

Félévi NagyHF Dokumentáció

Üzleti Intelligencia Labor

[Félév: 2025 tavasz]

S&P 500 Index Elemzés

Dancsó Marcell - (AZTVS7)

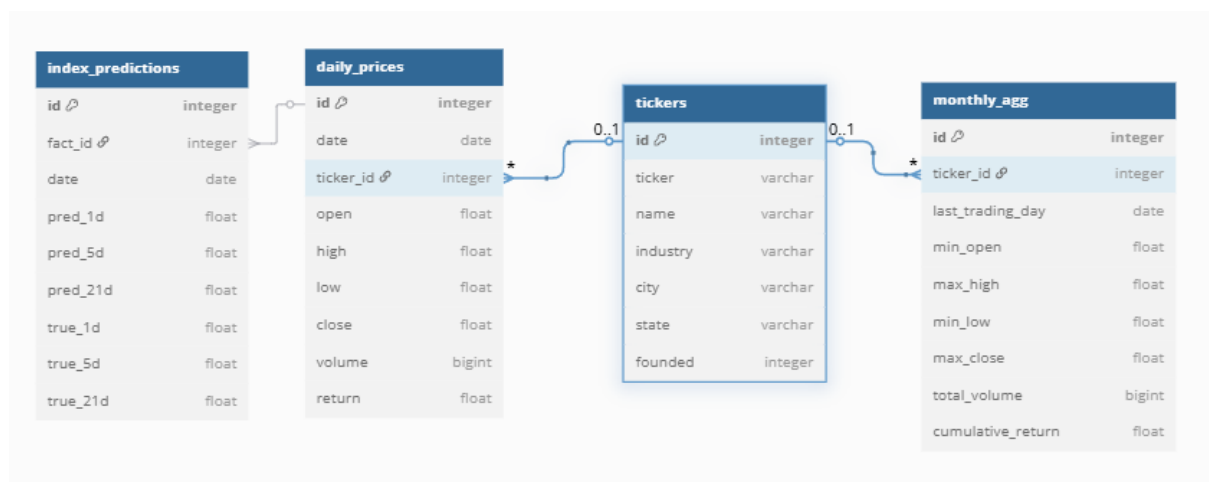
dancsomarci@gmail.com

Adathalmaz

Az S&P 500 Index 500+ vállalat gazdasági teljesítményét aggregálja. Az elemzéshez szükséges adatokat a [yfinance](#) python library segítségével kértem le dinamikusan. Ebbe beletartozik maga az index és az egyes céges idősoros adatai. A letöltött adatot csv formátumban kezeltem pandas library-vel. Az egyes céges részletes adatai [datahub](#)-ról származnak.

ETL

DB Séma:



Néhány ponton tervezői döntésből kifolyólag denormalizáltan tárolom az adatokat. Tény táblának minősül a „daily_prices” és dimenzióknak a „tickers”. A pontos típusok és kényszerek leírását a „create_db_schemas.sql” file tartalmazza.

Ütemezés: A választott technológiám: SSIS. Az adatok streamelt betöltését egy variable segítségével oldottam meg: az ütemezett ETL job csak adott dátum előtti értékeket tölthet be. Lehetett volna automatizálni a python script futtatását, amivel dinamikusan lehet letölteni az adatokat, de ki akartam próbálni ezt a variable-ös megoldást is. (máshol lesz még python script:)

ETL jobok/Transzformációk:

- Tickerek betöltése
 - Dinamikusan ellenőrzi mi van már betöltve az adatbázisba (Load + Merge Join a kis táblaméret miatt)
 - Szűrés + Transzformációk:
 - Állam kinyerése szövegből
 - Alapítás éve (megjegyzés levágása + parse)
 - Szimbólumok validálása (no whitespace)
- Napi árfolyamok betöltése
 - Max dátum dinamikus lekérése -> változóban tárolása
 - Változó, illetve egyéb formai kényszerek (semmi nem NULL) alapján szűrés

- Előző job-ban feltöltött sorok generált id-ját Merge Join segítségével a „daily prices” táblához csatolja.
- Hónapos aggregáció frissítése
 - Komplex/Többlépcsős SQL Utasítással update
 - A be nem fejezett hónapok, illetve már aggregáltak kihagyása
- Előrejelzés
 - Dinamikus python script futtatás
 - DB lekérdezés (Index-hez hiányzó napi predikciók megtalálása)
 - 1, 5, 21 napos előrejelzés LSTM modellel.
 - Valós eredmények SQL utasítással történő frissítése (külön SSIS lépésben)
 - Párhuzamosítva az aggregáció frissítésével.

