



Universidade Federal do Rio de Janeiro  
Centro de Ciências Matemáticas e da Natureza  
Observatório do Valongo  
Graduação em Astronomia



# Prevendo a Existência de Exoplanetas a Partir das Abundâncias Químicas das Estrelas

Augusto de Paula Baldo

Rio de Janeiro  
Julho de 2022

# Prevendo a Existência de Exoplanetas a Partir das Abundâncias Químicas das Estrelas

Autor: Augusto de Paula Baldo

Trabalho de conclusão de curso/Projeto final submetido ao Curso de Graduação em Astronomia, Observatório do Valongo, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Astrônomo.

Orientador: Luan Ghezzi Ferreira Pinho

Rio de Janeiro  
Julho de 2022

## CIP - Catalogação na Publicação

d178p de Paula Baldo, Augusto  
Prevendo a Existência de Exoplanetas a Partir  
das Abundâncias Químicas das Estrelas / Augusto de  
Paula Baldo. -- Rio de Janeiro, 2022.  
116 f.

Orientador: Luan Ghezzi Ferreira Pinho.  
Trabalho de conclusão de curso (graduação) -  
Universidade Federal do Rio de Janeiro, Observatório  
do Valongo, Bacharel em Astronomia, 2022.

1. Estrelas: Abundâncias Químicas. 2. Exoplanetas.  
3. Técnica: Aprendizado de Máquina. I. Ghezzi  
Ferreira Pinho, Luan, orient. II. Título.



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
CCMN - OBSERVATÓRIO DO VALONGO  
DEPARTAMENTO DE ASTRONOMIA



## PROJETO FINAL

### RELATÓRIO DA COMISSÃO JULGADORA

**ALUNA:** Augusto de Paula Baldo (DRE 117038324)

**TÍTULO DO TRABALHO:** “PREVENDO A EXISTÊNCIA DE EXOPLANETAS A PARTIR DAS ABUNDÂNCIAS QUÍMICAS DAS ESTRELAS”

**DATA DA DEFESA:** 29 de julho de 2022 às 14:00 h

### MEMBROS DA COMISSÃO JULGADORA:

*Luan Ghezzi F. Pinho*

Prof. Luan Ghezzi Perreira Pinho – Presidente/Orientador - (OV/UFRJ)

*Hélio Jaques Rocha Pinto*

Prof. Hélio Jaques Rocha Pinto – (OV/UFRJ)

*Simone Daflon Santos*

Dr.<sup>a</sup> Simone Daflon Santos – (ON/MCTI)

Prof. Wagner Luiz Ferreira Marcolino – (OV/UFRJ)

**CANDIDATO:**

*Augusto de Paula Baldo*  
Augusto de Paula Baldo

Rio de Janeiro, 29 de julho de 2022.

*PA*

Prof. Paulo Afrânio Augusto Lopes  
Coord. de Grad. do Curso de Astronomia

O presente trabalho foi realizado com apoio da  
Fundação Carlos Chagas Filho de Amparo à  
Pesquisa do Estado do Rio de Janeiro – (Brasil)  
FAPERJ – Código de Financiamento:  
E-26/200.796/2020.

*Àqueles que abriram o meu mundo.*

# *Agradecimentos*

À minha mãe, por ser minha primeira fonte de ideias e conhecimento, por me dar a segurança e o conforto de um lar, metafórico e físico, e por me ajudar a traçar um futuro tão incerto, mas ao mesmo tempo fascinante.

Ao meu pai, pelo infindável altruísmo, pelo companheirismo e compreensão de minhas características individuais, e pelo otimismo que me fez chegar até aqui, mesmo quando o copo estava nitidamente vazio.

A ambos os meus pais, por me deram a liberdade de seguir um sonho que, desde sempre, quis seguir.

À minha irmã, por ser sempre minha figura de referência, por me ajudar a crescer e entender que a vida precisa ser vivida, mesmo que passemos por obstáculos a cada curva. Sua coragem e resiliência me ajudam a seguir em frente, todos os dias.

À minha família como um todo, por parte de pai e mãe, que me acolheu e bajulou por tantos anos dificultando, positivamente, o processo do abrir das asas.

À minha prima, Gisele, por ter sido irmã e amiga quando precisei, e por ter me acolhido no momento de maior caos em meu espectro particular e me ensinado que eu nunca deixaria de ser quem eu era.

Ao meu amigo Vitor, por se fazer presente, por ser a fonte de suporte durante e depois de minha adolescência, e por ser o narrador onipresente de nossas aventuras que seguirão até o fim dos tempos.

Aos meus amigos da graduação: Júlia, Paulo, Gabi, Kadu, João, Érica, Mylenna, Marco e Jéssica, por me mostrarem um mundo tão grande e tão incrível que me moldou e transformou em alguém disposto a explorar cada canto dele com vocês.

Ao meu orientador Luan, por ter me confiado esse projeto, e ensinado não só a caminhar dentro da astronomia, mas a correr e voar, mesmo que para ele, seus alunos façam isso por conta própria.

Aos professores, por quebrarem a visão que tinha do ambiente universitário, e me mostrarem o companheirismo, além de ensinarem diariamente a lição do profissionalismo.

À equipe da administração do Observatório do Valongo, por sua dedicação no trabalho e disponibilidade para toda e qualquer pessoa que peça ajuda.

À FAPERJ, pela bolsa de iniciação científica durante todo o projeto.

“Por mais árduos e ásperos que sejam os caminhos até as estrelas, um dia lá  
chegaremos.”

Fernando Pessoa



# *Resumo*

## **Previendo a Existência de Exoplanetas a Partir das Abundâncias Químicas das Estrelas**

Augusto de Paula Baldo

Orientador: Luan Ghezzi Ferreira Pinho

TRABALHO DE CONCLUSÃO DE CURSO/PROJETO FINAL SUBMETIDO AO CURSO DE GRADUAÇÃO EM ASTRONOMIA, OBSERVATÓRIO DO VALONGO, DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO TÍTULO DE ASTRÔNOMO

A composição química das estrelas é fundamental para entendermos questões como a evolução e populações estelares, processos de nucleossíntese e a formação planetária. Por exemplo, sabemos que a formação de planetas gigantes ocorre preferencialmente ao redor de estrelas de sequência principal e tipos espectrais FGK com metalicidades mais altas. No entanto, ainda não está claro se esta correlação permanece válida para outros elementos químicos ou classes de planetas. Desta forma, o objetivo de nosso trabalho é investigar se determinados padrões de abundâncias influenciam a formação de planetas de variados tamanhos. Para atingir esse objetivo, adaptamos o código planetPrediction, o qual utiliza técnicas de aprendizado de máquina e abundâncias químicas para estimar a probabilidade da estrela possuir planetas em sua órbita. Utilizamos uma versão atualizada do catálogo Hypatia que possui mais de 9000 estrelas e, a partir desses dados, os resultados mostraram que o ferro é o elemento mais relevante para a previsão de planetas gigantes, assim como o oxigênio, o titânio, o alumínio e o carbono. No entanto, ele não parece ter uma importância alta para prever a existência de planetas netunianos e terrestres. Para esses, o sódio, o oxigênio, o vanádio, o alumínio e o carbono se mostraram os elementos mais importantes, e a taxa de sucesso das previsões foi de 86%. Apresentamos também, uma lista de estrelas cujas probabilidades de hospedarem planetas estão acima de 90%, e que pode ser futuramente usada para projetos de busca por exoplanetas.

**palavras chave:** *estrelas: abundâncias químicas, exoplanetas, técnicas: aprendizado de máquina*

Rio de Janeiro

Julho de 2022

# *Abstract*

## **Predicting the Existence of Exoplanets from the Chemical Abundances of Stars**

Augusto de Paula Baldo

Orientador: Luan Ghezzi Ferreira Pinho

TRABALHO DE CONCLUSÃO DE CURSO/PROJETO FINAL SUBMETIDO AO CURSO DE GRADUAÇÃO EM ASTRONOMIA, OBSERVATÓRIO DO VALONGO, DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO TÍTULO DE ASTRÔNOMO.

The chemical composition of stars is fundamental to understanding issues such as stellar evolution and populations, nucleosynthesis processes and planetary formation. As an example, we know that the formation of giant planets preferentially occurs around main-sequence and FGK stars with higher metallicities. However, it remains unclear whether this correlation holds for other chemical elements or classes of planets. Thus, the objective of our work is to investigate whether certain patterns of abundances influence the formation of planets of different sizes. To achieve this goal, we adapted the planetPrediction code, which uses machine learning techniques and chemical abundances to estimate the probability of a star having planets orbiting it. We use an updated version of the Hypatia catalog that has more than 9000 stars, and from these data, the results showed that iron is the most relevant element for predicting giant planets, as well as oxygen, titanium, aluminum and carbon. However, it does not seem to have a high importance for predicting the existence of neptunian and terrestrial planets. For these, sodium, oxygen, vanadium, aluminum and carbon proved to be the most important elements, and the success rate of the predictions was 86%. We also present a list of stars for which probabilities of hosting planets are above 90%, and which can be used in the future for exoplanet search projects.

**keywords:** *stars: chemical abundances, exoplanets, techniques: machine learning*

Rio de Janeiro

Julho de 2022

# Lista de Figuras

1.1	Correlação planeta-metalicidade (Fischer & Valenti 2005). . . . .	15
1.2	Diferença entre abundâncias químicas do Sol e os valores médios nas gêmeas solares (Meléndez et al. 2009). . . . .	17
2.1	Número de estrelas no catálogo Hypatia. . . . .	19
2.2	Distribuição de log (MaxPMass) para as estrelas do catálogo Hypatia. . .	22
2.3	Distribuição de abundâncias do ferro e do sódio para análise do parâmetro MaxPRad. . . . .	23
2.4	Distribuição de [Fe/H] após os cortes referentes à origem da amostra. . . .	26
3.1	Exemplo de árvore de decisão. . . . .	28
4.1	Distribuição de estrelas com $p > 90\%$ em Hinkel et al. (2019) e as de nosso trabalho. . . . .	37
4.2	Ordem de relevância de elementos para a previsão de planetas gigantes - Catálogos: original e atualizado. . . . .	38
4.3	Gráficos de distribuição das abundâncias químicas para a previsão de planetas gigantes - Catálogo original. . . . .	39
4.4	Diferença média das abundâncias entre catálogos. . . . .	41
4.5	Histograma das probabilidades das estrelas com $p > 90\%$ no teste com o catálogo original. . . . .	41
4.6	Gráficos de distribuição das abundâncias químicas para a previsão de planetas gigantes - Catálogo atualizado. . . . .	42
4.7	Gráficos de relevância dos elementos para a previsão de planetas gigantes - Teste variando MaxPMass. . . . .	45
4.8	Gráficos de relevância dos elementos para a previsão de planetas gigantes - Teste variando grupos de elementos. . . . .	47
4.9	Gráficos [X/H] <i>versus</i> [Fe/H] - inclusão de MaxPRad na amostra. . . . .	48
4.10	Gráficos de relevância dos elementos para a previsão de planetas gigantes - Teste com MaxPRad apenas. . . . .	49
4.11	Gráfico de relevância dos elementos para a previsão de planetas menores. . . . .	53
4.12	Gráficos de distribuição das abundâncias químicas para os 3 elementos mais relevantes para a previsão na Seção 4.2.1. . . . .	55
4.13	Gráficos de distribuição das abundâncias químicas para a previsão de planetas menores. . . . .	56
4.14	Gráficos de relevância dos elementos para a previsão de planetas menores - Teste variando MaxPRad. . . . .	59
4.15	Gráficos de relevância dos elementos para a previsão de planetas menores - Teste variando grupos de elementos. . . . .	61

---

4.16	Gráficos $[X/H]$ <i>versus</i> $[Fe/H]$ - inclusão de MaxPMass na amostra. . . . .	62
4.17	Gráfico de relevância dos elementos para o teste com planetas menores - Teste incluindo MaxPMass. . . . .	63

# Lista de Tabelas

2.1	Exemplo do Catálogo Hypatia . . . . .	24
2.2	Parâmetros para cálculo de origem das estrelas, retirados de Bensby et al. (2003) . . . . .	25
3.1	Parâmetros do XGBoost . . . . .	30
3.2	Intervalos de abundâncias das amostras de treinamento . . . . .	32
3.3	Descrição dos sets disponíveis no código . . . . .	33
4.1	Resumo dos Testes com Planetas Gigantes . . . . .	38
4.2	Teste Variando a Massa Mínima para Planetas Gigantes . . . . .	44
4.3	Teste variando os grupos de elementos (Planetas Gigantes) . . . . .	46
4.4	Resumo dos testes com planetas menores . . . . .	52
4.5	Porcentagem de estrelas, em relação ao total de cada grupo, que possuem as abundâncias de cada um dos elementos mais relevantes para a previsão de estrelas com planetas menores . . . . .	54
4.6	Teste variando os raios planetários máximos . . . . .	58
4.7	Teste variando os grupos de elementos (Planetas Menores) . . . . .	60
A.1	Estrelas e Probabilidades de abrigarem planetas gigantes, Teste 4.1.2 . . .	70
B.1	Estrelas e Probabilidades de abrigarem planetas gigantes, Teste 4.2.1 . . .	92

# Abreviações

<b>KOI</b>	<b>K</b> epler <b>O</b> bject of <b>I</b> nterest (Objeto de Interesse Kepler)
<b>NASA</b>	<b>N</b> ational <b>A</b> eronautics and <b>S</b> pace <b>A</b> dministration (Administração Nacional da Aeronáutica e Espaço)
<b>XGBoost</b>	<b>e</b> Xtreme <b>G</b> radient <b>B</b> oosting (Aumento de Gradiente Extremo)

# Constantes Físicas

Raio da Terra  $R_{\oplus} = 6378,137 \text{ km}$

Massa da Terra  $M_{\oplus} = 5,9724 \times 10^{24} \text{ kg}$

Massa de Júpiter  $M_J = 1,898 \times 10^{27} \text{ kg}$

# Sumário

<b>Resumo</b>	<b>4</b>
<b>Abstract</b>	<b>5</b>
<b>1 Introdução</b>	<b>13</b>
<b>2 Dados</b>	<b>19</b>
2.1 Catálogo Hypatia . . . . .	19
2.2 Refinamento dos dados . . . . .	23
<b>3 Metodologia</b>	<b>27</b>
3.1 Algoritmo . . . . .	27
3.2 XGBoost . . . . .	28
3.3 Execução . . . . .	30
3.4 Alteração do algoritmo . . . . .	33
<b>4 Resultados e Discussão</b>	<b>36</b>
4.1 Previsões para Planetas Gigantes . . . . .	36
4.1.1 Dados originais e grupo 1 de elementos . . . . .	36
4.1.2 Dados atualizados e grupo 1 de elementos . . . . .	39
4.1.3 Grupo 1 de elementos e diferentes limites de Massa Planetária Máxima . . . . .	43
4.1.4 Outros grupos de elementos . . . . .	45
4.1.5 Grupo 1 de elementos e inclusão do parâmetro Raio Planetário Máximo . . . . .	47
4.1.6 Grupo 1 de elementos e Raio Planetário Máximo como o único critério . . . . .	49
4.2 Testes com Planetas Menores . . . . .	50
4.2.1 Grupo 1 de elementos . . . . .	51
4.2.2 Grupo de elementos 1 e raio planetário máximo de $4,4 R_{\oplus}$ . . . . .	57
4.2.3 Grupo de elementos 1 com variação do limite de Raio Planetário Máximo . . . . .	58
4.2.4 Outros grupos de elementos . . . . .	59
4.2.5 Grupo de elementos 1 e inclusão do parâmetro Massa Planetária Máxima . . . . .	60
4.3 Estudo da Previsão do Sol . . . . .	62



<i>Sumário</i>	12
<b>5 Conclusões e perspectivas futuras</b>	<b>64</b>
<b>Referências Bibliográficas</b>	<b>67</b>
<b>A Lista de estrelas com probabilidades altas de abrigarem <i>planetas gigantes</i></b>	<b>70</b>
<b>B Lista de estrelas com probabilidades altas de abrigarem <i>planetas menores</i></b>	<b>92</b>

# Capítulo 1

## Introdução

O estudo das abundâncias químicas estelares teve grande desenvolvimento no século XX, com resultados que partiram de análises das atmosferas de estrelas individuais, como o nosso Sol, até grupos de estrelas maiores e mais distantes. Cecilia Payne foi uma precursora desses estudos, ao mostrar em seu doutorado (Payne 1925) que nosso Sol e outras estrelas eram compostos em maioria por hidrogênio e hélio, resultado esse que revolucionou a astronomia como um todo.

Com o desenvolvimento de melhores equipamentos de observação e análise dos espectros estelares, a coleta de abundâncias químicas de números cada vez maiores de estrelas passou a ser possível, e construiu-se, aos poucos, um maior entendimento sobre a evolução de nossa Galáxia. Edvardsson et al. (1993) colaboraram com essa pesquisa, ao estudarem a evolução do disco Galáctico a partir das abundâncias químicas de 189 estrelas vizinhas do Sol.

Os novos dados coletados também permitiram estudar se o Sol é uma estrela típica (Gustafsson 1998), enquanto a melhora na precisão das abundâncias permitiu investigar questões cada vez mais detalhadas. Anders & Grevesse (1989) por exemplo, forneceram um amplo estudo de comparação das abundâncias químicas dos meteoritos com as do Sol, que foi de grande impacto para o entendimento de processos que ocorreram durante a formação de nosso Sistema Solar.

Em 1995, encontrou-se o primeiro planeta ao redor de uma estrela do tipo solar (Mayor & Queloz 1995). Com esse e outros eventos similares que o seguiram,

a busca por relações entre a formação desses planetas e as abundâncias químicas de suas estrelas hospedeiras teve início. Gonzalez (1997) foi o pioneiro em encontrar resultados nessa área, quando reportou que duas estrelas que hospedavam planetas gigantes possuíam altas metalicidades ( $[Fe/H]$ ), ou seja, suas abundâncias do elemento ferro eram altas em relação à do Sol. Como o próprio autor comenta em seu artigo, essas abundâncias seguiram a tendência de outros dois sistemas estudados até então. A partir desse resultado, a comunidade astronômica produziu diversos estudos relacionados a ele, detalhando cada vez mais essa relação. Laws et al. (2003), por exemplo, estenderam a amostra de estrelas com planetas gigantes investigadas para 30, fazendo uma análise espectroscópica delas e corroborando os resultados de Gonzalez (1997). Em 2005, popularizou-se esse padrão como a correlação planeta-metalicidade, por Fischer & Valenti (2005), e seu artigo trouxe a Figura 1.1, bastante conhecida pela comunidade astronômica, e que mostra que a probabilidade de que as estrelas tenham um planeta gigante aumenta com a sua metalicidade.

Nos anos seguintes, o estudo continuou se aprofundando e explorando novos caminhos, como em Sousa et al. (2008) e Ghezzi et al. (2010a), por exemplo, que investigaram a relação planeta-metalicidade para planetas netunianos, sugerindo que a distribuição de metalicidade das estrelas que hospedam apenas estes planetas está deslocada para valores menores quando comparada com a obtida para estrelas hospedeiras de, pelo menos, um planeta joviano em órbita próxima. Isso apontaria que a metalicidade atua na definição da massa do planeta mais massivo.

Desde então, com o constante avanço tecnológico na área de detecção e caracterização de sistemas planetários, novas pesquisas coletavam cada vez mais valores de abundâncias e buscavam por uma relação entre a presença de planetas gigantes e outros elementos. Robinson et al. (2006), por exemplo, reportaram grande enriquecimento de silício e níquel nessas estrelas, enquanto Adibekyan et al. (2012b) apontaram o enriquecimento dos elementos alfa (assim chamados quando seus isótopos mais abundantes são múltiplos da massa do núcleo do hélio). Ghezzi et al. (2010b), dois anos antes, estudaram o comportamento das abundâncias do lítio, e notaram que, quando são consideradas estrelas com valores similares de

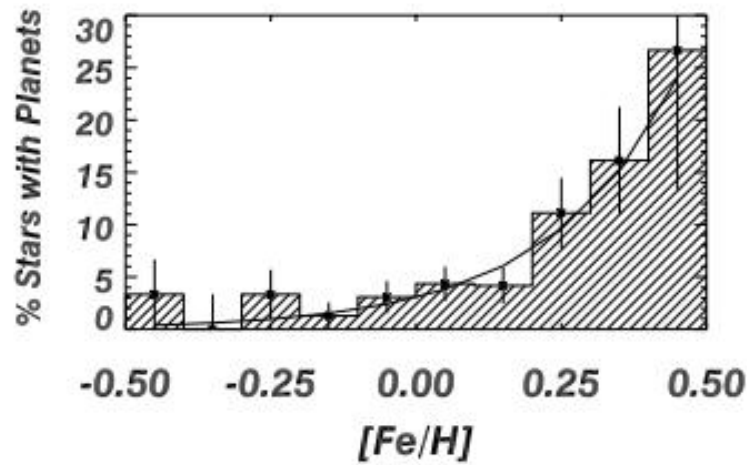


FIGURA 1.1. Clássica correlação planeta-metalicidade apresentada em Fischer & Valenti (2005). O eixo x expõe valores de metalicidade, e o eixo y mostra a porcentagem das estrelas com planetas que se encaixam nesses valores. Uma curva preta é ajustada no gráfico e aponta justamente para a relação estabelecida entre a presença de planetas gigantes ao redor das estrelas e suas metalicidades mais altas.

metalicidade e massa, o padrão de lítio se mantém o mesmo, sugerindo que a diferença nas abundâncias de lítio entre estrelas com e sem planetas conhecidos não existe. Apesar da ampla gama de resultados, a questão permanece em aberto (e.g., Figueira et al. 2014, Carlos et al. 2016). Essa inconsistência de resultados para elementos diferentes do ferro não permitiu que uma resposta fosse, até os dias de hoje, formulada.

Para planetas menores e rochosos, a relação ficou ainda mais turva. Com telescópios potentes, a observação dessas classes planetárias instigou a comunidade a explorar também a relação entre elas e as abundâncias de suas estrelas hospedeiras. O caminho mais natural para o início dessa investigação foi o estudo da relação entre o ferro e a presença de planetas menores ao redor das estrelas. Buchhave et al. (2014) e Buchhave & Latham (2015), por exemplo, sugeriram que estrelas que hospedam planetas menores não têm uma metalicidade diferente das estrelas sem planetas, enquanto Wang & Fischer (2015), por outro lado, apresentaram resultados que demonstraram o contrário. Petigura et al. (2018) exploram essa relação ainda mais profundamente, associando a alta metalicidade à ocorrência de planetas de determinados tamanhos e/ou períodos orbitais. Eles sugerem que essas altas abundâncias nos discos protoplanetários podem potencializar a produção de planetas maiores que  $1,7 R_{\oplus}$ .

O próximo passo natural para o estudo de abundâncias químicas seria encontrar uma relação entre a presença de planetas menores e as abundâncias dos demais elementos em suas estrelas hospedeiras. Meléndez et al. (2009) foram alguns dos pioneiros da área, trazendo resultados bem interessantes como os 20% de depleção dos elementos refratários em relação aos voláteis no Sol, quando comparado a gêmeas solares. Esse resultado, cuja Figura 1.2 destaca, mostraria que a presença de planetas rochosos em nosso sistema, e não nos das gêmeas solares, poderia estar ligada a essa falta dos refratários. Além deles, Brewer & Fischer (2018) forneceram um amplo catálogo de abundâncias com mais de 1000 estrelas com planetas de variadas classes, a fim de investigar tendências para a formação de planetas menores. Em sua pesquisa, eles usaram as abundâncias coletadas para contribuir com a descoberta de que o *radius gap*, uma escassez observada de planetas com raios entre 1,5 e 2,0  $R_{\oplus}$ , é um resultado da evolução dos sistemas planetários, e não necessariamente é causado por diferenças de composição entre populações diferentes de estrelas hospedeiras. Wilson et al. (2022), mais recentemente, trouxeram também um estudo dessa relação entre a presença de planetas menores e 10 elementos químicos, apontando que, para planetas com período orbital curto, um aumento na abundância de qualquer um desses elementos contribuiria para maiores ocorrências de superterras e subnetunos. Esse é um resultado muito importante para o desenvolvimento da área, pois entrega uma possível relação entre abundâncias químicas e a formação de planetas menores que, como mostramos, é há tanto buscada.

Frente às dificuldades mostradas nos parágrafos anteriores e usando o catálogo Hypatia (Hinkel et al. 2014), até então com estrelas hospedeiras de planetas majoritariamente gigantes ou sem planetas, Hinkel et al. (2019) propuseram um algoritmo que, através da técnica do aprendizado de máquina (a qual será melhor apresentada no Capítulo 4), seria capaz de prever, a partir dos valores de abundâncias químicas das estrelas, as probabilidades de uma estrela abrigar planetas gigantes. Esse trabalho, que inspirou o nosso, teve resultados muito interessantes, entregando previsões compatíveis com os estudos feitos nas décadas anteriores relacionados aos planetas gigantes, como a dependência de sua formação com a alta abundância do elemento ferro.

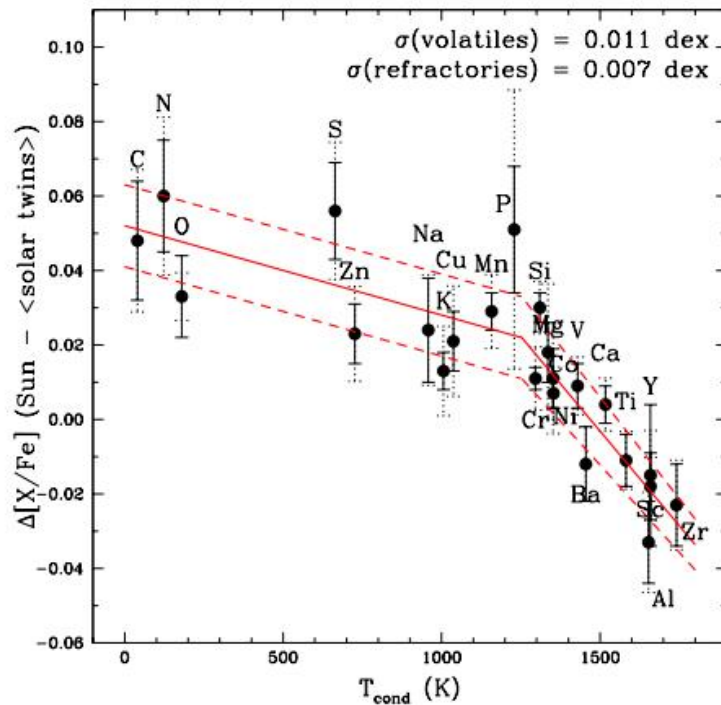


FIGURA 1.2. Diferença entre abundâncias químicas do Sol e os valores médios nas gêmeas solares em função da temperatura de condensação. Há uma baixa dispersão para cada elemento dos ajustes para os refratários e voláteis. Resultado de Meléndez et al. (2009).

Nosso projeto se inicia a partir dessa pesquisa, contribuindo para a incrementação dos estudos da relação entre as abundâncias químicas e a formação de planetas menores, bem como para a ampliação das pesquisas envolvendo a previsão de planetas gigantes e sua relação com outros elementos além do ferro. Tentaremos contribuir para a solução desses problemas ainda sem resposta, a partir de análises de previsões obtidas através da técnica de aprendizado de máquina.

