Explicação do Código: Análise de Indicadores Técnicos e Métricas Financeiras

Este código foi desenvolvido para calcular e adicionar uma série de indicadores técnicos e métricas financeiras a um DataFrame contendo dados de mercado financeiro, como preços de ativos e volumes. O objetivo principal é gerar novos recursos para análise técnica e quantitativa, que podem ser usados em modelos de previsão ou em estratégias de análise de investimentos. O código é modularizado, com várias funções responsáveis por calcular diferentes tipos de indicadores, como médias móveis, volatilidade, momentum, risco, entre outros.

# Funções Principais

**1. feature\_engineering\_pipeline(df, symbol, market\_returns=None):** Esta função principal orquestra o cálculo de todos os indicadores e métricas financeiras. Ela começa validando os dados de entrada, depois chama várias outras funções para adicionar diferentes indicadores ao DataFrame. Esses indicadores incluem médias móveis, Ichimoku Cloud, VW-MACD, estatísticas avançadas e métricas de risco. Por fim, a função retorna o DataFrame com as novas colunas calculadas, após remover quaisquer valores ausentes.

**Explicação do Pipeline:**

1. **Validação de Dados (validate\_inputs)**:
   * A função validate\_inputs é chamada para garantir que o DataFrame tenha todos os dados necessários e que não haja inconsistências. Isso pode incluir verificar a presença de colunas essenciais, como 'close', 'high', 'low', etc.
2. **Cálculo dos Indicadores Principais (calcular\_indicadores)**:
   * A função calcular\_indicadores calcula indicadores técnicos fundamentais, como **médias móveis**, **RSI**, **MACD**, etc. Esses indicadores são essenciais para identificar a tendência e a força de um ativo no mercado.
3. **Adição do Indicador Ichimoku Cloud (add\_ichimoku\_cloud)**:
   * O indicador **Ichimoku Cloud** é uma ferramenta de análise técnica que fornece sinais de compra e venda com base na posição das linhas da nuvem. A função calcula e adiciona esses componentes ao DataFrame.
4. **Adição do VW-MACD (add\_vwmacd)**:
   * O **VW-MACD** é uma versão do **MACD** que pondera os preços pelo **volume** de negociação. Ele é mais sensível às mudanças de volume e pode ajudar a identificar sinais de reversão de tendência mais cedo.
5. **Adição de Estatísticas Avançadas (add\_advanced\_statistics)**:
   * Funções como **Exponente de Hurst**, **assimetria** e **curtose** do volume ajudam a analisar o comportamento da série temporal em termos de sua distribuição, tendência e intensidade de flutuações.
6. **Adição de Métricas de Risco (add\_risk\_metrics)**:
   * As métricas de risco como **drawdown** (a perda máxima a partir de um pico histórico) e **VaR (Value at Risk)** são essenciais para avaliar a exposição do ativo a possíveis perdas.
7. **Adição de Features de Aprendizado de Máquina (add\_ml\_features)**:
   * Lags e outras variáveis temporais são calculadas e adicionadas ao DataFrame para ajudar modelos de aprendizado de máquina a capturar as dependências temporais (ou seja, o impacto do passado no futuro).
8. **Adição de Métricas Financeiras Adicionais (add\_additional\_metrics)**:
   * **Métricas financeiras** como o **Sharpe ratio** (que mede o retorno ajustado pelo risco) e o **Sortino ratio** (que foca em perdas) são adicionadas para avaliar o desempenho e risco do ativo.
9. **Remoção de Valores Ausentes (dropna())**:
   * Após todos os cálculos e a adição de novas colunas, a função garante que não haja valores ausentes no DataFrame final. Isso é essencial para evitar problemas durante a modelagem e análise.

**2. calcular\_indicadores(df, symbol):** Esta função calcula uma série de indicadores técnicos, como médias móveis (SMA e EMA), RSI, MACD, Bandas de Bollinger, entre outros. Também calcula estatísticas de volatilidade e adiciona informações como os preços de bid, ask e preço médio. Ela utiliza a biblioteca TA-Lib para calcular esses indicadores, além de outros cálculos manuais para detectar o comportamento do ativo no mercado.

