

Využitie techník dolovania dát (Data Mining) pri riadení výrobných procesov

Using techniques of Data Mining in control of production process

Eudmila Magyariová

Abstract: The article was created during the PhD. study at the Slovak University of Technology and it represents the introduction to Knowledge Discovery in Databases, the one of Business Intelligence tools. It provides brief characteristic of this field, defines the particular steps of Knowledge Discovery in Databases and describes potential using techniques of Data Mining in control of production process.

Key words: Data Mining, Business Intelligence, Knowledge Discovery in Databases, production process, control

Abstrakt: Tento príspevok bol vytvorený v priebehu doktorandského štúdia na Slovenskej technickej univerzite a predstavuje úvod do problematiky získavania znalostí z databáz, ktorá tvorí jeden z nástrojov Business Intelligence. Ďalej poskytuje stručný prehľad tejto oblasti, charakterizuje jednotlivé kroky procesu získavania znalostí a popisuje možné využitie techník dolovania dát (Data Mining) pri riadení výrobných procesov.

Kľúčové slová: dolovanie dát, Business Intelligence, získavanie znalostí z databáz, výrobný proces, riadenie

1. Úvod

Žijeme v dobe, v ktorej je čoraz dôležitejšie mať tie správne informácie v správnom čase. Práve na tejto skutočnosti je postavený úspech podniku či firmy. Je zrejmé, že snahou každého podniku alebo firmy je dosiahnuť maximálny zisk v čo najkratšom čase s využitím čo najnižších nákladov. Uspokojenie týchto potrieb je závislé napríklad od kvality poskytovaných služieb, od orientácie na toho správneho zákazníka alebo predikcie rôznych problémov, ktoré by mohli ohroziť ich dosiahnutie. Firmy a podniky narábajú s množstvom dát a na ich zhromažďovanie používajú databázy alebo dátové sklady, ktoré v sebe ukrývajú

užitočné informácie, im prvý pohľad neviditeľné. Takto získané informácie môžu mať využitie pri riešení širokej škály problémov. Na to nám slúži práve Data Mining. Data Mining je definovaný ako proces získavania platných, neznámych a potencionálne užitočných informácií z dát [1]. Ako slovenský ekvivalent pre Data Mining sa používa dolovanie dát. Teoreticky je možné dolovanie dát použiť všade tam, kde dochádza k zberu údajov. V súčasnosti má Data Mining najväčšie využitie v obchode, bankovom sektore či telekomunikáciách, no začíname vidieť jeho prienik aj do výrobného procesu.

2. Dátové sklady

Jadrom informačného systému podniku sú transakčné (prevádzkové) databázové systémy. V týchto systémoch sú vykonávané operácie v reálnom čase a obsahujú údaje platné v danom okamžiku alebo inak povedané aktuálne dáta. Sú tiež označované aj ako systémy OLTP (Online Transaction Processing). Takéto databázy nie sú prispôbolené a vhodné na rozsiahle analýzy. V podnikoch sa obvykle dáta z transakčných databáz zozbierajú, upravujú a v pravidelných intervaloch sú zavádzané do dátového skladu. Dátové sklady umožňujú online spracovanie dát, preto sú pomenované tiež pojmom systémy OLAP (Online Analytical Processing). Zber, úprava a zavedenie dát do dátového skladu sa realizuje prostredníctvom dátovej pumpy. Tento proces je tiež nazývaný ako ETL proces, kde E znamená Extraction čiže zozbieranie dát, T znamená Transformation čiže čistenie dát a ich úprava do potrebnej formy, L znamená Loading čiže zavedenie dát. Kým prevádzkové databázové systémy sú postavené na relačnom databázovom modeli, pri dátových skladoch ide o multidimenzionálny databázový model, ktorý je založený na dátovej kocke. Ten umožňuje, aby sa nad dátami dali robiť zložitejšie analýzy. Súbor techník, ktoré sa používajú na získanie informácií z dátových skladov nazývame OLAP operácie. V niektorých literatúrach sú OLAP operácie označované ako manuálne dolovanie v dátach [3] [4]. Medzi OLAP operácie zahrňame :

