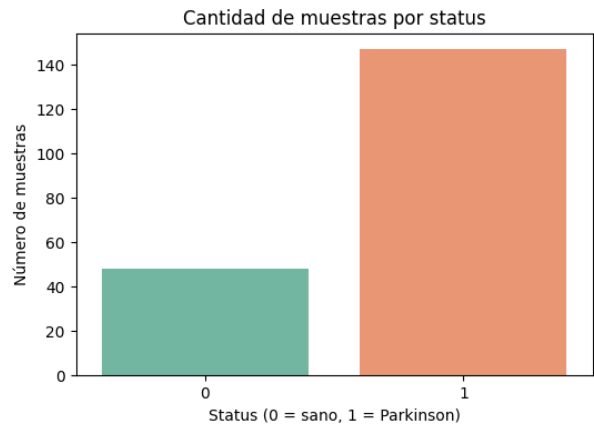


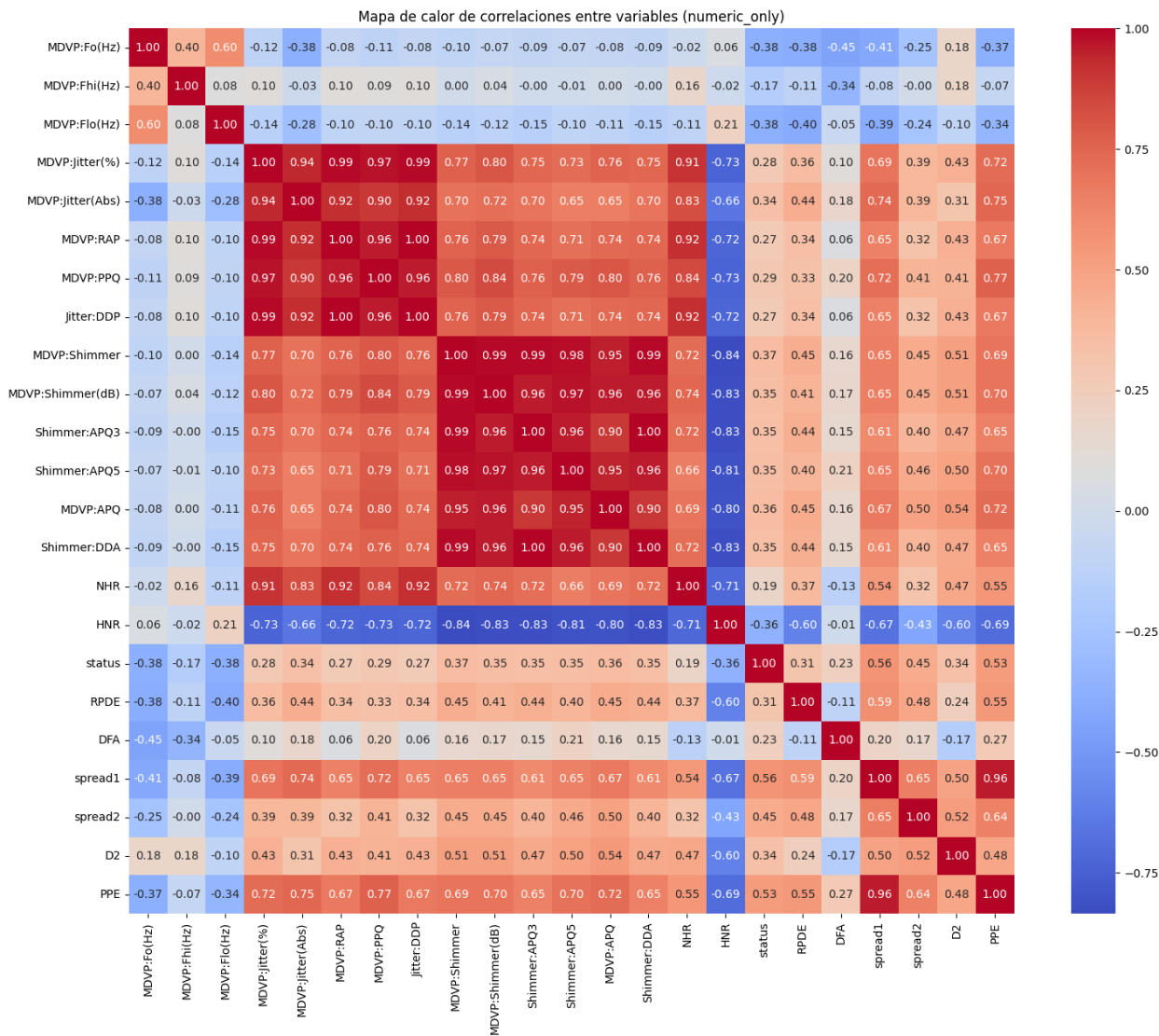
PREDIÇÃO DE PARKINSON

Relatório técnico-estatístico do software

Distribuição por status do alvo:



Mapa de correlação de Pearson entre variáveis::



Validação de pressupostos e escolha do teste estatístico

É fundamental verificar os pressupostos que sustentam os testes paramétricos clássicos. Caso esses pressupostos não sejam atendidos, optamos por métodos não paramétricos, mais robustos a desvios na distribuição dos dados.

Normalidade (Teste de Kolmogorov–Smirnov):

Para cada variável, dividimos as observações em dois grupos conforme o status (0 = saudável, 1 = Parkinson) e aplicamos o teste de Kolmogorov–Smirnov separadamente para cada subconjunto.

- **Hipótese nula (H_0):** os dados provêm de uma distribuição normal.
- **Decisão:**
 - Se o valor de p for maior ou igual a 0,05 ($p \geq 0,05$), não rejeitamos H_0 e consideramos que a amostra é suficientemente próxima da normalidade;
 - Se o valor de p for menor que 0,05 ($p < 0,05$), há evidência de que os dados se desviam significativamente da distribuição normal.

	Variável	0	Normalidade	1	Normalidade
0	MDVP:Fo(Hz)	<0.001	não é normal	<0.001	não é normal
1	MDVP:Fhi(Hz)	<0.001	não é normal	<0.001	não é normal
2	MDVP:Flo(Hz)	<0.001	não é normal	<0.001	não é normal