Adatelemzés

Cél: 1,5, 21 napos záró árfolyam előrejelzés, rövidtávú „day-trading” támogatására.

Adatbetöltés: pandas library használatával, illetve index adatokra szűrés.

Feature Engineering:

- Idő dimenziójú adatok ciklikus enkódolása
- Mozdó átlag, volatilitás származtatása
- 1, 5, 21 LAG bevezetése

EDA: Sok érdekes dolog nem derült ki, ami szembetűnő és megegyezik az intuícióval, hogy a „Return” szép normális, 0 várhatóértékkel.

SARIMA:

- Idősoros előrejelzés a napi záróárfolyamokra, csak az árfolyam figyelembevételével.
- Stacionaritás ellenőrzése (Dickiy-Fuller Teszt) + 1-szeres differenciálás után megfelelő az adatsor.
- ACF, PACF grafikon elemzésből látszik, hogy nem túl eredményes a modell.
- Potenciális továbbfejlesztésként a trendszűrést és szezonális elemzését látom.

Winter-Holt féle Exponenciális Simitás:

- Additív modell, additív szezonalitással és 252 napos (trading year) ciklusokkal egészen meggyőző trendvonalat adott néhány hónap távolságra.
- Bonyolultabb feature-ök kihasználása miatt nem ezt a modellt választottam végül.

LSTM:

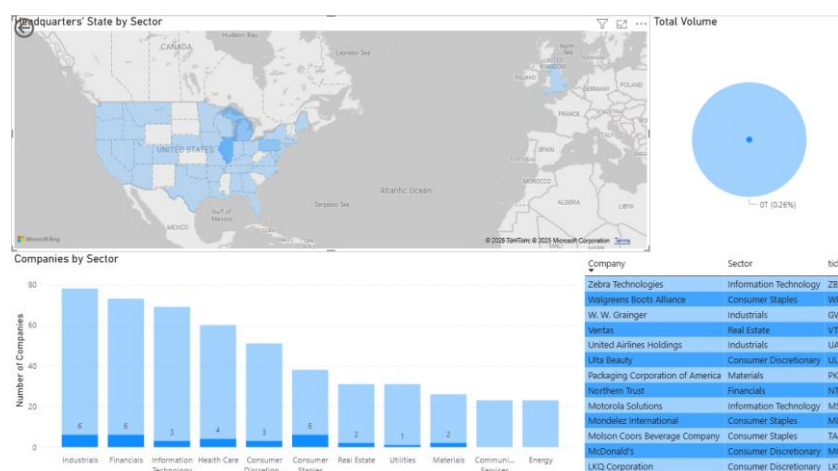
- Tran/Valid/Test split. (Az utóbbi évek elég zavarosak árfolyam tekintetében, ezért a vágási pontok: 2000-2022-2023-napjainkig)
- Standardizálás (scikitlearn StandardScaler)
- Empirikus hangolás után 10-es időablak használata hozta a legjobb eredményt. (21-es lag miatt kicsit messzebbre is lát a modell)

- 2 LSTM réteg egyenként 32 „unittel”.
- ModelCheckpoint és EarlyStopping használata
- Érdekesség: Múlt hónapban történt „érdekes” tőzsdei eseményekkel nem nagyon tudott mit kezdeni. (Manapság, aki tőzsdézik, annak érdemes X-et olvasgatni:)
- A kód notebook formájában érhető el a repóban.

Reportok

PowerBI Desktop alkalmazásban csatlakoztam a lokálisan MS-SQL adatbázisomhoz.

- **Ticker Market**
 - Cégenként szűrhető módon megtekinthető a napi záróárfolyam a teljes adatbázis által felölelt időtartamra.
 - Hónapos aggregált megtérülés, illetve annak számításához szükséges metrikák ábrázolása
- **Ticker Metrics**
 - Megtérülés és „Volume” ábrázolása
 - Évre, negyedévre szűrhető
 - Lefűrés „ticker” mentén -> Ticker Details
- **Ticker Details**
 - Pontos név és rövidítés megtekintése
 - Dinamikusan aggregált „Volume” és székhely térképes nézeten
- **Sectors**
 - Szektorok Államok szerinti megoszlása, illetve szektoronként Cégek megtekintése
 - Adott halmaz (szektor/állam) hozzájárulása a teljes „Volume”-hoz.
 - Lefűrés itt is alkalmazható az egyes cégekre (Ticker Details)
- **Forecasts**
 - 1 és 21 napos előrejelzés tekinthető meg
 - Valós és predikció összehasonlítása
 - Év/negyedévre szűrés



Sectors Report