Gáspár Bencéné Vér Katalin*

ADATBÁNYÁSZAT A GAZDASÁGI ÉLETBEN

Az *adatbányászat* egy *döntéstámogatási módszer*, olyan üzleti intelligencia megoldás, amely új üzleti lehetõségeket segít megtalálni és kiaknázni a *nagytömegû adathalmazokban rejlõ, nem ismert összefüggések feltárásával*. Az angol *Data mining* kifejezés tükörfordítása. Egyesíti az adatbázis-kezelés, a statisztika és a mesterséges intelligencia kutatások eredményeit. Az adatbányászat kifejezést a különbözõ informatikai cégek különbözõ értelemben használják, de a szigorúbb szakmai terminológia nem tekinti adatbányászatnak az adatokból egyszerû lekérdezésekkel, aggregálásokkal, illetve alap-statisztikai vizsgálatokkal történõ információ-nyerést (http://hu.wikipedia.org/wiki/Adatbányászat).

Az elmúlt évek során az adatbányászat iránti érdeklõdés rohamosan nõtt mind a korszerû informatikai szolgáltatásokat nyújtó cégek, mind az ilyen szolgáltatásokat igénybe vevõ vállalatok körében. Az adatbányászati funkció hatékony infrastruktúrát biztosít az egész vállalatra kiterjedõ, magas szintû üzleti adatelemzõ alkalmazások létrehozásához. Az új üzleti ismeretek kinyerésének és terjesztésének automatizálásával és integrálásával a vállalatok ki tudják használni az adataikban fekvõ információkat, hatékonyabban mûködhetnek, és nagyobb versenyelõnyhöz juthatnak.

Az adatbányászat mint az adatelemzés eszköze és lehetõsége – a statisztika hasonló kategóriáinak megfelelõen – két nagyobb kategóriába sorolható:

- a *leíró adatbányászat* az adatok alap jellemzõinek meghatározását jelenti.
- a *következtetéses adatbányászat* alapvetõen összefüggések feltárásával foglalkozik.

A statisztikai eszközökkel „kis és közepes" adatmennyiségek esetén meg lehet találni bizonyos szabályszerûségeket és korrelációkat, de ezek az eszközök igazán nagy mennyiségû adattal már nem képesek megbirkózni. Az adatbányászatnak nincsenek ilyen korlátai. Az adatbányászat az adatok mélyére hatol. (Gáspár, 2006) Az adatbányászat alapadatai egyaránt lehetnek üzleti, kutatási, mérési adatok. Lényegében bármilyen nagytömegû adathalmaz elemei képezhetik az adatbányászat alapadat-állományait. Az adatok két köréhez kötõdik az adatbányászat speciális területe: a szövegbányászat (textmining) és a webbányászat (webmining).

„A *szövegbányászat* a tudásmenedzsment egyik leghatékonyabb információtechnológiai eszköze. Tudásalapú technológia, amely képes a mesterséges intelligencia, a gépi tanulás, a természetes nyelvi feldolgozás, a nyelvtechnológia, a többváltozós matematikai statisztika, a valószínûségszámítás, a tartalomelmélet és még jónéhány tudományág legfrissebb eredményeinek produktív hasznosítására. A szövegbányászat segítségével olyan rejtett ismeretanyag nyerhetõ ki strukturálatlan szöveges dokumentum állományokból, amelyek kinyerése egyrészt emberi erõvel kivitelezhetetlen lenne, másrészt pedig olyan tudásvagyont képvisel, amelybõl rövid úton üzleti versenyelõny kovácsolható." (Vázsonyi, 2006)

A *web-bányászat* olyan gépesített dokumentum-feldolgozó szakterület, amely az internethez kapcsolódóan a weboldalakon található képi, szöveges és egyéb alakú adatok feldolgozhatóvá való átalakításával foglalkozik. Az adatok átalakításának a célja többnyire megfelelõ input-állományok kialakítása további, a felhasználó szempontjából értékes adatok kinyerése érdekében. Ezt a munkát speciális alakfelismerõ programok végzik. A megrendelõk többnyire a kereskedelmi és hírszerzés területérõl való cégek. (http://hu.wikipedia.org/wiki/Adatbányászat) Az adatbányászat kialakulásának indítéka a napjainkban szinte az élet valamennyi területén, de különösen a gazdasági szférában, a vállalatoknál, a napi tevékenység során keletkezõ, *nagytömegû adat*. Ez különösen igaz a *nagyvállalatok*ra és a *pénzügyi szektorra*, ahol *minden tevékenységi lépés adatbázison alapul*.