Em nosso trabalho, buscamos verificar, principalmente, se a presença de planetas de diferentes classes está associada a padrões específicos de abundância. Investigaremos, primeiramente, se os padrões referentes à relação entre a formação de planetas gigantes e a alta abundância de ferro nas estrelas hospedeiras são mantidos para um grupo de estrelas maior do catálogo Hypatia. Em seguida, mantendo a análise com as estrelas desse catálogo, nos aprofundaremos no estudo do efeito das outras abundâncias químicas além do ferro na formação de planetas gigantes. Em relação à pesquisa dos planetas menores, também exploraremos os efeitos das abundâncias químicas, averiguando seu comportamento em grupos para a previsão de estrelas hospedeiras dessa classe planetária.

---

Através de atualizações e complementações dos dados apresentados no Capítulo 2 e das técnicas discutidas no Capítulo 3, realizamos diversos testes para os mais variados cenários, sejam eles de previsões envolvendo planetas gigantes ou menores, que serão destrinchados no Capítulo 4. No mesmo capítulo, analisaremos todos os resultados obtidos, e passaremos ao Capítulo 5 com as conclusões e as possibilidades futuras de nosso projeto. Nos apêndices, disponibilizamos as previsões e probabilidades para a comunidade científica com o intuito de facilitar futuras missões de busca por determinadas classes de exoplanetas.

# Capítulo 2

## Dados

### 2.1 Catálogo Hypatia

Para esse trabalho, utilizamos o Catálogo Hypatia<sup>1</sup> (Hinkel et al. 2014), que abriga as abundâncias químicas de cerca de 80 elementos para aproximadamente 10000 de estrelas, dentre outras características físicas dessas. A Figura 2.1 apresenta todos esses elementos e o número de estrelas que possuem suas respectivas abundâncias. O catálogo é atualizado regularmente com novas estrelas e seus parâmetros, o que o mantém sempre como uma ótima opção para o estudo de um amplo grupo de estrelas.

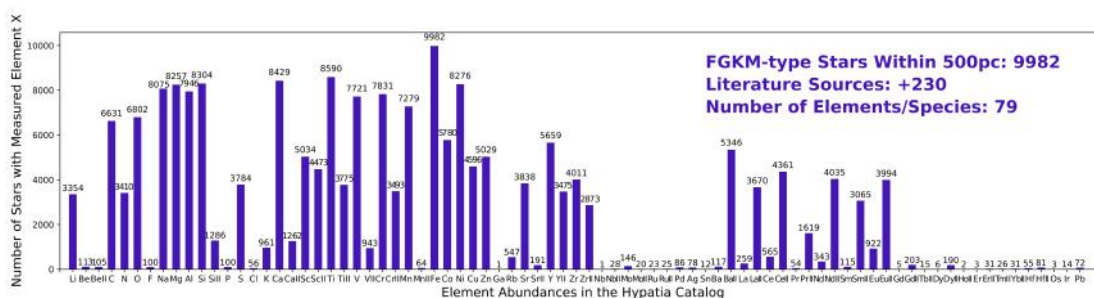


FIGURA 2.1. Número de estrelas contidas no catálogo Hypatia que possuem abundâncias para os diferentes elementos químicos - até junho de 2022. Versão em alta definição: <https://www.hyptiacatalog.com/hyptia/static/images/abundances.png>.

O catálogo é um compilado de informações sobre diversas estrelas, retiradas de diversas referências da literatura. Dentre as mais de 230 fontes utilizadas para a coleta dos dados, determinou-se muitas abundâncias usando códigos, espectros e

<sup>1</sup><https://www.hyptiacatalog.com>



escalas de abundâncias solares diferentes. Essa variedade de fontes para coleta de dados foi um dos maiores obstáculos enfrentados para a montagem do catálogo, já que usou-se diferentes valores de normalização em função da abundância solar para cada artigo. Diante disso, escolheu-se valores fixos para a normalização das abundâncias de todas as estrelas.

Para o catálogo Hypatia, esses valores de referência para o cálculo da renormalização foram os de Lodders et al. (2009). Usando o mesmo exemplo abordado em Hinkel et al. (2014), temos o artigo de Fischer & Valenti (2005), que aponta como valor de abundância do titânio para uma determinada estrela:  $[\text{Ti}/\text{H}] = \log \epsilon(\text{Ti})_{\star} - \log \epsilon(\text{Ti})_{\odot} = -0,28 \text{ dex}$ . Anders & Grevesse (1989) calcularam a abundância solar usada em seus cálculos para esse elemento:  $\log \epsilon(\text{Ti})_{\odot} = 4,99$ , portanto, sua abundância absoluta é  $\log \epsilon(\text{Ti})_{\star} = \log \epsilon(\text{Ti})_{\odot} + [\text{Ti}/\text{H}] = 4,99 - 0,28 = 4,71$ . Assim, para a renormalização no catálogo Hypatia, deve-se fazer a simples equação usando o  $\log \epsilon(\text{Ti})_{\odot}$  de Lodders et al. (2009): 4,93.

$$[\text{Ti}/\text{H}] = 4,71 - 4,93 = -0,22 \text{ dex} \quad (3.1)$$

Com isso, foi possível impor um padrão para reduzir o possível viés que surgiria das diferenças entre artigos. A renormalização não reduziu a dispersão das abundâncias gerada pela variedade de fontes usada na construção do catálogo, mas Hinkel et al. (2014) a fizeram para que a consistência fosse mantida. O catálogo também permite a escolha da normalização com a qual se quer obter os dados, e, para os valores de abundâncias com diferentes medições, há uma opção para escolher a média ou mediana deles. Caso um desses valores exceda sua própria barra de erro, a qual seus respectivos artigos mediram, e seja muito discrepante dos demais, ele é removido da amostra geral.

O catálogo Hypatia abriga, na versão que utilizamos para essa pesquisa (julho de 2020), mais de 9000 estrelas com seus respectivos parâmetros e abundâncias químicas, as quais estão, em maioria, na sequência principal e são de tipos espectrais FGKM. Para estarem inclusas nos dados, é preciso que haja ao menos a abundância química de um elemento além do ferro (Fe) para cada estrela. A

distância das estrelas também é um fator de seleção da amostra, de forma que apenas aquelas dentro de um raio de até 500 pc da Terra podem entrar no catálogo. Estrelas hospedeiras de planetas confirmados, no entanto, não são cortadas da amostra geral por esse fator e podem apresentar distâncias maiores. Atualmente, o catálogo possui estrelas pertencentes aos discos fino e espesso e ao halo da Galáxia que respeitam esses limites. A Tabela 2.1 no final dessa seção, fornece uma versão sucinta das abundâncias químicas do catálogo Hypatia, em que apenas dois dos quase 80 elementos presentes são apresentados, para exemplificação.

O catálogo lista, como comentamos acima, diversas estrelas com planetas, os quais pertencem a diferentes classes, incluindo netunianos e terrestres. As abundâncias de Brewer & Fischer (2018), coletadas de centenas de estrelas KOI (*Kepler Objects of Interest*), compõem grande parte dos valores das estrelas com planetas menores, as quais têm grande relevância para nosso estudo. Tal importância será melhor discutida na Seção 4.2 deste trabalho.

As estrelas com planetas são identificadas na tabela pelo número 1 na coluna Exo, enquanto aquelas sem planetas confirmados pelo número 0. Além disso, há outras colunas de extrema relevância para nossa análise, como a MaxPMass (ver Tabela 2.1). Essa apresenta a massa do maior planeta que orbita a estrela em questão, para os casos em que múltiplos planetas estão presentes. É a partir dela que impomos os limites de massa planetária para nossa pesquisa referente a previsão de planetas gigantes. A Figura 2.2 expõe a distribuição desse parâmetro nas estrelas com planetas da amostra. Podemos ver que a maior parte da amostra possui valores de massa acima da massa de Júpiter ( $\log(\text{MaxPMass}) = 0,0$ ), ou seja, é composta por planetas gigantes.

Seguindo um dos principais objetivos do trabalho, para fazermos previsões para planetas menores, o parâmetro da massa máxima planetária não seria suficiente para o estudo. Isso deve ao fato de que apenas 81 estrelas da amostra com planetas confirmados e com esse parâmetro medido possuíam valores de MaxP-Mass menores que  $30,0 M_{\oplus}$ . Para  $10,0 M_{\oplus}$ , esse número cai para 34 estrelas. Como será explicado na Seção 3.3, precisamos de um mínimo de 200 estrelas para a amostra de treinamento do código que iremos utilizar. Diante disso, e sabendo

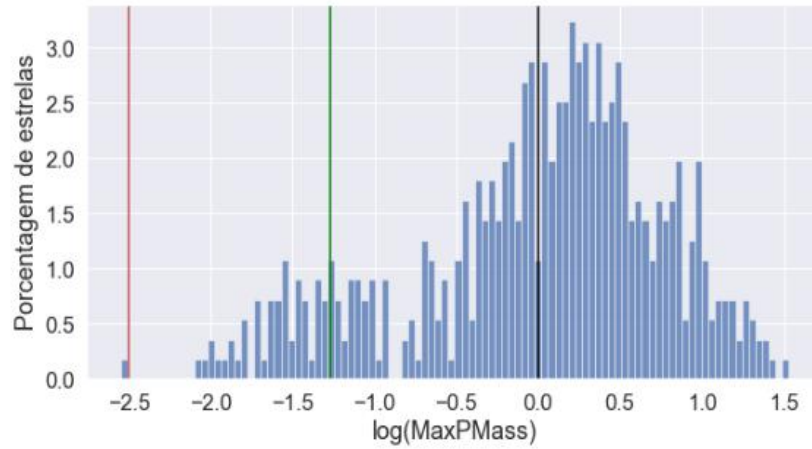


FIGURA 2.2. Distribuição de  $\log(\text{MaxPMass})$  para as estrelas do catálogo Hypatia. As linhas verticais indicam os valores de massa para a Terra em vermelho, Netuno em verde e Júpiter em preto.

que essa classe de planetas menores é detectada principalmente através do método de trânsito, o melhor parâmetro que precisávamos para inteirar nossa análise seria o raio desses planetas.

Incluímos então, nos dados, uma coluna extra com os valores de raio máximo planetário - MaxPRad. Buscamos, em fevereiro de 2022, os valores para essa coluna no NASA Exoplanet Archive<sup>2</sup>, um acervo digital com milhares de informações e parâmetros sobre os exoplanetas confirmados (e suas estrelas hospedeiras), seguindo os padrões da coluna de massas, ou seja, selecionamos os raios dos maiores planetas de cada estrela, para aqueles casos em que múltiplos planetas estão presentes.

A partir desses valores de raios, pudemos construir, usando as estrelas do catálogo Hypatia com planetas confirmados, gráficos mostrando as distribuições de abundâncias químicas para todos os elementos, e dividindo os planetas de acordo com as classes planetárias: planetas gigantes ( $R \geq 4,0 R_{\oplus}$ ) e menores ( $R < 4,0 R_{\oplus}$ ). O limite de valor 4,0 vem do trabalho de Petigura et al. (2018), que definem esse como um valor de transição de subsaturnos para subnetunos. Como exemplo, apresentamos a Figura 2.3, que expõe essa distribuição para o sódio, cuja relevância para nossa pesquisa ficará mais clara nos próximos capítulos. Esse modelo de figura atua como uma primeira análise de padrão de abundância para cada classe de planetas, e no exemplo que fornecemos, podemos notar que não há

<sup>2</sup><https://www.exoplanetarchive.ipac.caltech.edu>

uma diferença visível entre as classes e, diante disso, a análise que faremos com a ajuda do algoritmo pode contribuir para expô-la.

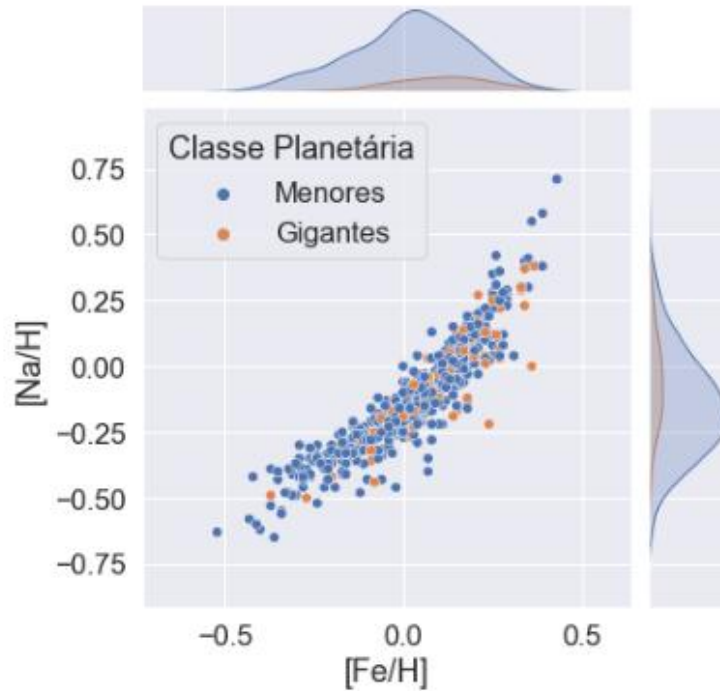


FIGURA 2.3. Abundâncias do ferro e do sódio, escolhido por apresentar maior relevância para a formação de planetas menores, como será investigado na Seção 4.2. Inclui as estrelas com planetas do catálogo Hypatia que possuem o parâmetro MaxPRad, as quais estão divididas de acordo com a classe planetária de seu maior planeta. Os painéis lateral e superior apresentam as distribuições relativas de abundância para as duas amostras.

A Tabela 2.1 reúne, de maneira sucinta, uma lista de estrelas e os parâmetros de maior importância para nossa pesquisa, os quais apresentamos nesta seção. Apesar do catálogo Hypatia fornecer os erros para alguns desses parâmetros, como não realizamos um estudo de seus efeitos nas previsões de nosso trabalho, não os incluímos nas tabelas. Os valores “NaN” em algumas colunas indicam a falta daquele parâmetro ou valor de abundância para a estrela em questão.

## 2.2 Refinamento dos dados

Fizemos algumas alterações na tabela com a amostra geral de estrelas, a fim de refinar os dados para a aplicação do algoritmo que será apresentado no Capítulo 4. Primeiramente, eliminamos algumas estrelas pertencentes ao disco espesso e ao halo da Galáxia. Realizamos essa remoção para evitar vieses causados pela diferença nas abundâncias das estrelas pertencentes a diferentes componentes da

TABELA 2.1. Exemplo do Catálogo Hypatia

Identificador	C	O	Exo	MaxPMass	MaxPRad
KOI-367	0.04	-0.01	1	0.153	4.52
HIP 80250	-0.06	0.38	1	6.560	NaN
HIP 99496	-0.32	-0.26	1	3.100	NaN
KOI-3248	-0.06	-0.12	0	NaN	NaN
KOI-3197	-0.00	-0.08	0	NaN	NaN
HIP 18538	-0.05	-0.24	0	NaN	NaN
KOI-82	0.13	0.00	1	0.028	2.48
KOI-262	-0.09	0.10	1	NaN	2.75
WASP-54	NaN	NaN	1	0.590	17.70
HIP 2049	0.16	0.05	0	NaN	NaN

Via Láctea. Para que esse corte fosse feito, calculamos a região de origem das estrelas (Bensby et al. 2003):

$$TD/D = \frac{X_{TD}}{X_D} \times \frac{f_{TD}}{f_D} \quad (3.2)$$

Nesta equação,  $X$  representa a fração observada de estrelas para as populações em nossa vizinhança solar,  $TD$  e  $D$  são siglas que se referem aos discos espesso e fino, respectivamente, e os valores de  $f$  são definidos pelas equações abaixo:

$$f(U, V, W) = k \times \exp \left( -\frac{U_{LSR}^2}{2\sigma_U^2} - \frac{(V_{LSR} - V_{asym})^2}{2\sigma_V^2} - \frac{W_{LSR}^2}{2\sigma_W^2} \right) \quad (3.3)$$

onde

$$k = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_U \sigma_V \sigma_W} \quad (3.4)$$

Nas equações anteriores,  $U_{LSR}$ ,  $V_{LSR}$  e  $W_{LSR}$ , são as componentes da velocidade galáctica, usando o Padrão Local de Repouso, ou LSR (*Local Standard of Rest*), como referencial, cuja origem descreve uma trajetória circular ao redor do centro da Via Láctea. Os parâmetros  $\sigma_U$ ,  $\sigma_V$  e  $\sigma_W$  são as dispersões de velocidade características de cada componente da velocidade galáctica e  $V_{asym}$  é a deriva assimétrica da população estelar em questão, que nada mais é que a diferença entre a velocidade média de rotação dessa população e a velocidade de um conjunto de

estrelas hipotéticas que possuem órbitas perfeitamente circulares. Os fatores  $X$ ,  $\sigma_U$ ,  $\sigma_V$ ,  $\sigma_W$  e  $V_{asym}$  têm os valores apresentados na Tabela 2.2.

TABELA 2.2. Parâmetros para cálculo de origem das estrelas, retirados de Bensby et al. (2003)

$f$	$X$	$\sigma_U$ [km/s]	$\sigma_V$ [km/s]	$\sigma_W$ [km/s]	$V_{asym}$ [km/s]
Disco Fino (D)	0.94	35	20	16	-15
Disco Espesso (TD)	0.06	67	38	35	-46

Para realizar o cálculo da origem, obtivemos as componentes da velocidade galáctica UVW a partir de 6 parâmetros - Ascensão Reta ( $\alpha$ ), Declinação ( $\delta$ ), Movimento Próprio em ascensão reta ( $\mu_\alpha$ ) e em declinação ( $\mu_\delta$ ), Velocidade Radial ( $V_{Rad}$ ) e Paralaxe ( $\pi$ ). Enquanto o catálogo Hypatia nos forneceu os primeiros dois parâmetros para todas as estrelas, usamos a base de dados astronômicos SIMBAD<sup>3</sup> (Wenger et al. 2000) para completar os demais parâmetros e, para as estrelas ainda sem valor de paralaxe, buscamos complementar com o catálogo de Anderson & Francis (2012) e o segundo Data Release do Gaia (DR2; Gaia Collaboration et al. 2018). Assim, pudemos usar o código *gal\_uvw*<sup>4</sup> que calcula e retorna as três componentes da velocidade galáctica para aplicarmos na Equação 3.3 e obter  $f$  para os discos fino (D) e espesso (TD). A partir desses valores de  $f$  e os valores de  $X$  da Tabela 2.2, conseguimos estimar o valor TD/D com a Equação 3.2.

As estrelas que seguem o padrão  $TD/D \leq 0,1$  provavelmente são pertencentes ao disco fino. Para a nossa pesquisa, removemos da amostra geral todas aquelas que não obedeceram esse limite, ou seja, obtiveram um valor de  $TD/D > 0,1$ . De um total de 9434 estrelas, removemos 2235. A Figura 2.4 apresenta a nova distribuição de  $[Fe/H]$  para as estrelas com e sem planetas da tabela após os cortes referentes à origem da amostra.

Um corte adicional, foi a remoção das abundâncias do elemento escândio (Sc) da tabela, cujos valores estavam presentes em menos de 50% das estrelas da amostra geral. Esse corte, sugerido em Hinkel et al. (2019), é de extrema importância para evitarmos os vieses que surgiriam pela tendência do algoritmo de selecionar os elementos com menos valores como os mais favoráveis à presença de determinados tipos de planetas. Removemos também 3 estrelas cujas abundâncias

<sup>3</sup><https://simbad.u-strasbg.fr/simbad/>

<sup>4</sup>[https://github.com/segasai/astrolibpy/blob/master/astrolib/gal\\_uvw.py](https://github.com/segasai/astrolibpy/blob/master/astrolib/gal_uvw.py)

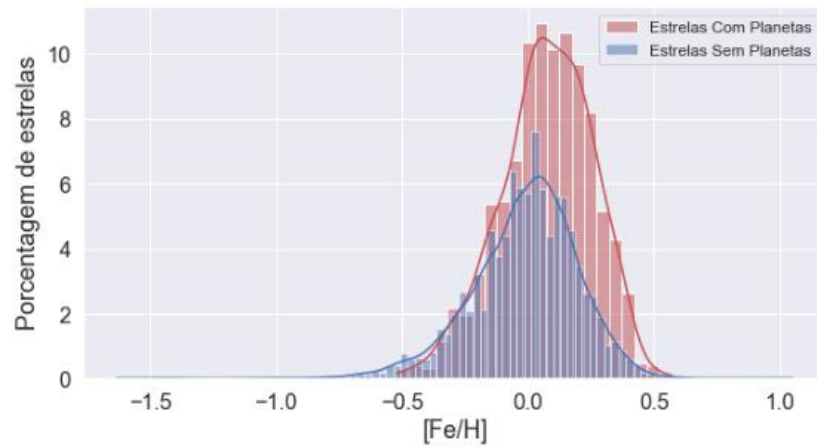


FIGURA 2.4. Distribuição de  $[Fe/H]$  da tabela após os cortes referentes à origem da amostra. As barras em vermelho indicam as abundâncias das estrelas com planetas, e as azuis as das estrelas sem planetas conhecidos.

para alguns elementos como o ferro estavam com valores muito abaixo dos valores típicos do restante das estrelas. Foram elas HIP 71908 ( $[Fe/H] = -2.06$ ), HIP 18339 ( $[Fe/H] = -5.06$ ) e HIP 113711 ( $[Fe/H] = -5.29$ ). Fizemos essa remoção como um cuidado extra com a amostra, já que, diante da amplitude dos dados, 3 estrelas não causam um grande impacto nos resultados. Após esses refinamentos da amostra, o número final de estrelas contidas em nossa tabela é de 7196.

## Capítulo 3

# Metodologia

### 3.1 Algoritmo

O aprendizado de máquina, ou *machine learning*, é a ferramenta que guia o nosso trabalho. Ela funciona a partir de várias áreas combinadas, como a estatística, a programação e a matemática de forma a nos entregar os resultados probabilísticos esperados. Em um procedimento de aprendizado de máquina, o código analisa os componentes de um grupo de treinamento, encontra um padrão em comum entre eles e, a partir dele, cria uma fórmula aplicável na amostra geral. Assim, o código consegue entregar as previsões para um novo grupo. Um exemplo bem simples dessa ferramenta, está nos aplicativos de tradução de línguas estrangeiras. As empresas responsáveis por essas traduções fornecem à máquina milhares de textos em inglês, por exemplo, e cada frase é estudada para contextos variados, desde uma conversa informal transcrita a um artigo científico. Essas são as amostras de treinamento. A partir delas, quando um cliente pede uma tradução daquela frase para o português, a máquina vai procurar os melhores textos em que utilizou-se a frase e entregará a tradução, colhida da mesma versão brasileira do livro ou de transcrições similares, de forma contextualizada e com baixas chances de erro.

O método das árvores de decisão, usado em nossa pesquisa, é outra ferramenta de aprendizado de máquina. Ele funciona dividindo os dados em subconjuntos menores baseados em seus parâmetros em comum, os chamados “galhos”



ou “nós”. Esses são então usados para a construção de outro conjunto em série, sucessivamente. Cada nó é treinado para corrigir os erros das árvores e nós anteriores, mantendo uma sucessão de verificações e, conseqüentemente, montando um modelo mais preciso de classificação. Quando um nó chega numa resposta, ou alcança o limite do processamento, ele se torna um “nó folha” e não se divide mais. A Figura 3.1 mostra uma ideia visual dessa divisória de nós.

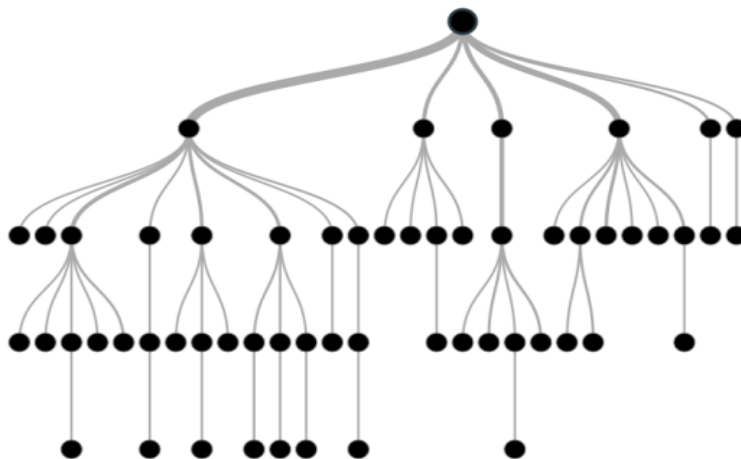


FIGURA 3.1. No método das árvores de decisão, essas, são estruturas de dados formadas por um conjunto de elementos que armazenam informações chamados nós ou galhos. Esses, são representados pelos círculos que se dividem. Além disso, toda árvore possui um nó chamado raiz, que fica no ponto de partida e tem ligações para os outros elementos. O nó que não possui repartição é conhecido como nó folha ou terminal.

O algoritmo usado em nossa pesquisa e que se baseia nesse método das árvores de decisão chama-se *planetPrediction*<sup>1</sup>, e Hinkel et al. (2019) o contruíram. Ao longo das próximas sessões, vamos explorar suas principais ferramentas que contribuíram para os resultados que encontramos e serão melhor discutidos nos capítulos seguintes.

## 3.2 XGBoost

O programa responsável pelas previsões de nossa pesquisa e do algoritmo, a partir dos parâmetros que entregamos a ele, se chama XGBoost (Extreme Gradient Boosting - Chen & Guestrin 2016). O XGBoost funciona a partir de uma série de parâmetros estatísticos responsáveis pelas previsões. São eles:

<sup>1</sup><https://www.github.com/nhinkel/planetPrediction>

- ‘learning\_rate’: é o responsável por controlar a etapa de impulsionamento da ferramenta, de maneira a evitar ajustes excessivos. Ele irá atuar nos pesos dos parâmetros, definindo quais deles têm maior relevância para a previsão do código e, assim, reduzir o esforço do algoritmo.
- ‘n\_estimators’: um dos mais importantes parâmetros para a análise de árvores de decisão. Fornece o número de árvores com que o código trabalhará. Quanto mais árvores, melhor a acurácia dos resultados, mas o processo pode se tornar lento com a escolha de um valor alto.
- ‘max\_depth’: outro parâmetro de alta importância, funciona de maneira similar ao anterior, mas age “aprofundando” as árvores de decisão. Em códigos com árvores rasas, pode haver a baixa captura de detalhes do problema, enquanto naqueles com árvores profundas há o aumento da complexidade do código, podendo também causar problemas de lentidão nas previsões.
- ‘min\_child\_weight’: atua como um peso estatístico que define se o código deve continuar ou não a partição dos “galhos” ou “folhas” das árvores de decisão.
- ‘gamma’: é conhecido como o “custo da complexidade ao introduzir folhas adicionais no código”. Também atua no conservadorismo do algoritmo instruindo as árvores a adicionarem divisões somente se o ganho associado à previsão for maior ou igual à gamma.
- ‘subsample’: define a proporção de dados de treinamento que serão pré-amostrados pelo XGBoost antes do cultivo efetivo das árvores. Isso contribui para evitar o sobreajuste do código.
- ‘colsample\_bytree’: funciona similarmente ao parâmetro anterior, mas define a proporção de parâmetros que serão pré-selecionados pelo XGBoost, no lugar das amostras de treinamento. Por isso, também colabora para evitar o sobreajuste.
- ‘objective’: trabalha com o modelo matemático e estatístico que o algoritmo utiliza. Em nossa pesquisa, esse critério é o ‘binary:logistic’, uma regressão logística para classificação binária, ideal para a medição de probabilidades.

Outros modelos como a regressão de Poisson, ou a função de densidade de probabilidade também estão disponíveis para escolha.

