



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

THIAGO BATISTA RODRIGUES SILVA

Predição de compatibilidade entre cultivares em processos de enxertia: um estudo comparativo de técnicas baseadas em sistemas de recomendação

Recife

2022

THIAGO BATISTA RODRIGUES SILVA

Predição de compatibilidade entre cultivares em processos de enxertia: um estudo comparativo de técnicas baseadas em sistemas de recomendação

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional.

Orientador: Ricardo B. C. Prudêncio

Coorientador: André C. A. Nascimento

Recife

2022

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S586p Silva, Thiago Batista Rodrigues
Predição de compatibilidade entre cultivares em processos de enxertia: um estudo comparativo de técnicas baseadas em sistemas de recomendação / Thiago Batista Rodrigues Silva. – 2022.
66 f.: il., fig., tab.

Orientador: Ricardo Bastos Cavalcante Prudêncio.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2022.

Inclui referências e apêndices.

1. Inteligência computacional. 2. Aprendizagem de máquina. I. Prudêncio, Ricardo Bastos Cavalcante (orientador). II. Título.

006.31 CDD (23. ed.) UFPE - CCEN 2022-161

Thiago Batista Rodrigues Silva

**“Predição de compatibilidade entre cultivares em processos de enxertia:
um estudo comparativo de técnicas baseadas em sistemas de
recomendação”**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Inteligência Computacional.

Aprovado em: 12 de agosto de 2022.

BANCA EXAMINADORA

Prof. Dr. Ricardo Bastos Cavalcante Prudêncio
Centro de Informática / UFPE
(Orientador)

Prof. Dr. Douglas Veras e Silva
Departamento de Computação / UFRPE

Profa. Dra. Rosimar dos Santos Musser
Departamento de Agronomia / UFRPE

Dedico esta dissertação aos meus pais, João e Elaine.

AGRADECIMENTOS

Primeiramente, agradeço aos professores Ricardo Prudêncio e André Câmara, que, de maneira calma e segura, me orientaram ao longo dos últimos anos e possibilitaram a realização deste sonho. Serei eternamente grato pela oportunidade concedida e pelos conhecimentos compartilhados, espero ter correspondido a contento toda confiança em mim depositada.

Agradeço aos demais professores e colegas alunos do CIn/UFPE, com quem vivi a fantástica experiência de me reconectar ao mundo acadêmico, após 15 anos de graduado. Agradeço também à Nina Iris, doutoranda da UFRPE, que gentilmente nos forneceu os dados e as informações necessárias acerca do domínio do problema.

Agradeço aos colegas de trabalho, na Procuradoria da República de Pernambuco, por todo apoio e compreensão em relação aos meus horários especiais durante a fase de disciplinas presenciais.

Por fim, nada disso teria sido possível sem o apoio e o amor incondicionais da minha família. Agradeço à minha querida esposa, Maria Eduarda, para mim, uma fonte inesgotável de inspiração e de alegria, com quem tive o prazer de compartilhar cada etapa dessa jornada. Agradeço ao meu cachorro, Luke Skywalker, que garantiu que eu acordaria cedo todos os dias para estudar e a quem devo valiosos insights durante os nossos passeios matinais e vespertinos. Agradeço à minha irmã, Isabella, que sempre foi um exemplo e uma grande incentivadora, graças a ela aprendi a aprender me divertindo. Agradeço especialmente aos meus pais, João e Elaine, que tanto se sacrificaram para me proporcionar as melhores oportunidades, que em absolutamente todos os momentos me estimularam a estudar e a ser grato por esse privilégio, pessoas cujo exemplo de vida me ensina a importância da humildade e da dedicação para se obter o verdadeiro êxito em qualquer aspecto da nossa existência.

"Nada verdadeiramente valioso surge da ambição ou de um mero senso de dever, mas de amor e devoção para com os homens e para com as coisas objetivas." (EINSTEIN, 1947).

RESUMO

A viticultura é a ciência que estuda o cultivo da uva para produção de sucos, vinhos e outros derivados. Tanto os produtos quanto a cadeia produtiva possuem elevada importância socioeconômica e cultural em grande parte do mundo. Recentemente, técnicas de enxertia vêm sendo aplicadas para aumentar a produtividade e a qualidade no setor, no entanto, o processo para encontrar cultivares de porta-enxerto que sejam compatíveis com enxertos de videiras é essencialmente experimental, lento e custoso. Embora a utilização de aprendizagem de máquina no Agronegócio não seja novidade, a literatura carece de trabalhos que demonstrem a aplicabilidade dessa técnica para apoiar especificamente processos de enxertia. Este trabalho, por meio de uma perspectiva de Sistema de Recomendação, oferece uma comparação entre abordagens de predição e de classificação para o problema da seleção de cultivares de enxerto e porta-enxerto. Além disso, também avalia os desempenhos de algoritmos baseados em filtragem colaborativa com os de algoritmos baseados em métodos de Kernel, para as tarefas de predição de *ratings* e de classificação de interações. Ao todo, 17 modelos baseados em algoritmos de filtragem colaborativa e métodos de kernel foram avaliados em um conjunto de dados de 251 interações rotuladas, atingindo o valor máximo de 96% para a métrica f1-score. Os resultados indicaram uma vantagem significativa para a abordagem de classificação, especialmente para os modelos baseados em kernel, bem como a viabilidade de uma ferramenta de apoio à decisão para orientar as escolhas de especialistas das melhores cultivares para enxertia.

Palavras-chaves: aprendizagem de máquina; sistemas de recomendação; filtragem colaborativa; métodos de kernel; agroinformática; enxertia.

ABSTRACT

Viticulture is the science that studies the cultivation of grapes for the production of juices, wines and other derivatives. Both the products and the production chain have high socio-economic and cultural importance in much of the world. Recently, grafting techniques have been applied to increase productivity and quality in the sector, however, the process to find rootstock cultivars that are compatible with vine grafts is essentially experimental, slow and expensive. Although the use of machine learning in Agribusiness is not new, the literature lacks works that demonstrate the applicability of this technique to specifically support grafting processes. This work, through a Recommendation System perspective, offers a comparison between prediction and classification approaches to the problem of scion and rootstock cultivar selection. Furthermore, it also evaluates the performance of algorithms based on collaborative filtering with those of algorithms based on Kernel methods, for rating prediction and interaction classification tasks. In all, 17 models based on collaborative filtering and kernel methods were evaluated on a dataset of 251 labeled interactions, reaching a maximum value of 96% for the f1-score metric. The results indicated a significant advantage for the classification approach, especially for kernel-based models, as well as the feasibility of a decision support tool to guide specialist choices of the best cultivars for grafting.

Keywords: machine learning; recommendation systems; collaborative filtering; kernel methods; agroinformatics; grafting.

LISTA DE FIGURAS

Figura 1 – Exemplo esquemático da operação de enxertia por estaca terminal	15
Figura 2 – Exemplo de enxertia em videira	19
Figura 3 – Filtragem Colaborativa (FC) baseada em usuários x FC baseada em itens .	21
Figura 4 – Filtragem baseada em conteúdo	22
Figura 5 – Representação do kernel <i>trick</i>	24
Figura 6 – Visão geral da solução proposta	29
Figura 7 – Histograma de distribuição das avaliações	32
Figura 8 – Representação em grafo bipartido: subconjuntos de alta-baixa conectividade	34
Figura 9 – Distribuição de graus dos nós do grafo de interação entre enxertos e porta- enxertos	34
Figura 10 – <i>Dataset</i> de interações - entrada dos modelos de FC na abordagem de pre- dição de <i>ratings</i>	35
Figura 11 – Validação cruzada k-folds	39
Figura 12 – Diagramas de dispersão dos modelos - Predição de <i>ratings</i>	43
Figura 13 – Diagramas de dispersão dos modelos - Classificação de interações	47
Figura 14 – Curva ROC - Classificação de interações	47

LISTA DE TABELAS

Tabela 1 – Exemplo da matriz de <i>ratings</i> original. As células destacadas em cinza representam as interações desconhecidas	31
Tabela 2 – Exemplo da matriz de <i>features</i> dos enxertos	31
Tabela 3 – Estatísticas descritivas do conjunto de dados	33
Tabela 4 – Características das cultivares	36
Tabela 5 – Resultados dos modelos - Predição de <i>ratings</i>	42
Tabela 6 – Erros tipo 1 e 2 - Predição de <i>ratings</i>	44
Tabela 7 – Resultados do modelos - Classificação de interações	46

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizagem de Máquina
AUC-ROC	<i>Area Under the ROC curve</i>
FBC	Filtragem Baseada em Conteúdo
FC	Filtragem Colaborativa
FN	Falsos Negativos
LOO	<i>Leave-one-out</i>
MAE	Erro absoluto médio
MK	Métodos de Kernel
RMSE	Erro quadrático médio
SR	Sistema de Recomendação
TFP	Taxa de Falso Positivo
TVP	Taxa de Verdadeiro Positivo
VC	Validação Cruzada k-fold

SUMÁRIO

1	INTRODUÇÃO	14
1.1	CONTEXTO GERAL	14
1.2	OBJETIVOS	15
1.3	ROTEIRO DA DISSERTAÇÃO	17
2	EMBASAMENTO TEÓRICO	18
2.1	ENXERTIA	18
2.2	SISTEMAS DE RECOMENDAÇÃO	20
2.2.1	Algoritmos de filtragem colaborativa	22
2.2.2	Algoritmos de filtragem baseada em conteúdo	24
2.3	SISTEMAS DE RECOMENDAÇÃO NA AGRICULTURA	25
2.3.1	Seleção de cultivos	25
2.3.2	Seleção de agentes químicos	26
2.3.3	Época de colheita/plantio	26
3	PREDIÇÃO DE COMPATIBILIDADE ENTRE CULTIVARES EM PROCESSOS DE ENXERTIA	28
3.1	SOLUÇÃO PROPOSTA	28
3.2	ESTUDO DE CASO: CONJUNTO DE DADOS	30
3.2.1	Análise exploratória dos dados	31
3.2.2	Pré-processamento e engenharia de <i>features</i>	33
3.3	ESTUDO DE CASO: MÉTODOS	36
3.3.1	Modelos baseados em filtragem colaborativa	37
3.3.2	Modelos baseados em métodos de Kernel	38
3.4	ESTUDO DE CASO: METODOLOGIA EXPERIMENTAL	39
3.4.1	Modelos baseados em FC	39
3.4.2	Modelos baseados em Kernels	39
3.4.3	Métricas	40
4	RESULTADOS	42
4.1	ABORDAGEM DE PREDIÇÃO DE <i>RATINGS</i>	42
4.1.1	Recomendações	44
4.2	ABORDAGEM DE CLASSIFICAÇÃO DE INTERAÇÕES	45

5	CONCLUSÃO	49
	REFERÊNCIAS	51
	APÊNDICE A – MATRIZ DE INTERAÇÕES	55
	APÊNDICE B – MATRIZ DE CARACTERÍSTICAS DOS ENXERTOS	56
	APÊNDICE C – MATRIZ DE CARACTERÍSTICAS DOS PORTA-ENXERTOS	58
	APÊNDICE D – TOP-5 RECOMENDAÇÕES DOS MELHORES MODELOS DA ABORDAGEM DE PREDIÇÃO DE RATINGS	59
	APÊNDICE E – DIAGRAMAS DE DISPERSÃO DOS MODELOS DA ABORDAGEM DE CLASSIFICAÇÃO BINÁRIA	63
	APÊNDICE F – CURVA ROC DOS MODELOS DA ABORDAGEM DE CLASSIFICAÇÃO BINÁRIA	65

1 INTRODUÇÃO

1.1 CONTEXTO GERAL

A videira (*Vitis* spp.) ocupa lugar de destaque entre as espécies agrícolas mais importantes do mundo, por ser uma frutífera perene versátil, com alto valor agregado aos seus produtos e grande importância socioeconômica (LEAO P.C.S., 2009; ZHANG L., 2016; DOUGHERTY, 2012; CANTU D., 2019). Na safra de 2019, o Brasil foi considerado o décimo quinto maior produtor de uvas do mundo, contribuindo com 1,9% das 77,1 milhões de toneladas de uvas frescas colhidas mundialmente (FAO..., 2021).

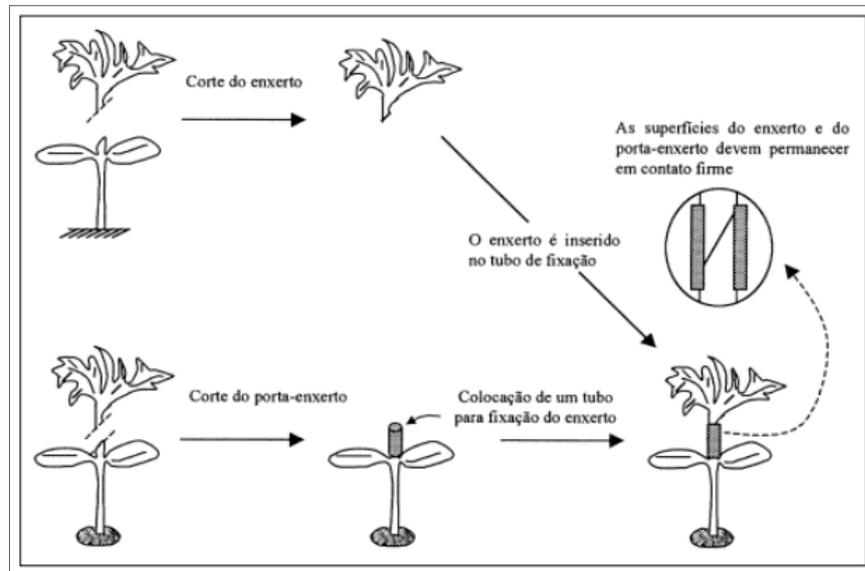
A uva, produzida pela videira, é considerada a fruta domesticada mais antiga conhecida (LEAO P.C.S., 2009). Apesar de ser cultivada e moldada pelo homem desde a antiguidade, a utilização de técnicas de enxertia é considerada relativamente recente no setor *vitivinícola*¹, tendo em conta a sua extensa história de cultivo (PORTA-ENXERTO..., 2021; DOUGHERTY, 2012; CANTU D., 2019).

A enxertia é um método de cultivo simples que consiste, basicamente, na união de duas plantas para formar uma planta híbrida, composta de duas partes: enxerto (ou copa) e porta-enxerto (DOUGHERTY, 2012; CANTU D., 2019; SOARES J.M., 2009) (Ver Figura 1). Diversos benefícios podem ser explorados por meio da enxertia, por exemplo: a capacidade do porta-enxerto em transmitir seu vigor às cultivares enxertadas, assim como a resistência à pragas, doenças e condições abióticas adversas; a influência na precocidade de produção; a modificação do tamanho da planta, promovendo uma forma de nanismo; o aumento da produtividade; a melhoria da qualidade dos frutos; e a adaptação a diferentes tipos de solos (PORTA-ENXERTO..., 2021; HERNANDES J.L., 2011; SOARES J.M., 2009).

Para os enxertos de videira, a escolha dos porta-enxertos deve ser feita com especial cautela, pois consiste em um processo lento e custoso que pode demorar entre 4 e 8 anos para ser totalmente analisado (DALBÓ M.A., 2019; VERSLYPE, 2021). Isso porque, muitas vezes, é necessário o desenvolvimento de pesquisas em cada possível região de plantio para avaliar a influência dos aspectos ambientais na resposta da combinação entre a copa de videira enxertada e o porta-enxerto selecionado. Todavia, a seleção das cultivares para um processo de enxertia, em geral, é feita de forma empírica, baseada na literatura ou em dados históricos. Além disso,

¹ Relativo à vitivinicultura, ciência que estuda o cultivo da uva e a produção de seus derivados, como: vinhos, espumantes, sucos de uva integrais ou concentrados e os destilados vínicos (VITIVINICULTURA, 2022).

Figura 1 – Exemplo esquemático da operação de enxertia por estaca terminal



Fonte: (PEIL, 2003)

estabelecer um vinhedo, per se, é um investimento de longo prazo (SOARES J.M., 2009; SERRA I., 2014), de modo que o resultado prático de tais combinações pode levar ainda mais tempo para ser totalmente compreendido. Assim, o desenvolvimento de métodos computacionais que possam acelerar tal processo é de alta relevância.

Com base na motivação acima, neste trabalho, o problema de compatibilidade entre enxerto e porta-enxerto foi modelado como uma tarefa de recomendação, que pode ser tratada por meio de Sistema de Recomendação (SR). Como será demonstrado na seção 2.3, na literatura, SR têm sido aplicados com diferentes propósitos para aumentar a eficiência no setor agrícola, por exemplo: para selecionar a melhor cultura para plantio, para detecção de doenças em plantações, para previsão de safras, para escolha do melhor período de colheita, entre outros (MOHANTY S., 2020; BONDRE D. A., 2019). No entanto, a investigação de SR ou qualquer outra técnica de Aprendizagem de Máquina (AM) no contexto de enxertia agrícola ainda é nova.

1.2 OBJETIVOS

Neste trabalho, foi utilizada uma abordagem de SR para investigar e comparar os resultados de algoritmos de FC e Métodos de Kernel (MK) (uma classe de algoritmos de Filtragem Baseada em Conteúdo (FBC)) na tarefa de identificar pares compatíveis de enxertos e porta-enxertos. Para tanto, o problema foi estudado sob duas perspectivas: predição de *ratings* de

interações e classificação binária de compatibilidade entre cultivares. Consideramos nesse trabalho um *rating* de interação como um valor numérico fornecido por um especialista, indicando o quão compatível é um dado porta-enxerto associado a um enxerto. A compatibilidade binária, por sua vez, é simplesmente um valor discretizado do *rating* de interação, a partir do qual são definidas classes de compatibilidade e de não-compatibilidade.

Para a realização dos experimentos, foi utilizada uma base de dados real, construída a partir de informações coletadas da literatura especializada sobre processos já conhecidos de enxertia de videiras. Ao todo, esta base de dados contém a avaliação do grau de compatibilidade de 251 interações entre 40 variedades de uvas (utilizadas como enxertos) e 31 variedades de porta-enxertos.