- ROLL UP - atribút dátovej kocky nadobúda všeobecnejší význam, postupuje sa z nižšej úrovne do vyššej, napríklad atribút mesiac postupuje na vyššiu úroveň rok
- DRILL DOWN - atribút dátovej kocky nadobúda konkrétnejší význam, postupuje sa z vyššej úrovne do nižšej, napríklad atribút rok postupuje smerom na nižšiu úroveň mesiac
- SLICE - výsledkom je podmnožina dátovej kocky, výber jednej dimenzie
- DICE - výsledkom je podmnožina dátovej kocky, výber dvoch alebo viacerých dimenzií

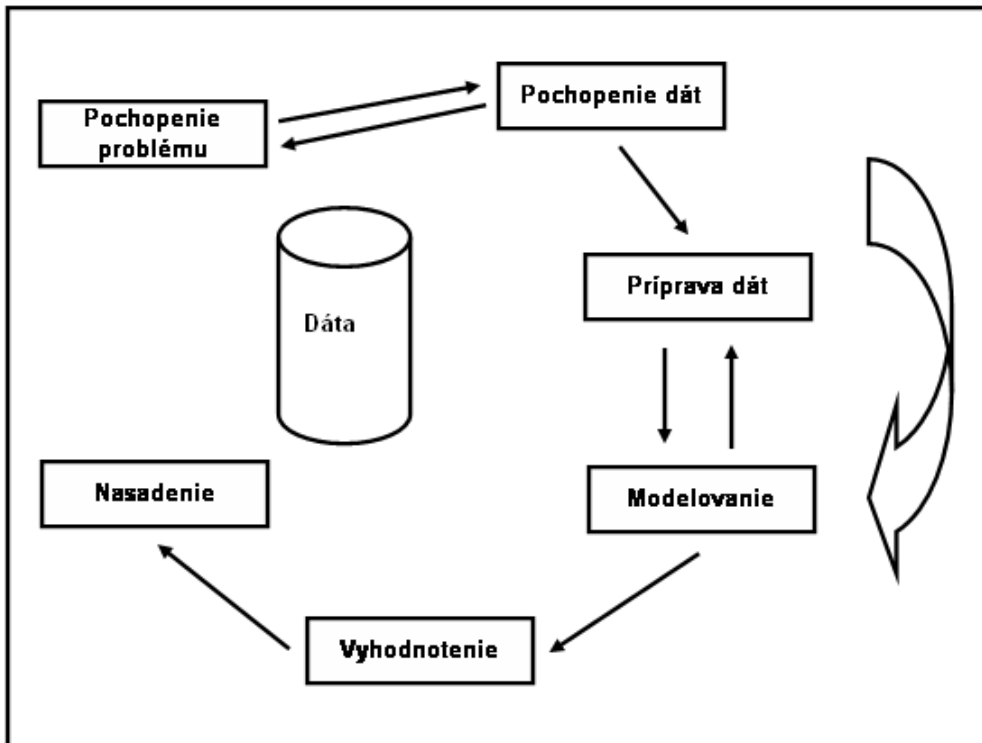
- PIVOT – rotácia osí dátovej kocky, poskytuje iné pohľady na dáta

3. Získavanie znalostí z databáz

Business Intelligence (BI) je proces transformácie dát na informácie a prevod týchto informácií na poznatky [2] . Data Mining (dolovanie dát) je charakterizovaný ako jeden z nástrojov BI. Často je možné sa stretnúť aj s pojmom Knowledge Discovery in Databases (získavanie znalostí z databáz – KDD [3], [4]) . Vzťah medzi KDD a dolovaním dát býva dvojaký. Buď je pojem Data Mining myslený ako synonymum ku KDD. Alebo je pojmom KDD označovaný celý proces získania znalostí z dát a Data Mining je iba jedným krokom v tomto procese, ide o aplikovanie zvolených techník na dané dáta. V príspevku bude použitý pojem Data Mining vo význame jedného kroku v procese získavania znalosti z databáz.

