

# AZ ELEMZÉS MŰVÉSZETE

*Az üzleti intelligencia egyik legérdekesebb szegmense az adatbányászat, amely általában kifejezetten komplex elemzési és előrejelzési problémákkal foglalkozik. Ezen a területen igen intenzív a fejlődés, mind az eszközök és módszerek, mind a megoldásra váró feladatok terén.*

## VERSENGÉS A PIAC ÉLÉN

Az adatbányászati szoftverek területén napjainkra két vezető cég emelkedett ki a versenytársak közül, akik egyértelműen dominálják ezt a szektort.

A SAS Institute 35 éves múltra tekinthet vissza, és ennek az időnek jelentős részében a cég az adatbányászatra és a fejlett analitikára koncentrált figyelmét. Bár az elmúlt években egyre nagyobb hangsúlyt kaptak a klasszikus BI és DW eszközök is, az évtizedes tapasztalatok birtokában a SAS várhatóan hosszú távon is megőrzi vezető szerepét.

A fő vetélytárs SPSS viszont 2009 óta már nem független, hanem az IBM kötelekébe tartozik. A kék óriás az üzleti analitikát célzó általános offenzívájára keretében vásárolta meg a céget, és az érte adott 1,2 milliárd dollár jól mutatja az adatbányászatban rejlő piaci potenciált.

A két vezető mellett természetesen számos kisebb cég is jelen van ebben a szegmensben. Ezek a cégek jellemzően valamilyen speciális részterületben erősek, így például a KXEN egyik specialitása a modellezési folyamat automatizálása, a Salford System pedig algoritmusainak erősségeire támaszkodik.

Végül említésre érdemes még, hogy mind az Oracle, mind a Microsoft kínál adatbáziskezelőjéhez csomagolva adatbányászati motort is. Az Oracle saját, grafikus felületű klienszoftvert is fejleszt, míg a Microsoft inkább az Excelt kínálja a felhasználók számára.

## NYÍLT FORRÁSKÓDÚ ALTERNATÍVÁK

Az elmúlt években az adatbányászat területén is megjelentek a klasszikus szállítók mellett a nyílt forráskódú üzleti modellel dolgozó vetélytársak.

Az egyik ilyen új szereplő a német Rapid-I, amelynek fő terméke a RapidMiner általános adatbányászati eszköz. A nagy vetélytársakhoz tudásban és megjelenésben egyre inkább hasonlító RapidMiner mellett a cég jelen van a specializált analitikus alkalmazások piacán is, így például forgalmaz megoldást a webes tartalmak elemzésére is.

A nyílt forráskódú szállítók között említésre érdemes még az új-zélandi egyetemi háttérrel rendelkező Weka, amely a Pentaho BI platformjának is része, illetve a különösen a gyógyszeriparban és az élettudományok területén használatos KNIME.

Markáns szereplője még a nyílt forráskód világának az R programozási nyelv, amely az alapvetően grafikus felhasználói felületet nyújtó szoftverekkel szemben procedurális környezetben teszi elérhetővé az analitikus algoritmusokat. A számos különböző programnyelvtől elemeket kölcsönző R sikerét az mutatja legjobban, hogy napjainkra szinte az összes nagyobb cég nyújt valamilyen szintű támogatást az R nyelvű elemzések futtatásához. Az R fejlesztésére és támogatására több üzleti vállalkozás is létrejött már, ezek közül a legismertebb a 2007-ben alapított amerikai Revolution Analytics.

## ÚJ PLATFORMOK

Az informatikai platformok terén zajló átalakulás az adatbányászatot sem hagyta érintetlenül. A szállítók két, egymásnak kicsit ellentmondó irányt követve próbálják a felhasználók igényeit – nagyobb teljesítmény, könnyebb üzemeltetés, alacsonyabb költségek – kielégíteni.

Az egyik irányzat szerint a hagyományosan külön, önálló szoftverként csomagolt analitikus algoritmusokat oda kell vinni, ahol az adatok vannak: az adatbázisok belsejébe. Az ezen a területen vezetők közül mind a SAS, mind az SPSS igen jelentős haladást ért már el. A SAS elsőként a Teradatával kötött stratégiai szövetséget a masszívan párhuzamos adatbáziskezelő belsejében történő modellfuttatásra, majd később más DW platformok felé is nyitott (Netezza, Aster Data). Hasonló taktikát követett az SPSS is, amely többek között az Oracle és az IBM adatbázisaiban képes algoritmusokat futtatni.

A nyílt forráskódú szoftverek természetesen ezen a területen is megtalálhatóak. Itt említésre érdemes a Berkeley egyetem és a Greenplum