

ADATBÁNYÁSZATI MÓDSZEREK OKTATÁSA A SAS[®] RENDSZERREL

Kiss Ferenc, hunfek@mvs.sas.com
SAS Institute

Abstract

This paper briefly summarises the core of Data Mining enlightening some application fields and techniques of it. A short list describes scientific and business fields which are regular application areas of the SAS System. These subjects are thought at a lot of well known universities worldwide using SAS as demonstration and/or problem solution tool. Figures demonstrates some analytical capabilities of SAS System.

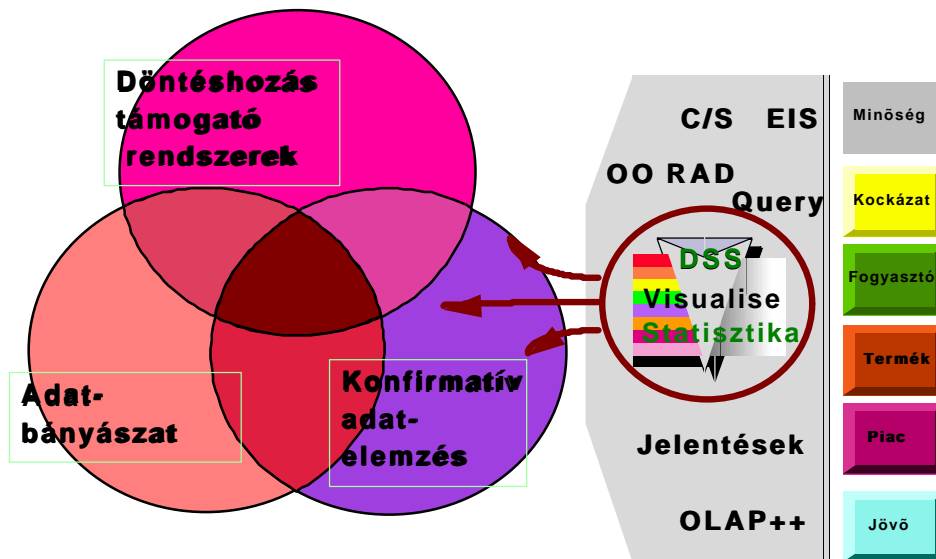
1. Mi az adatbányászat?

„A számítógépek a bölcsesség forrását ígérték nekünk, azonban csak az adatok áradatát szállítják” — írta 1991-ben megjelent könyvében Gregory Piatetsky-Shapiro. E megállapításával arra az alapvető tényre világít rá, hogy a kutatási, termelési, üzleti folyamatokból származó adatokat információvá kell alakítani ahhoz, hogy megfelelő következtetéseket vonhassunk le, és ezek alapján döntéseket hozhassunk. Az adatbányászat lényegét tekintve nem más, mint a nyers adatok halmaiból az össze függések, a kulcstényezők kiemelése.

Az alábbi néhány tipikus példa jól mutatja az adatbányászat alkalmazási területeinek sokszínűségét:

- célzott marketing: ügyfél célcsoportok meghatározása;
- ügyfél elégedettség-vizsgálat: kérdőívekre és demográfiai adatokra alapozva;
- értékesítési előrejelzés - korábbi forgalmi adatokra és befolyásoló tényezőkre alapozva;
- üzemi és ügyviteli folyamatok optimalizálása;
- hitelkérelem-elbírálás;
- visszaélés felderítés: biztosítási csalások, hitelkártya visszaélések;
- befektetési portfólió elemzés, árfolyam-előrejelzés;
- beruházás menedzsment;
- raktárkészlet tervezés és optimalizálás;
- beszállító és ügyfélminősítés;
- számítógépek és hálózatok terhelésének előrejelzése, feladatütemezés.

Az adatbányászati módszerek csak részben új eljárások: például minta keresés adatokban, felhívó jellegű elemzések, előrejelzések. Támogatják vagy automatizálják a mintakeresési folyamatokat és főleg nagy adathalmazokra használhatók, ezért nagy adathalmazok kezelésére szolgáló olyan informatikai technológiák jelenléte elengedhetetlenül szükséges, mint például az adatraktár koncepció megvalósítása.