- ‘nthread’: é o número de “fios” ou “tarefas paralelas” usadas pela máquina, para a execução do XGBoost. Quanto mais tarefas estão sendo usadas, mais rapidamente o código será executado, no entanto, como para alguns parâmetros já discutidos, aumentar seu valor pode custar o processamento da máquina.
- ‘scale\_pos\_weight’: controla o equilíbrio de sua amostra quando ela não é bem balanceada e o código pode acabar cometendo erros nas previsões. Ao atribuir pesos diferentes para os erros cometidos nas amostras, o parâmetro orienta as correções que serão feitas.
- ‘seed’: um parâmetro de seleção de valores aleatórios para o código. Para evitar o sobreajuste do algoritmo, sempre que a aleatoriedade é induzida no código, ela seguirá o valor escolhido por esse parâmetro.

A Tabela 3.1 apresenta os valores dispostos no código original, definidos em Hinkel et al. (2019), para cada um dos parâmetros em nossa análise.

TABELA 3.1. Parâmetros do XGBoost

Parâmetro	Valor
learning_rate	0.1
n_estimators	1000
max_depth	6
min_child_weight	1
gamma	0
subsample	0.8
colsample_bytree	0.8
objective	'binary:logistic'
nthread	1
scale_pos_weight	1
seed	27

### 3.3 Execução

Para o funcionamento do algoritmo, uma amostra de 400 estrelas é selecionada aleatoriamente do catálogo Hypatia, a partir de algumas condições. Essas

400 estrelas servirão como amostra de treinamento para a previsão do restante da amostra geral. Dentro desse grupo, 200 possuem planetas e serão usadas para o estudo de padrões de abundâncias, e as 200 restantes não possuem planetas conhecidos e servirão para o algoritmo identificar as diferenças entre as estrelas hospedeiras e não hospedeiras. A coluna de exoplanetas confirmados é então conferida pelo algoritmo, para que ele selecione aquelas estrelas que têm planetas ou não, a partir dos números 1 e 0, respectivamente. Para as estrelas com planetas, é aplicado um critério adicional, no qual definimos um valor  $\text{MaxPMass} = 0,0945 M_J$  ( $300,42 M_{\oplus}$ ) para expressar o limite mínimo de massa dos planetas gigantes (Mayor et al. 2011; Adibekyan et al. 2012a) e, para a análise dos planetas menores, impusemos um limite máximo  $\text{MaxPRad} = 4,0 R_{\oplus}$ , que é definido como o raio máximo de planetas da classe subnetunos por Petigura et al. (2018). Os testes com tais critérios, e outros com o uso de ambos serão melhor discutidos no Capítulo 5. A Tabela 3.2 apresenta os intervalos de abundâncias para cada amostra de treinamento. Essas amostras são selecionadas a partir desses intervalos, mas não necessariamente os cobrirão por completo devido à amostragem aleatória. O processo passa, então, por uma rodada de 3000 iterações, valor escolhido por não exigir excessivamente do processamento da máquina, e ainda assim entregar resultados satisfatórios, e seleciona novos dados para formar o subgrupo de 400 estrelas a cada uma delas. No total, são executadas 25 rodadas, valor este definido por entregar uma quantidade de previsões suficiente e eficaz para análise. No fim, obtemos as porcentagens para cada estrela abrigar ou não os planetas baseadas nas condições que impusemos anteriormente.

No início da seleção de amostras, outro subgrupo de estrelas também é selecionado, a *golden set*. Esse grupo é o responsável pela verificação de acurácia do código. Dez estrelas com planetas que seguem os limites de massa ou raio escolhidos são selecionadas. Ao final dos cálculos de previsão do código, estas estrelas também recebem valores de previsões e, por isso, servem como um fator de checagem da eficácia do código. Espera-se que as probabilidades sejam altas, já que as estrelas seguem as condições selecionadas. Caso não sejam, temos um indicativo da falta de acurácia do teste sendo feito.

Para a análise do comportamento probabilístico das abundâncias quando

TABELA 3.2. Intervalos de abundâncias das amostras de treinamento

Elemento	Sem Planetas Conhecidos	Planetas Gigantes (MaxPMass $\geq 0.0945 M_J$ )	Planetas Menores (MaxPRad $\leq 4.0 R_{\oplus}$ )
C	-1.38 – 1.28	-0.60 – 1.41	-0.42 – 0.76
O	-1.57 – 1.13	-0.35 – 0.60	-0.60 – 0.75
Na	-1.17 – 1.13	-0.53 – 0.68	-0.65 – 0.71
Mg	-1.52 – 0.83	-0.45 – 0.54	-0.42 – 0.36
Al	-1.39 – 1.12	-1.50 – 0.59	-0.80 – 0.40
Si	-1.00 – 0.73	-0.38 – 0.62	-0.39 – 0.40
Ca	-1.85 – 0.96	-0.71 – 0.51	-0.44 – 0.84
Ti	-1.58 – 1.10	-1.04 – 0.64	-0.38 – 0.37
V	-1.18 – 1.49	-0.46 – 0.92	-0.93 – 0.50
Cr	-1.47 – 1.50	-0.61 – 1.01	-0.54 – 0.39
Mn	-1.87 – 0.86	-0.73 – 0.85	-0.91 – 0.38
Fe	-1.63 – 1.05	-0.51 – 0.57	-0.52 – 0.43
Ni	-1.15 – 0.79	-0.46 – 0.53	-0.53 – 0.88
Y	-1.79 – 2.33	-0.87 – 0.68	-0.87 – 0.72
Co	-1.61 – 1.57	-0.46 – 0.73	–

atuam em grupos nas previsões, Hinkel et al. (2019) motaram 5 grupos (*sets*) com combinações de elementos para diferentes análises. Essas combinações ocorrem entre três classes de elementos, os siderófilos (S), litófilos (L) e os voláteis (Vl). A escolha dos elementos para essa análise vai além de sua disponibilidade no catálogo Hypatia. Os siderófilos, por exemplo, compostos por Cr, Co e Ni, são elementos presentes no núcleo terrestre juntamente com o Fe, e sua abundância nas estrelas pode estar relacionada com a formação de núcleos de planetas rochosos no sistema. Os litófilos, compostos por Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn e Y, são elementos altamente reativos e se combinam com o oxigênio para formar mineirais que são encontrados na crosta terrestre, podendo ser úteis também para a pesquisa de estrelas e suas abundâncias para a procura de planetas como a Terra. Além disso, também podem ajudar na busca por planetas gigantes, já que eles também são responsáveis por formar parte do núcleo para a acreção dos voláteis. Esses, compostos por C e O, por sua vez, têm sua significância para a formação das atmosferas planetárias. Separamos o ferro, que pertence ao grupo dos siderófilos, desse grupo, para que possamos estudar o comportamento das previsões quando ele não está presente, já que sabemos sua importância na formação de planetas gigantes. A Tabela 3.3 expõe os 5 grupos originais de elementos, e dois grupos,

isolados por uma linha horizontal, que adicionamos para estudarmos outras combinações. O primeiro (*set 6*) é sugerido no website do algoritmo *planetPrediction*<sup>2</sup> e explora a atuação de apenas 3 elementos (C, O e Fe) na previsão dos planetas. O segundo (*set 7*), criamos após notar a falta de uma análise sem a influência das abundâncias dos litófilos.

TABELA 3.3. Descrição dos sets disponíveis no código

Set	Grupos	Elementos
1	VI + L + S + Fe	C, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn, Y, Cr, Co, Ni, Fe
2	VI + L + S	C, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn, Y, Cr, Co, Ni
3	L + S + Fe	Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn, Y, Cr, Co, Ni, Fe
4	VI + L + Fe	C, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn, Y, Fe
5	L + S	Na, Mg, Al, Si, Ca, Sc, Ti, V, Mn, Y, Cr, Co, Ni
6	VI + Fe	C, O, Fe
7	VI + S + Fe	C, O, Cr, Co, Ni, Fe

Ao final da execução do código, há a opção, também, de gerar gráficos de distribuição das abundâncias químicas de um elemento X ( $[X/Fe]$ ) em função da metalicidade  $[Fe/H]$ , automaticamente. Nestes gráficos, todas as estrelas do catálogo estão dispostas, e são diferenciadas com respeito a serem estrelas com planetas confirmados, estrelas sem planetas confirmados, mas que o código previu com probabilidades acima de 90%, e as demais estrelas com menor probabilidade de abrigarem planetas. Eles estão dispostos no Capítulo 5.

### 3.4 Alteração do algoritmo

Nesta seção, apresentamos as alterações feitas diretamente no algoritmo, para que os testes apresentados no Capítulo 5 pudessem ser realizados.

Para os testes com planetas gigantes, usamos as seguintes linhas de código para estabelecer os limites de massa planetária máxima. Nesse primeiro exemplo, temos a linha para a definição da *golden set*.

```

1  if golden:
2      df2 = df.copy()

```

<sup>2</sup><https://www.github.com/nhinkel/planetPrediction>

```

3         df2.loc[df2[(df2['Exo']==1) & ...
        (df2['MaxPMass']>parameters['gas_giant_mass'])].sample(10, ...
        random_state=np.random.RandomState()).index, 'Exo'] = 0

```

Aqui, “df2” é uma cópia da tabela usada para coletar os dados de abundância e outros parâmetros, e “parameters” faz parte de um arquivo que compila parte dos valores de parâmetros que serão usados ao longo do algoritmo:

```

1  {'N.iterations': 3000,
2   'N.samples': 200,
3   'features': ['C',
4               'O', 'Na', 'Mg', 'Al',
5               'Si', 'Ca', 'Ti', 'V',
6               'Mn', 'Cr', 'Ni', 'Fe'],
7   'gas_giant_mass': 0.0945
8   'gas_giant_radius': 4.0}

```

Nesse exemplo, vemos os elementos que formam os grupos apresentados na Seção 3.3 no conjunto *features* e o número de iterações do código (*N\_iterations*). Os parâmetros “*gas\_giant\_mass*” e “*gas\_giant\_radius*” definem os limites de massa e raio mínimos para planetas gigantes que serão utilizados pelo código para as seleções das amostras de treinamento, cujo tamanho é definido pelo parâmetro *N\_samples*.

As linhas que vão definir as amostras de treinamento, seguem um padrão similar à da *golden set*:

```

1  for iteration in range(0, N_iterations):
2         #dataframe de 200 hospedeiras aleatorias com planetas gigantes
3         df_iter_with_exo = df[(df['Exo']==1) & ...
        (df['MaxPMass']>parameters['gas_giant_mass'])].sample(N_samples, ...
        random_state=np.random.RandomState())

```

Similarmente ao que ocorre com a *golden set*, aqui cria-se um novo grupo de estrelas da tabela com a coluna “Exo” = 1, o qual fará parte da amostra de treinamento.

Para a análise de planetas menores, as alterações ocorreram também para esse grupo de linhas. Nesse caso, incluímos os limites para o raio planetário máximo:

```

1 if golden:
2     df2 = df.copy()
3     df2.loc[df2[(df2['Exo']==1) & ...
4             (df2['MaxPRad']<parameters['gas-giant-radius'])].sample(10, ...
5             random_state=np.random.RandomState()).index, 'Exo'] = 0
6
7 for iteration in range(0, N_iterations):
8     #dataframe de 200 hospedeiras aleatorias com planetas menores
9     df_iter_with_exo = df[(df['Exo']==1) & ...
10    (df['MaxPRad']<parameters['gas-giant-radius'])].sample(N_samples, ...
11    random_state=np.random.RandomState())

```

Para os casos em que unimos as condições MaxPMass e MaxPRad, as linhas do código seguiram o padrão:

```

1 if golden:
2     df2 = df.copy()
3     df2.loc[df2[(df2['Exo']==1) & ...
4             ((df2['MaxPMass']>parameters['gas-giant-mass']) | ...
5             (df2['MaxPRad']>parameters['gas-giant-radius']))].sample(10, ...
6             random_state=np.random.RandomState()).index, 'Exo'] = 0
7
8 for iteration in range(0, N_iterations):
9     df_iter_with_exo = df[(df['Exo']==1) & ...
10    ((df['MaxPMass']>parameters['gas-giant-mass']) | ...
11    (df['MaxPRad']>parameters['gas-giant-radius']))].sample(N_samples, ...
12    random_state=np.random.RandomState())

```

Aqui, “|” estabelece a condição *ou* para a previsão de planetas usando um parâmetro ou o outro. Os testes envolvendo esse fator serão discutidos no próximo capítulo.



## Capítulo 4

# Resultados e Discussão

Com o algoritmo devidamente configurado, começamos a realizar testes variando diversos parâmetros em busca de previsões para planetas gigantes e menores.

### 4.1 Previsões para Planetas Gigantes

Nos primeiros testes, buscamos compreender melhor a previsão de planetas gigantes produzida pelo algoritmo e tabela originais (Hinkel et al. 2019), bem como as alterações que ocorreriam ao atualizarmos os dados com as novas estrelas adicionadas ao catálogo Hypatia.

#### 4.1.1 Dados originais e grupo 1 de elementos

Assim, o teste inicial, que tenta prever estrelas com planetas gigantes a partir do parâmetro da massa planetária máxima (MaxPMass), foi uma tentativa de alcançarmos resultados similares aos de Hinkel et al. (2019), usando o grupo 1 de elementos (ver Tabela 3.3) e o catálogo original com aproximadamente 4800 estrelas.

Após 25 rodadas, obtivemos resultados em acordo com a previsão original: previmos 8,8% (próximo às 9% do artigo de origem do código) das estrelas com probabilidades de abrigarem planetas gigantes acima de 90%, e dessas, Hinkel

et al. (2019) previram 85,3% também com esse valor. A Figura 4.1 apresenta essa comparação dos resultados de nosso teste com os do artigo. Podemos notar pelo gráfico que, das estrelas que ficaram abaixo de 90% no nosso teste, a maioria está bem próxima deste valor, com todas estando acima de 75%. Possivelmente, essa mínima queda de 15% é causada pela diferença no número de rodadas entre os testes (25 ao invés de 100, a fim de tornar os testes mais eficientes) e a natureza probabilística da análise.

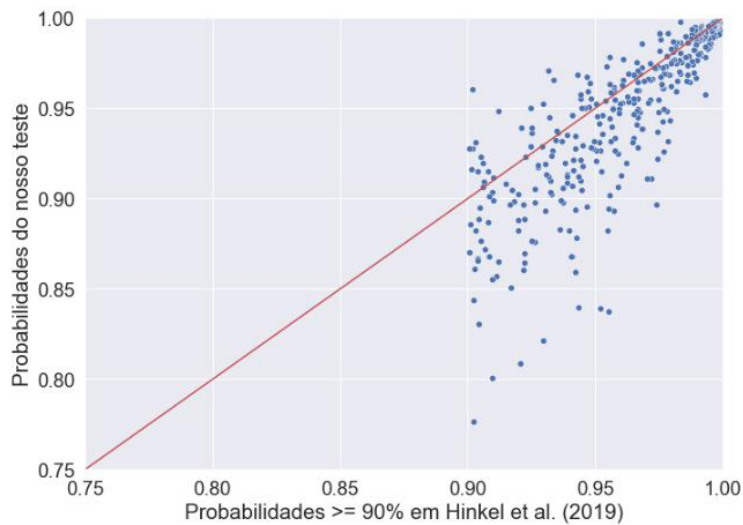


FIGURA 4.1. Gráfico de distribuição que apresenta as estrelas com probabilidades acima de 90% de abrigarem planetas gigantes, obtidas em Hinkel et al. (2019), em função das que obtivemos tentando replicar esses resultados. A linha no gráfico é uma função  $y=x$ , que ajuda a apontar a semelhança de resultados que obtivemos.

A *golden set* teve uma probabilidade média de 76,7%, em concordância com a encontrada no artigo (75%), e alterou-se razoavelmente a ordem de elementos considerados relevantes em Hinkel et al. (2019), mas quatro dos 5 elementos mais importantes para a previsão se mantiveram no topo do gráfico (Figura 4.2 [a]). Essa troca de posições é comum e pode ser relevada se também levarmos em conta o fato de termos usado um número de rodadas diferente para a previsão do algoritmo.

A Tabela 4.1 resume os principais resultados obtidos nesta e nas seções seguintes. As seções não representadas possuem tabelas próprias. As colunas apresentam os principais resultados para cada teste, dados pelos números das respectivas seções em que são apresentados: “Probabilidade  $\geq 90\%$ ” expõe a porcentagem de estrelas com probabilidade de abrigarem planetas gigantes igual ou maior do

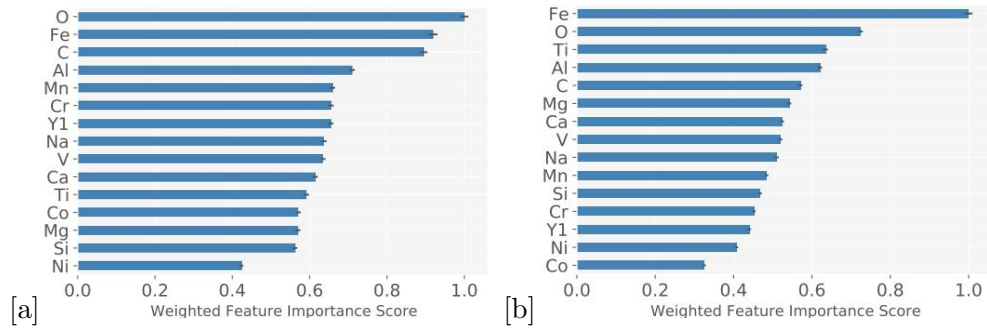


FIGURA 4.2. Gráficos que mostram a ordem de relevância de cada elemento para as previsões de planetas gigantes. *Weighted Feature Importance Score* denota a pontuação ponderada de importância de cada elemento, ou seja, quanto mais próximo de 1,0 maior foi a contribuição da abundância daquele elemento para a previsão sendo feita. Os elementos estão listados verticalmente em ordem de crescimento dessa pontuação. O painel [a] refere-se à previsão de planetas gigantes usando os dados originais do catálogo Hypatia e a massa planetária máxima (Seção 4.1.1); O painel [b] refere-se à previsão de planetas gigantes usando o catálogo Hypatia atualizado e a massa planetária máxima (Seção 4.1.2).

que 90%, “Estrelas Previstas” exibe a porcentagem de estrelas que obtiveram um valor de probabilidade diferente de 0,0, “Probabilidade Média *golden set*” indica a probabilidade média de todas as estrelas selecionadas para a *golden set* em cada teste, e “Elementos mais relevantes” exibe os elementos mais importantes para a previsão de planetas gigantes em cada teste.

TABELA 4.1. Resumo dos Testes com Planetas Gigantes

Teste	Probabilidade $\geq 90\%$ (%)	Estrelas Previstas (%)	Probabilidade Média <i>golden set</i> (%)	Elementos mais relevantes
4.1.1	8,8	98,9	76,7	O, Fe, C, Al, Mn
4.1.2	10,3	99,9	74,7	Fe, O, Ti, Al, C
4.1.5	10,4	99,8	79,7	Fe, O, Ti, Al, C
4.1.6	8,4	56,7	93,7	Na, O, V, C, Al

O algoritmo também gerou gráficos que apresentam a distribuição de todas as estrelas do catálogo Hypatia de acordo com as abundâncias de cada elemento em relação às do ferro, para cada um dos testes. Esses gráficos evidenciam as estrelas com planetas confirmados, as sem planetas confirmados que o algoritmo previu com  $p < 90\%$ , e aquelas previstas com  $p \geq 90\%$ . A partir desses gráficos, podemos visualizar os padrões de abundância seguidos por cada tipo de amostra. Apresentamos os gráficos, na Figura 4.3, para os quatro elementos mais relevantes (além do ferro) para o teste desta seção.

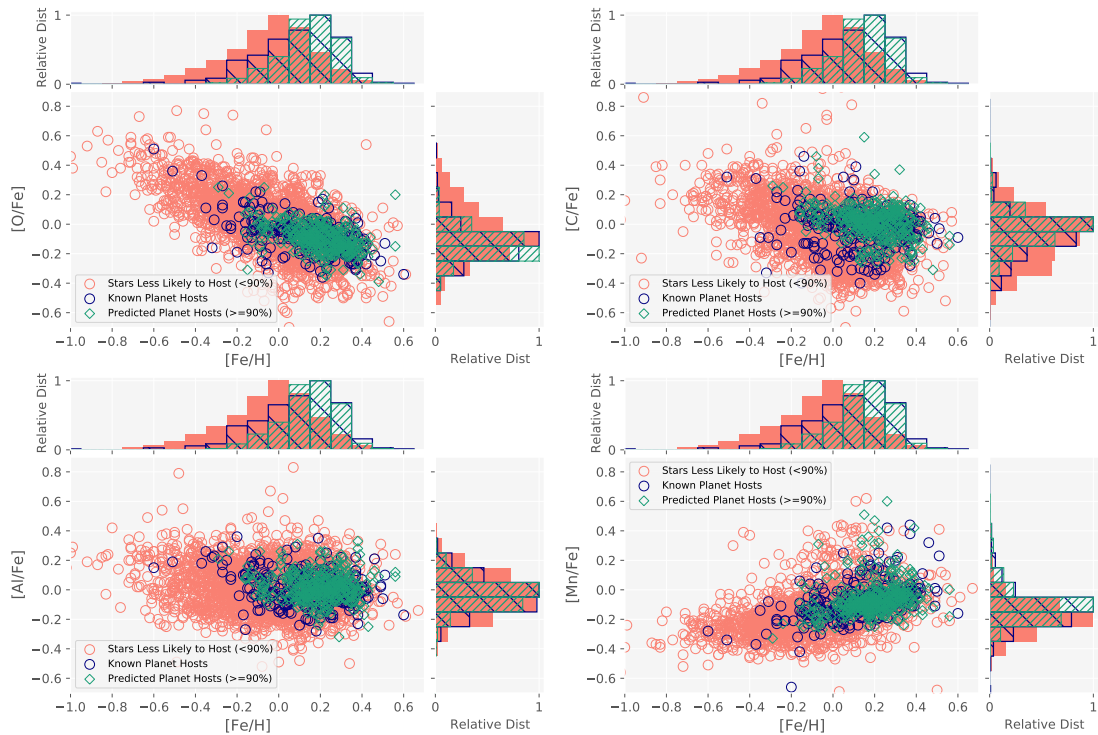


FIGURA 4.3. Gráficos de distribuição das abundâncias químicas de um elemento X ( $[X/Fe]$ ) em função da metalicidade  $[Fe/H]$  para a previsão de planetas gigantes. Aqui, todas as estrelas do catálogo estão dispostas, e são diferenciadas com respeito a serem estrelas com planetas confirmados (azul), estrelas sem planetas confirmados, mas que o código previu com probabilidades acima de 90% (verde), e as demais estrelas com menor probabilidade de abrigarem planetas (vermelho). Apresentamos também as distribuições marginais das abundâncias. Nesse primeiro gráfico, mostramos a previsão de planetas gigantes usando o catálogo Hypatia original. Os elementos mais relevantes para esse teste, além do ferro, foram: O, C, Al, Mn.

#### 4.1.2 Dados atualizados e grupo 1 de elementos

A partir do segundo teste, passamos a usar o nosso atual conjunto de aproximadamente 7000 estrelas (descrito no Capítulo 3), por isso, a partir da próxima seção, não indicaremos no título os dados sendo utilizados. Para a primeira análise, decidimos testar a previsão para planetas gigantes de maneira idêntica ao teste original, porém usando a nova tabela com mais dados. Assim, a previsão de planetas gigantes obteve os seguintes resultados: O algoritmo previu 10,3% das estrelas com probabilidades de abrigarem planetas gigantes acima de 90%, e a probabilidade média da *golden set* foi de 74,7%. As estrelas previstas nesse teste com probabilidades acima de 90% estão listadas no Apêndice A.

Aqui, vemos que se manteve um padrão nos resultados de previsão, quando comparados aos obtidos na seção anterior. Com essa atualização do catálogo,

alterou-se novamente a ordem dos elementos mais relevantes para a previsão (Figura 4.2 [b]), com o ferro, considerado o mais relevante para a formação de planetas gigantes (e.g., Fischer & Valenti 2005; Ghezzi et al. 2018) apresentando a maior importância, e o titânio subindo significativas posições. O carbono, por outro lado, caiu em relevância. Os resultados que descrevemos estão apresentados na Tabela 4.1.

Para esse teste, fizemos uma comparação com a previsão original, de forma que pudéssemos ver quais estrelas obtinham probabilidades acima de 90% em ambos os testes. Nesse caso, 50,1% das estrelas previstas no teste anterior se repetiram nessa análise. Acreditamos que o grupo de estrelas novamente não previsto possa ter alterado suas probabilidades com a inclusão de muitas novas estrelas, além da mudança que alguns valores de abundâncias possam ter sofrido com a inclusão de novos catálogos. Essa mudança ocorre pois, como explicamos no Capítulo 3, com a inclusão de novas abundâncias coletadas de artigos recém adicionados, uma nova mediana é obtida e os valores para cada elemento em cada estrela são atualizados. Para compararmos esses números entre as duas amostras, original e atual, construímos a Figura 4.4, que mostra a diferença média das abundâncias químicas de cada elemento para todas as estrelas que estavam no catálogo original. Todos os valores estão abaixo do módulo 0,02, e as barras de erro indicam algumas alterações mais relevantes. Assim, em alguns casos, as mudanças nas abundâncias das estrelas poderia ajudar a explicar a mudança na probabilidade.

Dessa forma, podemos considerar o aumento dos dados com a atualização da tabela e a consequente mudança na ordem de relevância dos elementos o principal motivo para a diminuição das probabilidades de estrelas que antes tinham  $p \geq 90\%$ . O histograma da Figura 4.5 mostra que, apesar dessa diferença de estrelas previstas entre os testes, a maioria (87,7%) das estrelas do catálogo original com  $p \geq 90\%$  teve probabilidades acima de 75% no novo teste.

Apresentamos os gráficos, na Figura 4.6, com a distribuição de todas as estrelas do catálogo Hypatia de acordo com as abundâncias de cada elemento em

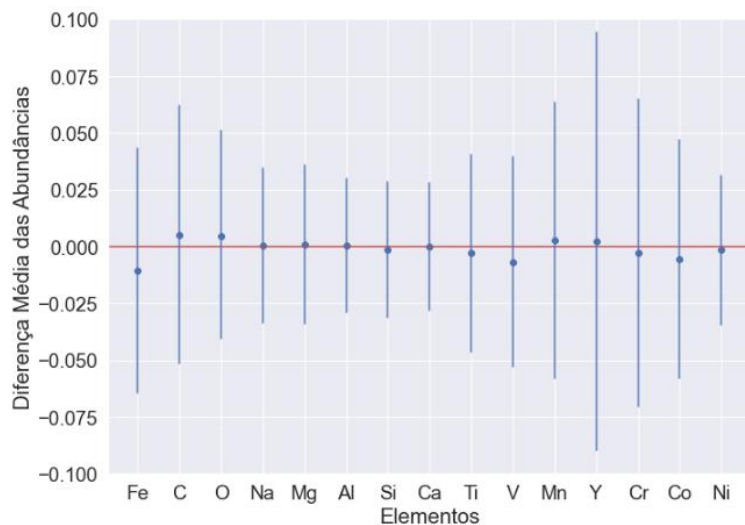


FIGURA 4.4. Diferença média das abundâncias entre os catálogos do Hypatia atualizado e original. As barras de erro representam o desvio padrão da média.

