

# Podpora metodológie Six Sigma s KDD

## Six Sigma methodology support with KDD

*Andrej Trnka, Katedra aplikovanej informatiky, FPV UCM v Trnave*

**Abstract:** This paper describes possibilities of improvement Six Sigma methodology with KDD (Knowledge discovery in databases), especially with Data Mining methods. Six Sigma methodology consists of 5 phases (DMAIC cycle): Define, Measure, Analyse, Improve and Control. Within those phases we can use many of mathematical and statistical methods. We could improve results of individual phases with patterns obtained from Data Mining.

**Key words:** Six Sigma, KDD, Data Mining, DMAIC

**Abstrakt:** Článok popisuje možnosti zlepšenia metodológie Six Sigma pomocou techník KDD (Knowledge discovery in databases – Získavanie znalostí z databáz). Metodológia Six Sigma pozostáva z piatich fáz (DMAIC cyklus): Definícia, Meranie, Analýza, Zlepšenie a Riadenie. Vo všetkých týchto fázach môžeme použiť matematické a štatistické metódy. Výsledky jednotlivých fáz by sme mohli zlepšiť pomocou vzoriek získaných z dolovania dát

**Kľúčové slová:** Six Sigma, KDD, dolovanie dát, DMAIC

### 1. Six Sigma

Six Sigma je metodológia určená na minimalizáciu chýb a maximalizáciu hodnoty procesu. Každá chyba podniku alebo ľudí má napokon za následok určitú cenu - stratu zákazníka, potrebu opakovať jednotlivé operácie procesu, výmenu pokazených častí, stratu času, zbytočné náklady na materiál, zníženú účinnosť procesu a nízku produkciu. Vo všeobecnosti je možné konštatovať, že chyby stoja podniky 20 až 30 percent ich príjmu. (1)

Podľa Lindermana (2) je Six Sigma metodológia pre organizované a systematické riešenie problémov pre strategické zlepšenie systému a nového produktu, ktorá sa spolieha na matematické a štatistické metódy s cieľom dosiahnutia dramatického zníženia zákazníkom definovaných medzí a/alebo zlepšenia kľúčových výstupných premenných procesu.

Hodnota 6 Sigma korešponduje s 3.4 nepodarkami na milión príležitostí (DPMO - Defects per million opportunities).