1. ábra

Az adatbányászat az adatraktár kiaknázásának egyik lehetősége

A Gartner Group szerint az adatbányászat az alábbi elemző eszköztárak felhasználását jelenti:

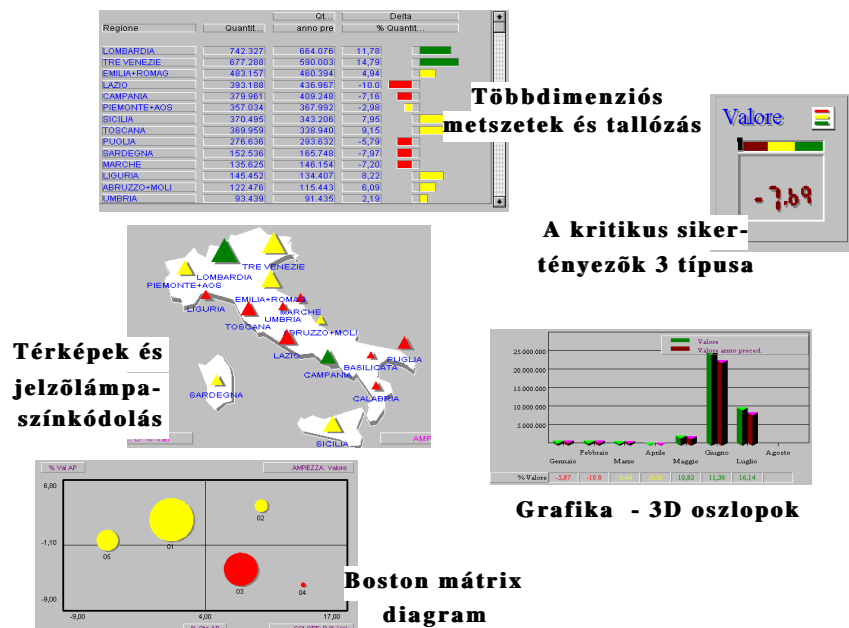
- vizualizáció;
- magasszintű „statisztikai” módszerek;
- következtető rendszerek;
- neurális hálózatok.

Vizualizáció alatt a két és háromdimenziós grafikonokat, tudományos és üzleti diagramokat, speciális ábrázolásokat, térképeket, térinformatikai eszközöket valamint multimédia használatát értik. A 2. ábrán néhány, a vezetői információs rendszerekben szokásos ábrázolás látható.

A „statisztikai” eljárások között nem csak a szorosan vett matematikai statisztikai módszerek találhatók. Ide tartoznak a statisztikai próbák, klaszter-, faktor-, diszkriminancia analízis, több dimenziós skálázás, lineáris és nemlineáris modellek, kontingenciatáblák, conjoint elemzés, preferencia térképek mellett például az idősorok elemzése, lineáris és nemlineáris regresszió-analízis, lineáris és nemlineáris programozás, az ökonometriai modellek, és szimulációs és egyéb speciális eljárások.

Szintén a Gartner Group elemzésében olvasható — és ez az oktatás szempontjából egyáltalán nem mellékes —, hogy milyen szempontokat érdemes figyelembe venni egy adatbányászati szoftver értékelésénél. A következő kérdéseket kell feltenni:

- mekkora adatmerekkel képes a rendszer dolgozni?
- képes-e a rendszer a szükséges előfeldolgozások elvégzésére, vagy ehhez más eszköz kell?
- "bottom-up" és/vagy "top-down" megközelítést kínál?
- milyen könnyen frissíthetők a modellek?
- használata milyen tudást igényel?
- van-e lehetőség felhasználói beavatkozásra az elemzés során?



