



**UNIVERSIDADE DE FORTALEZA
CENTRO DE PÓS GRADUAÇÃO
CURSO DE MBA EM CIÊNCIA DE DADOS**

José Freitas Alves Neto

**Aplicação de Modelo de *Machine Learning* para a Previsão de Topos e Fundos
em Gráficos de Renko no Contrato Futuro de Dólar**

FORTALEZA – CEARÁ
2025

RESUMO

Este projeto ataca um problema crucial para traders: o atraso na confirmação de topos e fundos de mercado. Focando no contrato futuro de dólar em gráficos de Renko, nosso objetivo foi antecipar estes sinais. Para isso, desenvolvemos um modelo de Machine Learning que classifica o movimento atual do preço. O modelo prevê se o "brick" corrente se tornará um topo, um fundo ou um movimento neutro. Utilizando indicadores técnicos, testamos e otimizamos diversos algoritmos de classificação. O modelo de Análise Discriminante Linear (LDA) apresentou a melhor performance. Alcançamos uma precisão superior a 70% na identificação desses pontos de reversão. Essa ferramenta foi projetada para ser um indicador de apoio estratégico e em tempo real. Ele permite otimizar os pontos de entrada, a gestão de risco e a tomada de decisão. O estudo visa, assim, aumentar a assertividade e a potencial rentabilidade do operador.

Palavras-chave: Machine Learning, Gráficos de Renko, Topos e Fundos.

ABSTRACT

This project tackles a critical challenge for traders: the inherent delay in confirming market tops and bottoms. Focusing on the Brazilian dollar futures contract using Renko charts, we aimed to anticipate these key turning points. We developed a Machine Learning model to classify the current price action in real-time. The model predicts whether the current Renko brick will form a top, a bottom, or a neutral move. Using established technical indicators, we trained and compared several classification models. The Linear Discriminant Analysis (LDA) model emerged as the most effective and robust. It achieved a precision rate exceeding 70% in identifying these reversal signals. The result is a powerful strategic indicator to support trading decisions. It enables traders to optimize their entry points and improve their risk management. Ultimately, this work provides a tool to enhance trading accuracy and potential profitability.

Keywords: Machine Learning, Renko Charts, Tops and Bottoms.

LISTA DE ILUSTRAÇÕES

1	Gráfico de Candles de 5 minutos para o ativo Dólar Futuros	12
2	Gráfico de 4 Renko para o ativo Dólar Futuro	13
3	Tabela da distribuição de classes do Target	14
4	Distribuição dos atributos usados para o treinamento do modelo.	15
5	Resultados do treinamento dos modelos via AutoML	17
6	Resultado da classificação do Logistic Regression	20
7	Matriz de confusão para o modelo do Logistic Regression	20
8	Curva ROC da classificação do Logistic Regression	21
9	Variação do Thresholds da classificação do Logistic Regression	22
10	Curva de aprendizagem para o modelo Logistic Regression	22
11	Resultado da classificação do Linear Discriminant Analysis	23
12	Matriz de confusão do Linear Discriminant Analysis	24
13	Variação da precisão ao longo Linear Discriminant Analysis	25
14	Curva de Aprendizado do Linear Discriminant Analysis	25
15	Resultado da classificação do Random Forest	26
16	Matriz de confusão do Random Forest	27
17	Curva ROC da classificação do Random Forest	27
18	Variação da precisão ao longo do thresholds do Random Forest	28
19	Curva de aprendizado do Random Florest	28
20	Resultado da classificação do Voting Classifier	29
21	Matriz de confusão do Voting Classifier	30
22	Variação da precisão para cada uma das classes do Voting Classifier	30
23	Resultado da classificação para o modelo Voting Classifier com limiar (threshold) de 70%.	31

SUMÁRIO

1	INTRODUÇÃO	10
2	PRÉ PROCESSAMENTO DE DADOS	11
2.1	Conhecendo o Dado	11
2.1.1	Tipo de gráfico	11
2.1.2	Indicadores usados	13
2.1.3	Origem da base de dados	13
2.2	Tratamento de dados	14
2.2.1	Variáveis desenvolvidas	14
2.2.2	Ajustando o Target	14
2.2.3	Normalizando e Padronização da base	15
2.2.4	Exportação de dados.	16
3	TREINAMENTO DO MODELO	17
3.1	Aplicação de AutoML	17
3.2	Ajuste final para treinar o modelo	18
3.3	Treinamento dos modelos	19
3.3.1	Logistic Regression	19
3.3.2	Linear Discriminant Analysis	23
3.3.3	Random Forest Classifier	26
3.3.4	Voting Classifier	29
4	Conclusão	32
5	CONSIDERAÇÕES FINAIS	33

1 INTRODUÇÃO

A busca por rentabilidade no mercado de capitais é um objetivo perseguido há séculos, levando inúmeros indivíduos a dedicarem extensos períodos de suas vidas ao desenvolvimento de estratégias para a geração de patrimônio. Nesse contexto, consolidaram-se diversas abordagens e áreas de estudo sobre o comportamento dos mercados, dentre as quais se destacam três principais: a análise técnica, a análise fundamentalista e a análise de fluxo de ordens (*tape reading*). O presente estudo concentra-se na análise técnica, metodologia que busca desenvolver estratégias operacionais a partir da interpretação de gráficos e indicadores.

Com efeito, a identificação de topos e fundos no gráfico de preços reveste-se de fundamental importância para os operadores do mercado. Essa prática permite a identificação da tendência vigente e da direção mais provável do ativo, subsidiando a tomada de decisão de compra ou venda. A relevância dessa identificação é exemplificada em estratégias consolidadas, como as Ondas de Elliott.

Contudo, é crucial salientar que os topos e fundos considerados "verdadeiros" — ou seja, pontos de reversão confirmados — possuem um inerente atraso (*delay*) em sua confirmação. Geralmente, são validados com segurança apenas após a formação de um número predeterminado de candles subsequentes, conforme estabelecido pelo critério de cada negociante. Tal defasagem temporal pode reduzir a eficiência da negociação, resultando em uma entrada tardia no movimento, o que pode erodir a margem de lucro e ampliar o stop financeiro da operação.

Portanto, a capacidade de antecipar a formação de um topo ou fundo verdadeiro é de suma importância para os negociantes, pois lhes permite otimizar a rentabilidade e maximizar o retorno esperado de suas estratégias ao capturar o início de um novo movimento de preço. Visando mitigar essa deficiência, foi desenvolvido um modelo de *Machine Learning* cujo objetivo é prever se o candle corrente se configurará como um topo, um fundo ou um movimento neutro, oferecendo assim uma ferramenta de apoio estratégico para os participantes da bolsa de valores.

