

Kivonat

A szakdolgozat a királis molekulák enantiomertiszta előállításához kapcsolódó vegyipari folyamatok optimalizálásával foglalkozik, a folyamatot egy újszerű, integrált kristályosító-nedvesítő készülékben hajtva végre. A szakdolgozat fő feladata egy publikációhoz tartozó komplex, szintetikus előállított vegyészmérnöki adatsor idősorainak klaszterezése.

A dolgozat három fő részre oszlik. Az első rész a királis molekulák és azok jelentőségének bemutatására fókuszál, kiemelve az enantiomertiszta előállítás gyógyszeripari relevanciáját. Itt ismertetjük annak a kémiai adatbázisnak a hátterét, illetve előállításának folyamatát, amely a munka alapját biztosítja. A második rész az adatbázis kiértékeléséhez használt gépi-tanulási eszköztárat mutatja be. Többek között részletezésre kerül az alkalmazott K -medoids klaszterező modell, de a klaszterezéshez szükséges legfőbb alapfogalmak is ismertetésre kerülnek.

Az utolsó fejezet pedig a szintetikus adatok elemzésének a lépésein vezet végig. A munka során használt idősoros adatokat különböző matematikai transzformációk segítségével dolgoztuk fel, hogy megértsük az adathalmaz strukturális sajátosságait. A végső 12 klaszter pontos meghatározása érdekében kétdimenziós dinamikus idővetemítést (2D Dynamic Time Warping, 2D DTW) alkalmaztunk, amely remekül illeszkedett az idősorok időbeli eltolódásaihoz.

A kész klaszterezés segítségével végzett osztályozási és regressziós modellek is bemutatásra kerülnek. A szakdolgozatban alkalmazott Catboost osztályozási modell feladata a minták klaszterek szerinti csoportosítása volt 8 különböző kémiai és fizikai paraméter alapján. A modell segítségével a minták körülbelül 63%-át sikerült helyesen klasszifikálni, ami figyelembe véve a csoportok nagy számát és a változatos adathalmazt, ígéretes eredménynek számít. Ezzel párhuzamosan a szintén döntési fa alapú XGBoost regressziós modell az előbbi 8 jellemzőt és a klaszterezés osztályait felhasználva ad mennyiségi becslést 5 olyan paraméterre, amely az előállított termék-kristályokkal kapcsolatos. Ilyen kulcsfontosságú attribútum például az enantiomertisztaság, de vannak közöttük a termék-kristályok méreteloszlásához kapcsolódó értékek is. A modell pontosságát több metrika szerint is kiértékeljük, az R^2 -pontszámok minden célváltozó esetében 86% fölött végeztek, ami kiváló illeszkedést mutat a modell és a valós adatok között.

A kutatás során kapott eredmények rávilágítanak arra, hogy a gépi tanulás alkalmazása jelentős előrelépést jelenthet a vegyipari folyamatok optimalizálásában, a technológiák nagy áteresztőképességű, korai stádiumú szűrésében. Ezen eredmények új műszaki kutatási területeket indíthatnak, valamint hozzájárulnak a fenntartható gyártási eljárások előmozdításához.