

Objetivos do desenvolvimento sustentável: abordagem com recurso à aprendizagem automática e dados de observação da terra

Bruno Ferreira^a, Muriel Iten^b, Rui Silva^c

^abaferreira@isq.pt

^bmciten@isq.pt

^crgasilva@gmail.com

RESUMO: O objetivo deste trabalho é avaliar, compreender e alcançar os Objetivos de Desenvolvimento Sustentável (ODS) através de dados de Observação da Terra (OT) e técnicas de Machine Learning (ML). Para o caso de estudo selecionado, os parâmetros analisados foram: diversos índices de vegetação e os valores das bandas espectrais, que foram extraídos de dados OT (Sentinel-2) e validados com diferentes abordagens de ML. Os resultados obtidos nas diferentes abordagens sugerem que a melhor técnica de classificação, bem como de regressão corresponde à Fusão de técnicas. No geral, constata-se que a OT desempenha um papel fundamental na monitorização e execução dos ODS, devido à sua relação custo-eficácia, à riqueza da informação, e o sucesso do ML na análise dos dados. A aplicabilidade das técnicas de ML aliadas aos dados de OT provou, com o caso de estudo, que estas podem contribuir para os ODS, assim como para outros fins.

PALAVRAS-CHAVE: Objetivos De Desenvolvimento Sustentável; Observação Da Terra; Dados de Observação da Terra; Machine Learning; Agricultura.

ABSTRACT: The main objective of this work is to evaluate and understand as well as to achieve the Sustainable Development Goals (SDGs) through Earth Observation (EO) data and Machine Learning (ML) techniques. For the selected case study, the parameters analyzed were: vegetation indices and the spectral bands' values, which were extracted from EO data (Sentinel-2) and validated with different ML approaches. The results obtained in the different ML approaches suggest that the best classification technique, as well as the best regression technique corresponds to the fusion of techniques. Overall, it is observed that EO plays a key role in monitoring and executing the SDGs, due to its cost-effectiveness, the wealth of information and the success of the ML in data analysis. The applicability of ML techniques combined with EO data proved, within the case study, that these can contribute to the SDGs and can be used for other purposes.

KEYWORDS: Sustainable Development Goals; Earth Observation; Earth Observation Data; Machine Learning; Agriculture.

1. Introdução

O conceito de Desenvolvimento Sustentável (DS), foi desenvolvido em 1960, quando se tornou evidente que os problemas ambientais podem ser causados pelo desenvolvimento económico e industrial. Em 1972, um primeiro relatório foi publicado e apresentado na ONU, que visava aumentar a compreensão pública dos problemas contemporâneos, abordando o impacto das 5 principais tendências de preocupação global (ou seja, industrialização acelerada, rápido crescimento populacional, desnutrição generalizada, esgotamento de recursos não-renováveis e um ambiente em deterioração), consequências do rápido desenvolvimento do mundo [1]. Este relatório, denominado Relatório Meadows [1], foi fortemente

criticado na época, por advogar o não crescimento para os países em desenvolvimento [2]. Mais tarde, em 1987, o Relatório Brundtland (BR) [3] definiu o conceito de DS como o desenvolvimento que atende às necessidades essenciais do presente sem comprometer as gerações futuras de atender às suas próprias necessidades. Em 2000, os Objetivos de Desenvolvimento do Milénio (ODM), estabeleceram 8 objetivos para combater a pobreza e a fome, alcançar a igualdade de género e melhorar o setor da saúde [4]. Até 2015, os ODM [5] impulsionaram o progresso do DS, incluindo melhorias nos serviços de saúde e educação, redução da fome e lacunas de equidade e níveis mais elevados de cobertura em intervenções com grandes investimentos [6], [7]. No entanto, permaneceu incompleto, e em 2012 novos objetivos foram estabelecidos, denominados Objetivos de

Desenvolvimento Sustentável (ODS) [5], definindo 17 objetivos únicos, representando uma chamada urgente para mudar o mundo para um caminho mais sustentável [8], [9].

