

SZAKDOLGOZAT KIVONATA

Gépi tanulási algoritmusok alkalmazása pénzügyi adatsorokon

Béres Boglárka, Matematika BSc

Témavezetők: Dr. Ottucsák György, Morgan Stanley

Dr. Simon Károly, BME-TTK Sztochasztika Tanszék

Az utóbbi évtizedekben a tárolókapacitások méretének gyors növekedése a nyers, feldolgozatlan adatok tömeges méretű felhalmozását eredményezte. Az adatokban rejlő összefüggések feltárása, az adathalmazok analizálása központi szerepet tölt be a cégek nyereségességének növelésében.

A gépi tanulási algoritmusok célja az adathalmazra legjobban illeszkedő hipotézis megkeresése, amely alkalmazható jövőbeli adatokra is. Ezen algoritmusok egyik alkalmazása a tőzsdei termékek árának előrejelzése, amellyel e dolgozat keretein belül foglalkoztunk.

A szakdolgozat első részében áttekintettük a gépi tanulás irodalmát, különként a regressziós becslés témakörét. Az algoritmusok elméleti hátterének ismertetése mellett gyakorlati alkalmazásuk lehetőségét is részleteztük.

A dolgozat második részében pedig egy Kaggle versenyfeladat megoldásában alkalmaztuk a korábban megismert algoritmusokat, amelyeket Python programnyelven implementáltunk. A verseny feladata az volt, hogy előrejelezzük 198 tőzsdei termék árának százalékos változását a piacon elérhető adatok alapján két órával későbbi időpontra.

Először Benchmark algoritmusokkal foglalkoztunk, melyek közül a Utolsó Megfigyelt Érték Benchmark becslése bizonyult a legjobbnak. Majd egyszerű lineáris modelleket alkalmaztunk, amelyeknél a véletlen erdő algoritmus változataival jobb eredményt tudtunk elérni. Azt is megállapítottuk, hogy szintén javul a becslés hatékonysága, ha nem használunk időparamétereket. Ezután visszatérve a lineáris modellekhez igyekeztünk azokat minél robusztusabbá tenni, mind a regresszió fajtáját, mind az adat kiugró értékeinek csökkentését illetően, ezzel jelentős javulást elérve a becslés pontosságában. A dolgozat zárásaként pedig ismertettük a további lépések lehetőségét.