2. ábra

Néhány SAS üzleti grafikai ábrázolásmód

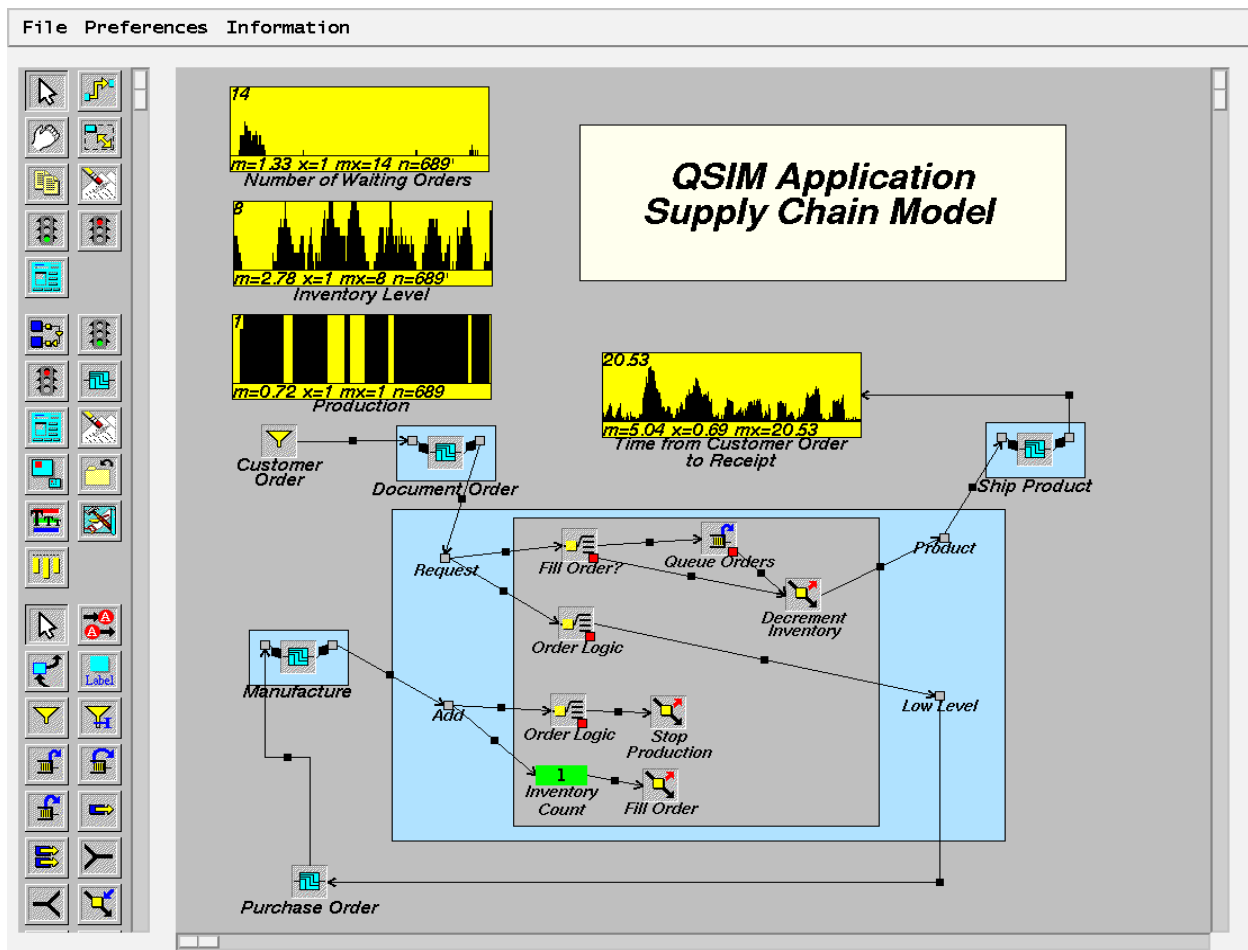
A SAS rendszerben az adatbányászati eszközök teljes skálája megtalálható úgy, hogy mind egyikük nagy adathalmazok feldolgozására van optimalizálva.

2. A SAS rendszerrel oktatható tárgyak és témakörök

A SAS rendszert széles körben alkalmazzák oktatási és kutatási célokra a világ mintegy 100 országának egyetemén. Ez részben annak köszönhető, hogy a SAS Institute évente bruttó bevételének mintegy 35%-át kutatás-fejlesztésre fordítja, ezáltal a legújabb tudományos és technológiai eredmények szinte késedelem nélkül a felhasználók rendelkezésére állnak. Emellett a SAS Institute közreműködésével számos elismert szakkönyv és egyetemi tankönyv jelent meg egyetemi oktatók tollából.

A dokumentáció elkészítésekor nagy gondot fordítottunk arra, hogy minden elemző eljárásnál képlettel és hivatkozással egyértelműen azonosítható legyen a módszer, és jellemző alkalmazási példákkal mutassuk be azok használatát.

A SAS rendszer fejlesztő eszközeivel oktató és demonstrációs programok is készíthetők. A SAS rendszer felhasználóinak járó támogatás többek között tartalmaz egy olyan ftp hozzáférést, ahonnan a dokumentáció példáin túl további példaprogramok, esettanulmányok tölthetők le.