A execução do trabalho foi orientada pelos seguintes objetivos:

1. Comparar os resultados da abordagem de predição de *ratings* e os da abordagem de classificação de compatibilidade;
2. Comparar o desempenho de algoritmos baseados em FC e de algoritmos baseados em MK;
3. Avaliar a viabilidade do uso de AM e SR para o problema da seleção de cultivares em processos de enxertia.

As contribuições deste trabalho, podem ser assim resumidas:

1. Modelagem do problema de predição de pares enxerto e porta-enxerto na vitivinicultura por meio de uma abordagem de SR;
2. Avaliação do uso de SR baseado em FC e MK para o problema específico da previsão de compatibilidade entre pares de enxertos e porta-enxertos;
3. Comparação do desempenho de diferentes algoritmos de FC para os dados disponíveis;
4. Comparação do desempenho de diferentes MK para os dados disponíveis;
5. Desenvolvimento de uma abordagem alternativa para apoiar processos de enxertia na vitivinicultura, extensível para outras culturas.

1.3 ROTEIRO DA DISSERTAÇÃO

Os próximos capítulos estão organizados conforme descrito abaixo:

- Capítulo 2 - Embasamento Teórico: apresenta de forma resumida os principais conceitos necessários à compreensão do trabalho desenvolvido, também apresenta trabalhos publicados que tenham alguma relação com o problema ou que apliquem metodologia semelhante em um contexto diferente;
- Capítulo 3 - Estudo de caso: formaliza o problema, descreve a solução concebida para enfrentá-lo e descreve os experimentos realizados;
- Capítulo 4 - Análise dos Resultados: apresenta os resultados e faz a comparação entre os métodos testados;
- Capítulo 5 - Conclusão: avalia o trabalho em relação aos objetos iniciais e propõe trabalhos futuros.

2 EMBASAMENTO TEÓRICO

Neste capítulo serão apresentados alguns conceitos básicos a respeito dos temas essenciais para a compreensão do trabalho desenvolvido.

2.1 ENXERTIA

A enxertia é uma técnica utilizada por especialistas em plantas, notadamente botânicos e agrônomos, que consiste na união dos tecidos de duas plantas (também chamadas de cultivares), até mesmo de espécies diferentes. A planta resultante não constitui uma nova espécie, as partes que a compõem conservam as características das plantas de origem. No entanto, essas partes cooperam para que a planta resultante continue o seu desenvolvimento como se fosse uma única planta. Abaixo, apresentamos uma definição um pouco mais detalhada que também introduz a terminologia essencial utilizada na área:

A enxertia é uma associação íntima entre duas partes de diferentes plantas que continuam seu crescimento como um ser único. São consideradas duas plantas: o cavalo ou porta-enxerto que é a planta que contribui com o sistema radicular, assegurando a nutrição mineral; e o cavaleiro ou enxerto que é a planta de características nobres que se quer reproduzir, que forma a copa e frutifica, sendo responsável pela absorção da luz do sol e do carbono do ar para transformação da seiva bruta em seiva elaborada, essencial à vida da planta (RIBEIRO G.D.; COSTA, 2005).

Existem diversos métodos para se realizar um processo de enxertia, tais como: enxertia por aproximação, enxertia por estaca, enxertia por fenda, entre outros. Cada método, obviamente, possui vantagens e desvantagens que devem ser consideradas em relação ao resultado desejado (PEIL, 2003). Na Figura 2, que ilustra um exemplo de enxertia em videira, é possível visualizar o enxerto, o porta-enxerto e como é feita a união dessas cultivares.

O uso da enxertia como forma de reprodução assexuada de plantas não é algo recente, existem indícios de que esta técnica já era empregada a mais de 1000 anos no Oriente Médio e na China (MUDGE K.; JANICK, 2009; MENG C.; XU, 2012). Atualmente, existem diversas aplicações em que essa técnica vem sendo bem empregada, por exemplo, na fruticultura, em pesquisas com floração e na recuperação de pomares.

Em sua maioria, os benefícios obtidos por meio da enxertia decorrem do aproveitamento de uma ou mais características inerentes a uma das cultivares, normalmente o porta-enxerto,

Figura 2 – Exemplo de enxertia em videira



Fonte: (COELHO, 2012)

que possam favorecer o desenvolvimento da outra cultivar, o enxerto. Alguns dos benefícios mais comuns são:

1. **Precocidade:** capacidade de induzir a frutificação em períodos mais curtos do que os obtidos naturalmente;
2. **Nanismo:** possibilidade de forçar um processo de nanismo na cultivar enxertada;
3. **Robustez da raiz:** capacidade de conferir maior tolerância às condições adversas oferecidas pelo terreno ou pelo clima para as copas cujas raízes são muito frágeis;
4. **Resistência às doenças e/ou pesticidas:** uso de porta-enxertos resistentes aos patógenos presentes no solo ou aos pesticidas utilizados para combatê-los, a fim de cultivar copas que naturalmente não sobreviveriam a esses agentes;
5. **Mudança de cultivares (Sobre enxertia):** consiste em enxertar novas cultivares em plantas de um pomar já estabelecido sobre enxertos anteriormente realizados, ao invés de replantar um pomar completamente novo.

A união dos tecidos do enxerto e do porta-enxerto pode transferir material celular de uma cultivar para a outra e o compartilhamento desse material genômico tem potencial para criar uma nova espécie. Por esse motivo, a enxertia é também considerada uma forma natural de engenharia genética (PAGE, 2022).

Embora a enxertia possa promover diversos benefícios, também existem algumas desvantagens que devem ser levadas em consideração, por exemplo: possibilidade de transmissão de

viroses, pequena longevidade da planta, desenvolvimento vegetativo mais lento durante a fase de crescimento e risco de rejeição do enxerto decorrente de incompatibilidade entre as cultivares enxertadas (LEÃO, 2004). Por esses motivos, a seleção de cultivares para um processo de enxertia é uma tarefa extremamente complexa, pois não basta que as espécies utilizadas sejam compatíveis, a planta enxertada resultante também precisará oferecer boa tolerância/adaptação ao ambiente em que será inserida, além de apresentar outros possíveis benefícios almejados no processo.

2.2 SISTEMAS DE RECOMENDAÇÃO

Um SR, como o nome sugere, é um *software* capaz de fornecer recomendações de itens para usuários (F. ROKACH L., 2020), por exemplo, quando o Netflix indica 'filmes que talvez você goste', quando o Spotify cria uma *playlist* de 'músicas recomendadas para você' ou quando a Amazon oferece uma lista de 'melhores escolhas de livros para você', todas essas plataformas estão executando um SR. É importante notar que o conceito de usuários e itens em um SR não precisa, necessariamente, fazer referência a pessoas e coisas, respectivamente, mas sim a qualquer tupla (A, B) cujos elementos tenham uma relação do tipo A tem (ou é provável que tenha) preferência sobre B .

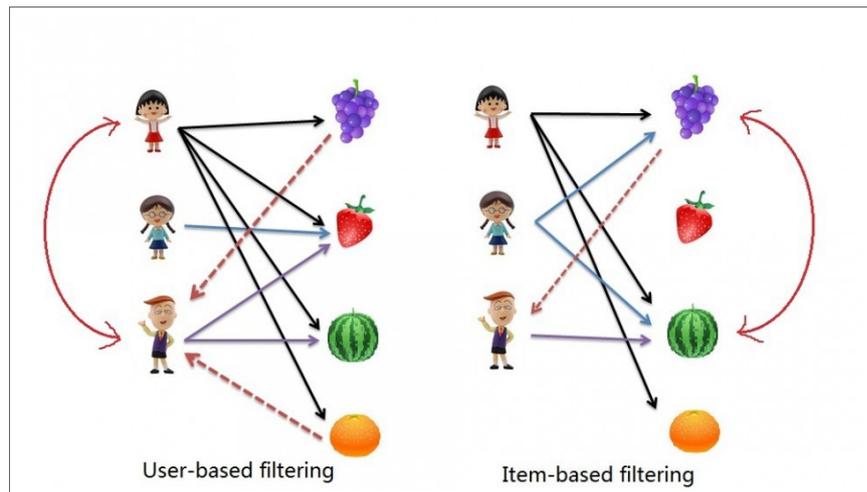
SR é uma área extremamente ativa na Ciência da Computação e encontra uma infinidade de aplicações práticas com alto valor comercial. Dentre as técnicas existentes para construção de um SR, duas se destacam: a FC, que é considerada a mais popular e amplamente implementada; e a FBC (BREESE J.S.; HECKERMAN, 1998).

FC, basicamente, utiliza as informações disponíveis sobre as interações entre usuários e itens para identificar as similaridades existentes entre os membros de cada grupo e, assim, realizar recomendações (F. ROKACH L., 2020). Os dois tipos de FC utilizados neste trabalho, e ilustrados na Figura 3, são:

1. **FC baseada em usuário (*User-based filtering*):** recomenda itens com base nas preferências de outros usuários que tenham interesses semelhantes. Para o exemplo da enxertia, considerando que os enxertos são usuários e os porta-enxertos são itens para recomendação, o sistema recomendaria porta-enxertos que demonstrem boa compatibilidade com outros enxertos similares ao usuário em questão;
2. **FC baseada em itens (*Item-based filtering*):** recomenda itens com base na similaridade

dade que eles tenham com outros itens para os quais o usuário já demonstrou interesse. Esta similaridade é calculada usando as classificações fornecidas pelos usuários. Para o exemplo da enxertia, o sistema recomendaria porta-enxertos similares àqueles compatíveis com o usuário em questão.

Figura 3 – FC baseada em usuários x FC baseada em itens



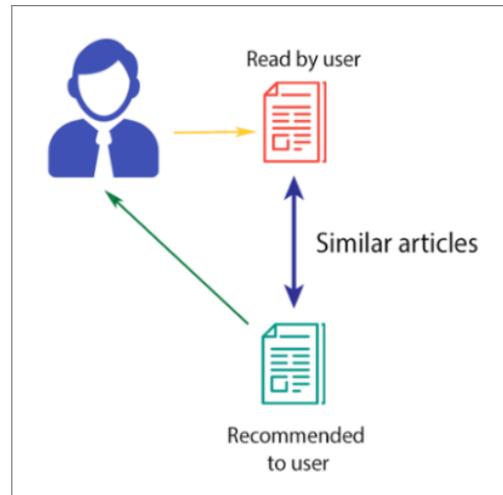
Fonte: (LE, 2018)

FBC, por sua vez, utiliza as informações disponíveis sobre os atributos dos itens e sobre o perfil do usuário para identificar os melhores itens para recomendação (AGGARWAL, 2016). Desta forma, trata a recomendação como um problema de classificação específico do usuário, baseado em suas preferências em relação às características dos itens. Para o exemplo da enxertia, o sistema recomendaria os porta-enxertos cuja configuração de atributos oferecesse maior compatibilidade com o perfil do usuário "enxerto". O conceito de FBC é ilustrado na Figura 4, a partir dela, é possível perceber que não existe dependência em relação às avaliações de outros usuários, apenas no conteúdo dos itens.

A FBC possui algumas vantagens em relação à FC: não requer dados de outros usuários para começar a fazer recomendações, as recomendações são altamente relevantes para o usuário, as recomendações fazem mais sentido para os usuários e, por último, evita o problema da "partida a frio" que torna difícil fazer recomendações para novos usuários ou de novos itens. Entre as desvantagens, podemos citar: falta de diversidade e de novidade nas recomendações; manter os registros de atributos atualizados é um desafio para a escalabilidade; e a forte dependência da correta etiquetagem dos atributos.

A seguir, serão apresentados os algoritmos de FC e FBC utilizados neste trabalho.

Figura 4 – Filtragem baseada em conteúdo



Fonte: (MOVIE..., 2018)

2.2.1 Algoritmos de filtragem colaborativa

Existe uma infinidade de algoritmos mencionados ou disponíveis em bibliotecas de linguagens que utilizam estratégias diferentes para implementar os conceitos de FC. A seguir, serão apresentados os 4 algoritmos utilizados neste trabalho:

- **k-NN:** *k-nearest neighbors* ou k-vizinhos mais próximos, são uma classe de algoritmos que podem ser utilizados em problemas de regressão ou de classificação. A ideia básica é que coisas semelhantes estão próximas umas das outras e o quão próximas elas estão é chamado de similaridade. Para o k-NN, em um problema de classificação, a classe de um exemplo x é determinada pela classe majoritária dos k vizinhos mais similares à x . Em um problema de regressão, o valor da propriedade de x poderia ser definido pela média dos valores das propriedades dos k vizinhos mais similares à x . Dois dos principais desafios relacionados ao uso do k-NN são justamente a escolha do valor k e da função de similaridade que maximizam o desempenho do modelo. No contexto de FC, quando o algoritmo KNN faz inferências sobre um usuário/item, ele calcula a “distância” entre aquele usuário/item e todos os outros usuários/itens na base de dados. Feito isso, classifica essas “distâncias” e retorna os k usuários/itens vizinhos mais próximos como as recomendações (D., 2012).
- **SVD:** o nome do algoritmo é um acrônimo para *Singular Value Decomposition*, um termo da Álgebra Linear que designa um método para fatoração de matrizes que permite

decompor uma matriz retangular A em 3 matrizes:

$$A = U\Sigma V^T \quad (2.1)$$

Onde: A é uma matriz ($m \times n$) real ou complexa; U é uma matriz ($m \times m$) ortogonal, real ou complexa; Σ é uma matriz ($m \times n$) diagonal retangular com números reais não negativos na diagonal; e V é uma matriz ($n \times n$) ortogonal, real ou complexa.

No contexto de FC, o algoritmo SVD utiliza fatoraçoão de matrizes para preencher os valores ausentes em uma matriz de *ratings* que relaciona usuários e itens. No entanto, essa substituição não é feita de forma arbitrária, ela utiliza a informação disponível dos *ratings* conhecidos (SARWAR G. KARYPIS; RIEDL, 2000).

- **NMF**: algoritmo similar ao SVD, cuja fundamentação algébrica também é baseada em fatorizaçoão de matrizes, mais especificamente fatorizaçoão de matriz não negativa (*Non-negative matrix factorization*), em uma transformaçoão do tipo:

$$A = WH \quad (2.2)$$

A principal nota distintiva é que todas as matrizes envolvidas não têm elementos negativos e essa característica torna as matrizes resultantes mais fáceis de inspecionar. No contexto de FC, o algoritmo NMF também utiliza fatoraçoão de matrizes para preencher os valores ausentes na matriz de *ratings* entre usuários e itens, muito semelhante ao que é feito pelo SVD.

- **SlopeOne**: algoritmo relativamente simples de FC, baseada em itens, utiliza um preditor da forma representada na Equação 2.3, em que b é uma constante e x a variável que representa os *ratings*, esse preditor pré-computa a diferença média entre as avaliações de um item e outro, para usuários que avaliaram ambos. Desta forma, o SlopeOne consegue fatorar separadamente os itens que o usuário gostou dos itens que o usuário não gostou. É considerado um algoritmo de fácil implementação, cuja precisão pode competir com a de algoritmos mais complicados e computacionalmente caros. (LEMIRE D., 2005).

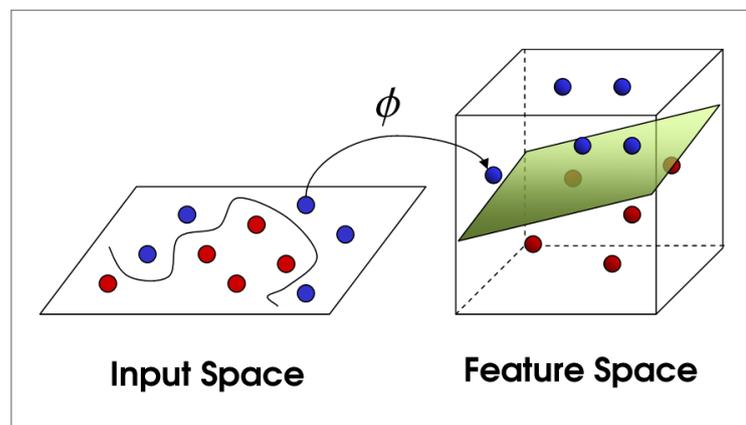
$$f(x) = x + b \quad (2.3)$$

2.2.2 Algoritmos de filtragem baseada em conteúdo

Os objetivos pretendidos pela FBC podem ser alcançados por meio de diversos métodos, como árvores de decisão, redes neurais, classificadores bayesianos, entre outros. Porém, neste trabalho, foram utilizados exclusivamente métodos baseados em kernel. A seguir, será apresentado o conceito dessa classe de algoritmos e as 3 funções de kernel utilizadas neste trabalho:

MK são uma classe de algoritmos de AM muito utilizados para análise de padrões que tentam otimizar o *tradeoff* viés-variância (KOHAVI R.; WOLPERT, 1996) mapeando o espaço de entrada dos dados para um espaço de características de dimensionalidade mais alta (THEODORIDIS, 2008; HOFMANN T.; SCHOLKOPF, 2008). Esta operação, também conhecida como *kernel trick*, permite que o modelo encontre um hiperplano de separação em uma dimensionalidade mais alta para um problema que não seria linearmente separável, conforme ilustrado na Figura 3.

Figura 5 – Representação do kernel *trick*



Fonte: (WILIMITIS, 2018)

Além disso, normalmente, o uso de MK resulta em modelos menos custosos, mais rápidos e, ainda assim, capazes de apreender relações complexas existentes entre as entradas e as saídas. Ou seja, uma combinação das virtudes de modelos mais simples e de modelos mais complexos (SCHÖLKOPF B.; SMOLA, 2018).