Keďže v procese získavania znalosti je možné použiť rôzne algoritmy pre postup práce, bola tu snaha o vytvorenie akejsi normy. Jednou takou normou je aj CRISP-DM (Cross-Industry Standard Process for Data Mining), kde je načrtnutý odporúčaný postup práce v procese získavania znalostí z databáz. Podľa CRISP-DM sa získavanie znalostí z databáz skladá z nasledujúcich krokov [5] :

- Pochopenie problému
- Pochopenie dát
- Príprava dát
- Modelovanie
- Vyhodnotenie
- Nasadenie



Obrázok 1: Fázy KDD podľa CRISP-DM

Fáza pochopenia problému je úvodnou fázou získavania znalostí. V tomto kroku je snaha o vymedzenie cieľov z obchodného aspektu, o ich pochopenie a preskúmanie faktorov, ktoré by mohli ovplyvniť proces KDD. Medzi tieto faktory sú zaradené dostupné zdroje dát, rôzne obmedzenia či riziká. Úspech celého procesu je dosť závislý na tejto fáze, pretože zle pochopenie obchodných cieľov či nedôkladné preskúmanie faktorov, vedie zákonite k neúspechu alebo k problémom. Nakoniec sú tieto obchodné ciele popísané prostredníctvom technických pojmov.

Vo fáze pochopenia dát ide o zber dát z dostupných zdrojov, ďalej táto fáza pokračuje aktivitami, ktorých zámerom je pochopiť a popísať dáta, ktoré máme k dispozícii a identifikovať ich kvalitu. Stretávame sa tu aj s prvými analýzami týchto dát prostredníctvom dotazov, reportov či vizualizácie. Tieto analýzy sú môžu byť nápomocné pri popise dát alebo prostredníctvom nich formulujeme prvé hypotézy.

Fáza prípravy dát zahŕňa všetky aktivity, ktoré sú potrebné na vytvorenie množiny dát pripravenej na modelovanie. Toto je pomerne náročná fáza, pretože dáta bývajú v rôznych formátoch z rôznych zdrojov, môžu obsahovať výrazne odklonené, chýbajúce, zašumené alebo nekonzistentné dáta. Okrem vyčistenia dát je vo fáze príprave dát realizované aj hľadanie nových odvodených atribútov, ktoré sú potrebné pre dosiahnutie cieľov. Ak sú vstupné dáta z rôznych zdrojov, ich integrácia sa uskutočňuje práve v tomto kroku. Na konci

tejto fázy je snaha o redukciu dát, pretože spracovanie obrovských dát je časovo náročné a často sa ani nedá realizovať. Je odporúčané venovať tomuto kroku dostatok času a byť dôsledný, pretože kvalita vstupných dát býva kľúčom k úspechu.

Vo fáze modelovania ide o aplikovanie už konkrétnej techniky Data Miningu na pripravené dáta a sú nastavované parametre danej techniky na optimálne hodnoty. Pokračuje sa vyhodnotením a testovaním kvality modelu. Ak sa výsledok nezhoduje s našimi predstavami, sú zmenené parametre techniky alebo samotná technika. Môže tak nastať opakovanie jedného alebo viacerých krokov v procese získavania znalostí z databáz, pretože napríklad zmenená technika môže klásť iné požiadavky na reprezentáciu dát, ktoré vstupujú do tejto fázy [4].

Kým v predchádzajúcej fáze je vyhodnocovaná kvalita modelu z pohľadu Data Miningu, fáza vyhodnotenia je orientovaná na vyhodnotenie výsledkov z hľadiska obchodných cieľov. V tejto fáze je potrebná úzka spolupráca s odborníkom na oblasť z ktorej bola zadaná úloha, ktorý najlepšie dokáže posúdiť, či tento proces priniesol kvalitné a použiteľné výsledky.