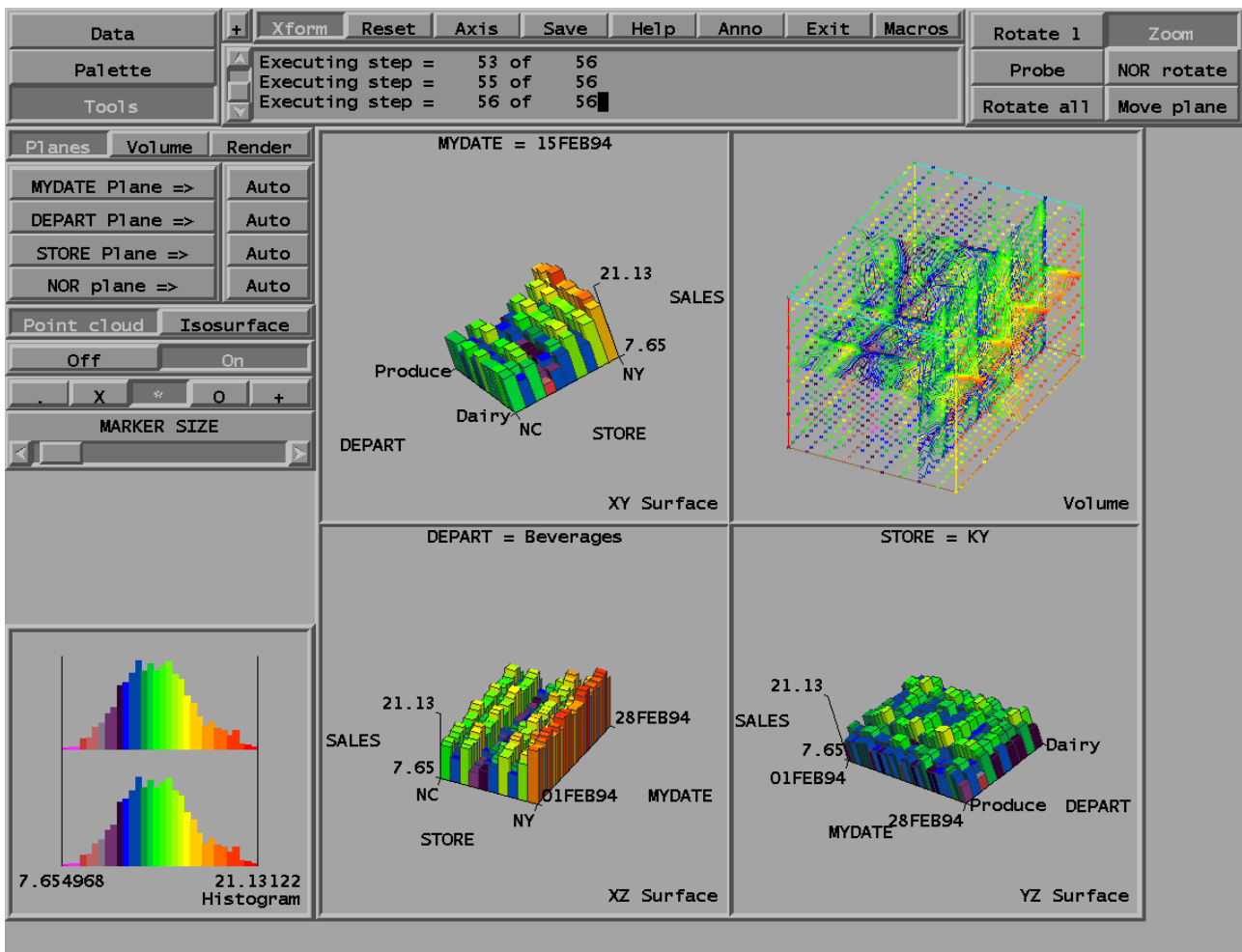
3. ábra

Egy raktári kiszolgáló rendszer vizsgálata a QSIM diszkrét szimulációs eszközzel

A SAS rendszerben található eszközök és eljárások legnagyobb hányada általános célú, paramé terezhető, ezáltal a különböző szakterületekről vett, de azonos logikával működő, illetve azonos matematikai háttérű problémák vizsgálhatók velük.