No contexto de FBC, em que a recomendação é tratada como um problema de classificação, é possível explorar a capacidade dos métodos de kernel de trabalhar em dimensionalidades mais altas para encontrar o hiperplano de separação que determina se um item deve ou não ser recomendado. Assim, um modelo baseado em kernel precisaria receber como entrada, além

dos *ratings* das interações, as matrizes com o valor dos atributos dos itens e com o perfil dos usuários. Por meio da informação adicional contida nessas matrizes, o modelo seria capaz de executar o *kernel trick* e separar os exemplos em classes positivas e negativas.

É possível utilizar diferentes funções de kernel para mapear o espaço de entrada. A seguir, foram listadas as funções de kernel utilizadas neste trabalho e uma breve descrição sobre cada uma delas (GENTON, 2001):

- **Kernel Linear:** é a função de kernel mais simples, dada pelo produto interno das entradas (x_1, x_2) mais uma constante opcional c (viés ou *bias*);

$$k(x_1, x_2) = x_1^T x_2 + c \quad (2.4)$$

- **Kernel Gaussiano:** é uma função de kernel de base radial, classificado como estacionário, pois depende exclusivamente do vetor que separa os dois exemplos:

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{\theta}\right) \quad (2.5)$$

- **Kernel Polinomial:** é uma função de kernel não estacionário, a classe mais geral de kernels existente, são adequados para problemas em que todos os dados de treinamento são normalizados. Os parâmetros ajustáveis são a inclinação α , o termo constante c e o grau polinomial d :

$$k(x_1, x_2) = (\alpha x_1^T x_2 + c)^d \quad (2.6)$$

2.3 SISTEMAS DE RECOMENDAÇÃO NA AGRICULTURA

SR são amplamente utilizados nas mais diversas áreas do conhecimento, tanto para exploração comercial quanto para o desenvolvimento científico (F. ROKACH L., 2020). Para o agronegócio, especificamente, é possível perceber que essa técnica vem sendo aplicada em muitos cenários diferentes, desde política pública até seleção de cultivo. Nesta seção, apresentaremos as áreas mais comuns de aplicação de SR na agricultura, reforçando o valor deste trabalho no contexto geral.

2.3.1 Seleção de cultivos

Foram encontradas muitas publicações de pesquisadores, principalmente indianos, relatando estudos envolvendo o uso de SR para subsidiar a escolha do melhor cultivo para uma

região, com base em sua geografia, clima ou uma combinação das duas condições (VIVEK M.V.R., 2019; PATEL K., 2020). Por exemplo, em (MOKARRAMA M.J., 2018), os autores apresentam um SR que usa a correlação de Pearson para determinar a similaridade entre uma região X qualquer com outras n regiões cujo histórico de cultivo é conhecido, com base nessas relações, o sistema é capaz de recomendar os cultivos mais apropriados para a região X. Em (PUDUMALAR S., 2017), um SR baseado em *ensemble*, que utiliza a técnica de voto majoritário, foi proposto para recomendar o melhor cultivo a ser explorado em um região, de acordo com as características do solo das propriedades rurais, por meio da análise dos resultados de modelos preditivos utilizando árvores de decisão, CHAID, k-NN e Naive Bayes.

2.3.2 Seleção de agentes químicos

Pesquisas relacionadas à escolha de fertilizantes, pesticidas e agentes químicos em geral, também são intensivas no uso de SR. Em (LACASTA J., 2018), os autores construíram uma ontologia que modela as interações entre cultivos, pragas e pesticidas e desenvolveram um SR para facilitar a identificação de pragas e a escolha do tratamento mais adequado. Em (BONDRE D. A., 2019), por meio da análise do banco de dados nacional de solos de Bangladesh, os autores desenvolveram um sistema capaz de gerar recomendações de fertilizantes específicos por localidade para culturas selecionadas. O software requer informações sobre a localização do campo do agricultor, tipo de solo e terra, tipo de cultura e variedade para gerar recomendações instantâneas de fertilizantes específicos para um cultivo em determina localidade.

2.3.3 Época de colheita/plantio

Outro subconjunto de aplicações envolvendo SR comuns na literatura é o da escolha da época mais adequada para iniciar o plantio ou a colheita. Em (H. EL-BENDARY N., 2020), os autores implementaram um sistema que usa regressão de vetores de suporte e dados sobre as condições climáticas e seus efeitos nas plantações para fornecer recomendações aos agricultores em relação ao momento adequado para iniciar o plantio ou para realizar a colheita. Em (AL., 2021), os autores mediram a eficácia de um sistema que utilizou aprendizagem de máquina para prever os tempos de colheita da safra de arroz. Este estudo indicou que as recomendações do sistema eram suficientemente semelhantes as de agricultores experientes e especialistas.

Obviamente, as áreas de aplicação listadas nas 3 seções deste capítulo não esgotam o

rol de publicações existentes sobre SR no agronegócio. Existem ainda várias outras iniciativas, porém mais isoladas, sobre aplicações de SR em contextos diferentes, por exemplo, em (JAISWAL S., 2020), um SR baseado em filtragem colaborativa foi utilizado para recomendar os melhores programas governamentais disponíveis para agricultores na Índia, com base no perfil dos agricultores e nas características de admissão dos programas.

No entanto, embora o rol de publicações sobre o uso de SR na agricultura seja imenso, até onde sabemos, não há estudos que demonstrem o uso dessa técnica para abordar o problema específico da compatibilidade entre enxerto e porta-enxerto, ainda mais que ofereçam uma comparação entre modelos baseados em FC e em MK.

3 PREDIÇÃO DE COMPATIBILIDADE ENTRE CULTIVARES EM PROCESSOS DE ENXERTIA

Conforme mencionado nos capítulos anteriores, o cultivo de videiras para a produção de sucos, vinhos e outros derivados tem grande relevância econômica, social e cultural, pois os impactos dessa cadeia de produção e consumo se estendem por todas as regiões do mundo. A enxertia é uma prática muito comum na vitivinicultura, porém, selecionar as cultivares que serão utilizadas no processo não é uma tarefa fácil. Em geral, isso é feito empiricamente e o resultado da experimentação é lento e custoso. Portanto, o desenvolvimento de ferramentas capazes de orientar a escolha de cultivares com maior chance de compatibilidade pode ser um recurso valioso para pesquisadores e empresários da área.

Este trabalho apresenta uma abordagem de SR baseada em FC e FBC para tratar, especificamente, o problema da escolha de pares de enxertos e porta-enxertos compatíveis para processos de enxertia em viticultura. Para tanto, foram utilizadas duas perspectivas distintas para estudar o problema: na primeira, ele é tratado como um problema de predição de *ratings* de interações e na segunda, ele é tratado como um problema de classificação de interações entre cultivares.

No decorrer deste capítulo, será detalhada a solução proposta e o estudo de caso selecionado.

3.1 SOLUÇÃO PROPOSTA

A Figura 6 apresenta uma visão geral do trabalho desenvolvido e cada etapa é detalhada a seguir:

1. **Coleta de dados:** Coletar dados de livros, artigos científicos e sites especializados sobre o grau de afinidade entre cultivares em processos de enxertia em viticultura;
2. **Matrizes de características e de interações:** Separar os dados em 3 matrizes que servirão como entradas para os modelos, são elas: uma matriz para armazenar as interações entre cultivares, enxerto e porta-enxerto, e seus graus de compatibilidade; e outras duas matrizes para armazenar as características dos enxertos e dos porta-enxertos;
3. **Pré-processamento:** Realizar as transformações e a engenharia de *features* necessárias para ajustar a estrutura das matrizes e o formato dos dados, a fim de utilizá-los para

Figura 6 – Visão geral da solução proposta

Coleta de Dados	<ul style="list-style-type: none"> • Interações entre enxertos e porta-enxertos • Fonte: Literatura especializada 	
Construção das Matrizes	<ul style="list-style-type: none"> • Matriz com avaliação das interações • Matriz de características dos enxertos • Matriz de características dos porta-enxertos 	
Pré-processamento	Predição de ratings	Classificação de interações
	<ul style="list-style-type: none"> • Dataset de enxerto, porta-enxerto, rating 	<ul style="list-style-type: none"> • Binarização dos dados • Modelos de FC <ul style="list-style-type: none"> ◦ Dataset de enxerto, porta-enxerto, rating • Modelos de MK <ul style="list-style-type: none"> ◦ Engenharia de features
Construção dos modelos	Filtragem Colaborativa <ul style="list-style-type: none"> • k-NN • SVD • NMF • SlopeOne 	Filtragem Colaborativa <ul style="list-style-type: none"> • k-NN • SVD • SlopeOne Métodos de Kernel <ul style="list-style-type: none"> • Linear • Gaussiano • Polinomial
Comparação e análise	<ul style="list-style-type: none"> • k-NN • SVD • NMF • SlopeOne 	
Recomendações	Melhor modelo <ul style="list-style-type: none"> • Top 5 recomendações de porta-enxertos • Avaliação 	

Fonte: Elaborada pelo autor (2022)

treinar e testar os modelos de acordo com as suas especificidades;

- Treinamento e teste dos modelos:** Treinar diferentes modelos regressores baseados em FC e diferentes modelos classificadores baseados em FC e em MK (uma classe de algoritmos de FBC), usando diferentes configurações de hiperparâmetros e técnicas de reamostragem;
- Comparação e análise:** Com base nas métricas calculadas nos testes, comparar os resultados obtidos por cada modelo, de cada abordagem;
- Recomendações:** Submeter as recomendações de alguns modelos à análise de um especialista no domínio do problema. Neste trabalho, essa etapa foi realizada apenas para os modelos de regressão, na perspectiva de predição de *ratings*.

A solução proposta neste trabalho é uma aplicação de FC e MK em uma abordagem de SR em um problema original. Até onde sabemos, não existem trabalhos anteriores na literatura

que tenham utilizado SR ou outra técnica de AM para abordar o problema de compatibilidade entre enxerto e porta-enxerto, seja em viticultura ou qualquer outra cultura que utilize enxertia.

3.2 ESTUDO DE CASO: CONJUNTO DE DADOS

Os dados utilizados neste experimento representam uma amostra, extraída da literatura, de 251 avaliações, ou *ratings*, do grau de afinidade entre pares de enxertos e porta-enxertos utilizados em processos de enxertia na vitivinicultura. O trabalho de coleta e estruturação desses dados, bem como a análise especializada no domínio do problema, foram realizados no trabalho de mestrado (VERSLYPE, 2021) de Nina Iris, na UFRPE. É importante ressaltar que as 251 avaliações mencionadas não constituem uma lista exaustiva de todas as interações possíveis e/ou conhecidas, elas representam apenas a amostra que foi possível coletar no trabalho desenvolvido em (VERSLYPE, 2021). Nessas 251 avaliações, foram identificadas interações entre 40 variedades de uva (utilizadas como enxerto) e 31 variedades de plantas (utilizadas como porta-enxertos). Para fins de organização, esses dados foram consolidados em 3 matrizes de acordo com a informação que representavam, da seguinte forma: matriz de interações, matriz de características das cultivares enxertos e matriz de características das cultivares porta-enxerto.

A matriz de interações \mathcal{I} é uma matriz esparsa $|E| \times |P|$, onde E representa o conjunto de enxertos e P representa o conjunto de porta-enxertos. Cada elemento a_{ij} da matriz \mathcal{I} , tal que $\{a_{ij} \in \mathbb{N} | 1 \leq a_{ij} \leq 5\}$, indica o grau de compatibilidade observado do enxerto i com o porta-enxerto j e pode receber valores inteiros entre 1 e 5, em ordem crescente de afinidade. O grau de compatibilidade também pode ser desconhecido, situação em que a_{ij} assumirá valor nulo na matriz, essas interações com valores nulos serão as candidatas para recomendação. A Tabela 1 apresenta uma amostra da matriz de interações, exibindo algumas interações com grau de compatibilidade conhecido e outras com valor nulo (destacadas em cinza), a versão completa desta matriz está disponível no Apêndice A. Na matriz de interações, graus de compatibilidade maiores que 4 são considerados bons ou excelentes; valores iguais a 3 representam baixa compatibilidade; e valores iguais a 1 e 2 representam estados de total incompatibilidade entre as cultivares.

A matriz de características, ou *features*, dos enxertos \mathcal{E} tem dimensão $|E| \times |C|$, em que E representa o conjunto de enxertos e C representa o conjunto de características dessas cultivares. De modo similar, a matriz de características dos porta-enxertos \mathcal{P} tem dimensão

Tabela 1 – Exemplo da matriz de *ratings* original. As células destacadas em cinza representam as interações desconhecidas

Enxertos	Porta-enxertos				
	1103P	110R	420A	5BB	SO4
BORDO	5		4		
ISABEL	5				
MERLOT	5	5	5		5
SYRAH	5	4	3		4
TANNAT	5		5		5

$|P| \times |C|$, em que P representa o conjunto de porta-enxertos e C representa o conjunto de características dessas cultivares. Nessas duas matrizes, cada elemento e_{ij} registrado na interseção entre $E \times C$ ou $P \times C$, representa o valor da característica j para o enxerto ou porta-enxerto i e esses elementos podem assumir valores numéricos, ordinais ou categóricos, dependendo da natureza da característica a que se referam. A Tabela 2 apresenta uma amostra da matriz de características dos enxertos, a versão completa desta matriz e da matriz de características dos porta-enxertos estão disponíveis nos Apêndices B e C.

Tabela 2 – Exemplo da matriz de *features* dos enxertos

Enxertos	Características dos enxertos			
	INCIDENCIA_ANTRACNOSE	INCIDENCIA_MILDIO	USO	TERMINO_DA_MATURACAO
BORDO	AUSENTE	FRACA	VINHO/SUCO	INTERMEDIARIA
CABERNET FRANC	MEDIA	MEDIA	VINHO	TARDIA
CABERNET SAUVIGNON	MEDIA	MEDIA	VINHO	TARDIA
CARIGNAN	MUITO FORTE	FORTE	VINHO	TARDIA
CENTENNIAL SEEDLESS	FORTE	FORTE	MESA	INTERMEDIARIA
CHARDONNAY	MEDIA	MEDIA	VINHO	INTERMEDIARIA

O conjunto de características C é composto por 6 atributos de plantas, comuns às cultivares enxertos e porta-enxertos, e outras 7 características específicas, de acordo com a finalidade da cultivar no processo de enxertia. Essas características referem-se a aspectos morfológicos, genéticos, imunológicos e do ciclo de vida das cultivares, tais como: espécie, vigor, resistência a nematoides, acidez, etc.

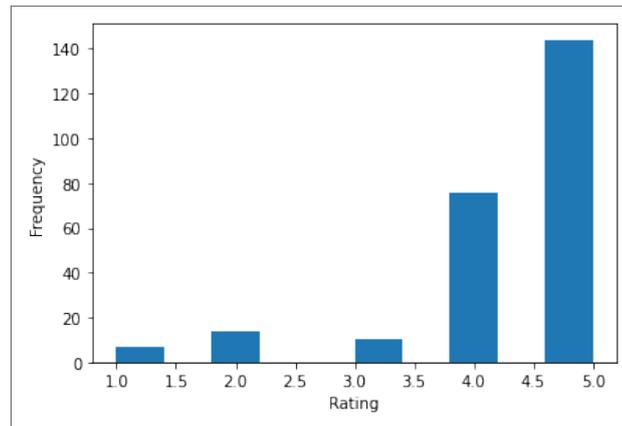
3.2.1 Análise exploratória dos dados

As 251 interações identificadas representam cerca de 20% do universo de 1.240 combinações possíveis entre as 40 cultivares enxerto e as 31 cultivares porta-enxerto que foram identificadas. A hipótese levantada é que talvez, nas quase 80% de combinações inexploradas, possam existir outras interações com alto grau de compatibilidade.

O histograma de distribuição de graus de compatibilidade, ou *ratings*, apresentado na

Figura 7, ilustra o alto nível de desbalanceamento dos dados, pois 220 *ratings* (87,5% do total) concentram-se entre os valores 4 e 5, dos quais 144 são iguais a 5 e 76 são iguais a 4, enquanto os 31 *ratings* restantes são iguais a 1, 2 ou 3.

Figura 7 – Histograma de distribuição das avaliações



Fonte: Elaborada pelo autor (2022)

O *rating* médio de $\approx 4,34$ e o desvio padrão de ≈ 1 refletem numericamente o desbalanceamento observado no histograma de *ratings*, reafirmando o fato de que a maioria dos exemplos observados está concentrada em torno das classificações mais altas. Como no conjunto das 251 interações observadas os *ratings* variam apenas entre 1 e 5, não existem *outliers* em relação a esses valores. A Tabela 3 apresenta outras estatísticas calculadas a partir da matriz de interações \mathcal{I} .

As interações entre as cultivares também podem ser representadas como um grafo bipartido não direcionado \mathcal{G} do tipo $\mathcal{G} = (E, P, A)$, onde E e P são conjuntos de vértices, cada elemento $e_i \in E$ representa um enxerto e cada elementos $p_j \in P$ representa um porta-enxerto, sendo A o conjunto de arestas entre os elementos de E e P , na forma $A \rightarrow \{(e_i, p_j) | a_{ij} \text{ is not missing}\}$. A representação em grafo permite calcular os graus de conectividade entre os nós enxerto e porta-enxerto, demonstrados na Tabela 3, cuja média para o grafo como um todo é de ≈ 7 arestas por vértice, com o mínimo de 1 e o máximo de 23 conexões. Uma densidade de arestas de 0,101 reafirma o fato de que um número baixo de interações é conhecido e o coeficiente de agrupamento médio igual a zero mostra que este grafo não possui *clusters*.