**Explicação da Função calcular\_indicadores:**

1. **Objetivo**:
   * A função calcular\_indicadores tem como objetivo calcular uma série de indicadores técnicos e adicionar novas colunas ao **DataFrame** fornecido. Esses indicadores são comumente usados na análise técnica para avaliar o comportamento dos ativos financeiros.
2. **Adição de Dados de Bid, Ask e Preço Médio**:
   * **Bid e Ask** são os preços de compra e venda de um ativo. A função utiliza o mt5.symbol\_info\_tick(symbol) para obter esses preços diretamente de uma plataforma de negociação MetaTrader 5 (MT5).
   * O preço médio (last\_price) é calculado como a média entre os preços de bid e ask.
3. **Indicadores Básicos**:
   * A **volatilidade** é calculada como a diferença entre o preço máximo e o preço mínimo do ativo no período.
   * O **percentual de mudança** entre o preço de fechamento e o preço de abertura é calculado para observar a variação percentual no preço do ativo.
   * O **preço médio** é a média simples dos preços de abertura, máxima, mínima e fechamento.
4. **Cálculo das Médias Móveis**:
   * A função utiliza a função calculate\_moving\_average para calcular tanto **SMA** (Média Móvel Simples) quanto **EMA** (Média Móvel Exponencial) para diferentes períodos. Isso ajuda a observar a tendência do preço em vários horizontes de tempo.
5. **Indicadores Técnicos Avançados**:
   * **RSI** (Índice de Força Relativa): Mede se um ativo está sobrecomprado ou sobrevendido. São calculados para 14 e 9 períodos.
   * **MACD** (Moving Average Convergence Divergence): Calcula a diferença entre duas médias móveis e ajuda a identificar possíveis reversões de tendência.
   * **Bandas de Bollinger**: Usadas para medir a volatilidade e identificar zonas de sobrecompra ou sobrevenda.
   * **ATR** (Average True Range): Mede a volatilidade do ativo.
   * **ROC** (Rate of Change): Mede a variação percentual do preço.
   * **OBV** (On-Balance Volume): Mede a pressão de compra e venda com base no volume.
   * **MFI** (Money Flow Index): Avalia a pressão de compra e venda, ponderada pelo volume.
   * **ADX** (Average Directional Index): Mede a força da tendência.
6. **VWAP** (Volume Weighted Average Price):
   * O **VWAP** é calculado como a média ponderada do preço de fechamento pelo volume, o que ajuda a identificar a tendência do preço em relação ao volume de negociação.
7. **Médias Móveis do OBV**:
   * São calculadas médias móveis do OBV para diferentes períodos (10, 20, 50, 100), proporcionando uma visão mais detalhada da pressão de compra e venda ao longo do tempo.
8. **Tratamento de Exceções**:
   * Caso ocorra algum erro durante o cálculo dos indicadores (por exemplo, se algum dado estiver faltando ou algum cálculo falhar), o erro é registrado no log, garantindo a rastreabilidade do problema.

**3. add\_ichimoku\_cloud(df):**  
 A função adiciona os indicadores do Ichimoku Cloud ao DataFrame. A nuvem Ichimoku é composta por quatro componentes (linha de conversão, linha base, leading span A e leading span B), que são usados para detectar a direção da tendência e potenciais sinais de compra ou venda. Essas linhas são calculadas com base em máximos e mínimos de diferentes períodos.

**Explicação dos indicadores:**

* **ichi\_conversion**: A linha de conversão (Tenkan-sen) é calculada com a média do maior máximo e menor mínimo nos últimos 9 períodos. Essa linha é usada para identificar a tendência de curto prazo.
* **ichi\_base**: A linha base (Kijun-sen) é calculada com a média do maior máximo e menor mínimo nos últimos 26 períodos. Ela é usada para identificar a tendência de médio/longo prazo.
* **ichi\_leading\_span\_a**: O Leading Span A (Senkou Span A) é a média das linhas de conversão e base, deslocada 26 períodos à frente. Esse valor ajuda a identificar a direção da tendência futura.
* **ichi\_leading\_span\_b**: O Leading Span B (Senkou Span B) é a média do maior máximo e menor mínimo nos últimos 52 períodos, também deslocado 26 períodos à frente. Esse indicador, junto com o Leading Span A, forma a "nuvem" Ichimoku, que é usada para determinar os níveis de suporte e resistência.