* *Fõiskolai tanár, Általános Vállalkozási Fõiskola*

***

Ahhoz, hogy olyan adathalmazunk/állományunk legyen, amellyel az adatbányászati munka megkezdõdhet egy adatelemzési, logikai folyamatnak kell végbemennie. Ennek lépései:

- adatbázis audit – adatforrások és tárhelyek, kapcsolódásaik feltérképezése,
- adattisztítás,
- adatintegrálás,
- adatkiválasztás,
- adattranszformálás,
- egységes adatmodell kidolgozása.

Megfelelõ adatállomány(ok) – ez sok esetben egy adattárházat jelent – birtokában kezdõdhet maga az *adatbányászat*, amely ugyancsak több lépésbõl áll. Elsõ lépés az adott üzleti probléma megértése. Ezt az adatok megismerése, és modellezés céljára történõ elõkészítése követi. Ezután kerül sor a statisztikai alapelemzések elvégzésére, majd a részletes adatelemzésre, amely magában foglalja az elemzési módszerek, szoftverek kiválasztását, és az elemzés elvégzését. A következõ lépés a modellezés, az adott elemzés céljából érdekes összefüggések feltárása, valamint a kapott eredmények értékelése, majd a kapott eredmények ismertetése, bemutatása. Az eredmények felhasználása ezután kezdõdhet el.

Mire keresi az adatbányászat a választ? A teljesség igénye nélkül csak néhány példa:

- személyek/szervezetek legvalószínûbb reakciói,
- együtt értékesített/vásárolt termékek/szolgáltatások,
- kitõl várható hogy a konkurenciához pártol,
- csalások felderítése,
- azonos terméket vásárlók közös tulajdonságai,
- mi érdekli a látogatót a legjobban egy Web oldalon.

Az adatbányászat jellemzõi

- Új távlatok (ahol a hagyományos módszerek nem elég jók, vagy nem elég gyorsak az igazán hatékony adatelemzéshez).
- Valós múltbeli összefüggések alapján segít a jövõ döntéseinek megalapozásában.
- Általános módszer, amely minden üzleti területen hatékonyan alkalmazható.
- Hatékony eszköze az elektronikus kereskedelem bevezetésének és fejlesztésének.

Az adatbányászat módszertana a

- társításon,
- csoportosításon,
- osztályozáson,
- visszafelé haladáson (neurális hálók, regresszió, kapcsolatok felderítése),
- összegzésen alapul.

***

Vizsgáljuk meg, hogy *milyen információkat bányásznak* a különbözõ módszerek segítségével?

A *társítás (association)* egy eseményt, vagy tárgyat, például árucikket rendel egy másikhoz. Például: az asszociáció a marketingben megtalálja (felismeri), hogy mely termékeket vásárolnak együtt.

A *csoportosítás (clustering)* az adatok korábban nem ismert rendszereinek, együvé-tartozásának a felismerése, az adatokat leíró véges kategóriasorok azonosítása. Azonos tulajdonságok alapján csoportosítja a vizsgált adatokat. *Ezt az információt használják fel csoport-képzésre további elemzésekhez.*

Az *osztályozás (classification)* az adatok újfajta szervezõdését eredményezõ minták észrevétele, amely alapján egy adat egy vagy több elõre meghatározott osztályba rendezhetõ. (Például: drága sportkocsit vásárlók tipikusan fiatal, városi diplomások, magas jövedelemmel.)

A *visszafelé haladás,* amely neurális hálókat is alkalmaz, az adatokat valós-értékû elõrejelzés változóhoz (*real-valued prediction variable*) társító funkció, illetve a változók közötti kapcsolatok felderítésére (*dependency modelling*) használatos módszer. Ilyen a hasonló idõsorok keresése. Felfedi a hasonló sorozatokat egy adott idõszakban, vagy felfedi a hasonló sorozat-párokat. Például: felfedi a hasonló árfolyam-mozgású részvényeket.

A *kivétel-keresés*, ahogy az elnevezése is mutatja, a „szokatlant” keresi. Például: felfedi a szokatlan hitelkártya tranzakciót, és ezáltal detektálja csalásokat.