3	MDVP:Jitter(%)	<0.001	não é normal	<0.001	não é normal
4	MDVP:Jitter(Abs)	<0.001	não é normal	<0.001	não é normal
5	MDVP:RAP	<0.001	não é normal	<0.001	não é normal
6	MDVP:PPQ	<0.001	não é normal	<0.001	não é normal
7	Jitter:DDP	<0.001	não é normal	<0.001	não é normal
8	MDVP:Shimmer	<0.001	não é normal	<0.001	não é normal
9	MDVP:Shimmer(dB)	<0.001	não é normal	<0.001	não é normal
10	Shimmer:APQ3	<0.001	não é normal	<0.001	não é normal
11	Shimmer:APQ5	<0.001	não é normal	<0.001	não é normal
12	MDVP:APQ	0.015	não é normal	<0.001	não é normal
13	Shimmer:DDA	<0.001	não é normal	<0.001	não é normal
14	NHR	<0.001	não é normal	<0.001	não é normal
15	HNR	0.013	não é normal	0.002	não é normal
16	RPDE	0.630	Normal	<0.001	não é normal
17	DFA	<0.001	não é normal	0.090	Normal
18	spread1	0.371	normal	0.009	não é normal
19	spread2	0.640	normal	0.520	normal

20	D2	0.917	normal	0.025	não é normal
21	PPE	0.030	não é normal	<0.001	não é normal

2. Homogeneidade de variâncias (Levene)

Mesmo que ambas as amostras sejam normais, testes paramétricos como o ANOVA exigem que as variâncias dos dois grupos sejam semelhantes. Com o teste de Levene verificamos:

- **H₀:** as variâncias dos dois grupos são iguais..
- **Decisão:** igual ao anterior, $p \geq 0.05 \rightarrow$ variâncias homogêneas; $p < 0.05 \rightarrow$ variâncias diferentes.