Fáza nasadenia sa zameriava na implementáciu výsledkov dolovania dát do každodenného života. Vytvára sa plán sledovania výsledkov Data Miningu. Toto sledovanie výsledkov predstavuje akúsi spätnú väzbu celého procesu.

4. Techniky prípravy dát

Výhodou dátových skladov z hľadiska dolovania dát je to, že dáta v nich už bývajú vyčistené a zoskupené z rôznych zdrojov, čím sa nám zníži časová náročnosť procesu získavania znalostí z databáz. Z toho vyplýva, že najlepším zdrojom dát pre proces dolovania dát sú práve dátové sklady. Pri ostatných zdrojoch sa využívajú rôzne metódy na to, aby sme dosiahli požadovanú kvalitu dát. Pri chýbajúcich hodnotách používame aritmetický priemer, medián, modus alebo regresiu. Hodnoty, ktoré sú odlišné od ostatných hodnôt tj. nachádzajú ďaleko od strednej hodnoty (outliers) môžeme nájsť prostredníctvom zhlukovania, alebo na vyhladenie takýchto dát je možné použiť regresnú analýzu. Ďalej je možné mať k dispozícii dáta z rôznych zdrojov. Vtedy je žiaduce, aby sme tieto dáta integrovali. Pri integrácii sú dôležité tzv. metadáta, čo sú dáta o dátach a obsahujú informácie o typoch jednotlivých atribútov, popisujú jednotlivé objekty, čím pomôžu identifikovať napríklad redundancie. Na redukciu dát je možné použiť napríklad agregáciu alebo vzorkovanie. Samozrejme, toto nie sú všetky dostupné techniky, ktoré je možné použiť pri príprave dát.

5. Dolovanie dát a techniky dolovania dát

Podľa problému, ktorý sa bude riešiť sa môžeme stretnúť s týmito typmi úloh dolovania dát [3][4] [9]:

- Klasifikácia
- Predikcia
- Asociačné pravidlá
- Zhlukovanie
- Deskripcia

Pri predikcii a klasifikácii je daná trénovacia množina dát, kde sú známe aj hodnoty atribútu, ktorý sa bude predikovať (cieľový atribút). Na základe tejto trénovacej množiny sa vytvorí model a pri novom objekte sa bude cieľový atribút predikovať na základe známych atribútov tohto objektu podľa modelu. Rozdiel medzi nimi je v iba tom, že pri klasifikácii atribút, ktorý sa bude predpovedať je diskretný, pri predikcii ide o predpovedanie spojitého (numerického) atribútu.

Asociačné pravidlá sú pravidlá, ktoré popisujú kategórie alebo triedy, vyjadrujú súvislosti medzi objektmi . Často používajú spojenie IF-THEN, logické spojky AND,OR a určuje sa spoľahlivosť pravidla pomocou pravdepodobnosti.

Zhlukovanie sa používa na zgrupovanie objektov do tried a to tak, že v tej istej grupe sú rovnaké alebo príbuzné objekty a dve grupy sa navzájom líšia skupinami rôznych objektov. Pri tomto type úlohy nie je dostupná žiadna trénovacia množina dát. Zhlukovanie sa nepoužíva len pri samotnom dolovaní, ale je aj technikou prípravy dát.

Pri deskripcii ide o popis množiny dát. Podľa popisu sú vytvárané pravidlá, na základe ktorých sa objekty zaraďujú do kategórii. Deskripcia využíva často generalizáciu. To značí, že atribúty objektov nadobúdajú všeobecnejší význam, na základe ktorého sa znižuje množina dát, až nadobudne takú formu, pri ktorej je možné vytvoriť kategórie. Tieto kategórie sa dajú popísať logickými pravidlami, alebo je možné použiť popis pomocou grafov či krížovej tabuľky [4]. Keďže generalizácia je aj technikou prípravy dát podobne ako zhlukovanie, často sa pri jej využití zjednotia fázy prípravy dát a modelovania.