**Tabuľka 1: Hodnoty Sigma, DPMO a ekvivalentný zisk (3)**

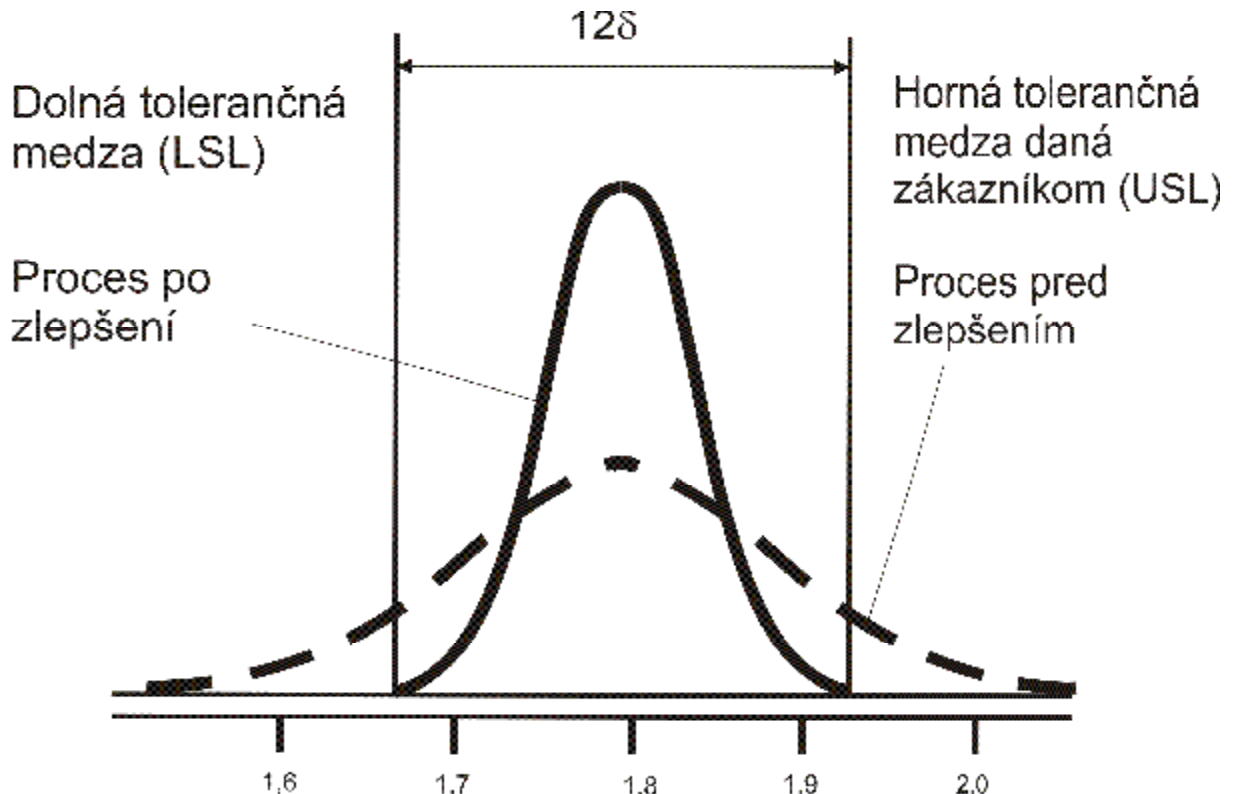
| SIX SIGMA hodnota | DPMO    | Zisk (%) |
|-------------------|---------|----------|
| 1                 | 691 462 | 30.85    |
| 2                 | 308 538 | 69.146   |
| 3                 | 66 807  | 93.319   |
| 4                 | 6 210   | 99.379   |
| 5                 | 233     | 99.9767  |
| 6                 | 3.4     | 99.99966 |

Proces je možné pomocou metodológie Six Sigma riadiť dvomi spôsobmi:

- DMAIC – pre existujúci proces,
- DMADV – pre nový proces.

Keďže v ďalšom výskume sa budeme zaoberať iba aplikáciou a zlepšením procesu už existujúceho, fázy DMADV nebudú súčasťou podpory metodológie Six Sigma.

Všeobecná interpretácia metodológie Six Sigma sa dá zobrazit' pomocou histogramu:



**Obrázok 1: Zlepšenie procesu pomocou metodológie Six Sigma**

## 1.1. DMAIC

Fázy DMAIC charakterizujú postup zavedenia metodológie Six Sigma do už existujúceho procesu. V jednotlivých fázach sa používajú rôzne matematické, štatistické ale aj manažérske metódy. Cieľom nášho výskumu je navrhnúť podpory metodológie Six Sigma pomocou matematických a štatistických metód s využitím metód a úloh dolovania dát.

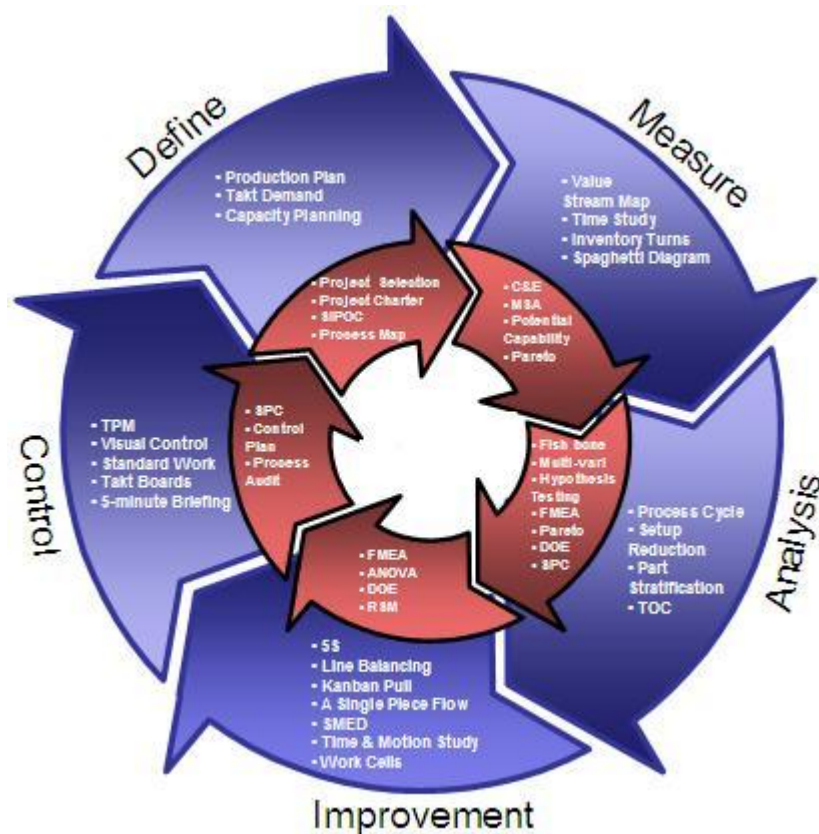
**D – Define** – definícia cieľov projektu a existujúcich procesov