3. Escolha do teste

Com os resultados dos testes de normalidade e homogeneidade:

- Se ambos os grupos passam no Shapiro ($p \geq 0.05$) e Levene ($p \geq 0.05$), usamos o teste **ANOVA-F**. O estatístico F mede a razão entre variabilidade “entre grupos” e “intra grupo”: quanto maior, mais discriminante é a variável.
- **Caso contrário**, utilizamos o **teste de Kruskal–Wallis**, que não assume normalidade nem igualdade de variâncias. Seu p-valor indica se as medianas dos dois grupos diferem significativamente.

	Variável	Prova	Estatística	Valor p
1	spread2	ANOVA–F	50.34	<0.001
2	PPE	Kruskal–Wallis	68.08	<0.001
3	spread1	Kruskal–Wallis	68.08	<0.001
4	MDVP:APQ	Kruskal–Wallis	45.88	<0.001
5	MDVP:Jitter(Abs)	Kruskal–Wallis	36.87	<0.001
6	MDVP:PPQ	Kruskal–Wallis	35.63	<0.001
7	MDVP:Shimmer(dB)	Kruskal–Wallis	35.11	<0.001
8	MDVP:Shimmer	Kruskal–Wallis	34.53	<0.001
9	MDVP:Jitter(%)	Kruskal–Wallis	33.32	<0.001
10	Jitter:DDP	Kruskal–Wallis	33.25	<0.001
11	MDVP:RAP	Kruskal–Wallis	33.13	<0.001
12	NHR	Kruskal–Wallis	32.24	<0.001
13	Shimmer:APQ5	Kruskal–Wallis	31.47	<0.001
14	Shimmer:APQ3	Kruskal–Wallis	28.05	<0.001
15	Shimmer:DDA	Kruskal–Wallis	28.02	<0.001
16	HNR	Kruskal–Wallis	24.46	<0.001

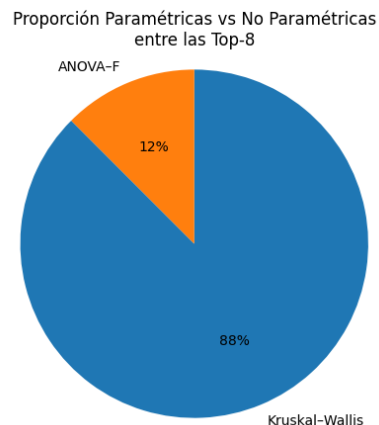
17	D2	Kruskal–Wallis	21.85	<0.001
18	RPDE	Kruskal–Wallis	18.55	<0.001
19	MDVP:F0(Hz)	Kruskal–Wallis	17.40	<0.001
20	MDVP:F10(Hz)	Kruskal–Wallis	16.81	<0.001
21	MDVP:Fhi(Hz)	Kruskal–Wallis	13.21	<0.001
22	DFA	Kruskal–Wallis	9.69	0.002

Para variáveis paramétricas (aquelas submetidas aos testes de Shapiro-Wilk e Levene), é a estatística F da ANOVA.

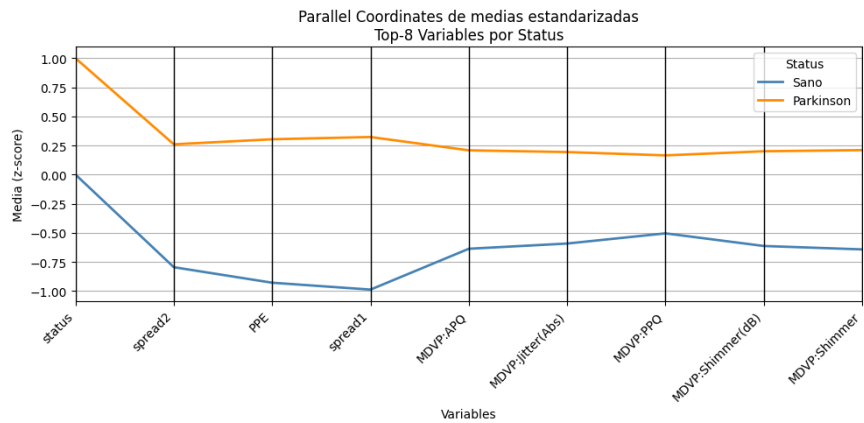
- Matematicamente, $F = (\text{variabilidade intergrupo}) / (\text{variabilidade intragrupo})$.
- Quanto maior o F, maior a diferença relativa nas médias entre indivíduos saudáveis e pacientes com Parkinson, e mais discriminante é a variável.