2 PRÉ PROCESSAMENTO DE DADOS

O propósito deste documento é detalhar a metodologia empregada no desenvolvimento do projeto. Abordaremos o tipo de gráfico utilizado, os indicadores selecionados e o processo de construção do pipeline de modelagem de dados para o treinamento do modelo.

2.1 Conhecendo o Dado

2.1.1 Tipo de gráfico

No âmbito da análise de mercado na bolsa de valores, utilizam-se diversos tipos de gráficos, cada qual concebido como um método de análise específico para destacar determinadas informações do mercado.

Como exemplo notório, o gráfico de *Candlestick* (ou de velas) representa a variação do preço de um ativo ao longo de um intervalo de tempo predefinido, que pode variar de meses e semanas a dias, horas, minutos e segundos. A principal função de cada *Candle* é exibir quatro informações cruciais para o período em questão: o preço de abertura, a cotação máxima, a cotação mínima e o preço de fechamento.

Contudo, uma limitação inerente a essa representação é a perda de visibilidade sobre a trajetória do preço dentro do intervalo de tempo consolidado pelo *Candle*. Para analisar essa dinâmica interna, torna-se necessário reduzir a periodicidade do gráfico (por exemplo, de um gráfico diário para um de 60 minutos), permitindo assim a observação do conceito de fractal do movimento de preços, no qual padrões gráficos se repetem em diferentes escalas de tempo.

Nessa análise detalhada, os topos (pontos de máxima) e os fundos (pontos de mínima) são frequentemente identificados e destacados, como ilustrado pela utilização das letras "T" para Topo e "F" para Fundo.

Figura 1 - Gráfico de *Candles* de 5 minutos para o ativo Dólar Futuros



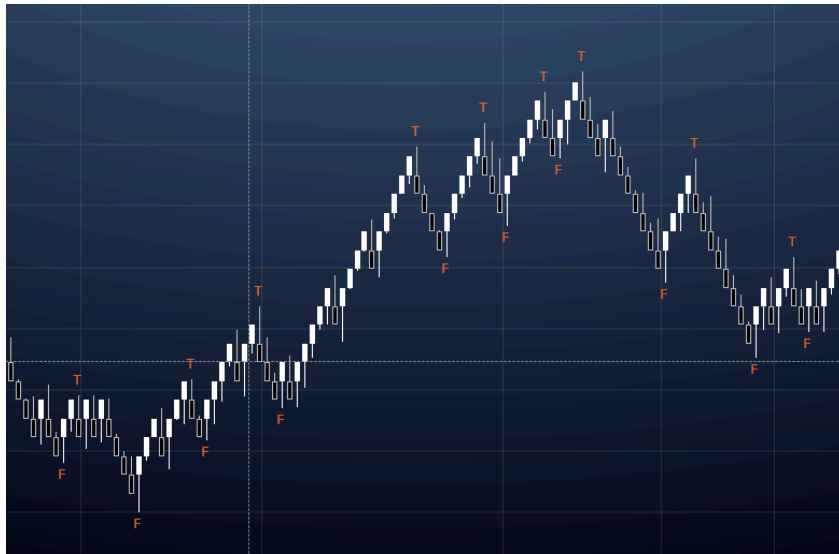
Fonte: Imagem retirada da plataforma ProfitChart Pro pelo autor.

O gráfico de *Renko* constitui uma ferramenta de análise gráfica atemporal, cuja construção independe de intervalos de tempo fixos, baseando-se exclusivamente na variação dos preços de um ativo. Em sua metodologia, a unidade gráfica fundamental não é o tradicional *Candle*, mas sim o "*Brick*".

A formação de cada "*Brick*" ocorre quando o preço se desloca uma magnitude predeterminada pelo usuário (em pontos ou *ticks*) em uma direção, seja ela ascendente ou descendente, a partir do fechamento anterior. Ao atingir esse limiar, um novo "*Brick*" é consolidado, possuindo uma estrutura análoga à do *Candle*, com valores de abertura, máxima, mínima e fechamento.

Uma vantagem proeminente desta representação gráfica é a notável simplificação na identificação de topos e fundos. Tal clareza visual otimiza a performance de modelos analíticos, facilitando a interpretação dos movimentos de mercado e a execução de negociações (*trades*).

Figura 2 - Gráfico de 4 *Renko* para o ativo Dólar Futuro



Fonte: Imagem retirada da plataforma ProfitChart Pro pelo autor.

A escolha recaiu sobre o gráfico de 4 *Renko* para a análise do ativo dólar futuro. Esta metodologia foi selecionada por sua capacidade de detalhar a movimentação dos preços, o que facilita a identificação precisa de topos e fundos na negociação do ativo na bolsa de valores brasileira (B3).

2.1.2 Indicadores usados

Para subsidiar o modelo na tarefa de classificação de topos e fundos, foram utilizados indicadores consagrados da análise técnica. Os instrumentos empregados incluem: Saldo de agressão, Índice de força relativa (IFR), Arron Oscilador, Williams %R, Volume negociado, MACD histograma, média móvel e o detector de topos e fundos com um delay de 2 candles.

2.1.3 Origem da base de dados

O conjunto de dados em questão foi obtido diretamente da plataforma de negociação ProfitChart Pro. Foram extraídos 20.000 *bricks* do gráfico *Renko* (4R) referente ao ativo financeiro DOLPRO. O período de análise dos dados compreende o intervalo de 10 de abril de 2025 a 04 de julho de 2025.

2.2 Tratamento de dados

2.2.1 Variáveis desenvolvidas

A partir dos indicadores previamente elencados, foram desenvolvidas métricas consideradas relevantes, tais como: saldo de agressão do *Brick*, distância absoluta do fechamento em relação à média móvel, a natureza positiva ou negativa dos últimos cinco *bricks* e a ocorrência de virada de *brick*.