Uma visualização de grafo é apresentada na Figura 8, filtrada para mostrar apenas interações boas ou excelente (maiores ou iguais a 4). Os vértices dos enxertos (círculos na cor verde) com alto grau de conectividade (> 6) estão posicionados à direita, os vértices dos enxertos com baixo grau de conectividade estão posicionados à esquerda, os vértices dos porta-enxertos

Tabela 3 – Estatísticas descritivas do conjunto de dados

Estatísticas	Valores
Total de enxertos (copas)	40
Total de porta-enxertos	31
<i>Ratings</i> conhecidos	251
<i>Ratings</i> desconhecidos	989
<i>Ratings</i> favoráveis (≥ 4)	220
<i>Ratings</i> desfavoráveis (< 4)	31
<i>Rating</i> médio	4.338645
Desvio padrão dos <i>ratings</i>	0.988363
Grau de conectividade médio	7.07
Grau de conectividade máximo dos enxertos	22
Grau de conectividade médio dos enxertos	6.27
Grau de conectividade máximo dos porta-enxertos	23
Grau de conectividade médio dos porta-enxertos	8.1
Densidade de arestas	0.101
Coefficiente de agrupamento médio	0.0

Fonte: Elaborada pelo autor (2022)

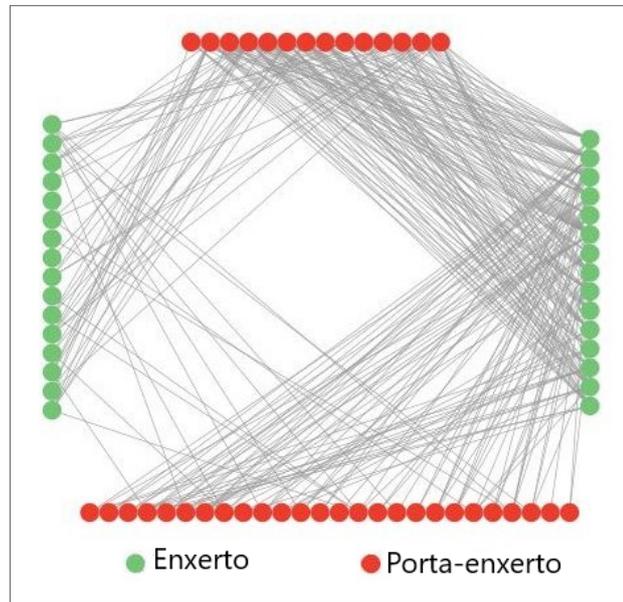
(círculos na cor vermelha) com alto grau de conectividade estão posicionados na parte superior e os vértices dos porta-enxertos com baixo grau de conectividade estão posicionados na parte inferior. A figura sugere a existência de elementos com maior predisposição para formar boas interações.

Os vértices correspondentes aos enxertos e porta-enxertos que possuem maior grau de conectividade (acima e à direita na Figura 8) parecem mais propensos a formar conexões positivas, não apenas entre si, mas também com os vértices de menor grau de conectividade do grupo oposto. Conforme esperado, o número de arestas entre os vértices com baixo grau de conectividade (abaixo e à esquerda na Figura 8) é visivelmente menor. A distribuição de frequência do grau de conectividade no grafo pode ser melhor percebida por meio da Figura 9, onde a maior parte dos vértices possui até 6 arestas.

3.2.2 Pré-processamento e engenharia de *features*

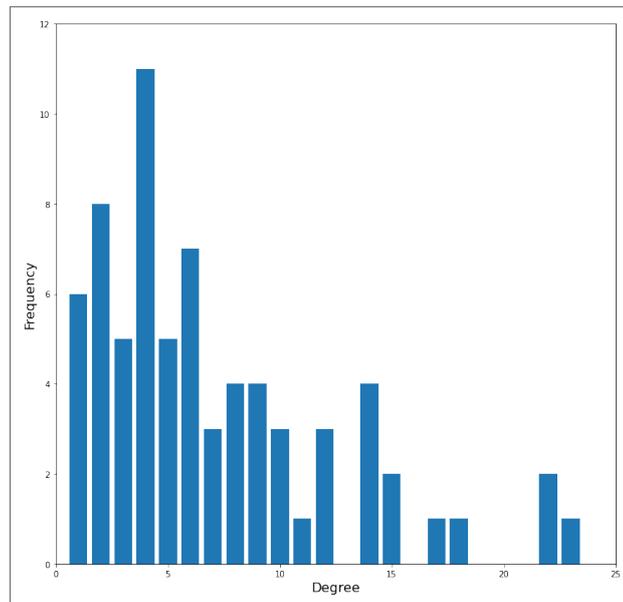
Os dados brutos, consolidados nas matrizes descritas no início da Seção 3.2, foram pré-processados a fim de prepará-los para utilização durante o treinamento dos modelos. Como cada abordagem e/ou classe de algoritmo possui requisitos específicos, as explicações sobre

Figura 8 – Representação em grafo bipartido: subconjuntos de alta-baixa conectividade



Fonte: Elaborada pelo autor (2022)

Figura 9 – Distribuição de graus dos nós do grafo de interação entre enxertos e porta-enxertos



Fonte: Elaborada pelo autor (2022)

as transformações realizadas serão agrupadas para melhor compreensão.

Para a abordagem de **predição de rating**, na qual foram utilizados exclusivamente modelos baseados em algoritmos de FC, foram realizadas as seguintes operações:

1. Transposição da matriz de interações para um formato de tabela com n linhas e 3 colunas, em que cada linha representa uma interação entre enxerto e porta-enxerto e as 3 colunas representam, respectivamente: `userId` (enxerto), `itemId` (porta-enxerto) e

rating;

2. A partir da tabela obtida no item 1, construção de uma nova tabela contendo apenas as interações com ratings avaliados, que servirá como *dataset* para as etapas de treinamento e teste dos modelos (Ver Figura 10);
3. A partir da tabela obtida no item 1, construção de uma nova tabela contendo apenas as interações com ratings nulos, que servirá como base para a predição de ratings dos exemplos não rotulados e posterior recomendação.

Figura 10 – *Dataset* de interações - entrada dos modelos de FC na abordagem de predição de *ratings*

	userID	itemID	rating
0	BENITAKA	5BB	4.0
1	BENITAKA	IAC313	4.0
2	BENITAKA	IAC572	4.0
3	BORDO	101-14MGT	5.0
4	BORDO	1103P	5.0
5	BORDO	420A	4.0
6	BORDO	GRAVESAC	5.0
7	BORDO	IAC313	3.0
8	BORDO	IAC572	5.0
9	BORDO	IAC766	5.0

Fonte: Elaborada pelo autor (2022)

Para a abordagem de **classificação de interações**, na qual foram utilizados modelos baseados em algoritmos de FC e modelos baseados em MK, primeiro, foi necessário binarizar o valor dos ratings na matriz de interações original, uma vez que o objetivo dos modelos era classificar as interações em boas ou ruins. Portanto, as interações cujo *rating* era menor ou igual a 3 (interações ruins) receberam valor 0 e as interações cujo *rating* era maior ou igual a 4 (interações boas) receberam valor 1.

Os *datasets* utilizados para treinamento e teste dos modelos baseados em FC foram construídos seguindo os mesmos passos descritos na abordagem de predição de *rating*. Para os modelos baseados em MK, foram utilizadas as matrizes de dados \mathcal{I} , \mathcal{E} e \mathcal{P} no seu formato original, sendo necessário apenas a realização de engenharia de *features* para tratar valores faltantes e valores categóricos (nominais e ordinais). Na Tabela 4, foram relacionadas as *features* que sofreram alguma modificação.

Tabela 4 – Características das cultivares

ENXERTOS			PORTA-ENXERTOS		
Característica	Tipo	Missing	Característica	Tipo	Missing
INCIDENTARIA_ANTRACNOSE	ORDINAL		INCIDENCIA_ANTRACNOSE	ORDINAL	
INCIDENCIA_MILDIO	ORDINAL		INCIDENCIA_MILDIO	ORDINAL	
CACHO	ORDINAL		RESISTENCIA_NEMATOIDE	ORDINAL	1
VIGOR	ORDINAL	8	RESISTENCIA_FILOXERA	ORDINAL	
USO	NOMINAL		VIGOR	ORDINAL	
TERMINO_DA_MATURACAO	NOMINAL		FLOR	NOMINAL	
ESPECIE	NOMINAL		CICLO	NOMINAL	
PAI1	NOMINAL		ESPECIE	NOMINAL	
PAI2	NOMINAL		PAI1	NOMINAL	
COR_DA_UVA	NOMINAL		PAI2	NOMINAL	
AUSENCIA_DE_SEMENTE	NOMINAL		BROTACAO_EM_DIAS	NUMÉRICO	1
GRAU_BRUX	NUMÉRICO		MATURACAO_EM_DIAS	NUMÉRICO	1
ACIDEZ_TOTAL	NUMÉRICO	2	%_RAIZES_EM_100CM	NUMÉRICO	9

Apenas 6 *features* tiveram valores faltantes, sendo 2 ordinais e 4 numéricas. Como a quantidade de valores faltantes em relação ao total, por *feature*, foi considerada pequena (no pior caso igual a 29% do total de exemplos), a estratégia adotada foi assumir o valor da média para as características numéricas e o valor da moda para as características ordinais. Além disso, foram realizados *encoding* das *features* ordinais e nominais e normalização das *features* numéricas, conforme descrito abaixo:

- *Features* ordinais: mapeadas para valores inteiros em ordem crescente de relevância. Ex.: baixo = 0, médio = 1, alto = 2;
- *Features* nominais: *one-hot encoding* para um array binário de tamanho n , onde n é a quantidade de valores que a *feature* pode assumir e apenas a posição do array que correspondente à característica identificada será preenchida com 1;
- *Features* numéricas: normalização L2 (WEISSTEIN,).

3.3 ESTUDO DE CASO: MÉTODOS

Nesta seção, apresentaremos os métodos de SR adotados neste trabalho para recomendar interações de enxertos e porta-enxertos, bem como as configurações de hiperparâmetros utilizadas.

3.3.1 Modelos baseados em filtragem colaborativa

Neste trabalho, foram utilizadas as implementações de 4 algoritmos de FC disponíveis na biblioteca *SurpriseLib*¹ para Python, uma biblioteca específica para construção de SR. Para a abordagem de predição de ratings, foram utilizados os algoritmos k-NN, SVD, NMF e SlopeOne, e para a abordagem de classificação de interações foram utilizados os algoritmos k-NN, SVD e SlopeOne.

Os hiperparâmetros desses algoritmos foram otimizados através da execução de um método *cross-validated grid-search*, utilizando Erro quadrático médio (RMSE) (para os modelos de predição de rating) e Fração de Pares Concordantes (para os modelos de classificação) como critérios de avaliação para selecionar a melhor combinação de hiperparâmetros. Abaixo, apresentamos os algoritmos de FC utilizados e os respectivos hiperparâmetros selecionados:

- **k-NN**

- **Hiperparâmetros - predição de ratings:**

- * Número máximo de vizinhos: 11;
- * Número mínimo de vizinhos: 5;
- * Função de similaridade: *Mean Squared Difference*;
- * Similaridade baseada em itens;
- * Número mínimo de itens comuns: 2.

- **Hiperparâmetros - classificação de interações:**

- * Número máximo de vizinhos: 9;
- * Número mínimo de vizinhos: 1;
- * Função de similaridade: Cosine (PAGE, 2022);
- * Similaridade baseada em usuário;
- * Número mínimo de itens comuns: 1.

- **SVD:**

- **Hiperparâmetros - predição de ratings:**

- * Número de iterações: 35;

¹ <<http://surpriselib.com/>>

- * Taxa de aprendizado: 0.03;
- * Termo de regularização para os parâmetros: 0.01;
- * Utiliza viés.

– **Hiperparâmetros - classificação de interações:**

- * Número de iterações: 40;
- * Taxa de aprendizado: 0.01;
- * Termo de regularização para os parâmetros: 0.03;
- * Utiliza viés.

▪ **NMF:**

– **Hiperparâmetros - predição de ratings:**

- * Número de iterações: 40;
- * Utiliza viés.

- **SlopeOne:** não há hiperparâmetros para configurar.

3.3.2 Modelos baseados em métodos de Kernel

Conforme apresentado em 2.2.2, MK são um exemplo de algoritmos de filtragem baseada em conteúdo. Neste trabalho, foram utilizadas as implementações de 3 métodos de kernel disponíveis no pacote RLScore², para Python, para calcular matrizes de kernel a partir dos dados de interações entre cultivares e dos dados de características dos enxertos e dos porta-enxertos. Abaixo, apresentamos os MK utilizados e os respectivos hiperparâmetros selecionados:

- **Kernel Linear:** dada a função de kernel expressa na Equação 2.4: Valor utilizado para o hiperparâmetro: $c = 1$;
- **Kernel Gaussiano:** dada a função de kernel expressa na Equação 2.5: Valor utilizado para o hiperparâmetro: $\theta = 1$;
- **Kernel Polinomial:** dada a função de kernel expressa na Equação 2.6: Valores utilizados para os hiperparâmetros: $\alpha = 1$, $c = 0$ e $d = 2$.

² <<http://staff.cs.utu.fi/~aatapa/software/RLScore/index.html>>

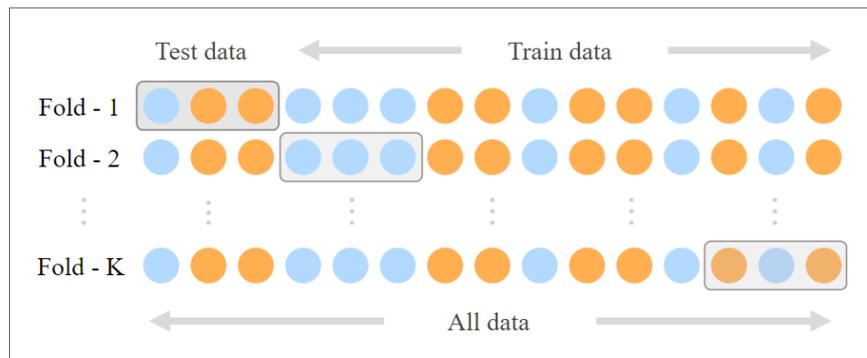
3.4 ESTUDO DE CASO: METODOLOGIA EXPERIMENTAL

Nesta seção, serão apresentados os métodos de reamostragem utilizados nas etapas de treinamento e teste dos modelos e as métricas adotadas para avaliar os resultados obtidos.

3.4.1 Modelos baseados em FC

Dois métodos de reamostragem foram utilizados no processo de treinamento e teste dos modelos baseados em FC: Validação Cruzada k-fold (VC) e *Leave-one-out* (LOO). A VC consiste em particionar aleatoriamente a amostra original em k subamostras de tamanhos iguais, também chamadas de *folds*, conforme ilustrado na Figura 11. O processo de validação cruzada é repetido k vezes, a cada iteração um *fold* diferente é usado como dado de validação e os outros são usados como dados de treinamento. Finalmente, a média de todos os k resultados para qualquer métrica escolhida é usada como estimativa (DEVIJVER PIERRE A.; KITTLER, 1982).

Figura 11 – Validação cruzada k-folds



Fonte: SCHMID (2020)

LOO é um caso particular de VC onde k é igual ao tamanho da amostra original. Neste projeto, o valor k para o método de VC foi definido como 5, enquanto o valor k para o LOO, por definição, teve que ser 251 (número de exemplos rotulados, consultar a Tabela 3).

3.4.2 Modelos baseados em Kernels

Para os modelos baseados em MK, a rotina de treinamento e teste foi realizada por meio de uma função específica, também disponível no pacote RLScore, chamada CGKronRLS. Essa

função é baseada no trabalho publicado em (AIROLA A., 2018) e consiste em um algoritmo Kronecker RLS de treinamento iterativo especialmente projetado para trabalhar com pares de dados de um *dataset* em que a relação entre os pares seja apenas parcialmente conhecida, por exemplo, a matriz de interações utilizada neste trabalho.

3.4.3 Métricas

Dois conjuntos de métricas diferentes foram utilizados de acordo com as características da abordagem em questão. Para a abordagem de **predição de ratings**, as médias das seguintes métricas foram calculadas após 10 execuções independentes de VC e LOO:

- **RMSE e Erro absoluto médio (MAE):** Como os algoritmos de FC utilizados tentam estimar o *rating* das combinações entre as cultivares, por meio do RMSE e do MAE é possível avaliar a acurácia desses modelos nesta tarefa, ou seja, quão próximos os valores estimados estiveram do valor real (CHAI; DRAXLER, 2014);
- **Precisão (*precision*) e revocação (*recall*):** são métricas de desempenho das estimativas, a precisão fornece a porcentagem de exemplos recuperados que são relevantes e a revocação fornece a porcentagem de exemplos relevantes que são recuperados. Essas métricas são mais comuns em problemas de classificação, em que uma matriz de confusão pode ser desenhada para uma classe positiva e outra negativa (POWERS, 2011). Portanto, para utilizá-las na abordagem de predição de *ratings* foi necessário adequar o resultado das estimativas ao conceito das métricas, para tanto, as estimativas de *rating* acima de 3.5 foram consideradas como pertencentes à classe positiva e as outras pertencentes à classe negativa.
- **F1 score:** também conhecido como F-score ou F-measure balanceado, é a média harmônica entre precisão e revocação, resultados próximos a 1 são desejáveis e indicam maior equilíbrio entre precisão e revocação (POWERS, 2011);

Para a abordagem de **classificação binária de interações**, as médias das seguintes métricas foram calculadas após 10 execuções independentes das rotinas de VC, LOO e CGKronRLS, de acordo com os modelos correspondentes:

- **RMSE:** embora a principal tarefa seja dar classificações corretas de boas e más interações entre copas e porta-enxertos, a métrica RMSE foi usada para avaliar a precisão

dos modelos em torno dos valores 0 e 1;

- **Precisão e revocação:** duas métricas de desempenho para avaliar classificadores, precisão indica quão bom o classificador é em **não** rotular uma amostra negativa como positiva, enquanto revocação indica quão bom o classificador é em encontrar todas as amostras positivas (POWERS, 2011);
- **F1 score:** mesmo conceito da métrica descrita para a abordagem de predição de *ratings*;
- **Area Under the ROC curve (AUC-ROC) score:** calcula a relação entre a especificidade e a sensibilidade do classificador, quanto mais próximo de 1 for a pontuação AUC-ROC *score*, melhor o desempenho do modelo na distinção entre as classes positivas e negativas (FAWCETT, 2006).

