



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA PODNIKATELSKÁ
ÚSTAV INFORMATIKY

FACULTY OF BUSINESS AND MANAGEMENT
INSTITUT OF INFORMATICS

ANALÝZA FIREMNÍCH PROCESŮ A NÁVRH OPTIMALIZACE S VYUŽITÍM DATA MININGU

BUSINESS PROCESS ANALYSIS AND OPTIMIZATION DESIGN USING DATA MINING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

BRANISLAV BARTÍK

VEDOUCÍ PRÁCE
SUPERVISOR

ING. JIŘÍ KŘÍŽ, PH.D.

BRNO 2011

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Bartík Branislav

Manažerská informatika (6209R021)

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách, Studijním a zkušebním řádem VUT v Brně a Směrnicí děkana pro realizaci bakalářských a magisterských studijních programů zadává bakalářskou práci s názvem:

Analýza firemních procesů a návrh optimalizace s využitím data miningu

v anglickém jazyce:

Business Process Analysis and Optimization Design Using Data Mining

Pokyny pro vypracování:

Úvod

Vymezení problému a cíle práce

Teoretická východiska práce

Analýza problému a současné situace

Vlastní návrhy řešení, přínos návrhů řešení

Závěr

Seznam použité literatury

Přílohy

Seznam odborné literatury:

DOSTÁL, P. Pokročilé metody analýz a modelování v podnikatelství a veřejné správě, Brno : CERM, 2008. 340 s. ISBN 978-80-7204-605-8.

DOSTÁL, P; RAIS, Karel; SOJKA, Zdeněk. Pokročilé metody manažerského rozhodování, Praha : Grada, 2005. 166 s. ISBN 80-247-1338-1.

LACKO, L. Business Intelligence v SQL Serveru 2008, Brno : Computer Press, 2009. 456 s. ISBN 978-80-251-2887-9.

LACKO, L. Business Intelligence v SQL Serveru 2005, Brno : Computer Press, 2006. 391 s. ISBN 80-251-1110-5.

Vedoucí bakalářské práce: Ing. Jiří Kříž, Ph.D.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2010/2011.

L.S.

Ing. Jiří Kříž, Ph.D.
Ředitel ústavu

doc. RNDr. Anna Putnová, Ph.D., MBA
Děkan fakulty

V Brně, dne 26.05.2011

Abstrakt

Táto bakalárska práca je zameraná na analýzu firemných procesov v spoločnosti Acision s.r.o. . Je vybraná určitá skupina špecialistov alebo tímov, nad ktorými sú vykonávané analýzy výkonnosti, ktoré môžu poslúžiť ako dobrý základ pre ďalšie manažérske rozhodovanie. Práca je rozdelená na teoretickú a praktickú časť. Teoretická časť obsahuje popis metód, ktoré sa používajú a môžu byť aplikované v praxi resp. v spoločnosti Acision s.r.o. . Praktická časť obsahuje popis firemných procesov a analýzu firemných dát s využitím moderných metód (Business Intelligence - Data mining) a vlastný návrh, ako by sa dal aplikovať data mining nad týmito procesmi.

Kľúčové slová

Business Intelligence, Data mining, datové sklady, analýza procesov, manažérske rozhodovanie

Abstract

This bachelor thesis is focused on business process analysis in Acision ltd. There is the group of specialists or teams chosen, on which the performance analysis is being performed that can serve as the base for further managerial decision making. The thesis is divided into the theoretical and the practical part. The theoretical part contains description of methods that are used and are applied in practice and might be applicable in Acision ltd. The practical part contains description of business process and business data analysis where modern methods are being used (Business Intelligence – Data mining) and contains own design of how to make data mining applicable upon these process.

Key words

Business Intelligence, Data mining, data warehouses, process analysis, managerial decision making

Bibliografická citácia

BARTÍK, B. *Analýza firemních procesů a návrh optimalizace s využitím data miningu*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2011. 79 s.

Vedúci bakalárskej práce Ing. Jiří Kříž, Ph.D..

Čestné prehlásenie

Prehlasujem, že som celú bakalársku prácu spracoval samostatne na základe uvedenej literatúry a pod vedením svojho vedúceho bakalárskej práce. Prehlasujem, že citácie použitých prameňov sú úplné, a že som v práci neporušil autorské práva (v zmysle zákona č. 121/2000 Sb. o právu autorskom a o právach souvisejících s právem autorským).

V Brne dňa 3. júna 2011

.....
Branislav Bartík

Pod'akovanie

Rád by som chcel týmto poďakovať vedúcemu bakalárskej práce pánovi Ing. Jiřímu Křížovi, Ph.D. za odbornú pomoc, vecné rady a pripomienky pri spracovaní bakalárskej práce. Ďalej ďakujem spoločnosti Acision s.r.o. za poskytnutie materiálov a dát, ktoré poslúžili ako základ a zdroj informácií. Taktiež by som rád poďakoval pánu Mgr. Jiřímu A. Randusovi a Ing. Antonínovi Klanicovi za ústretovosť a podporu, ktorú mi prejavili pri písaní tejto práce.

Obsah

ÚVOD.....	9
CIEĽ PRÁCE.....	10
1. TEORETICKÉ VÝCHODISKÁ.....	11
1.1. ZÁKLADNÉ POJMY	11
1.2. BUSINESS INTELLIGENCE.....	11
1.2.1. <i>Vhodná databáza pre Business Intelligence</i>	13
1.2.2. <i>Dátový sklad</i>	15
1.3. DATA MINING	17
1.3.1. <i>Procesy v data miningu</i>	17
1.4. ALGORITMY PRE DATA MINING.....	19
1.4.1. <i>Rozhodovacie stromy</i>	19
1.4.2. <i>Zhlukovanie</i>	21
1.4.3. <i>Sekvenčné zhlukovanie</i>	22
1.4.4. <i>Asociačné pravidlá</i>	22
1.4.5. <i>Časové rady</i>	22
1.4.6. <i>Neurónové siete</i>	24
1.4.7. <i>Naive Bayes</i>	25
1.5. TYPICKÉ OKRUHY ÚLOH A VÝBER ALGORITMOV PRE ICH RIEŠENIE	25
1.5.1. <i>Klasifikácia</i>	26
1.5.2. <i>Regresia</i>	26
1.5.3. <i>Segmentácia</i>	27
1.5.4. <i>Priradovanie</i>	27
1.5.5. <i>Analýzy textu</i>	27
2. ANALÝZA PROBLÉMU A SÚČASNEJ SITUÁCIE.....	29
2.1. ZÁKLADNÉ ÚDAJE O SPOLOČNOSTI.....	29
2.2. PRODUKTY A RIEŠENIA SPOLOČNOSTI ACISION	30
2.3. POPIS POZÍCIÍ V SPOLOČNOSTI.....	32
2.3.1. <i>Customer Services Engineer (CSE)</i>	33
2.3.2. <i>Sustaining Engineer (SE)</i>	35
2.3.3. <i>Research & Development Engineer (R&D)</i>	36
2.4. POPIS FIREMNÝCH PROCESOV	38
2.4.1. <i>Riešenie problému na strane supportu (CSE)</i>	40
2.4.2. <i>Riešenie problému na strane sustainingu (SE)</i>	41
2.4.3. <i>Riešenie problému na strane R&D</i>	42
2.5. ACTION REQUEST SYSTEM	44
2.6. FIREMNÉ RIEŠENIE ANALÝZY NAD PROCESMI	51
2.6.1. <i>Databáza v spoločnosti Acision</i>	51
2.6.2. <i>Webová interpretácia informácií</i>	55
2.7. ZHODNOTENIE SÚČASNÉHO STAVU	59
3. NÁVRH VLASTNÉHO RIEŠENIA.....	61
3.1. DEFINÍCIA PROBLÉMU.....	62
3.2. PRÍPRAVA DÁT.....	62
3.3. VÝBER ALGORITMU A MODELU	66
3.4. UČIACA FÁZA APLIKOVANÁ NA DOSIAĽ EXISTUJÚCICH PRÍPADOCH.....	69
3.5. TESTOVANIE MODELU, ANALÝZA A PREDIKCIA NOVÝCH PRÍPADOV	72
ZÁVER	76
ZOZNAM POUŽITEJ LITERATÚRY	77
ZOZNAM OBRÁZKOV, GRAFOV, TABULIEK	79

Úvod

Každá moderná firma, či už pôsobí v oblasti výroby alebo služieb, má určitú štruktúru a funguje na základe určitých procesov, ktorými sa riadia zamestnanci tejto spoločnosti. Na to, aby bola firma úspešná, potrebuje, aby jej zamestnanci pracovali s čo najvyššou efektivitou, preto sa musia viesť určité záznamy o výkone daného zamestnanca a následne aj celého tímu. Na to slúžia špeciálne aplikácie i reportovacie systémy. Postupnou analýzou takýchto dát sa zisťuje aktuálna situácia, ale nie len to. Štatistiky, grafy, prognózy, tvoria základ pre ďalšie rozhodnutia vedúcich pracovníkov, manažmentu popřípade top manažmentu.

Spoločnosť Acision Czech Republic s.r.o., ktorá podniká v oblasti informačných technológií a mobile messagingu, poskytuje svoje služby a riešenia už pre viac ako tristo mobilných operátorov po celom svete. Svojim zákazníkom, v jej prípade mobilným operátorom, dodáva softwarové produkty a okrem ich predaja musí zabezpečovať i neustálu podporu. Pre každý z týchto produktov sú vytvorené tímy zamestnancov, ktorí sú špecialisti a takúto podporu musia zabezpečiť.

Primárnou úlohou týchto špecialistov je problém kompletne zanalyzovať, popr. vyžiadať si potrebné dáta k ďalšej analýze, a v konečnom dôsledku problém vyriešiť v čo najkratšom čase. Je tu vzájomné prepojenie tímov a pri riešení problémov tímy spolupracujú. Bližšie role jednotlivých tímov budú popísané v tejto práci. Pre zákazníka je však najdôležitejší čas, za ktorý sa obnoví kompletná funkcionálna daného systému.

Sekundárnou úlohou je daný problém zaznamenať, a vykonávať všetky záznamy o výkone daného špecialistu, ktoré zahrňujú analýzy, komunikáciu medzi tímami, všetky dokumenty a dáta použité pri riešení problému a všetko, čo s daným incidentom súvisí, to znamená jeho stav, prioritu, časy a v závere popis riešenia problému, v niektorých prípadoch i príčinu.

V tejto práci sa zameriame na popis tejto sekundárnej úlohy. Bude tu popísaná štruktúra, role, ticketovací systém (použitý *Action Request System* od spoločnosti *BMC Software*). Za použitia moderných metód *Business Intelligence - Data Miningu* a následnou analýzou za použitia vhodných algoritmov sa v závere pokúsime zhrnúť výsledky, vyhodnotiť a navrhnúť model pre optimalizáciu procesov ktoré sú v danom tíme žiaduce i nežiaduce.

Cieľ práce

Cieľom tejto bakalárskej práce je popísať princíp fungovania firemných procesov a zhodnotiť aktuálny stav výkonnosti vybranej skupiny špecialistov s využitím nástrojov Business Intelligence, predovšetkým data miningu a následne použiť získané poznatky. Za pomoci vhodných algoritmov sa následne pokúsiť tieto poznatky využiť v návrhu na zlepšenie fungovania firemných procesov, ktorý takisto môže slúžiť i pri podpore rozhodovania manažmentu, poprípade top manažmentu spoločnosti Acision Czech Republic s.r.o.

1. TEORETICKÉ VÝCHODISKÁ

1.1. Základné pojmy

Business Intelligence ako proces premeny údajov na informácie a poznatky je jedným z najdynamickejších a najviac rastúcich odvetví informačných technológií. *Data mining*, ako jeden z jeho podprocesov, bol použitý k samotnej analýze nad procesmi v tejto spoločnosti a takisto s ním súvisiace algoritmy. *Datový sklad* dopĺňa trojicu pojmov, ktorými sa budeme zaoberať v nasledujúcich podkapitolách venovaných teoretickým východiskám tejto práce.

1.2. Business Intelligence

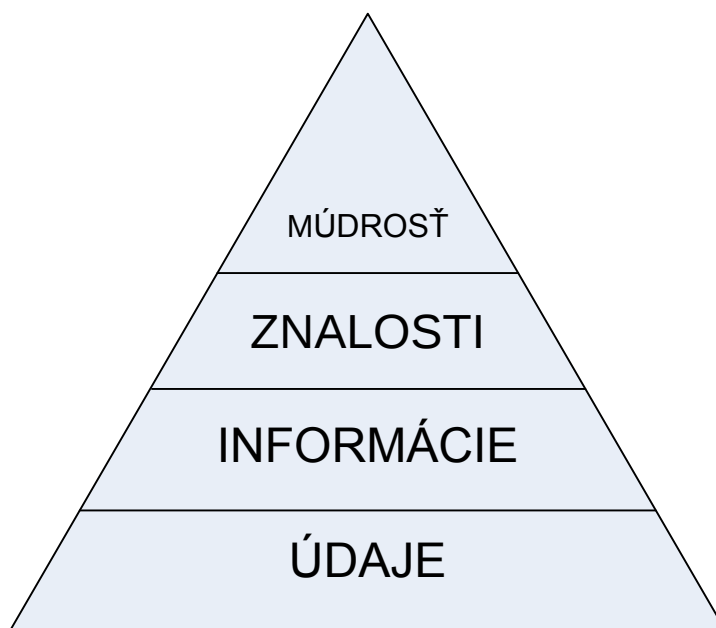
Termín Business Intelligence (BI) poprvýkrát definoval Howard Dresner zo spoločnosti Gartner Group takto:

Business Intelligence je množina konceptov a metodík, ktoré zlepšujú rozhodovací proces za použitia metrík, alebo systémov založených na metrikách. Účelom procesu je konvertovať veľké objemy dát na poznatky, ktoré sú potrebné pre koncových užívateľov. Tieto poznatky potom môžeme efektívne použiť napríklad v procese rozhodovania a môžu tvoriť veľmi významnú konkurenčnú výhodu.

Rýchlo sa meniaci trh, konkurencia, potreby zákazníkov, ponuky dodávateľov, to všetko sú faktory, ktoré si vyžadujú prijímanie rozhodnutí takmer okamžite. Predpokladom pre kvalitné rozhodovanie sú však kvalitné informácie, ktoré sú k dispozícii v správny čas, na správnom mieste a v požadovanej podobe. Aby mali manažéri pre svoje rozhodnutie k dispozícii všetky potrebné informácie, musia v organizácii fungovať procesy, ktoré zabezpečujú poskytovanie informácií pre podporu rozhodovania.

Proces transformácie dát na informácie a prevod týchto informácií na poznatky prostredníctvom objavovania nazývame Business Intelligence. Premenu dát na informácie, informácie na znalosti a budovanie „múdrosti“ na základe znalostí môžeme zobrazit' na hierarchickej pyramíde informačných úrovní (Obr. 1). Základom všetkého

sú dáta. Dáta obsahujú len jednoduché fakty, pri čom sa samozrejme predpokladá, že niekde vo vnútri množiny dát sú ukryté určité informácie. Tieto informácie však vystúpia na povrch až vtedy, keď k dátam pridáme znalosti. Keď do hry vstúpia okrem informácií i tvorivá inteligencia, získame znalosti. Pokiaľ tieto znalosti zobecníme, získame „múdrosť“, to znamená schopnosť presného zhodnotenia znalostí a ich následné uplatnenie v reálnej praxi. (3)



Obr. 1: Hierarchia informačných úrovní (3)

Často potrebujeme sledovať trend nejakej veličiny, napríklad pri obchodovaní s cennými papiermi, alebo potrebujeme nájsť medzi údajmi určité závislosti. Preto moderné databázové servery obsahujú rozsiahlu podporu pre budovanie *datových skladov* (anglicky *data warehouse*), OLAP (*On-line Analytical Processing*) analýzy a *data mining* (odkrývanie, dolovanie dát). Podľa rôznych prieskumov až 10 percent užívateľov používa výsledky analýz, 15-25 percent užívateľov tieto informácie skúma a hľadá v nich súvislosti. Najväčšia skupina užívateľov informácie používa vo forme rôznych výpisov a reportov. Na podnikovej úrovni sa generujú rôzne druhy reportov, napríklad pre obchodné oddelenie, finančné oddelenie, oddelenie ľudských zdrojov, vo sfére CRM¹ a podobne. Dáta sú buď v podnikových databázach, alebo v dátových

¹ Customer Relationship Management (riadenie vzťahu so zákazníkmi)

skladoch. Výhodou je, že dáta sú už predspracované a pretransformované v etape ETL² a prenesené z produkčných systémov do dátových skladov, prípadne dátových trhov (anglicky *data marts*). Reporty z reportovacích služieb potom vhodne dopĺňujú dáta z analytických *Business Intelligence* aplikácií. Nasadenie reportovacích služieb je v tomto prípade prevážne na úrovni podnikových portálov, takže koncoví užívatelia k ním prístupujú v rámci podnikového intranetu. (3)

1.2.1. **Vhodná databáza pre Business Intelligence**

Pokiaľ sa firma rozhodne pre centralizované koncipované BI riešenie, prvým krokom je voľba databázovej platformy. Tá sa volí podľa niekoľkých kritérií, ale určite najzákladnejší filter by mal byť nastavený. Pokiaľ chceme databázu pre inteligentné riešenie, i samotná databáza by nejakú inteligenciu mala mať. Inými slovami, nemali by sme si zavádzať databázu, ktorá nám umožní dáta iba ukladať, pokiaľ sú naše ciele ďalej a chceme s dátami pracovať. Nasledujúce hľadiská dávajú prehľad o tom, ako sa treba pozerať na databázu s perspektívou BI platformy:

1. Všetko v jednom pre BI - v ideálnom prípade treba nájsť databázu, ktorá bude zároveň efektívnym nástrojom pre dátovú transformáciu v heterogénnych prostrediach, plnohodnotným analytickým serverom poskytujúci i služby data miningu, nástrojom zaisťujúcim celý životný cyklus reportu.
2. Analytické služby databáze - databáza by mala pracovať s jedným, nedeliteľným modelom pre relačnú a multidimenzionálnu databázu, ktorý zaručí vždy pravdivé odpovede. Zároveň musí vedieť pracovať s nastavením pravidiel pre *KPI (key performance indicator)*. Dôležitá je podpora *real-time* BI - všetky zmeny v primárnych systémoch sú okamžite zohľadnené vo všetkých reportoch a analýzach.
3. Reporting - pokiaľ hovoríme o podpore celého životného cyklu reportu, potom sa počíta jeho návrh, definícia pravidiel komu, kedy a akou formou a výsledná distribúcia reportu. Riešenie reportingu musí byť pripravené pre náročnejších zákazníkov s veľkým počtom prístupov k reportom.