**4. add\_vwmacd(df):** A função adiciona o indicador VW-MACD ao DataFrame, que é uma versão do MACD, mas ponderada pelo volume de negociação. O cálculo é feito a partir do preço ponderado pelo volume (VWAP), o que torna o MACD mais sensível a mudanças no volume de negociação.

**Explicação do add\_vwmacd**:

* **VW-MACD**: O VW-MACD é uma variação do tradicional MACD (Moving Average Convergence Divergence), mas aqui ele usa o **preço ponderado pelo volume** (VW), ao invés de apenas o preço de fechamento.
* **Cálculo do VW-Preço**: O preço ponderado pelo volume é calculado somando o produto do preço de fechamento com o volume de cada período e dividindo pela soma acumulada do volume.
* **Cálculo do MACD**: O MACD é calculado usando o preço ponderado pelo volume, permitindo uma análise mais sensível à relação entre o volume e o preço.

5. add\_market\_regime(df):  
 Esta função adiciona indicadores que ajudam a avaliar o regime de mercado, como a força da tendência (utilizando o ADX) e a volatilidade do mercado. A volatilidade é comparada com a média móvel de 50 períodos para determinar se o mercado está em um regime de alta ou baixa volatilidade.

6. add\_basic\_features(df):  
 Adiciona características básicas ao DataFrame, como volatilidade, mudança percentual e preço médio. Também calcula o retorno logarítmico e o Z-score da volatilidade para avaliar variações extremas nos preços.

**7. add\_advanced\_statistics(df):**  
 Adiciona estatísticas avançadas, como assimetria e curtose do volume, além do cálculo do Exponente de Hurst, utilizado para avaliar a persistência de tendências no mercado. Essas estatísticas fornecem uma visão detalhada sobre a distribuição e a comportamento do volume de negociação.

**Explicação dos Indicadores:**

1. **volume\_skew (Assimetria do Volume)**:
   * Mede a assimetria da distribuição do volume nos últimos 30 períodos.
   * Uma assimetria positiva indica que a distribuição do volume tem mais volumes grandes (cauda à direita), e uma assimetria negativa indica que a maioria dos volumes está mais concentrada em valores baixos.
2. **volume\_kurtosis (Curtose do Volume)**:
   * Mede a "agudeza" ou "altura" das caudas da distribuição de volume.
   * A curtose ajuda a identificar a presença de **outliers** no volume. Uma curtose alta indica que o volume tem picos significativos (outliers), enquanto uma curtose baixa indica que a distribuição é mais suave.
3. **hurst\_30 (Exponente de Hurst)**:
   * O **Exponente de Hurst** é utilizado para medir a persistência de tendência de uma série temporal.
   * O valor do Exponente de Hurst varia entre 0 e 1:
     + **Hurst ~ 0.5**: A série é **aleatória** (sem tendência).
     + **Hurst > 0.5**: A série apresenta **tendência persistente** (continua em tendência).
     + **Hurst < 0.5**: A série possui **comportamento revertido** (tende a reverter).
   * A função calcula o Exponente de Hurst utilizando uma técnica de "R/S" (Rescaled Range), ajustando uma linha reta aos dados transformados.

**8. add\_risk\_metrics(df):**  
 Esta função adiciona métricas de risco, como o Expected Shortfall (ES) e o Omega Ratio, que são usados para avaliar o risco e a exposição a perdas no mercado.

**Explicação do add\_risk\_metrics**:

* **Expected Shortfall (ES)**: O **Expected Shortfall (ES)**, também conhecido como **Conditional VaR** (Value at Risk), é uma medida de risco que calcula a **média dos piores retornos** (percentil 5). Ele é útil para avaliar o risco extremo de uma série temporal de retornos.
* **Omega Ratio**: O **Omega Ratio** mede o potencial de **ganhos versus perdas** de um ativo. Ele calcula a razão entre a soma dos retornos positivos e a soma dos retornos negativos. Quanto maior o Omega Ratio, melhor o perfil de risco-retorno do ativo.