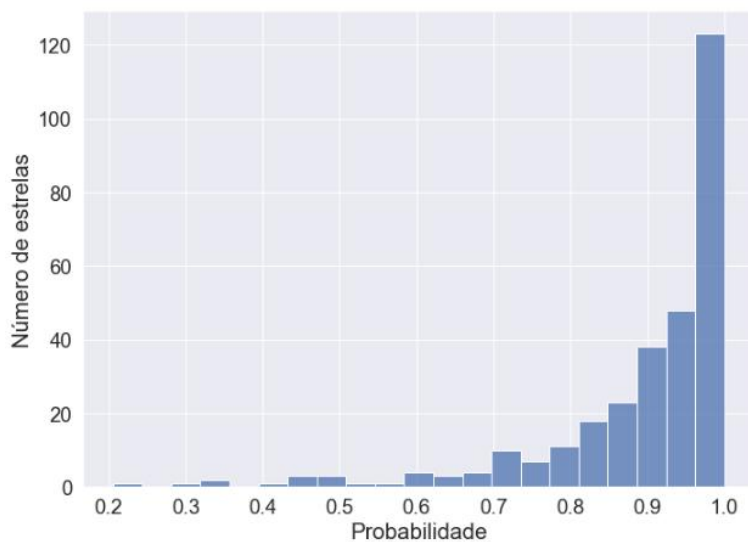


FIGURA 4.5. Histograma com as probabilidades das estrelas previstas no teste com o catálogo atualizado e que apresentaram  $p > 90\%$  no teste com o catálogo Hypatia original.

relação às do ferro para o teste realizado nesta seção. Analisando-os, podemos notar que a nova versão do catálogo contém um número maior (314, em comparação com 56 do catálogo original) de estrelas com planetas para  $-0,2 < [\text{Fe}/\text{H}] < 0,0$ , deixando esta região mais povoada. Apesar disso, as previsões do programa ficam mais concentradas na região  $[\text{Fe}/\text{H}] > 0,0$ , diferentemente dos resultados nos gráficos da Figura 4.3. Não sabemos exatamente a causa disso, mas é possível que ela esteja relacionada com a posição do ferro como o elemento mais importante e a conhecida correlação planeta-metalicidade.

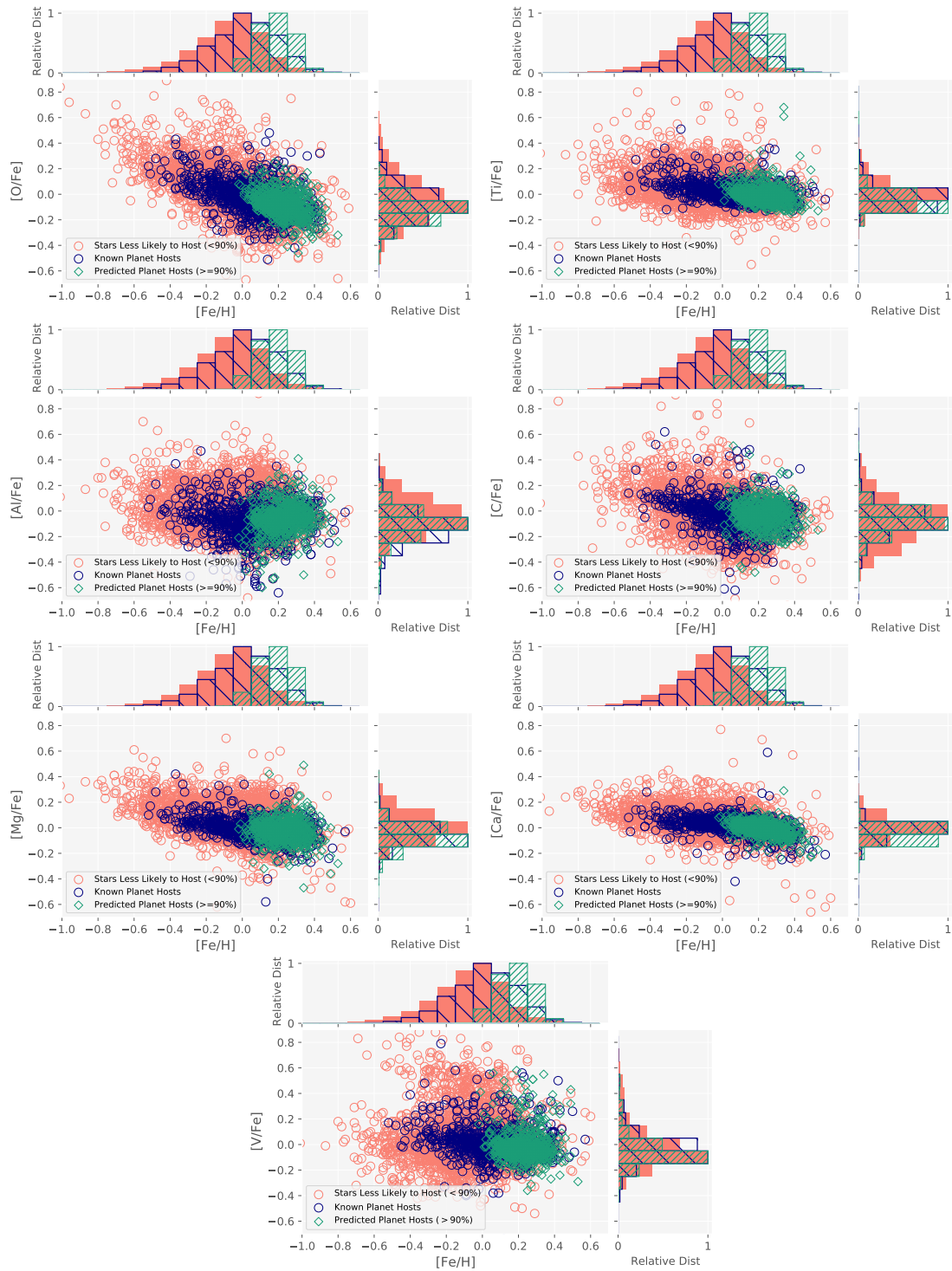


FIGURA 4.6. Gráficos de distribuição das abundâncias químicas de um elemento X ( $[X/Fe]$ ) em função da metalicidade  $[Fe/H]$ . Os painéis, símbolos e cores são os mesmos da Figura 4.3. Previsão usando o catálogo Hypatia atualizado.

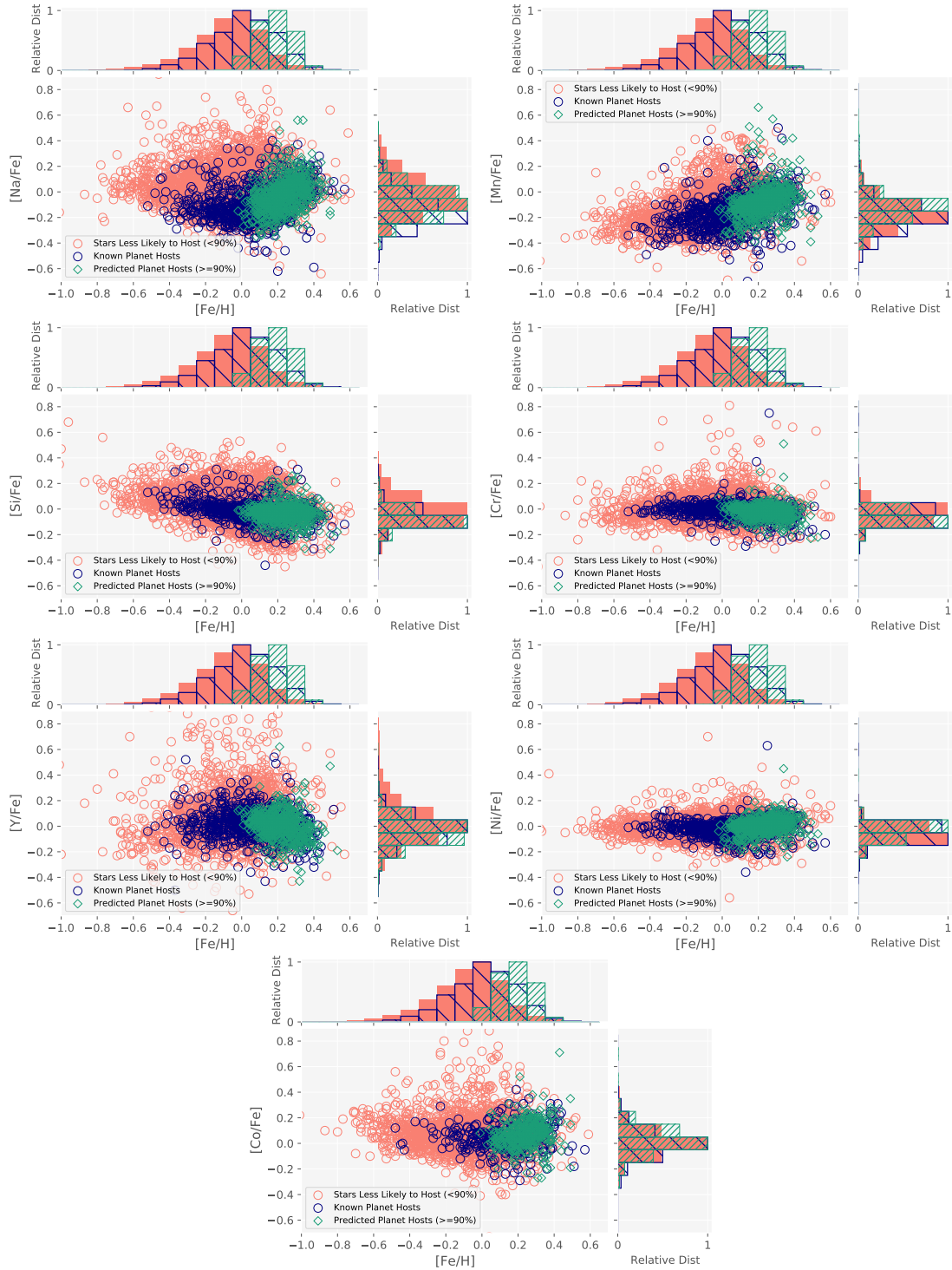


FIGURA 4.6. Continuação da Figura 4.6.

### 4.1.3 Grupo 1 de elementos e diferentes limites de Massa Planetária Máxima

Em seguida, a fim de investigar se o limite de massa escolhido em Hinkel et al. (2019) influenciava os resultados, variamos o limite mínimo para a inclusão



de estrelas com planetas na amostra de treinamento. No primeiro desses testes, impusemos um limite mínimo com a massa de Júpiter ( $317,8 M_{\oplus}$ ) e assim seguimos, diminuindo para metade da massa de Júpiter ( $158,9 M_{\oplus}$ ), uma massa de Saturno ( $95,16 M_{\oplus}$ ), Netuno ( $17,15 M_{\oplus}$ ), Urano ( $14,54 M_{\oplus}$ ), e, enfim, alcançamos  $10,0 M_{\oplus}$ , sendo que, abaixo desse limite, entramos no regime das superterras. Incluímos também outros dois limites mínimos, de  $5,0$  e  $2,0 M_{\oplus}$ , para testarmos a influência das abundâncias dessas estrelas com planetas menores na previsão. Os resultados desses testes estão dispostos na Tabela 4.2. Acrescentamos o valor com  $30,0 M_{\oplus}$  para efeitos de comparação e ele vem da Seção 4.1.2.

TABELA 4.2. Teste Variando a Massa Mínima para Planetas Gigantes

Massa Mínima ( $M_{\oplus}$ )	Probabilidade $\geq 90\%$ (%)	Probabilidade Média <i>golden set</i> (%)	Elementos mais relevantes
30,0	10,3	74,7	Fe, O, Ti, Al, C
317,8	9,6	74,3	Fe, O, Al, Ti, C
158,9	10,2	75,4	Fe, O, Ti, Al, C
95,16	10,5	74,2	Fe, Ti, O, Al, C
17,15	10,2	71,8	Fe, O, Ti, Al, C
14,54	10,4	72,6	Fe, O, Ti, Al, C
10,0	10,7	69,9	Fe, Ti, O, Al, C
5,0	9,5	76,6	Fe, Ti, O, Al, C
2,0	9,5	70,8	Fe, Ti, O, Al, C

O estudo da variação dos valores de massa mínima para o algoritmo procurar por planetas gigantes trouxe resultados que mostram a estabilidade das previsões, independentemente do limite utilizado. Manteve-se o padrão do ferro como elemento mais relevante, com pouca variância entre os outros elementos mais importantes (Figura 4.7), um resultado condizente com a literatura (e.g., Ghezzi et al. 2021). Assim, podemos notar que não há uma diferença visível quando incluímos apenas estrelas com gigantes gasosos na amostra de treinamento ou se consideramos gigantes gelados e subnetunos. Esses resultados podem ser uma consequência dos baixos números de estrelas que são incluídas quando diminuimos o limite. Quando trocamos o limite mínimo de MaxPMass da massa de Saturno para a de Netuno, por exemplo, 64 novas estrelas são adicionadas à possível amostra de treinamento. Já da massa de Urano para  $10,0 M_{\oplus}$ , apenas 14 estrelas entram nessa amostra, e o padrão segue para os dois últimos limites de massa.

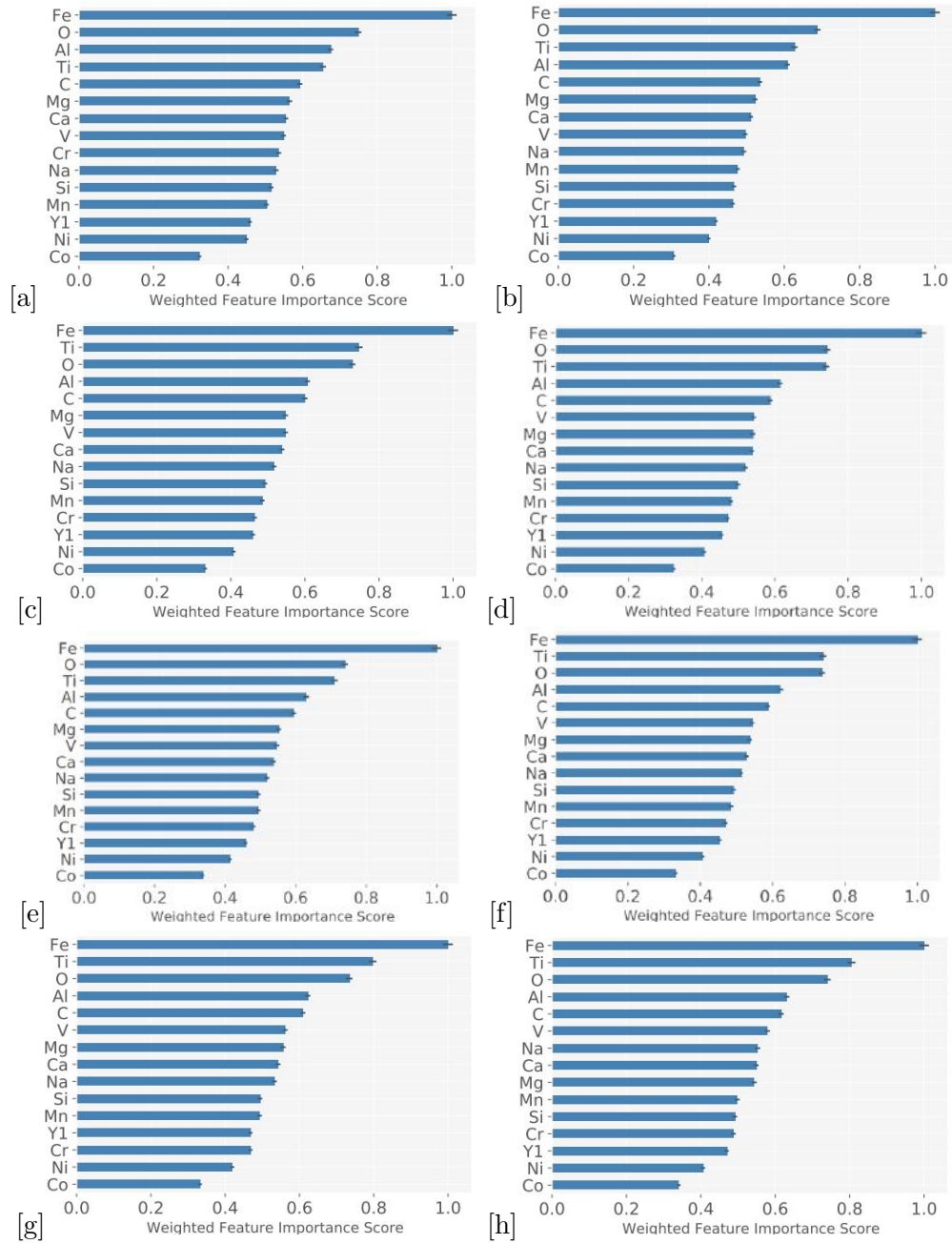


FIGURA 4.7. Gráficos de relevância dos elementos para os testes com planetas gigantes e variação da massa planetária máxima. [a]:  $M_J$ ; [b]:  $0,5 M_J$ ; [c]:  $M_{Saturno}$ ; [d]:  $M_{Urano}$ ; [e]:  $M_{Netuno}$ ; [f]:  $10,0 M_{\oplus}$ ; [g]:  $5,0 M_{\oplus}$ ; [h]:  $2,0 M_{\oplus}$ .

#### 4.1.4 Outros grupos de elementos

Fizemos, também, um estudo das previsões de planetas gigantes para cada um dos grupos de elementos dispostos na Tabela 3.3. Os resultados para esses testes estão dispostos na Tabela 4.3 e na Figura 4.8.

Analisando os resultados dos testes, encontramos valores interessantes entre

TABELA 4.3. Teste variando os grupos de elementos (Planetas Gigantes)

Grupo de Elementos	Probabilidade $\geq 90\%$ (%)	Estrelas Previstas (%)	Probabilidade Média <i>golden set</i> (%)	Elementos mais relevantes
2	9,2	90,4	74,8	Ti, O, Al, C, Mg
3	8,8	96,4	76,9	Fe, Ti, Mg, Na, V
4	10,1	55,5	75,2	Fe, O, C, Mg, Al
5	8,9	86,3	70,1	Ti, Mg, Na, Si, V
6	11,7	95,1	67,3	Fe, C, O
7	11,3	95,5	75,7	Fe, Cr, Ni, C, O

os grupos de elementos. Para os grupos 2, 3, 5, 6 e 7, a diferença em relação aos testes com o grupo 1, foi a singela queda no número de estrelas previstas. Anteriormente, quase 100% das estrelas recebiam valores de previsão diferentes de 0,0. Agora, esse número caiu entre 5% e 15% nos testes. Para esses casos, a queda no número de estrelas previstas pode ser facilmente justificada pelo corte do ferro da lista de abundâncias, já que as previsões dos grupos 2 e 5 (onde ele não está presente), sofreram o maior déficit. A maior alteração dos testes, no entanto, ocorreu na análise do grupo 4 de elementos, em que notamos uma queda no número de estrelas previstas notável ( $\approx 45\%$ ), similar ao observado na Seção 4.1.6. A diferença desse grupo de elementos em relação aos demais, é a falta dos siderófilos na amostra (Cr, Co e Ni). Assim, análises desse conjunto de elementos em projetos futuros para compreender essa queda no número de estrelas previstas são incentivadas.

A respeito da ordem dos elementos para todos os testes dessa seção, notamos que o titânio sobe em relevância quando o ferro não está presente, e os voláteis (C e O) mantêm também posições altas nessa ordem (ver Tabela 4.3 e Figura 4.8). Também é possível notar que, quando removemos os siderófilos da amostra (Figura 4.8 [c]), o titânio perde a relevância para os voláteis e, quando removemos os litófilos, o carbono e o oxigênio também perdem suas posições de relevância (Figura 4.8 [f]). No entanto, apesar dessas análises apresentarem resultados interessantes, é possível ver que a lista de elementos se mantém parecida entre os testes, dadas as especificidades de cada caso, com poucas alterações na ordem de importância em comparação com o teste com o grupo 1 de elementos.

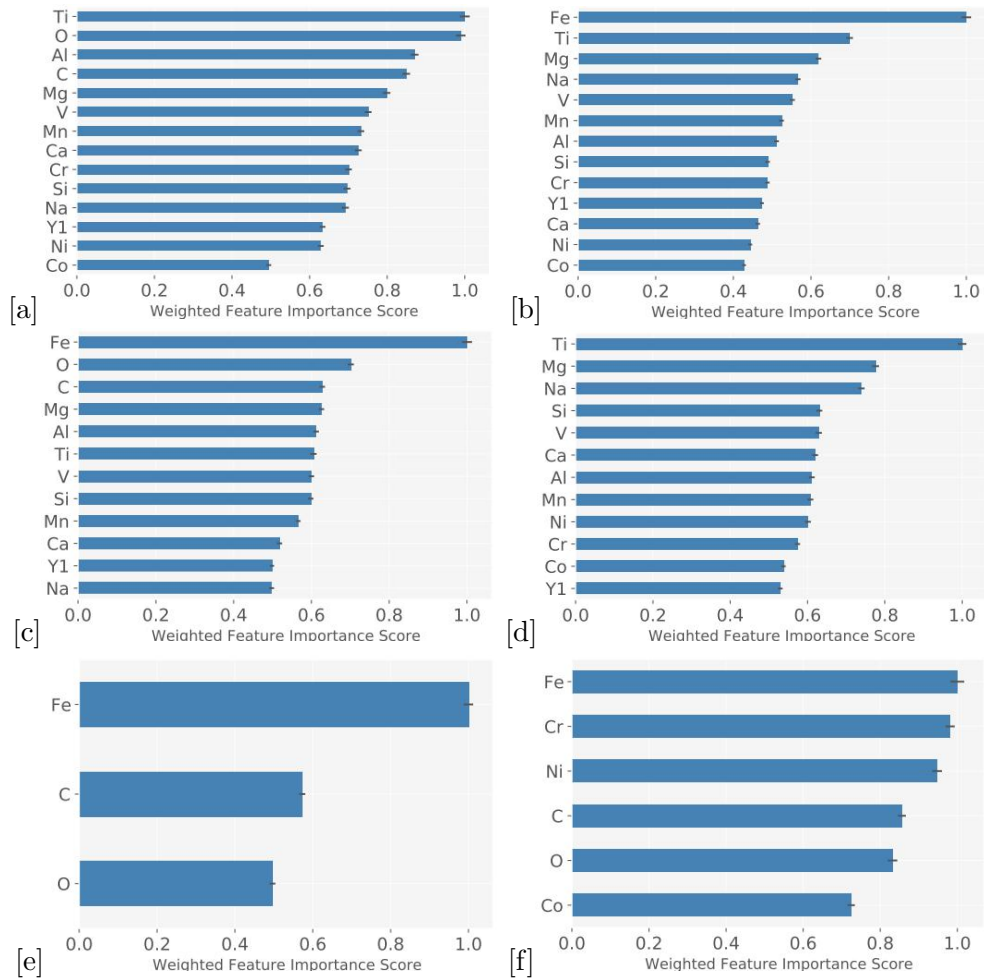


FIGURA 4.8. Gráficos de relevância dos elementos para os testes com planetas gigantes e a variação de grupos de elementos. [a]: Grupo 2; [b]: Grupo 3; [c]: Grupo 4; [d]: Grupo 5; [e]: Grupo 6; [f]: Grupo 7.

#### 4.1.5 Grupo 1 de elementos e inclusão do parâmetro Raio Planetário Máximo

Em seguida, realizamos o primeiro teste alterando o código original do algoritmo (ver Seção 3.4). Nessa análise, previmos planetas gigantes usando não só o critério da massa planetária máxima, mas também o raio máximo ou seja, o código optaria por selecionar aquelas estrelas que atendessem ao limite da massa *ou* do raio planetário máximos. Escolhemos usar  $4,0 R_{\oplus}$  (Seção 3.3) como limite mínimo para considerarmos que o sistema possui um planeta gigante. Dessa forma, definimos que um sistema possui um planeta gigante caso o maior planeta da estrela tenha  $\text{MaxPMass} \geq 0,0945 M_J$  *ou*  $\text{MaxPRad} \geq 4,0 R_{\oplus}$ . Escolhemos a condição *ou* pois poucas estrelas (33) possuem valores para ambos os parâmetros.

Para esse teste, 10,4% das estrelas têm alta chance ( $p \geq 0,9$ ) de abrigarem planetas gigantes, e a probabilidade média da *golden set* aumentou para 79,7% (ver Tabela 4.1). Fizemos, também, uma comparação entre as estrelas previstas com probabilidades acima de 90% nesse teste e no teste 4.1.2, e encontramos 86,9% de estrelas em comum, o que mostra um bom acordo. Podemos notar, na Figura 4.9, a faixa de estrelas que adicionamos para testagem quando acrescentamos o parâmetro MaxPRad na análise. Apesar das novas estrelas ocuparem intervalos similares aos da amostra com  $\text{MaxPMass} \geq 0,0945 M_J$ , elas contribuem para o aumento da porcentagem média da *golden set*.

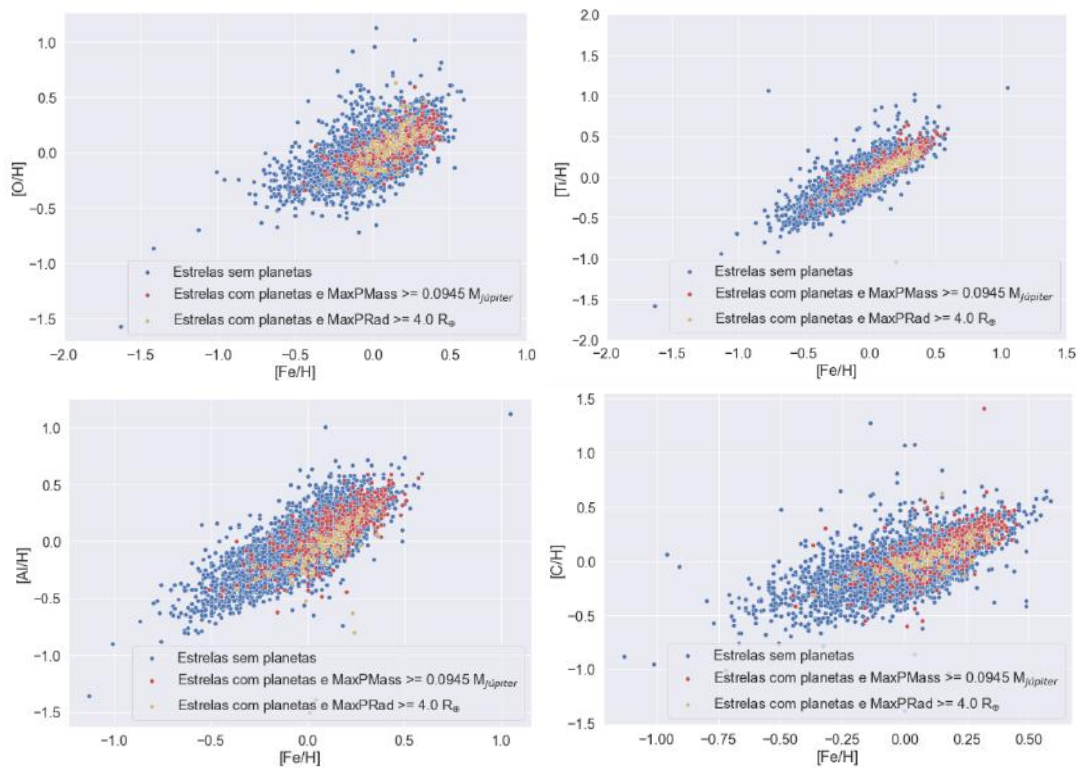


FIGURA 4.9. Gráficos  $[X/H]$  versus  $[Fe/H]$  para o teste discutido na Seção 4.1.5. Aqui apresentamos as estrelas sem planetas confirmados na cor azul, as com planetas e  $\text{MaxPMass} \geq 0,0945 M_J$  em vermelho e as estrelas adicionadas ao teste, com  $\text{MaxPRad} \geq 4,0 R_{\oplus}$ , em amarelo. A distribuição conta com os 4 elementos mais relevantes (além do ferro) para previsão.

