI: 10.3390/su142316139.
- BAFJAISH, S. S. Comparative Analysis of Naive Bayesian Techniques in Health-Related For Classification Task. Journal of Soft Computing and Data Mining, v. 1, n. 2, p. 1–10, 2020. DOI: 10.30880/jscdm.2020.01.02.001.
- BILAL, M.; ISRAR, H.; SHAHID, M.; KHAN, A. Sentiment classification of Roman-Urdu opinions using Naive Bayesian, Decision Tree and KNN classification techniques. J. King Saud Univ. Inf. Sci., v. 28, n. 3, p. 330-344, 2015. DOI: 10.1016/j.jksuci.2015.11.003.
- BUZABOON, A. et al. Automated Mapping of Environmental Higher Education Ranking Systems Indicators to SDGs Indicators using Natural Language Processing and Document Similarity. In: 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, Zallaq, Bahrain. Proceedings... Zallaq, Bahrain, 2021. p. 170-174. DOI: 10.1109/3ICT53449.2021.9581693.
- CHARBUTY, B.; ABDULAZEEZ, A. Classification based on decision tree algorithm for machine learning. J. Appl. Sci. Technol. Trends, v. 2, n. 01, p. 20–28, 2021. DOI: 10.38094/jastt20165.
- COBO MARTIN, Manuel Jesus et al. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. Journal of Informetrics, v. 5, n. 1, p. 146–166, 2011.
- DARVISH, Hamid. Bibliometric Analysis using Bibliometrix an R Package. Journal of Scientometric Research, v. 8, p. 156-160, 2020. DOI: 10.5530/jscires.8.3.32.
- DHIMAN, R.; MITEFF, S.; WANG, Y.; MA, S.-C.; AMIRIKAS, R.; FABIAN, B. Artificial Intelligence and Sustainability - A Review. Analytics, v. 3, p. 140-164, 2023. DOI: 10.20944/preprints202312.2117.v1.

GALLONI, A.; LENDÁK, I. Differentially Private Copulas, DAG and Hybrid Methods: A Comprehensive Data Utility Study. In: NGUYEN, N. T., et al. (Eds.). Computational Collective Intelligence. ICCCI 2023. Cham: Springer, 2023. Lecture Notes in Computer Science, vol. 14162. DOI: 10.1007/978-3-031-41456-5\_21.

GUPTA, S. et al. Operationalizing Digitainability: Encouraging Mindfulness to Harness the Power of Digitalization for Sustainable Development. Sustainability, v. 15, p. 6844, 2023. DOI: 10.3390/su15086844.

HAJIKHANI, A.; SUOMINEN, A. Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection. Scientometrics, v. 127, p. 6661–6693, 2022. DOI: 10.1007/s11192-022-04358-x.

HAJIKHANI, A.; SUOMINEN, A. The interrelation of sustainable development goals in publications and patents: A machine learning approach. CEUR Workshop Proceedings, v. 2871, p. 183-193, 2021.

HASSAN, Sayar; AHAMED, Jameel; AHMAD, Khaleel. Analytics of Machine Learning-based Algorithms for Text Classification. Sustainable Operations and Computers, v. 3, 2022. DOI: 10.1016/j.susoc.2022.03.001.

HASSAN, Sayar; AHAMED, Jameel; AHMAD, Khaleel. Analytics of Machine Learning-based Algorithms for Text Classification. Sustainable Operations and Computers, v. 3, 2022. DOI: 10.1016/j.susoc.2022.03.001.

HODSON, J.; SPEZZATTI, A. Hidden in Plain Sight: Building a Global Sustainable Development Data Catalogue. In: FONG, S.; DEY, N.; JOSHI, A. (Eds.). ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol. 154. Singapore: Springer, 2021. DOI: 10.1007/978-981-15-8354-4\_79.

INDRA, Rafly. Classification of User Comment Using Word2vec and SVM Classifier. International Journal of Advanced Trends in Computer Science and Engineering, v. 9, p. 643-648, 2020. DOI: 10.30534/ijatcse/2020/90912020.

JINDAL, Rajni; MALHOTRA, Ruchika; JAIN, Abha. Techniques for text classification: Literature review and current trends. Webology, v. 12, 2015.



JURAFSKY, D.; MARTIN, J. H. Speech and Language Processing. 3. ed. Pearson, 2020.

KOT, R. Review of Obstacle Detection Systems for Collision Avoidance of Autonomous Underwater Vehicles Tested in a Real Environment. *Electronics*, v. 11, n. 21, p. 3615, 2022. DOI: 10.3390/electronics11213615.

LAI, S.; XU, L.; LIU, K.; ZHAO, J. Recurrent Convolutional Neural Networks for Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 29, n. 1, 2015. DOI: 10.1609/aaai.v29i1.9513.

MEHMOOD, H.; LIAO, D.; MAHADEO, K. A Review of Artificial Intelligence Applications to Achieve Water-related Sustainable Development Goals. In: 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), 2020, Geneva, Switzerland. *Proceedings...* Geneva, Switzerland, 2020. p. 135-141. DOI: 10.1109/AI4G50087.2020.9311018.

MILLER, Lynn et al. AI for monitoring the Sustainable Development Goals and supporting and promoting action and policy development. 2020, p. 180-185. DOI: 10.1109/AI4G50087.2020.9311014.