5.1. Lineárna regresia

Patrí medzi techniky, ktorá využíva štatistické metódy. Z hľadiska Data Miningu je technikou predikcie, ktorá je založená na lineárnej závislosti dvoch spojitých atribútov, kde jedným je cieľový atribút a druhým je predikujúci atribút. Tento vzťah je vyjadrený rovnicou: $Y = a + bX$

kde Y je atribút, ktorý sa predikuje (cieľový) a X je predikujúci atribút.

Pri viacnásobnej lineárnej regresii atribút, ktorý sa predikuje Y je lineárne závislý na predikujúcich atribútoch X_1 až X_n a vzťah medzi nimi je vyjadrený polygómom prvého stupňa: $Y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_nx_{ni}$

kde Y_i je hodnota atribútu, ktorý sa predikuje Y pre i-tý príklad v trénovacej množine ($i = 1, \dots, k$) dát a x_{ji} je číselná hodnota predikujúceho atribútu X_j pre i-tý príklad z trénovacej množiny[4].

5.2. Logistická regresia

Je podobná lineárnej regresii s tým rozdielom, že sa používa na predikciu cieľového atribútu, ktorý nadobúda diskkrétne hodnoty. Aby sa dala využiť regresia, transformuje sa atribút, ktorý sa ide predikovať na spojitú hodnotu, ktorá je funkciou pravdepodobnosti výskytu udalosti. Vzťah je vyjadrený rovnicou: $\log(p/(1-p)) = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$

kde $p/(1-p)$ je pravdepodobnosť výskytu udalosti, b_0 až b_n sú koeficienty a X_1 až X_n sú predikujúce atribúty [7].

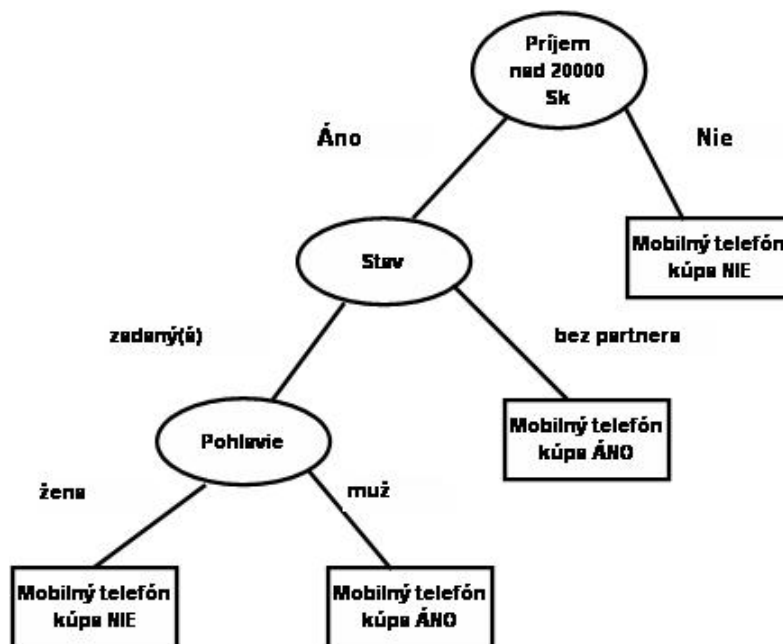
5.3. Rozhodovacie stromy

Rozhodovacie stromy sú štruktúry na predikovanie cieľového atribútu za pomoci jednoduchých rozhodovacích pravidiel. Sú technikou aj klasifikácie aj predikcie. Na zostrojenie rozhodovacieho stromu sa využíva testovacia množina dát. Nové objekty teda postupujú jednotlivými uzlami stromu až k listovému, ktorý určí hodnotu cieľového atribútu. Je vyjadrený vo forme grafu, ktorý má tieto prvky [3]:

- Koreňový uzol - predstavuje predikujúci atribút
- Vnútorňý uzol - predstavuje predikujúci atribút alebo atribúty
- Listový uzol - reprezentuje triedu alebo cieľový atribút
- Vetva - vyjadruje výsledok testu