Por fim, a adição do parâmetro do raio planetário máximo não gerou alterações muito significativas na ordem de relevância dos elementos em comparação ao teste em que apenas o critério da massa máxima é usado (Tabela 4.1 e Figura 4.10 [a]). Esse aumento de 86 estrelas disponíveis para a amostra de treinamento, que antes dispunha de 476 estrelas, contribuiu principalmente para o aumento da probabilidade média da *golden set*, ou seja, houve um leve ganho de acurácia nas previsões.

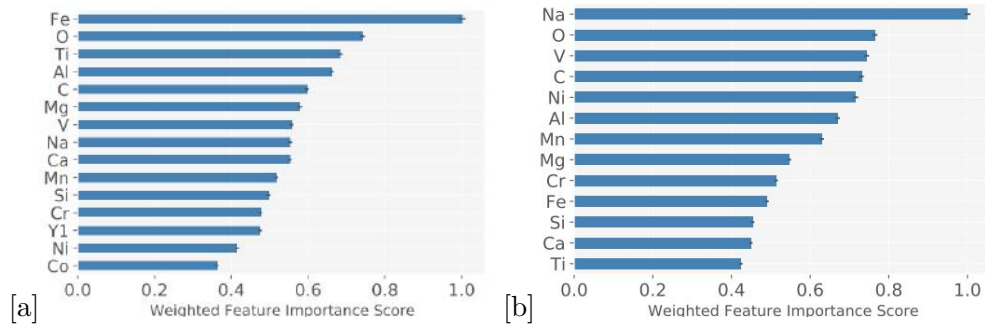


FIGURA 4.10. Gráficos de relevância dos elementos para os testes com planetas gigantes envolvendo o raio planetário máximo. [a]: Teste da Seção 4.1.5; [b]: Teste da Seção 4.1.6.

#### 4.1.6 Grupo 1 de elementos e Raio Planetário Máximo como o único critério

A fim de testarmos o comportamento do novo parâmetro que adicionamos ao algoritmo, fizemos um teste removendo a massa planetária máxima e usando apenas o limite de raio planetário máximo como critério para prevermos os planetas gigantes. Em primeira análise, para  $\text{MaxPRad} \geq 4,0 R_{\oplus}$  e  $\text{MaxPRad} \geq 3,0 R_{\oplus}$ , menos de 200 estrelas estavam disponíveis para a amostra de treinamento (86 e 159 estrelas, respectivamente), valor abaixo da quantidade necessária para o algoritmo funcionar. Frente a isso, decidimos utilizar o limite  $\text{MaxPRad} \geq 2,0 R_{\oplus}$ , que define aproximadamente o limite entre subnetunos e superterras (Martinez et al. 2019), aumentando assim, o número de estrelas disponíveis para treinamento, e, consequentemente, contribuindo para a acurácia do código.

Para esse teste, 8,3% das estrelas previstas tiveram probabilidades acima de 90% e a média da *golden set* ficou em 93,9%. No entanto, apenas 43% das estrelas receberam um valor de previsão, um resultado muito baixo, tendo em vista que nos testes anteriores, quase 100% das estrelas estavam sendo previstas, mesmo que com probabilidades próximas a 0,0. Além disso, o ítrio subiu muitas posições na ordem de relevância.

Apesar dos resultados da *golden set* serem bastante positivos, é improvável que o algoritmo tenha previsto corretamente as estrelas e suas probabilidades, já que, por exemplo, reconhecemos o ferro como um elemento de alta relevância para a previsão de planetas gigantes (e.g., Fischer & Valenti 2005), e, nesse teste, ele perdeu muitas posições na ordem de importância para as previsões.

Analisando a quantidade de estrelas da amostra geral que possuem valores de abundância para o ítrio, notamos que, apesar de estar presente em mais de 50% dos dados (53,7%), seus poucos valores possivelmente estavam gerando o viés nos gráficos de importância dos elementos. Além dele, o cobalto, apesar de não aparecer com alta relevância no gráfico, está presente em apenas 3 estrelas da amostra de treinamento, um número não aceitável para o funcionamento do código. Por isso, refizemos o teste removendo o Co e o Y do grupo 1 de elementos, e encontramos números similares aos anteriores de previsão: 8,4% das estrelas com  $p \geq 90\%$ , 56,7% com resultado de previsão (um aumento bom, mas ainda distante dos valores anteriores, próximos a 100%), e probabilidade média da *golden set* igual a 93,7% (ver Tabela 4.1). Apesar disso, a ordem de elementos mais relevantes se alterou, e agora o sódio, o oxigênio, o vanádio, o carbono e o níquel foram os maiores atuantes para a previsão de estrelas (ver Figura 4.10 [b]).

Essa nova ordem de relevância é interessante, pois o ferro ainda se manteve muitas posições abaixo dos principais elementos e o sódio assumiu a liderança, resultado esse não reconhecido na literatura. Suspeitamos que, ao adicionarmos estrelas com  $\text{MaxPRad} < 4,0 R_{\oplus}$  na amostra disponível para previsão, estamos incluindo os subnetunos, para os quais não há uma correlação planeta-metalicidade como no caso dos gigantes. O surgimento do sódio e do vanádio como elementos relevantes para a previsão de planetas será mais explorada nas seções seguintes, quando veremos que, para planetas menores, esse padrão parece se repetir.

## 4.2 Testes com Planetas Menores

Como mostrado no Capítulo 1, a busca por um padrão de abundâncias característico nas estrelas hospedeiras de planetas netunianos e terrestres segue inconclusiva (e.g., Petigura et al. 2018; Wilson et al. 2022). Nosso trabalho surge diante dessa inconclusão, como ferramenta para colaborar na procura. O aumento dos dados para planetas menores no catálogo Hypatia permitiu que esta análise fosse feita, e, apesar do código não ter sido originalmente construído para essas previsões, o adaptamos para que fosse (conferir Seção 3.4). Assim, nas seções

seguintes apresentaremos resultados que acreditamos que contribuirão significativamente para a área.

Dando início aos testes prevendo planetas menores, começaríamos usando apenas o critério do raio planetário máximo. Definimos então que estrelas com planetas cujo raio máximo é  $4,0 R_{\oplus}$  (Seção 3.3) entrariam para a amostra de treinamento.

Analisando resultados preliminares desses testes, percebemos que os elementos ítrio e cobalto estavam com uma importância surpreendentemente alta nos gráficos de relevância de elementos. O ítrio, como já mostrado na Seção 4.1.6, estava presente em um número próximo ao limite de 50% de presença de abundâncias na amostra geral (melhor discutido na Seção 2.2), e apresentou-se com alta relevância para a previsão dos planetas menores, indicando um claro viés nas probabilidades. Já o cobalto não estava presente em nenhuma estrela com o parâmetro raio planetário máximo  $\text{MaxPRad} \leq 4,0 R_{\oplus}$ , e por isso poderia gerar efeitos ainda mais enviesados que os que ocorreram com o ítrio. Diante disso, removemos esses elementos dos grupos existentes no código, de forma a evitarmos esses possíveis vieses e obtermos resultados mais confiáveis. Optamos por manter esses elementos nas análises anteriores com planetas gigantes, pois o ítrio respeita os critérios e possui resultados esperados, e o cobalto está presente em estrelas suficientes que vão para a amostra de treinamento, mesmo não sendo tão relevante para as previsões. A exceção ocorreu no teste da Seção 4.1.6, no qual, como já comentamos anteriormente, resultados diferentes do esperado surgiram.

#### 4.2.1 Grupo 1 de elementos

No primeiro teste com planetas menores, o algoritmo previu 9,6% das estrelas com mais de 90% de probabilidade de abrigarem planetas menores, e a *golden set* teve probabilidade média de 86%. Resultado esse muito positivo, tendo em vista que a *golden set* é um fator de eficácia do algoritmo e, aqui, sua probabilidade média sofreu um aumento considerável. A Tabela 4.4 resume os principais resultados obtidos nesta e nas seções seguintes com testes envolvendo as previsões



de planetas menores. As colunas seguem os padrões da Tabela 4.1, apresentando os principais resultados para cada teste.

TABELA 4.4. Resumo dos testes com planetas menores

Teste	Probabilidade $\geq 90\%$ (%)	Estrelas Previstas (%)	Probabilidade Média <i>golden set</i> (%)	Elementos mais relevantes
4.2.1	9,6	57,1	86,0	Na, O, V, Al, C
4.2.2	7,9	55,6	92,2	Na, O, V, C, Al
4.2.5	10,0	69,7	88,9	Na, V, O, C, Ti

Podemos ver que a ordem dos elementos mais importantes teve alterações nítidas, em comparação às análises dos gigantes. Os voláteis se mostraram os mais relevantes, juntamente com o sódio, o vanádio e o alumínio enquanto, por outro lado, o ferro, elemento de extrema importância na previsão de planetas gigantes, caiu significativas posições (Figura 4.11). Esses resultados são muito importantes e podem apresentar novos elementos relevantes nas estrelas para a formação de planetas menores já que para alguns desses, até onde sabemos, não há resultados na literatura. Para o carbono e o alumínio, por outro lado, Wilson et al. (2022) mostraram que abundâncias mais elevadas deles podem favorecer a ocorrência de subnetunos e superterras. Já Da Silva et al. (2015) sugeriram que o vanádio pode ser importante, mas não para a formação de planetas menores e sim para gigantes. Investigamos a presença do sódio e do vanádio na amostra de estrelas sem planetas confirmados, e notamos que os valores de suas abundâncias estão presentes em mais de 70% dessas estrelas, excluindo, portanto, o viés gerado pela pouca presença na amostra. O fato do ferro ter decaído em relevância, também é um resultado interessante e a literatura já o sugeriu (e.g., Buchhave & Latham 2015). Além disso, a presença de sua abundância em todas as estrelas da amostra também garante parte da acurácia do teste.

Quando comparamos os dados do teste nesta seção com os da Seção 4.1.2, notamos que o algoritmo previu aqui, 25% das estrelas previstas no primeiro teste com probabilidades altas ( $p \geq 0,9$ ). A partir desse resultado, é possível especular que essas estrelas podem abrigar planetas de diferentes tipos, mas é mais provável que ele indique a dificuldade do código em encaixar essas estrelas em um padrão específico de abundâncias. Assim, ele resulta em probabilidades altas para os

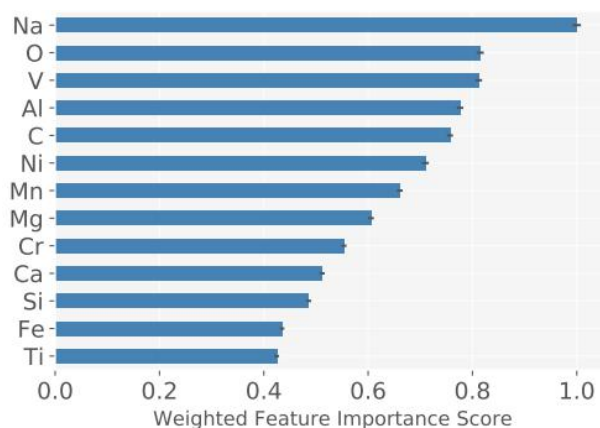


FIGURA 4.11. Gráfico de relevância dos elementos para o teste com planetas menores e o parâmetro raio planetário máximo.

dois testes visto que C e O, por exemplo, são relevantes para a presença tanto de planetas gigantes quanto menores.

Outro resultado intrigante, no entanto, foi que apenas cerca de 57,1% das estrelas tiveram um resultado de previsão, acontecimento similar ao discutido na Seção 4.1.6. A discrepância na quantidade de estrelas com previsões nos testes usando ou o parâmetro da massa planetária máxima ou o raio planetário máximo gerou um alerta para que estudássemos melhor o caso dos planetas menores.

Para isso, supusemos que parte das estrelas não estavam obtendo resultados de previsão por terem probabilidade final de abrigarem planetas menores igual a 0,0, que é um resultado aceitável, mas devido a uma linha no código do algoritmo, não é retornado na tabela final com os valores de probabilidade. Frente a isso, fizemos um teste removendo essa condição do algoritmo e confirmamos que o algoritmo previu as estrelas ocultas nas previsões envolvendo o parâmetro do raio planetário máximo com probabilidades iguais a 0,0. Notamos que essas estrelas eram sim amostradas pelo código em todas as rodadas, mas não eram previstas nenhuma vez.

Apesar de encontrarmos uma possível resposta para o problema das previsões de planetas menores, devido à grande amostra de estrelas com esse valor nulo de previsão, decidimos estudar mais a fundo o porquê do código decidir com tanta precisão quais estrelas tinham 0% de chance de não hospedarem planetas menores. Para isso, realizamos alguns testes adicionais.

Em uma primeira análise, estudamos as abundâncias para os dois grupos dessas estrelas: aquelas com resultados de previsão e aquelas sem resultados de previsão. Notamos que muitas das estrelas que o algoritmo não previu, não têm as abundâncias dos elementos destacados como mais relevantes para a análise dos planetas menores. Por outro lado, as estrelas previstas têm, em sua maioria, esses valores de abundância. A Tabela 4.5 expõe esses dados e apresenta o número de estrelas de cada grupo (previstas e não previstas) que possuem abundância para cada um dos 5 elementos mais relevantes na previsão de planetas menores. Com isso, um dos motivos para o código não prever essas estrelas pode estar nessa falta das abundâncias consideradas relevantes pelo algoritmo. Uma solução para esse problema está na determinação das abundâncias desses elementos com mais valores nulos para as estrelas do catálogo. Assim, com mais dados, o algoritmo trabalharia com as novas informações e seria capaz de entregar probabilidades com maior precisão. Esse é um projeto que pode ser trabalhado no futuro.

TABELA 4.5. Porcentagem de estrelas, em relação ao total de cada grupo, que possuem as abundâncias de cada um dos elementos mais relevantes para a previsão de estrelas com planetas menores

Elemento	Estrelas Previstas (%)	Estrelas Não Previstas (%)
C	94,8	16,8
O	88,7	25,4
Na	99,5	50,7
Al	97,8	49,2
V	96,4	41,2

Seguindo com o raciocínio, quando levamos em consideração que o mesmo grupo de estrelas e abundâncias conseguiu ser previsto na análise dos gigantes, chegamos a uma hipótese: para as previsões dos gigantes, há uma alta dependência do elemento ferro, e, como ele está presente em todas as estrelas de nossa amostra, não afetou-se a previsão dos gigantes e ela pôde se completar para 99% dos dados. Em uma tentativa de testar essa hipótese de maneira rápida e simples, alteramos as abundâncias ausentes para o carbono e o oxigênio (elementos com menos abundâncias entre as estrelas que não estavam sendo previstas) para 0,0, de forma a seguirmos o padrão de escala da abundância solar. Assim, o código seria capaz de ler essas abundâncias e levá-las em consideração para o cálculo das previsões. Ao fim da testagem, notamos um crescimento nas estrelas previstas de

57,1% para 61,0%, o que nos mostra que a presença dos valores faz diferença nas previsões, mas não é suficiente para explicar todo o problema.

Partindo do pequeno aumento que observamos no teste anterior, suspeitamos que as estrelas com probabilidades nulas possivelmente teriam padrões de abundâncias diferentes das estrelas com planetas. Frente a isso, montamos diversos gráficos  $[X/H]$  versus  $[Y/H]$ , onde X e Y são os 5 elementos mais importantes do teste desta seção. Como exemplo, mostraremos os gráficos  $[C/H]$  versus  $[Na/H]$  e  $[C/H]$  versus  $[Al/H]$ , apresentados na Figura 4.12. Podemos notar, pelos gráficos, que a maioria das estrelas com  $p = 0$  (em azul) estão bem separadas da população de estrelas com planetas. Padrões similares são observados em outros gráficos ( $[O/H]$  vs  $[Na/H]$ ,  $[O/H]$  vs  $[Al/H]$  e  $[C/H]$  vs  $[V/H]$ ). Isso possivelmente explica o fato de que o algoritmo as previu com probabilidade 0,0. No entanto, mais testes são necessários para compreender melhor este aspecto do código.

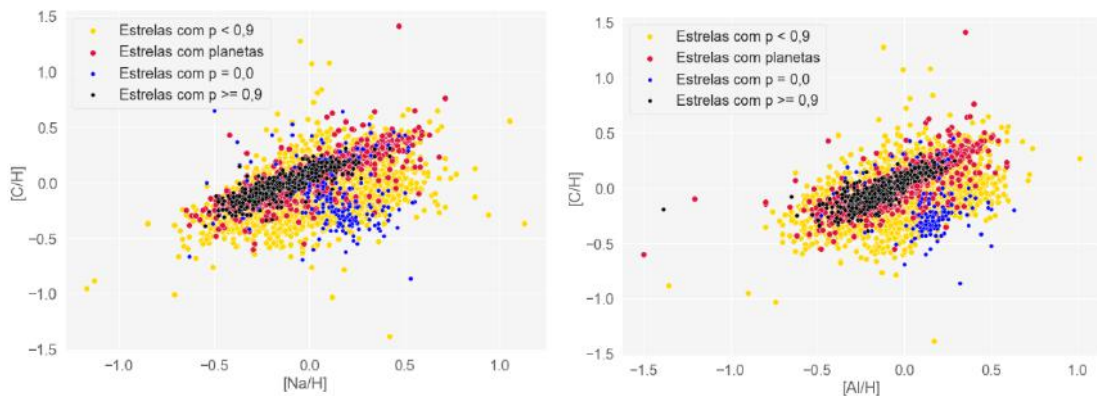


FIGURA 4.12. Gráficos de distribuição das abundâncias químicas  $[X/H]$  versus  $[Y/H]$  onde X e Y são 3 dos elementos mais importantes para o teste da Seção 4.2.1. Aqui, as estrelas são diferenciadas com respeito a serem estrelas com planetas confirmados (vermelho), estrelas sem planetas confirmados, mas que o algoritmo previu com probabilidades acima de 90% (preto), estrelas com probabilidade menor do que 90% de abrigarem planetas (amarelo) e as demais estrelas que o algoritmo previu com probabilidade igual a 0% (azul).

O teste dessa seção resultou em uma lista de estrelas previstas com probabilidades acima de 90% de abrigarem planetas menores que está no Apêndice B.

Apresentamos os gráficos, na Figura 4.13, com a distribuição de todas as estrelas do catálogo Hypatia de acordo com as abundâncias de cada elemento em relação às do ferro para o teste realizado na Seção 4.2.1.

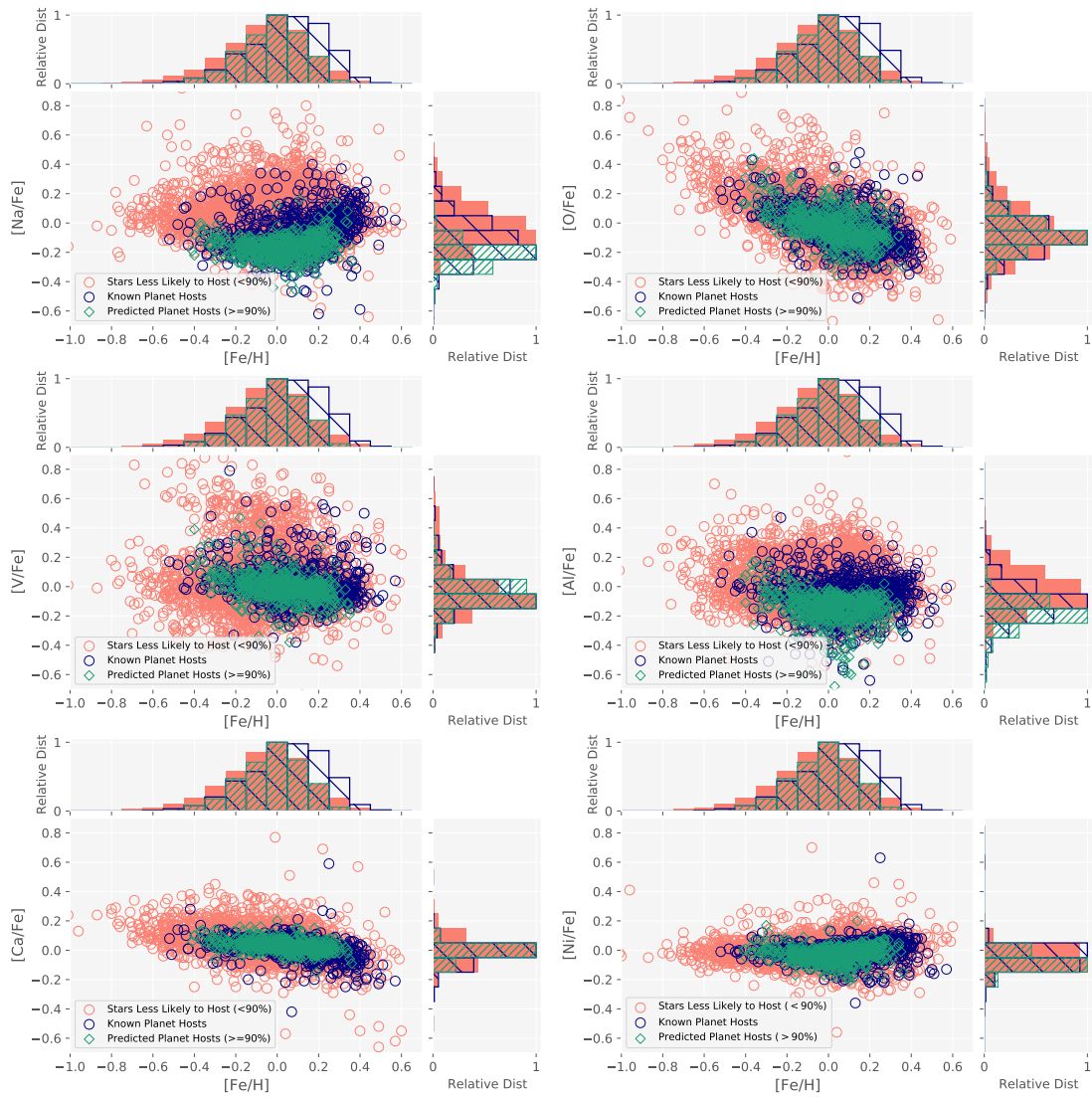


FIGURA 4.13. Gráficos de distribuição das abundâncias químicas de um elemento X ( $[X/Fe]$ ) em função da metalicidade  $[Fe/H]$ . Os painéis, símbolos e cores são similares aos da Figura 4.3. Previsão de planetas menores.

Em uma análise da comparação entre esses gráficos e os apresentados na Seção 4.1.2, notamos que, aqui, as estrelas com planetas menores previstos ocuparam um intervalo maior de metalicidades, indo até aproximadamente -0,4. Não sabemos exatamente o porquê de, nesse teste, mais estrelas terem sido previstas nessa região, mas acreditamos que isso ocorra devido a baixa dependência com o ferro para esses testes.

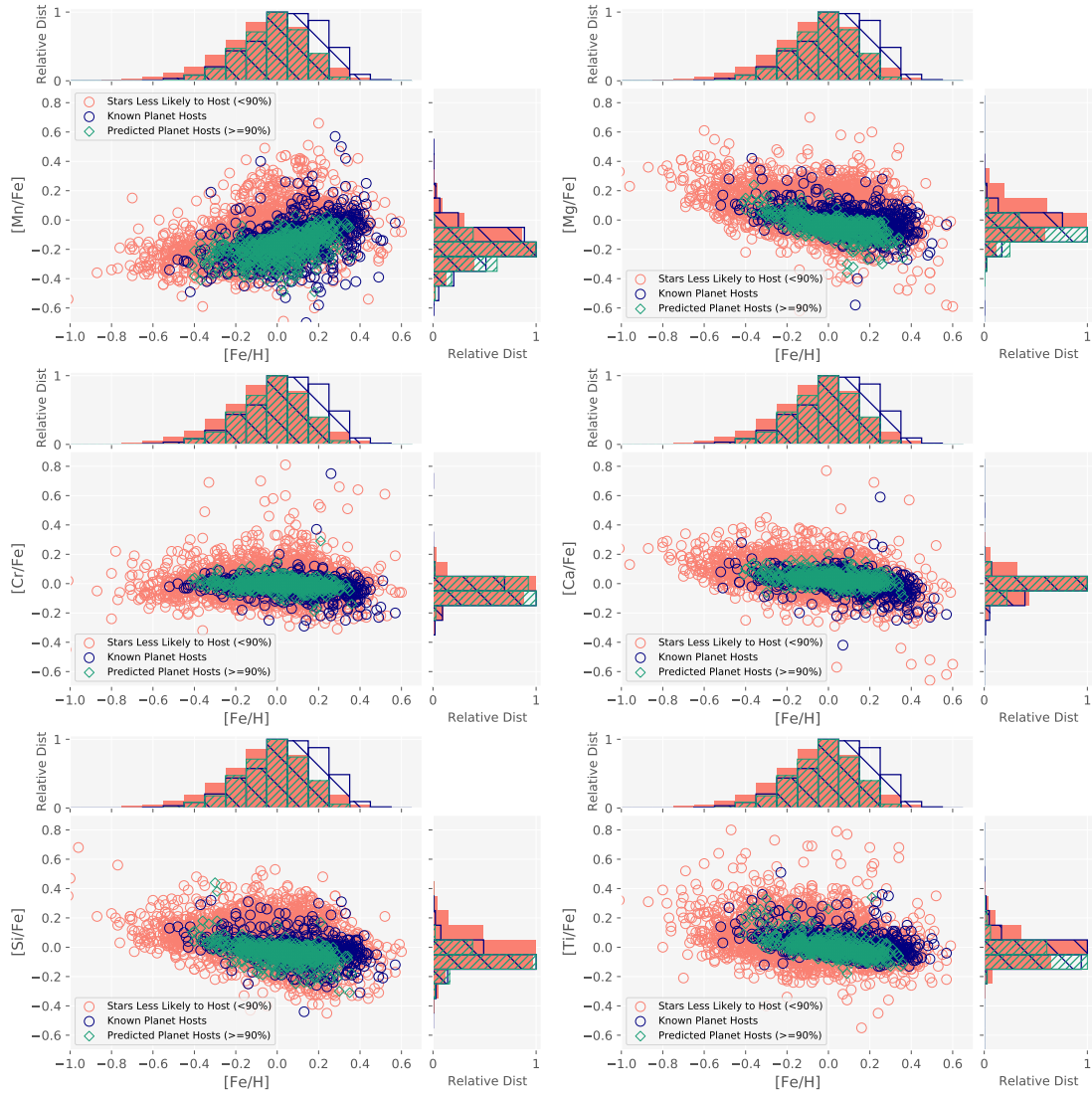


FIGURA 4.13. Continuação da Figura 4.13.

#### 4.2.2 Grupo de elementos 1 e raio planetário máximo de $4,4 R_{\oplus}$

A fim de seguir o limite máximo do raio de planetas menores estipulado em Ghezzi et al. (2021), alteramos o valor do parâmetro do raio planetário máximo para  $4,4 R_{\oplus}$ . Para esse teste, 7,9% das estrelas tiveram probabilidades  $\geq 90\%$ , com 55,6% do total sendo previstas, e a *golden set* teve probabilidade média igual a 92,2% (ver Tabela 4.4). Como era esperado, por ser uma mudança pequena de valores, não houve alteração significativa nos resultados.