2.2.2 Ajustando o *Target*

O objetivo do estudo consiste em detectar topos e fundos do mercado financeiro. Tais eventos, entretanto, não ocorrem em todos os movimentos de mercado. Dessa forma, ao desenvolver a variável alvo (*target*), atribuiu-se o valor 1 para a ocorrência de um topo, -1 para a de um fundo e 0 para uma condição neutra. Constata-se que a base de dados possui um desbalanceamento significativo para a classe neutra, apresentando a seguinte distribuição:

Figura 3 - Tabela da distribuição de classes do *Target*

Target	Quantidade
Neutro	16647
Fundos	1677
Topos	1677

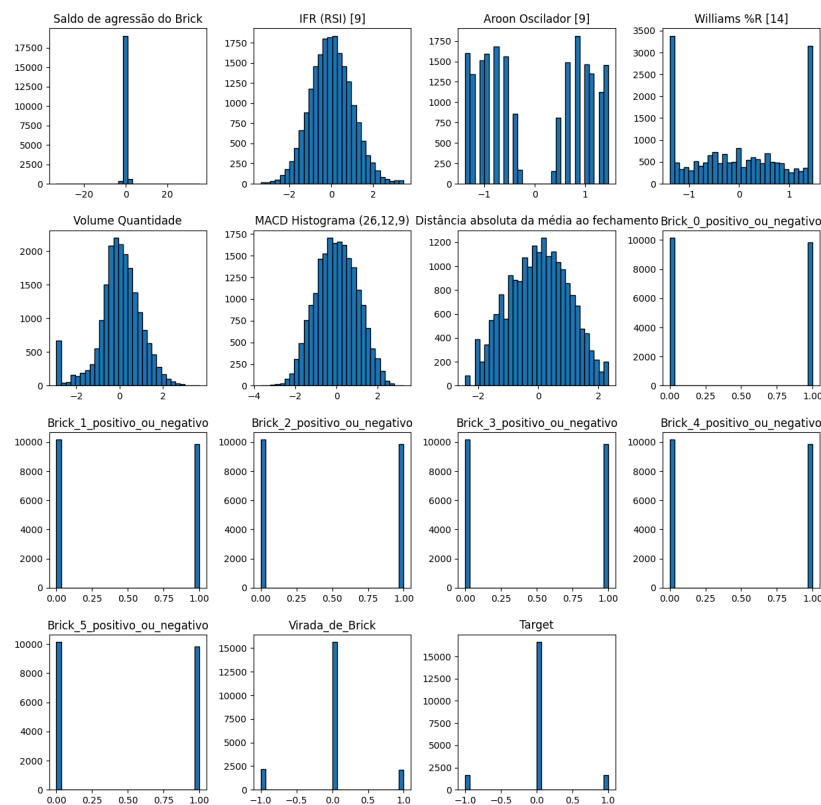
Fonte: Tabela gerada pelo autor

Diante disso, temos um problema de multiclass, onde 3 classes são possíveis.

2.2.3 Normalizando e Padronização da base

Ao analisar a distribuição de todos os atributos do conjunto de dados, identificou-se que duas distribuições apresentaram potencial para a aplicação de normalização. Consequentemente, realizou-se uma transformação de raiz quarta na coluna “Volume Quantidade” e de raiz quadrada na coluna “Distância Absoluta da Média ao Fechamento”, a fim de aproximá-las de uma distribuição normal. Optou-se pela não remoção dos outliers, uma vez que os topos e fundos são considerados eventos extremos e frequentemente se manifestam em tais pontos, tornando sua permanência na base de dados essencial para a análise.

Figura 4 - Distribuição dos atributos usados para o treinamento do modelo



Fonte: Imagem gerada pelo autor.

Adicionalmente, aplicou-se a ferramenta *StandardScaler* a todos os atributos de entrada não binários, com exceção da coluna “Virada de *Brick*”, com o objetivo de padronizar seus respectivos valores.

2.2.4 Exportação de dados.

Finalmente, para fins de organização e visando à sua utilização em um notebook distinto, o conjunto de dados tratado foi salvo. Neste novo arquivo, foram exportadas apenas as colunas que serão utilizadas na modelagem, a saber: 'Saldo de agressão do Brick', 'IFR (RSI) [9]', 'Aroon Oscilador [9]', 'Williams %R [14]', 'Volume Quantidade', 'MACD Histograma (26,12,9)', 'Distância absoluta da média ao fechamento', 'Brick_0_positivo_ou_negativo', 'Brick_1_positivo_ou_negativo', 'Brick_2_positivo_ou_negativo', 'Brick_3_positivo_ou_negativo', 'Brick_4_positivo_ou_negativo', 'Brick_5_positivo_ou_negativo', 'Virada_de_Brick' e 'Target'.

3 TREINAMENTO DO MODELO

A presente seção descreve as etapas do treinamento do modelo, com o aprofundamento dos conceitos utilizados e a justificativa para cada decisão adotada no desenvolvimento da estrutura de treinamento.

3.1 Aplicação de *AutoML*

A determinação a priori do modelo com o melhor desempenho é impraticável, tornando necessária a avaliação de múltiplas abordagens para identificar a mais adequada ao conjunto de dados em questão. Embora a experiência prévia possa fornecer estimativas sobre os modelos mais promissores, não há garantia de certeza absoluta. Diante desse cenário, optou-se pela aplicação inicial de um processo de *Automated Machine Learning (AutoML)* por meio da biblioteca *PyCaret*, na qual foram treinados e avaliados diversos modelos de classificação.

Figura 5 - Resultados do treinamento dos modelos via *AutoML*

Model	Accuracy	AUC	Recall	Prec
Linear Discriminant Analysis	0.9361	0	0.9361	0.9489
Logistic Regression	0.9360	0	0.9360	0.9487
Random Forest Classifier	0.9333	0.9653	0.9333	0.9443
Gradient Boosting Classifier	0.9326	0	0.9326	0.9430
Extra Trees Classifier	0.9300	0.9668	0.9300	0.9398
K Neighbors	0.9286	0.9596	0.9286	0.9352

Classifier				
Light Gradient Boosting Machine	0.9268	0.9679	0.9268	0.9339
Ada Boost Classifier	0.9235	0	0.9235	0.9268
Naive Bayes	0.9136	0.9668	0.9136	0.9417
Decision Tree Classifier	0.8975	0.8088	0.8975	0.8961
SVM - Linear Kernel	0.8815	0	0.8815	0.8631
Quadratic Discriminant Analysis	0.8436	0	0.8436	0.7712
Dummy Classifier	0.8323	0.5	0.8323	0.6927
Ridge Classifier	0.8321	0	0.8321	0.6927

Fonte: Tabela gerada pelo autor

Diante dos resultados expostos, optou-se por realizar um treinamento mais aprofundado dos três modelos que apresentaram a maior precisão (*Precision*). Dessa forma, os modelos *Linear Discriminant Analysis*, *Logistic Regression* e *Random Forest Classifier* foram otimizados por meio da técnica *Grid Search*, com o objetivo de aprimorar os resultados previamente encontrados.

3.2 Ajuste final para treinar o modelo

Para o treinamento aprofundado dos modelos, o conjunto de dados foi segmentado em amostras de treinamento e teste, nas proporções de 70% e 30%, respectivamente. A divisão foi realizada de forma estratificada com base

na variável alvo (*target*), garantindo a representatividade das classes em ambas as amostras por meio de aleatorização.

Subsequentemente, a técnica de subamostragem aleatória (*Random Undersampling*) foi aplicada exclusivamente ao conjunto de dados de treinamento. Este procedimento teve como objetivo equalizar o número de instâncias entre as classes e, assim, evitar um possível viés do modelo para a classe majoritária. O conjunto de teste foi mantido em seu estado original, sem a aplicação de balanceamento, a fim de simular um ambiente de aplicação real e permitir uma avaliação imparcial da performance do modelo.

