

# CLASSIFICAÇÃO DA EMISSÃO NUCLEAR DE GALÁXIAS COM DADOS FOTOMÉTRICOS DO SURVEY S-PLUS

Thiago Corrêa e Silva de Moraes <sup>1</sup>; Roberto Bertoldo Menezes <sup>2</sup>

<sup>1</sup> Aluno do Instituto Mauá de Tecnologia;

<sup>2</sup> Professor do Instituto Mauá de tecnologia.

**Resumo.** *Nas últimas décadas, o volume e a complexidade dos dados astronômicos têm aumentado consideravelmente. Como consequência, técnicas de machine learning têm sido cada vez mais utilizadas para a análise desses dados. Nesse projeto, foram avaliadas as técnicas de machine learning mais eficientes (e também os parâmetros mais adequados) para classificar a emissão nuclear de galáxias como sendo característica de galáxias Seyfert, Low Ionization Nuclear Emission-Line Regions (LINERs) ou regiões H II, usando dados fotométricos obtidos com Southern Photometric Local Universe Survey (S-PLUS). Ao final, constatou-se que os algoritmos supervisionados de machine learning Random Forests e Gradient Boosted Decision Trees foram os mais eficazes. Estes apresentaram assertividades para a classificação da emissão nuclear de galáxias, a partir de dados fotométricos, superiores a 80%, que é um valor muito adequado para análises desse tipo, apesar de inferior às assertividades obtidas, com as mesmas técnicas, a partir de dados espectroscópicos (por volta de 98%). A metodologia de classificação da emissão nuclear de galáxias, com base em dados fotométricos e técnicas de machine learning, é inédita e poderá ser de grande utilidade em estudos focados nesse tópico da Astrofísica Extragaláctica.*

## Introdução

Nas últimas décadas, o volume e a complexidade dos dados astronômicos têm aumentado consideravelmente. Como consequência, técnicas de *machine learning* têm sido cada vez mais utilizadas para a análise desses dados. Nesse projeto, propõe-se avaliar as técnicas de *machine learning* mais eficientes (e também os parâmetros mais adequados) para classificar a emissão nuclear de galáxias como sendo característica de galáxias Seyfert, *Low Ionization Nuclear Emission-Line Regions* (LINERs) ou regiões HII, usando dados fotométricos obtidos com o *Southern Photometric Local Universe Survey* (S-PLUS; Mendes de Oliveira et al., 2019).

Primeiramente, é importante introduzir os conceitos de fotometria e espectroscopia, pois ambos são mencionados com frequência ao longo de todo o texto.

A fotometria consiste da medida da intensidade, ou fluxo, da luz (fótons) proveniente de um objeto. As medições são feitas em relação a um filtro (banda espectral) que se refere a um intervalo em torno de um comprimento de onda específico. Por sua vez, a espectroscopia corresponde ao estudo da luz através de seus componentes em diferentes comprimentos de onda, cuja sequência é chamada de espectro, quando ela passa por uma rede de difração. O espectro oferece, dentre outras informações, a composição química do objeto, pois cada elemento químico apresenta a sua própria linha de emissão, que será detectada no espectro (Oliveira Filho e Saraiva, 2017).

De modo geral, a espectroscopia apresenta uma análise mais “precisa” da luz emitida pelo objeto do que a fotometria, por proporcionar uma análise mais detalhada dos comprimentos de ondas. Em contrapartida, os dados fotométricos são mais simples e rápidos de serem obtidos.

A emissão nuclear observada em algumas galáxias não pode ser explicada apenas pelas estrelas que as constituem. Tais objetos hospedam os chamados Núcleos Ativos de Galáxias (*Active Galactic Nuclei* – AGNs). Atualmente, acredita-se que a radiação emitida pelos AGNs se origina de um disco de matéria que rodeia um buraco negro supermassivo (*supermassive black hole* – SMBH) central. Essa matéria espirala e cai em direção ao SMBH, ocasionando a emissão intensa de radiação (Oliveira Filho e Saraiva, 2017).