**M – Measure** – meranie kľúčových aspektov procesov a zber relevantných dát

**A – Analyze** – analýza dát na verifikáciu závislostí typu príčina - dôsledok.

**I – Improve** – zlepšenie alebo optimalizácia procesu

**C – Control** – riadenie procesu tak, aby každá odchýlka od cieľovej hodnoty bola korigovaná skôr, než bude mať za následok chybu (nepodarok). (4)



Obrázok 2: Cyklus DMAIC s jednotlivými nástrojmi (5)

## 2. KDD

Získavanie znalostí z databáz (KDD) je proces identifikácie validity, originality, užitočnosti a pochopiteľnosti vzoriek údajov z veľkých databáz. Dolovanie dát je jadrom procesu dobývania znalostí. Proces získavanie znalostí z databáz sa riadi modelom CRISP-DM. Tento model sa snaží nájsť univerzálne použiteľný postup pre dolovanie dát. Metodika CRISP-DM uvádza nasledujúce etapy:

- porozumenie problematike (Business Understanding)
- porozumenie dátam (Data Understanding)
- príprava dát (Data Preparation)
- modelovanie (Modeling)
- vyhodnotenie výsledkov (Evaluation)
- využitie výsledkov (Deployment) (6)

Samotný proces dolovanie dát sa uskutočňuje vo fáze modelovania. V tejto fáze sa na dáta aplikujú rôzne metódy a algoritmy dolovania dát a získavajú sa nové vzorky, ktoré môžu byť pretransformované na novoobjavené vedomosti. Táto fáza je z pohľadu využiteľnosti KDD v metodológii Six Sigma podstatná, pretože nesprávne určený algoritmus alebo úloha môže priniesť nesprávne výsledky.

Proces dolovania dát a využitia znalostí je možné použiť pri riadení procesov v rozličných sektoroch priemyselných odvetví. (7)

## 3. Využitie procesu KDD

Samotné využitie procesu Získavania znalostí z databáz a hlavne metód dolovania dát môže priniesť nové znalosti z jednotlivých fáz metodológie Six Sigma. Nájdenie nových vzoriek, súvislostí alebo predikcií stavu procesu nie je pomocou bežných nástrojov použitých vo fázach DMAIC celkom dobre uskutočniteľné. Aj keď sa nejedná o riadenie procesu v reálnom čase, novo objavené vzorky v dátach môžu umožniť zlepšenie výsledkov fáz DMAIC. To má priamy vplyv na hodnotu DPMO, ktorá zase určuje hodnotu Sigma samotného procesu.

Zlepšovanie kvality procesu a jeho optimalizácia dokážu bez dodatočných vstupných nákladov do procesu zvýšiť zisk podniku.

Na zlepšenie fáz DMAIC je možné viacero úloh dolovania dát:

- identifikácia cieľových zákazníkov,
- predikcia správania sa zákazníka na určitú ponuku,
- vytvorenie správnej ponuky pre zákazníka,

- príprava dát pre analýzu,
- klasifikácia dát,
- identifikácia dôležitých premenných procesu,
- predikcia poruchovosti stavu výrobných zariadení,
- analýza predaja tovaru,
- detekcie podvodov,
- ...