² Extract, Transform and Load (extrakcia, transformácia, zavedenie)

4. Dáta odkiaľkoľvek, tzn. integrácia - databáza musí byť platformou pre budovanie vysokodostupných riešení pre dátovú integráciu, workflow a pre extrakciu, transformáciu a zavedenie dát (ELT) do dátových skladov. Súčasťou pokročilých platforiem sú pripravené funkcie pre transformáciu dát z najrôznejších dátových zdrojov.
5. Bezpečnosť a dostupnosť dát - dáta v analýzach bývajú často to najrýdzejšie zlato firmy a ich strata by mohla výrazne ovplyvniť ďalšiu existenciu takéhoto podniku. Preto je treba dbať na samotnú spoľahlivosť DB platforiem. Napr. uloženie dát na viac oddelených serveroch zaručí dostupnosť i pri úplnej skaze, potom ale obnovenie dát musí prebiehať za plnej prevádzky, bez nutnosti výpadku. Identifikácia a autorizácia užívateľov by mala spĺňať najvyššie požiadavky na zabezpečenie, pomocou proti zneužitiu môže byť i šifrovanie databázových súborov.
6. Plná kompatibilita s ďalšími celkami - môžeme vyvinúť sebe lepšiu analytickú platformu, pokiaľ však celé riešenie a vlastné informácie zneprístupníme relevantným ľuďom, naša práca bola zbytočná. Takže nami zvolená databáza by mala umožňovať integráciu informácií do bežného kancelárskeho prostredia alebo bežného internetového okna v rámci firemného intranetu. Tu platí, čím jednoduchší prostriedok pre interpretáciu informácií to bude, tým úspešnejšie bude riešenie BI ako celok prijaté. (8)

1.2.2. Dátový sklad

Asi najznámejšia definícia dátového skladu pochádza od Billa Inmona:

„Dátový sklad je podnikovo štrukturovaný depozitár subjektovo orientovaných, integrovaných, časovo premenných, historických dát použitých pre získavanie informácií a podporu rozhodovania. V dátovom sklade sú uložené atomické i sumárne dáta.“

Dáta sú získavané a ukladajú sa do produkčných (operačných) databází, ktoré môžu byť na rôznych oddeleniach firmy, alebo dokonca v rozličných geografických lokalitách. Tieto dáta v pravidelných intervaloch zozbierame, spracujeme a zavedieme do dátového skladu. Dátový sklad je v podstate taktiež databáza, len je organizovaná v podľa trochu iných pravidiel. Tabuľky napríklad nemusia byť normalizované a podobne. Dátový sklad je teda súbor technológií pre efektívne skladovanie dát takých, aby tieto dáta po ich premene na informácie slúžili pre podporu rozhodovania. (3)

V nasledujúcej tabuľke sú zobrazené hlavné rozdiely medzi produkčnou databázou a dátovým sklado.

Tab. 1: Rozdiely medzi produkčnou databázou a dátovým sklado (3)

Vlastnosť	Produkčná databáza	Dátový sklad
Čas odozvy	Zlomky sekúnd až sekundy	Sekundy až hodiny
Operácie	DML (data manipulation language)	Primárne iba pre čítanie
Pôvod dát	30-60 dní	Série snímkov za časový úsek
Organizácia dát	Podľa aplikácie	Podľa predmetu, času...
Veľkosť	Malá až veľká	Veľká až veľmi veľká
Zdroje dát	Operačné, interné	Operačné, interné, externé
Činnosti	Procesy	Analýza

Definícia dátového skladu podľa Billa Inmona je síce stručná, ale za to výstižná. Pojmy, ktoré túto definíciu tvoria sú: subjektová orientácia, integrovanosť, časová variabilita a nemennosť.

- **Subjektová orientácia:** dáta sa do dátového skladu zapisujú väčšinou podľa predmetu záujmu, než podľa aplikácie, v ktorej boli vytvorené. Pri orientácii na subjekt sú dáta v dátovom sklade kategorizované podľa subjektu, ktorým môže byť napr. zákazník, dodávateľ, zamestnanec, výrobok a podobne. Orientácia na aplikáciu oproti tomu znamená, že dáta sú v systéme uložené podľa jednotlivých aplikácií, napr. dáta aplikácie pre odbyť, dáta aplikácie pre fakturáciu, dáta aplikácie pre personalistiku.
- **Integrovanosť:** dátový sklad musí byť jednotný a integrovaný. To znamená, že dáta týkajúce sa konkrétneho predmetu sa do dátového skladu ukladajú len raz. Preto musíme zaviesť jednotnú terminológiu, jednotné a konzistentné jednotky veličín. Nie je to jednoduchá úloha, pretože dáta prichádzajú do dátového skladu z nekonzistentného a neintegrovaného operačného prostredia. Preto musia byť dáta v etape prípravy a zavedenia upravené, vyčistené a zjednotené. Pokiaľ dáta nie sú konzistentné a dôveryhodné, dátový sklad stráca význam.
- **Časová variabilita:** dáta sa ukladajú do dátového skladu ako séria snímok, z ktorých každý reprezentuje určitý časový úsek. Na rozdiel od operačného prostredia, kde sú dáta platné v okamihu prístupu, v dátových skladoch sú dáta platné pre určitý časový moment, časový snímok. Zatiaľ čo v operačnom databázovom prostredí sú uložené dáta za kratšie časové obdobie, väčšinou niekoľko dní, maximálne mesiacov, v dátovom sklade sú dáta za dlhšie časové obdobie, typicky za niekoľko rokov. Kľúčové atribúty v dátovom sklade obsahujú čas, ktorý v operačných databázach nemusí byť uvádzaný. Akonáhle je v dátovom sklade zaznamenaný konkrétny snímok dát z operatívnej databáze, nemôžu už byť tieto dáta v dátovom sklade modifikované.
- **Nemennosť:** v operačných transakčných databázach sú dáta do databáze jednak vkladané, jednak modifikované a zmazávané. Dáta v dátovom sklade sa obvykle nemenia ani neodstraňujú, iba sú v pravidelných intervaloch pridávané nové dáta. Preto je manipulácia s dátami v dátových skladoch ďaleko jednoduchšia. V zásade môžeme pripustiť len dva typy operácií. Zavedenie dát do dátového skladu a prístup k týmto dátam. Žiadne zmeny dát nie sú prípustné. Z toho vyplýva že väčšina metód pre optimalizáciu a normalizáciu dát a transakčný prístup k dátam je v dátovom sklade nepotrebná. (3)

1.3.Data mining

Data mining, alebo aj dolovanie z dát je odborom, ktorý zastrešuje širokú škálu techník používaných v rade odvetví. Pretože dochádza k zostreniu konkurencie v boji o zisky a tržných podielov, stal sa tento odbor nevyhnutným k udržaniu konkurenčnej schopnosti. Tento odbor sa začal rozvíjať počiatkom 90. rokov dvadsiateho storočia a to za účelom vylepšenia riadenia rizík u bánk špecializovaných na kreditné karty. K rozšíreniu došlo v ďalších odvetviach ako je poisťovníctvo, verejné služby, zásielkové služby, energetika, maloobchod, zábavný priemysel atď. (1)

Pojem datamining alebo data mining (niekedy sa technológia dolovania dát označuje i skratkou KDD – *Knowledge discovery in databases*), je inými slovami proces analýzy dát z rôznych perspektív a ich premena na užitočné informácie. Z matematického a štatistického hľadiska ide o hľadanie korelácií, teda vzájomných vzťahov alebo vzorov v dátach. Data mining je proces, ktorého cieľom je ťažba informácií v databázach. Využíva štatistické metódy a ďalšie metódy hraničiace s oblasťou umelej inteligencie. (3)

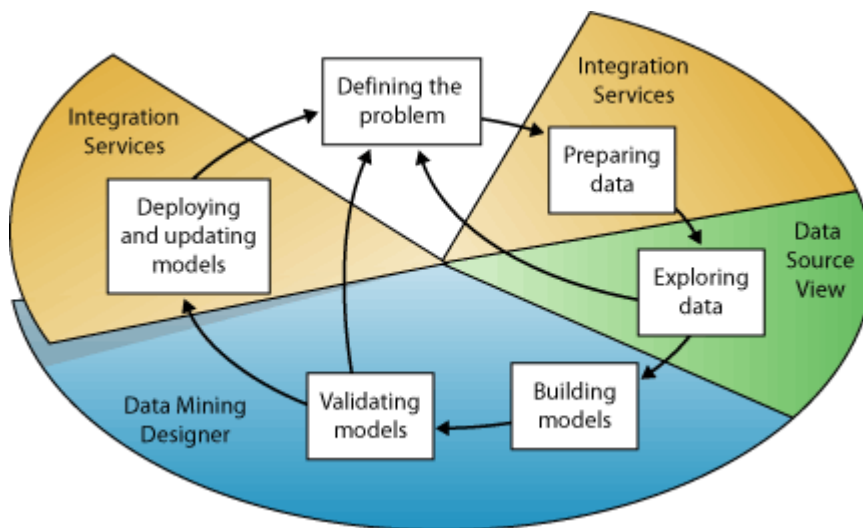
Všetky procesy a algoritmy, ktoré možno použiť v data miningu, sú popísané v nasledujúcich podkapitolách (popisované algoritmy sú súčasťou platformy SQL Server 2008, na ktorej bola táto práca realizovaná).

1.3.1. Procesy v data miningu

Data mining sa skladá z viacerých procesov, nakoľko ich musíme nad dátami vykonať hneď niekoľko a presne dodržanom poradí. Proces dataminingu môžeme rozdeliť do niekoľkých etáp (schéma na Obr. 2):

1. Definícia problému.
2. Príprava dát.
3. Výber algoritmu a modelu.
4. Učiaci fáza aplikovaná na doposiaľ existujúcich prípadoch.
5. Testovanie modelu.
6. Analýza a predikcia nových prípadov.

Inak povedané, na základe prieskumu dát sú v nich vyhľadávané určité vzory a na základe týchto vzorov sú spracované modely pre predikciu budúcich prípadov. Vzory sa samozrejme podarí odhaliť iba vtedy, pokiaľ nejaké existujú. Preto napríklad nemá význam aplikovať data mining nad rôznymi cvičiacimi databázami obsahujúce náhodne generované dáta.



Obr. 2: Procesná schéma data miningu na platforme SQL Server 2008 (3)

V učiacej sa fázi sa vybraný data miningový model učí a spresňuje svoje parametre na množine dát získaných z doposiaľ existujúcich prípadov. Ako podklad pre učiacu fázu slúžia doposiaľ zozbierané (namerané) a vyhodnotené dáta. Výberom a prípravou dát, ktoré budú slúžiť ako podklad pre učiacu sa fázu, musíme venovať veľkú pozornosť, aby dáta boli dostatočne reprezentatívne a očistené od prípadných chýb. Najvhodnejšie je pripraviť a predspracovať tieto dáta v procese ETL. Učiacu fázu je veľmi náročná na výber správneho modelu a taktiež na výber a predspracovanie dát.

V etape analýzy a predikcie aplikujeme už naučený data miningový model na množinu vstupných dát, z ktorých potrebujeme získať súvislosti. Na rozdiel od učiacej fáze je táto fáza náročná skôr na výpočtovú kapacitu použitého hardwaru a softwaru. (3)

1.4.Algoritmy pre data mining

Data mining je proces analýzy dát z rôznych perspektív a ich premena na užitočné informácie. Z matematického a štatistického hľadiska ide o hľadanie korelácií, teda vzájomných vzťahov alebo vzorov v dátach. SQL Server 2008 má implementované tieto algoritmy:

1. Rozpadové stromy.
2. Zhlukovanie.
3. Sekvenčné zhlukovanie.
4. Asociačné pravidlá.
5. Časové rady.
6. Neurónové siete.
7. Naive Bayes.
8. Lineárna regresia.
9. Logistická regresia.

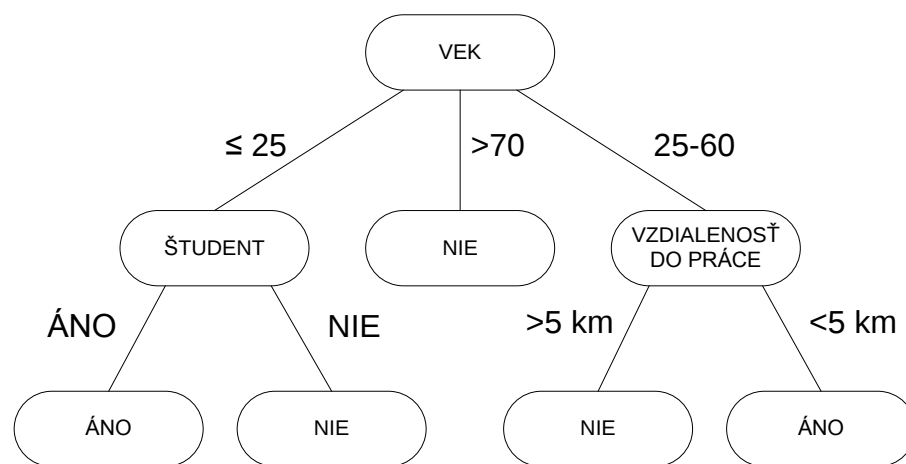
Jednoznačné doporučenie pre výber najvhodnejšieho algoritmu pre tú ktorú úlohu neexistuje, rozhodnutie závisí na konkrétnom prípade. Pre uľahčenie orientácie popíšeme jednotlivé algorimy, a potom priradíme vhodné algorimy k jednotlivým typom úloh. Pri popise algoritmov uvádzame aj zoznam parametrov, pomocou ktorých špecifikujeme úlohu, ktorú chceme pomocou algoritmu riešiť. (3)

1.4.1. Rozhodovacie stromy

V pôvodnej terminológii sa tento algoritmus nazýva *Microsoft Decision Tree Algorithm*. Tento typ diagramov odhaduje závislosti a vyhľadáva špecifické vlastnosti, ktoré potom poslúžia pre zostavovanie predikčného modelu rozhodovania na jednotlivých úrovniach hierarchickej štruktúry. Hlavnou výhodou je zrozumiteľnosť, prehľadnosť a jednoduchá interpretácia, ktorá umožňuje užívateľom rýchlo a ľahko vyhodnocovať získané výsledky, identifikovať kľúčové položky a vyhľadávať zaujímavé segmenty prípadov. Pretože sa jedná o hierarchickú stromovú štruktúru, tento

algoritmus je veľmi rýchly. Cieľom algoritmu je identifikovať entity popísané rôznymi atribútmi a rozdeliť ich do tried. Rozhodovací strom sa vytvorí z množiny daných objektov, ktoré sú zaradované do skupín.

Každý uzol stromu predstavuje určitú vlastnosť entít. Z tohoto uzlu vedie konečný počet hrán. Preto je nutné vlastnosti najprv diskretizovať z reálnych čísel do konečného počtu intervalov. Na Obr. 3 je príklad vytvorenia uzlu rozhodovacieho stromu z databáze zákazníkov, kde sa skúma, či si zákazník príslušný produkt kúpi, alebo nekúpi. Jeden z rozpadových uzlov stromu je vytvorený na základe veku zákazníka.



Obr. 3: Jednoduchý rozpadový strom posudzujúci, či si zákazník kúpi bicykel (3)

V nevyváženom rozpadovom strome sú dáta rozdeľované do vetví bez straty. To znamená, že počet záznamov splňujúcich túto podmienku rodičovského uzlu sa rovná súčtu záznamov splňujúcich podmienku potomkov. Keby sme teda postupovali smerom dolu, počet záznamov, ktoré vyhovovali podmienkam cesty v strome, sa znižuje.

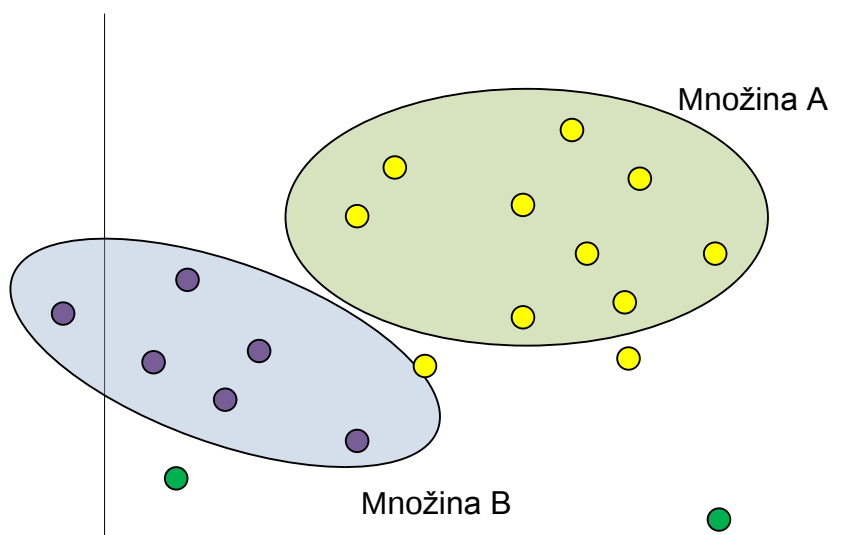
Nevyvážený rozpadový strom rozdeľuje N dimenzionálny priestor na oblasti s možnými riešeniami. Počet dimenzií je určený počtom atribútov použitých v strome.