Segundo suas propriedades, os AGNs recebem diferentes classificações como: *Quasi Stellar Radio Sources* (quasares); rádio-galáxias; galáxias Seyfert; *Low Ionization Nuclear Emission-Line*

*Regions* (LINERs). Para o presente relatório, as duas últimas classificações são mais relevantes e serão discutidas com mais detalhes junto com as chamadas regiões HII, que precisam ser diferenciadas dos AGNs durante a classificação (Peterson, 1997).

As galáxias Seyfert são AGNs de baixa luminosidade caracterizados pela presença de fortes linhas de emissão de alta ionização (Peterson, 1997).

Os LINERs são AGNs caracterizados por uma luminosidade muito baixa e por uma forte presença de linhas de emissão de baixa ionização, como [OI] $\lambda$ 6300 e [NII] $\lambda$ 6583 (Peterson, 1997).

As regiões HII são nebulosas de emissão, objetos caracterizados pela presença de gás ionizado pela radiação de estrelas, gerando a emissão de radiação por parte do gás. Por serem compostas principalmente por hidrogênio ionizado, tais nebulosas são frequentemente referidas como regiões HII (Chaisson e Mcmillan, 2017).

Uma técnica comum de classificação de emissão nuclear utilizando características espectroscópicas dos objetos foi proposta por Baldwin et al. (1981) e envolve as razões de fluxos de dois pares de linhas. Tal técnica gera os chamados diagramas de diagnóstico BPT (de Baldwin, Phillips e Terlevich) que podem ser utilizados para distinguir galáxias Seyfert, LINERs e regiões HII (Peterson, 1997). Apesar dos diagramas possuírem regiões com classificações não muito claras (regiões de transição), o método é consideravelmente eficiente. Esta técnica foi utilizada no projeto e está descrita na seção “Material e Métodos”.

Apesar da pesquisa aqui relatada ser voltada para dados fotométricos e não espectroscópicos, o uso do diagrama BPT serve para determinar objetos com baixa incerteza na classificação e avaliar se os algoritmos de *machine learning* a serem utilizados produzirão resultados condizentes.

Para este projeto, consideraram-se objetos com baixa incerteza na classificação aqueles que apresentaram a mesma categoria segundo todos os diagramas BPT de pares de linhas de emissão distintos utilizados. Isso evita que objetos que apresentem classificação incerta (ou seja, que estejam na região de transição) em algum dos diagramas sejam considerados na análise e, conseqüentemente, faz com que os nossos *labels* sejam mais confiáveis.

A importância do *machine learning* para o projeto se dá pela impossibilidade de análise da quantidade imensa de dados obtidos atualmente que devem ser, por exemplo, classificados quanto a sua emissão nuclear, apresentando uma alternativa para tornar os processos de análise mais rápidos e eficientes.

As técnicas desta área da computação são geralmente divididas em supervisionadas e não supervisionadas. Os algoritmos supervisionados contêm exemplos das soluções desejadas (chamados *labels*) dentre os dados de treinamento. Já os algoritmos não supervisionados não apresentam *labels* (Géron, 2022). Exemplos de algoritmos não supervisionados são os *Gaussian Mixture Models* e *K-means*. Já exemplos de algoritmos supervisionados incluem *Random Forests* e *Gradient Boosted Decision Trees*, sendo a utilização destes dois discutida durante a seção “Material e Métodos”.

Quanto aos dados a serem utilizados nos algoritmos, o survey S-PLUS será a fonte dos dados fotométricos a serem usados para o treinamento de máquina, mais especificamente o seu *Data Release 4*. A grande relevância desse projeto está associada ao fato de que a obtenção de dados fotométricos, em geral, é bem mais simples e rápida do que a obtenção de dados espectroscópicos de galáxias. Assim, a elaboração de uma metodologia para classificar a emissão nuclear de galáxias a partir de dados fotométricos (ao invés dos usuais dados espectroscópicos) será uma contribuição muito significativa para estudos na área da Astrofísica Extragaláctica.

## Material e Métodos

Como material para o projeto, foram necessárias fontes de dados tanto espectroscópicos quanto fotométricos, além de códigos de *machine learning* escritos na linguagem de programação Python.