### 4.2.3 Grupo de elementos 1 com variação do limite de Raio Planetário Máximo

Visando testar se a escolha do limite do raio planetário máximo (MaxPRad) influencia os resultados, fizemos 3 testes diminuindo o valor máximo do raio em  $0,5 R_{\oplus}$ , até que não houvesse estrelas suficientes para serem usadas como amostra de treinamento. Esse limite se fixou em  $2,5 R_{\oplus}$ , o que infelizmente impediu o teste exclusivamente de planetas terrestres, que, de acordo com Martinez et al. (2019), têm um limite de aproximadamente  $2,0 R_{\oplus}$ . A Tabela 4.6 mostra nossos resultados.

TABELA 4.6. Teste variando os raios planetários máximos

Valor limite de raio ( $R_{\oplus}$ )	Probabilidade $\geq 90\%$ (%)	Estrelas Previstas (%)	Elementos mais relevantes
3,5	7,9	55,1	Na, O, V, Al, C
3,0	7,8	55,0	Na, V, O, Al, C
2,5	8,0	54,5	Na, Al, V, O, C

Em análise, o estudo da variação dos valores de raio mínimo para o algoritmo procurar por planetas cada vez menores trouxe valores de probabilidade que seguiram os padrões do teste anterior. Isso pode ter ocorrido porque ainda há muitos planetas gasosos na amostra, já que, das estrelas com planetas com raios menores que  $2,5 R_{\oplus}$ , 53% têm planetas com raios maiores que  $1,6 R_{\oplus}$ , valor acima do qual a maioria dos planetas não devem ser terrestres (Rogers 2015). Dessa forma, os subnetunos ainda contaminam a amostra quando usamos apenas o limite superior de raio e, assim, não conseguimos ver uma possível influência dos planetas terrestres nos padrões de abundâncias. Além disso, os comportamentos observados na literatura (e.g., Petigura et al. 2018; Ghezzi et al. 2021; Wilson et al. 2022) para as classes de subnetunos e superterras não são radicalmente diferentes, o que poderia explicar os resultados. Para esses testes, também é mantida a lista de elementos relevantes com apenas algumas alterações de posição, quando comparada ao teste da Seção 4.2.1. Essas variações podem ser vistas na Figura 4.14.

Realizamos também um teste adicional selecionando, para a amostra de treinamento, estrelas com  $\text{MaxPRad} = 1,9 - 4,0 R_{\oplus}$ , para estudarmos o comportamento exclusivamente dos planetas do tipo subnetuno (Martinez et al. 2019).

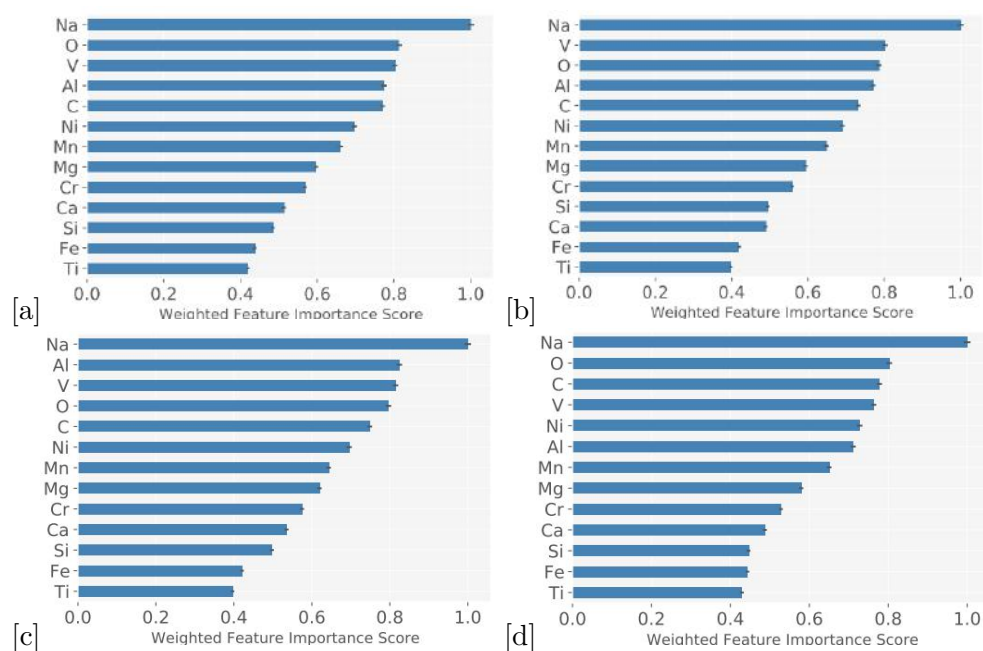


FIGURA 4.14. Gráficos de relevância dos elementos para os testes com planetas menores e variação do raio planetário máximo. [a]: 3,5 R<sub>⊕</sub>; [b]: 3,0 R<sub>⊕</sub>; [c]: 2,5 R<sub>⊕</sub>; [d]: 1,9 R<sub>⊕</sub> a 4,0 R<sub>⊕</sub>.

Aqui, 8,3% das estrelas tiveram probabilidades acima de 90% de abrigarem esses planetas, 54,5% obtiveram resultados de previsão e os elementos Na, O, C, V e Ni foram os mais relevantes (Figura 4.14 [d]). Dessa forma podemos concluir que, eliminando as superterras da amostra, o resultado fica parecido com os anteriores, o que reforça a suspeita de que as estrelas com subnetunos estão ditando os padrões de abundâncias.

#### 4.2.4 Outros grupos de elementos

Seguindo os padrões dos testes com gigantes, fizemos também um estudo das previsões de planetas menores para cada um dos grupos de elementos, dispostos na Tabela 3.3. Os resultados para esses testes estão dispostos na Tabela 4.7 e na Figura 4.15.

Analisando os resultados dos testes, encontramos valores muito similares àqueles resultantes do teste com o grupo 1 de elementos. Para todos os grupos, a porcentagem de estrelas com resultado de previsão e a média das probabilidades da *golden set* se mantiveram próximas de 60% e 90%, respectivamente. Os testes



TABELA 4.7. Teste variando os grupos de elementos (Planetas Menores)

Grupo de Elementos	Probabilidade $\geq 90\%$ (%)	Estrelas Previstas (%)	Probabilidade Média <i>golden set</i> (%)	Elementos mais relevantes
2	7,9	55,1	92,8	Na, O, C, V, Al
3	8,2	65,5	89,5	Na, Al, Ni, V, Mg
4	8,3	54,3	90,8	Na, V, C, O, Al
5	8,0	65,8	89,3	Na, Ni, Al, V, Mn
6	15,8	62,2	87,5	Fe, C, O
7	15,7	60,9	87,4	C, Ni, Cr, O, Fe

com menor número de estrelas previstas, foram os testes com os grupos 2 e 4 de elementos. No entanto, a queda não é muito brusca, como ocorreu na Seção 4.1.4.

A respeito da ordem dos elementos, notamos que o sódio predomina em relevância em todos os testes que está envolvido, assim como no teste com o grupo 1 de elementos (ver Figura 4.15). O alumínio, o vanádio e os voláteis (C e O) também se mantêm altos na ordem de importância quando estão presentes, no entanto, quando apenas C, O e Fe estão disponíveis para análise (ver Figura 4.15 [e]), o ferro passa a ser o elemento mais relevante para a previsão de planetas menores, um resultado intrigante, quando vemos que em todos os outros testes sempre aparece abaixo de C e O. Um outro elemento interessante que subiu algumas posições quando os voláteis não estavam presentes foi o níquel. Desses elementos, a literatura apontou poucos como relevantes para a formação de planetas menores na literatura (e.g., Wilson et al. 2022).

#### 4.2.5 Grupo de elementos 1 e inclusão do parâmetro Massa Planetária Máxima

Seguindo o padrão de testes, tentamos prever planetas pequenos usando o parâmetro do raio máximo *ou* da massa planetária máxima, com ambas as condições disponíveis para a previsão. Fizemos isso acrescentando e invertendo o limite da massa usado nas seções de testes com planetas gigantes para menor que  $0,0945 M_J$  à lista de condições, além de  $\text{MaxPRad} \leq 4,0 R_{\oplus}$ . Aqui, 10,0% das estrelas tiveram probabilidades altas ( $p \geq 0,9$ ) de abrigarem planetas menores, e a *golden set* manteve probabilidade média alta, com 88,9% (ver Tabela 4.4). Notamos que, nesse teste, o algoritmo previu 69,7% das estrelas, indicando que,

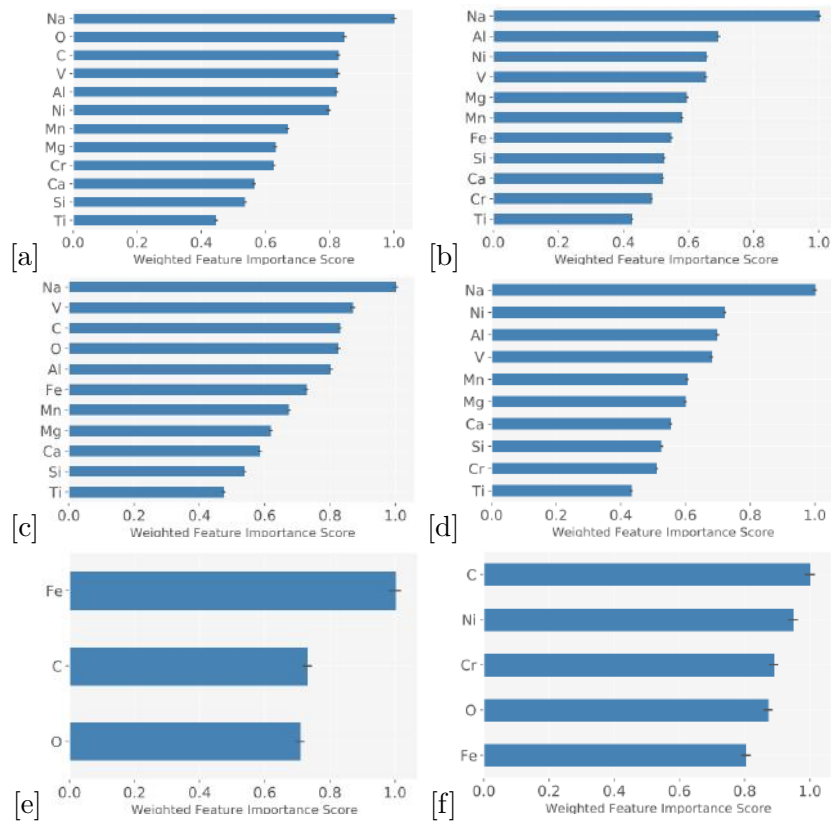


FIGURA 4.15. Gráficos de relevância dos elementos para os testes com planetas menores e a variação de grupos de elementos. [a]: Grupo 2; [b]: Grupo 3; [c]: Grupo 4; [d]: Grupo 5; [e]: Grupo 6; [f]: Grupo 7.

com a inclusão da massa planetária máxima, ele conseguiu prever novas estrelas. Podemos notar na Figura 4.16, a faixa de estrelas que adicionamos para teste quando acrescentamos o parâmetro  $\text{MaxPMass}$  na análise. Algumas dessas estão fora dos intervalos ocupados pelas que possuem  $\text{MaxPRad} \leq 4,0 R_{\oplus}$ , o que pode ser um indicativo do aumento da porcentagem de estrelas previstas.

Os elementos mais relevantes, no entanto, permaneceram similares aos do teste da Seção 4.2.1, com mínima alteração na ordem de importância, como pode ser visto na Figura 4.17. Fizemos, também, uma comparação com os resultados do teste da Seção 4.2.1, e o código previu 80,7% das estrelas em ambos os testes com probabilidades acima de 90%, o que representa um bom acordo.

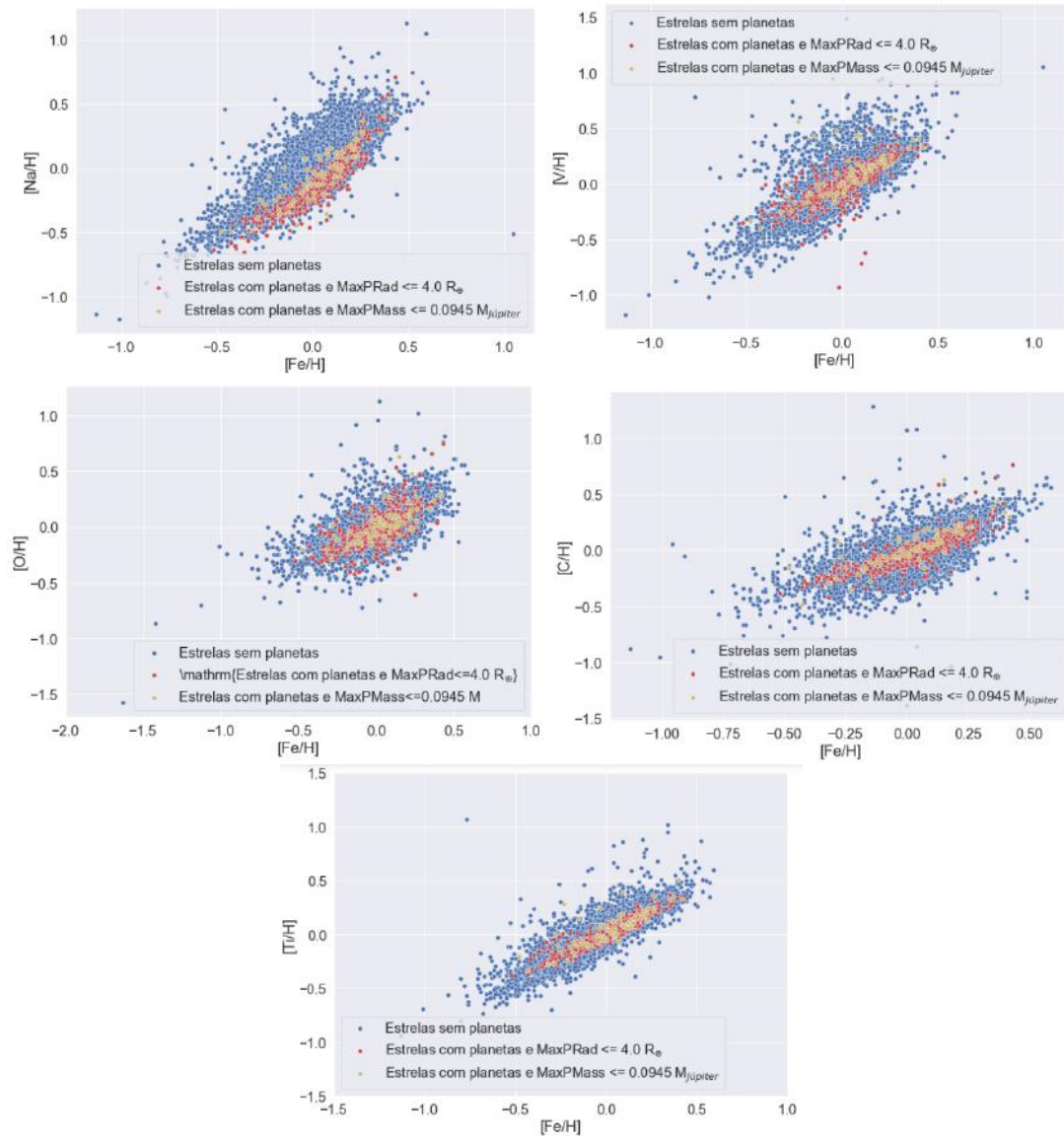


FIGURA 4.16. Gráficos  $[X/H]$  versus  $[Fe/H]$ . Aqui apresentamos as estrelas sem planetas confirmados na cor azul, as com planetas e  $\text{MaxPRad} \leq 4,0 R_{\oplus}$  em vermelho e as estrelas adicionadas ao teste, com  $\text{MaxPMass} \leq 0,0945 M_J$  em amarelo. Apresentamos essa distribuição para os 5 elementos mais relevantes para a previsão da Seção 4.2.5.

### 4.3 Estudo da Previsão do Sol

Realizamos, também, um pequeno estudo das previsões para o Sol, ao inserirmos uma estrela nos dados com valores de abundância iguais a 0,0 para todos os elementos. Isso se deve por causa da definição de  $[X/H] = \log \epsilon(X)_* - \log \epsilon(X)_{\odot}$ . Assim, se uma estrela tem  $[Fe/H]=0$ , por exemplo, ela tem a mesma abundância de ferro que o Sol. Com isso, usando abundâncias iguais a zero para todos os elementos, replicaríamos uma estrela idêntica a ele na amostra geral e poderíamos, assim, testar se o algoritmo conseguiria prever os planetas de nosso Sistema Solar.

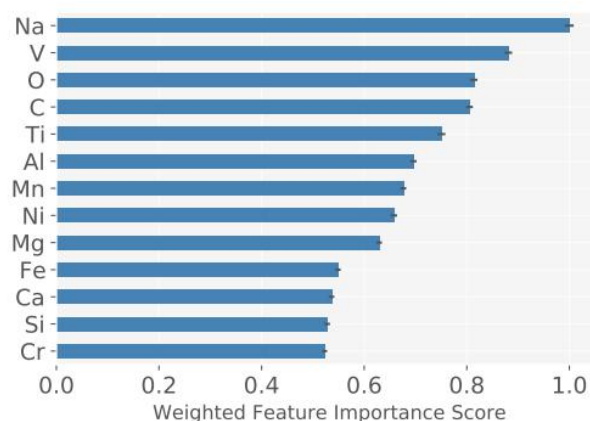


FIGURA 4.17. Gráfico de relevância dos elementos para o teste com planetas menores e os parâmetros massa e raio planetário máximos.

Fizemos essa análise para o teste da Seção 4.1.2 com planetas gigantes e para o da Seção 4.2.1 com planetas menores. No primeiro, o algoritmo previu o Sol com probabilidade 45,6%, enquanto no segundo, ele teve uma probabilidade de 33,6%. Percebemos que o código não retornou previsões altas para o Sol, apesar de confirmadamente ele hospedar planetas gigantes e menores.

Os motivos para essas probabilidades baixas giram em torno dos tipos de exoplanetas que temos descobertos hoje em dia. O período orbital da maioria dos planetas descobertos fora do Sistema Solar é diferente dos que temos por aqui. Júpiteres quentes, subnetunos e superterras com períodos curtos compõem a maioria dos exoplanetas de nossos dados e a baixa previsão deles para o nosso Sol é coerente. Além disso, o tamanho dos planetas também é um fator limitante. O Sistema Solar, por exemplo, não possui superterras, subnetunos ou planetas maiores que Júpiter (os quais dominam a amostra de gigantes; ver Figura 2.2), independentemente do período orbital. À medida que a tecnologia de detecção planetária evoluir e planetas mais parecidos com os nossos forem descobertos, poderemos criar um grupo de estrelas grande o suficiente para montar uma amostra de treinamento e, assim, prevermos mais sistemas parecidos com o Solar.

## Capítulo 5

# Conclusões e perspectivas futuras

As possíveis conexões entre as abundâncias químicas das estrelas e a presença de planetas de diferentes tipos ainda estão em aberto, com a exceção da alta metalicidade para estrelas de tipo solar com planetas gigantes ao seu redor. O objetivo de nosso projeto era investigar estas possíveis correlações para diversos elementos além do ferro e, também, para classes de planetas diferentes.

Para atingirmos esse objetivo, utilizamos as estrelas disponíveis no catálogo Hypatia (Hinkel et al. 2014) em sua versão atualizada, juntamente com acréscimos e cortes que realizamos, e previmos uma lista de estrelas com as probabilidades de abrigarem planetas de classes diferentes. Fizemos essa previsão através do algoritmo *planetPrediction* (Hinkel et al. 2019) que usa a técnica de aprendizado de máquina para prever, a partir das abundâncias químicas de uma lista de estrelas, planetas gigantes. Manipulamos esse código para que pudéssemos usar novos parâmetros além da massa planetária máxima, a fim de prever planetas menores.

Para os primeiros testes, investigamos novos padrões de abundância relevantes para a previsão de planetas gigantes e demos suporte às pesquisas que listam o ferro como um elemento associado à presença desses planetas. Ao final das previsões, encontramos uma lista mais ampla de possíveis elementos relevantes para a formação dessa classe planetária como o oxigênio, o titânio, o alumínio e o carbono. Além disso, quando removemos o ferro das abundâncias disponíveis para o algoritmo encontrar um padrão de previsões, o titânio se tornou o elemento mais

relevante. Já a lista de cinco elementos menos importantes para a previsão dessa classe planetária inclui o cobalto, o níquel, o ítrio, o cromo e o silício. Em testes adicionais, notamos que os limites de MaxPMass não alteram os resultados de maneira significativa.

No caso dos planetas menores, incluímos o parâmetro do raio planetário máximo no algoritmo e os testes nos entregaram resultados interessantes, como a baixa importância do ferro para a previsão desses planetas, a qual a literatura já havia sugerido (e.g., Buchhave & Latham 2015), e o surgimento de novos elementos importantes. Dentre eles, temos o sódio, o oxigênio, o vanádio, o alumínio e o carbono. Para os dois últimos, há referências na literatura que sustentam sua relevância para a presença de planetas menores (e.g., Wilson et al. 2022), no entanto, para o outro grupo, esses resultados aparentam ser inéditos, pois não encontramos referências na literatura que os listem. A lista dos cinco elementos menos relevantes para a previsão dessa classe planetária inclui o titânio, o ferro, o silício, o cálcio e o cromo. Em testes adicionais, notamos que a variação dos limites de MaxPRad não afeta significativamente os resultados.

Realizamos, também, um teste de previsão dos planetas do Sistema Solar, ao incluirmos uma estrela com parâmetros idênticos ao Sol nos dados. Nesse teste, percebemos que o código não retornou previsões altas quando procuramos por planetas gigantes ou menores, apesar de confirmadamente o Sol hospedar ambas as classes planetárias. Para nós, o possível motivo para as baixas probabilidades gira em torno dos tipos de exoplanetas que temos descobertos hoje em dia, com tamanhos e períodos orbitais diferentes dos que temos no Sistema Solar.

Como resultado geral, notamos que o carbono e o oxigênio foram elementos com alta relevância para a previsão não só de planetas gigantes, como também de menores. Essa consequência é interessante por ter sido corroborada em diversos testes que realizamos, variando limites de parâmetros e os grupos de elementos usados pelo algoritmo para a previsão. Além dessa análise construímos, também, listas de possíveis alvos para buscas de exoplanetas. Essas constituem-se das estrelas e suas probabilidades acima de 90% que encontramos nos testes para a previsão de planetas gigantes e menores.

Para o futuro, como já apresentamos em capítulos anteriores, o catálogo Hypatia está sempre em atualização, com novos valores de abundâncias e cada vez mais estrelas com planetas confirmados sendo adicionados à tabela. Com esses novos dados, a nossa pesquisa pode se tornar mais acurada e precisa, sendo possível complementá-la a procura de mais estrelas e uma melhor determinação dos parâmetros que estão associados à presença ou não de planetas de variadas classes, inclusive os do Sistema Solar. A aplicação do algoritmo em outros catálogos também é estimulada, de forma que possamos obter mais estrelas com valores de previsão. Além disso, incentivamos a utilização do algoritmo com nossas alterações para realizar esses estudos e instigamos a alteração dos parâmetros do XGBoost para testar se é possível melhorar ainda mais a qualidade das previsões e/ou o tempo de execução do código.

A determinação futura de abundâncias faltantes no catálogo Hypatia também contribuiria significativamente para nossa pesquisa, pois com todos os novos valores inseridos na tabela usada pelo algoritmo, os padrões de abundância poderiam ser encontrados com incertezas muito menores e a eficácia do código aumentaria, nos dando elementos relevantes para cada estudo sem um viés gerado pela escassez de dados.

Esperamos que essas e outras maneiras de contribuir com a comunidade científica que propomos sejam aproveitadas e acreditamos que muitas outras descobertas e estudos possam ser feitos a partir de nossos resultados. Incentivamos essa busca e futuros projetos derivados deste podem e devem ser construídos.

# Referências Bibliográficas

- Adibekyan, V. Z., Sousa, S., Santos, N., et al. 2012a, *Astronomy & Astrophysics*, 545, A32
- Adibekyan, V. Z., Santos, N., Sousa, S., et al. 2012b, *Astronomy & Astrophysics*, 543, A89
- Anders, E., & Grevesse, N. 1989, *Geochimica et Cosmochimica acta*, 53, 197
- Anderson, E., & Francis, C. 2012, *Astronomy Letters*, 38, 331
- Bensby, T., Feltzing, S., & Lundström, I. 2003, *Astronomy & Astrophysics*, 410, 527
- Brewer, J. M., & Fischer, D. A. 2018, *The Astrophysical Journal Supplement Series*, 237, 38
- Buchhave, L. A., & Latham, D. W. 2015, *The Astrophysical Journal*, 808, 187
- Buchhave, L. A., Bizzarro, M., Latham, D. W., et al. 2014, *Nature*, 509, 593
- Carlos, M., Nissen, P. E., & Meléndez, J. 2016, *Astronomy & Astrophysics*, 587, A100
- Chen, T., & Guestrin, C. 2016, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794
- Da Silva, R., Milone, A. d. C., & Rocha-Pinto, H. J. 2015, *Astronomy & Astrophysics*, 580, A24
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *Astronomy and Astrophysics*, 275, 101
- Figueira, P., Faria, J., Delgado-Mena, E., et al. 2014, *Astronomy & Astrophysics*, 570, A21



- Fischer, D. A., & Valenti, J. 2005, *The Astrophysical Journal*, 622, 1102
- Gaia Collaboration, C., Brown, A., Vallenari, A., et al. 2018, *Astronomy & Astrophysics*, 616
- Ghezzi, L., Cunha, K., Smith, V., et al. 2010a, *The Astrophysical Journal*, 720, 1290
- Ghezzi, L., Cunha, K., Smith, V., & de La Reza, R. 2010b, *The Astrophysical Journal*, 724, 154
- Ghezzi, L., Martinez, C. F., Wilson, R. F., et al. 2021, *The Astrophysical Journal*, 920, 19
- Ghezzi, L., Montet, B. T., & Johnson, J. A. 2018, *The Astrophysical Journal*, 860, 109
- Gonzalez, G. 1997, *Monthly Notices of the Royal Astronomical Society*, 285, 403
- Gustafsson, B. 1998, *Solar Composition and its Evolution—from Core to Corona*, 419
- Hinkel, N. R., Timmes, F., Young, P. A., Pagano, M. D., & Turnbull, M. C. 2014, *The Astronomical Journal*, 148, 54
- Hinkel, N. R., Unterborn, C., Kane, S. R., Somers, G., & Galvez, R. 2019, *The Astrophysical Journal*, 880, 49
- Laws, C., Gonzalez, G., Walker, K. M., et al. 2003, *The Astronomical Journal*, 125, 2664
- Lodders, K., Palme, H., & Gail, H. 2009, *JE Trümper*, 4, 44
- Martinez, C. F., Cunha, K., Ghezzi, L., & Smith, V. V. 2019, *The Astrophysical Journal*, 875, 29
- Mayor, M., & Queloz, D. 1995, *Nature*, 378, 355
- Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv preprint arXiv:1109.2497
- Meléndez, J., Asplund, M., Gustafsson, B., & Yong, D. 2009, *The Astrophysical Journal*, 704, L66
- Payne, C. H. 1925
- Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, *The Astronomical Journal*, 155, 89

- 
- Robinson, S. E., Laughlin, G., Bodenheimer, P., & Fischer, D. 2006, *The Astrophysical Journal*, 643, 484
- Rogers, L. A. 2015, *The Astrophysical Journal*, 801, 41
- Sousa, S., Santos, N., Mayor, M., et al. 2008, *Astronomy & Astrophysics*, 487, 373
- Wang, J., & Fischer, D. A. 2015, *The Astronomical Journal*, 149, 14
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *Astronomy and Astrophysics Supplement Series*, 143, 9, doi: 10.1051/aas:2000332
- Wilson, R. F., Cañas, C. I., Majewski, S. R., et al. 2022, *The Astronomical Journal*, 163, 128

# Apêndice A

## Lista de estrelas com probabilidades altas de abrigarem *planetas gigantes*

Este apêndice contém uma listagem de estrelas e suas probabilidades acima de 90% de abrigarem planetas gigantes. Aqui, apresentamos as probabilidades obtidas a partir do teste usando a massa máxima planetária (Seção 4.1.2).