Az e*lõrejelzés (trendek)* az adatminták alapján becsli a jövõbeli értékeket. A gazdasági menedzserek számára legfontosabb információ az elõrejelzés, de gyakran kombinálják az elõzõekben ismertetett információkat is.

Az *összegzés* (*summarization*) az adatok egy meghatározott részhalmazának tömör leírására vonatkozik.

***

Az adatbányászat technikái a
- „felügyelt tanulás”
- „felügyelet nélküli tanulás”

Felügyelt tanulás

A *felügyelt tanulásnál* az adatelemzõnek ki kell jelölnie a célmezõt vagy a függõ változót. A felügyelt tanulási technikával ezután átvizsgálja az adatokat, hogy *szabályszerûségeket* és *összefüggéseket fedezzen fel* a független változók és a függõ változók között. *Felügyelt tanulási modellnél* az adatbányászati algoritmus aprólékosan átvizsgálja az adatokat, rejtett szabályszerûségeket tár fel, olyan *modellt* alkot, amely a lehetõ legjobban leírja a függõségeket.

Az adatokat általában három részre osztják:
- betanítási – kezdeti modell,
- tesztelési – modellfinomítás,
- kiértékelési adatok, elõrejelzések készítése.

Felügyelet nélküli tanulás

A *felügyelet nélküli tanulásnál* a felhasználó nem adja meg a célt az adatbányászati algoritmus számára. Az *adattársítási* és *csoportosító algoritmusok* ilyenkor *semmit* sem *feltételeznek a célmezõrõl*. Ilyen esetben az adatbányászati algoritmusok minden elõre definiált üzleti cél nélkül keresnek kapcsolatokat és csoportokat.

***

Az adatbányászat talán leggyakoribb gyakorlati felhasználási területei
- a direktmarketing,
- a kockázatelemzés,
- a keresztértékesítés,
- a visszaélések felderítése,
- az ügyfélmegtartás,
- az internetes oldalak vizsgálata.

Üzleti döntések adatbányászat segítségével

A cégek négy lehetõség közül választhatnak, amennyiben üzleti döntéseiket adatbányászat segítségével kívánják megtámogatni:

1. Megvásárolhatják az adott üzleti problémára létrehozott adatbányászati modell eredményeit, amely szabályok, ügyfélllisták, pontszámok formáját öltheti. Ilyenkor nincs saját modell, lényegében információt vásárol a cég.
2. Megvásárolhatnak egy olyan, beágyazott adatbányászati alkalmazást tartalmazó szoftvert, amellyel kifejezetten a saját üzleti problémájukra vonatkozó modellezéseket lehet végezni, többek között csalásfelderítésre, lojalitásnövelésre vagy kampánymenedzselésre vonatkozóan.
3. Külsõ adatbányászati szakértõket bízhatnak meg elõre definiált feladatokra egy-egy projekt keretei között.
4. A cégen belül saját adatbányászati csoportot hozhatnak létre.

Mind a négy verziónak vannak pozitívumai és negatívumai. Mindig az adott körülményektõl függ, hogy melyiket, vagy melyek kombinációit célszerû egy cégnek választania. A különbözõ felhasználók különbözõ területeken várnak választ és hasznot az adatbányászat alkalmazásából:

Kiskereskedõk (retail szektor)
– A vásárlói szokások és preferenciák megismerése piaci vagy vásárlói kosárelemzéssel.
– A tisztességtelen fizetési magatartás felderítése.

Direktmarketing / Telemarketing
– Nagy megtakarítások pontosan megcélzott ügyfelek révén.

Gyártók
– A termelési folyamat ellenõrzése és ütemezése.

Légitársaságok
– Felfedezni az ügyfelek igényeit a stratégiai változtatásokhoz (pl. új útvonalak felkínálása vagy a szolgáltatások bõvítése, stb.)

Telekommunikációs társaságok
– Annak „kitalálása", milyen szolgáltatások lennének népszerûek az ügyfelek között
– A hívási csalások felfedezése szokatlan minták felderítésével.

***

A pénzügyi felhasználás területén kiemelhetõk:

Bankok
- Célzott marketing
- Kölcsön visszafizetés teljesítés elõrejelzése

A nagy bankok (pl. Bank of American, Dresdner Bank) általában saját adatbányász munkatársakkal rendelkeznek. A kisebb bankok, amelyek korlátozott erõforrásokkal és technológiával rendelkeznek, többnyire megrendelik az adatbányászatot, illetve az adattárház-létrehozást illetve mûködtetést.