Para os dados espectroscópicos, foi utilizado o catálogo do survey *Mapping Nearby Galaxies at APO* (MaNGA; Bundy et al., 2015), disponível em [https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=J/A%2bA/674/A85&-out.max=50&-out.form=HTML%20Table&-out.add=\\_r&-](https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=J/A%2bA/674/A85&-out.max=50&-out.form=HTML%20Table&-out.add=_r&-)

out.add= \_RAJ, DEJ&-sort= r&-oc.form=sexa. A partir dele, foram coletados objetos com declinação negativa (hemisfério sul) em conjunto com seus respectivos parâmetros espectroscópicos: as razões de linhas de emissão  $[NII]/H\alpha$ ,  $[SII]/H\alpha$ ,  $[OIII]/H\beta$  e  $[OI]/H\alpha$ .

Já para os dados fotométricos, foco principal deste projeto, foi utilizado o *Data Release 4* (iDR4 - <https://splus.cloud>) do survey S-PLUS como fonte.

Antes de retirar dados do S-PLUS, obteve-se uma estimativa do *redshift* (indicativo da distância do objeto) das galáxias a serem pesquisadas no survey de modo a ter uma maior certeza de que se está, de fato, encontrando o objeto correto. Para isso, foi usado o banco de dados *NASA/IPAC Extragalactic Database* (NED), disponível em <https://ned.ipac.caltech.edu/byname>.

Porém, para pesquisar o objeto no NED, necessita-se de seu *identifier* (nome do objeto), que foi obtido com o banco de dados *Set of Identifications, Measurements and Bibliography for Astronomical Data* (SIMBAD), disponível em <https://simbad.cds.unistra.fr/simbad/sim-fbasic>. Neste site, a galáxia de interesse é pesquisada através de suas coordenadas e uma lista de objetos ordenados a partir da menor distância em relação às coordenadas fornecidas é apresentada. O primeiro objeto dessa lista, por ter a menor distância, é o objeto buscado.

Com o *identifier* em mãos, pode-se buscá-lo no NED e obter o *redshift* de interesse, que aparece como  $z(Helio)$ .

Em posse da estimativa dos *redshifts* de todos os objetos, um código é adicionado na plataforma do S-PLUS, que permite que uma galáxia, através de suas coordenadas (declinação e ascensão de reta, ambos em graus) e um raio de busca, seja procurada. O código gera uma lista de vários objetos com seus respectivos parâmetros fotométricos, um parâmetro *zml* (*redshift*), um parâmetro *odds*, que representa a chance, de 0 a 1, do *redshift zml* apresentado estar correto e um parâmetro *CLASS*, que indica a classificação (tipo) do objeto. Se este último parâmetro for igual a “0”, trata-se de um quasar, se for igual a “1”, trata-se de uma estrela e, se for igual a “2”, trata-se de uma galáxia. Ou seja, deve-se filtrar a lista para apenas conter objetos com *CLASS* igual a “2”.

O objeto a ser selecionado é o com o valor de *zml* mais próximo do  $z(Helio)$  obtido através do NED, pois um *redshift* mais próximo do esperado aumenta a chance de estar-se de fato escolhendo o mesmo objeto já obtido no MaNGA. O parâmetro *odds* é utilizado como um fator de desempate para objetos com *zml* próximos. Este procedimento foi aplicado a cada uma das galáxias da amostra.

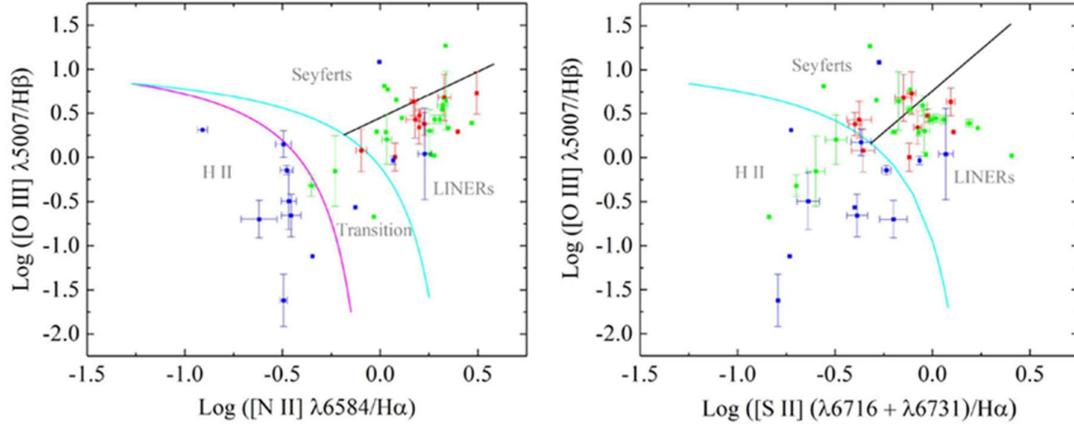
Os parâmetros fotométricos retirados do S-PLUS foram as magnitudes das galáxias nos filtros de bandas largas, g, i, r, u e z, e bandas estreitas, J378, J395, J410, J430, J515, J660 e J861.