V existujúcom strome sa požadovaná hodnota predpovedá na základe známych atribútov postupne od koreňa smerom dole. Pomocou atribútov je volená najpravdepodobnejšia cesta k výsledku. Tento proces je cyklický a vyžaduje jednoduchú rozhodovaciu podmienku. Pri použití tohoto atribútu je potrebné zo skúsenosti odhadnúť optimálnu veľkosť množiny testovacích dát. Keď je príliš malá, strom nemusí byť správne špecifický, keď je príliš veľká, strom môže byť preučný. Algoritmus podporuje predikciu diskretných i spojitých atribútov. (3)

1.4.2. Zhlukovanie

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Clustering Algorithm*. Používa sa pre odhaľovanie zhlukov dát. Je to proces organizovania objektov do skupín, pričom objekty v skupinách sú nejakým spôsobom podobné. Algoritmus vyhľadáva prirodzené zhluky (klastry) dát. Kritériá zhlukovania nie sú predom dané, ale odhaľujú sa z prirodzenej štruktúry dát. Inštancie v rámci zhuku sú si určitým spôsobom podobné. Jednotlivé zhluky sa môžu a nemusia prekrývať, to znamená, že jedna inštancia môže patriť do jedného alebo viacerých zhlukov. Pomocou algoritmu je možné rozdeliť množiny prípadov na čo najohraničenejšie skupiny, tzv. ostrovy podobnosti. Algoritmus je vhodný napríklad pri identifikácii zákazníckych segmentov, ktoré sú založené na spoločných charakteristikách, napr. demografických, sociálnych, profesných a podobne.

Zhluky sú odhaľované na základe aplikovania a analýzy kriviek pravdepodobnosti, pričom sa skúma, či jednotlivé dáta, alebo skupiny dát nespĺňujú podmienku napr. Gaussovho normálneho rozdelenia a podobne. (3)



Obr. 4: Princíp odhaľovania zhlukov prípadov (3)

1.4.3. **Sekvenčné zhlukovanie**

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Sequence Clustering Algorithm*. Špecifickým príkladom zhlukovania je sekvenčné zhlukovanie. Z matematického hľadiska sa jedná o aplikáciu Markových procesov a modelov. (3)

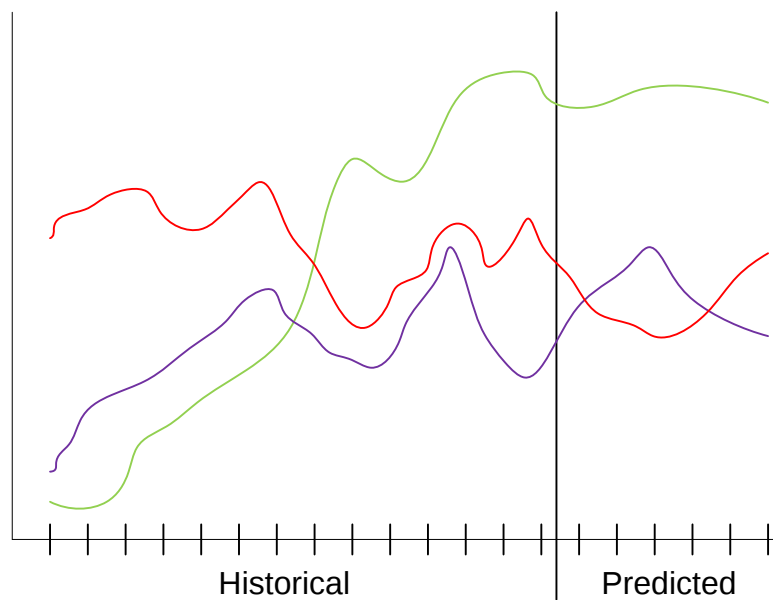
1.4.4. **Asociačné pravidlá**

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Association Algorithm*. Analýza z hľadiska asociačných pravidiel je zameraná na odkrývaní vzťahov v dátach a odhaľovania rôznych súvislostí priradovania. Najčastejšie sa pravdepodobne používa pri analýze nákupného košíku, teda hľadaní odpovede na otázku, aký tovar si zákazníci kupujú najčastejšie spoločne. Nájdenie týchto súvislostí umožní efektívnejšie rozmiestnenie tovaru v supermarketoch, prípadne zostavenie akčných ponúk. Z hľadiska matematickej štatistiky sa jedná o skúmanie korelácie, či už pozitívnej, alebo negatívnej. Pozitívna korelácia udáva, že vysoká úroveň jednej premennej bude sprevádzaná vysokou úrovňou korelačnej premennej. Naopak negatívna korelácia udáva, že vysoká úroveň jednej premennej bude sprevádzaná nízkou úrovňou korelačnej premennej. Nájdenie pozitívnej korelácie je dôležité napr. pri rozmiestňovaní tovaru alebo marketingových rozhodnutí, ktorý tovar je možné predávať spoločne, prípadne čo je potrebné ktorému zákazníkovi ponúknuť. (3)

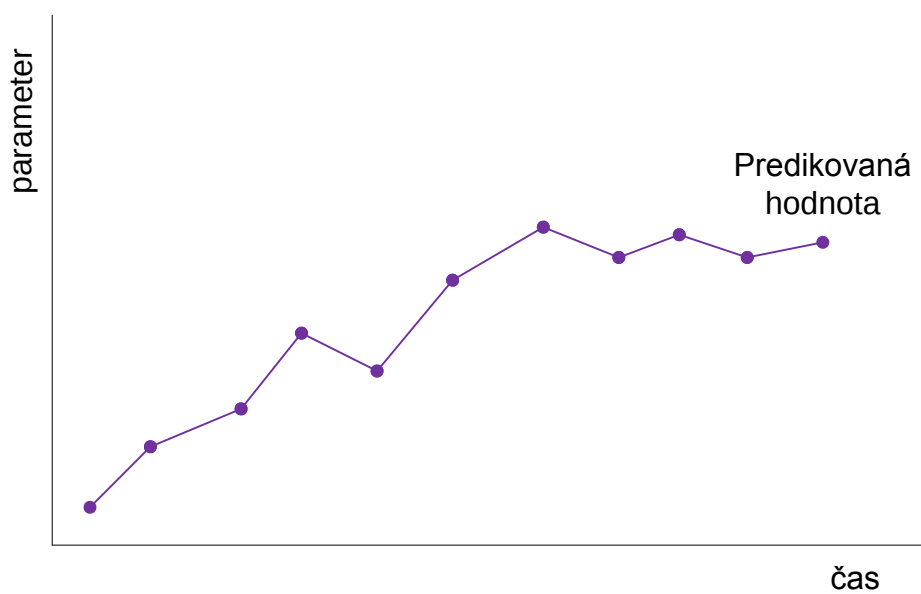
1.4.5. **Časové rady**

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Time Series Algorithm*. Algoritmus analýzy časových radov sa spravidla vzťahuje k nejakej premennej, napr. obratu, zisku, nákladom a pod. Na základe analýzy dát z minulosti a súčasnosti je možné definovať určité pravidlá, pomocou ktorých potom predpovedáme budúci trend v sebe zahrňujúci i krátkodobé cyklické fluktuácie. Napr. u predikcie trendu vývoja nezamestnanosti pre budúci rok na základe dát z minulých rokov sa

v dobrej predpovedi určite prejavia i logické fluktuácie, napr. nárast nezamestnanosti po ukončení školskej dochádzky, jej pokles pri sezónnych prácach a podobne. (3)



Obr. 5: Časové rady (3)



Obr. 6: Predikovaná hodnota na základe priebehu veličiny v minulosti (3)

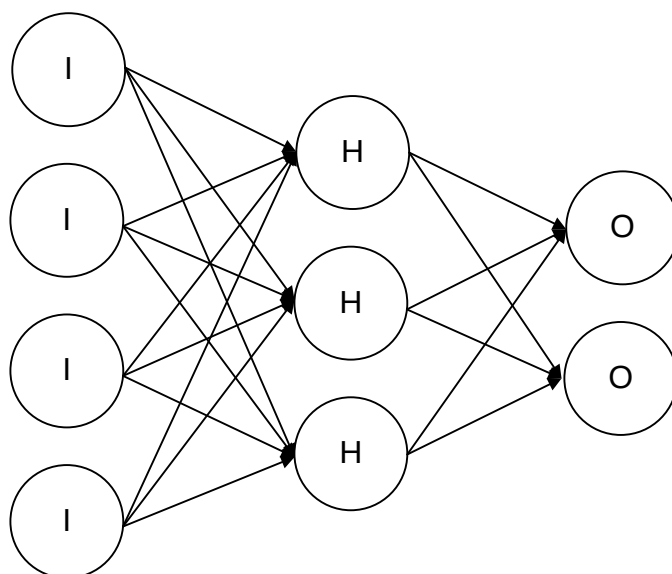
1.4.6. Neurónové siete

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Neural Network Algorithm*. Spracovanie pomocou neurónových sietí nevychádza zo žiadneho štatistického rozdelenia, ale pracuje podobne ako ľudský mozog, na princípe rozpoznania vzorov a minimalizácie chýb. Tento proces si môžeme predstaviť ako prijímanie informácií a poučenie sa z každej skúsenosti. Neurónovú sieť tvoria uzly usporiadané do vrstiev. Predtým, než sa začne vlastný proces, rozdelia sa údaje do trénovacej a testovacej množiny. Behom každej iterácie sú výstupy spracované systémom a sú porovnané so skutočnou hodnotou. Zmeria sa chyba a odovzdá sa k spracovaniu systému, aby upravil pôvodné váhy. Proces končí väčšinou v momente dosiahnutia predom určenej minimálnej chyby. Schéma neurónovej siete s jednou skrytou vrstvou obsahuje tri typy uzlov: vstupné uzly, uzly skrytej vrstvy, výstupné uzly. (3)

Na schéme (Obr. 7) jednoduchkej neurónovej siete už na prvý pohľad vidíme veľkú výhodu tejto architektúry. Dáta sú totiž v jednotlivých vrstvách spracovávané paralelne s mnohými neurónami. Schéma neurónovej siete s jednou skrytou vrstvou obsahuje tri typy uzlov:

- I – vstupné uzly.
- H – uzly skrytej vrstvy.
- O – výstupné uzly.

Algoritmus síce podporuje predikciu diskretných i spojitých atribútov, ale najvýhodnejšie je použiť len diskrétne alebo diskretizované atribúty.



Obr. 7: Schéma neurónovej siete s jednou skrytou vrstvou (3)

1.4.7. Naïve Bayes

V originálnej terminológii sa tento algoritmus nazýva *Microsoft Naïve Bayes Algorithm*. Je to veľmi rýchly a pritom pomerne presný algoritmus založený na Bayesovej vete, pomocou ktorej je možné pracovať priamo s pravdepodobnosťami. Tento algoritmus používa matematické pravidlo vysvetľujúce, ako sa má zmeniť existujúca domienka pri konfrontácii s novými skutočnosťami. Tento algoritmus kombinuje nové dáta s predchádzajúcimi znalosťami. Používa sa pre zložitejšie analýzy, napr. pre výpočet paralelnej korelácie a podobne. Algoritmus podporuje len diskkrétne alebo diskretizované atribúty. (3)

1.5. Typické okruhy úloh a výber algoritmov pre ich riešenie

Výber algoritmov pre riešenie typických okruhov úloh nie je jednoduchý a už nie jednoznačný. Vymenovanie a stručný popis algoritmov nám nie vždy pomôže pri základnej orientácii, ktorý algoritmus sa hodí k akému typu úloh. Vymenujeme teda typické úlohy a k nim priradíme vhodné algoritmy. Pre väčšinu úloh je možné použiť i viacero druhov algoritmov.

1.5.1. **Klasifikácia**

Do tejto skupiny úloh patria odpovede na otázky typu „Aké typy kampaní a členských kariet má ponúkať obchodná firma?“, „Aký druh záujemcov osloví reklamná a marketingová kampaň?“, „Prečo firma stráca zákazníkov?“, „Bude zamýšľaný produkt úspešný?“ a podobne. Klasifikácia sa dá použiť pri testovaní a riadení i pre detekciu chýb v aplikáciách. Proces klasifikácie má dve fázy. V prvej sa na základe trénovacej množiny vytvorí model a tento sa potom aplikuje na klasifikovanie dát, ktoré neboli v trénovacej množine. Modely by mali byť čo najjednoduchšie, čo zvyšuje ich zrozumiteľnosť a znižuje výpočtovú náročnosť.

Najvhodnejšie algoritmy: rozhodovacie stromy, Naive Bayes, neurónové siete.

Použiteľné algoritmy: zhľukovanie, sekvenčné zhľukovanie. (3)

1.5.2. **Regresia**

Zo štatistického hľadiska ide o metódu, ktorá kvantifikuje závislosť medzi dvomi spojitými premennými: závislou premennou ktorú je potrebné predikovať, a nezávislou, teda prediktívnou premennou. Cieľom procesu je nájsť priamku prechádzajúcu jednotlivými bodmi, pre ktorú platí, že súčet druhých mocnín odchýlok od každého bodu je minimálny. Keď je vzťah medzi závislou a nezávislou premennou nelineárny, spravidla je potrebné transformovať prediktívnu premennú tak, aby umožnila nájsť lepšie preloženie. Logická regresia je veľmi podobná lineárnej regresii, s tým rozdielom, že závislá premenná nie je spojitá, ale diskretná. Preto ju transformujeme spojitou hodnotou, ktorá je funkciou pravdepodobnosti výskytu udalostí. Regresiu je možné použiť k predikovaní výsledku dvoch, alebo viac úrovní, napr. prečo nastane jav nesplácanie pôžičiek, akú tržbu je možno očakávať od ktorého zákazníka a podobne. (3)

Najvhodnejšie algoritmy: rozhodovacie stromy, neurónové siete.

Použiteľné algoritmy: zhľukovanie, sekvenčné zhľukovanie.

1.5.3. **Segmentácia**

Rozdelenie nejakej množiny entít, napr. zákazníkov do rôznych skupín a segmentov podľa určitých kritérií, umožňuje dosiahnuť cielený personalizovaný prístup ku každému segmentu. Iným okruhom využitia segmentácie je organizácia dát tak, aby boli použiteľné s čo najvyššou efektivitou. (3)

Najvhodnejšie algoritmy: zhľukovanie, sekvenčné zhľukovanie.

Použiteľné algoritmy: neurónové siete.

1.5.4. **Prirad'ovanie**

Okruh úloh pre prirad'ovanie zahŕňa typické obchodnícke otázky, napr. aké typické produkty sa najčastejšie kupujú spoločne s výrobkom, ktorého predaj analyzujeme, aké produkty sa typicky nakupujú spoločne. Veľmi typickou je i otázka: „Aké produkty je výhodné doporučiť zákazníkovi?“ (3)

Najvhodnejšie algoritmy: asociačné pravidlá, rozhodovacie stromy.

Použiteľné algoritmy: zhľukovanie a sekvenčné zhľukovanie, Naive Bayes, neurónové siete.

1.5.5. **Analýzy textu**

Textové záznamy sú typickým príkladom neštrukturovaných dát. I keď v nich určité pravidlá dokážeme identifikovať, týkajú sa skôr syntaxu a sémantiky prirodzeného jazyka (čeština, angličtina apod.), v ktorom bol text vytvorený. Niektoré heuristické data miningové algoritmy si však poradia i s bežným neštrukturovaným textom. (3)

Obr. 8 zobrazuje použitie data miningových algoritmov pre niektoré typické druhy úloh v prehľadne usporiadanej tabuľke.

<i>Decision Trees</i>	<i>Naïve Bayes</i>	<i>Clustering</i>	<i>Seq. Clustering</i>	<i>Time Series</i>	<i>Association rules</i>	<i>Neural Network</i>	
✓	✓	✓	✓		✓	✓	Classification
✓	✓	✓	✓			✓	Regression
		✓	✓			✓	Segmentation
✓	✓	✓	✓		✓	✓	Assoc. Analysis
		✓	✓			✓	Anomaly Detect
			✓				Seq. Analysis
				✓			Time series

✓ - Najlepší ✓ - Vhodný

Obr. 8: Použitie algoritmov pre typické okruhy úloh (3)

2. Analýza problému a súčasnej situácie

V tejto kapitole budeme analyzovať súčasnú situáciu v spoločnosti Acision. Uvedieme základné údaje o firme, predmet podnikania, krátku históriu a štatistiky. Následne budeme definovať kľúčové skupiny zamestnancov a taktiež procesy s nimi spojené.

Definícia zamestnaneckých pozícií a náplne práce sú obsiahnuté v interných dokumentoch spoločnosti. Je potrebné poznamenať, že tu samozrejme existujú aj iné pozície, z technického hľadiska však nie sú kľúčové. Inými slovami, uvedené pozície sú základný stavebný kameň firmy a realizujú to, čo firme zabezpečuje hlavný príjem z podnikania (pozn. práca nezahrňuje popis manažérskych, ekonomických, marketingových, administratívnych a HR³ pozícií).

V poslednej časti tejto kapitoly priblížime jednotlivé procesy, akým spôsobom sa zaznamenávajú operácie jednotlivých špecialistov, aký systém firma používa (vrátane jeho popisu) a nakoniec samotnú analýzu nad týmito procesmi, tzn. aké bolo rozhodnutie manažmentu ohľadom implementácie aktuálne používaného riešenia. V závere kapitoly sa pokúsime stanoviť klady a zápory tohto riešenia.