9. add\_moving\_averages(df):  
 Adiciona médias móveis simples (SMA) para diferentes períodos (9, 21, 50, 100, 200). As médias móveis são usadas para identificar tendências de curto, médio e longo prazo no mercado.

**Explicação da add\_moving\_averages**:

* **Médias Móveis Simples (SMA)**: As **médias móveis** são usadas para suavizar as flutuações dos preços e ajudar a identificar a tendência geral de um ativo. Elas são calculadas com a média dos preços de fechamento de um determinado número de períodos.
  + **SMA de 9 períodos**: A média dos últimos 9 preços de fechamento.
  + **SMA de 21 períodos**: A média dos últimos 21 preços de fechamento.
  + **SMA de 50, 100, 200 períodos**: Médias móveis de diferentes períodos ajudam a identificar a tendência de curto, médio e longo prazo.
* **Uso do ta.SMA**: A função ta.SMA() da biblioteca TA-Lib é usada para calcular as médias móveis simples para os períodos definidos.
* **Tratamento de Exceções**: Caso algum erro ocorra durante o cálculo das médias móveis, a função irá registrar o erro no log usando logging.exception().

**10. add\_dynamic\_indicators**(df, config={'sma\_periods': [9,21,50], 'rsi\_period': 14}):  
 Adiciona indicadores dinâmicos ao DataFrame, como médias móveis (SMA) e o Índice de Força Relativa (RSI). A função permite personalizar os períodos dessas médias móveis e do RSI.

**Explicação da add\_dynamic\_indicators**:

* **Indicadores Dinâmicos**: Esta função é projetada para ser mais flexível do que a anterior, permitindo que os períodos das médias móveis e do **Índice de Força Relativa (RSI)** sejam configurados dinamicamente através do parâmetro config.
* **Médias Móveis Simples (SMA)**: Assim como na função anterior, as médias móveis simples são calculadas, mas agora os períodos são passados através do parâmetro config['sma\_periods'], o que permite maior flexibilidade. A função itera sobre a lista de períodos para calcular as SMAs e adicioná-las ao DataFrame.
* **Índice de Força Relativa (RSI)**: O **RSI** é calculado para o período especificado em config['rsi\_period']. O RSI é um indicador de momentum que mede a velocidade e a mudança dos movimentos de preço. É geralmente usado para identificar condições de sobrecompra ou sobrevenda.
* **Uso de ta.SMA e ta.RSI**: A função utiliza as funções ta.SMA() para calcular as médias móveis e ta.RSI() para calcular o RSI.

**11. add\_volatility\_indicators(df):** Adiciona indicadores de volatilidade como o ATR (Average True Range), Bandas de Bollinger, Keltner Channels e volatilidade rolling. Esses indicadores ajudam a medir a magnitude das flutuações de preço no mercado.

**Explicação dos Indicadores:**

1. **ATR (Average True Range)**:
   * O **ATR** é um indicador de volatilidade que calcula a média da diferença entre o **high** e o **low** de um ativo, considerando também gaps de preços de um dia para o outro.
   * O **ATR de 14 e 50 períodos** ajuda a medir a volatilidade de curto e longo prazo, respectivamente.
2. **Volatilidade Rolling de 20 períodos**:
   * Este indicador calcula o **desvio padrão** dos preços de fechamento nos últimos 20 períodos, que serve como uma medida da **volatilidade histórica**. Quanto maior o desvio padrão, maior a volatilidade.
3. **Bandas de Bollinger**:
   * As **Bandas de Bollinger** consistem em uma média móvel simples (SMA) com duas bandas (superior e inferior) calculadas com base em dois desvios padrão.
   * A **largura das bandas** (bb\_width) ajuda a entender a volatilidade: se as bandas estão mais distantes, há maior volatilidade; se estão mais próximas, há menor volatilidade.
   * O **Percentual B** (bollinger\_percent\_b) indica a posição do preço em relação às bandas. Um valor próximo de 1 indica que o preço está próximo da banda superior (potencial sobrecompra), enquanto um valor próximo de 0 indica que está próximo da banda inferior (potencial sobrevenda).
4. **Keltner Channels**:
   * Os **Keltner Channels** consistem em uma média exponencial (EMA) com bandas superior e inferior baseadas no **ATR**.
   * O **keltner\_break** é um sinal que indica se o preço está rompendo as bandas superior ou inferior, sugerindo uma possível mudança de tendência ou confirmação de uma tendência forte.