A Observação da Terra (OT) desempenha um papel importante no apoio ao progresso de muitos dos ODS [10], [11]. De acordo com as Nações Unidas [12], é vantajoso usar dados de OT para produzir e apoiar estatísticas oficiais para complementar as fontes tradicionais de dados. Imagens de satélite podem ser uma das únicas tecnologias com boa relação custo-benefício capazes de fornecer dados em escala global [13], [14]. A enorme quantidade de dados, fornecidos por fontes de OT, precisa ser analisada e processada com métodos e ferramentas apropriados para fornecer indicadores robustos sobre o DS. O crescimento da área de Machine Learning (ML), que está presente no nosso quotidiano [15], [19], está constantemente a criar novas oportunidades para monitorizar e avaliar de esforços humanitários, desempenha um papel essencial na análise de imagens de satélite aplicadas aos ODS. Na verdade, a maioria dos métodos usados para processar dados de OT são baseados em ML [10], [20] dado a sua capacidade de processar enormes quantidades de dados e pelas suas características únicas de classificação e previsão.

O trabalho está estruturado da seguinte forma: capítulo 1 (o atual), introdução de conceitos relativos aos principais tópicos de investigação deste trabalho, bem como os objetivos do mesmo. O capítulo 2 descreve os métodos e materiais utilizados, bem como os desenvolvimentos práticos e a metodologia a ser aplicada ao caso de estudo. Em seguida, no capítulo 3, o caso de estudo é descrito e os resultados apresentados e discutidos. O capítulo 4 apresenta as principais conclusões deste trabalho, considerando os resultados alcançados.

1.1 Objetivos

Este trabalho pretende explorar novas metodologias e técnicas de ML e OT para otimizar (por exemplo, tempo, custos e recursos) e aprimorar os processos de monitorização para atingir os ODS. Isto é alcançado

através de 2 casos de estudo, onde é necessário mapear e/ou prever recursos naturais, adquirindo dados de OT disponíveis e trabalhando-os com técnicas de ML de forma a apoiar a realização e monitorização dos ODS (2, 8, 9 e 12) [5], bem como os objetivos da Estratégia Europeia para o Espaço (maximizar os benefícios do espaço para a sociedade e a economia da EU; promover um setor espacial europeu inovador e competitivo a nível mundial; reforçar a autonomia da Europa no acesso e utilização do espaço num ambiente seguro e protegido, reforçar o papel da Europa como ator global e promover a cooperação internacional) [21]. O objetivo principal deste trabalho é alcançado através de diferentes objetivos:

- Revisão e análise do estado da arte atual em Objetivos de Desenvolvimento Sustentável (ODS), Observação da Terra (OT) e Machine Learning (ML);
- Análise de fontes de dados de OT europeias e públicas;
- Identificação das técnicas de ML mais adequadas para a análise e classificação de dados e imagens de OT (levantamento na literatura);
- Desenvolvimento e avaliação de técnicas de ML;
- Previsão e classificação de problemas relacionados com a agricultura (ou seja, problemas fitossanitários - culturas saudáveis ou infectadas), para uma produção sustentável e suficiente.

1.2 Alinhamento com os objetivos do desenvolvimento sustentável

Os resultados deste trabalho contribuem para os seguintes ODS:

- ODS 2 - Erradicar a fome e promover a agricultura sustentável: através de resultados de ML que permitam a previsão de doenças, qualidade e rendimento das culturas. Assim, é possível alcançar maior eficiência de recursos e produção;
- ODS 8 - Promover uma economia inclusiva e sustentável, emprego e trabalho decente para todos: mantendo os empregos, pois pragas como a da Xylella Fastidiosa colocam em risco vários milhares empregos na área em toda a Europa;
- ODS 9 - Inovação industrial: através da disponibilização de informação sobre estado da

cultura, permitindo processos e procedimentos agrícolas mais eficazes, responsáveis e recomendados;

- ODS 12 - Consumo e produção sustentáveis: através de informações que permitem a redução do impacto ambiental ou da pegada ecológica, como a previsão de doenças de árvores, que podem contribuir para a redução dos produtos químicos utilizados na agricultura;

1.3 Metodologia aplicada

A base teórica desta pesquisa corresponde a uma metodologia de ação-investigação, pois esta permite ao investigador atuar sobre um problema real, tentando resolvê-lo, à medida que este é estudado/investigado. Este aspeto é crucial para este trabalho devido à novidade dos campos e problemas investigados. O procedimento de pesquisa, bem como a metodologia aplicada (descrita neste capítulo), consideram cinco etapas principais (**figura 1**), a serem, posteriormente, aplicadas e adaptadas ao caso de estudo cuja o objetivo assenta em mapear e prever culturas agrícolas para fornecer apoio à monitorização através de mapas de risco.