Začlenenie úloh dolovania dát do jednotlivých fáz metodológie Six Sigma si vyžaduje historické dáta. Vhodný spôsob na uloženie takýchto dát sú dátové sklady. Správne využitie získaných vedomostí môže byť podporené práve implementáciou dátových skladov do celého procesu. Implementácia dátových skladov však vyžaduje precízny návrh testovania dátového skladu a zohľadňovanie bezpečnostných rizík nových informačných technológií zavádzaných do procesu riadenia.(8),(9)

Z dôvodu nutnej potreby historických dát je budúci výskum podpory metodológie Six Sigma zameraný práve na existujúce procesy (DMAIC). Dôslednou analýzou existujúceho procesu je však možné vytvoriť modely, ktoré môžu byť použité aj pri zavádzaní nových procesov (DMADV).

Pri využití metód dolovania dát pri jednotlivých úlohách je potrebné vykonať cyklus metodiky CRISP-DM a tým zabezpečiť nielen správne namodelovanie, ale aj vyhodnotenie a využitie výsledkov.

Pri nasadení metód a úloh dolovania dát do procesu jednotlivých fáz DMAIC je potrebné kvôli budúcim analýzám zabezpečiť vo fáze C – Riadenie zber údajov z procesu.

#### **4. Zoznam bibliografických odkazov**

- (1) Gygi, Craig – DeCarlo, Neil – Wiliams, Bruce: Six Sigma for Dummies. Wiley Publishing, 2005, 344 p., ISBN 0-7645-6798-5
- (2) Linderman K., et. al: Six Sigma: A Goal-theoretic Perspective. In: The Journal of Operations Management 21, 2003, p. 193-203, ISSN 0272-6963
- (3) Truscott Wiliam: Six Sigma: Continual Improvement for Businesses. Butterworth-Heinemann, 2003, 250 p., ISBN 0-7506-57650

- (4) De Feo, Joseph - Barnard, William: JURAN Institute's Six Sigma Breakthrough and Beyond-Quality Performance Breakthrough Methods, McGraw-Hill Publishing, 2005, ISBN 0-07-059881-9
- (5) DMAIC, on-line [<http://www.6sixsigma.com>], [cit.:2.12.2009]
- (6) CRISP-DM: Process model, on-line [<http://www.crisp-dm.org>], [cit.:2.12.2009]
- (7) Makyš, Peter - Kebísek, Michal: Možnosti spracovania dát pre podporu rozhodovania pri riadení výrobných systémov. Data processing possibilities decision support in production system control. In: Informačné technológie v manažmente výrobných systémov = Information Technologies in Management of Production Systems : Zborník. Medzinárodná vedecká konferencia. Nitra, 20.- 21.5.2004. - Nitra : Slovenská poľnohospodárska univerzita v Nitre, 2004. - ISBN 80-8069-364-1. - s. 144-149
- (8) Tanuška, Pavol - Moravčík, Oliver - Važan, Pavol - Miksa, František: The Proposal of Data Warehouse Testing Activities. - Vega 1/4078/07. In: CECIIS. - ISSN 1847-2001. - 20th Central European Conference on Information and Intelligent Systems : Conference Proceedings. September 23-25, 2009, Varaždin, Croatia. - Zagreb : University of Zagreb, 2009
- (9) Halenár, Igor - Kebísek, Michal - Kuník, Stanislav: Bezpečnostné riziká informačných technológií v ASR. In: Process Control 2006 : 7th International Scientific-Technical Conference. Kouty nad Desnou, Czech Republic, 13.-16.6.2006. - Pardubice : University of Pardubice, 2006. - ISBN 80-7194-860-8. - R155-1/R155-5

## **5. Adresa autora:**

Andrej Trnka, Ing.  
Katedra aplikovanej informatiky FPV UCM  
Nám. J. Herdu 2  
917 01 Trnava  
[andrej.trnka@ucm.sk](mailto:andrej.trnka@ucm.sk)