

BovDBv2: Atualização e Expansão do Conjunto de Dados para Análise Intraday

Souza, Andrey Santos; Lucca, Giancarlo;
Borges, Eduardor Nunes; Dalmazo, Bruno Lopes

Berri, Rafael Alceste
andreyvinicius@furg.com
Universidade Federal do Rio Grande – Furg

Palavras-chave: Banco de Dados; Intraday; Séries Temporais; Ações.

1 Introdução

Com o avanço das novas tecnologias emergentes aplicadas ao mercado financeiro, como técnicas de *machine learning*, o uso de análises técnicas e estatísticas tornou-se fundamental para a compreensão das dinâmicas do mercado de ações. Contudo, apesar do uso crescente dessas tecnologias, a complexidade dos dados e a falta de informações acessíveis e organizadas podem dificultar a realização de análises precisas e previsões robustas.

Com o intuito de atender a essa demanda, foi criado o **BovDBv2**, uma versão aprimorada do dataset anterior, BovDB(CARDOSO et al., 2022). Este novo conjunto de dados é mais abrangente, contendo informações sobre a precificação diária e dados de títulos negociados intrapregão, ou seja, durante o horário normal das negociações, com uma nova tabela que armazena esses **dados intraday**. Esses dados representam as flutuações dos preços dos ativos ao longo do dia em intervalos regulares de tempo, permitindo uma análise mais detalhada do comportamento dos preços em curto prazo (menos de um dia). Além disso, o BovDBv2 inclui atualizações de todas as empresas e títulos listados na B3 entre 1995-01-01 e 2024-06-28.

2 Detalhamento do Dataset Intraday

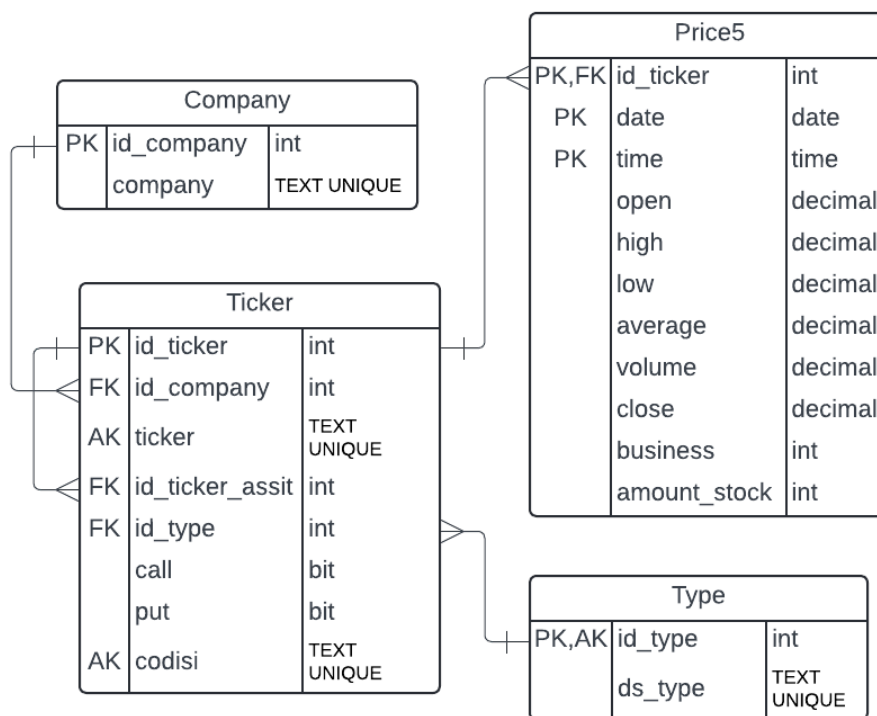
Para a formação do conjunto de dados, utilizamos um sistema de gerenciamento de banco de dados relacional (SGBDR) conhecido como SQLite¹(ALLEN; OWENS, 2010), o mesmo empregado na versão anterior. As atualizações do novo dataset são compostas por quatro tabelas diferentes, descritas a seguir:

- **Company:** Armazena dados referentes às empresas.
- **Ticker:** Armazena dados de um determinado título.
- **Price5:** Armazena dados intradiários das negociações, incluindo informações detalhadas sobre preços e volumes em intervalos de 5 minutos ao longo do dia.
- **Type:** Classifica os tickers em diferentes categorias, como ações, BDRs, instrumentos financeiros e opções.

¹<<https://www.sqlite.org/index.html>>

A Figura 1 mostra como as tabelas do banco estão relacionadas. Essa estrutura facilita a integridade dos dados, assegurando que as atualizações em uma tabela se propaguem de maneira controlada para as demais tabelas, evitando inconsistências.

Figura 1: Modelo Relacional do Conjunto de Dados BovDBv2.



Fonte: Autoria própria.