A 3. ábrán egy gyári késztermékraktár logisztikai láncának szimulációs vizsgálatára látható példa. Elemezhető, hogy a lánc melyik pontján jöhetnek létre leginkább késések, hol van gyakran szabad kapacitás, milyen átszervezésekkel lehetne optimálisabbá tenni a folyamatot, stb. Ugyanezzel az eszközzel vizsgálhatók és optimalizálhatók olyan sorbanállási és tömeg kiszolgálási modellek és folyamatok is, mint például:

- a számítógép hálózatok szervei ki tudják-e szolgálni a felhasználókat adott topológia, teljesít mény-paraméterek mellett?
- termelési folyamatban milyen ütemezésű logisztikai lánc mellett lesz biztonságos a gyártósor kiszolgálása?
- egy ismert ügyiratkezelési folyamatban hol torlódnak az akták, hol vannak szabad kapacitások, hol vannak ismétlődő, párhuzamos folyamatok?
- ismert erőforrás, probléma eloszlás, populáció és ellátási körzet adatok mellett melyik kórházat (hivatal, iskolát, üzletet, szolgáltatót) kell bezárni, és melyiket bővíteni?



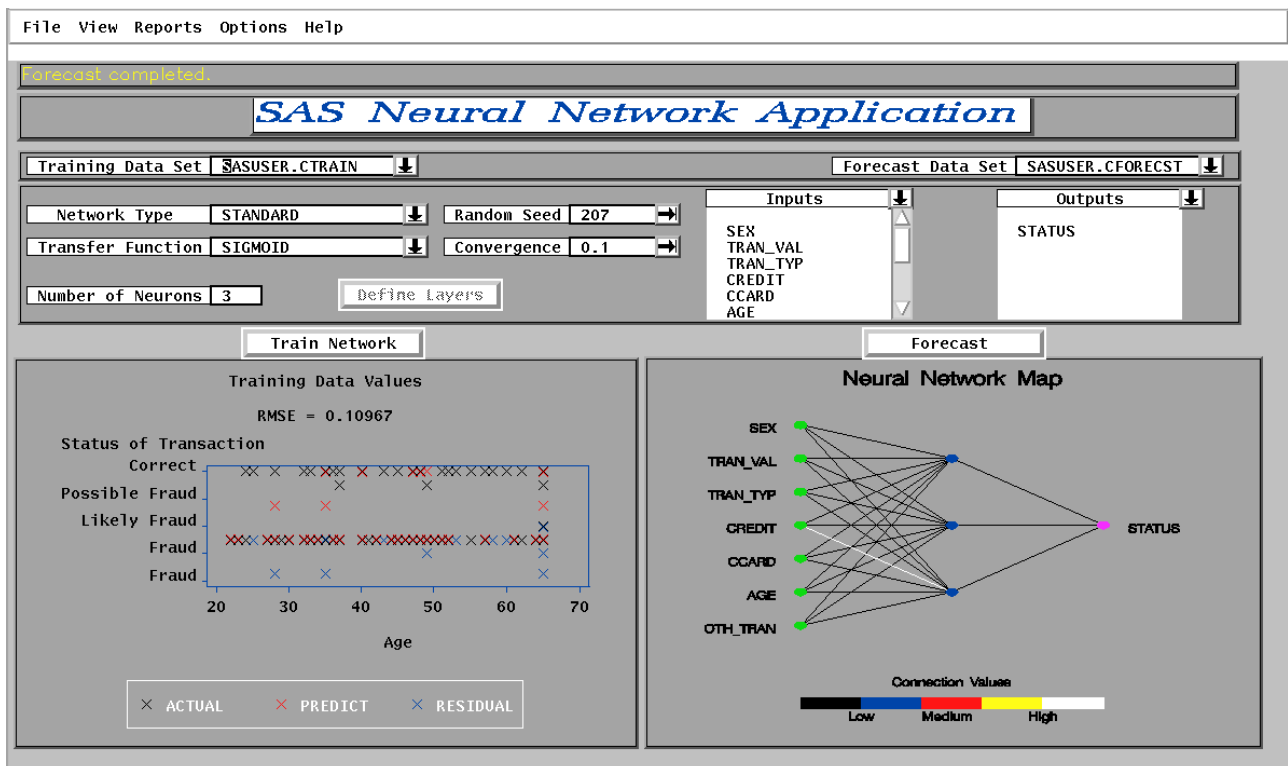
4. ábra

Pontfelhők háromdimenziós ábrázolása vizuális feltáró elemzéshez a SAS/Spectraview[®] rendszerrel