2.1. Základné údaje o spoločnosti

V nasledujúcej tabuľke (Tab. 2) sú údaje o českej pobočke spoločnosti Acision (v obchodnom registri ČR zaregistrovaná ako Acision Czech Republic s.r.o., ďalej len Acision) zo Štatistického úradu ČR. Spoločnosť má v súčasnosti na území ČR dve pobočky a to v Brne a v Prahe.

³ Human Resources (ľudské zdroje)

Tab. 2: Základné údaje o spoločnosti Acision Czech Republic s.r.o. (9)

NÁZOV	Acision Czech Republic s.r.o.
SKR. NÁZOV	ACISION CR,SRO,PHA
IČO	26772825
PRÁVNÁ FORMA	Spoločnosť s ručením obmedzeným
ADRESA	Na okraji 335/42, Praha 6 16200
DÁTUM VZNIKU	2.4.2003
POČET ZAMESTNANCOV	500-999 zamestnancov
ZÁKL. KAPITÁL	29.8 miliónu Kč
HL. ČINNOST'	Programovanie
PREDMET PODNIKANIA	Poradenstvo v oblasti riadenia
	Vydávanie softwaru
	Informačné činnosti
	Architektonické a inžinierske činnosti a súvisiace technické poradenstvo
	Prenájom a leasing výrobkov pre osobnú potrebu a prevažne pre domácnosť
	Prekladateľské a tlmočnicke činnosti
	Sprostredkovanie veľkoobchodu a veľkoobchod v zastúpení
	Nešpecializovaný veľkoobchod
	Ostatný maloobchod s novým tovarom v špecializovaných predajniach

2.2. Produkty a riešenia spoločnosti Acision

Spoločnosť Acision ako svetová jednotka v mobilných komunikáciách , poskytuje *real-time* mobilné dátové riešenia, ktoré zároveň podporujú inovácie v tomto odbore, zvyšujú zisk mobilných operátorov a poskytovateľov mobilných služieb po celom svete. Sila týchto inovácií a odborných znalostí spoločnosti Acision rozširuje jej komplexné portfólio *real-time* mobilných dátových riešení (Obr. 9). Jej riešenia mobilných dátových služieb, riadenia dátovej komunikácie a spoplatňovania dátových služieb boli navrhnuté tak, aby splňovali všetky požiadavky svojich zákazníkov - od riadenia a optimalizácie mobilných dátových sietí až po služby a aplikácie zabezpečujúce vysoký príjem z týchto riešení.



Obr. 9: Portfólio real-time mobilných dátových riešení spoločnosti Acision (5)

Spoločnosť Acision:

- odbavuje viac než 1,8 miliardy predplatiteľov globálne,
- doručuje viac ako 900 miliárd textových a multimediálnych správ ročne,
- drží si vedúce postavenie vo svojom odbore podnikania,
- má viac ako 230 zákazníkov vo viac ako 100 krajinách,
- viac ako 2000 nasadených produkčných systémov,
- približne 1600 zamestnancov po celom svete,
- 31 kancelárií a stredísk celosvetovo,
- 650 vývojárov v USA, Holandsku, Indii a Českej Republike.

História spoločnosti siaha do roku 1988, samotná Acision vznikla oddelením od spoločnosti Logica v roku 2007 (do tej doby fungovala ako jedna z divízií tejto firmy). Na nasledujúcom obrázku Obr. 10 je znázornené rozmiestnenie zákazníkov spoločnosti Acision v svete. Medzi jej zákazníkov patrí 8 z 10 TOP mobilných operátorov na svete.

(5)



Obr. 10: Zákazníci spoločnosti Acision vo svete (10)

2.3. Popis pozícií v spoločnosti

V nasledujúcich podkapitolách budú popísané tri kľúčové pozície v spoločnosti Acision a náplň práce tak, ako sú definované samotnou spoločnosťou a ktoré neodmysliteľne súvisia s procesmi popísaných v ďalších kapitolách. Sú to *Customer Services Engineer*, *Sustaining Engineer* a *Research and Development Engineer*. Títo zamestnanci majú v rámci svojich povinností zaznamenávať všetky úkony do aplikácie, nad ktorou sú následne vykonávané štatistiky a analýzy, ktoré slúžia vedúcim a manažérskym pracovníkom v ich ďalšom rozhodovaní.

2.3.1. Customer Services Engineer (CSE)

Náplň práce (Job Purpose):

Je zodpovedný za podporu, zavádzanie nových systémov a ostatné zákaznícke a servisné aktivity. Pracuje samostatne s minimálnym dohľadom.

Kľúčové povinnosti (Key Accountabilities):

- Rieši požiadavky zákazníka podľa SLA⁴,
- implementuje zákaznícke systémy podľa dohodnutých plánov, rozpočtu a kvality,
- zaisťuje, že všetky aktivity sú vykonávané v súlade s procesmi kvality,
- prispieva k spokojnosti zákazníka a plní stanovené ciele,
- identifikuje produktové a procesné zlepšenia,
- školí zamestnancov (*juniorské* pozície) v spoločnosti.

Kľúčové úlohy a aktivity (Key tasks & activities):

Zákaznícky servis (Support):

- Zaznamenáva problém (*incident logging*), overuje oprávnenie a kategorizuje,
- vykonáva podporu širokého spektra produktov (*cross product support*),
- skúma, overuje, zavádza systém do pôvodného stavu na základe známych riešení,
- je schopný analyzovať výpisy súborov aplikácií (*log files and debug*) tak, aby prispel k identifikácii problému a jeho riešeniu,
- určuje komplexnú diagnózu problémov s kritickým dopadom, implementuje dočasné riešenie (*workaround*) tak, aby zaistil firemnú kontinuitu,
- vykonáva analýzy komplexných problémov tretích strán (*3rd party issues*) a spravuje tieto problémy prostredníctvom týchto tretích strán,
- popisuje a kontroluje metódy jednotlivých procedúr,
- popisuje, kontroluje a schvaľuje články vyriešených problémov,

⁴ Service Level Agreement, zmluva o garantovanej úrovni služieb

- navrhuje zmeny produktovej dokumentácie,
- vykonáva audit a pravidelné kontroly (*health checking*) zákazníckych systémov podľa predom definovaných pokynov,
- podieľa sa na zlepšovaní produktu na základe skúseností v danej oblasti,
- podieľa sa na nepretržitej prevádzke (*24x7 on-call coverage*) i mimo pracovnej doby.

Zavádzanie systémov (*deployment*):

- Vykonáva aktivity spojené so zavádzaním nových systémov, pracuje nezávisle s minimálnym dohľadom,
- vylepšuje a vytvára technickú a produktovú dokumentáciu,
- odsúhlasuje testovanie, inštaláciu, integráciu softvéru a dáva pokyny zákazníkovi,
- diskutuje jednotlivé aktivity s lokálnymi technikmi a manažmentom,
- popisuje a kontroluje procedúry. (11)

2.3.2. Sustaining Engineer (SE)

Náplň práce:

Poskytuje technickú podporu najvyššej úrovne (*3rd line support*) produktov spoločnosti za podmienok stanovených v SLA. Overuje produktové operačné schopnosti predtým, ako je produkt vydaný a dostupný zákazníkovi.

Kľúčové povinnosti:

- Nezávisle rieši incidenty tretej úrovne podľa stanovených SLA,
- podieľa sa na uvedení nových produktov,
- kontroluje technickú a produktovú dokumentáciu,
- kontroluje úplnosť a správnosť vedľajšieho softvéru a dokumentácie a výsledkov produktov spoločnosti (hotfixy, opravné balíky, CR⁵ a balíkov s novou funkcionalitou),
- poskytuje detailný vstup pre analýzu príčin (RCA⁶),
- navrhuje produktové/procesné zlepšenia,
- školí zamestnancov (*juniorské* pozície) v spoločnosti.

Kľúčové úlohy a aktivity:

- Poskytuje podporu na najvyššej technickej úrovni pre operatívne problémy podľa definovaných operačných levelov (*OLA*⁷), analyzuje a rieši incidenty a poskytuje dočasné riešenia, diagnostikuje a skúma verzie balíkov, poskytuje finálne riešenie,
- veľmi úzko spolupracuje s vývojárskym tímom (R&D) na definícií, plánovaní, kvalifikovaní a distribúcií vyvíjaných produktov a komponent (pravidelné *patches* a *hot-fixy*),
- proaktívne diskutuje s operačnými tímami (CSE) incidenty a iné technické problémy jednotlivých produktov,
- kvalifikuje, preberá a distribuuje nové softvérové balíky firemným operačným oddeleniam,

⁵ Change Request, požiadavka na pridanie novej funkcionality, popr. zmenu

⁶ Root Cause Analysis, analýza príčiny problému

⁷ Operation Level Agreement, dohoda o úrovni operácie, zmluva

- pravidelne reportuje stav a postup, vrátane denného postupu v reportovacích aplikáciách a nástrojoch pre sledovanie (*tracking tools*),
- podieľa sa na nepretržitej prevádzke (24x7),
- vo všeobecnosti prispieva k zlepšovaniu procesov, úrovni služieb, procesov a produktovej kvality,
- prispieva svojimi vedomosťami a schopnosťami k rozvíjaniu iných tímov (*guiding, coaching, training*),
- definuje riziká smerom k manažmentu,
- postupuje podľa firemných štandardov kvality. (12)

2.3.3. **Research & Development Engineer (R&D)**

Náplň práce:

Vyvíja softvérové produkty a ich komponenty za účelom splnenia požiadaviek na funkcionality, vykonáva údržbu softvéru, zúčastňuje sa analýzy a návrhu produktu.

Kľúčové povinnosti:

- Prispieva k analýze požiadaviek, predkladá požiadavky užívateľa a prípady systémových požiadaviek,
- prispieva k návrhu celého systému, predkladá systémové požiadavky do návrhu systému a komponentov,
- programuje na základe dohodnutých štandardov,
- vykonáva testovanie komponent – testuje softvérové komponenty na funkcionality, stabilitu, dizajn,
- udržiava zdrojový kód,
- rieši softvérové problémy identifikované pri zavádzaní produktu u zákazníka a doručuje opravné balíky,
- vyjadruje stanovisko a odhady z pohľadu vývoja, kontroluje návrhy na zmeny a požiadavky zákazníka.

Kľúčové úlohy a aktivity:

- Spolupracuje na špecifikácií systémových požiadaviek,
- pripravuje návrhy (UML⁸),
- implementuje novú funkcionálnu produkt,
- odstraňuje existujúce problémy na úrovni zdrojového kódu,
- pripravuje a vykonáva proces testovania (*Unit Tests*),
- vykonáva kontrolu kódu (*Code Review*),
- pripravuje integračné testy komponentov (*Component Integration Tests*),
- pripravuje alebo modifikuje konfiguráciu, inštaláciu a manuály súvisiace s daným produktom,
- odhaduje vývoj softvéru v časovom horizonte,
- inštaluje a konfiguruje testovacie a vývojové systémy,
- podieľa sa na aktivitách *sustaining* tímu,
- poskytuje technickú podporu počas testovania a zavádzania produktu. (13)

(R&D tím sa skladá z niekoľkých tímov, ktoré sa venujú špecifickým oblastiam vývoja produktu, ale z pohľadu CSE a SE sa berie R&D ako celok. Uvedený popis zodpovedá pozícií *Software Engineer*. V R&D sa teda vyskytujú i ďalší členovia: *Configuration Engineer*, *Technical Writer*, *Verification Engineer*, *Requirements Engineer*, *Product Architect*, *Software Architect*.)

⁸ Unified Modeling Language, grafický jazyk pre vizualizáciu, špecifikáciu a dokumentáciu programových systémov.

2.4. Popis firemných procesov

V predošlej podkapitole bola detailne popísaná náplň práce a oblasti zodpovednosti jednotlivých špecialistov. Aby však bola zaručená plná podpora, ktorú zákazník vyžaduje, je potrebná a nevyhnutná taktiež kooperácia medzi jednotlivými tímami. V tejto podkapitole sa zameriame na medzi tímové procesy a ich previazanosť v rámci spoločnosti. Sústredíme sa na technických špecialistov, CSE, SE, R&D (ďalej len support, sustaining, R&D).

Keďže spoločnosť vyvinula už veľký počet produktov, uvedení špecialisti sú združení a spolupracujú v rámci jedného produktu alebo skupiny produktov, tzn. že ak máme napríklad produkt s názvom PSA (Pre-delivery Service Agent) a produkt s názvom ACE (Acision Charging Engine), patria do skupiny „charging“. Tým pádom existujú pre skupinu „charging“ tímy support, sustaining i R&D, ktoré majú dané produkty na starosti a musia preto medzi sebou spolupracovať.

Táto spolupráca sa nezaobíde bez softvéru, ktorý by zabezpečoval komunikáciu medzi jednotlivými tímami. Preto sa spoločnosť rozhodla zakúpiť a používať softvér na realizáciu týchto medzi tímových procesov. Osvedčený software od spoločnosti BMC Software s názvom Remedy, ktorý používa rada známych spoločností, ako je napríklad IBM, Hewlet-Packard či Dell, sa spoločnosť Acision rozhodla používať taktiež. Jednou z výhod je tiež fakt, že ak zamestnanci, ktorí do spoločnosti prídu alebo ju opustia nemajú problém adaptovať sa v novom prostredí. Tu je však dôležité pripomenúť, že vo väčšine prípadov si spoločnosť, ktorá si tento softvér zakúpi, tak si ho taktiež prispôbi svojím požiadavkám a ďalej softvér vyvíja, tzn. prebieha tzv. kustomizácia (*customization*). Tá zahŕňa zmenu určitých častí GUI⁹, zmenu názvu a usporiadania jednotlivých polí, sekcií a pod., čo vyžaduje špecializovaný tím, ktorý sa o požiadavky na zmeny a administratívu stará a taktiež tím, ktorý novú funkcionálnu implementuje.

Spoločnosť Acision vytvorila a udržiava taktiež svoju vlastnú kustomizáciu aplikácie BMC Remedy User pod názvom NiKiTa, ktorú používajú support, sustaining aj R&D všetkých produktov a prostredníctvom nej prebieha celá technická komunikácia.

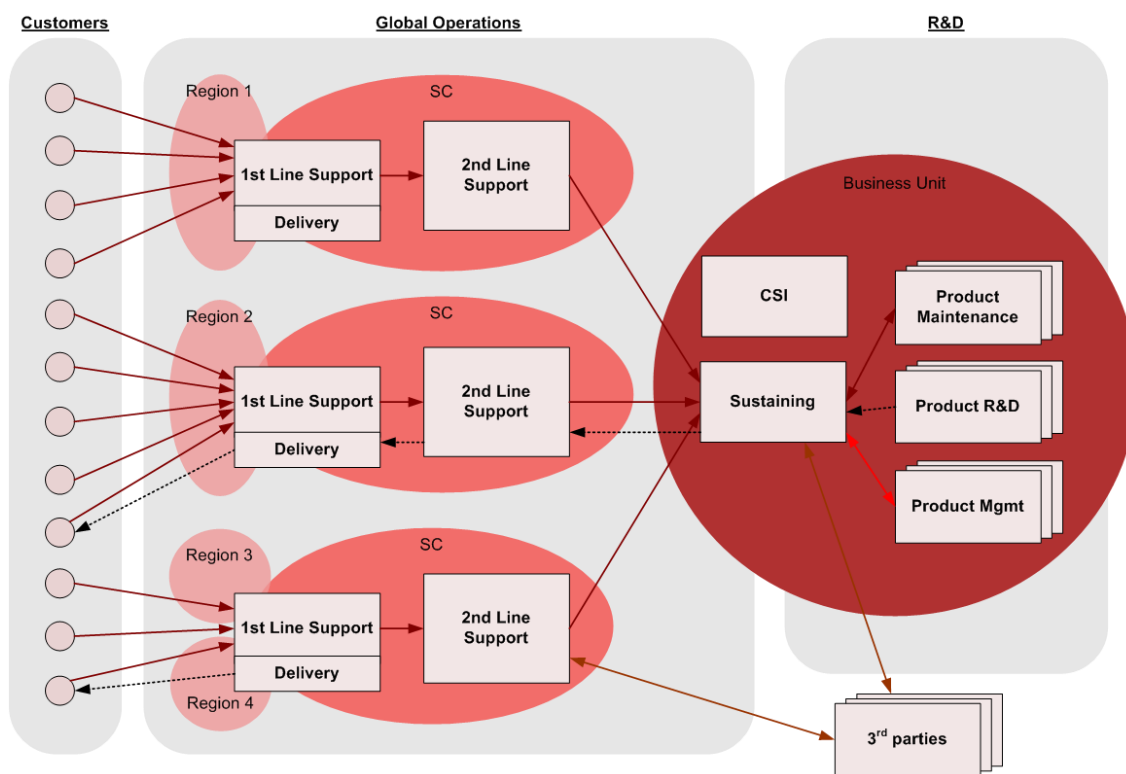
V nasledujúcich podkapitolách si priblížime túto komunikáciu na príkladoch, ktoré objasňujú životný cyklus daného incidentu od jeho vzniku, cez operácie nad ním až do

⁹ Graphical User Interface, grafické užívateľské rozhranie

konca tohto cyklu, tzn. vyriešení a uzavretiu daného problému. Tento scenár sa však môže líšiť od skutočnosti, pretože ani v praxi nikdy nie je vopred jasné, na ktorej technickej úrovni a v ktorej skupine sa daný incident podarí vyriešiť. Preto všeobecne môžeme uvažovať tri základné scenáre a to tie, kde je problém identifikovaný a v samotnom závere vyriešený:

1. Problém sa vyrieši na strane supportu (CSE).
2. Problém sa vyrieši na strane sustainingu (SE).
3. Problém sa vyrieši dodaním hotfixu zo strany R&D.

Na nasledujúcom obrázku (Obr. 11) je znázornené, akým spôsobom jednotlivé tímy komunikujú daný problém, pre lepšie pochopenie jednotlivých scenárov v nasledujúcich podkapitolách.