3.3 Treinamento dos modelos

3.3.1 Logistic Regression

O modelo de Regressão Logística (*Logistic Regression*), intrinsecamente binário, foi adaptado para o problema de classificação multiclasse por meio da estratégia *One-vs-Rest* (OvR), implementada pela classe *OneVsRestClassifier*. A utilização dessa abordagem é um requisito para habilitar o método *predict_proba*, que calcula as probabilidades de cada classe e será utilizado posteriormente na implementação do *VotingClassifier*.

A fim de otimizar o desempenho do modelo, realizou-se a sintonia de hiperparâmetros com a utilização da técnica *Grid Search*, associada a um processo de validação cruzada com cinco partições (*5-folds*). Tal procedimento resultou em uma acurácia de validação de 96,6%, valor que representa um aprimoramento em relação ao resultado obtido na etapa de *AutoML*.

Figura 6 - Resultado da classificação do *Logistic Regression*

```

--- RELATÓRIO DE CLASSIFICAÇÃO ---
precision    recall  f1-score   support

-1          0.64      1.00      0.78        490
0           1.00      0.90      0.95       5009
1           0.71      1.00      0.83        500

 accuracy          0.92        5999
 macro avg          0.78      0.97      0.85        5999
 weighted avg          0.95      0.92      0.93        5999

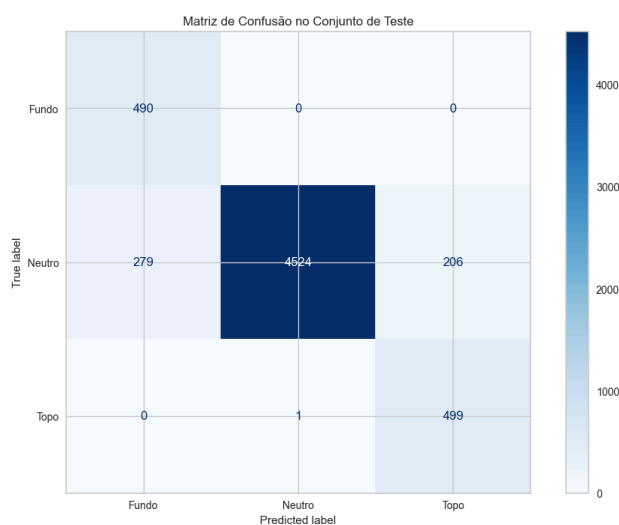
```

Fonte: Imagem gerada pelo autor.

A avaliação do desempenho do modelo no conjunto de teste teve como métrica principal a precisão (*Precision*), em virtude de sua relevância para os objetivos do estudo. O classificador alcançou uma precisão de 64% para a classe -1 (fundo) e de 71% para a classe 1 (topo).

A análise da matriz de confusão indica que o modelo assimila os padrões de forma adequada. Observa-se que as classificações incorretas ocorrem, majoritariamente, entre as classes extremas (topo e fundo) e a classe neutra. Contudo, é relevante notar que o modelo não confunde um evento de topo com um de fundo, e vice-versa, demonstrando um comportamento desejável para a aplicação final.

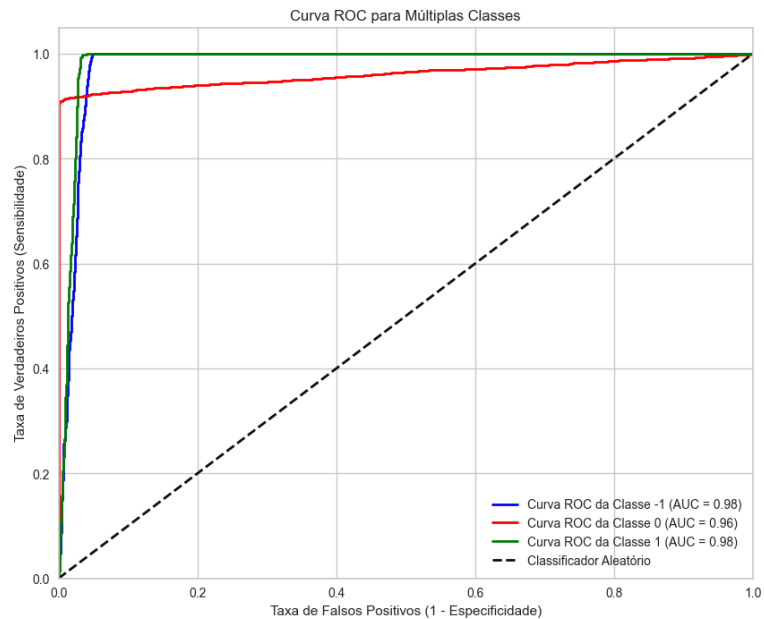
Figura 7 - Matriz de confusão para o modelo do *Logistic Regression*



Fonte: Imagem gerada pelo autor.

A análise da curva ROC (*Receiver Operating Characteristic*) revela um desempenho de alta performance do classificador, o que é evidenciado por um valor de Área Sob a Curva (AUC) superior ou igual a 0,96 ($AUC \geq 0,96$) para todas as classes.

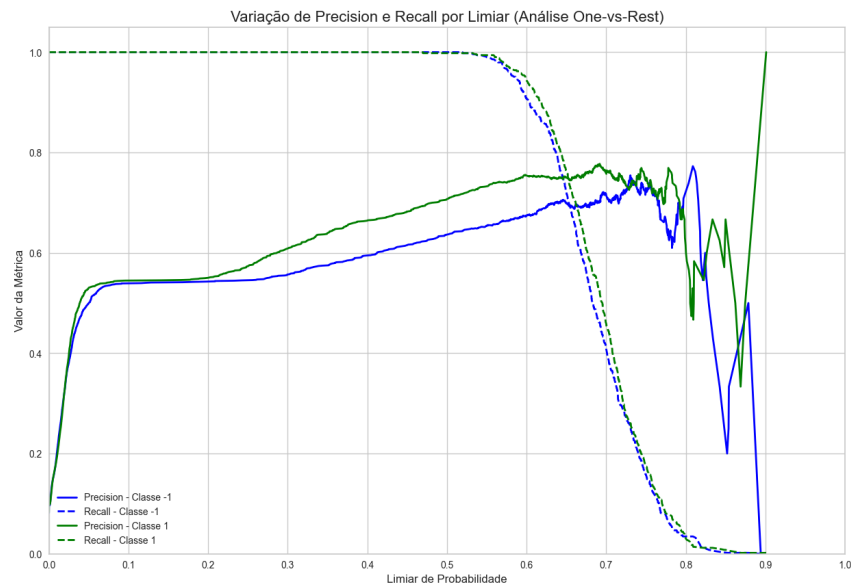
Figura 8 - Curva ROC da classificação do *Logistic Regression*



Fonte: Imagem gerada pelo autor

Ao variar o limiar de decisão (*threshold*), é possível otimizar a métrica de precisão. Com o aumento do valor desse limiar, alcançaram-se resultados superiores, atingindo-se valores de precisão próximos a 80% para a classe de topo e de aproximadamente 70% para a classe de fundo.