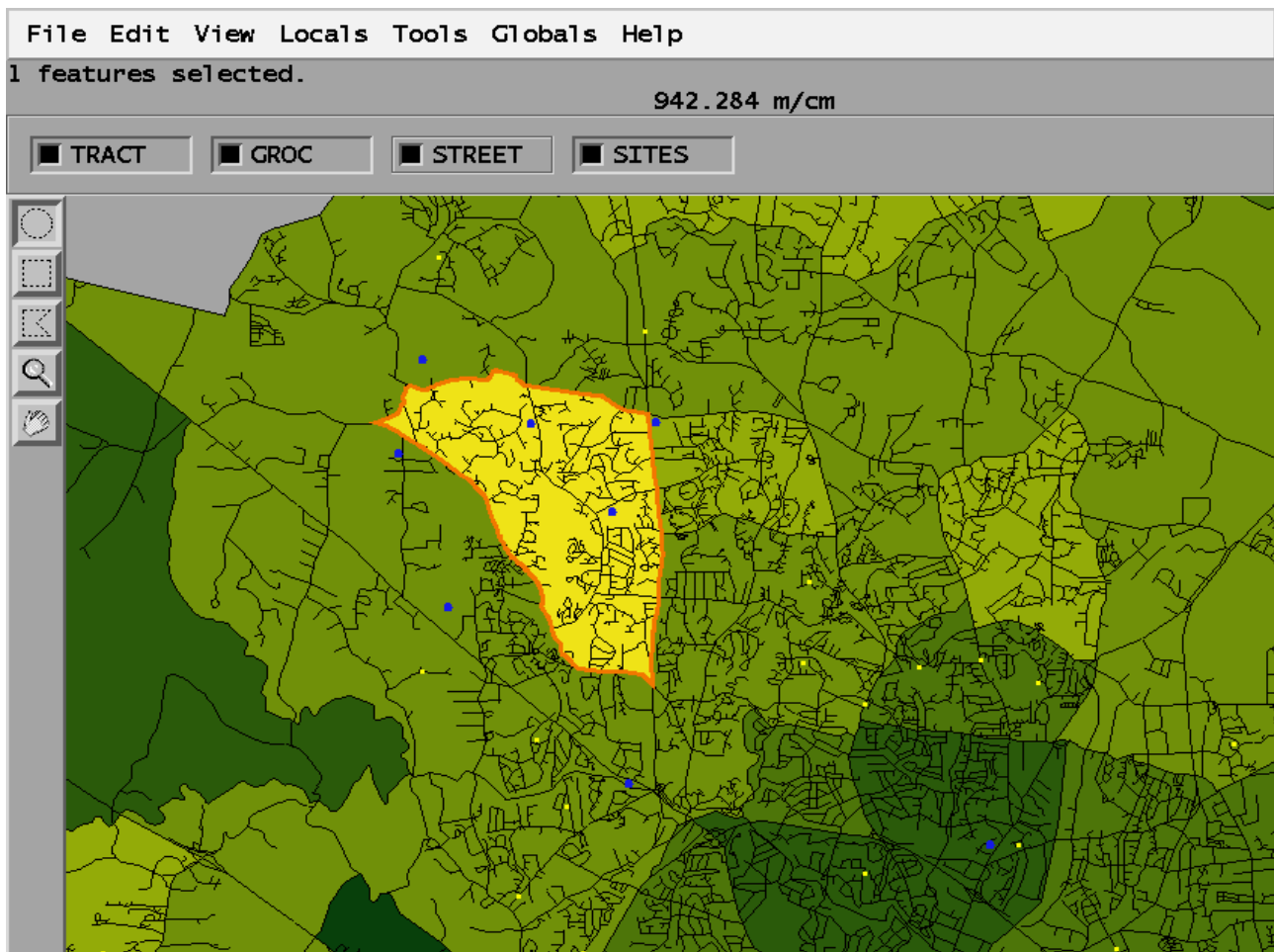
A következő felsorolás néhány olyan tárgyat és témakört foglal össze a teljesség igénye nélkül, amelyek oktatásánál a SAS rendszer jól alkalmazható mint demonstrációs, vagy mint feladat megoldó eszköz.