Obr. 11: Oblasti, vzťahy, a komunikácia jednotlivých tímov (10)

2.4.1. Riešenie problému na strane supportu (CSE)

Majme situáciu, kde nastane problém u zákazníka (*customer*), ktorý je detekovaný tímom, ktorý sa obvyčajne označuje ako *monitoring* alebo aj *1st Line support*. Oblasť, v ktorej pôsobí sa označuje ako región (*region*), tj. určitá skupina zákazníkov. Tím reprezentujúci región zváži, či problém:

- a) Dokáže vyriešiť na základe dostupnej dokumentácie a vlastných znalostí.
- b) Nedokáže vyriešiť a bude potrebovať podporu CSE.

Pokiaľ problém vyriešiť nedokáže, bude potrebné v NiKiTe vytvoriť tzv. incident so všetkými náležitosťami (popis problému, priorita, dokumenty potrebné k analýze, ...) a predať problém support tímu. Akonáhle support tím, označovaný i ako *2nd Line Support* (oblasť v ktorej pôsobí tento tím sa označuje ako SC – *Service Centre*, Obr. 11), tento incident dostane, začne ho riešiť na základe nastavenej priority. Support tím komunikuje s regiónom všetkými dostupnými prostriedkami (email, telefón, IM¹⁰) a v prípade potreby si vyžiada údaje a ďalšie informácie týkajúce sa daného problému. Vo všetkých týchto prípadoch je support engineer pripojený na zákaznicke systémy cez terminál a vykonáva potrebné úkony na základe svojich technických znalostí a vedomostí, má k dispozícii dokumentáciu s ktorou pracuje a podľa ktorej postupuje.

V tomto scenári sa problém podarí vyriešiť už na tejto technickej úrovni. Support engineer problém identifikuje a vyrieši, poprípade doporučí regiónu úkony vedúce k tomuto riešeniu, vyplní v NiKiTe všetky potrebné náležitosti a incident uzavrie. Tento špecialista navyše môže obohatiť vedomostnú databázu (*knowledge base*), poučiť kolegov o postupe a riešení a k všetkým týmto informáciám sa môže kedykoľvek prostredníctvom NiKiTy vrátiť.

¹⁰ Instant Messaging (napr. MS Office Communicator, Jabber, Skype)

2.4.2. Riešenie problému na strane sustainingu (SE)

V tomto prípade sa postupuje ako v predošlom scenári, ale nastáva zmena. 1st Line support engineer ani 2nd line support engineer nie sú schopní so svojimi technickými znalosťami a ani za pomoci dostupnej dokumentácie daný problém vyriešiť. Preto sa zvažuje pomoc *sustaining engineera*, tzn. špecialistu s najvyššími technickými znalosťami v rámci daného produktu popr. produktov.

Support má v tomto prípade za úlohu popísať všetky svoje doterajšie aktivity a prácu na tomto incidente, priložiť povinné dokumenty do NiKiTy, kontaktovať sustaining. K svojej analýze prikladá tzv. *sustaining intake form*, čo je predávací protokol, s ktorým potom support transferuje daný incident do sustainingu (príslušná skupina).

Sustaining má za úlohu skontrolovať tento priložený protokol a v prípade nejasností alebo zistených nedostatkov má oprávnenie daný incident neprijíť. Vo väčšine prípadov však protokol dané náležitosti splňuje a tak sustaining môže začať analýzu na tejto úrovni. V tomto prípade má sustaining možnosť komunikovať tak so stranou supportu ako i so stranou R&D. Je tu pravidlo, že support a R&D spolu komunikovať nesmú. Sustaining postupne žiada ďalšie potrebné materiály a výstupy aplikácií, sieťové záznamy (*debug logs, traces, tcpdumps*) od support tímu a problém analyzuje. Na rozdiel od supportu, sustaining má prístup k zdrojovým kódom daného produktu, ktorých pochopenie taktiež napomáhajú k dosiahnutiu riešenia.

Špeciálna vymoženosť tímu sustaining je, že disponuje tzv. prostrediami na simuláciu, inými slovami má k dispozícii laboratórium, ktoré používa v prípade, kedy nie je možné prísť na daný problém priamo, ale zároveň môže daný problém nasimulovať, tzn. vykonať replikáciu problému. Väčšinou postupuje tým spôsobom, že nastaví svoje prostredie (*test system*, alebo i *testbed*) tak, aby sa aby simuloval zákaznícke prostredie (tzn. rovnaký operačný systém, produkt a komponent, produkty tretích strán). Výsledkom je väčšinou identifikácia problému. Replikácia poslúži ako dôkaz a je silným prostriedkom, za pomoci ktorého sa dôjde k návrhu finálneho riešenia. V sustainingu sa objavujú problémy komplexného typu. V tomto scenári sa problém identifikuje a nie je potrebné zasahovať do zdrojového kódu produktu. Sustaining popíše postup riešenia a navrhne support tímu, ako postupovať a problém odstrániť. Incident v NiKiTe popíše, a transferuje ho späť na support, ktorý u incidentu

vyplní potrebné náležitosti tak, ako v predošlom scenári a problém uzavrie. Väčšinou sa jedná o zlú konfiguráciu, zle navrhnuté prostredie z hľadiska výkonu (*sizing*) ako napr. zle navrhnutá architektúra siete, hardvéru a pod., problémy s databázou nad ktorou daný produkt beží alebo je identifikovaný neodborný zásah niektorého z regiónov atď.

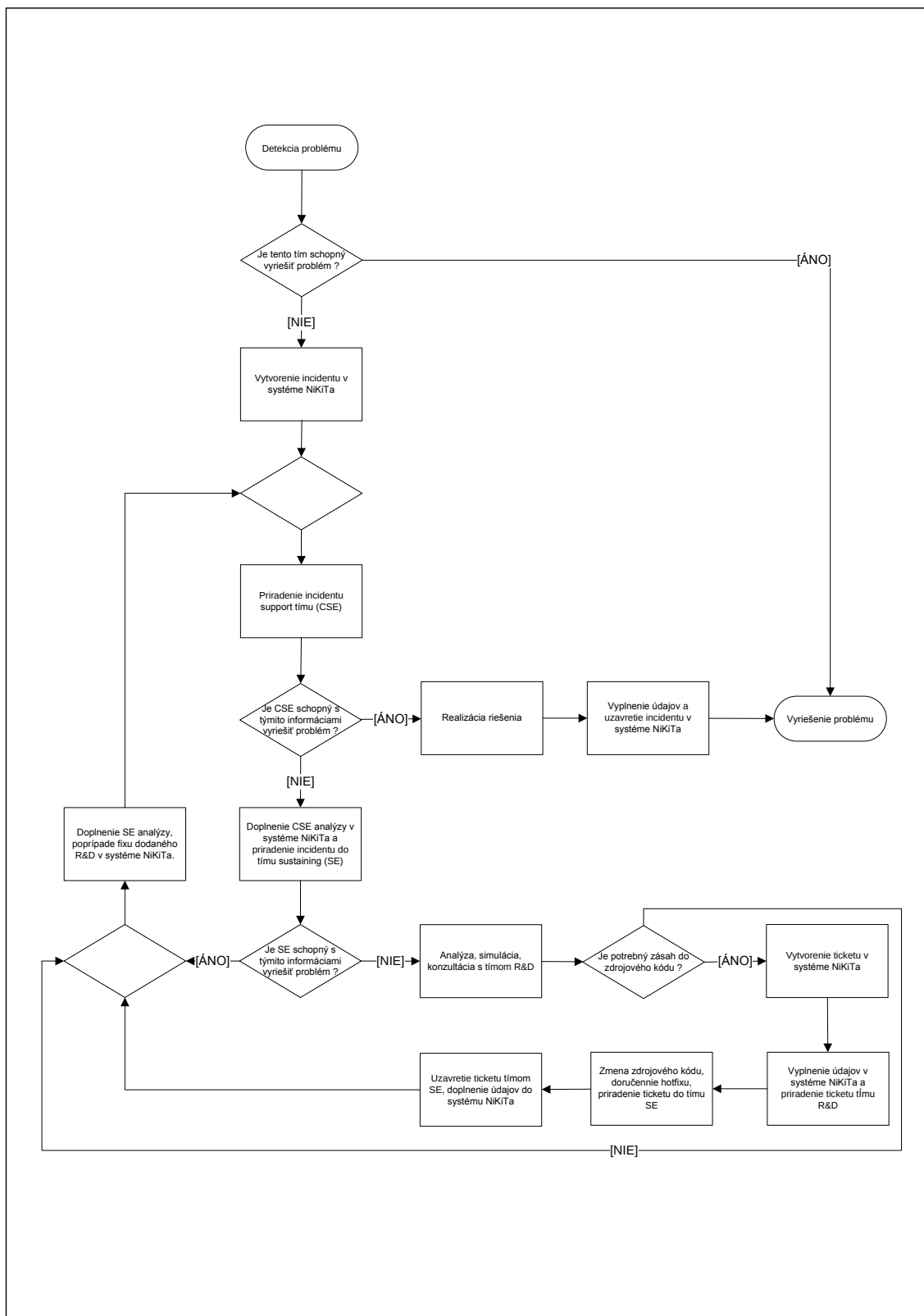
2.4.3. Riešenie problému na strane R&D

V niektorých prípadoch dôjde k situácii, že nie je možné dospieť k riešeniu problému bez zásahu do zdrojového kódu. Pokiaľ na to sustaining príde, je problém potrebné eskalovať to tímu R&D. V prípade, že sa tak sustaining rozhodne, procesne to znamená, že problém musí byť popísaný v tzv. *sustaining handover*, čo je predávací protokol medzi sustainingom a R&D (obdobne to bolo medzi supportom a sustainingom). Taktiež musí obsahovať vopred schválené náležitosti a v systéme NiKiTa musia byť priložené dokumenty súvisiace s analýzou, ktorú vykonal sustaining tím. Tu nastáva zmena, nakoľko incident zostáva u sustainingu.

Sustaining vytvára tzv. *ticket*, ktorý v NiKiTe spojí s daným incidentom. V tomto momente sa vytvára medzi incidentom a ticketom vzťah, tzn. v databázi nad ktorou NiKiTa beží. Novo vytvorený ticket automaticky zdedí všetky údaje z pôvodného incidentu a sustaining do tohoto ticketu doplní údaje potrebné na to, aby R&D začal pracovať na oprave zdrojového kódu. R&D môže počas doby riešenia požadovať ďalšie informácie od sustainingu. Sustaining sa v tomto smere stáva prostredníkom medzi regiónom a R&D, pretože nemá priamy prístup na zákaznícke systémy.

V prípade, že tím R&D opraví zdrojový kód, vytvorí tzv. *build ticket*. Po ukončení všetkých úkonov jednotlivých sub-tímov v R&D, sa v poslednej fáze dostane do stavu *released*, čo znamená, že je nový *hotfix (patch)* pripravený na distribúciu. Tým práca R&D skončila a *ticket* môže byť uzavretý. Sustaining uloží daný *hotfix* na distribučný server a vráti daný incident späť do support tímu. Support následne prevezme tento *hotfix* z distribučného serveru a aplikuje ho najprv na testovací systém zákazníka a ak sa potvrdí správnosť testovaním, fix môže byť aplikovaný na reálny zákaznícky systém.

Všetky horeuvedené scenáre možno zjednodušene graficky reprezentovať na nasledujúcom UML diagrame aktivít. (Obr. 12)



Obr. 12: Riešenie problému – UML diagram aktivít (vlastné spracovanie)

2.5.Action Request System

Spoločnosť Acision, ako už bolo spomenuté v predošlých kapitolách, nasadila pre podporu procesov a medzi-tímovej komunikácie softvér od spoločnosti BMC Software, konkrétne aplikáciu BMC Remedy User.

AR System poskytuje konsolidovanú platformu servisného procesného manažmentu pre automatizáciu a riadenie podnikových procesov servisného manažmentu.

- Nahradzuje zastarané manuálne systémy automatizáciou procesov,
- zahŕňa workflow modulov bežne používaných pri automatizácii servisných procesov ako sú notifikácie, eskalácie a schvaľovanie,
- integruje procesy so systémami v rámci celého podniku,
- prispôsobuje a vyvíja procesy, neustále ich zladzuje s potrebami podniku,
- riadi výkon podnikových procesov v reálnom čase,
- rýchlo sa modeluje, nasadzuje, udržiava aplikácie servisného manažmentu,
- zachytuje a sleduje kritické podnikové dáta.(5)

Tento systém sa dá prispôbiť požiadavkám danej spoločnosti a pretože každá podniká v inej sfére, má inú štruktúru a rozdelenia tímov, spoločnosť Acision si aplikáciu prispôbila taktiež. Tento ARS systém pomenovala, ako už bolo spomenuté, NiKiTa, aktuálne sa používa vo verzii 3.4.12. Vyznačuje sa bohatou funkcionalitou a prehľadným grafickým užívateľským rozhraním. Na strane klienta sa využíva aplikácia BMC Remedy User vo verzii 7.0. V niektorých spoločnostiach sa používa i webové rozhranie z rôznych dôvodov (napr. ako alternatíva v prípade výpadku klasického klienta). Acision má samozrejme k dispozícii špeciálny tím, ktorý NiKiTu udržiava tak, aby bola flexibilná k požiadavkám užívateľov, implementuje novú funkcionalitu a aplikáciu udržiava vo funkčnom stave. Jej výpadok by mal obrovský dopad v spoločnosti a to hlavne na organizáciu procesov v jednotlivých tímoch, preto sa kladie veľký dôraz na bezproblémový chod tohto systému (existujú záložné systémy / servery) a každá väčšia údržba je vopred plánovaná a užívatelia sú o nej informovaní.

Na Obr. 13 je ukážka GUI aplikácie z pohľadu sustainingu (v tomto prípade skupina *BU Messaging*, tím *Sustaining Payments*).

Team Incidents

Incident ID	Priority	Weight	Status	Customer	Assigned Group	Assigned Team	Assigned Engineer	Type	Workaround
INC00207862	40	A	Investigating	ACE statistics document	Utone	Sustaining Payments	Branišlav Bartík	Internal	No
INC00207868	30	B	Investigating	disable smg_sa_sub_tra	Utone	Sustaining Payments	Branišlav Bartík	Internal	No
INC00205630	30	B	Investigating	SPBP Session	Korvdel	Sustaining Payments	Branišlav Bartík	Support	No
INC00204308	30	B	Investigating	Providirect core dumping	Cosmote	Sustaining Payments	Branišlav Bartík	Support	No
INC00207860	30	A	Investigating	CurrencyCode in UCAP r	Globul	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00207187	30	C	Investigating	Stuck calls during faced	Setar	Sustaining Payments	Branišlav Bartík	Support	No
INC00202655	30	B	Investigating	master vxmf file has ap	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00206141	30	A	Investigating	ACE COREDR data gettin	Acision	Sustaining Payments	Branišlav Bartík	Deployment	No
INC00206973	30	C	Investigating	Wrong IC charging after	Cosmote	Sustaining Payments	Branišlav Bartík	Support	No
INC00207074	30	B	Investigating	Third Party Recharge do	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00207998	30	B	Investigating	FinF customer pack call fi	Cosmote	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202830	30	B	Investigating	master vxmf form not trigg	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202834	30	B	Investigating	master vxmf form - form i	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202835	30	B	Investigating	master vxmf does not str	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202838	30	B	Investigating	using an invalid recharg	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202840	30	B	Investigating	transferCalltoCSR plays	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202852	30	B	Investigating	SetApplicationData vxmf	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202854	30	B	Investigating	501003 vxmf falls when	Setar	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00202858	30	B	Investigating	Third Party Recharge in	Setar	Sustaining Payments	Branišlav Bartík	Deployment	No
INC00202900	30	A	Investigating	SPBP logging at debug	Telenor Pakistan	Sustaining Payments	Branišlav Bartík	Deployment	Yes
INC00204416	30	A	Investigating	ACE 2.7 FOA TransServ	Telenor Pakistan	Sustaining Payments	Zbynek Michovsky	Internal	Yes
INC00202403	30	A	Investigating	Tranidty Process keeps	Globacom	Sustaining Payments	Zbynek Michovsky	Internal	Yes
INC00202684	30	B	Investigating	Sending the SSNIP D to f	DATECOM India	Sustaining Payments	Zbynek Michovsky	Support	Yes
INC00207131	30	B	Investigating	Propose an alternative f	Bateico (Bahamas)	Sustaining Payments	Zbynek Michovsky	Internal	Yes
INC00201558	20	A	Investigating	Query Customer has as	STC	Sustaining Payments	Zbynek Michovsky	Support	No
INC00203858	10	Q	Investigating	INTERNAL - incorrect ex	Setar	Sustaining Payments	Zbynek Michovsky	Internal	Yes
INC00204979	10	Q	Investigating	INTERNAL - third party r	Setar	Sustaining Payments	Zbynek Michovsky	Internal	Yes
INC002021453	10	Q	Investigating	INTERNAL - Ticket to tra	Setar	Sustaining Payments	Zbynek Michovsky	Internal	No

Obr. 13: NiKiTa, záložka Incident (vlastné spracovanie)

Jednotlivé incidenty, ktoré boli na supporte vytvorené a následne priradené do tejto sustainingovej skupiny, sú prehľadne zobrazené formou zoznamu. Stĺpce reprezentujú nasledovné hodnoty: Identifikačné číslo incidentu *Incident ID*, váha *Priority Weight*, priorita *Priority*, dátum vytvorenia *Create Date*, krátky popis *Short Description*, stav *Status*, zákazník *Customer*, priradená skupina *Assigned Group*, priradený tím *Assigned Team*, priradený špecialista *Assigned Engineer*, typ *Type*, dočasné riešenie *Workaround*.

Tento jednoduchý zoznam dáva členovi priradeného tímu prehľad o tom, na aké incidenty by sa mal práve zamerať. Aby bolo jednoznačne identifikované, kto na incidente pracuje, je možné vidieť v poli *Assigned Engineer* a stav práce v poli *Status*. Na Obr. 13 sú v tomto prípade všetky incidenty priradené jednotlivým špecialistom a sú v stave vyšetřovaný *Investigating*, čo znamená, že nad nimi aktuálne prebieha analýza.