**12. add\_momentum\_indicators(df):**  
 Adiciona indicadores de momentum como o ROC, Williams %R, RSI, MACD, estocástico, ADX, CCI e outros. Esses indicadores ajudam a medir a força de uma tendência e identificar condições de sobrecompra ou sobrevenda no mercado.

**xplicação dos Indicadores de Momentum:**

1. **ROC (Rate of Change)**:
   * **roc\_5 e roc\_20**: O ROC calcula a variação percentual do preço de um ativo entre dois períodos. Indicadores de **momentum** como o ROC ajudam a identificar a força e a velocidade das mudanças de preço.
2. **Williams %R**:
   * **williams\_r\_14**: O **Williams %R** é um oscilador que compara o preço de fechamento com a faixa de preço durante um período de tempo, ajudando a identificar condições de **sobrecompra** ou **sobrevenda**.
3. **Momentum**:
   * **momentum\_10**: O **indicador de momentum** mede a **velocidade** da mudança nos preços. Quanto maior o valor do momentum, mais forte a tendência de preço.
4. **RSI (Índice de Força Relativa)**:
   * **rsi\_14**: O **RSI** é um indicador de **momentum** que mede a velocidade e a mudança dos movimentos de preço. Valores acima de 70 indicam **sobrecompra**, enquanto valores abaixo de 30 indicam **sobrevenda**.
5. **MACD (Moving Average Convergence Divergence)**:
   * **macd**: O **MACD** é usado para identificar mudanças na **direção** e **força** de uma tendência. Ele é calculado subtraindo a média móvel exponencial de 26 períodos da média móvel exponencial de 12 períodos.
   * **macd\_signal**: A **linha de sinal** é uma média móvel exponencial do MACD.
   * **macd\_hist**: O **histograma do MACD** representa a diferença entre o MACD e sua linha de sinal.
6. **Estocástico**:
   * **stoch\_k e stoch\_d**: O **Oscilador Estocástico** compara o preço de fechamento com o intervalo de preços durante um período. A linha %K (rápida) e a linha %D (lenta) são usadas para identificar sinais de **sobrecompra** ou **sobrevenda**.
7. **ADX (Average Directional Index)**:
   * **adx\_14**: O **ADX** é usado para medir a **força de uma tendência**, independentemente da direção (se é uma tendência de alta ou baixa).
8. **CCI (Commodity Channel Index)**:
   * **cci\_14**: O **CCI** mede a variação do preço em relação à média, indicando condições de **sobrecompra** ou **sobrevenda**.
9. **Chande Momentum Oscillator**:
   * **cmo\_14**: O **CMO** mede o **momentum** e é usado para avaliar a força de uma tendência.
10. **Vortex Indicator**:
    * O **Vortex Indicator** (calculado pela função vortex\_indicator) mede a direção e a intensidade da **tendência**.
11. **Elder Ray Index**:
    * **bull\_power e bear\_power**: O **Elder Ray** calcula a diferença entre o preço máximo e a média exponencial (EMA), ajudando a identificar a **força dos touros** (bulls) e dos **ursos** (bears).

**13. add\_volume\_indicators(df):** Adiciona indicadores baseados em volume como o Chaikin Money Flow (CMF), Money Flow Index (MFI), VWAP, OBV, VROC, Volume Profile e Accumulation/Distribution Line (ADL). Esses indicadores são utilizados para avaliar o fluxo de volume e identificar movimentos significativos de compra ou venda.