Ao longo de todo o processo de tratamento dos dados, foi utilizado o software Excel, da Microsoft.

Como comentado na seção anterior, a classificação de emissão nuclear de dados espectroscópicos utilizada no trabalho foi realizada através do diagrama BPT. O uso do diagrama é explicado a seguir.

A Figura 1 abaixo, adaptada de Menezes et al. (2022), mostra os dois diagramas de diagnóstico aqui utilizados para a classificação da emissão nuclear dos objetos da amostra do survey MaNGA:

Figura 1 – Exemplos dos diagramas de diagnóstico utilizados nesse trabalho para a classificação da emissão nuclear da amostra de objetos selecionada do survey MaNGA. A curva em magenta corresponde à divisão empírica entre regiões HII e Núcleos Ativos de Galáxias (*Active Galactic Nuclei* – AGNs) obtida por Kauffmann et al. (2003), as curvas em ciano representam o limite máximo de ionização por um *starburst* determinado por Kewley et al. (2001), a curva preta no diagrama de  $[SII](\lambda 6716 + \lambda 6731)/H\alpha$  x  $[OIII] \lambda 5007/H\beta$  corresponde à divisão entre LINERs e Seyferts obtida por Kewley et al. (2006) e a curva preta no diagrama de  $[NII]\lambda 6583/H\alpha$  x  $[OIII]\lambda 5007/H\beta$  representa a divisão entre LINERs e Seyferts determinada por Schawinski et al. (2007). Figura adaptada de Menezes et al. (2022)



As equações correspondentes às linhas divisórias entre as diferentes categorias de objetos na Figura 1 são dadas por:

- Divisão entre regiões HII e AGNs (Kauffmann et al. 2003):  $\log\left(\frac{[OIII]\lambda5007}{H\beta}\right) = \frac{0.61}{\log\left(\frac{[NII]\lambda6583}{H\alpha}\right) - 0.05} + 1.3$

- Limite máximo de ionização por um *starburst* (Kewley et al. 2001):  $\log\left(\frac{[OIII]\lambda5007}{H\beta}\right) = \frac{0.61}{\log\left(\frac{[NII]\lambda6583}{H\alpha}\right) - 0.47} + 1.19$  e  $\log\left(\frac{[OIII]\lambda5007}{H\beta}\right) = \frac{0.72}{\log\left(\frac{[SII](\lambda6716+\lambda6731)}{H\alpha}\right) - 0.32} + 1.3$

- Divisão entre LINERs e Seyferts (Kewley et al. 2006):  $\log\left(\frac{[OIII]\lambda5007}{H\beta}\right) = 1.89 \cdot \log\left(\frac{[SII](\lambda6716+\lambda6731)}{H\alpha}\right) + 0.76$

- Divisão entre LINERs e Seyferts (Schawinski et al. 2007):  $\log\left(\frac{[OIII]\lambda5007}{H\beta}\right) = 1.05 \cdot \log\left(\frac{[NII]\lambda6583}{H\alpha}\right) + 0.45$

Como busca-se minimizar as incertezas, considerou-se um objeto com classificação de baixa incerteza aquele que obteve a mesma classificação nos diagramas BPT segundo todas as pares de linhas de emissão utilizados no projeto.

Agora, segue uma explicação dos algoritmos de *machine learning* utilizados.

As chamadas *Decision Trees* são um método no qual são construídos modelos que preveem os valores de determinadas variáveis a partir de decisões simples tomadas com base em características específicas dos dados.

