

Zadání úlohy v rámci přijímacího řízení na pozici Quantitative/Machine Learning Analyst

Qminers, s.r.o.

1 Zadání

1. V příloze emailu jsou data akciového indexu S&P 500 za období 1. 1. 2000–31. 12. 2018.
2. Navrhněte a nakalibrujte několik různých modelů $\mathbb{E}[v_{d+1} \mid \mathcal{F}_d]$, kde $d = 1, \dots, n$ jsou obchodní dny, v_{d+1} je objem obchodů (volume) v den $d+1$ a \mathcal{F}_d je veškerá informace do dne d (včetně). Odhadněte přesnost modelů (jako kritérium použijte SSE, resp. R2) na testovacích (out-of-sample) datech a porovnejte tuto přesnost s referenčním modelem $\hat{\mathbb{E}}[v_{d+1}] = v_d$. Jako testovací množinu použijte data z období 1. 1. 2017–31. 12. 2018.

2 Poznámky

- Pro vypracování použijte programovací jazyk Python.
- Vedle informací obsažených přímo v časové řadě v_1, \dots, v_n použijte i další informace, jako např. sezónnost (objem může být závislý na ročním období) nebo kalendář očekávaných eventů (objem může záviset na relativní důležitosti daného eventu) apod.
- Zaměřte se na praktické vlastnosti modelu, tzn. vedle přesnosti na historických datech, analyzujte např. Vaši důvěru v použitelnost modelu na datech budoucích, robustnost, interpretovatelnost nebo chování při extrémních hodnotách vstupů.
- Upřednostňujeme kvalitu analýzy před kvantitou modelů.
- Uvědomujeme si, že dataset není zcela vhodný pro aplikaci některých populárních metod (obzvláště z oblasti strojového učení). Chytrá, nekonvenční řešení vítáme a dobře podloženou analýzu, proč daná třída modelů není vhodná, považujeme také za validní výsledek.
- Jako výstup očekáváme:
 - Technický report ve formátu PDF (k tvorbě reportu doporučujeme sázecí nástroj L^AT_EX) obsahující popis dat, formální popis modelu, tabulky a grafy s výsledky (předpokládaný rozsah 5–10 stran).
 - Zdrojové kódy (kvalita kódu bude součástí hodnocení).