**Explicação dos Indicadores Baseados em Volume:**

1. **Chaikin Money Flow (CMF)**:
   * O **Chaikin Money Flow (CMF)** calcula o fluxo de dinheiro com base no volume e nos preços relativos de fechamento, máximo e mínimo.
   * O **MFM** (Money Flow Multiplier) é calculado e, em seguida, multiplicado pelo **volume** para determinar o **fluxo de dinheiro**. A soma desse fluxo nos últimos 20 períodos é dividida pela soma do volume, formando o **CMF**.
   * **Interpretabilidade**: Valores positivos indicam que o fluxo de dinheiro é positivo, enquanto valores negativos indicam um fluxo de dinheiro negativo.
2. **Money Flow Index (MFI)**:
   * O **Money Flow Index (MFI)** é um indicador de **momentum** que avalia a pressão de compra e venda com base no volume e nos preços.
   * Ele pode ser usado para identificar condições de **sobrecompra** ou **sobrevenda**.
3. **VWAP (Volume Weighted Average Price)**:
   * O **VWAP** é o **preço médio ponderado pelo volume**. Ele é muito utilizado para determinar se o preço atual está acima ou abaixo da média ponderada, ajudando a identificar se o mercado está em uma tendência de alta ou baixa.
   * **VWAP** é frequentemente usado em negociações diárias para identificar o preço de **equilíbrio**.
4. **On-Balance Volume (OBV)**:
   * O **OBV** é um indicador de **volume acumulado**, calculado somando ou subtraindo o volume dependendo se o preço de fechamento subiu ou desceu. Ele ajuda a identificar se o volume confirma ou não a tendência do preço.
5. **Volume Rate of Change (VROC)**:
   * O **VROC** mede a **variação percentual do volume** ao longo do tempo, fornecendo insights sobre mudanças no fluxo de volume. Ele pode ser útil para identificar uma mudança no **interesse dos investidores** ou na força de uma tendência.
6. **Volume Profile**:
   * O **Volume Profile** mostra o volume negociado em diferentes **intervalos de preços**, fornecendo informações sobre **suporte** e **resistência**. Se o volume for maior em certos níveis de preço, isso pode indicar um suporte ou resistência significativa.
7. **Accumulation/Distribution Line (ADL)**:
   * O **ADL** calcula o **fluxo de dinheiro acumulado** usando o volume e as variações de preço. Ele ajuda a identificar o quanto um ativo está sendo acumulado ou distribuído ao longo do tempo.
   * A linha ADL é um bom indicador de divergências com o preço.
8. **VWAP RSI**:
   * O **VWAP RSI** combina o **RSI** com o **VWAP** para calcular uma **versão ponderada pelo volume** do Índice de Força Relativa. Ele dá mais peso aos preços com maior volume, o que pode ser útil para filtrar sinais de sobrecompra ou sobrevenda mais **relevantes**.

**14. add\_crossover\_indicators(df):** Adiciona indicadores de cruzamento de médias móveis, como o cruzamento entre as SMAs de 50 e 200 períodos, e entre as SMAs de 9 e 21 períodos. Esses cruzamentos são usados para identificar possíveis mudanças de tendência no mercado.  
**Explicação dos Indicadores de Cruzamento de Médias Móveis (SMA):**

1. **Cruzamento SMA 50 e 200**:
   * O **cruzamento da SMA de 50 períodos com a SMA de 200 períodos** é uma das estratégias mais comuns de análise técnica para identificar **tendências de longo prazo**.
   * Quando a **SMA de 50 períodos** (curto prazo) cruza para **acima** da **SMA de 200 períodos** (longo prazo), isso é visto como um **sinal de compra**, indicando uma tendência de **alta** no mercado.
   * Quando a **SMA de 50 períodos** cruza para **abaixo** da **SMA de 200 períodos**, isso é visto como um **sinal de venda**, indicando uma tendência de **baixa**.
   * **sma\_diff\_50\_200**: Calcula a diferença entre a SMA de 50 e 200 períodos.
   * **sma\_cross\_signal\_50\_200**: Aplique a função lambda para gerar um sinal de **compra (1)** ou **venda (-1)**, dependendo de a SMA de 50 períodos estar acima ou abaixo da SMA de 200. A função diff() é usada para detectar mudanças no cruzamento.
2. **Cruzamento SMA 9 e 21**:
   * O **cruzamento da SMA de 9 períodos com a SMA de 21 períodos** é um indicador de **curto prazo** frequentemente usado para identificar **entradas e saídas rápidas** do mercado.
   * Quando a **SMA de 9 períodos** cruza para **acima** da **SMA de 21 períodos**, isso é considerado um **sinal de compra**, indicando uma tendência de **alta** no curto prazo.
   * Quando a **SMA de 9 períodos** cruza para **abaixo** da **SMA de 21 períodos**, isso é considerado um **sinal de venda**, indicando uma tendência de **baixa** no curto prazo.
   * **sma\_diff\_9\_21**: Calcula a diferença entre a SMA de 9 e 21 períodos.
   * **sma\_cross\_signal\_9\_21**: Gera um sinal de **compra (1)** ou **venda (-1)** baseado na direção do cruzamento da SMA de 9 com a SMA de 21, usando a função diff() para capturar o momento exato do cruzamento.