Para variáveis não paramétricas (las que no cumplían los supuestos), es el **estadístico H** de Kruskal–Wallis.

- H mede a dispersão dos rankings entre os dois grupos
- H maior indica maior diferença entre as distribuições



Após ranquear as variáveis, selecionamos as oito mais discriminantes, padronizamos para z-scores e calculamos a média por grupo (saudáveis vs. Parkinson). Em seguida, traçamos um gráfico de coordenadas paralelas, onde a linha de Parkinson (laranja) permanece acima da linha dos saudáveis (azul), mostrando especialmente em spread2 e spread1 a maior separação entre os grupos.



Treinamento:

Utilizou-se hold-out com 80% dos dados para treino e 20% para teste.

Tabla de validación cruzada (CV)

A tabela mostra o tempo médio de treinamento (TrainTime) e as métricas de avaliação (Acurácia, Precisão, Recall, F1 e MCC) para cada modelo, calculados com base em uma abordagem de CV Estratificado de 5 vezes. A tabela é classificada por MCC em ordem decrescente para destacar qual modelo apresentou o melhor desempenho de classificação equilibrada entre as classes.

Modelo	Tiempo (s)	Accuracy	Precisão	Recall	F1	MCC
Random Forest	0.58	0.85	0.89	0.92	0.9	0.59
Naïve Bayes	0.01	0.76	0.97	0.71	0.82	0.54
XGBoost	0.12	0.83	0.89	0.88	0.88	0.54
k-NN	0.01	0.81	0.87	0.88	0.87	0.47
SVM (RBF)	0.01	0.82	0.85	0.93	0.89	0.46
Logistic L2	0.02	0.81	0.85	0.91	0.88	0.44

Tabela de recursos do sistema:

Ele detalha o ambiente onde os experimentos foram executados: sistema operacional, processador e RAM. Isso fornece contexto sobre os tempos de treinamento e permite a reprodutibilidade dos resultados.

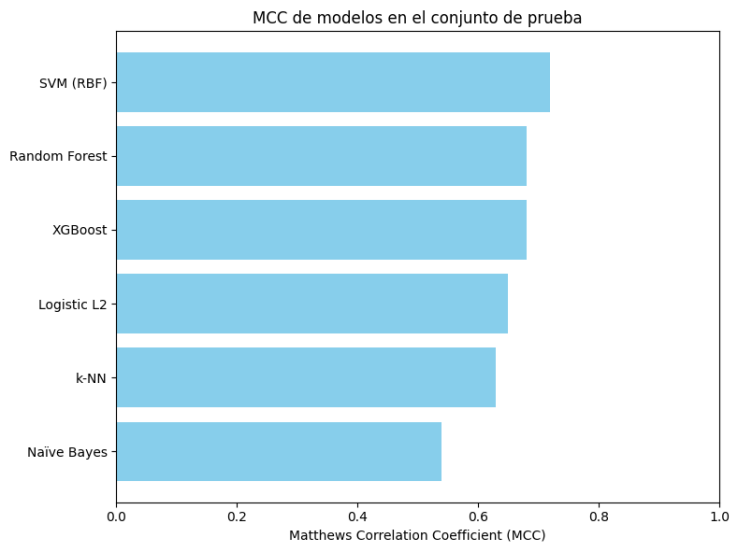
Recurso	Especificação
OS	Linux 6.1.123+
CPU	x86_64
RAM	12.67 GB

Evidência:

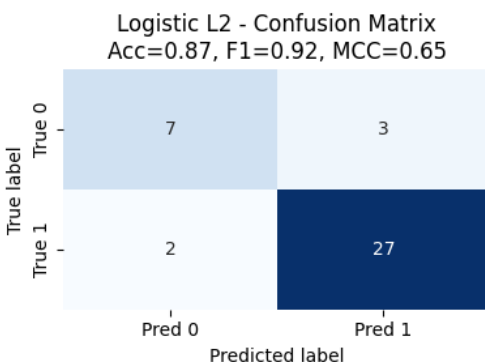
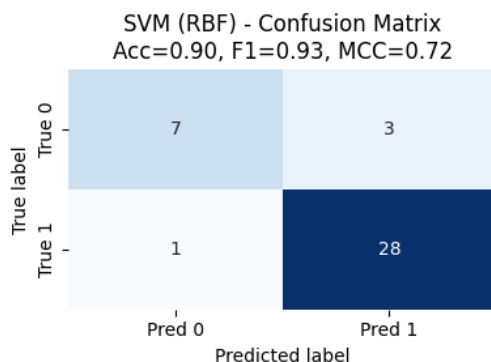
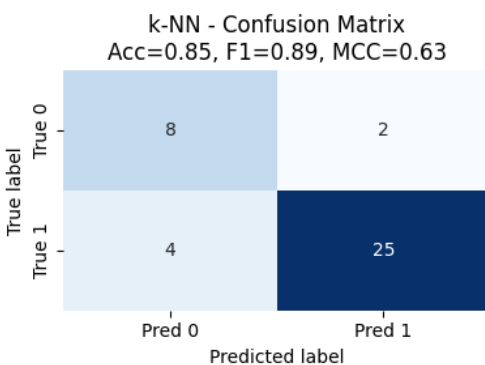
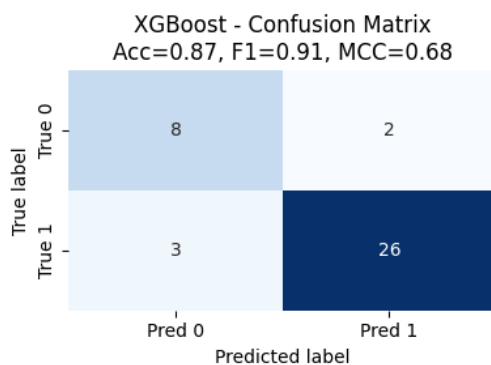
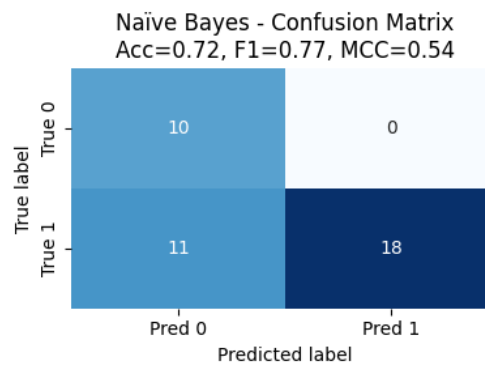
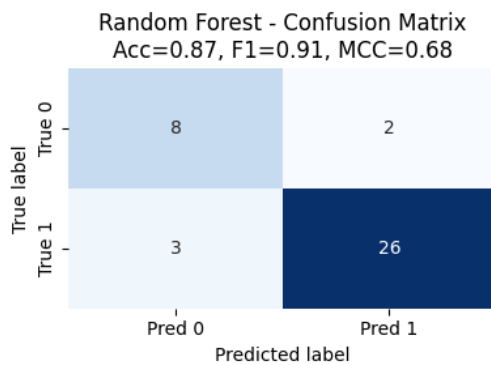
Na fase final, treinamos cada pipeline completo em 80% dos dados, serializamos em disco para garantir a repetibilidade e, em seguida, avaliamos seu desempenho nos 20% restantes ("conjunto de teste"). Para cada modelo, calculamos a Acurácia, a Precisão, a Recall, o F1 e o MCC, e pintamos sua matriz de confusão como um mapa de calor anotado com Verdadeiro/Falso, Positivos e Negativos. Por fim, compilamos todas essas métricas em uma tabela ordenada por MCC, que reflete o equilíbrio geral das classes.