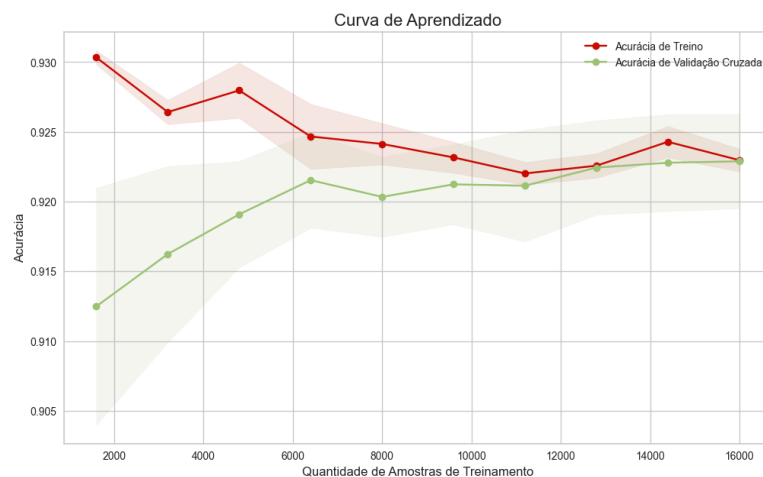
Figura 9 - Variação do *Thresholds* da classificação do *Logistic Regression*



Forte: Imagem gerada pelo autor.

A curva de aprendizado obtida durante o treinamento do modelo apresenta uma convergência adequada, com as métricas de desempenho nos conjuntos de treino e de validação seguindo trajetórias próximas. Tal comportamento evidencia que não ocorreu o fenômeno de sobreajuste (*overfitting*).

Figura 10 - Curva de aprendizagem nos dados de teste para o modelo *Logistic Regression*



Fonte: Imagem gerada pelo autor.

Portanto, o modelo de *Machine Learning* desenvolvido apresenta um desempenho promissor, mostrando-se viável para aplicação no mercado financeiro como um indicador de negociação. Não obstante, com o intuito de otimizar os resultados, proceder-se-á com a avaliação de algoritmos alternativos.

3.3.2 *Linear Discriminant Analysis*

Para o treinamento do modelo de Análise Discriminante Linear (*Linear Discriminant Analysis*), empregou-se a técnica de otimização de hiperparâmetros *Grid Search*, associada a um processo de validação cruzada de cinco partições (*5-folds*). Nesta etapa, utilizou-se o mesmo conjunto de dados de treinamento do modelo anterior, testando-se uma gama diversificada de parâmetros.

O processo de busca pelo melhor conjunto de hiperparâmetros envolveu um total de 1.780 ajustes (*fits*), resultando em uma acurácia média de 95,75% na validação cruzada. O desempenho do modelo resultante, quando aplicado ao conjunto de dados de teste, é apresentado a seguir.

Figura 11 - Resultado da classificação do *Linear Discriminant Analysis*

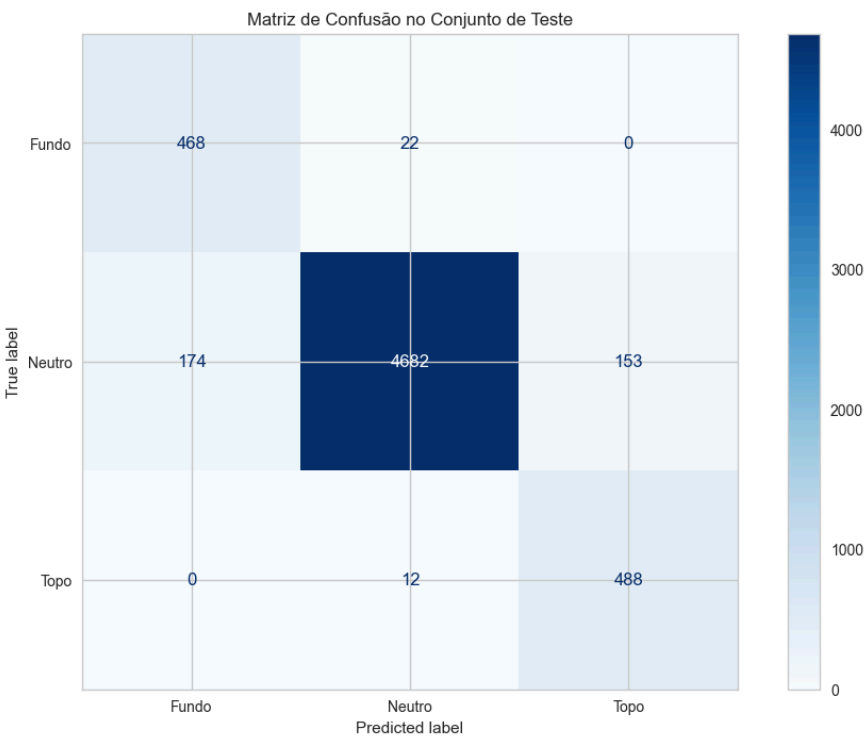
--- RELATÓRIO DE CLASSIFICAÇÃO ---				
	precision	recall	f1-score	support
-1	0.73	0.96	0.83	490
0	0.99	0.93	0.96	5009
1	0.76	0.98	0.86	500
accuracy			0.94	5999
macro avg	0.83	0.96	0.88	5999
weighted avg	0.95	0.94	0.94	5999

Fonte: Imagem gerada pelo autor.

Em comparação ao modelo de Regressão Logística, o classificador de Análise Discriminante Linear obteve valores de precisão superiores, tornando-se, assim, um candidato mais favorável para a aplicação em um ambiente de negociação real. A análise da matriz de confusão corrobora seu desempenho, demonstrando que as classificações incorretas se concentram

entre as classes extremas (topo e fundo) e a classe neutra, sem que haja confusão entre os eventos de topo e de fundo.