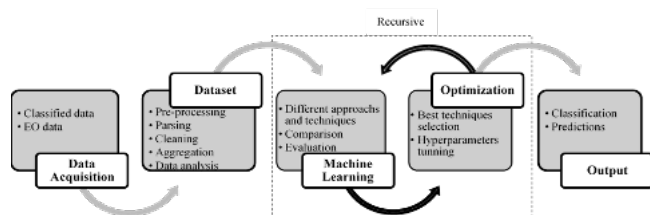


Fig. 1 – Passos da metodologia aplicada.

Tendo estabelecido a base da perspetiva teórica (estado da arte em ODS, técnicas de OT e modelos de ML), a abordagem geral apresentada na **figura 1** é implementada para ambos os casos de estudo. Em primeiro lugar, a aquisição de dados, dos dados classificados para o caso de estudo, é realizada através de contactos com especialistas e investigadores. Os restantes dados (ou seja, dados de input), necessários para construir o conjunto de dados para os algoritmos de ML, são recolhidos de fontes de dados de OT

(europeias e públicas) adequadas ao propósito de cada caso de estudo. Em seguida, para construir o conjunto de dados, algumas tarefas (ou seja, processamento, limpeza, filtragem e agregação de dados de diferentes fontes de dados) devem ser realizadas. De seguida, vários algoritmos de ML são aplicados ao conjunto de dados final, onde uma comparação e seleção das categorias e técnicas de ML mais adequadas é realizada, considerando também, um melhor entendimento sobre quais as características mais relevantes (aquelas que melhor explicam e contribuem para o resultado/objetivo de cada caso de estudo).

Cada modelo de ML selecionado deve ser aplicado 3 vezes, usando 80% dos dados para treinar e 20% para testar as técnicas. De forma garantir uma boa representação e as mesmas condições para cada técnica de ML comparada, os dados devem ser previamente “misturados/embaralhados”. A seleção das categorias de ML usadas neste trabalho (ou seja, técnicas de classificação e regressão) é baseada no tipo de saída (por exemplo, categórica, contínua ou binária) e o resultado esperado em cada caso de estudo (ou seja, classificação ou previsão).

Por último, no que diz respeito ao processo de otimização, todas as abordagens são aplicadas em primeiro lugar sem qualquer otimização (ou seja, usando os valores padrão dos parâmetros). Em seguida, é realizada a otimização das técnicas, onde os parâmetros são ajustados para obter melhores resultados. Após este processo, é produzido o resultado final, onde as técnicas devem ser aplicadas com seus melhores parâmetros.

2. Métodos e materiais

Uma revisão sistemática e análise de artigos publicados, em língua inglesa, foi conduzida usando ScienceDirect e Google Scholar. A pesquisa considera tópicos de pesquisa como: *desenvolvimento sustentável, objetivos do desenvolvimento sustentável, observação da Terra e machine learning*. Para explorar uma ampla gama de diferentes aplicações, o nome de vários campos (por exemplo, *agricultura, clima, floresta e prospecção mineral*) foi pesquisado em combinação com

observação da Terra, e em comparação com a adição de *machine learning*. A pesquisa foi refinada para sustentar resultados de relevância, considerando as pesquisas e casos de estudo mais recentes, mantendo relatos e acordos históricos.

3. Metodologia desenvolvimento para o caso de estudo

De forma a atingir o objetivo principal deste estudo e dar respostas à lacuna identificada por Ferreira [22] (elevada necessidade de novos métodos e técnicas de ML a serem aplicados em dados de OT para a execução e monitorização dos ODS), além das abordagens comuns, novas técnicas de ML são desenvolvidas: (i) fusão de técnicas e (ii) Random Forest baseada na árvore de decisão não binária. Para avaliar os resultados e desempenho destas técnicas, as (i) técnicas de classificação comum e (ii) de regressão, juntamente com a (iii) Árvore de Decisão Não Binária (previamente desenvolvida por Ferreira [23]) são implementadas.