Modelo	Nome do arquivo	Precisão	Precisão	Recall	F1	MCC
SVM (RBF)	SVM_(RBF).h	0.9	0.9	0.97	0.93	0.72
Random Forest	Random_Forest.h	0.87	0.93	0.9	0.91	0.68
XGBoost	XGBoost.h	0.87	0.93	0.9	0.91	0.68
Logistic L2	Logistic_L2.h	0.87	0.9	0.93	0.92	0.65
k-NN	k-NN.h	0.85	0.93	0.86	0.89	0.63
Naïve Bayes	Naïve_Bayes.h	0.72	1.0	0.62	0.77	0.54

MCC (Coeficiente de correlação de Matthews): métrica de correlação para classificação binária que considera TP, TN, FP e FN, e fornece um único valor interpretativo (−1 a +1). O critério principal é escolhido porque equilibra a eficácia da previsão em conjuntos de dados com possível desequilíbrio de classes.



Matriz de confusão de todos os modelos



Na etapa final, escolhemos o melhor modelo com base no MCC no conjunto de teste (neste caso, SVM com kernel RBF, $MCC = 0,720$) e salvamos seu pipeline. Para verificar se tudo funciona, carregamos esse pipeline e realizamos uma inferência de exemplo em uma amostra de teste, obtendo tanto o rótulo previsto quanto a probabilidade associada.

Em seguida, para garantir que sua superioridade não fosse uma coincidência, aplicamos o teste de McNemar entre o melhor modelo e cada um dos outros: construímos tabelas de contingência a partir de suas previsões no mesmo conjunto de teste, calculamos o valor de p e determinamos se as diferenças no número de acertos/erros eram estatisticamente significativas ($\alpha = 0,05$).

Teste de McNemar (melhor vs. outros):

SVM (RBF) vs. Random Forest: valor de $p = 0,250 \rightarrow$ diferença significativa: não

SVM (RBF) vs. XGBoost: valor de $p = 0,375 \rightarrow$ diferença significativa: não

SVM (RBF) vs. Logistic L2: valor de $p = 1,000 \rightarrow$ diferença significativa: não

SVM (RBF) vs. k-NN: valor de $p = 0,219 \rightarrow$ diferença significativa: não

SVM (RBF) vs. Naïve Bayes: valor de $p = 0,000 \rightarrow$ diferença significativa: sim

Otimização de Hiperparâmetros

Utilizamos o GridSearchCV com estratificação de 5 vezes (o mesmo da validação anterior) para cada uma das principais métricas: AUC-ROC, Precisão, Precisão, Recall, F1 e MCC.

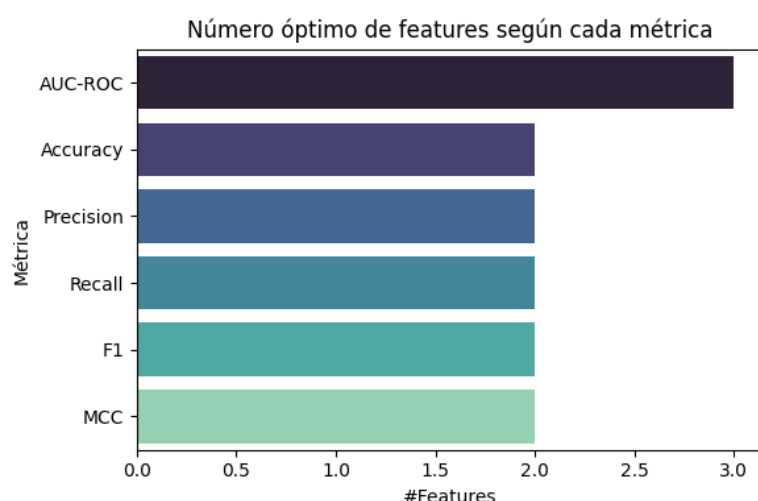
Métricas	Parâmetros	CV_score
AUC-ROC	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel': 'linear'}	0.90
Accuracy	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.84
Precision	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.88
Recall	{'clf__C': 0.1, 'clf__gamma': 'scale', 'clf__kernel': 'rbf'}	1.0
F1	{'clf__C': 1, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.89
MCC	{'clf__C': 100, 'clf__gamma': 1, 'clf__kernel': 'rbf'}	0.56

Otimização de MCC e Seleção de Variáveis com RFECV

Para refinar nosso SVM, realizamos duas etapas encadeadas:

1. **GridSearchCV** otimizando a métrica MCC (Coeficiente de Correlação de Matthews), que equilibra verdadeiros/falsos positivos e negativos.
2. **RFECV** com SVM linear, para determinar o número ótimo de variáveis de acordo com cada métrica (AUC-ROC, Exatidão, Precisão, Recall, F1 e MCC).