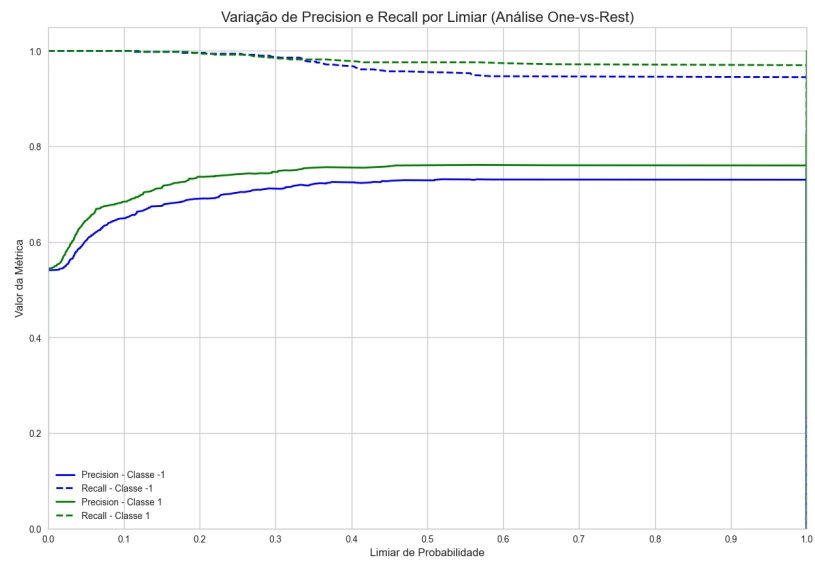
Figura 12 - Matriz de confusão do *Linear Discriminant Analysis*



Fonte: Imagem gerada pelo autor.

O modelo exibe uma curva ROC (*Receiver Operating Characteristic*) de alta performance, com um valor de Área Sob a Curva (AUC) superior ou igual a 0,97 ($AUC \geq 0,97$) para todas as classes. Adicionalmente, a análise da variação do limiar de classificação demonstra que as métricas de desempenho apresentam estabilidade, com pouca variação para valores de limiar acima de 0,5.

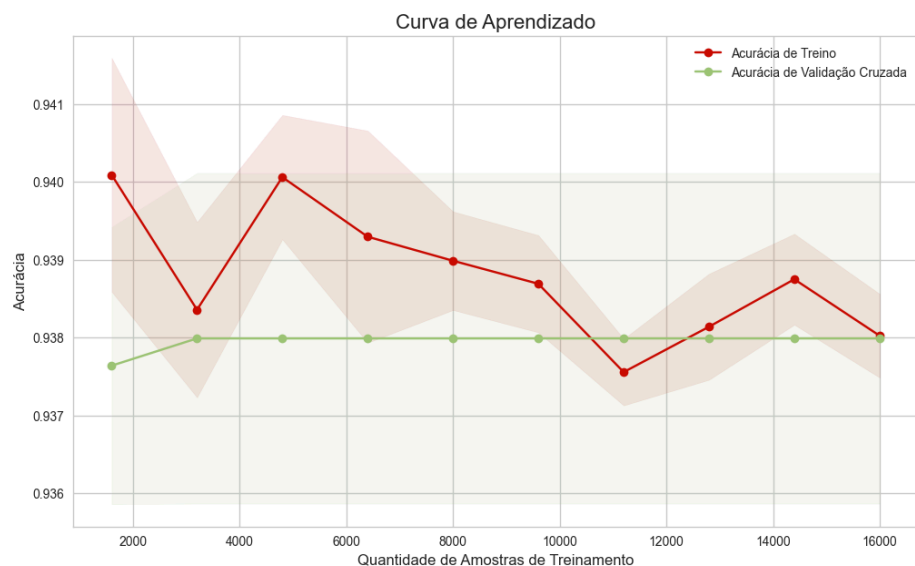
Figura 13 - Variação da precisão ao longo do *Thereshold* do *Linear Discriminant Analysis*



Fonte: Imagem gerada pelo autor.

A curva de aprendizado do modelo demonstra um comportamento estável à medida que o número de amostras de treinamento é aumentado.

Figura 14 - Curva de Aprendizado do *Linear Discriminant Analysis*



Fonte: Imagem gerada pelo autor.

3.3.3 *Random Forest Classifier*

De maneira análoga, o modelo *Random Forest* foi submetido a um treinamento aprofundado com o propósito de otimizar seus hiperparâmetros. Empregou-se, para isso, a técnica *Grid Search* associada a uma validação cruzada de cinco partições (*5-folds*).

O processo de busca exaustiva compreendeu um total de 1.200 ajustes (*fits*). O melhor modelo resultante alcançou uma acurácia média de 97,42% durante a etapa de validação cruzada. O desempenho final do referido modelo, quando avaliado no conjunto de dados de teste, é demonstrado a seguir.

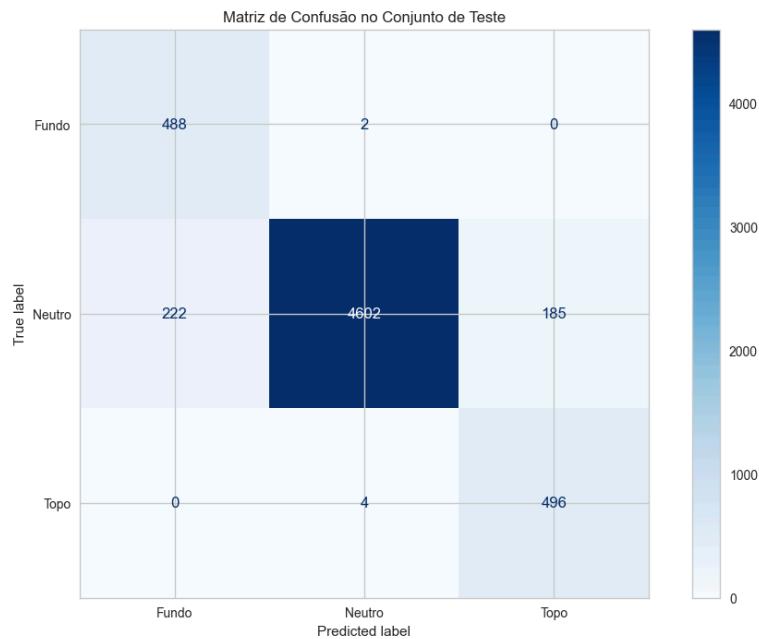
Figura 15 - Resultado da classificação do *Random Forest*

--- RELATÓRIO DE CLASSIFICAÇÃO ---				
	precision	recall	f1-score	support
-1	0.69	1.00	0.81	490
0	1.00	0.92	0.96	5009
1	0.73	0.99	0.84	500
accuracy			0.93	5999
macro avg	0.80	0.97	0.87	5999
weighted avg	0.95	0.93	0.94	5999

Fonte: Imagem gerada pelo autor.