- Técnicas de classificação: Para a avaliação dos resultados, 16 técnicas de classificação da biblioteca *Scikit-learn* (por exemplo, *Random Forest* e *K-Nearest Neighbor*), foram implementadas e aplicadas aos conjuntos de dados. Em seguida, uma comparação e otimização das técnicas é realizada, a fim de escolher a técnica mais adequada para esta aplicação.
- Técnicas de regressão: Esta abordagem é igual à anterior, porém em vez de utilizar técnicas de classificação, utiliza 16 de regressão (por exemplo, *Regression Trees* e *Support Vector Regression*). Para além disso, os dados usados em combinação com as técnicas de regressão são normalizados (ou seja, normalização Min-Max [24], onde os atributos/variáveis são redimensionados para um intervalo entre 0 e 1.), com o propósito de melhorar o seu desempenho [25].
- Árvore de Decisão não binária (NBDT): De forma a superar as limitações inerentes às árvores de decisão ID3 [26] (manipulação de variáveis contínuas) uma

adaptação do algoritmo original foi desenvolvida e implementada utilizando as mesmas métricas, mas permitindo a sua aplicação em dados com variáveis contínuas. Além disso, esta técnica também se mostra eficiente e precisa utilizando apenas uma pequena parte do número de atributos presentes no conjunto de dados original [23].

- Random Forest baseada na árvore de decisão não binária (NBRF): Este algoritmo proposto, foi desenvolvido com base no NBDT [23], apresentado acima, e na RF original [27]. Deste modo, em vez de produzir uma única árvore, produz n árvores e, para cada árvore, uma série de observações e atributos são removidos aleatoriamente, produzindo assim árvores diferentes. Após concluídas todas as árvores, a moda (valor mais frequente) é aplicada nos resultados de todas as árvores para cada observação. Se duas ou mais saídas possuírem o mesmo valor de moda, é selecionado o resultado da árvore com o melhor desempenho.
- Fusão de técnicas: Esta abordagem proposta é baseada no conceito da *Random Forest* (RF) [27], onde cada árvore produzida contribui para o resultado final. Assim, nesta abordagem, diferentes técnicas de classificação ou regressão são utilizadas com um conceito semelhante, o que permite produzir resultados mais confiáveis quando comparado com a aplicação de uma única técnica, pois o outlier de uma técnica será minimizado pelas restantes.

4. Caso De Estudo – *Xylella Fastidiosa* em Olivais

Sendo a Europa o maior produtor de azeite do mundo, onde Portugal desempenha um papel fundamental, e o facto da espécie mais afetada pela *Xylella Fastidiosa* (*Xf*) ser as oliveiras, o presente caso de estudo tem como objetivo prever ou classificar áreas com baixo e alto risco de ocorrência de *Xf* em olivais. A identificação da *Xf* é fundamental para a sua erradicação, porém, é bastante difícil, pois os sintomas visuais podem ser confundidos com inúmeras outras doenças e pragas, ou mesmo com seca [28]. Além disso, esta doença pode só manifestar

sintomas visuais após anos da data da infecção. A única forma fidedigna de identificar *Xf* é através de testes de PCR. Devido à sua taxa de disseminação, difícil identificação e ao fato de não haver cura, tornou-se importante entender seu comportamento e ter alertas precoces de áreas com maior risco de *Xf*. Deste modo, os dados de OT têm se tornado cada vez mais utilizados [28], [29], onde o cálculo de índices de vegetação permite a produção de diversos insights sobre avaliações quantitativas e qualitativas da cobertura vegetal, vigor, saúde e dinâmica de crescimento [30].

Neste caso de estudo foi necessário dois tipos de dados: dados in-situ: relativos à presença ou ausência de *Xf* e dados OT: relacionados com a condição e saúde das árvores (Sentinel-2).

4.1 Resultados

Para comparação das diferentes abordagens, a métrica *True Negative accuracy* foi a escolhida para as técnicas de classificação (considerou-se mais relevante saber com exatidão as áreas em que não há ocorrência da *Xf*, por este motivo a *accuracy* geral vem em segundo plano) enquanto que para as técnicas de regressão foi o *Root Mean Squared Error*.