- bank- és pénzügytan pénzügyi számítások, pénzügyi modellek, banki és tőzsdei alkalmazások (pl. hitelminősítés, árfolyamelőrejelzés, hozamszámítások, likviditásmenedzsment, visszaélés-felderítés, portfóliómenedzsment, kockázatkezelés);
- diplomatervezés a SAS Institute kidolgozandó diplomaterveket ajánl;
- informatika adatbázisok menedzselése, adatbázis hozzáférés, SQL, információ tárház, vezetői információrendszerek, jelen téskészítés, döntéstámogatás, adat vizualizálás, prezentáció, 4GL nyelvek és eszközök, intelligens ügyfél-kiszolgáló rendszerek, számítóközpontok teljesítmény menedzsmentje, hálózatok optimalizálása, szimuláció, OLAP, RAD, OOAD, multi média, térinformatika, mesterséges intelligencia, programozás, tanuló algoritmusok;
- ipari folyamatok folyamat szabályozás, sorbanállási problémák vizsgálata, szimuláció, mérési adatok feldolgozása és kiértékelése;
- kontrolling, számvitel pénzügyi elemzések, jelentések, mutatók, konszolidáció, pénzügyi modellek, pénzügyi vezetői rendszerek, vállalati likviditásmenedzsment, kockázatkezelés, portfóliómenedzsment;

- közgazdaságtan mikro- és makrogazdasági modellezés, gazdasági döntések, elemzések, pénzügyi számítások;
- labor/mérési gyakorlat laboratóriumi elemzések, adatfeldolgozás, adatvizualizáció, dokumentálás;
- logisztika szállítási feladatok, sorbanállási problémák, tömegkiszolgáló rendszerek, készletmodellek;
- marketing előrejelzések, trendszámítás, piackutatás, piacelemzés számítógépes támogatása, „mi lenne ha” elemzések, piaci potenciál vizsgálata, szimuláció, ügyfélkifogások elemzése, politikai marketing elemzések;
- matematika klasszikus és többváltozós, alkalmazott statisztika, operációkutatás, matematikai programozás (pl. biztosítási matematika, biostatisztika, pénzügyi matematika, döntéstámogatás);
- menedzsment erőforrás- és projektmenedzsment, humánerőforrás menedzsment, üzleti folyamatok újratervezése (BPR), TQM, vezetői információrendszerek, döntéstámogatás, ügyfélminősítés;
- minőségügy Pareto elemzés, Ishikawa diagram, ellenőrző kártyák tervezése és használata, kísérlettervezés, SPC, QC, QI, TQM, LIMS, mérési adatok kiértékelése;
- szociológia demográfiai és szociológiai elemzések, adatvizualizáció.



5. ábra
Hitelkártya-visszaélések felderítése neurális hálózatokkal



6. ábra

A SAS/GIS® térinformatikai alkalmazása: egy piacszegmentációs esettanulmány

3. Összefoglalás

A SAS rendszer széles körben alkalmazható az oktatásban demonstrációs és feladatmegoldó eszközként egyaránt. A SAS Institute által nyújtott felhasználói támogatás, valamint a cég fejlesztések iránti elkötelezettsége biztosítja, hogy a SAS rendszer mindig a legkorszerűbb eszközöket és technológiákat nyújtja használóinak. Ezáltal lehetőséget ad a folyamatosan versenyképes, világ színvonalú oktatás szakmai háttérének fenntartására.

Irodalom:

Gartner Group: Peak Performance, Gatabase Marketing and Data Mining. March, 1996.

Gartner Group: Using Data Mining to Gain Competitive Edge, Part 2. Inside Gartner Group This Week, Aug. 30, 1995.

Piatetsky-Shapiro, Gregory: Knowledge Discovery in Databases, AAAI, 1991.

SAS Institute White Paper: Data Minng with the SAS System: From Data to Business Advantage.

További információk:

SAS Institute Magyarországi Képviselet
1122 Budapest, Határőr út 36.
Tel: (1) 202-6247, (1) 202-6047
Fax: (1) 202-5847

Internet:

web: <http://www.sas.com>
ftp: <ftp.sas.com>
BBS: 00 1 919-677-8155