MOHI, A.; DIN, U.; SYED, K.; RABANI, T.; RAYEES, Q. Machine learning based approaches for detecting COVID-19 using clinical text data. *Int. J. Inf. Technol.*, v. 12, n. 3, p. 731–739, 2020. DOI: 10.1007/s41870-020-00495-9.

NASIR, Osama et al. Artificial intelligence and sustainable development goals nexus via four vantage points. *Technology in Society*, v. 72, p. 102171, 2022. DOI: 10.1016/j.techsoc.2022.102171.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Transformando nosso mundo: A agenda 2030 para o desenvolvimento sustentável. 2015. Disponível em: <https://brasil.un.org/sites/default/files/2020-09/agenda2030-pt-br.pdf>. Acesso em: 19 mar. 2024.

RODENBURG, K.; DE SILVA, V.; CHRISTENSEN HUGHES, J. SDGs: A Responsible Research Assessment Tool toward Impactful Business Research. *Sustainability*, v. 13, 2021, p. 14019. DOI: 10.3390/su132414019.

ROMA, Júlio César. Os objetivos de desenvolvimento do milênio e sua transição para os objetivos de desenvolvimento sustentável. *Ciência & Cultura*, São Paulo, v. 71, n. 1, p. 33-39, jan. 2019. DOI: 10.21800/2317-66602019000100011.

SÁNCHEZ-RONCERO, A. et al. The Sustainable Development Goals and Aerospace Engineering: A critical note through Artificial Intelligence. *Results in Engineering*, v. 17, 2023, p. 100940. ISSN 2590-1230. DOI: 10.1016/j.rineng.2023.100940.

SANTOS DIVINO, Sthéfano Bruno. Reflexões sobre a inteligência artificial na agenda 2030 para o desenvolvimento sustentável. *Revista Eletrônica Direito e Política*, Programa de Pós-Graduação Stricto Sensu em Ciência Jurídica da UNIVALI, Itajaí, v. 16, n. 1, 1º quadrimestre de 2021.

SERAFIM, M. P.; LEITE, J. P. de A. O papel das Universidades no alcance dos ODS no cenário do "pós"-pandemia. *Avaliação: Revista Da Avaliação Da Educação Superior (Campinas)*, v. 26, n. 2, p. 343–346, 2021. DOI: 10.1590/S1414-40772021000200001.

SILVA, M. da V. C. e; CORREIA, I. S.; SILVA, L. B. da; RODRIGUES, W.; BAZZOLI, J. A. Revisão bibliométrica: o papel da inteligência artificial na implementação dos objetivos de desenvolvimento sustentável em tempos de COVID-19. *Revista Baru - Revista Brasileira de Assuntos Regionais e Urbanos*, Goiânia, Brasil, v. 8, n. 1, p. 21 páginas, 2022. DOI: 10.18224/baru.v8i1.9145.

SINGH, A. et al. Artificial intelligence for Sustainable Development Goals: Bibliometric patterns and concept evolution trajectories. *Sustainable Development*, v. 32, n. 1, p. 724–754, 2024. DOI: 10.1002/sd.2706.

SMITH, T. B. et al. Discovering new pathways toward integration between health and sustainable development goals with natural language processing and network science. *Global Health*, v. 19, p. 44, 2023. DOI: 10.1186/s12992-023-00943-8.

SOARES, S. V.; PICOLLI, I. R. A.; CASAGRANDE, J. L. Pesquisa bibliográfica, pesquisa bibliométrica, artigo de revisão e ensaio teórico em administração e contabilidade. *Administração: ensino e pesquisa*, v. 19, n. 2, p. 308-339, 2018. DOI: 10.13058/raep.2018.v19n2.970.

SONG, J.; JANG, C.-H. Unpacking the sustainable development goals (SDGs) interlinkages: A semantic network analysis of the SDGs targets. *Sustainable Development*, v. 31, n. 4, p. 2784–2796, 2023. DOI: 10.1002/sd.2547.

SUMMERS, Alexander et al. Constructing process maps for pulsed wave laser additive manufacturing with interpretable machine learning. *Journal of Manufacturing Processes*, v. 104, p. 138-149, 2023. DOI: 10.1016/j.jmapro.2023.09.018.

VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. *Ciência da Informação*, v. 31, n. 2, p. 369-379, 2002.

WANG, Y. et al. Massive Open Online Courses Role in Promoting United Nations Sustainable Development Goals. In: 2022 IEEE Learning with MOOCS (LWMOOCS), 2022, Antigua Guatemala, Guatemala. Proceedings... Antigua Guatemala, Guatemala, 2022. p. 126-130. DOI: 10.1109/LWMOOCS53067.2022.9927947.

YOON, S.; KIM, M.; LEE, WW. Long Short-Term Memory-Based Deep Learning Models for Screening Parkinson's Disease Using Sequential Diagnostic Codes. *J Clin Neurol*, v. 19, n. 3, p. 270-279, maio 2023. DOI: 10.3988/jcn.2022.0160.

ZHENG, Yuhan. An Exploration on Text Classification with Classical Machine Learning Algorithm. In: International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), 2019, Taiyuan, China. Proceedings... Taiyuan, China, 2019. p. 81-85. DOI: 10.1109/MLBDBI48998.2019.00023.