Člen tímu si taktiež môže svoje incidenty vyfiltrovať, tzn. zobraziť len tie, ktoré spĺňajú určité kritériá, napr. spadajú pod jeho zodpovednosť (tlačidlo *My Incidents*), alebo taktiež len tie s určitým stavom. Ten môže byť nasledovný: nový *New*, priradený *Assigned*, vyšetřovaný *Investigating*, čakajúci *Wait*, vyriešený *Resolved*, uzavretý *Closed*, zamietnutý *Cancelled*. Stav incidentu sa v priebehu času mení v závislosti na situácií, napr. zo stavu vyšetřujúci do stavu čakajúci v prípade, že sustaining očakáva od daného regiónu informácie alebo potrebné súbory, bez ktorých analýza nie je možná.

Zmena stavu do stavu čakajúci sa deje pod súhlasom zákazníka (kvôli časomiere SLA). Na Obr. 14 je už pohľad na samotný incident. Ten musí obsahovať potrebné náležitosti. Niektoré polia sú automaticky generované systémom, ostatné však musí vyplniť špecialista (región), ktorý daný incident vytvorí.

The screenshot shows the 'Incident (Modify)' form in the BMC Remedy User interface. The form is titled 'Incident INC00286006 (Modify)'. It contains several sections for data entry:

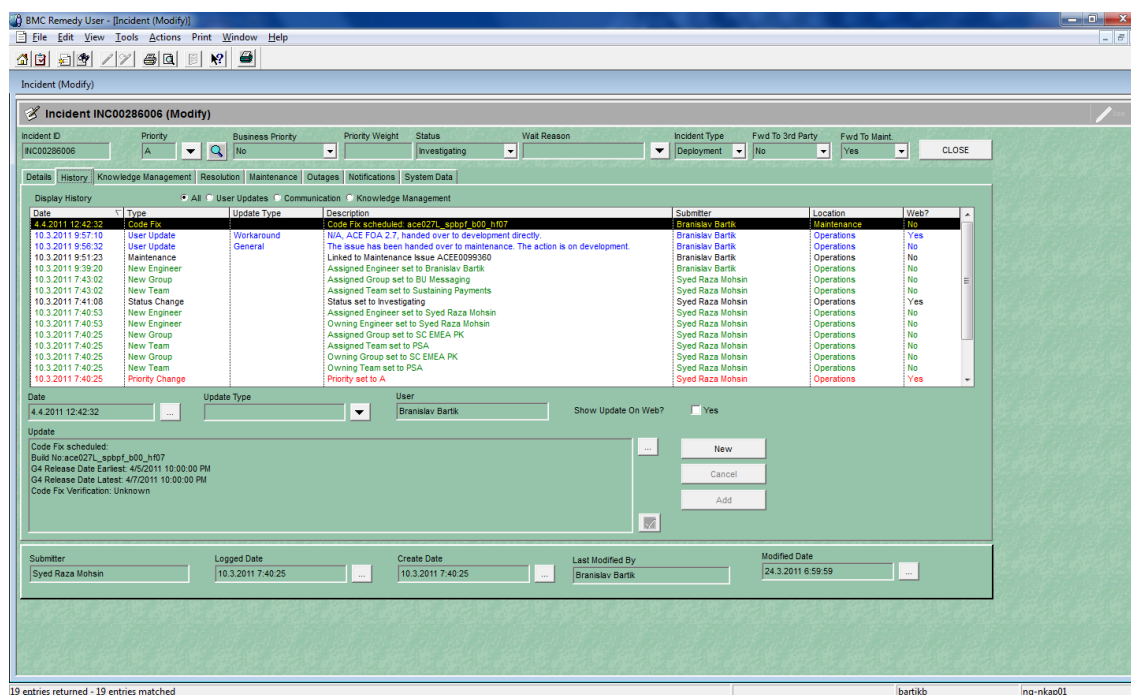
- Incident Details:** Incident ID (INC00286006), Priority (A), Business Priority (No), Priority Weight, Status (Investigating), Wait Reason, Incident Type (Deployment), Fwd To 3rd Party (No), Fwd To Maint. (Yes).
- Incident Details:** System Name (ACPKA11), System Alias (6542043024), Service Tag, Product Solution (Action Charging Engine), Component, Release (2.7), Build No. (ace027L_RR600_h02), Workaround (Yes).
- Short Description:** SPBPF logging at debug level 0.
- Detailed Description:** SPBPF keeps logging the following even at debug level 0, this causes the exhaustion of debuglogs...
- Attachments:** A table listing attachments with columns for Description, Create Date, and Size (KB). Attachments include 'sys_log', 'spbp112920.txt', 'rpm_log', 'env', 'charging_sustaining_intake_form.txt', and 'ace.pc'.
- Customer:** Customer (Telenor Pakistan), Contract Status, Support Partner, Contact Name (TP VAS), Telephone (+92345500106), Mobile (+923456503378), Email (syed.mohsin@telenor.com.pk).
- Assignment:** Owning Group (SC EMEA PK), Assigned Group (BU Messaging), Owning Team (PSA), Assigned Team, Owning Engineer (Syed Raza Mohsin), Assigned Engineer (Branislav Bartik).
- Submission:** Submitter (Syed Raza Mohsin), Logged Date (10.3.2011 7:40:25), Create Date (10.3.2011 7:40:25), Last Modified By (Branislav Bartik), Modified Date (24.3.2011 6:59:59).

Obr. 14: NiKiTa Incident, záložka Details (vlastné spracovanie)

Incident obsahuje niekoľko záložiek, medzi najdôležitejšiu však patria detaily (*Details*). Tá obsahuje 3 hlavné podskupiny: *Incident* (údaje o incidente, z nich najdôležitejšie produkt *Product*, komponenta *Component*, *Release*, *Build No.*, krátky popis *Short Description*, detailný popis *Detailed Description*, prílohy *Attachments*), *Customer* (údaje o zákazníkovi vrátane potrebných kontaktov) a *Assignment* (údaje o tom, ktorá daný incident vytvoril, vlastníca skupina *Owning Group*, vlastníci tím *Owning Team*, vlastníci špecialista *Owning Engineer* a taktiež údaje, komu je aktuálny incident priradený, priradená skupina *Assigned Group*, priradený tím *Assigned Team*, priradený špecialista *Assigned Engineer*).

Ak do skupiny sustainingu príde incident (z regiónu), sustaining špecialista si incident otvorí a skontroluje, či spĺňa kritériá a obsahuje potrebné náležitosti. V prípade že áno, incident si na seba priradí a začne problém analyzovať. V opačnom prípade priradí incident do pôvodnej, vlastniacej skupiny s relevantným oddôvodnením.

Sustaining si prečíta predávací protokol, detailný popis problému, študuje priložené dokumenty a zahajuje tak počiatočnú analýzu. Kooperácia medzi sustainingom a regiónom je nevyhnutná, a taktiež sa musí zaznamenávať (v incidente záložka *History* Obr. 15). To zahrňuje všetku komunikáciu, zmeny údajov týkajúce sa incidentu napr. zmena priority, stavu, atp. Záložka histórie incidentu má svoj praktický význam. Okrem zainteresovaných (región, sustaining) sa ktokoľvek a kedykoľvek môže pozrieť na stav analýzy daného incidentu, ako často bola história aktualizovaná a zákazník, popr. manažment taktiež môžu nahliadnuť a sledovať tak prácu jednotlivých špecialistov.



Obr. 15: NiKiTa Incident, záložka History (vlastné spracovanie)

Ďalšia záložka incidentu je tzv. *Resolution* alebo "Riešenie". Najdôležitejšia sekcia je tzv. "zmluva o garantovanej úrovni služieb" *Service Level Agreement* (ďalej len SLA).

Je to pojem, ktorého vznik bol vynútený potrebou čo najpresnejšie definovať rozsah, úroveň a intenzitu externe poskytovaných služieb. Jednalo sa hlavne o servisné zmluvy v oblasti IT. Neodmysliteľne je SLA zviazané s outsourcingom¹¹, kde predstavuje hlavný parameter rozhrania medzi externým poskytovateľom a príjemcom služieb, tzn. zákazníkom.(7)

¹¹ Outsourcing je uskutočňovanie činností pomocou vonkajších zdrojov (externých firiem)

SLA definuje, aké služby budú dodávané, kedy (napr. časové rozmedzie od 09:00 do 18:00; čas do opravy, príjazdu technika apod.) a kde (miesto plnenia služby). SLA samozrejme môže obsahovať postihy za nedodržanie definovaných parametrov a niekedy taktiež bonifikáciu za ich prekročenie. Pre to, aby bolo možné preukázať nedodržanie, alebo naopak prekročenie garantovaných parametrov, slúžia nástroje merania služby - požiadaviek a ich vyriešenie (napr. *helpdesk*). Cieľom SLA je dosiahnutie užívateľskej, teda zákazníckej spokojnosti. Spokojnosť užívateľov je pre nás jedným z najdôležitejších faktorov. (7)

V ARS systéme NiKiTa sú zadefinované 3 dôležité SLA položky (Obr. 16) a s nimi súvisiace dátumy a časy, do kedy sa má daná skutočnosť naplniť (*Target Date*), dátumy a časy uskutočnenia, tzn. kedy bola daná udalosť splnená (*Actual Date*), informácia o tom či zmluvná časomiera ešte stále beží (*Running*) a či bola prekročená (*Exceeded*). Položky sú nasledovné:

- dátum a čas registrácie prvej odpovede špecialistu, *Response*,
- dátum a čas registrácie dočasného riešenia, *Workaround*,
- dátum a čas registrácie finálneho riešenia, *Solution*.

Incident (Modify)

Incident INC00286006 (Modify)

Incident ID: INC00286006, Priority: A, Business Priority: No, Priority Weight: 1, Status: Investigating, Wait Reason: Deployment, Incident Type: Deployment, Fwd To 3rd Party: No, Fwd To Maint: Yes, CLOSE

Details | History | Knowledge Management | Resolution | Maintenance | Outages | Notifications | System Data

Resolution

Issue Classification: [dropdown], Classification: [dropdown], Total Time spent on call: 1, Operational Change Request Date: [dropdown], Due Date: [dropdown]

Solution: [text area]

Service Level Agreement

Solution target time expired by 23 days 7 hours 17 minutes

	Target Date	Actual Date	Running	Exceeded	Remaining Time (min)	Time Type
Response	10.3.2011 9:10:25	10.3.2011 7:41:08	No	No		Internal
Workaround	15.3.2011 12:40:25	10.3.2011 9:57:10	No	No		Internal
Solution	24.3.2011 8:40:25		Yes	Yes		Internal

Relationships

Related Incident	Short Description	Customer	Priority	Status	Relationship
[empty table body]					

Break Link

Submitter: Syed Raza Mohsin, Logged Date: 10.3.2011 7:40:25, Create Date: 10.3.2011 7:40:25, Last Modified By: Branislav Bartik, Modified Date: 24.3.2011 6:59:59

1 entries returned - 1 entries matched

Obr. 16: NiKiTa Incident, záložka Resolution (vlastné spracovanie)

Tieto časomiere sú generované systémom automaticky na základe hodnoty priority incidentu a zákazníka. Tieto hodnoty sú však vypočítané podľa toho, ako sú definované v karte zákazníka, tzn. na základe samotnej zmluvy medzi zákazníkom a spoločnosťou Acision. Prvá odpoveď špecialistu *Response*, zjednodušene, je registrácia prvej aktualizácie špecialistu v histórii incidentu. Ak špecialista napríklad zaregistruje informáciu, že práve inicializuje vyšetřovanie a pod., tým pádom je splnená prvá z troch podmienok SLA.

Položka *Workaround* - dočasné riešenie, možno väčšinou považovať za riešenie, ale nie je tomu tak. Táto položka sa vytvorila za účelom nájdenia takého riešenia, ktoré ak je to možné, aplikovať čo najskôr na daný problém bez toho, aby to malo dopad na servis. Je naň dokonca kladený väčší dôraz ako na finálne riešenie, pretože sa tu jedná o rýchlosť poskytnutia riešenia, a vo väčšine prípadov sa *workaround* taktiež stáva i finálnym riešením (pre predstavu, napr. vyradenie služby posielania SMS správ u veľmi veľkého zákazníka na pár hodín by mohlo znamenať škodu rádovo stotisíc dolárov a tým pádom obrovskú nespokojnosť koncových užívateľov).

Položka *Solution* - finálne riešenie, je automaticky zaregistrovaná systémom v prípade, že špecialista nájde toto riešenie a uvedie tak incident do stavu *resolved* (záložka *Incident, Action -> Resolve*). Tým je splnená posledná podmienka SLA, a incident tak môže byť uzavretý (stav *Closed*). Povinnosť uzatvoriť incident má špecialista, ktorý daný incident vytvoril.

V prípadoch, kde je potrebné do procesu zahrnúť aj tím R&D, sustaining vytvorí nad incidentom tzv. *development ticket*. Ak je teda potrebná intervencia R&D tímu (tzn. doručenie opravného balíku, zmena zdrojového kódu produktu, komponenty), sustaining vytvorí tento *development ticket* (záložka *Incident, Action -> Forward to Maintenance*). ARS systém vytvorí v rámci incidentu novú záložku s názvom *Maintenance*, v pod ktorou je možné nájsť tento *ticket* a zobrazíť jeho detaily. V databázi, nad ktorou systém ARS beží sa vytvorí vzťah medzi ním a týmto incidentom. Sustaining v tomto *tickete* vyplní taktiež potrebné náležitosti (podobne ako región pri vytváraní incidentu smerom na sustaining) a priradí ho danej skupine R&D. Tu však nastáva zmena, vlastníaca skupina (tím, špecialista) je v tomto prípade sustaining, a priradená skupina (tím, špecialista) je R&D. *Development ticket* (Obr. 17) má tú vlastnosť, že pri vytvorení zdedí všetky informácie ktoré boli v incidente, s ktorým je

teraz vo vzťahu (produkt, krátky popis, detailný popis, všetky prílohy, ale i celú doterajšiu históriu, tzn. komunikáciu medzi regiónom a sustainingom, atp.). Od tohto momentu sa však história *development ticketu* nededí. Je to tak kvôli tomu, aby sa zachovali určité procesné stanovy. Zjednodušene, sustaining komunikuje tak ako s regiónom, tak i s R&D, ale R&D priamo s regiónom komunikovať nesmie. Jedno z možných vysvetlení je i fakt, že sustaining a R&D môžu diskutovať časti zdrojového kódu, ku ktorému by regióny nemali mať prístup.

Obr. 17: NiKiTa Development Ticket, záložka Details (vlastné spracovanie)

Pokiaľ R&D potrebuje údaje z daného systému, pošle túto požiadavku sustainingu, a ten ju prepošle regiónu. V prípade, že dané informácie/súbory región dodá, sustaining ich prepošle tímu R&D. Tieto úkony samozrejme prebiehajú v čase, tzn. SLA budíky neustále bežia a často sa stáva, že niektorá zo strán (región, sustaining, R&D) má meškanie. S prihliadnutím na prioritu, tím, ktorý je označený ako priradený (*Assigned*) preradí tento *incident/ticket* na tím, od ktorého je očakávaná aktivita a na ktorí tím sa vlastne v celom procese riešenia problému čaká. Samozrejme, mení sa i stav incidentu, tzn. z stavu vyšetrujúci do stavu čakajúci a pod. Životný cyklus incidentu alebo ticketu je v závere ukončený, ak skončí v stave uzavretý *closed*, a tým pádom na nich končí práca jednotlivých tímov.

2.6. Firemné riešenie analýzy nad procesmi

Štatistiky a analýzy nad špecifickými procesmi sú nevyhnutné v každej spoločnosti za účelom zhodnotenia minulého stavu, zhodnotenia aktuálneho stavu a s tým spojené rozhodovanie manažmentu a to nie len v IT priemysle, ale v ktoromkoľvek inom odvetví. Výsledkom takýchto rozhodnutí môže byť napr.: naberanie nových zamestnancov (v prípade veľkého množstva práce), prepustenie zamestnancov (prípade nadbytočnosti), reorganizácia práce, zistenie silných a slabých článkov v jednotlivých tímoch, zabezpečenie školenia pre zamestnancov, povýšenie atp., no hlavným cieľom je však zvýšenie efektivity jednotlivých procesov.

Moderným a vysoko použiteľným prostriedkom, ktorý podporuje proces rozhodovania je práve *Business Intelligence* a ním spojený *Data mining*. V tejto podkapitole je popísané riešenie v spoločnosti Acision a v závere taktiež zhodnotenie vhodnosti tohto riešenia v rámci týchto metód.

2.6.1. Databáza v spoločnosti Acision

V predošlej podkapitole bolo popísané, čím by mala daná databáza disponovať a čo všetko by mala spĺňať, aby bolo možné pomocou prostriedkov BI dosiahnuť stanovené ciele, tzn. i správne rozhodnutia.

V kapitole 2.5 bol popísaný ARS systém NiKiTa a procesy, ktoré sú realizované prostredníctvom tohto systému a taktiež informácie, ktoré jednotliví špecialisti/tímy do NiKiTy zadávajú. Systém beží nad databázou SQL od spoločnosti Microsoft, tzn. všetky dáta ktoré do NiKiTy za použitia BMC Remedy User klientov vstupujú alebo z nej vystupujú, sú uložené práve v tejto databázi. MSSQL server je systém komerčný a disponuje širokou škálou analytických nástrojov a taktiež práve tých, ktoré podporujú samotný BI.