4 RESULTADOS

Neste capítulo, serão analisados os resultados obtidos pelos modelos de cada abordagem adotada para tratar o problema da identificação de pares compatíveis de enxerto e porta-enxerto. Para facilitar a compreensão das comparações entre os modelos, as análises foram separadas por abordagem, em duas seções.

4.1 ABORDAGEM DE PREDIÇÃO DE RATINGS

Nesta abordagem, foram testados apenas modelos baseados em FC. Os resultados obtidos, de acordo com as métricas mencionadas na Subseção 3.4.3, estão dispostos na Tabela 5. Os melhores resultados gerais para cada métrica estão destacados em amarelo e negrito. Do ponto de vista da acurácia, o modelo que obteve o melhor resultado foi o SVD-LOO (RMSE= 0,908794 e MAE= 0,639274). Em relação à qualidade das predições, o modelo SVD-LOO foi o melhor em relação à precisão (= 0,585435) e o modelo KNN-LOO foi o melhor em relação à revocação (= 0,548207) e ao F1-score (= 0,549594).

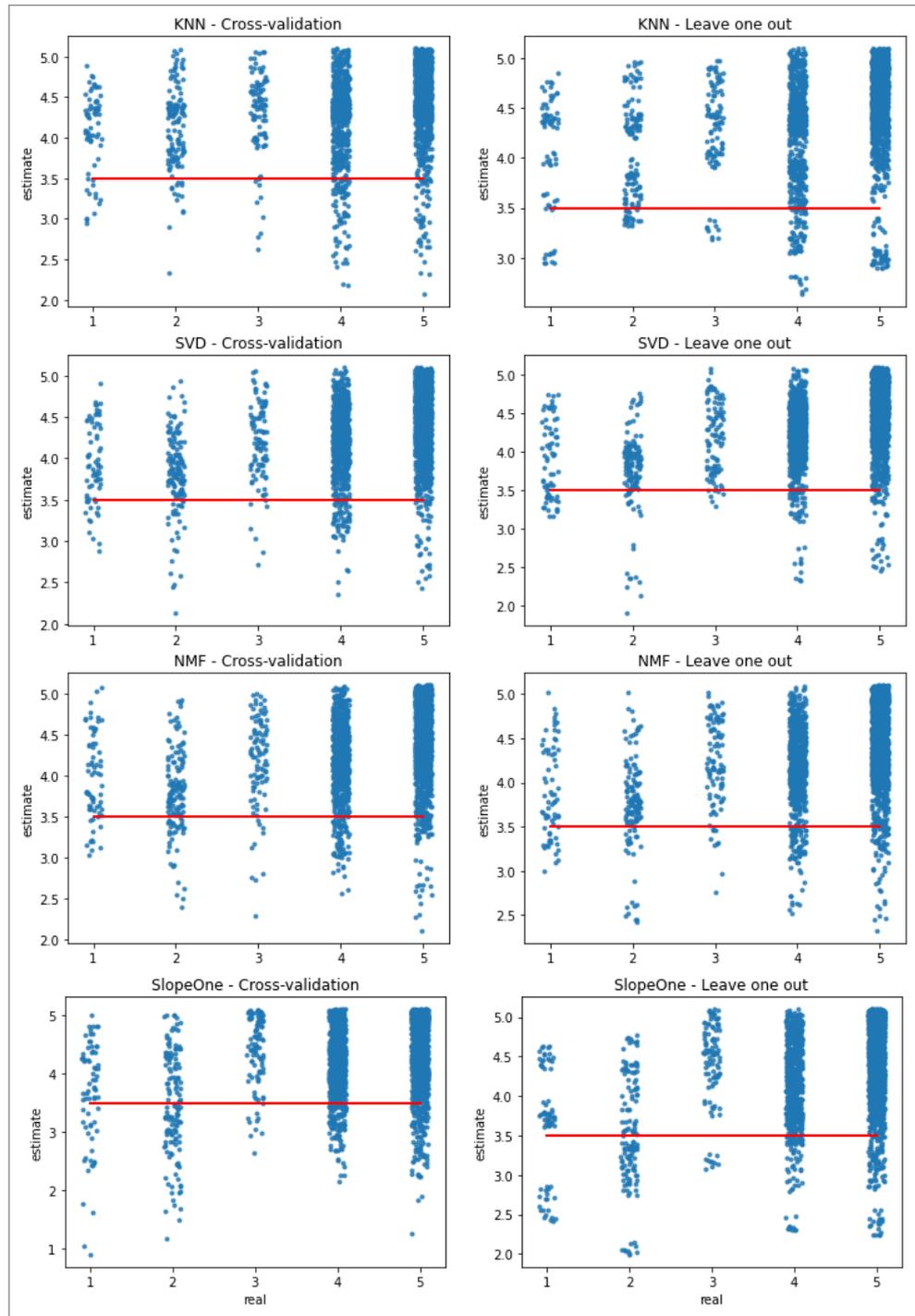
Tabela 5 – Resultados dos modelos - Predição de *ratings*

	KNN		SVD	
	VC	LOO	VC	LOO
MAE	0.675273	0.652598	0.651706	0.639274
RMSE	0.977068	0.951245	0.923663	0.908794
Precision	0.550038	0.550989	0.569431	0.585435
Recall	0.516733	0.548207	0.509163	0.516733
F1	0.532866	0.549594	0.537613	0.548943

	NMF		SlopeOne	
	VC	LOO	VC	LOO
MAE	0.668483	0.656761	0.694417	0.663068
RMSE	0.949563	0.932261	0.992074	0.948315
Precision	0.559021	0.581675	0.555666	0.533929
Recall	0.493227	0.497211	0.483266	0.498007
F1	0.524067	0.536137	0.516943	0.515343

A Figura 12 apresenta os diagramas de dispersão das predições de *ratings* de cada modelo, nessa ilustração é possível visualizar os dois tipos de situações mais onerosas para o problema da enxertia, pois necessariamente produziriam recomendações ruins:

1. *Rating* estimado maior que 3,5 quando o valor real for 1 ou 2;
2. *Rating* estimado menor que 3,5 quando o valor real for 4 ou 5.

Figura 12 – Diagramas de dispersão dos modelos - Predição de *ratings*

Fonte: Elaborada pelo autor (2022)

A situação 1 caracteriza o pior tipo de falso positivo (erro tipo I de uma matriz de confusão) para o problema proposto. Uma vez que as avaliações 1 e 2 desqualificam completamente uma combinação de enxerto e porta-enxerto e não deveriam ser estimadas como 4 ou 5. Da mesma forma, a situação 2 caracteriza o pior tipo de falso negativo (erro tipo II de uma matriz de confusão), neste caso, excelentes combinações poderiam ser perdidas devido a valores

excessivamente subestimados. A Tabela 6, apresenta os valores percentuais da ocorrência dos erros tipo 1 e 2 nos testes de cada modelo, os melhores resultados estão destacados em amarelo.

Tabela 6 – Erros tipo 1 e 2 - Predição de *ratings*

	KNN		SVD		NMF		SlopeOne	
	VC	LOO	VC	LOO	VC	LOO	VC	LOO
Erro 1	86.19%	85.71%	80.95%	79.04%	78.09%	80.00%	54.28%	57.14%
Erro 2	5.95%	5.45%	4.54%	4.04%	6.50%	6.59%	9.81%	7.72%

Com exceção dos modelos que utilizaram SlopeOne, todos os outros incorreram no erro do tipo 1 a uma alta taxa de 78 – 86%, sugerindo que esses modelos tendem a superestimar os *ratings*. Talvez, como consequência do desbalanceamento existente na base de dados. De modo similar, todos os modelos incorreram no erro do tipo 2 com uma taxa relativamente baixa, entre 4 – 10%, reforçando a hipótese de que a tendência predominante é de superestimar os *ratings*. Ainda assim, os modelos que apresentaram o melhor *tradeoff* para esses erros foram os dois que utilizaram SlopeOne, em grande parte por terem sido mais eficientes em não superestimar os *ratings*.

4.1.1 Recomendações

Por fim, foram realizadas as estimativas de *ratings* para as interações desconhecidas e os 5 porta-enxertos que obtiveram a maior pontuação foram recomendados, em ordem decrescente do valor estimado. No Apêndice D, estão disponíveis as top-5 recomendações de porta-enxertos dos melhores modelos. Essas recomendações foram submetidas à análise da especialista Nina Iris Verslype, doutoranda na UFRPE, e algumas observações interessantes puderam ser extraídas, são elas:

1. As cultivares enxerto **Niagara Branca e Niagara Rosada**, tiveram, em média, 3 porta-enxertos em comum entre as 5 primeiras recomendações, para cada modelo. Ainda, 2 dessas recomendações foram praticamente unânimes entre todos os modelos, os porta-enxertos 110R e 3309C. Uma provável explicação é que ambas as cultivares possuem características semelhantes, considerando que Niagara Rosada é uma mutação de Niagara Branca;

2. Outro caso interessante foi o do enxerto **Marselan**, originado do cruzamento entre as cultivares **Grenache e Cabernet Sauvignon**, que teve, em média, 3 porta-enxertos em comum com a cultivar Grenache e 2 porta-enxertos em comum com a cultivar Cabernet Sauvignon.
3. Situação semelhante ocorreu com as copas **Brs Melodia e Brs Vitória**, originárias do cruzamento das cultivares **CNPUV 681-29 e BRS Linda**, elas (Brs Melodia e Brs Vitória) compartilharam, em média, quatro dos cinco porta-enxertos mais recomendados.
4. Os resultados também demonstram a complexidade das relações entre os dados e da tarefa de recomendação, uma vez que a cultivar Carignan que, mesmo possuindo 4 interações conhecidas com *ratings* 4 ou 5, recebeu apenas 8 recomendações com estimativas acima de 4, das 20 possíveis se forem considerando todos os modelos.

É importante ressaltar que essas recomendações não foram realizadas com o objetivo principal de validar as predições dos modelos para o conjunto de interações desconhecidas, para tanto, seria necessário uma avaliação muito mais ampla. Sendo assim, essas recomendações serviram principalmente para demonstrar a aplicabilidade dos modelos em um SR e para avaliar, superficialmente, se as recomendações demonstram alguma coerência.

Diante da análise dos resultados apresentada e considerando apenas os algoritmos e o conjunto de dados utilizados nesses experimentos, não houve um modelo que se destacasse absolutamente em relação aos demais. Desta forma, um SR que utilizasse a abordagem de predição de *ratings* baseada em algoritmos de FC, poderia, por exemplo, adotar o critério do voto majoritário entre as recomendações de diferentes modelos. Assim, as vantagens e desvantagens de cada modelo poderiam ser equalizadas e o consenso obtido seria a recomendação final.

4.2 ABORDAGEM DE CLASSIFICAÇÃO DE INTERAÇÕES

A Tabela 7 apresenta os resultados de cada modelo para as métricas mencionadas na Subseção 3.4.3, os melhores resultados gerais para cada métrica estão destacados em negrito e amarelo. Em relação ao RMSE, o modelo que obteve o melhor resultado foi aquele treinado com Kernel Polinomial (0,315074). Em relação a performance de classificação, o modelo SlopeOne-LOO obteve o melhor resultado de precisão (0,954545), mas ao custo de uma revocação muito

baixa (0,544090). O modelo treinado com Kernel Gaussiano obteve os melhores resultados para revocação, f1-score e AUC-ROC score, também obteve um valor alto para precisão 0,938462, mesmo a sua pontuação para RMSE (0,318331) foi aproximadamente igual a do modelo Kernel Polinomial.

Tabela 7 – Resultados do modelos - Classificação de interações

Modelos baseados em FC						
	KNN		SVD		SlopeOne	
	VC	LOO	VC	LOO	VC	LOO
RMSE	0.342427	0.324082	0.327919	0.322533	0.346852	0.329052
Precision	0.914859	0.945553	0.931515	0.931670	0.954545	0.954198
Recall	0.786363	0.710454	0.636818	0.780909	0.544090	0.568181
F1-score	0.845758	0.811315	0.756479	0.849653	0.693109	0.712250
AUC-ROC score	0.629300	0.698027	0.701468	0.726308	0.716509	0.727639

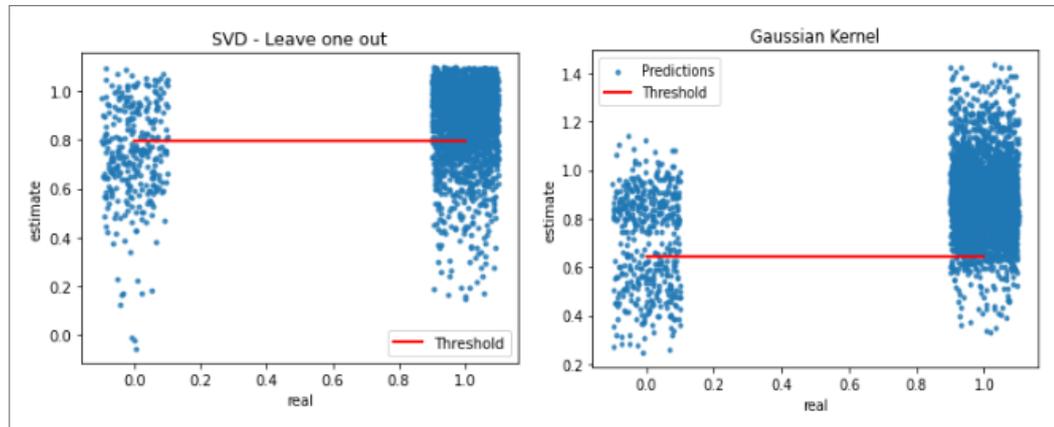
Modelos baseados em MK			
	Linear Kernel	Gaussian Kernel	Polynomial Kernel
	RMSE	0.329891	0.318331
Precision	0.933333	0.938462	0.935484
Recall	0.903226	0.983871	0.935484
F1-score	0.918033	0.960630	0.935484
AUC-ROC score	0.729391	0.769713	0.745520

Como, em média, os resultados dos modelos pertencentes à mesma categoria (FC ou MK) foram muito semelhantes, selecionamos o melhor modelo de cada categoria e realizamos uma investigação mais profunda das diferenças de desempenho. O modelo baseado em Kernel Gaussiano foi escolhido entre os métodos baseados em kernel, devido aos resultados discutidos acima. Por sua vez, o modelo SVD-LOO foi considerado o melhor modelo dentro da classe de algoritmos de FC, considerando os resultados globais disponíveis na Tabela 7.

A Figura 13 apresenta os diagramas de dispersão das classificações dos modelos SVD-LOO e Kernel Gaussiano sobre o conjunto de teste (Os diagramas de dispersão de todos os modelos estão disponíveis no Apêndice E). Como a saída dos classificadores é dada em termos de aproximações ou probabilidades, no espaço contínuo, em relação às classes alvo 0 (interações ruins) e 1 (interações boas). Portanto, um limiar (*threshold*) foi calculado usando a curva ROC e a estatística J de Youden, conforme descrito em (BROWNLEE, 2020), para estabelecer o limite entre as classes preditas. Assim, as previsões situadas abaixo da linha de limite pertencem à classe 0 e as previsões acima da linha de limite pertencem à classe 1.

É possível visualizar através da Figura 13 a diferença entre os classificadores Kernel Gaussiano e SVD-LOO na tarefa de identificar exemplos da classe positiva. A métrica de recall expressa essa diferença em números (consultar a Tabela 7), pois os modelos baseados em

Figura 13 – Diagramas de dispersão dos modelos - Classificação de interações

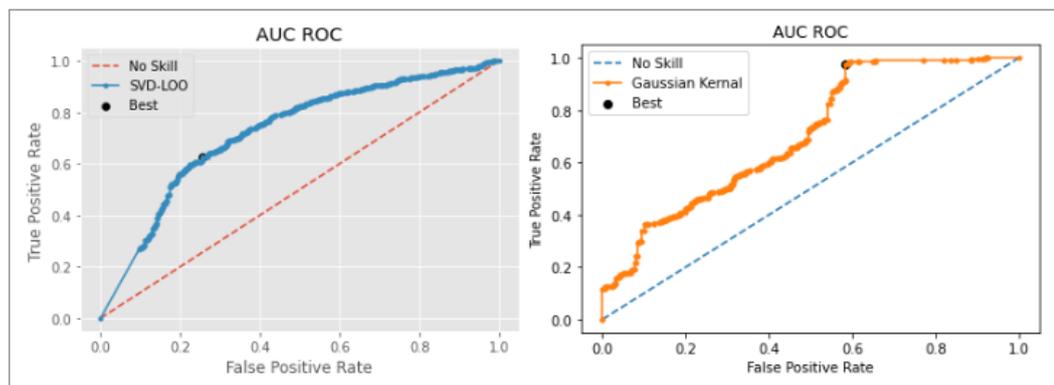


Fonte: Elaborada pelo autor (2022)

kernel tiveram pontuações de recall mais altas. O mesmo padrão foi observado para os outros modelos baseados em kernel e FC. Olhando para a questão de outra perspectiva, os modelos baseados em FC cometeram Falsos Negativos (FN) a uma taxa entre 40 – 50%, enquanto os modelos baseados em kernel entre 3 – 10%.

A Figura 14 mostra as curvas ROC dos modelos SVD-LOO e de Kernel Gaussiano, por meio das quais é possível visualizar a relação entre a Taxa de Verdadeiro Positivo (TVP) e a Taxa de Falso Positivo (TFP) obtida por esses classificadores (As curvas ROC de todos os modelos estão disponíveis no Apêndice F). Mais uma vez, o padrão foi semelhante para os outros modelos baseados em FC e MK estudados. Ao contrário do que sugere a análise isolada da métrica AUC-ROC score (consultar Tabela 7), que apresenta uma diferença relativamente baixa para essa métrica entre os 2 modelos, inferior a 0,05, a análise gráfica da curva ROC indica que o *tradeoff* entre TVP e TFP é mais custoso para o modelo SVD-LOO, pois um pequeno aumento na TVP implica em um grande aumento na TFP.