Observando os resultados das cinco abordagens (apresentados na **tabela 1**), é possível inferir que, entre todas as abordagens, a melhor para classificação corresponde à Fusão de técnicas de classificação (*True Negative accuracy* de 100% e *accuracy* geral de 92,87%) e Extra Tree Classifier otimizado (*True Negative accuracy* de 99,178% e uma *accuracy* geral de 93,727%); e a melhor técnica de regressão é a Fusão de técnicas de regressão (RMSE 0,148) seguida pelo *K-Nearest Neighbour Regressor* otimizado (RMSE 0,206). As abordagens NBDT e NBRF foram descartadas após os primeiros testes, pois os seus resultados, apesar de positivos (especialmente a NBRF com 84,01% de *accuracy* geral), ficaram aquém das restantes abordagens. Contudo, é importante destacar o fato de que as técnicas propostas (ou seja, Fusão de técnicas e NBRF) alcançaram um desempenho superior

quando comparadas às técnicas anteriores (ou seja, técnicas de classificação e regressão e, NBDT).

Tab. 1 – resultados das diferentes abordagens de ml.

Abordagem	Accuracy	True Negative Accuracy	True Positive Accuracy	RMSE
Classificação* (ETC)	92,74%	99,72%	43,56%	-
Classificação Otimizada* (ETC)	93,73%	99,18%	45,78%	-
Fusão de Classificação	92,87%	100%	41,11%	-
Regressão* (ETR)	-	-	-	0,228
Regressão Otimizada* (K-NN R)	-	-	-	0,206
Fusão de Regressão	-	-	-	0,148
NBDT** (0,1)	56,71%	-	-	-
NBRF** (0,1)	84,01%	-	-	-

*resultado da melhor técnica.

** resolução dos dados.

4.2 Discussão

Os resultados apresentados acima, possibilitaram a previsão ou classificação e mapeamento de áreas com maior e menor risco de *Xf*, permitem:

- Reduzir os custos de monitorização e de teste;
- Reduzir o uso de produtos químicos;
- Manter empregos;
- Aumento ou mantimento da produção.

A qualidade dos dados tem grande influência no desempenho de todas as técnicas de ML. Alguns dos dados *in-situ* carecem de precisão na localização, levando a ter, não uma única árvore, mas um cluster de árvores, o que significa que os modelos irão prever ou classificar um cluster em vez de uma única árvore. Além disso, existe outro problema relacionado ao tempo de infecção de *Xf*, uma vez que não é possível estabelecer uma data inicial. Os sintomas presentes na árvore podem não ser visualmente detectados por muitos anos. Além destes requisitos, dados balanceados - entre árvores infectadas e não infectadas - também são necessários para treinar os algoritmos de uma forma não tendenciosa.

5. Conclusões

O Desenvolvimento Sustentável (DS) é incontornável para o crescimento de qualquer sociedade e país, e que salienta o facto dos Objetivos do Desenvolvimento Sustentável (ODS), serem cruciais para o futuro do nosso planeta. Este trabalho destaca a importância de alcançar o DS através de dados de Observação da Terra (OT) e técnicas de *Machine Learning* (ML). Os aspetos de monitorização e avaliação relacionados ao DS alavancaram os métodos de recolha de dados OT. A OT é uma tecnologia econômica, capaz de fornecer dados em nível global e assim, possibilitar uma perspectiva diferente dos ODS. Devido à elevada quantidade de dados de OT a sua análise torna-se mais poderosa e possível quando aliada a técnicas de ML. Para este trabalho, dois casos de estudo foram utilizados para demonstrar os benefícios práticos do uso de técnicas de ML em dados de OT para alcançar os ODS.

Os resultados obtidos sugerem que, no Caso de estudo (com o objetivo de mapear e prever o risco ou ocorrência de *Xylella Fastidiosa* em oliveiras), de todas as técnicas aplicadas, às que se revelaram mais adequadas

foram a Fusão das técnicas de classificação e o Extra Tree Classifier. No que diz respeito à melhor técnica de regressão, foi identificada a Fusão de técnicas de regressão. Estes resultados podem alavancar e otimizar os processos nestes ambientes e indústrias, contribuindo assim para a monitorização e realização de ODS.