Pri tvorbe rozhodovacích stromov je dôležité určiť rozdelenie jednotlivých predikujúcich atribútov do uzlov a vyhodnotiť kvalitu takéhoto zatriedenia. Pre výber atribútu do uzla sa používajú rôzne kritéria ako informačný zisk, informačná entropia, Gini-ho index, Chí-kvadrát test pri diskretných cieľových atribútoch, alebo F- test a redukcia rozptylu pri spojitých cieľových atribútoch [9]. Pri tvorbe rozhodovacích stromov sa aplikujú rôzne algoritmy napríklad ID3 [3] [9] , C4.5 [3] , C5.0 [9], CART (classification and regression trees)[9].

Problémom pri rozhodovacích stromoch býva aj to, že môžu byť príliš rozvetvené a tak sa stavajú zložité. Riešením sú techniky orezávania a to buď počas vytvárania stromu, alebo po vytvorení. Tie by mali zaručiť pri zjednodušenjšom znázornení rovnakú presnosť predikcie cieľového atribútu.

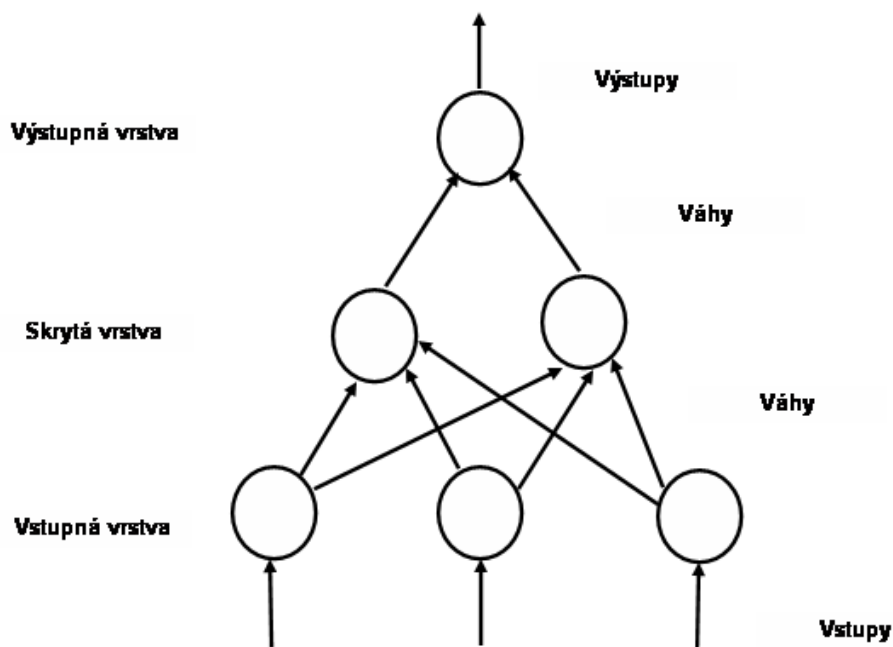


Obrázok 2: Príklad rozhodovacieho stromu

5.4. Neurónové siete

Sú technikou zhlukovania, klasifikácie aj predikcie. Neurónové siete sú využiteľné pri identifikovaní súvislosti medzi množinou dát a na nájdenie vzorov. Patria do oblasti umelej inteligencie a simulujú funkcie ľudského mozgu. Tento proces je možné si predstaviť ako prijímanie informácie a poučenie sa s každej skúsenosti. [7]. Najvyužívanejším typom neurónových sietí pri dolovaní dát sú viacvrstvové dopredné neurónové siete. Takáto neurónová sieť obsahuje určitý počet jednotiek nazývaných neuróny (uzly) , ktoré sú

usporiadané vo vrstvách. Každý neurón vykonáva jednoduchú úlohu spracovania informácií transformovaním vstupnej informácie na spracovanú výstupnú informáciu. V trojvrstvových dopredných neurónových sieťach sú uzly usporiadané do troch vrstiev : vstupná , skrytá a výstupná. Vstupná vrstva reprezentuje predikujúce atribúty, výstupná reprezentuje atribút alebo atribúty , ktoré chceme predikovať. Skrytá vrstva spája vstupnú a výstupnú vrstvu, vyjadruje vzťah medzi nimi a spracováva informácie zo vstupov.[10]. V neurónových sieťach sú znalosti reprezentované v podobné váh medzi jednotlivými uzlami, ktoré sú znázornené hranami.