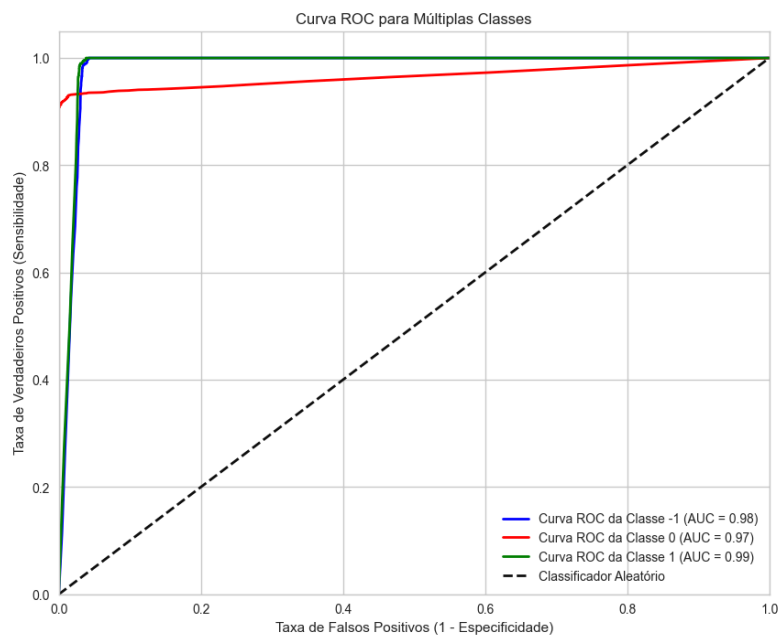
Verificando a matriz de confusão, tivemos valores satisfatórios e bem similares aos modelos anteriores.

Figura 16 - Matriz de confusão do *Random Forest*



A curva ROC do modelo apresenta um desempenho satisfatório, uma vez que obteve um valor de AUC de, no mínimo, 97% em todas as classes.

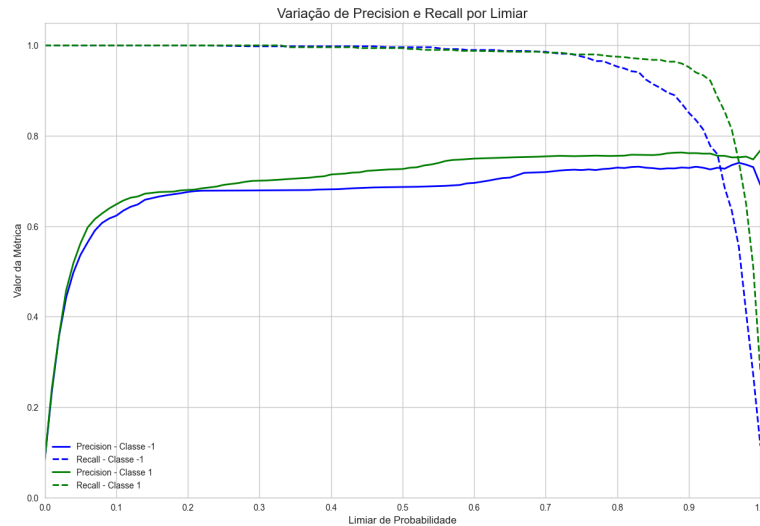
Figura 17 - Curva ROC da classificação do *Random Forest*



Fonte: Imagem gerada pelo autor.

A análise da relação entre o limiar e a precisão revelou uma progressão aproximadamente linear, o que indica que a seleção de valores de limiar mais elevados tende a aumentar a precisão.

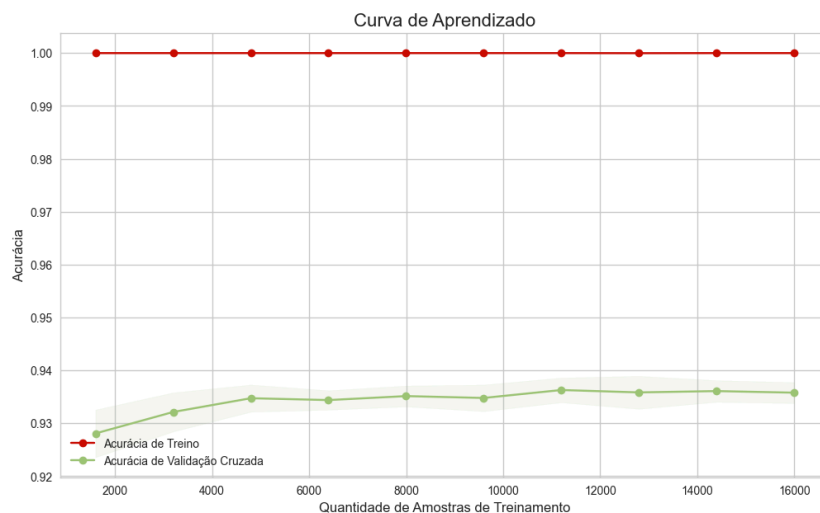
Figura 18 -Variação da precisão ao longo do *thresholds* do *Random Forest*



Fonte: Imagem gerada pelo autor.

A curva de aprendizagem apresenta uma acurácia constante para os dados de treinamento e uma leve progressão para os dados de teste.

Figura 19 - Curva de aprendizado do *Random Florest*



Fonte: Imagem gerada pelo autor.

3.3.4 Voting Classifier

Com o propósito de aprimorar a performance e obter um modelo superior, optou-se pela implementação de um *Voting Classifier*. Essa abordagem consiste em agrupar os três modelos previamente treinados que apresentaram o melhor desempenho, a fim de verificar a possibilidade de alcançar métricas mais elevadas.

Para tal, os modelos foram combinados utilizando o sistema de votação "soft", no qual se atribui peso idêntico a cada modelo. A predição final é definida pela classe que alcança a maior probabilidade média, calculada a partir das estimativas individuais de cada modelo.

Contudo, os resultados obtidos com o modelo agregado foram semelhantes aos registrados pelo melhor modelo individual, não demonstrando uma melhora expressiva em relação às métricas anteriores.