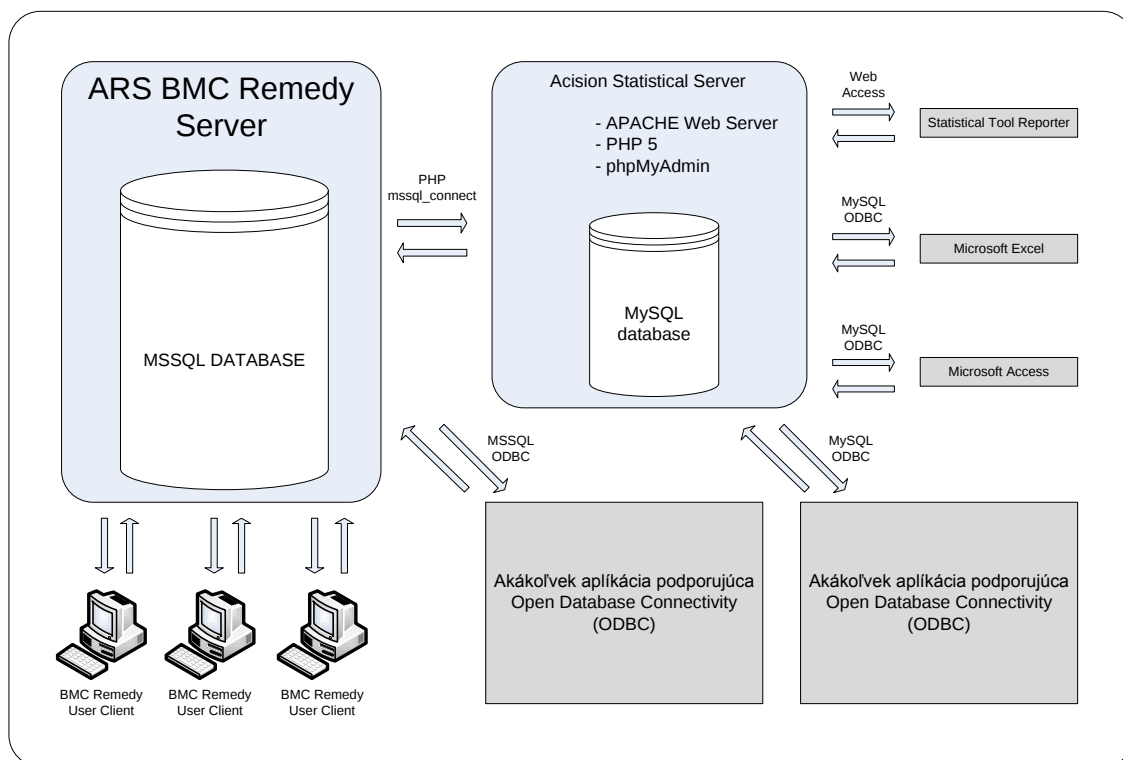
Spoločnosť Acision sa to však rozhodla riešiť iným spôsobom, rozhodla sa svoje štatistiky a analýzy realizovať na nekomerčných, voľne dostupných riešeniach a pôvodnú databázu nechala nedotknutú. Vedľa systému, na ktorom beží NiKiTa teda vytvorila server s databázou MySQL, webovým serverom Apache vrátane PHP a nástroja phpMyAdmin, ktorý slúži na jednoduchú správu obsahu databáze MySQL.

Pomocou PHP a špeciálne implementovaných skriptov v tomto jazyku sa, tak kopírujú dáta databáze zo systému MSSQL do systému MySQL v pravidelných intervaloch, vždy v stredu v čase 23:59:59 a na konci každého kalendárneho mesiaca, taktiež v čase 23:59:59. Do oboch databází je možné pristupovať prostredníctvom tzv. ODBC¹² aplikáciami podporujúcimi túto funkcionality, avšak cieľom tohto firemného riešenia je nezaťažovať systém ARS požiadavkami SQL zo strany týchto, alebo akýchkoľvek iných aplikácií, napr. spoločnosťou vyvinutá webová aplikácia Statistical Tool Reporter, ktorá beží na rovnakom serveri ako databáza MySQL. Takto bola formulovaná požiadavka manažmentu, a schéma tohto riešenia je zobrazená na nasledujúcom obrázku (Obr. 18).

Tento server bol vytvorený za účelom získavať len určité dáta, alebo lepšie povedané len určitú skupinu dát. S týmito dátami môžu následne pracovať analytici (v Acision je to *Performance Team*) i manažment. To znamená, že sa prostredníctvom aplikácií, ktoré podporujú pripojenie do MySQL databáze cez ODBC. Príkladom môžu byť aplikácie Microsoft Excel alebo Microsoft Access (pozn. systém musí obsahovať ovládač MySQL ODBC connector), ktoré sa do tejto databáze pripoja.

Tieto aplikácie aktívne využívajú manažéri v spoločnosti Acision, vytvárajú jednotlivé pohľady do databáze, zasielajú požiadavky na dáta a následne ich podľa potreby v tých aplikáciách spracovávajú a vyhodnocujú. Vytvárajú, štatistiky, grafy, tabuľky, ktoré môžu použiť vo svojich analýzach, alebo taktiež v prezentáciách, na mítingoch a pod.

¹² Open Database Connectivity je štandardizované softvérové aplikačné rozhranie pre prístup k databázovým systémom.



Obr. 18: Databáze v Acision s.r.o. (vlastné spracovanie)

Hlavný zámer tohto firemného riešenia bol nezaťažovať systém ARS požiadavkami SQL zo strany týchto, alebo akýchkoľvek iných aplikácií. Manažment sústreďuje svoju pozornosť na nasledujúce merateľné údaje jednotlivých tímov, skupín, produktov, špecialistov:

- *Response* (bol prekročený čas na poskytnutie prvej odpovede ?).
- *Workaround* (bol prekročený čas na poskytnutie dočasného riešenia ?).
- *Solution* (bol prekročený čas na poskytnutie finálneho riešenia?).
- *Workload* (aké je zaťaženie jednotlivých tímov ? Počet vytvorených incidentov a pridatých do skupín, počet odchádzajúcich a vyriešených incidentov).

To, či bola meraná hodnota (čas) prekročená alebo nie, bolo pri implementácii označované ako pozitívna alebo negatívna udalosť. Na základe tejto informácie potom môžeme určiť úspešnosť jednotlivých složík SLA za určité časové obdobie (najčastejšie týždne / mesiace). Tento vzťah nám vyjadrí nasledujúci vzorec:

$$Effectivity(\%) = \left(1 - \frac{\sum negative_{events}}{\sum negative_{events} + \sum positive_{events}}\right) \cdot 100$$

To znamená, že úspešnosť napr. workaround za mesiac vypočítame ako počet všetkých incidentov, ktoré prekročili čas na *workaround* a vydelíme ich počtom incidentov, u ktorých bol prekročený čas a incidentov, u ktorých bol tento čas splnený (súčet). Nakoniec vynásobíme výsledok číslom 100 a výjde nám úspešnosť v percentách. Na nasledujúcom príklade vypočítame úspešnosť *workaround* pre prípad, kde bolo 65 pozitívnych udalostí, a 35 negatívnych za časové obdobie, napr. mesiac:

$$EffectivityWA(\%) = \left(1 - \frac{\sum negative_{WAevents}}{\sum negative_{WAevents} + \sum positive_{WAevents}}\right) \cdot 100$$

$$\left(1 - \frac{35}{35 + 65}\right) \cdot 100 = 65$$

Úspešnosť *workaround* bola teda 65%. SLA sa však skladá z troch merateľných zložiek, to znamená *response*, *workaround* a *solution*. Tým pádom je potrebné do výpočtu zahrnúť ešte zostávajúce – *response* a *solution*. Vzorec pre výpočet celkovej úspešnosti SLA za časové obdobie bude teda:

$$Effectivity Total(\%) =$$

$$1 - \left(\frac{\sum neg_{RE} + \sum neg_{WA} + \sum neg_{SOL}}{(\sum neg_{RE} + \sum neg_{WA} + \sum neg_{SOL}) + (\sum pos_{RE} + \sum pos_{WA} + \sum pos_{SOL})} \right) \cdot 100$$

To znamená, že napr. pri súčte všetkých negatívnych *response time* $\sum neg_{RE} = 35$, *workaround time* $\sum neg_{WA} = 25$, *solution time* $\sum neg_{SOL} = 10$ a pri súčte všetkých pozitívnych *response time* $\sum pos_{RE} = 65$, *workaround time* $\sum pos_{WA} = 80$ a *solution time* $\sum pos_{SOL} = 90$ za dané časové obdobie, dostávame:

$$1 - \left(\frac{35 + 25 + 10}{(35 + 25 + 10) + (65 + 80 + 90)} \right) \cdot 100 \cong 70$$

Úspešnosť SLA bola za toto časové obdobie približne 70%.

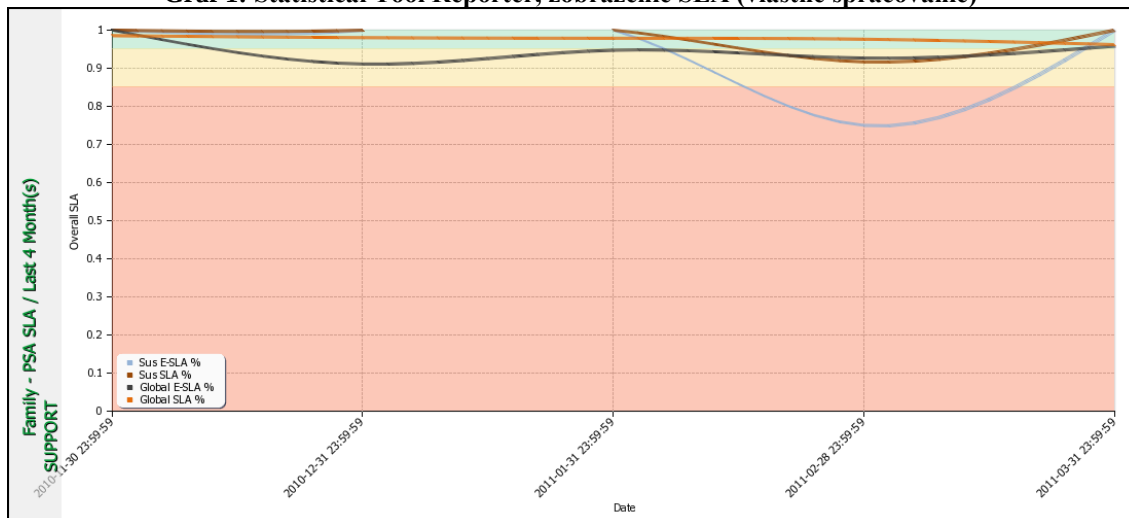
2.6.2. Webová interpretácia informácií

Ďalšou požiadavkou manažmentu bola implementácia jednoduchého webového rozhrania, ktoré dokáže reprezentovať dáta v databáze MySQL v grafickej podobe. Mal byť jednoduchý, prehľadný a má zachycovať *Response*, *Workaround*, *Solution*, *Workload* za určité časové obdobia a zároveň mal byť dostupný z intranetu všetkým zamestnancom spoločnosti. Táto webová aplikácia dostala názov *Statistical Tool Reporter* a je zobrazená na Obr. 19. Tu je ako príklad požiadavka na grafické zobrazenie s nasledujúcimi parametrami: detail produktu PSA (*Response*, *Workaround*, *Solution*) všetkých incidentov v sustainingu tohoto produktu typu *Support*, vrátane *emergency* incidentov za posledné 4 mesiace.

Obr. 19: Statistical Tool Reporter, výber SLA (vlastné spracovanie)

Výstup tohto požiadavku je zobrazený na nasledujúcom grafe (Graf 1). Na ose x sú zobrazené jednotlivé dátumy, tzn. rok, mesiac, deň a čas, kedy bola vykonaná aktualizácia tejto databáze, a na ose y percentuálne zastúpenie SLA k týmto dňom. Z tohto výstupu je zrejmé, že úspešnosť SLA nespadá pod hranicu 90%, čo možno považovať za prijateľný stav.

Graf 1: Statistical Tool Reporter, zobrazenie SLA (vlastné spracovanie)



Takisto bolo implementované zobrazenie záťaže jednotlivých tímov za určité časové obdobie. Nasledujúca požiadavka (Obr. 20) zobrazuje výber skupiny *Sustaining Payments* za posledné 4 mesiace (pozn. posledné 4 mesiace sa počítajú v čase požiadavky, v tomto prípade bola vyslaná 10.4.2011, takže posledná aktualizácia bola 31.3.2011 v čase 23:59:59 a zároveň to bude i posledný zobrazený mesiac).

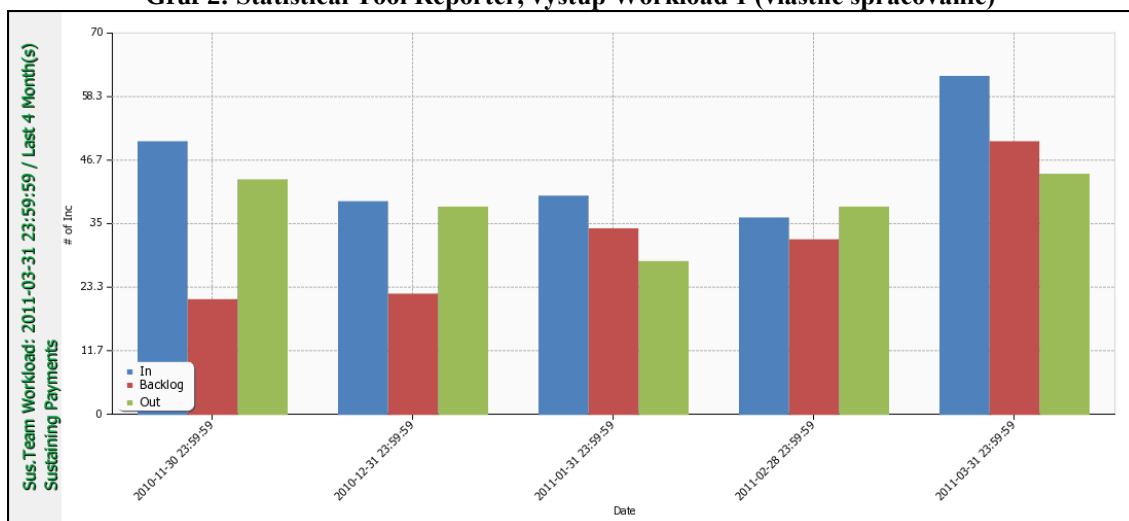
Obr. 20: Statistical Tool Reporter, výber Workload (vlastné spracovanie)

Výstupom tohto požiadavku vygenerované 2 grafy. Graf 2 zobrazuje za jednotlivé časové intervaly (osa x) počet incidentov, ktoré so tímu *Sustaining Payments* boli priradené (*In*, modrá farba), počet incidentov bez zmeny (*Backlog*, červená farba) a počet incidentov, ktoré z tohto tímu odišli (*Out*, zelená farba). Osa y reprezentuje počet týchto takýchto incidentov. Inými slovami, je tu vykreslené zaťaženie a čiastočne i

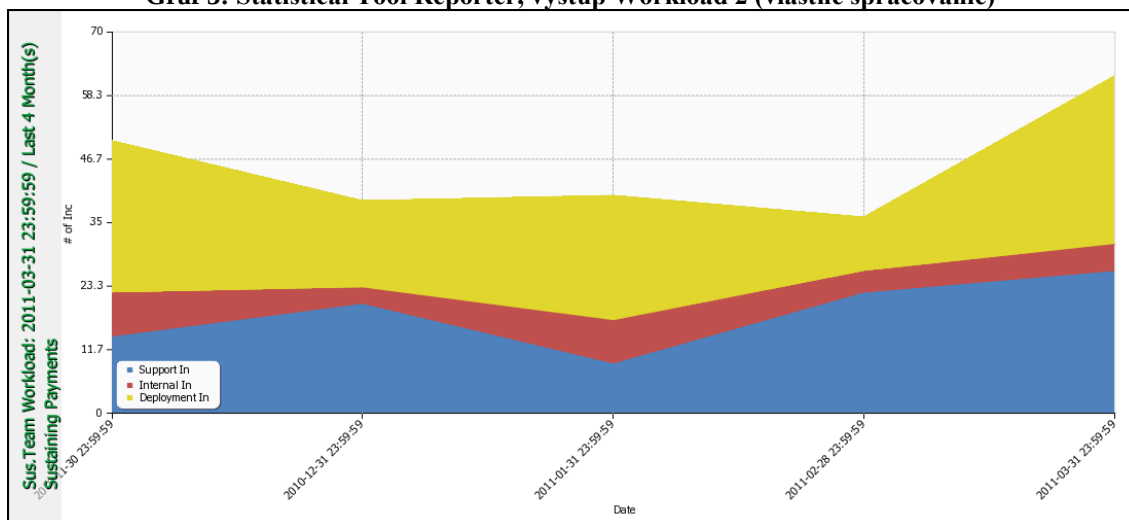
výkon tohto tímu. V tomto prípade je možné sledovať veľký nárast počtu priradených incidentov do tímu za mesiac marec oproti predošlému mesiacu takmer o polovicu.

Ďalší z grafov (Graf 3) v tých istých časových úsekoch zobrazuje ešte počty incidentov jednotlivých typov, tzn. *Support Incidents* (modré), *Internal Incidents* (červené) a *Deployment Incidents* (žlté). Z tohoto grafu je jasné, že v poslednom mesiaci stúpol počet incidentov typu *Deployment*, takže k predošlému tvrdeniu by bolo vhodné dodať, že celkovo stúpol počet práve tohto typu incidentov. (pozn. *Deployment* incidenty vznikajú pri nasadzovaní nového produktu u zákazníka, *Internal* incidenty sú vytvorené väčšinou na požiadanie samotnej spoločnosti Acision, ale jedine u incidentov typu *Support* sa počíta celkové SLA, tj. podľa aktuálneho rozhodnutia spoločnosti Acision).