Obrázok 3: Príklad neurónovej siete

Techník dolovania dát je mnoho, no vzhľadom na obmedzený priestor príspevku nie je možné ich predstaviť všetky. Medzi techniky dolovania dát patrí aj zhľukovanie a asociačné pravidlá, ktoré boli vyššie spomenuté aj ako typy úloh dolovania dát. Podrobný popis spomenutých techník, ako aj ďalších, je možné nájsť v literatúre uvedenej na konci príspevku.

6. Využitie techník dolovania dát (Data Mining) pri riadení výrobných procesov

V každom výrobnom podniku prebieha proces zberu a spracovania dát, ktoré súvisia s jeho činnosťou. Ich analýzou je teoreticky možné získať informácie a znalosti, ktoré sa budú dať využiť v budúcnosti pri optimalizácii výrobného procesu vo výrobných podnikoch.

Výrobný proces obsahuje všetky činnosti, ktorými sa dosiahne transformácia vstupov na výstupy. Je to súhrn výrobných, skladovacích, manipulačných a dopravných operácií. Finálnym výstupom výrobného procesu môže byť výrobok alebo služba. Cieľom výrobného procesu je poskytnúť zákazníkovi kvalitný výrobok alebo službu pri minimalizácii nákladov.

Jednou z hlavných požiadaviek na výrobný proces je jeho plynulosť. Analýzou dát, ktoré vznikajú v priebehu výrobného procesu a detekciou odchýlok sa môžu definovať problematické časti výrobného procesu. Prostredníctvom sledovania a analýzy parametrov výroby sa môže zabezpečiť nielen aplikovanie kontrolných mechanizmov v procese výroby na zníženie alebo celkové odstránenie porúch a stavov ohrozujúcich plynulosť výroby, ale aj podpora automatizovaného riadenia výroby.

Dolovanie dát môže poskytnúť napríklad aj poznatky, na základe ktorých bude možné zvýšiť alebo optimalizovať časové aj výkonové využitie kapacít, napríklad identifikovať prestoje a nečinnosť.

Analýzou dát sa môže dosiahnuť aj optimalizácia materiálového toku. Úlohou je teda hľadať také riešenie, ktoré by zabezpečilo nepretržitý pohyb materiálu pri minimálnych nákladoch na prepravu.

Iným problémom, ktorý sa môže riešiť využitím dolovania dát je zásobovanie a skladovanie materiálu. V súčasnosti sa na Slovensku využíva hlavne filozofia JIT (Just in Time), ktorá predpokladá minimálne zásoby materiálu vstupujúceho do výrobného procesu. Analýzou dát sa môžu dať predpovedať poruchy v dodávke materiálu a zistiť minimálne zásoby, ktoré treba zaistiť, aby nebola ohrozená plynulosť výroby.

Ďalšou požiadavkou na výrobný proces je kvalita výstupov. Využitím procesu získavania znalostí sa môže realizovať zavedenie účinných opatrení na zabezpečenie kvality výrobkov alebo služieb. Na základe údajov z výrobného procesu, pravdaže za pomoci techník dolovania dát, sa môže zistiť pri akej fáze alebo fázach výrobného procesu nastávajú chyby v kvalite výrobku. Ďalej sa môžeme zamerať na to, ktoré činitele sú z hľadiska kvality výrobku alebo služieb tie kritické, na základe týchto poznatkov eliminovať tieto chyby a nasadiť kontrolné mechanizmy.