Para realizar o levantamento dos dados, foi desenvolvida uma API personalizada que automatiza a busca de arquivos intradiários no formato TXT, disponibilizados pela B3² (Brasil Bolsa Balcão). Após o download, os arquivos são armazenados e processados. Foram utilizados scripts em Python para corrigir, formatar, organizar, analisar e inserir os dados nas tabelas do banco de dados. Além disso, foram realizados cálculos na tabela **Price5**, como preço médio (*average*), volume (*volume*) e quantidade negociada (*amount_stock*). É importante destacar que o **BovDBv2**, como uma versão atualizada do dataset anterior, aprimora as capacidades analíticas e expande as possibilidades ao incluir a categorização detalhada dos novos *tickers*, além da atualização das demais tabelas e da inclusão de dados intradiários.

3 Análise e discussões

A tabela **Price5** armazena informações sobre negociações diárias em intervalos de 5 minutos, oferecendo uma granularidade essencial para análises avançadas, como modelagem preditiva e estudos de volatilidade. Os aprimoramentos visam atender às

²<<https://www.b3.com.br>>

demandas por análises mais sofisticadas no mercado financeiro, uma vez que operações de compra e venda de ações em prazos curtos, como no day trade, tornaram-se bastante comuns.

Tabela 1: Amostra de dados da tabela Price5

id_ticker	date	time	open	close	high	low	average	volume	business	amountstock
107	2024-01-02	10:00:00	37.44	37.52	37.59	37.4	37.49	958964.4	443	522500
107	2024-01-02	10:05:00	37.52	37.58	37.59	37.5	37.56	623421.7	171	166000
107	2024-01-02	10:10:00	37.58	37.58	37.64	37.52	37.59	72483.2	212	285300
107	2024-01-02	10:15:00	37.5	37.5	37.57	37.48	37.51	832453.1	229	221900
107	2024-01-02	10:20:00	37.5	37.5	37.54	37.49	37.51	232947.4	69	62100
...

- **id_ticker**: Identificador único do título negociado.
- **date**: Data da negociação.
- **time**: Horário específico da negociação em intervalos de 5 minutos.
- **open**: Preço de abertura do título no intervalo.
- **close**: Preço de fechamento do título no intervalo.
- **high**: Preço máximo atingido pelo título no intervalo.
- **low**: Preço mínimo atingido pelo título no intervalo.
- **average**: Preço médio das negociações realizadas no intervalo.
- **volume**: Volume total negociado do intervalo.
- **business**: Número de negócios realizados no intervalo.
- **amount_stock**: Quantidade total de ações negociadas no intervalo.

Uma visualização detalhada mostra que os dados disponibilizados são cruciais para o estudo do mercado de ações. Para esses estudos, são utilizados diversos tipos de conjuntos de dados que dependem de diferentes janelas de tempo. A tabela **Price5** abre caminho para uma série de aplicações, tanto em conjunto quanto específicas, incluindo a plotagem de gráficos de candlestick para análise de padrões gráficos e a construção de indicadores técnicos, como médias móveis simples e exponenciais, além de bandas de Bollinger, como exemplificado na Figura 2.

O gráfico demonstra como o BovDBv2 pode ser utilizado para fornecer uma visão detalhada da oscilação dos preços de abertura e fechamento em intervalos de 5 minutos ao longo do dia, tornando-se fundamental para a tomada de decisões rápidas em operações de curto prazo. O uso de indicadores, como médias móveis simples (SMA) e exponenciais (EMA), juntamente com as Bandas de Bollinger, ajuda a suavizar as flutuações de preço, identificar tendências e períodos de maior volatilidade no mercado.

Figura 2: Candlestick PETR4 com SMA, EMA e Bandas de Bollinger - 2024-05-29.



Fonte: Autoria própria.

4 Considerações finais

O BovDBv2, como a inclusão de dados intraday, proporcionam uma base de dados mais detalhada e precisa, essencial para análises de curto prazo, como operações de day trade e modelagem preditiva, o tratamento dos dados intraday permite o desenvolvimento de ferramentas analíticas e algoritmos de machine learning para prever tendências de mercado. Durante o desenvolvimento dessas atividades, foi possível aprofundar o conhecimento sobre técnicas de processamento de dados e a importância da consistência e precisão na construção de conjuntos de dados robustos e úteis para a comunidade acadêmica e o mercado financeiro.

5 REFERÊNCIAS

ALLEN, G.; OWENS, M. **The Definitive Guide to SQLite**. New York, USA: Apress, 2010.

CARDOSO, F. C. et al. Bovdb: a data set of stock prices of all companies in b3 from 1995 to 2020. **Journal of Information and Data Management**, Local, v. 13, n. 1, p. 78–92, jan. 2022. Disponível em: <<https://doi.org/10.5753/jidm.2022.2345>>. Acesso em: 30 agost. 2024.