Figura 14 – Curva ROC - Classificação de interações



Fonte: Elaborada pelo autor (2022)

Portanto, os resultados obtidos indicam uma vantagem para os modelos que utilizaram o algoritmo SVD, dentre os modelos baseados em CF, em um contexto onde as características das entradas não foram utilizadas. Além disso, dentre os modelos baseados em métodos de kernel, que aproveitaram as características das entradas, aquele que utilizou o kernel gaussiano apresentou a melhor resposta. Uma hipótese que os resultados podem sugerir é que os modelos baseados em métodos de kernel com características de entrada foram melhores do que aqueles baseados em CF. No entanto, como os métodos de reamostragem utilizados foram diferentes para as duas classes de algoritmos, não é possível afirmar isso com total certeza.

Não fez parte do escopo deste trabalho avaliar o efeito do desbalanceamento dos dados no resultado dos modelos, principalmente porque essa é uma característica esperada em conjuntos de dados para este domínio de problema. Portanto, como regra, seria desejável que o mecanismo preditivo de um SR de porta-enxertos fosse capaz de ter um bom desempenho, mesmo que o desbalanceamento dos dados fosse grande e que houvesse poucos exemplos rotulados disponíveis.

5 CONCLUSÃO

No presente trabalho, utilizamos uma abordagem de SR para analisar o desempenho de algoritmos de FC e métodos baseados em kernel para resolver o problema de identificação de pares compatíveis de enxertos e porta-enxertos em viticultura. Esta abordagem foi pouco explorada até então, provavelmente inédita, neste domínio. Em nosso estudo, duas perspectivas foram exploradas para tratar o problema: predição de ratings e classificação binária de interações.

Na perspectiva de predição de *ratings*, os resultados obtidos demonstram que não houve um modelo que se destacasse em absoluto em relação aos demais. Ainda assim, foi possível observar que as recomendações produzidas por esses modelos sobre as interações desconhecidas possuem coerência em relação aos valores esperados.

Na perspectiva de classificação binária de interações, os resultados obtidos, considerando exclusivamente os dados disponíveis e os métodos escolhidos, sugerem que, em um contexto em que as *features* das entradas não estão disponíveis, classificadores baseados em FC usando algoritmos SVD são a melhor escolha. No entanto, se as *features* das entradas estiverem disponíveis, os classificadores baseados em Kernel Gaussiano são preferíveis. Além disso, o f1-score de 96% para o modelo de Kernel Gaussiano e 85% para o modelo SVD-LOO são bons indicadores de que eles poderiam ser testados em uma ferramenta de apoio à decisão para pesquisadores da área de viticultura.

O objetivo deste trabalho não foi esgotar as possibilidades e encontrar a melhor solução possível para o problema em análise, mas demonstrar a viabilidade de um caminho alternativo através de AM e oferecer uma comparação entre duas técnicas bem conhecidas. Os resultados obtidos são encorajadores, vide o F1-score de aproximadamente 96% do modelo de Kernel Gaussiano, um bom indicador de que esse modelo poderia servir como motor preditivo de um SR de porta-enxertos na viticultura.

Obviamente, outras abordagens e métodos poderiam ter sido aplicados e serão listados, aqui, como sugestões de trabalhos futuros:

- Aplicar técnicas para mitigar os efeitos do desbalanceamento dos dados;
- Testar outras classes de algoritmos;
- Utilizar aprendizagem com múltiplos kernels;

- Abordar o problema sob a perspectiva da teoria dos grafos, como um problema de predição de links, no qual o objetivo seria prever as arestas faltantes e o seu peso;
- Acrescentar mais dados de interações entre enxertos e porta-enxertos, inclusive de outros tipos de cultivo, ao *dataset* existente, bem como as *features* dessas cultivares. O objetivo seria construir modelos preditivos de propósito geral para processos de enxertia;
- Avaliar o desempenho de modelos sensíveis ao contexto que, além de utilizarem as *features* das entradas, também pudessem considerar informações ambientais capazes de interferir no sucesso da enxertia, como clima, solo e relevo.

Sistema de Recomendação é uma técnica poderosa e versátil. Tentamos através deste trabalho explorá-lo em um contexto diferente, ainda carente de referências na literatura, e escolhemos um problema, de inegável relevância econômica e social, para oferecer um método baseado em aprendizado de máquina para planejamento de processos de enxertia agrícola. Por fim, o desenvolvimento deste trabalho de mestrado resultou em um artigo intitulado "Predicting Compatibility of Cultivars in Grafting Processes using Kernel Methods and Collaborative Filtering", publicado no *11th Brazilian Conference on Intelligent Systems (BRACIS 2022)*.

REFERÊNCIAS

- AGGARWAL, C. C. *Recommender Systems: The Textbook*. [S.l.]: Springer, 2016.
- AIROLA A., P. T. Fast kronecker product kernel methods via generalized vec trick. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 29, Issue 8, p. 3374–3387, 2018.
- AL., D. R. S. et. Estimating the efficiency of machine learning in forecasting harvesting time of rice. *International Journal of Modern Agriculture*, p. 1930–1937, 2021. Disponível em: <<http://www.modern-journals.com/index.php/ijma/article/view/940>>.
- BONDRE D. A., M. S. Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *International Journal of Engineering Applied Sciences and Technology*, 2019.
- BREESE J.S.; HECKERMAN, D. K. C. Empirical analysis of predictive algorithms for collaborative filtering. *Fourteenth conference on Uncertainty in artificial intelligence*, 1998.
- BROWNLEE, J. *Imbalanced classification with python: Better metrics, balance skewed classes, cost-sensitive learning*. [S.l.]: Machine Learning Mastery, 2020.
- CANTU D., W. M. *The grape genome*. [S.l.]: Springer, 2019.
- CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae) - arguments against avoiding rmse in the literature. *Geoscientific model development*, Copernicus GmbH, v. 7, n. 3, p. 1247–1250, 2014.
- COELHO, A. *Vinhos: você sabia que*. 2012. Disponível em: <<http://www.decantandoavida.com/2012/01/13/vinhos-voce-sabia-que-3>>. Acesso em: 30 august 2022.
- D., W. J. M. C. *Machine Learning for Hackers*. [S.l.]: O'Reilly, 2012.
- DALBÓ M.A., F. N. Comportamento agrônomo de porta-enxertos de videira com resistência ao declínio de plantas jovens nas condições do estado de santa catarina. *Agropecuária Catarinense, Florianópolis*, v.32, n.2, p. 68–72, 2019. Disponível em: <<http://dx.doi.org/10.22491/RAC.2019.v32n2.10>>.
- DEVIJVER PIERRE A.; KITTLER, J. *Pattern Recognition: A Statistical Approach*. [S.l.]: Prentice-Hall, 1982.
- DOUGHERTY, P. *The geography of wine regions, terroir and techniques*. [S.l.]: Springer Science and Business Media, 2012.
- EINSTEIN, A. 1947.
- F. ROKACH L., S. B. R. *Recommender Systems Handbook*. [S.l.]: Springer, 2020.
- FAO - Food and Agriculture Organization of the United Nations. 2021. Disponível em: <<http://www.Fao.org/faostat/en/#data/QC/visualize>>. Acesso em: 16 Jun 2021.
- FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters* 27, p. 861–874, 2006.

- GENTON, M. G. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research* 2, p. 299–312, 2001. Disponível em: <<https://jmlr.csail.mit.edu/papers/volume2/genton01a/genton01a.pdf>>.
- H. EL-BENDARY N., F. E. E. M. O. Ontology based recommendation system for predicting cultivation and harvesting timings using support vector regression. *Computational Methods in Systems and Software*, 2020.
- HERNANDES J.L., M.-F. P. J. M. *Uso de porta-enxertos: Tecnologia simples e fundamental na cultura da videira*. [S.l.]: Instituto Agrônômico de Campinas, 2011.
- HOFMANN T.; SCHOLKOPF, B. S. A. J. Kernel methods in machine learning. *The Annals of Statistics*, 2008.
- JAISWAL S., K. T. K. N. S. S. Collaborative recommendation system for agriculture sector. *ITM Web of Conferences*, 2020.
- KOHAVI R.; WOLPERT, D. H. Bias plus variance decomposition for zero-one loss functions. *ICML*, 1996.
- LACASTA J., L.-P. F. E.-G. B. N.-I. J. Z.-S. F. Agricultural recommendation system for crop protection. *Computers and Electronics in Agriculture*, p. 82–89, 2018.
- LE, J. *The 4 Recommendation Engines That Can Predict Your Movie Tastes*. 2018. Disponível em: <<https://towardsdatascience.com/the-4-recommendation-engines-that-can-predict-your-movie-tastes-109dc4e10c52>>. Acesso em: 28 may 2022.
- LEAO P.C.S., B.-R. Melhoramento genético da videira. *Embrapa Semiárido-Documentos*, 2009.
- LEMIRE D., M.-A. Slope one predictors for online rating-based collaborative filtering. *Proceedings of the SIAM Data Mining Conference*, 2005.
- LEÃO, P. d. S. *Cultivo da Videira - Produção de mudas*. 2004. Disponível em: <http://www.cpatsa.embrapa.br:8080/sistema_producao/spvideira/mudas.htm>. Acesso em: 07 jun 2022.
- MENG C.; XU, D. S. Y.-J. . K.-C. Simulation-based economic feasibility analysis of grafting technology for propagation operation. *Proceedings of the 2012 Industrial and Systems Engineering Research Conference. IIE Annual Conference*, 2012.
- MOHANTY S., C. J. J. S.-E. A. G.-P. *Recommender System with Machine Learning and Artificial Intelligence*. [S.l.]: Wiley-Scrivener, 2020.
- MOKARRAMA M.J., A. M. A recommendation system for farmers. *5th IEEE Region 10 Humanitarian Technology Conference*, p. 843–850, 2018.
- MOVIE Recommendation and Rating Prediction using K-Nearest Neighbors. 2018. Disponível em: <<https://www.analyticsvidhya.com/blog/2020/08/recommendation-system-k-nearest-neighbors/>>. Acesso em: 25 june 2022.
- MUDGE K.; JANICK, J. S. S. G. A history of grafting. *Horticultural Reviews*, p. 449–475, 2009. Disponível em: <[doi:10.1002/9780470593776.ch9](https://doi.org/10.1002/9780470593776.ch9)>.

- PAGE, M. L. *Farmers may have been accidentally making GMOs for millennia*. 2022. Disponível em: <<https://www.newscientist.com/article/2079813-farmers-may-have-been-accidentally-making-gmos-for-millennia/>>. Acesso em: 18 apr. 2022.
- PATEL K., P. H. A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 2020.
- PEIL, R. M. A enxertia na produção de mudas de hortaliças. *Ciência Rural, Santa Maria*, v.33, n.6, p. 1169–1177, 2003. Disponível em: <<http://www.abhorticultura.com.br/downloads/enxertiaHort.pdf>>.
- PORTA-ENXERTO e cultivares. 2021. Disponível em: <<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Uva/UvaAmericanaHibridaClimaTemperado/cultivar.htm>>. Acesso em: 22 Mai 2021.
- POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies* 2, p. 37–63, 2011.
- PUDUMALAR S., R. E. R. R. K. C. K.-T. N. J. Crop recommendation system for precision agriculture. *8th International Conference on Advanced Computing*, p. 32–36, 2017.
- RIBEIRO G.D.; COSTA, J. V. A. S. M. Enxertia em fruteiras. *EMBRAPA - Recomendações Técnicas*, 2005. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/859550/1/rt92enxertiadefruteiras.pdf>>.
- SARWAR G. KARYPIS, J. A. K. B. M.; RIEDL, J. T. Application of dimensionality reduction in recommender system-a case study. *ACM WebKDD 2000 Web Mining for E-commerce Workshop*, 2000.
- SCHMID, P. *K-Fold as Cross-Validation with a BERT Text-Classification Example*. 2020. Disponível em: <<https://www.philschmid.de/k-fold-as-cross-validation-with-a-bert-text-classification-example>>. Acesso em: 22 apr. 2022.
- SCHÖLKOPF B.; SMOLA, A. J. B. F. *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. [S.l.]: MIT Press, 2018.
- SERRA I., S. A. M. P. D. A. The interaction between rootstocks and cultivars (*Vitis vinifera* l.) to enhance drought tolerance in grapevine. *Australian Journal of Grape and Wine Research*, 2014.
- SOARES J.M., L. P. *A vitivinicultura no Semiárido Brasileiro*. [S.l.]: Embrapa Semiárido, 2009.
- THEODORIDIS, S. *Pattern Recognition*. [S.l.]: Elsevier B.V., 2008.
- VERSLYPE, N. Avaliação e seleção de porta-enxertos de videira (*vitis* spp.) tolerantes ao déficit hídrico através de aprendizagem de máquina. *Dissertação (Mestrado) - Curso de Pós-Graduação em Agronomia: Melhoramento Genético de Plantas, Universidade Federal Rural de Pernambuco*, p. 140, 2021.
- VITIVINICULTURA, U. B. de. *UVIBRA*. 2022. Disponível em: <<http://www.uvibra.com.br/>>. Acesso em: 8 apr. 2022.

VIVEK M.V.R., H. D. A survey on crop recommendation using machine learning. *International Journal of Recent Technology and Engineering*, 2019.

WEISSTEIN, E. W. *L2-Norm*. Disponível em: <<https://mathworld.wolfram.com/L2-Norm.html>>. Acesso em: 9 september 2022.

WILIMITIS, D. *The Kernel Trick in Support Vector Classification*. 2018. Disponível em: <<https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>>. Acesso em: 25 june 2022.

ZHANG L., M. E. R. L. O. N. . G. G. A. The influence of grapevine rootstocks on scion growth and drought resistance. *Theoretical and Experimental Plant Physiology*, 2016.

APÊNDICE A – MATRIZ DE INTERAÇÕES

ENXERTOS	PORTA-ENXERTOS																															
	101-14MGT	1103P	110R	1202C	140RU	161-49C	1613C	3306C	3309C	41B	420A	44-53M	5BB	5C	99R	AXR#1	DOGRIDGE	FERCAL	FREEDOM	GOLIA	GRAVESAC	HARMONY	IAC313	IAC572	IAC766	RAMSEY	RIPARIA	TRAVIU	RUPESTRIS	SCHWARZMANT	SO4	
BENITAKA													4										4	4								
BORDO	5	5									4										5		3	5	5			4				
CABERNET FRANC	5	5						5		5			1																			5
CABERNET SAUVIGNON	5	5	5	5	5	5		5		5		4	4	5	5	5	5			5	3				5	4	5		5	4	4	
CARIGNAN			5		2	2																									1	
CENTENNIAL SEEDLESS											4												4	4	5							
CHARDONNAY	5	5	5	5		5			5		5										5							5			5	
CHENIN	4						4								4		4									4	5					
COLOMBARD	5	5				5			5																							
CRIMSON SEEDLESS		5	5				4				4		1						5			4		2	5	5						
FANTASY SEEDLESS											4														5							
GRENACHE	5		5			2						5																	4			
ISABEL		5																						4								
ITALIA											4		4										5	4	5							
MAGNA		4																	5			4	4	4	4						4	
MALBEC		5			5				4			1										5									5	
MARROO SEEDLESS		4					5				4												4		5							
MARSELAM			5			2																										
MAXIMO		3																			3			5	4	5			3			
MELODIA		4																														
MERLOT	5	5	5			5			5	4	5			5													4	5		4	5	
MOSCATO CANELLI																			5			5										
MOURVÈDRE						2																										
NIAGARA BRANCA											5														5			4				
NIAGARA ROSADA	5	5									5		4								5		5	5	5			4		4		
PATRICIA												4											5		5			4				
PAULISTINHA																									5			4				
PETIT VERDOT	5																						4									
PINOT NOIR	5	5				5			5	4	1																				5	
RED GLOBE		5					5													5			5	4	4	5	5					
RIESLING		5							5						4																5	
SAUVIGNON						2				5																			5			
SUGRAONE		5	5								2									5			4	5	2		5				4	
SYRAH	3	5	4		3	4	5		5	4	3				2		5	2				4	5			5				5	4	
TANNAT		5				5				5	5																	5				5
TEMPRANILLO	4	2	5		5	5			5	5	5											4		4		4					5	
THOMPSON SEEDLESS	5	5	5								4		4		5		3		5			4	4	2		4			4		3	
VENUS											5														5							
VIOLETA		4									5														4	4						
VITORIA		5									5									5			5	5	4	5						4