TABELA A.1. Estrelas e Probabilidades de abrigarem planetas gigantes, Teste 4.1.2

Estrela	Probabilidade
HIP 44137	0.9997850511330284
HIP 42214	0.999718917667678
HIP 45982	0.9995855781185248
HIP 22576	0.9994952796233628
HIP 6712	0.9994628409926698
HIP 1931	0.9993818256748404
HIP 20489	0.9991565089393514
HIP 117427	0.9990079611120756
HIP 60081	0.9989652317880796
HIP 7080	0.9989011170599512

TABELA A.1. Continuação

Estrela	Probabilidade
Gaia DR2 2312679845530628096	0.998766193707588
HIP 83983	0.9987590486039296
HIP 40761	0.9986689594735364
HIP 61379	0.9986094787200582
HIP 8525	0.998586297836457
HIP 92922	0.9983876702248168
HIP 14810	0.9983440281515216
Gaia DR2 4439085988769170432	0.9983419689119172
HIP 62345	0.9983225774465166
HIP 919	0.9982712556969984
Gaia DR2 1291120362349158016	0.9981416477390048
HIP 109169	0.9981404190255796
HIP 30860	0.9979332437738968
HIP 95849	0.9978919696443628
HIP 105606	0.9978268773135908
Gaia DR2 1485265047475984256	0.9977920381059492
HIP 101399	0.997751991801382
HIP 17960	0.9977310231023102
HIP 26973	0.9975591593579348
HIP 45983	0.997500786176534
KOI-4409	0.9974207628589852
HIP 109836	0.997386204785892
HIP 30243	0.997288045904784
HIP 58576	0.99724800211565
HIP 27384	0.9972156337011446
HIP 86974	0.9970981423150376
HIP 105483	0.9970318804153714
HIP 47080	0.996981125833292
HIP 42011	0.9969761562484508
HIP 60096	0.9969514965755972
HIP 87710	0.9969478150822602

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 24176	0.996898904279512
HIP 114638	0.9968919455098532
HIP 16405	0.9968828806482284
HIP 107040	0.9968430534780128
HIP 51258	0.9968343967996828
HIP 55846	0.9967987426586152
HIP 76641	0.9967168918807164
HIP 74500	0.996690454028338
HIP 113137	0.9964926758819888
HIP 5189	0.9963932663275014
HIP 84636	0.9963110923632996
HIP 65808	0.9963001027749228
HIP 95740	0.9962794543199668
Gaia DR2 1298310687356960896	0.9962394518691164
HIP 90896	0.996163989450971
Gaia DR2 359058441314838528	0.996067880794702
HIP 27706	0.9958575184799324
HIP 15005	0.9958265152642104
HIP 81347	0.9957939098458868
Gaia DR2 2076615077014515968	0.9957401050498368
HIP 42446	0.9956656346749226
HIP 95319	0.9955838205741032
HIP 51078	0.9955752944290064
HIP 55486	0.9955468102734052
KOI-17	0.9954629820581564
HIP 77641	0.995427182218106
Gaia DR2 2100417167287796992	0.995367651316497
KOI-4129	0.9952868411913542
HIP 80484	0.9952182952182952
HIP 23286	0.995202527771574
KOI-148	0.995060712080675

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 58725	0.9949571356530508
HIP 113948	0.9949311194535864
HIP 103682	0.9947895559543796
HIP 41022	0.994786847323199
HIP 104809	0.9947626218331652
HIP 10531	0.9946867847197548
HIP 11924	0.9945032691083724
Gaia DR2 2101493039418001536	0.994475138121547
HIP 87601	0.9944352844187964
HIP 30905	0.9944145635084816
HIP 69881	0.9943816970750072
HIP 104318	0.9943354255046432
HIP 83906	0.994333598041162
HIP 116217	0.994242143872055
HIP 13908	0.994239478991041
HIP 36272	0.9940783379648008
Gaia DR2 2132757996148004736	0.9939925030409852
HIP 51949	0.993949163050217
HIP 75722	0.9937838396362884
HIP 105066	0.993503380556428
HIP 682	0.9933082426899376
HIP 19232	0.9932863154935632
HIP 110028	0.9932681347364248
HIP 93889	0.9932264227406936
HIP 31246	0.9932028836251288
HIP 78955	0.9931963791179268
HIP 77301	0.9931902600082544
HIP 107397	0.9931684724175006
Gaia DR2 2116898606069151104	0.993143547159824
HIP 47463	0.9930805585214246
HIP 20800	0.9930526838144074

TABELA A.1. Continuação

Estrela	Probabilidade
KOI-135	0.9929460580912864
KOI-2581	0.9928740896277496
Gaia DR2 2129016564238409728	0.9928309188331018
HIP 21731	0.9927264001059182
Gaia DR2 111322601672419712	0.9925956396544632
HIP 59525	0.992570063720588
HIP 93607	0.992527134602508
KOI-1474	0.9923458833264376
HIP 76704	0.992259985115356
HIP 9353	0.992226522443849
HIP 49680	0.9921759986767016
HIP 16727	0.992136401597526
HIP 118319	0.992135761589404
HIP 91949	0.9921066754737232
Gaia DR2 2132546305800554112	0.992022156084656
HIP 41479	0.9919672740795836
HIP 54195	0.9919371511267314
Gaia DR2 2077773790469159680	0.9919312879382252
HIP 54381	0.991590245677287
HIP 97769	0.9915481344052772
HIP 105522	0.9915358711930036
Gaia DR1 2105185470039774336	0.991472342957246
HIP 15310	0.9914562400846112
HIP 27090	0.9913387713948048
HIP 74653	0.9912837815487412
HIP 112219	0.9912639189927034
HIP 68461	0.9912058638604851
HIP 111978	0.9912055000785962
HIP 41226	0.9911743054004648
HIP 83181	0.9911136461329896
HIP 60094	0.9911017755100352

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 102815	0.9909554960848448
Gaia DR2 3435282862461427072	0.990916597853014
HIP 17486	0.990915027817044
KOI-456	0.9908209570957096
HIP 92742	0.9907940731681114
HIP 12198	0.9906729233392092
HIP 83541	0.990655209452202
HIP 99240	0.9905309602895246
HIP 61595	0.990503715937242
HIP 184	0.9904980855585784
Gaia DR2 2100787084231447424	0.9903843768861256
HIP 39242	0.9903365655094816
HIP 113026	0.9903231409264972
Gaia DR2 5574766468155514752	0.990313272877164
HIP 108961	0.9902848402166274
HIP 107822	0.9902795457175448
HIP 22729	0.9899257468771472
HIP 34879	0.9898757174119492
HIP 83458	0.9897580144777662
HIP 84028	0.9895692134757166
HIP 13642	0.9895378817929392
Gaia DR2 3798552815560689792	0.9894736842105264
HIP 81022	0.9894519131334024
HIP 110716	0.989357306935929
HIP 81767	0.9892714859570528
HIP 58079	0.9890208009524124
KOI-85	0.9889474227869022
Gaia DR2 2666015878575546496	0.9888175605715468
HIP 92961	0.9886742226961964
HIP 75535	0.9886382416129564
Gaia DR2 4534144923690481408	0.9886340152924158



TABELA A.1. Continuação

Estrela	Probabilidade
HIP 34877	0.9884624927632124
Gaia DR2 5086537022856406272	0.988374506954536
HIP 38216	0.9882023446600738
HIP 86882	0.9880834160873884
Gaia DR2 4596481632339898240	0.987946799667498
HIP 49266	0.9878958864215508
Gaia DR2 1909762228985058944	0.9878725590955808
HIP 69972	0.987864759857816
HIP 71803	0.9878394811511958
HIP 16581	0.9878369005152596
HIP 87691	0.9877999354844046
HIP 18806	0.9876711536395713
HIP 53647	0.9875793335977786
HIP 59110	0.9875118885167268
HIP 42356	0.987463431565377
HIP 59968	0.9873145737779432
HIP 100474	0.9872146647052016
HIP 99702	0.9870357364180954
HIP 87272	0.986977013395072
HIP 5891	0.9869001768682744
HIP 22429	0.9868541882109616
Gaia DR2 2076223101122787328	0.9868463478153032
HIP 98589	0.9868115522869064
HIP 88790	0.986769661293394
HIP 59572	0.9866826485905597
HIP 71683	0.9866758196314552
HIP 25094	0.986598696223344
HIP 43177	0.9865424430641822
KOI-3083	0.9865138791213336
Gaia DR2 2052184512764630272	0.9864490578829443
HIP 62769	0.9863172254144104

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 7276	0.9862843820531264
KOI-3184	0.986128343838404
HIP 2742	0.9860795196069512
Gaia DR2 2131260667469179904	0.9860767965890216
HIP 6197	0.985616444014838
HIP 726	0.9856023538746363
HIP 58656	0.9854460753136692
HIP 4752	0.9853208582669356
Gaia DR2 135197672034753920	0.985162763001191
HIP 24110	0.9851294351170292
HIP 113207	0.9851012667641736
Gaia DR2 377245055126435072	0.9849860537489344
HIP 117184	0.9848263208786848
HIP 3533	0.9847843903408198
HIP 20741	0.9847421623810824
HIP 26779	0.9846049541951912
Gaia DR2 2133226319381976832	0.9845804051260852
HIP 70027	0.9844849106043198
HIP 70608	0.9844729745379944
HIP 110161	0.9842341783151676
HIP 1306	0.983928733901852
HIP 99174	0.9838969670416876
HIP 112426	0.983875233482652
HIP 53707	0.9834968482702712
HIP 48423	0.9834531005796788
Gaia DR2 4756649415309914240	0.9834402815152142
HIP 51579	0.9834352427208566
HIP 107001	0.9833489118476668
HIP 81229	0.9833250620347396
HIP 31540	0.9832575444398512
Gaia DR2 2133630011947261824	0.9830382322050298

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 12894	0.9824253947259652
HIP 93297	0.9821933903718536
Gaia DR2 4177560894633502720	0.9821835508597472
HIP 84171	0.9821687746216048
HIP 39039	0.982085674331626
HIP 35449	0.9820154884993346
Gaia DR2 2077247876015217920	0.9819860782725154
HIP 25166	0.9819172509591216
HIP 43634	0.9819024370060822
HIP 68805	0.9818707890712188
Gaia DR2 2119828460957942528	0.9816129032258064
HIP 63147	0.9814344875472196
KOI-3225	0.98140849260058
HIP 37309	0.9812827593048828
HIP 56907	0.9811213145648562
HIP 114244	0.9810461406206088
HIP 101966	0.981023102310231
KOI-1947	0.9808692345213588
HIP 23884	0.9804938783980078
HIP 93547	0.9804245887675016
HIP 88217	0.9803589569928886
HIP 116422	0.9792026726426252
HIP 106913	0.9791921462334314
HIP 77655	0.978960396039604
HIP 11941	0.9788631063625688
HIP 74118	0.9788111772486772
HIP 15669	0.9785012945545988
HIP 19633	0.9784301085932464
HIP 7562	0.978375581996896
HIP 117902	0.9783476442732824
KOI-246	0.9783326454808088

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 57090	0.9782522474881016
Gaia DR2 2133335823867642752	0.978181607731929
KOI-102	0.97777704289163
HIP 10599	0.97758125760045
HIP 116559	0.9775444278290468
HIP 55076	0.977500682173363
HIP 10990	0.9774576088845488
HIP 89478	0.9767858765603444
HIP 64475	0.976727651518283
HIP 90485	0.9765238879736408
KOI-288	0.9764566857558404
HIP 62886	0.9764343161322852
HIP 2156	0.976419813309742
HIP 99661	0.976216269857678
HIP 61597	0.9760911066850688
HIP 45406	0.9760033098882912
HIP 12158	0.9759541416447466
Gaia DR2 2605161444735454464	0.9758862324814508
HIP 57221	0.9758070519341951
HIP 60753	0.9757668863670153
HIP 60098	0.9757429997685874
HIP 81269	0.9756988039384128
HIP 109162	0.9756955072259954
HIP 23311	0.9755643760853386
HIP 116984	0.975496475294001
HIP 54406	0.9752437613617584
HIP 33690	0.9750142622096916
HIP 83398	0.9748489532106224
HIP 36866	0.9747598043677952
HIP 84132	0.9745101769069812
HIP 1803	0.9744640761615126

TABELA A.1. Continuação

Estrela	Probabilidade
Gaia DR2 2082282818938558848	0.97444588662274
HIP 30377	0.9744281059265264
HIP 79585	0.9742862576169729
Gaia DR2 1999550886837136640	0.9742300439974858
HIP 52733	0.9737901145392632
Gaia DR2 2128409045404259328	0.9734741648125326
HIP 7245	0.9731848184818482
HIP 81952	0.9731450399351776
HIP 57148	0.973092020990868
KOI-367	0.972945064023131
Gaia DR2 2085242360641759616	0.9729264863970892
Gaia DR2 4573892028507342336	0.9729017222571156
HIP 19921	0.9724580658521432
HIP 5176	0.972309727562115
Gaia DR2 2101918516057137024	0.9722766384804226
Gaia DR2 1014520826353577088	0.9720439014288672
Gaia DR2 2119602202081351168	0.9719182325005162
HIP 42634	0.9715586927413088
HIP 9960	0.9715405238193442
Gaia DR2 6237190612936085120	0.9714864250588868
Gaia DR2 2086570708126781312	0.9714226633581472
HIP 2902	0.9713241495642396
HIP 107246	0.9712034750724662
HIP 21654	0.9710473417323616
HIP 65429	0.9710270717090308
Gaia DR2 736173925863826944	0.970928598588173
HIP 56572	0.9709200041395012
HIP 63995	0.970629081590477
HIP 81375	0.9705034009107664
Gaia DR2 2126992363331351296	0.9701285283299582
HIP 42532	0.9698548136389192

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 78408	0.9698310073417554
HIP 65356	0.9695158249826372
Gaia DR2 4453211899986180352	0.9693264248704664
HIP 7539	0.9691603002943414
HIP 42074	0.9691477404211708
Gaia DR2 2007417244005768064	0.968408865365531
HIP 10492	0.9683008319687724
HIP 14113	0.968163616507264
HIP 17778	0.9678968864196306
HIP 85017	0.967665086731022
HIP 94751	0.9673249580824476
HIP 114420	0.9672564908218952
HIP 75253	0.9672426473626364
HIP 96561	0.9670187095068096
Gaia DR2 2101041513795585024	0.9670157111333338
Gaia DR2 1961681064973138176	0.9669576575236016
HIP 70616	0.9668603738024188
HIP 32892	0.966856859340114
HIP 64408	0.9665323830182831
HIP 19793	0.9662743865626836
HIP 115125	0.9656572841793564
HIP 24681	0.9656378600823046
HIP 51237	0.9654958421248604
Gaia DR2 2007417175286304000	0.9653406462697416
HIP 39102	0.9653258099745176
HIP 77358	0.9653244680411176
HIP 61743	0.9651250258317834
HIP 10218	0.9650317855282846
HIP 53084	0.9648940126441056
KOI-2706	0.964796415995768
HIP 16107	0.9647667793975376

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 53818	0.9647280604567328
HIP 67275	0.9647204456364762
HIP 96501	0.964469440079398
HIP 58983	0.9643769690655156
Gaia DR2 1424011082893734272	0.9643671017194944
HIP 100649	0.964342123760616
HIP 8587	0.9642085750461358
HIP 8830	0.963923747168626
HIP 67371	0.9637800868947516
HIP 109355	0.9633339117931428
HIP 76804	0.9631493143769808
Gaia DR2 2557461610961429760	0.9630918529295068
HIP 49161	0.9629801094558622
HIP 34271	0.9629553048598934
Gaia DR2 363702817083391232	0.9628712871287128
HIP 10540	0.9628367198358594
HIP 109381	0.9626427909848718
HIP 115861	0.9625152948179504
HIP 69564	0.9621758788570106
HIP 16517	0.9619071248397636
HIP 544	0.9614218314921464
HIP 53437	0.9613514586227426
HIP 48583	0.9610934204540756
HIP 21327	0.9610558164142584
HIP 114590	0.961038853792477
Gaia DR2 4224062406762625152	0.9610362694300518
HIP 19255	0.9610260735815824
HIP 9073	0.96083939193655
HIP 61662	0.9605182649534488
HIP 14774	0.960490992144456
HIP 80999	0.960165618465963

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 67388	0.9598762112022242
Gaia DR2 5160557726183065984	0.9595267745952676
HIP 85969	0.9592371536274056
HIP 116823	0.958985134386142
Gaia DR2 2127096572119535232	0.9589649073507778
HIP 44668	0.9588749782790378
HIP 7607	0.9578471182712108
HIP 21923	0.9578144895900514
HIP 25873	0.9573111217720724
KOI-834	0.9570135746606336
HIP 83289	0.9568035985976052
HIP 64844	0.9567345049349448
HIP 50174	0.9562285109759324
Gaia DR2 2103850667223592064	0.9561798124472818
HIP 10626	0.9561603989195928
HIP 41777	0.9561206846629218
HIP 14194	0.9558069531275842
HIP 51386	0.955606490637231
HIP 13889	0.9554142760296964
HIP 88268	0.9553903345724908
HIP 86456	0.9553891671212046
HIP 72772	0.9553214885120668
Gaia DR2 2078741876099629056	0.9550068972353236
HIP 99729	0.9548797620031404
HIP 66487	0.954719788394776
Gaia DR2 6038073970589665536	0.9543387302164816
Gaia DR2 2073655432234604288	0.9543379750603646
HIP 29193	0.954229077803128
HIP 80543	0.9540855130452036
HIP 12048	0.9540300968872396
HIP 89771	0.9536105069017708



TABELA A.1. Continuação

Estrela	Probabilidade
Gaia DR2 2105494780705108608	0.9535925668749792
HIP 94061	0.9535441405087156
KOI-2612	0.9534662553856588
HIP 5862	0.9529926548438328
HIP 13402	0.952817943566722
HIP 50839	0.9526337550649135
HIP 114456	0.9519665451780592
KOI-4188	0.951880544692391
HIP 96854	0.9518402117277314
HIP 78998	0.9516678518861528
HIP 70873	0.950918520111496
HIP 90055	0.9507068437012004
HIP 7862	0.9505238071595252
HIP 305	0.950521587346233
Gaia DR2 2075132209486310912	0.9503925737782228
HIP 2926	0.950165480721496
HIP 92962	0.9501600668392796
HIP 9553	0.9499413407359673
HIP 17378	0.949860701206154
HIP 31655	0.9496696190074512
Gaia DR2 2105620159389754624	0.9494608439468044
HIP 13976	0.949364159939184
HIP 71719	0.9491717365427852
HIP 96895	0.9488048681675748
HIP 112201	0.948537566417824
HIP 102532	0.9482300775142496
HIP 84551	0.9481874286470656
HIP 24923	0.9481458202388436
HIP 112100	0.9478826553807748
HIP 101345	0.9477357102105952
HIP 89826	0.9477145811789038

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 61123	0.947478557190475
HIP 94071	0.9474437813992368
HIP 17253	0.9465809845176656
HIP 50784	0.9460416477070596
KOI-4011	0.9459922101763876
HIP 58558	0.945934545815448
HIP 31197	0.9457964784657352
HIP 17356	0.945656128050797
Gaia DR2 65072435323295872	0.94563787812363
HIP 2292	0.9454737694220268
HIP 91582	0.9448410499061106
Gaia DR2 1928431764627661312	0.944809385134688
HIP 71743	0.9443209703138812
HIP 99727	0.9442317709409906
HIP 65530	0.9438287736941348
Gaia DR2 2135179700570225024	0.9428377730376838
HIP 19386	0.9422333352652366
HIP 87154	0.941915266569652
HIP 88595	0.941847489150651
HIP 93016	0.9417007730786722
HIP 54287	0.941585558468192
HIP 71631	0.9412558243468042
HIP 74271	0.940989627276788
HIP 66192	0.9407881163606354
HIP 81062	0.9407184290880204
HIP 53087	0.9406980012234034
HIP 10723	0.9400570884871552
HIP 77810	0.9400292782175024
HIP 117320	0.9400284241142252
HIP 60370	0.940023983790266
HIP 94244	0.9399302340960192

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 7240	0.939672700223159
HIP 98106	0.9395763398541498
HIP 7404	0.9395480366134994
HIP 20827	0.9392283914904872
HIP 103250	0.9389685334920712
HIP 28390	0.9388493666807128
HIP 7244	0.9383869021545987
HIP 1976	0.937966738064324
HIP 117411	0.9379444940623864
Gaia DR2 2119132916774654080	0.9378617496279148
HIP 74890	0.9376872997829906
Gaia DR2 4088006054816932352	0.9375832099827172
HIP 102580	0.9367623934453364
HIP 13718	0.9366870901859616
HIP 42291	0.9366477296220034
Gaia DR2 1838417801379850112	0.9365662381496004
Gaia DR2 2078450196282665088	0.936007475028114
HIP 105388	0.935847560522338
HIP 62127	0.9357828533598146
HIP 28393	0.9354073314704278
HIP 32103	0.9352965471666942
HIP 97071	0.9349776896380764
HIP 20889	0.93488179178764
HIP 15304	0.9346528925619836
HIP 38811	0.9345863108761204
HIP 104226	0.9344240610454784
HIP 17819	0.9342684609220756
Gaia DR2 2073865026641816064	0.9338190900745176
HIP 111928	0.9337073981712386
HIP 20826	0.9336628834533326
HIP 15323	0.933477193243545

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 20480	0.9333421559788588
HIP 9519	0.933003504432076
KOI-687	0.9327503987702176
HIP 43726	0.9327260854830844
HIP 62523	0.9322781933137732
Gaia DR2 2052070369713335808	0.9322477993139644
Gaia DR2 4290415081653653632	0.9319987617376948
HIP 72848	0.9319107790715933
HIP 94465	0.9316398182569188
KOI-92	0.9310980042294476
HIP 91300	0.9307672585620828
HIP 87815	0.9305846245481092
HIP 35209	0.9289905618274682
HIP 72190	0.9288412109391142
HIP 48113	0.9283211582322416
HIP 52498	0.9274663186327554
HIP 29788	0.9273903369094983
KOI-2219	0.9273820154115818
Gaia DR2 898130030131443584	0.9271564176640528
HIP 25670	0.9271550013231014
HIP 15457	0.9269804025248688
HIP 43565	0.926688666683208
HIP 16908	0.9266356917277072
HIP 65882	0.926533582521542
Gaia DR2 2128586655192626048	0.9264043876531176
HIP 72779	0.926220635746924
Gaia DR2 410965186967609856	0.9261479191124192
HIP 116517	0.9256645067442476
HIP 53882	0.925401790886671
HIP 21143	0.9253897826140842
HIP 4310	0.9252633753356744

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 17351	0.9248271412975154
HIP 70459	0.924656627502896
HIP 12611	0.9240687323850534
HIP 46901	0.9239239156492052
HIP 29094	0.9231761253255074
HIP 48351	0.92308901780025
HIP 100500	0.9226723560495176
HIP 96229	0.9225642888995326
HIP 44657	0.9223549064686328
HIP 46580	0.921958120641014
HIP 42172	0.9218570519548132
HIP 97125	0.921750992063492
Gaia DR2 2102991776844251264	0.9217242491133212
HIP 407	0.92149460135978
HIP 46677	0.921262576575534
KOI-2803	0.9208545622973598
HIP 29761	0.9207985450938247
HIP 6116	0.9207461133048062
HIP 19983	0.9205496213499124
HIP 98599	0.9201811585219712
HIP 2498	0.9195772096831554
HIP 98920	0.919541465754671
HIP 6456	0.919505283421046
HIP 84920	0.9194820005950216
HIP 102791	0.9192845506941976
HIP 98921	0.9187351634808644
Gaia DR2 2127790368255394048	0.9186858316221764
HIP 33109	0.9179909053327822
HIP 108416	0.9178748356256152
HIP 22919	0.9178484552630272
HIP 19126	0.9172896268854194

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 59280	0.9169934316109926
HIP 213	0.916987251081714
HIP 91258	0.9167699297810822
HIP 40110	0.916696292009425
HIP 42202	0.9165634035522512
HIP 27890	0.9162974840480048
HIP 4423	0.9161782575989288
HIP 34524	0.9154463746785624
HIP 51260	0.9151740676424378
HIP 114924	0.9147910499594832
HIP 75132	0.91476658050557
HIP 32480	0.9144966066150366
HIP 108065	0.9144182124789209
HIP 118008	0.9143654183636304
HIP 99913	0.91435395626374
HIP 38531	0.9140317512816272
HIP 62145	0.9139071205382976
HIP 9269	0.9138989661925973
HIP 113779	0.9138549832952928
HIP 14355	0.9133695499383988
KOI-18	0.913115432354158
HIP 27244	0.9123635461462124
HIP 44463	0.9116399798298764
HIP 93007	0.9113858154504664
HIP 24331	0.9113491001041202
HIP 22938	0.9110945705661
HIP 117880	0.9108841286907616
HIP 20130	0.910649909045808
HIP 61546	0.910493368349768
HIP 22787	0.910373924553276
HIP 7918	0.9103361670838808

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 4151	0.9098458280950176
KOI-3876	0.9097641817445428
Gaia DR2 509868109012678016	0.9095808878234272
HIP 57757	0.908900826446281
HIP 114796	0.908895467098844
HIP 490	0.9086435070306038
Gaia DR2 2129804600536996480	0.90851489734809
HIP 10090	0.9082984226138832
HIP 102753	0.9081793889656428
HIP 23214	0.9076358345183372
HIP 4801	0.9075043630017452
HIP 30220	0.906971362452994
HIP 18262	0.906786671519804
HIP 80000	0.9065753944988976
HIP 60973	0.90571386069882
Gaia DR2 736174028943041920	0.9056583485661822
HIP 23575	0.905543435144784
Gaia DR2 1932997143065038464	0.9054896191483734
HIP 80683	0.9052369695667268
HIP 94951	0.905027932960894
HIP 106213	0.9045610203980348
HIP 51987	0.9032124797567505
HIP 92026	0.9029361424165502
HIP 37170	0.9026939922320468
HIP 81300	0.9026304256146068
HIP 40283	0.9025597382423902
HIP 107528	0.9020864517089388
HIP 73269	0.901872095523178
HIP 108950	0.9017482951022938
Gaia DR2 4583004815239650176	0.9013675922088686
HIP 77152	0.90113134520915

TABELA A.1. Continuação

Estrela	Probabilidade
HIP 69316	0.9010088324478852
HIP 117541	0.9010015631332136
HIP 68184	0.9006502143972505
HIP 24479	0.9004971009338218
HIP 49841	0.9001778549861439



## Apêndice B

# Lista de estrelas com probabilidades altas de abrigarem *planetas menores*

Este apêndice contém uma listagem de estrelas e suas probabilidades acima de 90% de abrigarem planetas menores. Aqui, apresentamos as probabilidades obtidas a partir do teste usando o raio máximo planetário (Seção 4.2.1).