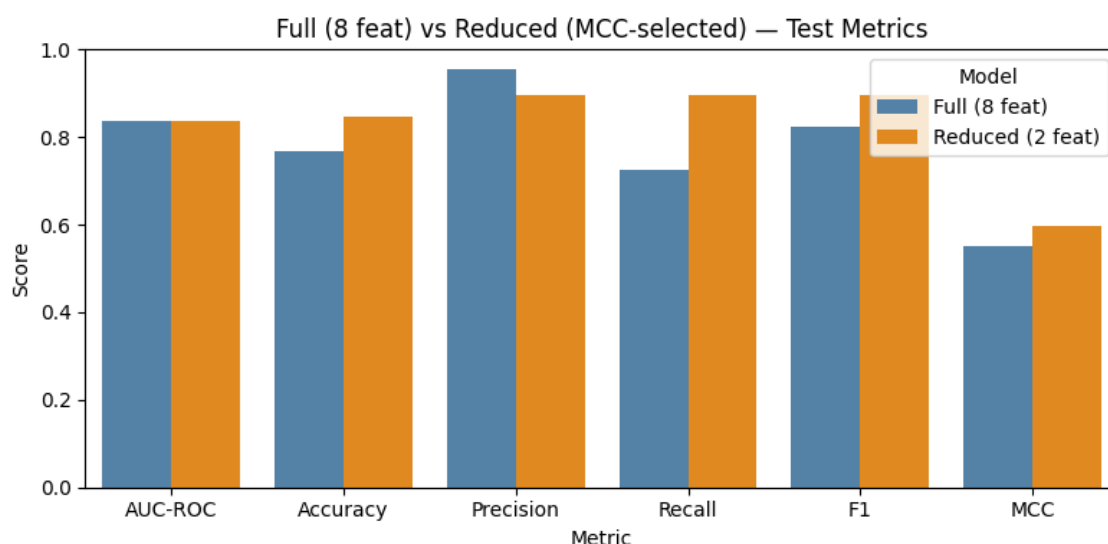
Métrica	#Características	CV_Score	Test_Score	Características
AUC-ROC	3	0.912	0.921	spread1, MDVP:APQ, MDVP:Shimmer
Accuracy	2	0.847	0.846	MDVP:APQ, MDVP:Shimmer
Precision	2	0.876	0.897	MDVP:APQ, MDVP:Shimmer
Recall	2	0.933	0.897	MDVP:APQ, MDVP:Shimmer
F1	2	0.902	0.897	MDVP:APQ, MDVP:Shimmer
MCC	2	0.573	0.597	MDVP:APQ, MDVP:Shimmer



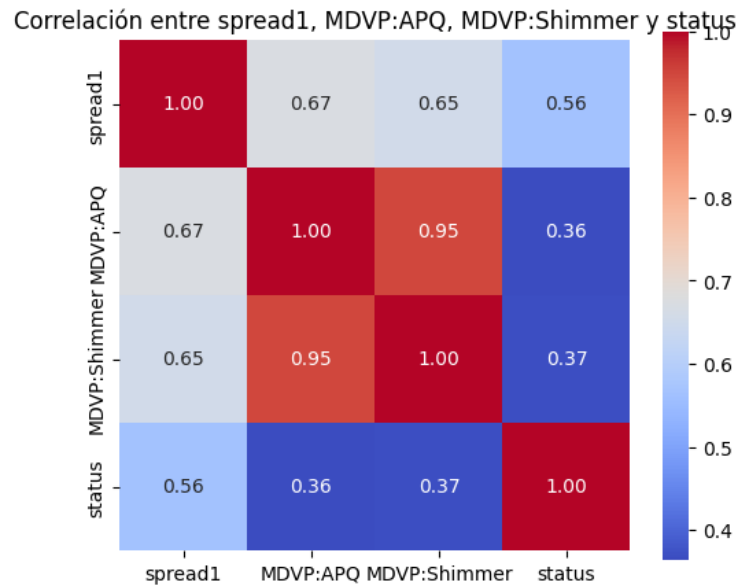
Comparação de desempenho no teste: Completo (8 recursos) vs Reduzido (2 recursos)