As *Random Forests* consistem em uma técnica de *machine learning* que utiliza uma coleção de *Decision Trees*, cada uma das quais realiza o aprendizado com diferentes subconjuntos de variáveis e dados do *dataset* original escolhidos aleatoriamente. Ao final, os resultados de cada *Decision Tree* são combinados para gerar o resultado definitivo.

O *Gradient Boosted Decision Trees* também corresponde a uma técnica baseada em *Decision Trees*. Nesse caso, o processo é realizado em iterações e, a cada iteração, uma *Decision Tree* é construída. De uma iteração para a próxima, é feito o ajuste de coeficientes e pesos das entradas, com o propósito de minimizar a função de perda (que mede a diferença entre o valor previsto e o valor real).

## Resultados e Discussão

A primeira etapa do trabalho envolveu a coleta de dados para a obtenção de uma amostra adequada, o que se mostrou uma tarefa difícil. Tal amostra consiste em galáxias que foram observadas pelo survey S-PLUS e cujas classificações da emissão nuclear sejam bem definidas, ou seja, que não tenham incertezas significativas (classificação igual em todos os diagramas BPT utilizados).

Inicialmente, tentou-se coletar uma amostra de objetos do *Deep IFS View of Nuclei of Galaxies* (DIVING<sup>3D</sup>) survey (Steiner et al. 2022), com classificações da emissão nuclear bem estabelecidas. Porém, o número de objetos desse survey é baixo e, dentre eles, a quantidade que também foi observada pelo S-PLUS é menor ainda, apenas 28 objetos. Assim, a possibilidade de usar apenas dados do DIVING<sup>3D</sup> mostrou-se inviável.

Com isto em mente, realizou-se uma busca por novas fontes de dados e decidiu-se usar o catálogo do survey MaNGA. Deste modo, coletou-se uma amostra de objetos com declinação negativa (hemisfério sul), pois o S-PLUS não observou objetos com declinações positivas, juntamente com vários parâmetros espectroscópicos de cada objeto, como as razões de linhas  $[\text{NII}]/\text{H}\alpha$ ,  $[\text{SII}]/\text{H}\alpha$ ,  $[\text{OIII}]/\text{H}\beta$  e  $[\text{OI}]/\text{H}\alpha$ .

Na sequência, elaboraram-se scripts em Python para realizar a classificação automática da emissão nuclear dos objetos obtidos do MaNGA, com base em uma análise de diagramas de diagnóstico, conforme explicado nas seções de “Introdução” e “Material e Métodos”. A princípio, pretendia-se classificar a emissão nuclear das galáxias com base nos diagramas de diagnóstico  $[\text{NII}]\lambda 6583/\text{H}\alpha$  x  $[\text{OIII}]\lambda 5007/\text{H}\beta$ ,  $[\text{SII}](\lambda 6716+\lambda 6731)/\text{H}\alpha$  x  $[\text{OIII}]\lambda 5007/\text{H}\beta$  e  $[\text{OI}]\lambda 6300/\text{H}\alpha$  x  $[\text{OIII}]\lambda 5007/\text{H}\beta$  e então selecionar apenas os objetos que apresentaram a mesma classificação nos três diagramas (esses seriam os objetos com classificações bem estabelecidas). No entanto, como o valor da razão  $[\text{OI}]\lambda 6300/\text{H}\alpha$  não estava disponível para um grande número de galáxias, optou-se por utilizar apenas os diagramas de diagnóstico  $[\text{NII}]\lambda 6583/\text{H}\alpha$  x  $[\text{OIII}]\lambda 5007/\text{H}\beta$  e  $[\text{SII}](\lambda 6716+\lambda 6731)/\text{H}\alpha$  x  $[\text{OIII}]\lambda 5007/\text{H}\beta$ , tomando-se como objetos com classificação bem estabelecida aqueles que apresentaram a mesma classificação nesses dois diagramas, ou seja, apresentam baixa incerteza em relação à sua classificação.

Após classificar os objetos da amostra do survey MaNGA, com base nas equações anteriores, e tomar aqueles que apresentaram a mesma classificação nos dois diagramas, chegou-se a uma amostra com 399 galáxias, das quais 27 eram Seyferts, 90 eram LINERs e 282 eram regiões HII.