TABELA B.1. Estrelas e Probabilidades de abrigarem planetas gigantes, Teste 4.2.1

Estrela	Probabilidade
KOI-4561	1.0
KOI-4383	1.0
KOI-361	1.0
KOI-2072	1.0
KOI-4617	1.0
KOI-3015	1.0
Gaia DR2 2104172854190854528	1.0
Gaia DR2 2099073667159859840	1.0
KOI-2352	1.0
KOI-4400	1.0

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2073595195312122880	0.9999862224794026
Gaia DR2 2105620159389754624	0.9999586412264256
KOI-3196	0.9999448997864868
Gaia DR2 2075162102458709632	0.9999448785260516
HIP 10510	0.9999311370647862
KOI-2636	0.99991722082724
KOI-4287	0.9999035972015644
Gaia DR2 2073865026641816064	0.999903583923308
HIP 19255	0.9999035414082954
KOI-2687	0.9999035307736832
Gaia DR2 2127096572119535232	0.9998759236792764
Gaia DR2 59670637774212224	0.9998483804272916
Gaia DR2 2119132916774654208	0.9998483574353106
Gaia DR2 2007417175286304000	0.9998348972235216
HIP 34271	0.9998346788636928
KOI-3225	0.9998070828165908
Gaia DR2 2099574421688108928	0.9998070668651122
HIP 96561	0.9997935732470928
Gaia DR2 2128525907175071872	0.9997518439374096
KOI-92	0.9997246506505127
KOI-3232	0.9997245292894232
HIP 80315	0.9996829597772448
Gaia DR2 2052579374867447936	0.9996695352839932
HIP 42491	0.9996695216325632
HIP 105522	0.9996693622825024
Gaia DR2 1928431764627661312	0.999669303056191
KOI-299	0.999656946826758
Gaia DR2 2100338792724606720	0.9996563573883162
KOI-257	0.999656002751978
KOI-4657	0.9996558749036448
Gaia DR2 2104071978297949056	0.9996555287633484

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2130326460543631360	0.9996550534667128
KOI-2722	0.9996549344375432
KOI-1809	0.9996548153261996
KOI-167	0.9996546961325968
Gaia DR2 2100408508633587712	0.9996540989277066
Gaia DR2 2117369540643692544	0.9995728085769012
Gaia DR2 2178963941519859456	0.999572802689965
Gaia DR2 2075132209486310912	0.9995589309589116
Gaia DR2 377244848968006016	0.9995173146143344
Gaia DR2 2130160399927772672	0.9995038930613932
KOI-4613	0.9994902598297192
KOI-312	0.9994840928632848
KOI-3060	0.999482580200069
HIP 58876	0.9994629726529152
KOI-1808	0.9994213839943792
Gaia DR2 2131314401800665344	0.9994071500461872
KOI-3111	0.9993803189291912
HIP 7830	0.999379806499628
KOI-370	0.9993143640726774
KOI-2246	0.9993143640726774
Gaia DR2 2103314346067173504	0.9993117024351968
Gaia DR2 2125713940547834496	0.9993105825577387
Gaia DR2 2119635153070609920	0.999310392386732
KOI-2561	0.9993072393488048
Gaia DR2 2129299035647555968	0.9992701735059212
HIP 107457	0.99922818236948
HIP 10218	0.9991884121765684
HIP 108809	0.99918639767227
Gaia DR2 2073655432234604288	0.99917332598512
HIP 4041	0.999144921939648
HIP 27225	0.9991313100671512

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2099518758909586304	0.9991181293318452
HIP 86456	0.9991173996386856
HIP 9073	0.9990630511463844
HIP 45859	0.9990629736805844
HIP 20218	0.9990358126721764
HIP 104587	0.9990221867210204
HIP 38340	0.9989938667217972
Gaia DR2 2078450196282665088	0.9989798028537948
KOI-2261	0.9989697802197802
KOI-323	0.9989687177724303
KOI-714	0.9989669421487604
KOI-4246	0.998965160400138
Gaia DR2 2104354690224334592	0.9989630141721396
Gaia DR2 2116862833283266432	0.998961937716263
Gaia DR2 2086505630781210752	0.9989597780859916
Gaia DR2 2107586872155346304	0.9989527352900648
Gaia DR2 2076615077014515968	0.9989252645502644
KOI-3197	0.998924657062108
HIP 6878	0.998898147510502
Gaia DR2 2131263759845581696	0.9988556143834104
KOI-1904	0.998789764868603
Gaia DR2 2103639835867512832	0.9986902513338756
Gaia DR2 2101241285613832192	0.9986249570299072
KOI-517	0.998624011007912
Gaia DR2 1341681095315252352	0.99862279819862
Gaia DR2 2133230786147644928	0.9986221530236852
KOI-987	0.9986206896551724
HIP 30687	0.9986077990819744
Gaia DR2 2052120191332583424	0.9985795454545454
Gaia DR1 3909309847346094208	0.9985653985157392
Gaia DR2 1999551612692096128	0.9985415319418264

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2127619222398354304	0.9985123352204636
HIP 35145	0.9985122327529204
HIP 58314	0.9984814600071784
HIP 117184	0.9984287565123908
Gaia DR2 2079613685738988800	0.998373580328592
HIP 95849	0.998332230675929
Gaia DR2 64804841680675200	0.9983064616062456
Gaia DR2 6237190612936085120	0.9983031908289534
Gaia DR2 2077299694799977856	0.9982915639079096
HIP 27417	0.998291163662422
KOI-1438	0.9982853223593964
Gaia DR2 2101689061725421312	0.9982806052269602
Gaia DR2 2126648731589350272	0.9982788296041308
Gaia DR2 2127704842568104320	0.9982771690441734
KOI-752	0.9982770503101308
KOI-1218	0.9982758620689656
Gaia DR2 2135377135926967424	0.9982731825246072
Gaia DR2 2053235577156249088	0.998265093684941
KOI-974	0.9982351156826518
Gaia DR2 2119132916774654080	0.9981828443991684
HIP 60753	0.9980726341584296
Gaia DR2 2127832695152371584	0.9980161463643128
HIP 14613	0.9979882325382718
Gaia DR2 1842775651294526720	0.9979744812191694
HIP 20480	0.99794791279318
KOI-1628	0.997936016511868
Gaia DR2 2077119748545602304	0.9979328165374676
KOI-759	0.9979324603721572
HIP 108947	0.9979210837899606
KOI-2583	0.9979176147725236
Gaia DR2 377245055126435072	0.9978543724039282

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2080195945866437504	0.9978228817879928
HIP 48423	0.9977953537079396
KOI-2169	0.9977644024075666
KOI-2743	0.9977574607555632
HIP 48524	0.9977408601261812
HIP 104225	0.9977249538096684
HIP 94071	0.9976973139926092
KOI-3876	0.997686651427942
HIP 61205	0.9976153390215998
KOI-1819	0.997592019263846
KOI-1837	0.997584541062802
Gaia DR2 2053183453434348672	0.9975820379965458
Gaia DR2 2076237253025202688	0.9975741871459484
HIP 118312	0.9975465872753336
Gaia DR2 1500133124742943104	0.997517309867314
HIP 682	0.9974653552635204
HIP 106913	0.9974652854309016
HIP 78161	0.997408540787914
Gaia DR2 2131133261560058496	0.9973679858890973
KOI-3165	0.997314271547806
Gaia DR2 2838213864935858816	0.9973125318706156
Gaia DR2 2121031257959210624	0.9972996059629108
Gaia DR2 2077773790469159680	0.9972971854875408
Gaia DR2 2052879266666889216	0.9972866882446112
Gaia DR2 2131060521993182976	0.9972731779871096
KOI-1534	0.9972536903535874
Gaia DR2 2085286890863482752	0.9972451790633609
HIP 109176	0.9971564242725416
HIP 65882	0.9971056439942112
Gaia DR2 2077386315701921920	0.997064410291832
Gaia DR2 2105943072207122688	0.9969679428870696

TABELA B.1. Continuação

Estrela	Probabilidade
KOI-3224	0.996934929379116
KOI-327	0.996917808219178
HIP 102791	0.9969164693574144
Gaia DR2 65072435323295872	0.9968335260752244
HIP 12685	0.996817875001722
HIP 68805	0.9967905451940824
HIP 2567	0.9967753937740296
Gaia DR2 3394546197171292672	0.9967096658727644
HIP 28764	0.9966236701394632
Gaia DR2 2080505900764913792	0.9965993887160284
KOI-2153	0.9965647543799382
KOI-2017	0.9965612104539202
HIP 8399	0.9965294992563212
HIP 57148	0.9965261024799767
HIP 41967	0.996516159237686
HIP 97527	0.9964880386728918
Gaia DR2 2132546305800554112	0.9964598112817687
HIP 12638	0.9964309337449012
HIP 101785	0.9964162152476256
KOI-69	0.9963985594237696
HIP 77790	0.9962238668155068
Gaia DR2 2127396880530950272	0.9962225274725276
Gaia DR2 2105389708623762304	0.9961959892495348
HIP 35093	0.996184625556122
HIP 25033	0.9961818908599706
HIP 28902	0.9960553356412838
HIP 76704	0.9960201608460946
HIP 56960	0.996004519026756
KOI-3473	0.9959620181640276
HIP 61739	0.9959240440093086
KOI-4276	0.995893223819302

TABELA B.1. Continuação

Estrela	Probabilidade
KOI-2545	0.9958734525447044
KOI-1931	0.9958720330237358
Gaia DR2 2101719263935952768	0.9958318860715526
HIP 919	0.9957676188375428
KOI-3991	0.9956824013021768
Gaia DR2 65289623229979136	0.9956585856636851
KOI-3039	0.9955234159779616
KOI-907	0.9955234159779616
Gaia DR2 1838417801379850112	0.9954118327867948
HIP 42109	0.9953387025595763
HIP 26129	0.9951954184391304
KOI-4212	0.9951823812801102
KOI-2001	0.9951388793257776
HIP 105918	0.995107969187096
Gaia DR2 1999550886837136640	0.99502700022041
KOI-103	0.9950102689216944
HIP 37170	0.9950006886103844
KOI-2273	0.9949965493443754
HIP 112812	0.9949321067562727
HIP 21276	0.994885227821052
Gaia DR2 2078794755737702144	0.9948471363028892
Gaia DR2 2129064874031303936	0.9948222298929928
Gaia DR2 2133335823867642752	0.9947168770260018
HIP 38647	0.9945701606901683
KOI-2755	0.9945017182130584
HIP 5521	0.9942678406569936
KOI-1216	0.9941600824458948
HIP 24864	0.9941433433930024
Gaia DR2 2100416823690394624	0.9941175660224
Gaia DR2 2129804600536996480	0.9939532513326262
Gaia DR2 2076223101122787328	0.9936022061358152



TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2079749956462030720	0.9934831075287258
Gaia DR2 2053546910744629504	0.9934414911977908
Gaia DR2 2132195557292587648	0.9934241363147592
HIP 52498	0.9932113300560445
KOI-1820	0.9931058255773872
KOI-708	0.9930963065239904
HIP 6575	0.9929450346547614
Gaia DR2 131624808999798656	0.9927756024926928
Gaia DR2 2127749342727327616	0.9927410992049776
Gaia DR2 2073881583725328896	0.9927198521909384
Gaia DR2 2082282818938558848	0.9926910665526656
HIP 37349	0.9926833301182242
Gaia DR2 2131178238457347712	0.9923901764095469
KOI-1929	0.9921070693205216
HIP 51260	0.9920968787498448
KOI-171	0.9920935029219664
Gaia DR2 2125795166969907968	0.9920913762934184
Gaia DR2 2117151420726319872	0.9920880632954936
Gaia DR2 2134950078738454016	0.9920716994139952
Gaia DR2 4573892028507342336	0.99200441135925
KOI-958	0.9919885274610112
HIP 17778	0.9918849285625716
HIP 53943	0.9918665563826854
HIP 67904	0.9917400956990582
KOI-1930	0.991722710812209
Gaia DR2 2104705193915993216	0.9917155678287884
Gaia DR2 2073677040206349696	0.9916987885462556
HIP 22787	0.9915799627919796
HIP 60641	0.9915591479283422
HIP 50061	0.9915162032254956
HIP 54426	0.9914900648572726

TABELA B.1. Continuação

Estrela	Probabilidade
HIP 104809	0.9910976366016676
Gaia DR2 2103877364740680704	0.9910896551724138
KOI-386	0.9910836762688614
KOI-4146	0.9910158949550796
HIP 89771	0.9909842845326716
Gaia DR2 2129299207446242304	0.9907945869966652
KOI-2675	0.9907565902088326
HIP 12158	0.9906863960761622
Gaia DR2 2101479879638683392	0.9904516520157624
Gaia DR2 2133108190601426176	0.9903514817367332
Gaia DR2 2080560635824924672	0.9901557976009928
KOI-2029	0.9900684931506848
HIP 21654	0.9900649019580824
KOI-639	0.9900582790538224
HIP 116217	0.9898533142163892
HIP 14300	0.9897818472655064
Gaia DR2 1470535165076817408	0.9897435897435898
Gaia DR2 2128586655192626048	0.989741468459152
HIP 112219	0.9896048749551934
KOI-2698	0.9893507385778084
HIP 8486	0.989185093338844
Gaia DR2 2053226403105909376	0.9890677989467588
KOI-1270	0.9889807162534436
KOI-3017	0.988897011158159
HIP 63995	0.9887991843932548
HIP 73269	0.98873802467434
HIP 117946	0.9886582878562372
HIP 88790	0.9886296372771698
Gaia DR2 2079403880876910464	0.988243077474398
KOI-440	0.9882352941176472
Gaia DR2 65284469269256704	0.9881954296891142

TABELA B.1. Continuação

Estrela	Probabilidade
HIP 33580	0.9880061210141032
Gaia DR2 2101719057777467008	0.9879795431675008
Gaia DR2 2104106647274122880	0.9879351947604276
Gaia DR2 1298310687356960896	0.987878787878788
HIP 10276	0.9878648466232316
KOI-1947	0.9876793318724936
KOI-2581	0.9874746806663636
KOI-1751	0.987293956043956
Gaia DR2 2129558760909448064	0.9871616932685636
KOI-1305	0.986819285466528
HIP 63458	0.9864991458643302
HIP 58333	0.9863799283154122
KOI-3371	0.9862282664830436
KOI-3008	0.986206896551724
Gaia DR2 2007417244005768064	0.9857091670800948
HIP 40419	0.9855545457049216
KOI-1797	0.985228443833734
HIP 112245	0.9849131992284376
HIP 63452	0.984838878628332
HIP 19793	0.9848165447305696
KOI-2105	0.9844666896789782
HIP 80262	0.9844184829030392
KOI-321	0.9842357779300892
Gaia DR2 2076904282933527424	0.984232868405094
HIP 13968	0.9842076277020516
KOI-1856	0.9841851999944848
KOI-1647	0.9841761265909872
Gaia DR2 2100642876410617856	0.9840340533392108
HIP 58619	0.983985887735836
HIP 67125	0.9836654867987936
KOI-3867	0.9834938101788172

TABELA B.1. Continuação

Estrela	Probabilidade
KOI-1964	0.9833588396685556
HIP 29650	0.983302565232965
HIP 86722	0.9831556003145392
KOI-2632	0.983068417415342
KOI-2147	0.9824259131633356
HIP 35209	0.9819051294065764
Gaia DR2 2076084974972891392	0.981849003555776
HIP 22960	0.981631386609667
Gaia DR2 4604548126378880256	0.9812496553245464
HIP 8553	0.9811762760445376
KOI-543	0.9810410203378146
HIP 74118	0.9809296617531472
HIP 77810	0.9808319907464784
Gaia DR2 2135179700570225024	0.9808198778318602
KOI-2260	0.9806429312132734
Gaia DR2 2100020484108391808	0.9806360093101406
HIP 19386	0.9806181246726036
HIP 116350	0.9804226826848892
KOI-2707	0.9804056376761774
HIP 10990	0.9802393313044796
Gaia DR2 2080314315164334464	0.9802298850574712
HIP 57090	0.9801685671789788
HIP 79308	0.9800916938581636
HIP 24137	0.9794838585226724
HIP 5476	0.979203369301916
HIP 60074	0.9791635064881338
Gaia DR2 2117335077826003840	0.979148554474574
HIP 107252	0.9789951315045444
HIP 6712	0.9789844793631994
HIP 84132	0.9789383844501954
HIP 112100	0.978922555209396

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2557461610961429760	0.9786117614653628
Gaia DR2 4088006054816932352	0.9783342286131428
HIP 43852	0.978069330102701
KOI-2754	0.977961432506887
Gaia DR2 1962056410755418624	0.9779305404400116
KOI-288	0.977839602242208
Gaia DR2 2073697419836171264	0.9776634908378948
Gaia DR2 2104890633423618048	0.9773351648351648
HIP 61218	0.9772492528886012
HIP 71631	0.9769732757550256
HIP 73577	0.9765634687719588
Gaia DR2 3426117131673547392	0.9763764342879182
KOI-102	0.976313320825516
KOI-1175	0.9762886597938144
Gaia DR2 2129016564238409728	0.9760080478460988
HIP 20826	0.9758284439302536
HIP 184	0.9757909974495071
KOI-2173	0.9755425421977264
HIP 104318	0.975535716749686
Gaia DR2 2132082999087248384	0.9753086419753086
HIP 92961	0.9752282334797516
KOI-3184	0.9749769452052908
KOI-166	0.9749742886527254
KOI-4409	0.9743437741259512
HIP 97071	0.974226448900834
HIP 7244	0.9740524460114588
Gaia DR2 2126578942662817536	0.9739344985388984
KOI-427	0.973738769868694
Gaia DR2 2128191964879794560	0.973437435452555
HIP 53747	0.9733992726471236
HIP 7576	0.9733346195187828

TABELA B.1. Continuação

Estrela	Probabilidade
KOI-1825	0.9730103806228374
KOI-442	0.9726738152888272
KOI-315	0.972588972258553
HIP 108525	0.9724219632476512
KOI-3248	0.9723178625533672
HIP 93297	0.9722792749328004
KOI-598	0.972184065934066
Gaia DR2 2132757996148004736	0.9721275979533576
HIP 46639	0.9721054590057746
HIP 68090	0.9720776538620404
Gaia DR2 2107201909944691968	0.9717559249211616
HIP 92615	0.9716548907043496
HIP 95428	0.9711566275017562
HIP 49366	0.9710055096418732
HIP 3712	0.970772241208246
KOI-687	0.9698086131004126
HIP 65641	0.9697654548962324
Gaia DR2 2100451630105040256	0.969530421367274
KOI-4129	0.9692972794360613
Gaia DR2 2076805120730177280	0.9692038471077796
KOI-573	0.968787722021038
Gaia DR2 2101695418277116672	0.9686931598963788
HIP 13134	0.9682782369146006
HIP 105455	0.9678272366681356
HIP 97295	0.9675134039943214
HIP 54406	0.9669981106314904
HIP 22729	0.9666804407713498
HIP 47839	0.966597307390208
HIP 112527	0.9665342382607736
HIP 113438	0.9660806789375068
HIP 15323	0.965739554490226

TABELA B.1. Continuação

Estrela	Probabilidade
HIP 30552	0.965489223989534
KOI-2010	0.9654817104585264
HIP 96501	0.9653727970457332
HIP 2156	0.9649816163813878
Gaia DR2 2125814129244571264	0.9649644686828625
Gaia DR2 2130908686309827200	0.9647326037580012
Gaia DR2 2133630011947261824	0.9644609194135156
HIP 33109	0.9643944298287902
KOI-245	0.9641379310344828
Gaia DR2 2129520930836942848	0.9640947182710332
HIP 3236	0.9637868811812996
HIP 54069	0.9637704127334116
HIP 14194	0.9622477551414502
HIP 52942	0.961784140969163
Gaia DR2 2102495759660459648	0.961186742251831
HIP 110508	0.9611829944547134
HIP 2902	0.961158221302999
Gaia DR2 2077247876015217920	0.961124657326666
HIP 70557	0.9602626170673508
HIP 57271	0.9602369798842656
KOI-2623	0.9601785101270168
HIP 107822	0.9593835220229732
HIP 14258	0.959305128876276
HIP 9328	0.9592899864943084
Gaia DR2 2119736892255743744	0.958935832931284
Gaia DR2 2101392537183148416	0.9581071772971484
KOI-649	0.9578947368421052
KOI-2706	0.9577956099713214
HIP 11881	0.9573899262729966
HIP 17356	0.9565989672977624
KOI-1278	0.9565816678152996

TABELA B.1. Continuação

Estrela	Probabilidade
HIP 87579	0.956379167814825
HIP 102290	0.9562170423097576
Gaia DR2 2105597825559486336	0.956043956043956
Gaia DR2 4464207428579939840	0.9555984555984556
HIP 90993	0.9551698507955734
HIP 99710	0.954777745636872
HIP 77372	0.9545172883876564
KOI-2714	0.9542640990371388
KOI-645	0.9534482758620688
HIP 78913	0.9531262917137424
HIP 116984	0.953076710816777
Gaia DR2 135197672034753920	0.9528696993910672
KOI-2803	0.9528684007494764
HIP 51386	0.9524085176762456
HIP 10940	0.9519266560970566
HIP 83006	0.9519135325907137
HIP 48537	0.9517333553791888
HIP 85423	0.950910519568244
HIP 79165	0.9508018296043204
HIP 18806	0.950558081852005
HIP 112768	0.9498323606109524
HIP 79190	0.9497617539317488
Gaia DR2 2078741876099629056	0.94966489229666
HIP 490	0.9485204981265152
Gaia DR2 2130211080541825024	0.9482936918304034
HIP 16517	0.9480937891533918
HIP 36272	0.9478590636420052
HIP 54498	0.9478234532889122
HIP 213	0.9474845542806708
KOI-304	0.946896551724138
KOI-409	0.9465148378191856



TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 1485265047475984256	0.9462081128747796
HIP 57536	0.9461625313248698
HIP 33848	0.9457957296940348
HIP 4148	0.9455176785074544
KOI-2753	0.9444636081407382
HIP 25873	0.9443212342447826
HIP 82860	0.9442028187623872
Gaia DR2 2130889204338693504	0.9439446366782008
HIP 95270	0.9438127919957
HIP 61252	0.943545388185712
HIP 82515	0.9433062181166414
HIP 30630	0.9429708588112158
HIP 65355	0.9426850780001104
HIP 41022	0.9424652629549556
Gaia DR2 2073750986669231232	0.9415204678362572
HIP 107038	0.9412534615543584
HIP 20741	0.9411797103045656
HIP 2742	0.94100558966875
HIP 23786	0.9409689687482776
KOI-2073	0.9399793032080028
HIP 58656	0.9399190662335516
Gaia DR2 1932997143065038464	0.9395349477538996
Gaia DR2 2079005651508492032	0.9392782789729356
HIP 95703	0.9390490886295688
HIP 11451	0.9386263199196002
HIP 12740	0.9385588839414952
HIP 105038	0.9384985894171884
HIP 108961	0.938425038277446
HIP 2498	0.9380394319591892
HIP 110028	0.9376584371211286
HIP 3850	0.936981849230176

TABELA B.1. Continuação

Estrela	Probabilidade
KOI-2342	0.936532507739938
HIP 112136	0.9364753533739288
HIP 73754	0.9356376310460264
HIP 16908	0.935245856505699
KOI-2202	0.935127674258109
HIP 96113	0.9347319989530672
HIP 22576	0.9346449321254507
HIP 53707	0.9335317870481388
KOI-4407	0.9331461756529876
HIP 21099	0.9331231673572776
Gaia DR2 2119828460957942528	0.932574828854392
Gaia DR2 2106402496271967360	0.9322558459422284
HIP 110716	0.9319786037278042
HIP 23750	0.9318225650916104
HIP 81062	0.9317740402495972
KOI-343	0.9317005864091066
Gaia DR2 2132325200882204544	0.9316386121607696
HIP 116258	0.9308889838805312
HIP 83181	0.9308349407550288
HIP 107152	0.9308227194492256
KOI-2414	0.9307745030843044
HIP 10540	0.9293452791178498
Gaia DR2 2086551397953774976	0.9279988982233852
HIP 5891	0.9273753786835582
HIP 54381	0.9271292419568092
HIP 16581	0.926959562721152
HIP 63406	0.9265162064163898
KOI-316	0.9262688614540466
HIP 8227	0.9259438061214638
HIP 14684	0.9229835929960016
HIP 113026	0.9229381798155032

TABELA B.1. Continuação

Estrela	Probabilidade
HIP 35702	0.9227604804424694
Gaia DR2 2104827888243104128	0.9220935043442284
KOI-2832	0.921917808219178
Gaia DR2 2119583510383666560	0.921775694492314
HIP 118008	0.9216999462565628
HIP 36948	0.92167339126124
HIP 2339	0.921282316442606
Gaia DR2 2131701365472538112	0.9197999310106934
HIP 23553	0.9195359539260668
HIP 33955	0.9193588366520696
HIP 72659	0.919293455778936
KOI-2219	0.9192770087685436
Gaia DR2 2086830772685401216	0.9178727330313003
Gaia DR2 2052070369713335808	0.9176475452766216
HIP 19428	0.9165289256198348
KOI-4411	0.9164086687306502
KOI-3202	0.9160015989193512
HIP 24923	0.915856714116578
HIP 7235	0.9152617140651492
HIP 93185	0.9151113869922064
HIP 100017	0.9150956397111516
HIP 99702	0.9140899629175224
HIP 86614	0.9139633869168148
HIP 93427	0.913959988434372
HIP 67388	0.9139257257864408
KOI-3438	0.9137633666781648
HIP 26973	0.9135116503515788
KOI-3083	0.9132250196370252
HIP 103250	0.9129081569415944
HIP 22938	0.912830126261234
HIP 61546	0.9122444481181462

TABELA B.1. Continuação

Estrela	Probabilidade
Gaia DR2 2128256492466658176	0.912121906507791
HIP 21327	0.9110257046036864
HIP 20130	0.9109515408787696
KOI-579	0.9106896551724138
HIP 108996	0.9105551039528252
KOI-4640	0.9104152328515714
Gaia DR2 2051842839520109568	0.9098786828422876
HIP 30220	0.909770083255224
HIP 51949	0.9095982911872116
Gaia DR2 2078049767888231936	0.9094664371772806
HIP 52462	0.9092401500938088
HIP 35314	0.9068305860301008
HIP 56352	0.9062504307077388
HIP 32322	0.9050606729178158
KOI-1615	0.9047454104606858
Gaia DR2 2079597124345617280	0.90417817679558
HIP 114796	0.904042145329674
HIP 23214	0.9035449072096772
KOI-749	0.9033816425120772
HIP 91120	0.903273789028606
KOI-4011	0.9028064951892592
Gaia DR2 509868109012678016	0.9026743561862552
KOI-2399	0.902530779753762
HIP 107001	0.9022143782919227
KOI-4222	0.9020281883808868
Gaia DR2 2099728391973441920	0.9015775034293552
KOI-1862	0.901180786426688