APÊNDICE B – MATRIZ DE CARACTERÍSTICAS DOS ENXERTOS

Enxertos	Características												
	INCIDENCIA ANTRACNOSE	INCIDENCIA MILDIO	USO	TERMINO DA MATURACAO	ESPECIE	PAI1	PAI2	COR DA UVA	AUSENCIA DE SEMENTE	GRAU BRIX	ACIDEZ TOTAL	CACHO	VIGOR
BENITAKA	MEDIA	MEDIA	MESA	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	VERMELHA	NÃO	14,1	121	GRANDE	ALTO
BORDO	AUSENTE	FRACA	VINHO/SUCO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	15,1	91	PEQUENO	MEDIO
CABERNET FRANC	MEDIA	MEDIA	VINHO	TARDIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	20,5	83	PEQUENO	ALTO
CABERNET SAUVIGNON	MEDIA	MEDIA	VINHO	TARDIA	Vitis vinifera	CABERNET FRANC	SAUVIGNON	PRETA	NÃO	18,9	118	MEDIO	ALTO
CARIGNAN	MUITO FORTE	FORTE	VINHO	TARDIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	18	133	GRANDE	ALTO
CENTENNIAL SEEDLESS	FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	GOLD	Q25-6	BRANCA	SIM	16,4	116	GRANDE	MUITO ALTO
CHARDONNAY	MEDIA	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	HEUNISCH WEISS	PINOT NOIR	BRANCA	NÃO	17,8	168	PEQUENO	MEDIO
CHENIN BLANC	FORTE	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	BRANCA	NÃO	18,9	152	MEDIO	ALTO
COLOBARD	FORTE	FORTE	VINHO	INTERMEDIARIA	Vitis vinifera	HEUNISCH	CHENIN	BRANCA	NÃO	18,5	154	MEDIO	MUITO ALTO
CRIMSON SEEDLESS	MUITO FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	EMPEROR	C33-199	VERMELHA	SIM	16,3	123	MEDIO	ALTO
FANTASY SEEDLESS	MUITO FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	B36-27	C78-68	PRETA	SIM	18	104	MEDIO	MUITO ALTO
GRENACHE	FORTE	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	19,7	87	MEDIO	MUITO ALTO
ISABEL	AUSENTE	MEDIA	VINHO/SUCO/MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	LABRUSCA	MESLIER PETIT	PRETA	NÃO	18,6	52	PEQUENO	ALTO
ITALIA	FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	BICANE	MUSCAT HAMBURG	BRANCA	NÃO	17,3	133	GRANDE	ALTO
MAGNA	MEDIA	MEDIA	SUCO	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	BRS Rúbea	TRAVIU	PRETA	NÃO	19	90	MEDIO	MEDIO
MALBEC	MEDIA	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	MAGDELEINE NOIR DES CHARENTES	PRUNELARD	PRETA	NÃO	17,4	103	PEQUENO	NaN
MARROO SEEDLESS	FORTE	MEDIA	MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	CAROLINA BLACKROSE	RUBY SEEDLESS	PRETA	SIM	14,8	112	GRANDE	NaN
MARSELAM	FRACA	FORTE	VINHO/MESA	INTERMEDIARIA	Vitis vinifera	CABERNET SAUVIGNON	GRENACHE	PRETA	NÃO	19,8	117	MEDIO	NaN
MAXIMO	MEDIA	FRACA	SUCO	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	SEIBEL 11-342	SYRAH	PRETA	SIM	14,2	153	MEDIO	NaN
MELODIA	FRACA	FORTE	MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	CNPUV 681-29	BRS Linda	ROSADA	SIM	11	NaN	GRANDE	MEDIO
MERLOT	FRACA	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	16,4	112	MEDIO	ALTO-MEDIO
MOSCATO CANELLI	FORTE	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	BRANCA	NÃO	18,2	97	PEQUENO	NaN
MOURVÈDRE	FRACA	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	14,2	133	MEDIO	ALTO-MEDIO
NIAGARA BRANCA	AUSENTE	FRACA	VINHO/MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	CONCORD	CASSADY	BRANCA	NÃO	16,9	62	MEDIO	MEDIO
NIAGARA ROSADA	AUSENTE	FRACA	VINHO/MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	NIAGARA MUTATION	NIAGARA MUTATION	ROSADA	NÃO	16,4	61	MEDIO	MEDIO
PATRICIA	FRACA	FORTE	MESA	TARDIA	VITIS INTERSPECIFIC CROSSING	SORAYA	IAC544-14	VERMELHA	NÃO	18,9	110	GRANDE	NaN
PAULISTINHA	FORTE	MEDIA	MESA	INTERMEDIARIA	VITIS INTERSPECIFIC CROSSING	NIAGRA BRANCA	SULTANINA	BRANCA	SIM	19	91	MEDIO	NaN
PETIT VERDOT	MEDIA	MEDIA	VINHO	TARDIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	20,6	152	MEDIO	NaN
PINOT NOIR	FRACA	FORTE	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	17,2	153	PEQUENO	MEDIO-BAIXO
RED GLOBE	MUITO FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	L12-80	S45-48	ROSADA	NÃO	15,3	105	GRANDE	ALTO-MEDIO
RIESLING	MEDIA	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	BRANCA	NÃO	17,6	109	PEQUENO	MEDIO
SAUVIGNON	FORTE	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	18,9	118	MEDIO	ALTO
SUGRAONE	MUITO FORTE	MEDIA	MESA	PRECOCE	Vitis vinifera	Vitis vinifera	Vitis vinifera	BRANCA	SIM	14,8	115	MEDIO	ALTO
SYRAH	MEDIA	FORTE	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	17,8	122	MEDIO	ALTO
TANNAT	MEDIA	FORTE	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	20,3	183	MEDIO	MEDIO
TEMPRANILLO	MUITO FORTE	MEDIA	VINHO	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	NÃO	17,4	126	MEDIO	ALTO
THOMPSON SEEDLESS	MUITO FORTE	FORTE	MESA	INTERMEDIARIA	Vitis vinifera	Vitis vinifera	Vitis vinifera	BRANCA	SIM	21,3	145	MEDIO	ALTO
VENUS	FORTE	MEDIA	MESA	PRECOCE	Vitis vinifera	Vitis vinifera	Vitis vinifera	PRETA	SIM	17,1	83	MEDIO	ALTO

VIOLETA	FRACA	MEDIA	VINHO/S UCO	PRECOCE	VITIS INTERSPECIFIC CROSSING	BRS Rúbea	IAC 1398-21	PRETA	NÃO	21	60	MEDIO	MEDIO
VITORIA	FRACA	FORTE	MESA	PRECOCE	VITIS INTERSPECIFIC CROSSING	CNPUV 681- 29	BRS Linda	PRETA	SIM	23	NaN	MEDIO	ALTO

APÊNDICE C – MATRIZ DE CARACTERÍSTICAS DOS PORTA-ENXERTOS

Enxertos	Características												
	INCIDENCIA ANTRACNOSE	INCIDENCIA MILDIO	FLOR	CICLO	BROTACAO EM DIAS	MATURACAO EM DIAS	ESPECIE	PAI1	PAI2	RESISTENCIA NEMATÓIDE	RESISTENCIA FLOXERA	VIGOR	% RAIZES EM 100CM
101-14MGT	MUITO FORTE	FRACA	FEMININA	INTERMEDIARIA	129,00	120,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	Vitis rupestris	MEDIA	MEDIA	MEDIO	93,6
1103P	FORTE	FRACA	MASCULINA	INTERMEDIARIA	107,00	104,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis rupestris	MEDIA-BAIXA	ALTA	ALTO-MEDIO	88,4
110R	FORTE	FRACA	MASCULINA	INTERMEDIARIA	106,00	134,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis rupestris	MEDIA-BAIXA	ALTA-MEDIA	ALTO-MEDIO	90,4
1202C	MEDIA	MEDIA	HERMAFRODITA	INTERMEDIARIA	119,00	103,00	VITIS INTERSPECIFIC CROSSING	MONASTRELL	RUPESTRIS GANZIN	BAIXA	ALTA	ALTO	82,6
140RU	FORTE	FRACA	MASCULINA	INTERMEDIARIA	122,00	124,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis rupestris	MEDIA-BAIXA	BAIXA	ALTO	86,5
1613C	FORTE	FRACA	FEMININA	INTERMEDIARIA	120,00	143,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	OTHELLO	ALTA	MEDIA	MEDIO	99,3
161-49C	MUITO FORTE	MEDIA	FEMININA	INTERMEDIARIA	108,00	102,00	VITIS INTERSPECIFIC CROSSING	RIPARIA	BERLANDIERI	MEDIA-BAIXA	ALTA	MEDIO	NaN
3306C	FORTE	FRACA	MASCULINA	INTERMEDIARIA	101,00	125,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	Vitis rupestris	BAIXA	ALTA-MEDIA	MEDIO	87,9
3309C	FORTE	FRACA	MASCULINA	INTERMEDIARIA	109,00	125,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	Vitis rupestris	BAIXA	ALTA-MEDIA	MEDIO-BAIXO	87,8
41B	MEDIA	FORTE	HERMAFRODITA	INTERMEDIARIA	110,00	91,00	VITIS INTERSPECIFIC CROSSING	Vitis vinifera	Vitis berlandieri	BAIXA	ALTA-MEDIA	ALTO-MEDIO	NaN
420A	MEDIA	FRACA	MASCULINA	INTERMEDIARIA	111,00	115,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis riparia	MEDIA	ALTA	MEDIO-BAIXO	81,3
44-53M	MEDIA	FRACA	MASCULINA	INTERMEDIARIA	106,00	89,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	MALEGUE 144	MEDIA	ALTA-MEDIA	MEDIO	NaN
5BB	FORTE	FRACA	FEMININA	INTERMEDIARIA	123,00	127,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis riparia	ALTA-MEDIA	ALTA	ALTO	66,9
5C	MUITO FORTE	FRACA	MASCULINA	INTERMEDIARIA	118,00	116,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis riparia	ALTA-MEDIA	ALTA	ALTO-MEDIO	85,2
99R	FORTE	FRACA	MASCULINA	INTERMEDIARIA	106,00	112,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis rupestris	ALTA-MEDIA	ALTA	ALTO-MEDIO	89,6
AxR#1	MEDIA	MEDIA	MASCULINA	INTERMEDIARIA	124,00	119,00	VITIS INTERSPECIFIC CROSSING	ARAMON NOIR	Vitis rupestris	BAIXA	ALTA-MEDIA	ALTO-MEDIO	77,7
DOGRIDGE	MEDIA	FRACA	FEMININA	INTERMEDIARIA	119,00	133,00	VITIS INTERSPECIFIC CROSSING	Vitis rupestris	Vitis candicans	ALTA-MEDIA	MEDIA	ALTO	94,3
FERCAL	FRACA	MEDIA	FEMININA	INTERMEDIARIA	109,00	114,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	31R	ALTA-MEDIA	ALTA	ALTO-MEDIO	NaN
FREEDOM	MEDIA	MEDIA	HERMAFRODITA	INTERMEDIARIA	136,00	117,00	VITIS INTERSPECIFIC CROSSING	FRESNO 1613-59	DOGRIDGE	ALTA	MEDIA-BAIXA	ALTO-MEDIO	NaN
GOLIA	FORTE	MEDIA	HERMAFRODITA	INTERMEDIARIA	106,00	109,00	VITIS INTERSPECIFIC CROSSING	CASTEL 156-12	Vitis rupestris	NaN	MEDIA	ALTO	NaN
GRAVESAC	FRACA	MEDIA	MASCULINA	INTERMEDIARIA	118,00	107,00	VITIS INTERSPECIFIC CROSSING	161-49C	3309C	BAIXA	ALTA-MEDIA	ALTO-MEDIO	NaN
HARMONY	MEDIA	FRACA	FEMININA	INTERMEDIARIA	115,00	107,00	VITIS INTERSPECIFIC CROSSING	1613C	DOGRIDGE	ALTA-MEDIA	MEDIA-BAIXA	MEDIO	NaN
IAC313	MEDIA	AUSENTE	MASCULINA	PRECOCE	138,00	145,00	VITIS INTERSPECIFIC CROSSING	GOLIA	Vitis cinerea	ALTA	ALTA	ALTO	100
IAC572	MEDIA	FRACA	MASCULINA	PRECOCE	160,00	132,00	VITIS INTERSPECIFIC CROSSING	Vitis caribaea	101- 14MGT	ALTA	ALTA	ALTO	95,7
IAC766	MEDIA	AUSENTE	MASCULINA	PRECOCE	140,00	120,00	VITIS INTERSPECIFIC CROSSING	106- 8MGT	Vitis caribaea	ALTA	ALTA	ALTO	99,94
RAMSEY	FRACA	AUSENTE	FEMININA	INTERMEDIARIA	137,00	116,00	VITIS CHAMPINII	Vitis champinii	Vitis champinii	ALTA	MEDIA-BAIXA	ALTO-MEDIO	61,1
RIPARIA	MUITO FORTE	FRACA	MASCULINA	PRECOCE	110,00	126,00	VITIS RIPARIA	Vitis riparia	Vitis riparia	MEDIA	ALTA	BAIXO	79,1
RUPESTRIS	FORTE	FRACA	MASCULINA	INTERMEDIARIA	133,00	138,00	VITIS RUPESTRIS	Vitis rupestris	Vitis rupestris	BAIXA	ALTA	ALTO-MEDIO	74,1
SCHWARZMANN	MUITO FORTE	MEDIA	MASCULINA	INTERMEDIARIA	120,00	113,00	VITIS INTERSPECIFIC CROSSING	Vitis riparia	Vitis rupestris	MEDIA	ALTA	MEDIO	100
SO4	MUITO FORTE	FRACA	MASCULINA	INTERMEDIARIA	116,00	120,00	VITIS INTERSPECIFIC CROSSING	Vitis berlandieri	Vitis riparia	ALTA-MEDIA	ALTA	MEDIO	91,1
TRAVIU	FORTE	FRACA	FEMININA	INTERMEDIARIA	NaN	NaN	VITIS INTERSPECIFIC CROSSING	IAC32-7	IAC470-26	MEDIA	ALTA-MEDIA	ALTO-MEDIO	NaN

**APÊNDICE D – TOP-5 RECOMENDAÇÕES DOS MELHORES MODELOS DA
ABORDAGEM DE PREDIÇÃO DE RATINGS**

Recomendações: KNN-LOO					
Enxerto	Recomendações de porta-enxerto				
	1º	2º	3º	4º	5º
BENITAKA	101-14MGT	1103P	110R	FREEDOM	IAC766
BORDO	FREEDOM	RIPARIA	110R	3309C	161-49C
CABERNET FRANC	IAC766	RIPARIA	FREEDOM	110R	HARMONY
CABERNET SAUVIGNON	FREEDOM	1613C	IAC313	3306C	44-53M
CARIGNAN	RIPARIA	FREEDOM	1103P	3309C	101-14MGT
CENTENNIAL SEEDLESS	110R	FREEDOM	1103P	RAMSEY	101-14MGT
CHARDONNAY	IAC766	FREEDOM	RAMSEY	GRAVESAC	AxR#1
CHENIN BLANC	1103P	IAC766	3309C	FREEDOM	110R
COLOMBARD	IAC766	RIPARIA	FREEDOM	110R	HARMONY
CRIMSON SEEDLESS	RIPARIA	3309C	101-14MGT	GRAVESAC	1202C
FANTASY SEEDLESS	140RU	RIPARIA	110R	3309C	161-49C
GRENACHE	1103P	RIPARIA	3309C	HARMONY	RAMSEY
ISABEL	110R	FREEDOM	IAC766	RAMSEY	IAC313
ITALIA	110R	FREEDOM	RAMSEY	1103P	101-14MGT
MAGNA	140RU	RIPARIA	110R	3309C	101-14MGT
MALBEC	IAC766	RIPARIA	FREEDOM	110R	101-14MGT
MARROO SEEDLESS	140RU	RIPARIA	FREEDOM	3309C	110R
MARSELAM	101-14MGT	3309C	SO4	5BB	IAC313
MAXIMO	3309C	RIPARIA	FREEDOM	110R	161-49C
MELODIA	IAC766	140RU	RIPARIA	FREEDOM	3309C
MERLOT	IAC766	FREEDOM	1613C	HARMONY	GRAVESAC
MOSCATO CANELLI	1103P	110R	RAMSEY	5BB	IAC313
MOURVÈDRE	5BB	IAC313	IAC572	101-14MGT	1103P
NIAGARA BRANCA	140RU	RIPARIA	FREEDOM	110R	3309C
NIAGARA ROSADA	140RU	RIPARIA	FREEDOM	110R	3309C
PATRICIA	FREEDOM	420A	1103P	101-14MGT	GRAVESAC
PAULISTINHA	420A	1103P	IAC313	IAC572	GRAVESAC
PETIT VERDOT	5BB	IAC572	1103P	420A	GRAVESAC
PINOT NOIR	IAC766	RIPARIA	FREEDOM	110R	HARMONY
RED GLOBE	RIPARIA	110R	420A	GRAVESAC	3309C
RIESLING	RIPARIA	FREEDOM	IAC766	110R	101-14MGT
SAUVIGNON	110R	161-49C	SO4	101-14MGT	1103P
SUGRAONE	RIPARIA	3309C	IAC766	1613C	101-14MGT
SYRAH	IAC766	RIPARIA	FREEDOM	1202C	AxR#1
TANNAT	IAC766	FREEDOM	110R	101-14MGT	HARMONY
TEMPRANILLO	RIPARIA	FREEDOM	HARMONY	RAMSEY	1202C
THOMPSON SEEDLESS	RIPARIA	IAC766	3309C	1613C	GRAVESAC
VENUS	5BB	IAC313	IAC572	101-14MGT	1103P
VIOLETA	110R	FREEDOM	RAMSEY	101-14MGT	IAC313
VITORIA	140RU	RIPARIA	110R	3309C	1613C