Figura 20 - Resultado da classificação do *Voting Classifier*

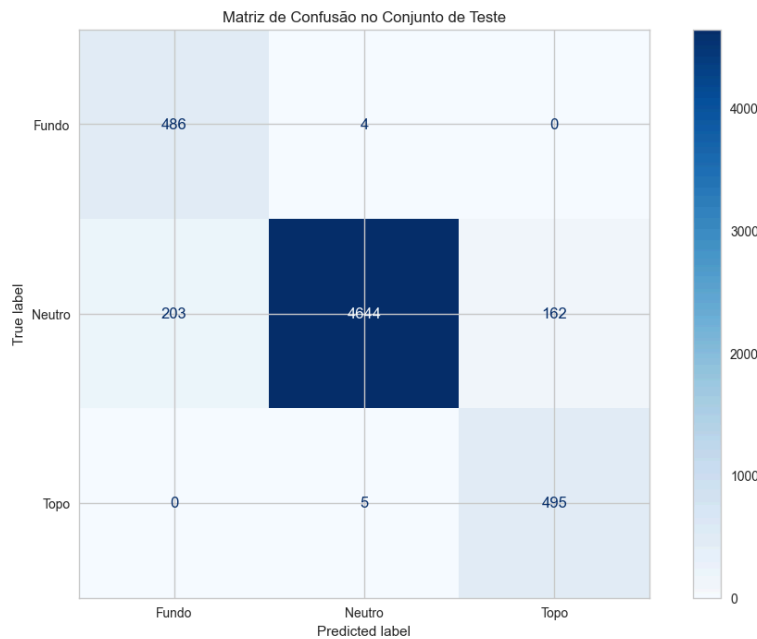
```
--- RELATÓRIO DE CLASSIFICAÇÃO ---
```

	precision	recall	f1-score	support
-1	0.71	0.99	0.82	490
0	1.00	0.93	0.96	5009
1	0.75	0.99	0.86	500
accuracy			0.94	5999
macro avg	0.82	0.97	0.88	5999
weighted avg	0.95	0.94	0.94	5999

Fonte: Imagem gerada pelo autor.

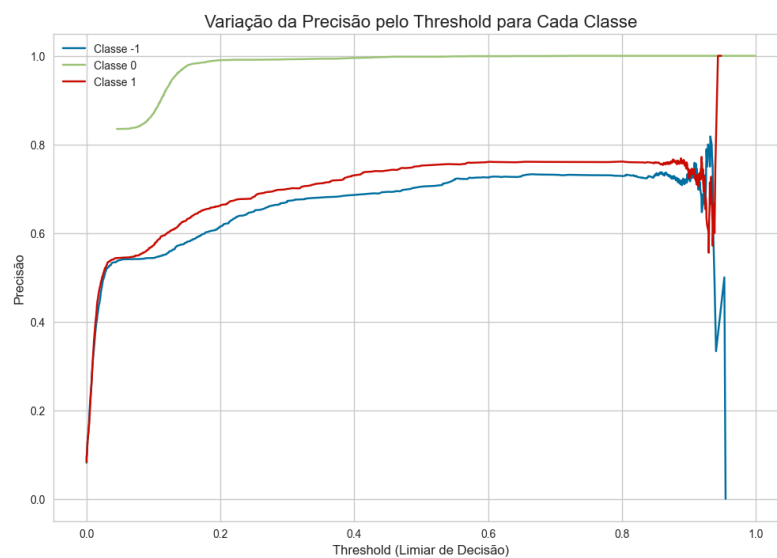
A matriz de confusão resultante apresenta resultados semelhantes aos dos três modelos anteriores, não demonstrando variações ou mudanças significativas.

Figura 21 - Matriz de confusão do *Voting Classifier*



A análise da variação do limiar revela um pico de precisão em aproximadamente 70%, sendo este valor adotado como referência para a avaliação.

Figura 22 - Variação da precisão para cada uma das classes do *Voting Classifier*



Fonte: Imagem gerada pelo autor.

Nota-se que os resultados obtidos são semelhantes aos do modelo Linear Discriminant Analysis ao se empregar um valor de limiar equivalente.

Figura 23 – Resultado da classificação para o modelo *Voting Classifier* com limiar de 70%.

--- AVALIAÇÃO COM REGRA CUSTOMIZADA (LIMIAIR 0.7) ---				
	precision	recall	f1-score	support
-1	0.73	0.96	0.83	490
0	0.99	0.94	0.96	5009
1	0.76	0.97	0.85	500
accuracy			0.94	5999
macro avg	0.83	0.95	0.88	5999
weighted avg	0.95	0.94	0.94	5999

Fonte: Imagem gerada pelo autor.

Portanto, obtém-se um modelo com desempenho satisfatório e precisão elevada, entretanto, sem apresentar uma diferença significativa em comparação com os demais modelos analisados.

4. CONCLUSÃO

Dentre os quatro modelos que apresentaram desempenho satisfatório, optou-se pela seleção do *Linear Discriminant Analysis* como o modelo principal, com base no princípio da Navalha de Occam, que favorece a simplicidade. Este modelo demonstrou maior precisão na identificação de topos e fundos, viabilizando a aplicação do indicador em operações reais com uma assertividade superior a 70%.

Vale salientar que perante ao período gráfico de 4 renko, essa ferramenta será na maioria das vezes melhor utilizada dentro do contexto de Daytrade, visto que 4r se equipara a um gráfico de 1-5 minutos de candle.

Diante disso, o usuário que utilizar este modelo disporá de uma maior segurança em seus pontos de entrada, fundamentada em uma probabilidade de acerto de aproximadamente 70%. O indicador pode ser implementado de forma complementar a outras análises e indicadores do mercado de capitais para otimizar a assertividade e a potencial rentabilidade das negociações.

A implementação do modelo requer a importação do arquivo exportado via joblib em um ambiente *Python* e a subsequente integração com a biblioteca do MetaTrader 5, a qual fornece acesso a uma vasta base de dados de negociações. A partir destes dados, o usuário deverá recriar os indicadores utilizados no treinamento do modelo, incluindo uma função para a geração de gráficos *Renko*. Subsequentemente, os dados gerados devem alimentar o modelo para que este realize as previsões.

Com a importação do modelo para dentro da plataforma, pode-se usar para desenvolver estratégias derivadas a partir do modelo base e otimizar com objetivo de encontrar melhores parâmetros para o modelo e assim ter uma estratégia vencedora na bolsa de valores.

Também recomenda-se a plotagem de um gráfico com notificações visuais (possível topo, fundo ou neutro) para facilitar a análise, fornecendo assim um indicador adicional para aprimorar o desempenho operacional na bolsa de valores.

5. CONSIDERAÇÕES FINAIS

O mercado de venda de produtos vinculado a bolsa de valores e a IA ainda está no início e não amplamente divulgado, existem várias maneiras de monetizar esse desenvolvimento e alcançar um público grande disposto a comprar indicadores e robôs de investimentos. Para isso recomendamos estudar o mercado e validar bem os próprios modelos, visto que existe uma grande quantidade de cursos e ferramentas sendo vendidas que não tem uma validação prática e entregam falsas promessas.

Por fim, ressalta-se que a bolsa de valores é um ambiente complexo e volátil, influenciado por múltiplas notícias e pelo comportamento coletivo de milhares de investidores em escala global. Portanto, é imperativo que o investimento seja realizado de forma responsável e que a utilização de qualquer ferramenta analítica seja associada a métricas de gestão de risco robustas, visando à obtenção de retornos consistentes a longo prazo.