Modelo	AUC-ROC	Accuracy	Precision	Recall	F1	MCC
Completo (8 talentos)	0.838	0.769	0.955	0.724	0.824	0.55
Reduzido (2 talentos)	0.838	0.846	0.897	0.897	0.897	0.597

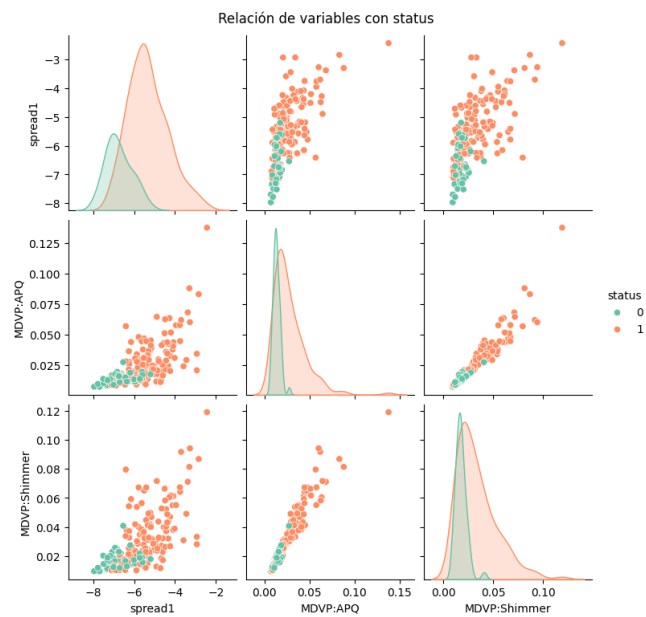
Reduzir para apenas MDVP:APQ e MDVP:Shimmer não só simplifica o modelo em 75% menos variáveis, como também melhora sua capacidade de classificação balanceada, refletida no aumento do MCC. Isso confirma que esses dois recursos capturam a maior parte das informações discriminativas sem sacrificar a confiabilidade geral do classificador.



Mapa de correlação das variáveis finais por status:



Discriminação de variáveis entre pacientes que têm e aqueles que não têm Parkinson:

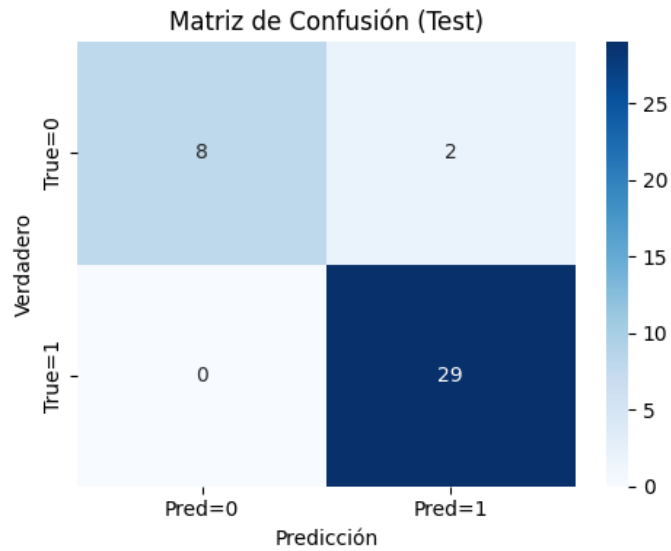


Com o modelo final (SVM-RBF otimizado por MCC) treinado nas três variáveis finais (spread1, MDVP:APQ e MDVP:Shimmer):

Métrica	Train	Test
AUC-ROC	0.951	0.921
Accuracy	0.885	0.949

Precisao	0.879	0.935
Recall	0.983	1.0
F1	0.928	0.967
MCC	0.669	0.865

Matriz de confusão final:



Selección de modelo Híbrido

Model	Auc	Accuracy	Precision	Recall	F1	Mcc
Soft_voting_parkinson	0.917	0.846	0.897	0.897	0.897	0.597
Stacking_parkinson	0.914	0.821	0.867	0.897	0.881	0.515