

NEHÉZ MUNKA A KINCS FELSZÍNRE HOZÁSA

Az adatmenedzsment hét sötét titka

Amikor a szervezet elkezd komolyan venni a működéséből származó információkat, akkor válik elérendő céllá az adatvezérelt vállalat. Viszont az adatmenedzsment-projektek során nagyon hamar nagyon komoly akadályokba ütközik a cég, és ezekről a marketinganyagok nem beszélnek. Összeállításunkban a Cio.com magazin cikke alapján beszéljük ki az adatmenedzsment sötét titkait.



Az elcsépeelt metaforáknak se szeri se száma, amikor az adatról van szó: az adat az új olaj, vagy az adat új arany. A művészi kifejezések minőségét és eredetiségét figyelmen kívül hagyva abban egyetérthetünk, hogy az adatok és azok elemzése fontos versenyelőnyt hoz a vállalatoknak. Ebbe előkészítés nélkül fejest ugrani bátor vállalkozás.

A jó hír, hogy az adatmenedzsment rengeteg területe kidolgozott, megértett és évtizedek óta fejlődő elvek alapján szerveződik. A matematika és a statisztika kutatóinak és tudósainak köszönhetően egy sor logisztikai keretrendszer áll a vállalatok rendelkezésére az adatok elemzésére és a következtetések levonására. Ami talán még fontosabb, statisztikai modellek segítségével az elemzések határait is pontosan definiálhatjuk.

A tudományos eredmények és kutatások ellenére sokan csak a fejüket vakargatják, amikor adatokkal kell dolgozni, a vállalatok sok akadályba is ütköznek. A gyakorlati kihívások közül elég megemlíteni az adatok gyűjtését és rendszerezését, minőségi szintjének

fenntartását. A különböző szabályozások és elvárások is megnehezítik az adatok jogszerű gyűjtését, tárolását és felhasználását. Rengeteg olyan sötét titok van az adatok menedzselésével kapcsolatban, amelyekre csak a mindennapi munka során derül fény.

1. Nem strukturált adatot nehéz elemezni

A vállalatnál fellelhető adatok többsége semmilyen struktúrával nem rendelkezik, ezért rendkívül nehéz bármilyen információt kinyerni belőlük. Egy banki call centernél MI segítségével próbálták elemezni a kollégák a hívásokkal kapcsolatos szöveges feljegyzéseit. Azt remélték, hogy az adatokból kinyerhető információkkal sikerül a hitelfelvételi folyamatot javítani. A jegyzeteket közel száz különböző ember készítette, akiknek különböző ötleteik támadtak az adott hívással, ügyfél-interakcióval kapcsolatban. Sokan rengeteg információt jegyeztek le, mások alig valamennyit. A természetes nyelvi szövegek struktúrájával számos tudományterület régóta foglalkozik, de amikor a közel száz ember telefonhívásokkal kapcsolatos jegyzeteiben kellene a struktúrát felfedezni, akkor még az MI-nek sincs könnyű feladata.

2. A strukturált adat gyakran valójában nem strukturált

Az adatbázis-adminisztrátorok tisztességes munkát végeznek, amikor az adatmezők strukturáját és típusát meghatározzák. Gyakran korlátozzák, hogy az adott mezőbe milyen értékeket lehet bevinni, sok esetben listából kell kiválasztani az értékeket. De még így is, az adatokat feltöltő emberek megtalálják a módját annak, hogy hibás, rossz adatokat vigyenek fel. Néha üresen hagynak mezőket. Ha egy kérdésről nincs elegendő információ, az „n. a.” megjegyzéssel letudják a munkát. Az is előfordul, hogy ugyanaz a személy különböző helyeken másképp írja nevét. A jó fejlesztők validációval sok ilyen problémát megfogják. A jó adattudósok pedig tisztítással csökkentik az adatokkal kapcsolatos bizonytalanságokat. De frusztráló, hogy még a legjobban kitalált strukturába is kerülnek megkérdőjelezhető adatok, amelyek bizonytalanságot, rossz esetben hibát visznek a következtetésekbe.

3. Túl szigorúak az adatvédelmi törvények

Az adatvédelmi törvények elég komolyak, és egyre több országban érvényesek. A GDPR és a HIIPA között nem is olyan egyszerű adatokat gyűjteni. Veszélyes csak úgy, adatokat tárolni valahol a cégen belül arra várva, hogy egy kiberbűnöző betörjön és elloplja azokat. Sok esetben egyszerűbb adatvédelmi ügyvédekkel megfizetni, mint programozókat vagy adattudósokat. Ezen problémák miatt sok vállalat nem vállalja fel az adatokkal kapcsolatos kockázatokat, nem hasznosítja azokat a mindennapokban.

4. Rendkívül magasak lehetnek az adattisztítás költségei

A legtöbb adatszakterember azt látja, hogy a munka 90 százalékát teszi ki az adatok összegyűjtése, konzisztens tárolása és a hibák menedzselése. Az adattal rendelkező személy mindig azt mondja, az adatok ott állnak készen, CSV-ben, csak elemezni kell. Az üres mezőkről vagy félregépelésről egy szó sem esik. Az adatok tisztításával könnyen eltelhet százszor hosszabb idő, mint R-ben vagy Pythonban kidolgozni és lefuttatni egy lekérdezést.

5. Van, amikor egyszerűen nem éri meg

Az adatvezérelt vállalat célja a működést, a folyamatokat hatékonyabbá és profitábilissá tenni. Az algoritmus képes a feladatokat ellátni, de gyakran semmi értelme a válasznak. (Lásd Douglas Adams: „Galaxis útikalauz stopposoknak”...) Manapság a fogyasztók és a vállalatok is megkérdőjelezzik a célzott marketing erejét. Gyakran látunk újabb és újabb hirdetéseket olyan termékekre, melyeket már rég



megvásároltunk, csak erre a hirdetést kiszolgálók nem jöttek rá. Hasonló sors várhat más megoldásokra is. Egy alapos adatelemzés kiszúrhatja a legrosszabbul teljesítő gyárat, de semmi értelme, ha a vállalat 30 éves bérleti szerződést írt alá.

6. Robbanásszerűen növekednek az adattárolás költségei

Igen, a kapacitás folyamatosan nő, a TB/\$ (fajlagos tárolási költség) folyamatosan csökken, de a generált adatok mennyisége ennél sokkal gyorsabban nő. (Lásd az infografikát!) A szervezetek szinte válogatás nélkül gyűjtik és tárolják az adatokat, hátha szükség lesz valamire alapon. Érdemes lenne átgondolni és szortírozni őket, de mindez időt és felelősséget jelent. Nem meglepő, hogy egyre csökken azon adatok aránya, amelyeket tárolás után nem használnak fel.

7. Sok esetben az adatvezérelt döntések valójában megérzések

A számok valóban precízek és sohasem hazudnak, de nem az algoritmus dönt, hanem az ember. Például hiába 130 km/h a sebességhatár az autópályán, van olyan helyzet, amikor egy baleset elkerülése érdekében ezt az értéket meg kell haladni. A bírások is csak bizonyos, nem matematikai alapon megállapított százalékos túllépés után születnek meg. A matematika, a statisztika és a rengeteg tudományos eredmény ellenére sok adatvezéreltnek hitt folyamatnál szükség van az ember megérzésére és döntésére is.

Vass Enikő