**15. add\_lag\_features(df, lags=[1, 2, 3]):** Adiciona os lags para os preços low, high, open e close ao DataFrame. Lags são usados para capturar as observações anteriores de uma série temporal e são úteis em modelos de previsão.

**Explicação da Função add\_lag\_features:**

1. **Objetivo**:
   * A função **adiciona colunas de lags** para os preços **low**, **high**, **open** e **close** no DataFrame. Lags são **observações passadas** da série temporal, e elas são úteis em modelos de previsão, pois ajudam a capturar a **dependência temporal** dos dados.
2. **Parâmetros**:
   * **df**: O DataFrame contendo as colunas de preços **low**, **high**, **open**, e **close**.
   * **lags**: Lista de períodos de lags que o usuário deseja adicionar. O valor padrão é [1, 2, 3], o que significa que as colunas de preços serão deslocadas para 1, 2 e 3 períodos atrás.
3. **Operação**:
   * A função percorre os valores de **lag** (1, 2 e 3) e, para cada valor, cria uma nova coluna para cada preço (low, high, open, close). A função shift(i) do pandas é usada para **deslocar os dados** para os **i períodos anteriores**, criando os lags.
   * Para cada preço (low, high, open, close), são criadas novas colunas como low\_lag1, high\_lag1, etc., que contêm os valores de **1, 2 e 3 períodos atrás**.
4. **Concatenando as Colunas**:
   * As colunas criadas para os lags são **concatenadas ao DataFrame original** utilizando a função pd.concat(). Isso adiciona as novas colunas de lags ao DataFrame original, mantendo os dados anteriores.
5. **Exceções**:
   * Caso ocorra algum erro durante o processo, ele será registrado no log de exceções com a mensagem "Erro em add\_lag\_features".

**Exemplos de uso dos lags:**

* **Lag 1**: Os preços de 1 período atrás.
* **Lag 2**: Os preços de 2 períodos atrás.
* **Lag 3**: Os preços de 3 períodos atrás.

**Utilidade:**

* **Lags** são comumente usados em modelos de previsão para capturar a **dependência temporal**. Por exemplo, ao prever os preços futuros de uma ação, podemos usar os preços passados (lags) como **variáveis independentes** no modelo para prever o preço de fechamento futuro.

**Exemplo:**

Se tivermos um DataFrame com os preços de **low**, **high**, **open** e **close**, após aplicar a função add\_lag\_features(df), o DataFrame resultante terá colunas adicionais como:

* low\_lag1, high\_lag1, open\_lag1, close\_lag1
* low\_lag2, high\_lag2, open\_lag2, close\_lag2
* low\_lag3, high\_lag3, open\_lag3, close\_lag3

Essas novas colunas são úteis para analisar como os preços passados podem influenciar os preços futuros.

**16. add\_hull\_moving\_average(df):** Adiciona as médias móveis de Hull (HMA) ao DataFrame para diferentes períodos (9, 14, 21, 50 e 100). A HMA é uma versão mais suavizada da média móvel ponderada, que visa reduzir o atraso nas respostas às mudanças de preço.  
**Explicação da Função add\_hull\_moving\_average:**

1. **Objetivo**:
   * A função **adiciona as médias móveis de Hull (HMA)** ao DataFrame para diferentes períodos (9, 14, 21, 50 e 100). A **HMA** é uma versão aprimorada da média móvel simples (SMA) ou ponderada (WMA), projetada para reduzir o **atraso** e **aumentar a suavização**.
2. **Cálculo da HMA**:
   * **Média Móvel de Hull (HMA)** é uma combinação de **Médias Móveis Ponderadas (WMA)**. A HMA tem como objetivo ser mais rápida, sem sacrificar a suavização. O cálculo da HMA utiliza a fórmula abaixo:

HMA=WMA(2×WMA(close,n/2)−WMA(close,n)),nHMA = WMA(2 \times WMA(\text{close}, n/2) - WMA(\text{close}, n)), \sqrt{n}HMA=WMA(2×WMA(close,n/2)−WMA(close,n)),n​

Onde:

* + - nnn é o período da HMA.
    - **WMA** é a média móvel ponderada.
  + A função hull\_ma (presumivelmente definida em outro lugar no seu código) é responsável por calcular a HMA para um determinado período.

1. **Períodos de Cálculo**:
   * A função calcula a HMA para os seguintes períodos: **9, 14, 21, 50 e 100**. Esses são períodos comuns usados para capturar diferentes horizontes de tempo, desde o curto até o longo prazo.
2. **Concatenando as Colunas**:
   * Após calcular as HMAs para os diferentes períodos, as novas colunas (hma\_9, hma\_14, hma\_21, hma\_50, hma\_100) são **adicionadas ao DataFrame** utilizando pd.concat(). Isso garante que todas as colunas de HMA sejam adicionadas ao DataFrame de uma vez, de forma eficiente.
3. **Tratamento de Exceções**:
   * Caso algum erro ocorra durante o processo (por exemplo, se a função hull\_ma falhar), o erro é registrado usando logging.exception(). Isso ajuda a rastrear e depurar possíveis problemas no cálculo das HMAs.
4. **Resultado**:
   * O **DataFrame** é retornado com as novas colunas de HMA adicionadas. Isso permite que você use as HMAs para análise de tendências e para identificar mudanças de direção nos preços.

**17. add\_risk\_features(df):**  
 Adiciona métricas de risco como Drawdown, Value at Risk (VaR) e Estatísticas de ADF para avaliar o risco de mercado e a exposição a perdas.  
**Explicação dos Indicadores de Risco:**

1. **Cumulative Max (Máximo Acumulado)**:
   * **cum\_max**: O **máximo acumulado** dos preços de fechamento ao longo do tempo. Ele é calculado utilizando o método cummax(), que fornece o valor máximo de fechamento até cada ponto no tempo. Este indicador é útil para ver o valor mais alto alcançado em qualquer ponto no passado.
2. **Drawdown**:
   * **drawdown**: O **drawdown** é a diferença entre o valor máximo acumulado (de cum\_max) e o preço de fechamento atual. Ele é normalizado pelo valor máximo acumulado, expressando o tamanho da **perda relativa** em relação ao valor mais alto histórico. Este indicador ajuda a avaliar a magnitude das **quedas** em relação ao pico histórico de um ativo.
     + Um **drawdown** alto indica que o ativo sofreu uma grande perda em relação ao seu valor mais alto.
     + Um **drawdown** baixo indica que o ativo tem uma recuperação mais estável, sem grandes perdas em relação aos picos.
3. **Value at Risk (VaR) de 95%**:
   * **var\_95**: O **Value at Risk (VaR)** de 95% é uma medida de risco que estima a **perda máxima** que um ativo pode ter com 95% de confiança em um determinado período. No código, o VaR é calculado como o **percentil 5** dos **retornos logarítmicos** dos últimos 20 períodos.
     + O **VaR de 95%** indica a **perda máxima esperada** no pior 5% dos cenários.
4. **Estatística de ADF (Augmented Dickey-Fuller)**:
   * **adf\_stat\_30**: O teste **ADF** (Augmented Dickey-Fuller) verifica se uma série temporal é **estacionária**. Uma série estacionária não tem tendência e a média e a variância são constantes ao longo do tempo. O teste ADF ajuda a entender a **natureza da série temporal** (se ela contém uma tendência de longo prazo ou não).
     + Se o valor da estatística ADF for **negativo** o suficiente, a hipótese nula (de que a série tem uma raiz unitária) é rejeitada, indicando que a série é **estacionária**.