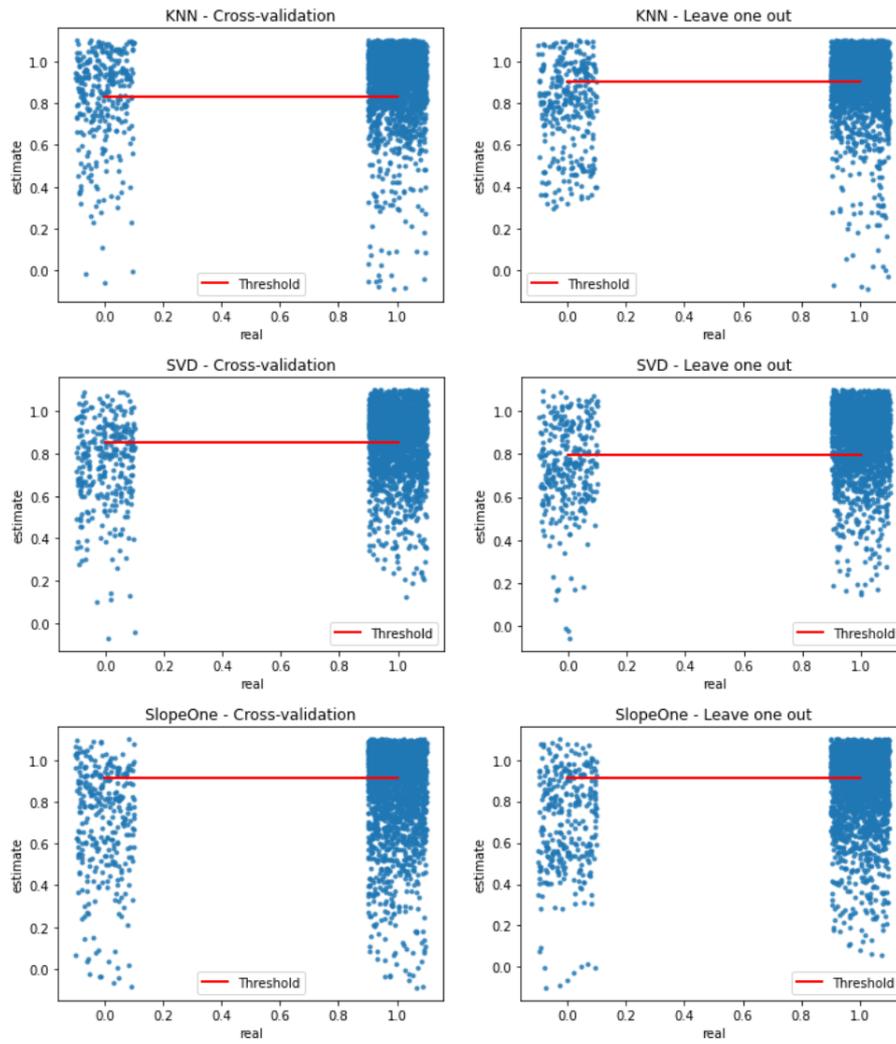
Recomendações: SVD-LOO					
Enxerto	Recomendações de porta-enxerto				
	1º	2º	3º	4º	5º
BENITAKA	110R	IAC766	HARMONY	1613C	FREEDOM
BORDO	3309C	FREEDOM	HARMONY	RIPARIA	41B
CABERNET FRANC	IAC766	RIPARIA	FREEDOM	44-53M	GRAVESAC
CABERNET SAUVIGNON	FREEDOM	41B	1613C	44-53M	3306C
CARIGNAN	1103P	99R	3309C	FREEDOM	44-53M
CENTENNIAL SEEDLESS	3309C	110R	1613C	GRAVESAC	RIPARIA
CHARDONNAY	IAC766	1613C	FREEDOM	41B	GRAVESAC
CHENIN BLANC	1103P	IAC766	FREEDOM	3309C	44-53M
COLOMBARD	IAC766	110R	FREEDOM	41B	RIPARIA
CRIMSON SEEDLESS	RIPARIA	1202C	AxR#1	3309C	41B
FANTASY SEEDLESS	FREEDOM	110R	101-14MGT	3309C	1103P
GRENACHE	3309C	FREEDOM	1103P	IAC766	GRAVESAC
ISABEL	3309C	FREEDOM	110R	HARMONY	44-53M
ITALIA	3309C	110R	HARMONY	FREEDOM	1103P
MAGNA	3309C	110R	44-53M	101-14MGT	1613C
MALBEC	SCHWARZMANN	1613C	IAC766	FREEDOM	161-49C
MARROO SEEDLESS	3309C	110R	FREEDOM	41B	RIPARIA
MARSELAM	3309C	IAC766	44-53M	FREEDOM	1103P
MAXIMO	3309C	RIPARIA	1103P	FREEDOM	1613C
MELODIA	110R	FREEDOM	3309C	IAC766	HARMONY
MERLOT	IAC766	FREEDOM	GRAVESAC	AxR#1	1202C
MOSCATO CANELLI	110R	3309C	IAC766	41B	44-53M
MOURVÈDRE	FREEDOM	110R	3309C	IAC766	1103P
NIAGARA BRANCA	110R	FREEDOM	44-53M	3309C	1103P
NIAGARA ROSADA	3309C	110R	FREEDOM	41B	44-53M
PATRICIA	FREEDOM	1103P	110R	HARMONY	3309C
PAULISTINHA	FREEDOM	110R	3309C	RIPARIA	44-53M
PETIT VERDOT	FREEDOM	110R	3309C	IAC766	1103P
PINOT NOIR	HARMONY	1613C	FREEDOM	SCHWARZMANN	IAC766
RED GLOBE	110R	3309C	RIPARIA	44-53M	1202C
RIESLING	FREEDOM	RIPARIA	HARMONY	IAC766	41B
SAUVIGNON	110R	FREEDOM	44-53M	1103P	101-14MGT
SUGRAONE	AxR#1	3309C	1202C	RIPARIA	IAC766
SYRAH	IAC766	RIPARIA	FREEDOM	44-53M	AxR#1
TANNAT	GRAVESAC	IAC766	FREEDOM	44-53M	41B
TEMPRANILLO	DOGRIDGE	RAMSEY	AxR#1	FREEDOM	44-53M
THOMPSON SEEDLESS	RIPARIA	AxR#1	1202C	IAC766	3309C
VENUS	FREEDOM	110R	3309C	1103P	RIPARIA
VIOLETA	110R	3309C	HARMONY	41B	44-53M
VITORIA	110R	RIPARIA	44-53M	3309C	1202C

Recomendações: NMF-CV					
Enxerto	Recomendações de porta-enxerto				
	1º	2º	3º	4º	5º
BENITAKA	110R	FREEDOM	IAC766	1613C	3309C
BORDO	FREEDOM	110R	1613C	HARMONY	44-53M
CABERNET FRANC	110R	161-49C	FREEDOM	RIPARIA	41B
CABERNET SAUVIGNON	FREEDOM	44-53M	1613C	41B	IAC313
CARIGNAN	FREEDOM	IAC766	3309C	RIPARIA	5C
CENTENNIAL SEEDLESS	110R	FREEDOM	GRAVESAC	1613C	3309C
CHARDONNAY	FREEDOM	44-53M	GRAVESAC	41B	IAC766
CHENIN BLANC	110R	FREEDOM	3309C	GRAVESAC	IAC766
COLOMBARD	GRAVESAC	110R	RIPARIA	FREEDOM	44-53M
CRIMSON SEEDLESS	44-53M	3309C	1202C	RIPARIA	GRAVESAC
FANTASY SEEDLESS	110R	FREEDOM	3309C	44-53M	RIPARIA
GRENACHE	FREEDOM	3309C	1103P	RIPARIA	1202C
ISABEL	110R	FREEDOM	1613C	RIPARIA	IAC766
ITALIA	110R	FREEDOM	3309C	HARMONY	1613C
MAGNA	110R	3309C	RIPARIA	44-53M	GRAVESAC
MALBEC	FREEDOM	AxR#1	GRAVESAC	RIPARIA	110R
MARROO SEEDLESS	110R	FREEDOM	3309C	DOGRIDGE	AxR#1
MARSELAM	FREEDOM	3309C	44-53M	3306C	RIPARIA
MAXIMO	SCHWARZMANN	1613C	HARMONY	DOGRIDGE	110R
MELODIA	110R	FREEDOM	3309C	44-53M	RIPARIA
MERLOT	FREEDOM	GRAVESAC	IAC766	HARMONY	AxR#1
MOSCATO CANELLI	1613C	110R	3309C	IAC766	44-53M
MOURVÈDRE	110R	FREEDOM	3309C	44-53M	3306C
NIAGARA BRANCA	110R	RIPARIA	FREEDOM	44-53M	3309C
NIAGARA ROSADA	3309C	110R	1202C	161-49C	HARMONY
PATRICIA	110R	FREEDOM	1613C	HARMONY	3309C
PAULISTINHA	FREEDOM	3309C	110R	1613C	HARMONY
PETIT VERDOT	110R	FREEDOM	3309C	44-53M	3306C
PINOT NOIR	FREEDOM	44-53M	IAC766	110R	HARMONY
RED GLOBE	110R	44-53M	3309C	GRAVESAC	RIPARIA
RIESLING	HARMONY	FREEDOM	110R	44-53M	1613C
SAUVIGNON	FREEDOM	44-53M	110R	1202C	IAC766
SUGRAONE	44-53M	3309C	RIPARIA	101-14MGT	1202C
SYRAH	IAC766	FREEDOM	IAC313	44-53M	1202C
TANNAT	IAC766	110R	HARMONY	FREEDOM	41B
TEMPRANILLO	RIPARIA	FREEDOM	AxR#1	HARMONY	1202C
THOMPSON SEEDLESS	44-53M	3309C	RIPARIA	IAC766	GRAVESAC
VENUS	110R	FREEDOM	3309C	44-53M	3306C
VIOLETA	110R	FREEDOM	44-53M	3309C	1202C
VITORIA	110R	RIPARIA	3309C	1613C	GRAVESAC

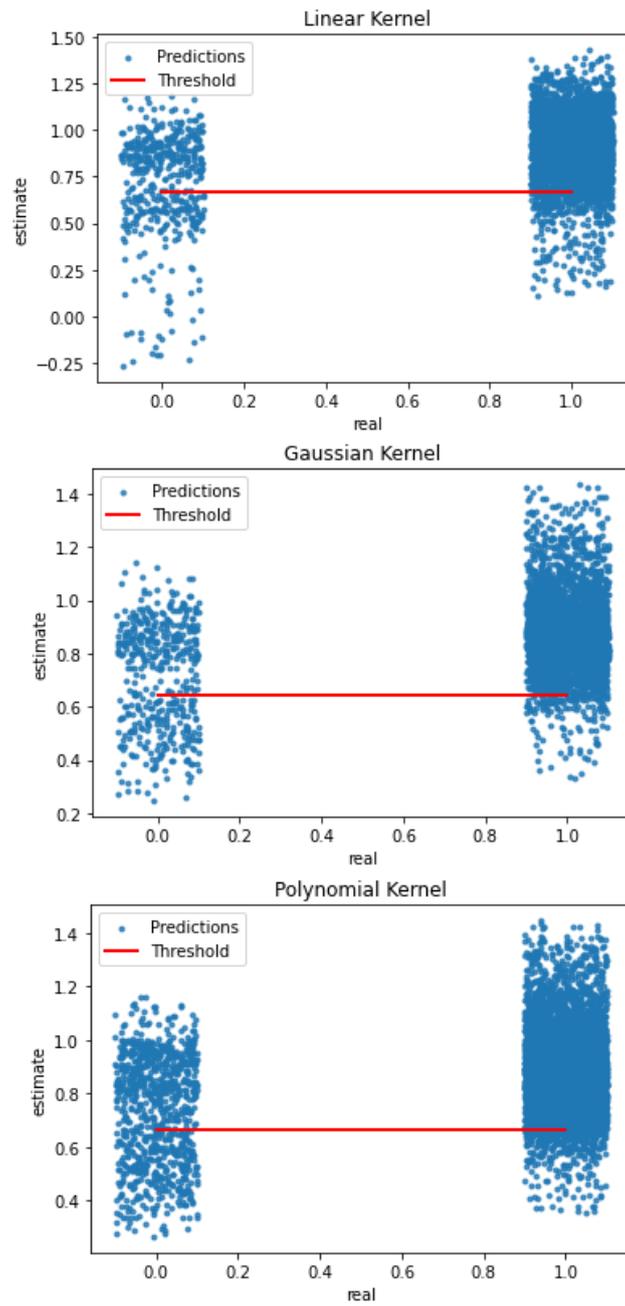
Recomendações: SlopeOne-LOO					
Enxerto	Recomendações de porta-enxerto				
	1º	2º	3º	4º	5º
BENITAKA	1103P	3309C	110R	1202C	140RU
BORDO	3309C	110R	161-49C	1613C	FREEDOM
CABERNET FRANC	110R	1613C	FREEDOM	RIPARIA	161-49C
CABERNET SAUVIGNON	1613C	FREEDOM	44-53M	41B	IAC313
CARIGNAN	44-53M	1613C	FREEDOM	DOGRIDGE	RIPARIA
CENTENNIAL SEEDLESS	3309C	110R	1613C	FREEDOM	41B
CHARDONNAY	IAC766	AxR#1	DOGRIDGE	RUPESTRIS	1613C
CHENIN BLANC	1613C	3309C	FREEDOM	110R	IAC766
COLOMBARD	110R	1202C	AxR#1	RIPARIA	3306C
CRIMSON SEEDLESS	3309C	RIPARIA	AxR#1	1202C	161-49C
FANTASY SEEDLESS	3309C	110R	161-49C	FREEDOM	41B
GRENACHE	1613C	3309C	RIPARIA	FREEDOM	41B
ISABEL	IAC766	110R	99R	RAMSEY	RUPESTRIS
ITALIA	1103P	3309C	110R	140RU	161-49C
MAGNA	110R	AxR#1	161-49C	140RU	1613C
MALBEC	FREEDOM	AxR#1	110R	1613C	1202C
MARROO SEEDLESS	FREEDOM	AxR#1	110R	3309C	1202C
MARSELAM	44-53M	1613C	RIPARIA	3309C	DOGRIDGE
MAXIMO	110R	1613C	FREEDOM	41B	3309C
MELODIA	140RU	IAC766	110R	161-49C	GRAVESAC
MERLOT	1202C	AxR#1	1613C	FREEDOM	DOGRIDGE
MOSCATO CANELLI	101-14MGT	1103P	GRAVESAC	IAC766	TRAVIU
MOURVÈDRE	44-53M	1613C	110R	RIPARIA	3309C
NIAGARA BRANCA	1103P	GRAVESAC	3309C	110R	161-49C
NIAGARA ROSADA	3309C	110R	161-49C	RAMSEY	RIPARIA
PATRICIA	101-14MGT	1103P	GRAVESAC	3309C	110R
PAULISTINHA	GRAVESAC	3309C	140RU	161-49C	41B
PETIT VERDOT	3309C	161-49C	1613C	41B	110R
PINOT NOIR	1613C	FREEDOM	110R	DOGRIDGE	RIPARIA
RED GLOBE	3309C	110R	1202C	AxR#1	RIPARIA
RIESLING	IAC766	110R	1202C	AxR#1	DOGRIDGE
SAUVIGNON	FREEDOM	44-53M	1613C	110R	1103P
SUGRAONE	AxR#1	3309C	1613C	1202C	RIPARIA
SYRAH	44-53M	FREEDOM	RIPARIA	AxR#1	1202C
TANNAT	110R	1202C	AxR#1	DOGRIDGE	RUPESTRIS
TEMPRANILLO	1613C	44-53M	FREEDOM	RIPARIA	DOGRIDGE
THOMPSON SEEDLESS	1613C	3309C	RIPARIA	AxR#1	161-49C
VENUS	101-14MGT	1103P	GRAVESAC	3309C	SO4
VIOLETA	110R	FREEDOM	41B	3309C	1613C
VITORIA	3309C	110R	1202C	140RU	161-49C

APÊNDICE E – DIAGRAMAS DE DISPERSÃO DOS MODELOS DA ABORDAGEM DE CLASSIFICAÇÃO BINÁRIA

Modelos baseados em Filtragem Colaborativa

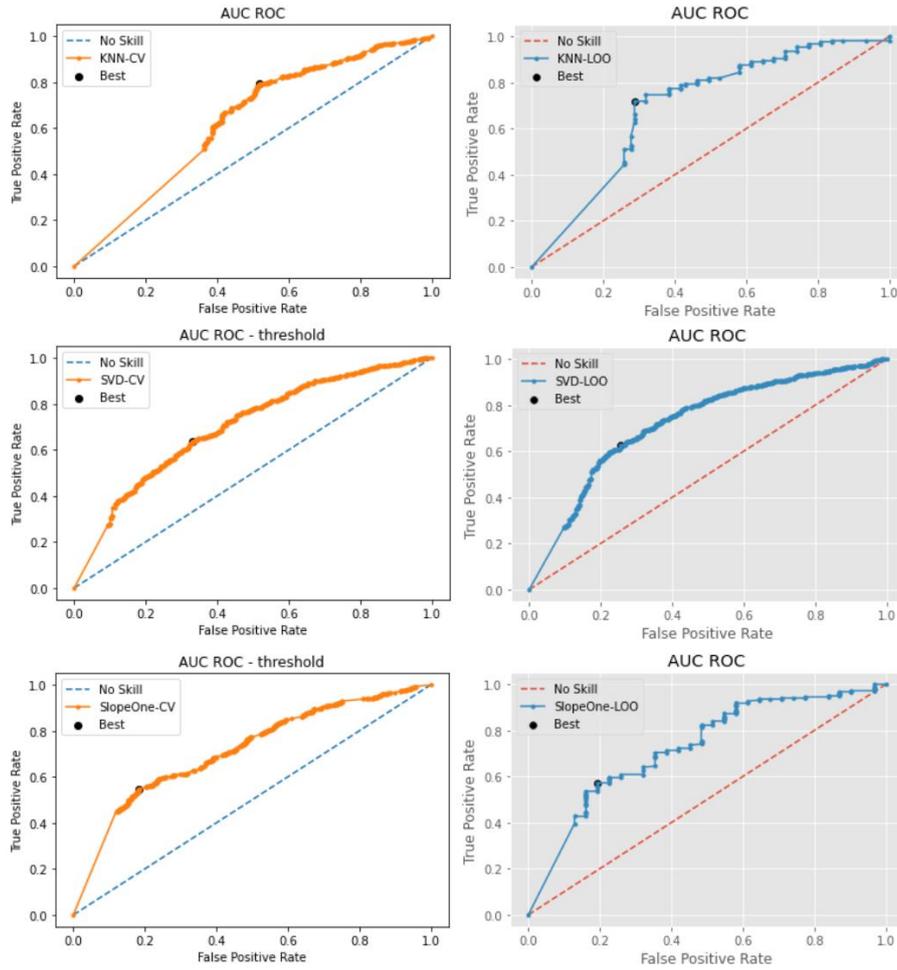


Modelos baseados em Métodos de Kernel



APÊNDICE F – CURVA ROC DOS MODELOS DA ABORDAGEM DE CLASSIFICAÇÃO BINÁRIA

Modelos baseados em Filtragem Colaborativa



Modelos baseados em Métodos de Kernel