Nesta etapa, embora os dados do survey S-PLUS ainda não tivessem sido obtidos para os objetos com classificações da emissão nuclear bem definidas, optou-se por realizar uma avaliação preliminar das técnicas de *machine learning* com maior potencial para fornecer resultados satisfatórios na próxima etapa da análise (com dados fotométricos do survey S-PLUS). Assim, elaboraram-se scripts, em linguagem Python, para classificar a emissão nuclear das galáxias da amostra selecionada do survey MaNGA com base, simultaneamente, nas razões de linhas  $[\text{NII}]\lambda 6583/\text{H}\alpha$ ,  $[\text{SII}](\lambda 6716+\lambda 6731)/\text{H}\alpha$  e  $[\text{OIII}]\lambda 5007/\text{H}\beta$ . O objetivo principal desse teste inicial era apenas avaliar se as técnicas de *machine learning* a serem testadas seriam capazes de reproduzir, com alta precisão, a classificação da emissão nuclear obtida com base em diagramas de diagnóstico.

As duas técnicas de *machine learning* que apresentaram os melhores resultados nesse teste inicial foram *Random Forests* e *Gradient Boosted Decision Trees*. Os scripts receberam os mesmos valores de entrada: 3 arquivos de texto (“.txt” separado por tabulações), um para cada classificação (HII, Seyfert e LINER), contendo os valores de  $\log([\text{OIII}]\lambda 5007/\text{H}\beta)$ ,  $\log([\text{SII}](\lambda 6716+\lambda 6731)/\text{H}\alpha)$  e  $\log([\text{NII}]\lambda 6583/\text{H}\alpha)$  (uma coluna para cada) para cada galáxia (uma linha cada). Após o teste, obteve-se uma taxa de assertividade extremamente elevada, de cerca de 98%, para ambas as técnicas, o que confirma que são adequadas para serem utilizadas nesse trabalho.

Proseguindo para a coleta de dados do survey S-PLUS, foi necessário, primeiramente, pesquisar os *identifiers* de todos os 399 objetos no site do SIMBAD, utilizando suas coordenadas. Com os *identifiers*, o *redshift*  $z(\text{Helio})$  de cada objeto foi pesquisado no site do NED.

Após esta etapa, alguns poucos itens acabaram sendo perdidos pelo fato de não estarem presentes no banco de dados do NED.

Com os *redshifts* para o restante das galáxias, a coordenada de um objeto foi adicionada, junto com um raio de busca de, a princípio, 0,05 graus, a um código no site do survey S-PLUS, que retornava uma lista de objetos dentro do raio de busca que tem a coordenada da galáxia sendo pesquisada como centro. A lista obtida pelo código era copiada para um arquivo “.txt” e este era aberto em uma planilha do Excel, onde a coluna contendo o tipo de objeto (*CLASS*) era filtrada para mostrar apenas valores iguais a “2”, que são galáxias.

Em várias ocasiões, a primeira busca não apresentava nenhum resultado com a classificação de interesse. Quando isto ocorria, voltava-se para o código do site e aumentava-se o raio de busca. Esse procedimento era realizado até achar um objeto com *CLASS* igual a “2” e um *redshift zml* condizente.

Quando um objeto satisfatório era encontrado, os parâmetros fotométricos, *zml* e *odds* pertencentes a ele eram copiados para a linha da galáxia buscada em uma outra planilha.

Ao longo desse processo, notou-se que alguns objetos diferiam bastante do *redshift* esperado, mesmo selecionando o mais próximo possível. Para resolver este problema havia duas soluções: aumentar o raio de busca para buscar um *redshift* mais próximo do desejado ou descartar o objeto. Como o aumento no raio de busca também acaba aumentando a incerteza de que o objeto em questão é de fato o procurado (pois pode-se estar selecionando um objeto com *redshift* similar ao procurado, mas que é diferente dele), optou-se por excluir os dados com essa característica.

Seguindo o mesmo raciocínio de evitar dados com altas incertezas, a coluna do parâmetro *odds* foi filtrada para só mostrar valores maiores do que 0,9. Deste modo, garante-se uma certeza de ao menos 90% de que os valores dos *redshifts* são de fato as medidas *zml* apresentadas.