Em síntese, é possível afirmar que a metodologia inovadora, apresentada neste estudo - combinação dos dados de OT e técnicas de ML (em particular as técnicas comuns de classificação e regressão, e a técnica de fusão para classificação e regressão), permitem, com sucesso, contribuir e monitorizar os ODS (2, 8, 9 e 12). Por um lado, os dados da OT desempenham um papel fundamental na tradução de problemas do mundo real em variáveis e dados que podem ser utilizados pelas técnicas ML. Além disso, as abordagens ou técnicas de ML propostas (ou seja, *Random Forest* baseada na Árvore de decisão não binária, e a fusão das técnicas) obtiveram um melhor desempenho quando comparadas com as genéricas / originais (ou seja, a Árvore de decisão não binária e, técnicas comuns de classificação e regressão. Por último, os resultados obtidos também sugerem que estas metodologias e técnicas podem ser aplicadas a outros problemas, ambientais e até, a outros ODSs.

Referências

- [1] D. H. MEADOWS, D. L. MEADOWS, J. RANDERS, AND W. W. BEHRENS III, "The Limits to Growth: A Report to The Club of Rome," *Formaliz. Math.*, 1972, doi: 10.2478/v10037-009-0024-8.
- [2] M. MESAROVIC AND E. PESTEL, "Mankind at the turning point: The second report of the club of Rome," *Econ. Anal. Policy*, 1975, doi: 10.1016/S0313-5926(76)50005-1.
- [3] WCED, "Report of the World Commission on Environment and Development: Our Common Future: Report of the World Commission on Environment and Development," 1987. doi: 10.1080/07488008808408783.
- [4] United Nations, "United Nations Millennium Declaration," *Gen. Assem.*, no. September, p. 9, 2000, [Online]. Available: http://www.un.org/en/events/pastevents/millennium_summit.shtml.
- [5] United Nations, "Transforming Our World: The 2030 Agenda for Sustainable Development," *A New Era Glob. Heal.*, no. 1, pp. 1–64, 2015, doi: 10.1891/9780826190123.ap02.
- [6] C. MOUCHERAUD ET AL. "Countdown to 2015 country case studies: what have we learned about processes and progress towards MDGs 4 and 5?," *BMC Public Health*, vol. 16, no. S2, p. 794, Sep. 2016, doi: 10.1186/s12889-016-3401-6.
- [7] United Nations, *The Millennium Development Goals Report 2015*. New York: UN, 2015.
- [8] R. G. GUSMÃO CAIADO, W. LEAL FILHO, O. L. G. QUELHAS, D. LUIZ DE MATTOS NASCIMENTO, AND L. V. ÁVILA, "A literature-based review on potentials and constraints in the implementation of the sustainable development goals," *Journal of Cleaner Production*, vol. 198. Elsevier Ltd, pp. 1276–1288, Oct. 10, 2018, doi: 10.1016/j.jclepro.2018.07.102.