Biztosítótársaságok
- Az ügyfelek jobb megismerése.
- A biztosítási csalások hatékonyabb felderítése.

Pénzügyi szolgáltatások
Értékpapír-elemzõk intenzíven használják nagytömegû pénzügyi adat elemzésére kereskedési és kockázati modellek létrehozása céljából, befektetési stratégiák kialakításához az adatbányászatot. Ez a felhasználási terület nagyon széles, és olyan alkalmazásokat foglal magába, mint:
- a devizakereskedelem,
- a részvény-kiválasztás,
- a jelzálog kiválasztás.

Hitelkártya kibocsátó társaságok (pl.: American Express, Citibank)
- Hitelkártya igénylések jóváhagyása.
- Vásárlásokat jóváhagyó döntések.
- A kártyatulajdonosok vásárlási szokásainak elemzése.
- Csalásfelderítés.

Pénzügyi felügyeletek és szabályozó hatóságok (pl: NASD, Internal Revenue Service, FBI)
- Pénzmosás felderítése.
- Bennfentes kereskedésre utaló minták felfedezése.
- A piac szabályainak megsértése.

Az adatbányászat eredményeinek *vizuális* megjelenítése elsõsorban a real-time (valósidejû) pénzügyi alkalmazásokban, például a pénzpiaci, tõzsdei adatbányászatnál különösen fontos lehet. Ennek oka a

nagy mennyiségû információ
- Fundamentális adatok, jövedelmek, becslések, vetítések, osztalékok, könyvszerinti érték stb.
- Technikai jellemzõk, pl. részvények teljesítménye, kvantitatív elemzések az ázsiai derivatív piacokra.
- Különbözõ elemzõk és brókerek véleménye stb.

a döntési elemek nagy száma, amelyeket az egyes értékpapírok esetében figyelembe kell venni, ami azt eredményezi, hogy nehéz az *anomáliák azonosítása és kiküszöbölése* az egyes piacokon.

A különbözõ információk egyszerû integrálása nem jelent segítséget a nagytömegû adat miatt. Az adatok tradicionális ábrázolása (táblázatkezelõ oldalak ipari csoportok és országok szerint, oszlop-, körvagy akár 3D-s diagramok) nem tudják az adatokat megfelelõen integrálni és prezentálni. A megoldás a *valósidejû vizuális adatbányászat:* különbözõ színek, formák, méretek, prezentációs stílus (pl. villogás és forgás) segítségével hajtják végre az adatelemek vizuális megjelenítését oly módon, hogy ezeket a megjelenítési módokat rendelik az egyedi adatelemekhez azok fontos numerikus értékeinek vagy nemnumerikus tulajdonságainak leírásához. (Például egy nyíl megjelenítése jelölheti minden egyedi adatelemhez a jelenlegi feltételek egy halmazát.) Ilyen vizuális adatbányászati megoldás például a NASD Advanced Detection System (ADS), amelynek célja a piacok integritásának védelme a belterjes kereskedelemtõl, és más szabályok megsértésétõl. Az eszköz, amellyel ezt az adatbányászatot és a vizualizációt végzik a Metaphor Mixer (MM) a Maxus Systems International Inc. terméke. (Metaphor Mixer)

***

Összefoglalásként elmondhatjuk, hogy **a**z adatbányászati funkció hatékony infrastruktúrát biztosít az egész vállalatra kiterjedõ, magas szintû üzleti adatelemzõ alkalmazások létrehozásához. Az új üzleti ismeretek kinyerésének és terjesztésének automatizálásával és integrálásával a vállalatok ki tudják használni nagytömegû adataikban fekvõ információkat, hatékonyabban mûködhetnek, és nagyobb versenyelõnyhöz juthatnak (Gáspár, 2006).

IRODALOM

Gáspár Bencéné dr. Vér Katalin (2006): *Üzleti intelligencia rendszerek.* Budapest. ÁVF.

Metaphor Mixer (MM) by Maxus Systems International Inc.: http://www.maxussystems.com/

Vázsonyi Miklós (2006): http://www.szovegbanyaszat.hu/Szövegbányászat

http://hu.wikipedia.org/wiki/Adatbányászat

http://www.datamining.hu/

http://www.eflow.hu/Tudasmenedzsment/Intelligens_portalok/Adatbanyaszat/adatbanyaszat.html