Samozrejme techniky dolovania dát sa dajú využiť na predpovedanie množstva a druhu potreby (dopytu), a tak zabezpečiť čo najrýchlejšie uspokojenie požiadaviek zákazníkov. Dolovanie dát môže byť užitočným nástrojom aj pri plánovaní výroby, kde analýzou historických dát je možné koordinovať všetky aktivity vo výrobnom procese, tak aby sa

zabezpečilo upokojenie dopytu po výrobkoch či službách. Ďalej sa dolovanie dát využíva na orientovanie marketingovej kampane na tú správnu skupinu zákazníkov.

Využitím techník dolovania dát pri riadení výrobných procesov sa môže predpovedať vznik porúch, havarijných stavov alebo stavov, ktoré ohrozujú plynulosť výrobného procesu. Získa sa tak východisková a podporná informácia pri rozhodovaní v kritických situáciách riadenia výrobného procesu. Ďalej poznatky, ktoré získame prostredníctvom dolovania dát sa dajú využiť pri znižovaní výrobných nákladov, maximalizácii zisku, získaní vyššieho podielu na trhu, či predbehnutí konkurencie v kvalite svojich výrobkov alebo služieb.

Ako vidieť, dolovanie dát môže mať uplatnenie aj vo výrobnom procese. Základom pre použitie techník dolovania dát je monitorovanie celého priebehu výrobného procesu a zbieranie údajov o všetkých krokoch v rámci neho. V konečnom dôsledku jedine implementácia výsledkov dolovania dát do praxe určí skutočnú mieru využitia dolovania dát v jednotlivých krokoch výrobného procesu.

Je preto vhodné zistiť, ktorá technika dolovania dát je vhodná na riešenie takýchto konkrétnych problémov, ktoré sa môžu vyskytnúť vo výrobnom procese a či výsledky dolovania dát efektívne riešia problémy vo výrobných podnikoch a ak áno, tak v akej miere.

7. Zoznam bibliografických odkazov

- (1) FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P. From Data Mining to Knowledge Discovery in Databases. AI- Magazine, 1996, dostupné [on-line] : <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf> .
- (2) LACKO, L. Oracle : Správa, programovaní a použití databázového systému. Computer Press Brno, 2003. 464 s. ISBN 80-7226-699-3.
- (3) HAN, J., KAMBER, M. Data Mining – Concepts and Techniques. San Francisco: Morgan Kaufmann Publ, 2001. 550 s. ISBN 1-55860-489-8.
- (4) PARALIČ, J. Objavovanie znalostí v databázach, Košice, 2003. 80 s. ISBN 80-89066-60-7.
- (5) CRISP DM, dostupné [on-line] : <http://www.crisp-dm.org> .
- (6) NOVOTNÝ, O., POUR, J., SLÁNSKY D. Business Intelligence: Jak využití bohatství ve vašich datech. Grada Publishing a.s., 2004. 254 s. ISBN 8024710943.

- (7) RUD, O.P. Data Mining: Praktický průvodce dolováním dat pro efektivní prodej, cílený marketing a podporu zákazníku. Praha: Computer Press, 2001. 330 s. ISBN 8072265776.
- (8) WITTEN I.H., FRANK, E. Data Mining : Practical Machine Learning Tools and Techniques. Elsevier, 2005. 525 s. ISBN: 0-12-088407-0.
- (9) BERRY, M.J.A, LINOFF G.S. Data Mining Techniques for Marketing, Sales and Customer Relationship Management: Second Edition. Wiley Publishing, 2004. 672 s. ISBN:0471470643.
- (10) MAIMON, O., ROKACH, L. Data Mining and Knowledge Discovery Handbook. Springer Science + Business Media, Inc., 2005. 1383 s. ISBN- 10: 0-387-24435-2.

8. Adresa autora:

Ludmila Magyariová, Ing.
Sociálna Poist'ovňa, pobočka Trnava
Ul. V. Clementisa 24/A
917 22 Trnava
ludmila.magyariova@gmail.com