- [9] United Nations Development Programme, "Background of the Sustainable Development Goals," United Nations Development Programme, 2017. <https://www.undp.org/content/undp/en/home/sustainable-development-goals/background.html> (accessed Feb. 15, 2019).
- [10] A. ANDRIES, S. MORSE, R. MURPHY, J. LYNCH, E. WOOLLIAMS, AND J. FONWEBAN, "Translation of Earth observation data into sustainable development indicators: An analytical framework," *Sustain. Dev.*, vol. 27, no. 3, pp. 366–376, May 2019, doi: 10.1002/sd.1908.
- [11] GEO, Earth Observations in support of the 2030 Agenda for Sustainable Development, no. March. Japan Aerospace Exploration Agency, 2017.
- [12] United Nations, "Earth Observations for Official Statistics: Satellite Imagery and Geospatial Data Task Team report," no. December, p. 170, 2017, Accessed: May 20, 2019. [Online]. Available: https://unstats.un.org/bigdata/taskteams/satellite/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf.
- [13] K. MURTHY, M. SHEARN, B. D. SMILEY, A. H. CHAU, J. LEVINE, AND D. ROBINSON, "SkySat-1: very high-resolution imagery from a small satellite," Oct. 2014, doi: 10.1117/12.2074163.
- [14] M. XIE, N. JEAN, M. BURKE, D. LOBELL, AND S. ERMON, "Testing the Race Model Inequality in Redundant Stimuli With Variable Onset Asynchrony," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 35, no. 2, pp. 575–579, 2016, doi: 10.1037/a0013620.
- [15] A. ALGHOUL, S. AL AJRAMI, G. AL JAROUSH, G. HARB, AND S. S. ABU-NASER, "Email Classification Using Artificial Neural Network," 2018. Accessed: Jul. 07, 2020. [Online]. Available: www.ijeais.org/ijaer.
- [16] D. VAN THIEL AND W. F. F. VAN RAAIJ, "Artificial intelligence credit risk prediction: An empirical study of analytical artificial intelligence tools for credit risk prediction in a digital era," *Journal of Risk Management in Financial Institutions*, 2019. <https://www.ingentaconnect.com/content/hsp/jrmfi/2019/00000012/00000003/art00008> (accessed Jul. 07, 2020).
- [17] D. TRONG, H. W. LAUW, Y. FANG, AND D.-T. LE, "Modeling contemporaneous basket sequences with twin Modeling contemporaneous basket sequences with twin networks for next-item recommendation networks for next-item recommendation Part of the Artificial Intelligence and Robotics Commons, Databases and Information Systems Commons, and the E-Commerce Commons Citation Citation Modeling Contemporaneous Basket Sequences with Twin Networks for Next-Item Recommendation," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul. 2018, vol. 3414, no. 3420, p. 3414, Accessed: Jul. 07, 2020. [Online]. Available: https://ink.library.smu.edu.sg/sis_research.
- [18] S. KUMER ROY, A. KANTI GHOSH, A. SIDDIKA ASA, M. PALASH UDDIN, M. RASHEDUL ISLAM, AND M. IBN AFJAL, "Bengali consonants-voice to text conversion using machine learning tool," 2017. Accessed: Jul. 07, 2020. [Online]. Available: <http://www.ijrcee.org>.
- [19] G. LUGANO, "Virtual assistants and self-driving cars," Jul. 2017, doi: 10.1109/ITST.2017.7972192.
- [20] T. LANDRY ET AL., "Applying Machine Learning to Earth Observations In A Standards Based Workflow," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2019, pp. 5567–5570, doi: 10.1109/IGARSS.2019.8898032.
- [21] European Commission, "Space Strategy for Europe," Brussels, 2016.
- [22] B. FERREIRA, M. ITEN, AND R. G. SILVA, "Monitoring sustainable development by means of earth observation data and machine learning: a review," *Environ. Sci. Eur.*, vol. 32, no. 1, p. 120, Dec. 2020, doi: 10.1186/s12302-020-00397-4.
- [23] B. FERREIRA, R. G. SILVA, AND V. PEREIRA, "Feature selection using non-binary decision trees applied to condition monitoring," *IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA*, vol. 1, pp. 1–7, 2017, doi: 10.1109/ETFA.2017.8247642.
- [24] L. AL SHALABI, Z. SHAABAN, AND B. KASASBEH, "Data Mining: A Preprocessing Engine," *J. Comput. Sci.*, vol. 2, no. 9, pp. 735–739, Sep. 2006, doi: 10.3844/jcssp.2006.735.739.
- [25] C. CHEN, S. A. VANNOY, A. ADEYEMO, AND H. WIMMER, "Effects of Normalization Techniques on Logistic Regression in Data Science," *J. Inf. Syst. Appl. Res.*, vol. 12, no. 2, p. 37, Aug. 2019, Accessed: Jan. 03, 2022. [Online]. Available: <http://conisar.org>.
- [26] J. R. QUINLAN, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1023/A:1022643204877.

- [27] L. BREIMAN, “Random Forests,” 2001, Accessed: Oct. 21, 2017. [Online]. Available: <https://www.stat.berkeley.edu/%7B~%7Dbreiman/randomforest2001.pdf>.
- [28] P. J. ZARCO-TEJADA ET AL., “Previsual symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations,” *Nat. Plants*, vol. 4, no. 7, pp. 432–439, Jul. 2018, doi: 10.1038/s41477-018-0189-7.
- [29] A. HORNERO ET AL., “Using sentinel-2 imagery to track changes produced by *xylella fastidiosa* in olive trees,” pp. 9060–9062, 2018.
- [30] J. XUE AND B. SU, “Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications,” *J. Sensors*, vol. 2017, pp. 1–17, 2017, doi: 10.1155/2017/1353691.