Ao final, restaram 72 regiões HII (dentre as quais se encontrava a maior quantidade de incertezas em relação aos *redshifts*), 10 galáxias Seyfert e 45 LINERs, totalizando 127 objetos, que são poucos, porém são suficientes para a análise com *machine learning*.

Finalmente, utilizando um script tanto para a técnica *Random Forests* quanto para a técnica *Gradient Boosted Decision Trees* foram realizados 3 testes distintos. O primeiro utilizava 13 parâmetros, as magnitudes nos 12 filtros e o *redshift zml*, e obteve uma assertividade de 86% no *Gradient Boosted Decision Trees* e 83% no *Random Forests*. O segundo utilizava seis parâmetros, as cinco magnitudes nas bandas largas (g, i, r, u, z) e o *redshift zml*, e obteve uma assertividade de 79% no *Gradient Boosted Decision Trees* e 81% no *Random Forests*. O terceiro utilizava oito parâmetros, as sete magnitudes nas bandas estreitas (J378, J395, J410, J430, J515, J660 e J861) e o *redshift zml*, e obteve uma assertividade de 81% no *Gradient Boosted Decision Trees* e 83% no *Random Forests*.

## Conclusões

Com base nos resultados obtidos, pode-se observar que, nos testes preliminares com os scripts de técnicas de *machine learning*, com dados espectroscópicos, a assertividade obtida foi de 98% e utilizando os mesmos scripts nos dados fotométricos, obteve-se uma assertividade de cerca de 80%. A queda de assertividade já é esperada, mas 80% de assertividade ainda é um valor muito satisfatório para análises deste tipo.

Além disso, constatou-se que houve uma redução na assertividade ao utilizar apenas as bandas largas, porém ela não foi muito significativa. Tal fato é surpreendente, em vista que as bandas estreitas incluem a extensão de apenas uma ou duas linhas de emissão, o que faz com que seja esperada uma maior proximidade para com os resultados dos diagramas de diagnóstico. Por outro lado, as bandas largas apresentam uma extensão espectral com várias linhas de emissão juntas, o que, a princípio, sugere que elas não são bons indicativos das intensidades de linhas de emissão individuais e que seus resultados de classificação tenderiam a diferir dos diagramas de diagnóstico. Mas, pelos resultados obtidos, tanto as bandas estreitas quanto as bandas largas apresentam potencial para uma classificação de emissão nuclear adequada. A metodologia de classificação da emissão nuclear de galáxias, com base em dados fotométricos e técnicas de *machine learning*, é inédita e poderá ser de grande utilidade em estudos focados nesse tópico da Astrofísica Extragaláctica.

## Referências Bibliográficas

- Baldwin, J. A., Phillips, M. M. & Terlevich, R. 1981, *PASP*, 93, 5.
- Bundy, K., Bershady, M. A., Law, D. R., Yan, R., Drory, N., et al. 2015, *ApJ*, 798, 7.
- Chaisson, E. & Mcmillan, S. *Astronomy today*. Boston: Pearson, 2017.
- Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, Ca: O'reilly Media, 2022.
- Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., et al. 2003, *MNRAS*, 346, 1055.
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A. & Trevena, J. 2001, *ApJ*, 556, 121.
- Kewley, L. J., Groves, B., Kauffmann, G. & Heckman, T. 2006, *MNRAS*, 372, 961.
- Mendes de Oliveira, C., Ribeiro, T., Schoenell, W., Kanaan, A., Overzier, R.A., et al. 2019, *MNRAS*, 489, 241.
- Menezes, R. B., Steiner, J. E., Ricci, T. V. & da Silva, P. 2022, *MNRAS*, 513, 5935.
- Oliveira Filho, K. S. & Saraiva, M. F. O.. *Astronomia e Astrofísica*. Porto Alegre: Editora Livraria da Física, 2017.
- Peterson, B. M. *An introduction to active galactic nuclei*. Cambridge: Cambridge University Press, 1997.
- Schawinski, K., Thomas, D., Sarzi, M., Maraston, C., Kaviraj, S., et al. 2007, *MNRAS*, 382, 1415.
- Steiner, J. E., et al. 2022, *MNRAS*, 510, 5780.
- Veilleux, S. & Osterbrock, D. E. 1987, *ApJS*, 63, 295.