Graf 2: Statistical Tool Reporter, výstup Workload 1 (vlastné spracovanie)



Graf 3: Statistical Tool Reporter, výstup Workload 2 (vlastné spracovanie)



2.7. Zhodnotenie súčasného stavu

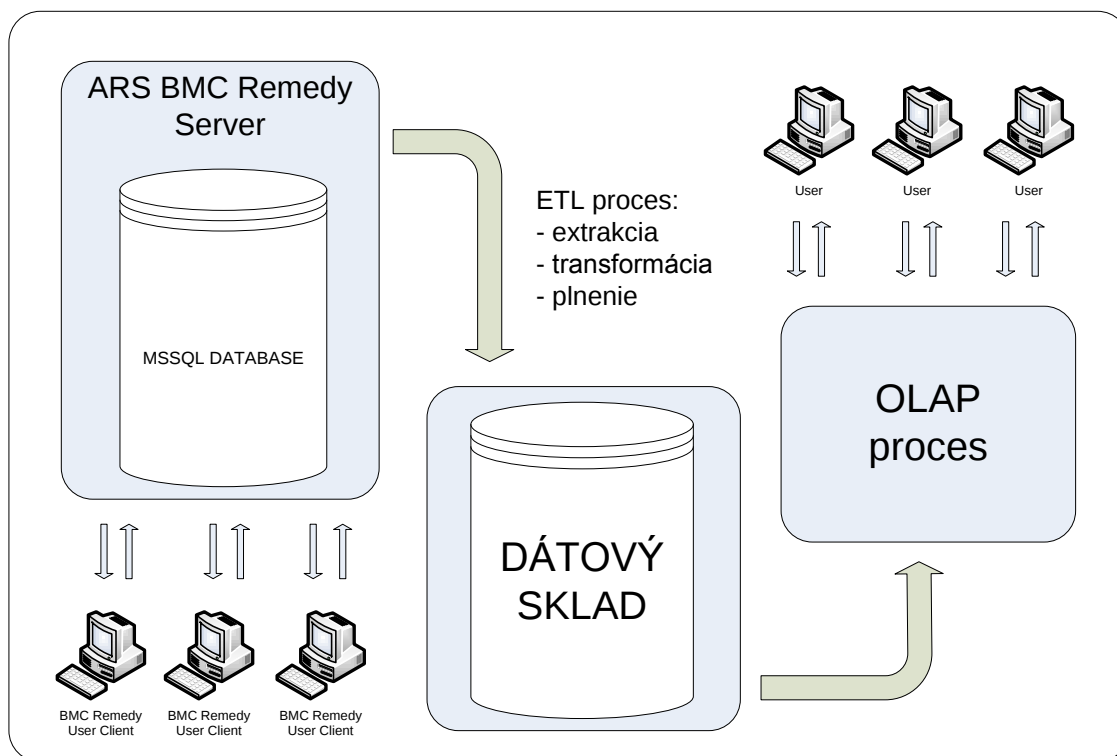
Na základe teoretických východísk v prvej kapitole a po následnej analýze súčasného stavu popísaných v tejto kapitole by sa dalo povedať, že rozhodnutie manažmentu bolo síce cielené, ale z pohľadu návrhu databázy a spôsobu riešenia analýzy nad dátami neefektívne. Riešenie je implementované pomocou voľne dostupných riešení, tzn. investícia do softwaru a hardwaru bola pomerne nízka. To sa však môže odraziť v kvalite analýzy, štatistík a vôbec rozsahu, ktoré tieto voľne dostupné, nekomerčné riešenia poskytujú. Bola tu vlastne snaha vybrať určité dáta z prostredia MSSQL a pretransformovať a uložiť ich do databázy MySQL a tým simulovať funkciu dátového skladu. A tu nastáva hlavný problém, slovo „určité“.

Platforma MSSQL Server 2008, ktorá bola použitá i v návrhu riešenia v tejto práci, ponúka obrovské možnosti, integračné služby, analytické nástroje, Business Intelligence vrátane samotného data miningu a taktiež nástroje pre vizualizáciu. Je to platforma síce komerčná, ale v tejto firme dominujúca voči ostatným komerčným riešeniam..

Samotný ARS systém (BMC Remedy User) totižto beží na platforme MSSQL, tzn. aktuálne použitým riešením prichádzame o možnosť využívať plnú funkcionality analytických nástrojov, ktoré sú priamo obsiahnuté v MSSQL servi. Okrem výhod, ktoré táto platforma ponúka je taktiež potrebné dodať, že prakticky odpadá problém s kompatibilitou pri pripájaní do tejto databázy napr. prostredníctvom ODBC. Počas analýzy za použitia MySQL ODBC ovládača totižto dochádzalo k neustálým výpadkom pri používaní sql queries z MS Access alebo MS Excel cez MS Query, či už na kvôli nízkemu timeout zo strany MySQL servra, alebo problémov medzi platformami samotnými a taktiež samotného ovládača.

Pokiaľ by sme vytvorili dátový sklad (podľa definície splňujúci subjektovú orientáciu, integrovanosť, časovú variabilitu a nemennosť) vedľa tejto operačnej databázy a za použitia ETL procesu, mohli by sme nad týmto dátovým skladom vykonávať OLAP analýzy, používať v nich pokročilé pohľady, dátové kocky, a následne dáta reprezentovať v reportovacích systémoch. Výsek iba „iba pre nás potrebných dát“ a uložiť ich znovu do ďalšej operačnej databázy nám znemožňuje vykonávať pokročilé analýzy a pohľady, ako i samotný data mining. Potrebujeme totižto v dátach dolovať, ale nie v obmedzenej množine dát, ako je to v momentálne dostupnom riešení. Preto nám neexistencia skutočného dátového skladu bráni vo

využívaní toho, čo nám samotný Business Intelligence ponúka. Na nasledujúcom obrázku je znázornené, aký spôsob by skutočne firme mohol zabezpečiť prostredie pre kvalitnú analýzu nad dátami a využívať tak tieto nástroje. Projekt dátového skladu nemôže byť neúspešný.



Obr. 21: Ideálne riešenie dátového skladu v Acision (vlastné spracovanie)

V snahe vyhnúť sa zaťažovaniu MSSQL systému sql príkazmi uložených pohľadov, ktoré aktuálne slúžia k analýze firemných procesov, by sme mohli MySQL databázu považovať za pokus o skutočný dátový sklad, avšak ten by nespĺňal už podmienku nemennosti. Taktiež implementácia plnenia tejto databáze dátami za použitia skriptovacieho jazyka PHP nie je vhodná. Tu chýba proces ETL. V návrhu riešenia sa pokúsime vykonať analýzu nad dátami uložených v databázi MySQL.

3. Návrh vlastného riešenia

V návrhu riešenia som sa pokúsil na základe možností použiť techniku data miningu nad databázou MySQL, v ktorých sú uložené údaje o všetkých procesoch popisovaných v predošlých kapitolách. K jednotlivých databázam som pristupoval pomocou rozhrania ODBC. Pracoval som v prostredí SQL Business Intelligence Development Studio, ktorý je súčasťou súčastou MSSQL Server 2008. Vlastný návrh riešenia som rozdelil do piatich fáz:

1. Definícia problému.
2. Príprava dát.
3. Výber algoritmu a modelu.
4. Učiacia fáza aplikovaná na dosiaľ existujúcich prípadoch.
5. Testovanie modelu, analýza a predikcia nových prípadov.

Definícia problému, je odpoveďou na otázku „čo chcem vlastne zmerať?“ Už v tejto fázi je nevyhnutné uvažovať, či je skutočný zámer realizovateľný a bude mať v konečnom dôsledku nejaký prínos. Pri riešení problému som sa následovne presunul do fáze 2, čo je zhromažďovanie relevantných dát k analýze. V drtivej väčšine prípadov som sa však musel vrátiť k predchádzajúcej fáze, pretože dáta buď neboli dostupné, neboli úplné, validné, bolo ich málo k samotnému meraniu alebo jednoducho neexistovali. Niektoré prípady spomeniem v samotnom závere tejto práce.

Úprava dát do podoby, v akej mohli byť merané spadala do fáze prípravy dát, a dáta boli transformované na požadovaný tvar už v pri tvorbe samotného sql príkazu, alebo v prostredí procesoru MS Excel (doplnenie chýbajúcich hodnôt, vzorkovanie, zmena kvalitatívnych atribútov na kvantitatívne a pod.). Výsledkom tejto fáze boli štrukturované tabuľky, ktoré boli importované v prostredí MSSQL serveru do testovacej databáze. Táto databáza bola potom použitá ako zdroj dát v analytických službách SQL BI Development Studio. V tomto prostredí som potom výberom jednotlivých slúpcov tabuliek určil, ktoré dáta použijem vo svojom modeli a aký algoritmus zvolím pre fázu učenia. Nie všetky algoritmy sú totiž vhodné pre každý typ úlohy. Nasledujúci postup vysvetľuje jednotlivé kroky vedúce k riešeniu. Použitý príklad je viac teoreticky

orientovaný, ale dáta použité v tomto postupe sú z extrahované z reálneho prostredia spoločnosti a použitý algoritmus dosahuje priaznivé výsledky, preto som vybral práve tento prípad.

3.1. Definícia problému

Vyskytne sa problém, a regionálny inžinier potrebuje určiť, ako moc je tento problém urgentný. Bude sa teda snažiť vytvoriť incident a priradiť mu určitú prioritu z množiny [1, 2, 3, 4, A, B, C, E1, E2, P0, P1, P2, P3, Q, S1, S2, S3], kde E1 a E2 znamenajú stav pohotovosti.

3.2. Príprava dát

Zo systému ARS som vybral položky, ktoré sa týkajú určitého incidentu, alebo incidentov. Zaujímali ma nasledujúce hodnoty :

- idINC (identifikácia incidentu).
- Priority (priorita incidentu).
- Emergency (pohotovosť).
- WeightPrior (váha).
- Target_Date_Response (čas na prvú odpoveď).
- Target_Date_Workaround (čas na doručenie dočasného riešenia).
- Target_Date_Solution (čas na doručenie finálneho riešenia).

V prostredí MS Excel som za použitia nástroja Microsoft SQL Query pripojením do databáze MySQL cez ODBC vykonal SQL príkaz, ktorý by sa dal slovne vyjadriť nasledovne:

Zobraz všetky incidenty vytvorené v rozmedzí rokov 2007-2009, kde hodnoty časov na Response, Workaround a Solution nie sú prázdne, a u týchto incidentov zobraz identifikačné číslo incidentu, prioritu, pohotovosť, váhu, čas ktorý bol daný na doručenie Response, Workaround a Solution (v minútach).

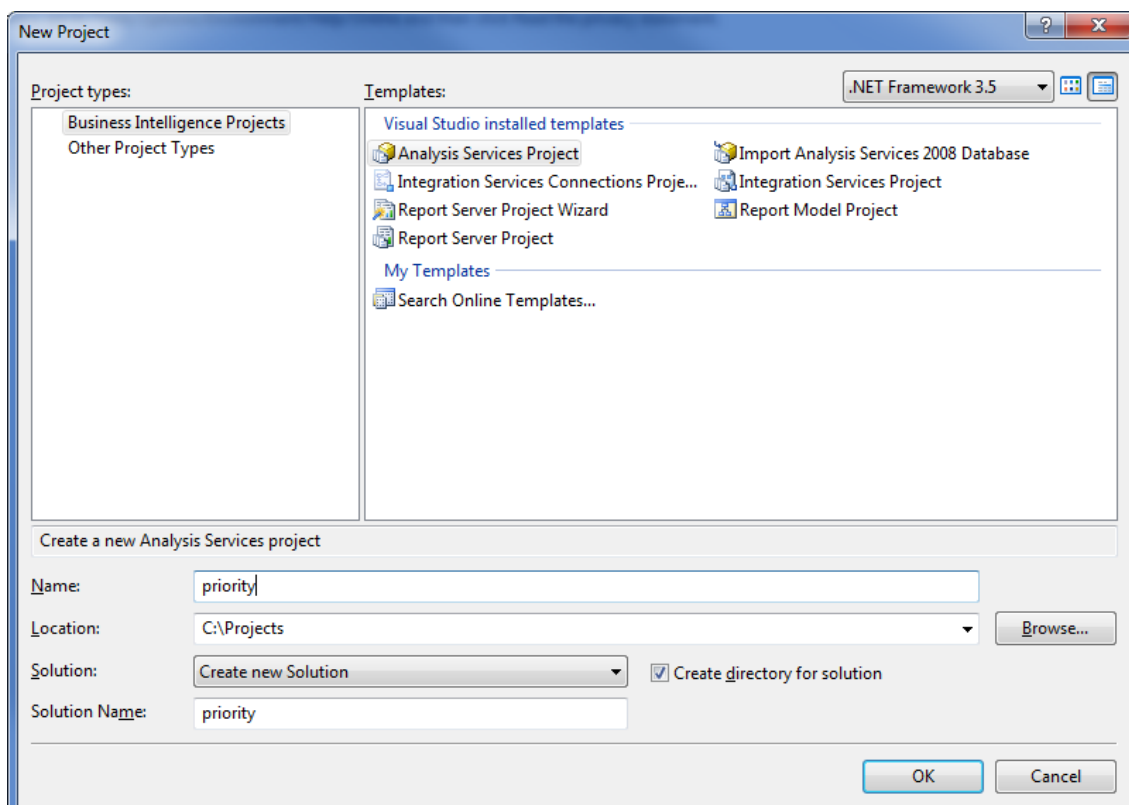
Ukážka tohto príkazu:

```
SELECT
idINC,
Priority,
Emergency,
WeightPrior,
TIMESTAMPDIFF(MINUTE, DateCreate, Target_Date_Response),
TIMESTAMPDIFF(MINUTE, DateCreate, Target_Date_Workaround),
TIMESTAMPDIFF(MINUTE, DateCreate, Target_Date_Solution)
FROM
incidents
WHERE
DateCreate > "2007-01-01 00:00:00"
AND DateCreate < "2009-12-31 23:59:59"
AND Target_Date_Response IS NOT NULL
AND Target_Date_Workaround IS NOT NULL
AND Target_Date_Solution IS NOT NULL
```

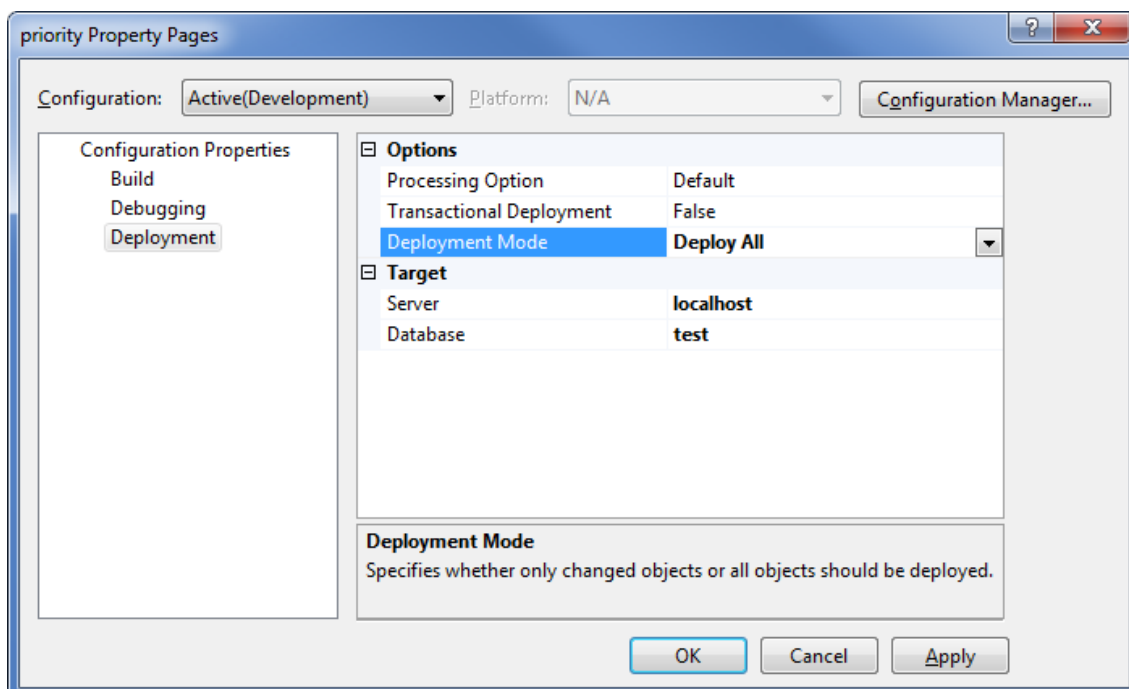
Výstup príkazu je reprezentovaný v 7 stĺpcoch vo forme tabuľky. Tú som importoval do prostredia MS Excel. Jednotlivé hodnoty však bolo potrebné prečistiť a vyfiltrovať za pomoci dostupných nástrojov a filtrov. Vyfiltrované boli napr. záznamy obsahujúce chýbajúce hodnoty, a taktiež všetky hodnoty priorít boli prevedené na kvantitatívne hodnoty (tzn. priority 1-S3 na čísla od 1 po 14). Takto upravené dáta sú uložené v dokumente MS Excel a sú pripravené na spracovanie.

V prostredí SQL Server 2008 som vytvoril pokusnú databázu „test“. V nej som vytvoril tabuľku s názvom „Metrika02B“, do ktorej som za použitia SQL Server Import and Export Wizard naimportoval všetky záznamy z dokumentu MS Excel. V tejto tabuľke sú teda pripravené dáta na spracovanie v prostredí SQL Business Intelligence Development Studio.

V tomto prostredí je potrebné vytvoriť projekt typu **Analysis Services Project** zo zložky BI projektov (Obr. 22), ktorý som pomenoval „priority“. Pre zrýchlenie nasadenia projektu na analytickú server a zrýchlenie ladenia som zmenil metódu nasadenia. V okne **Solution Explorer** cez miestnu ponuku som aktivoval dialóg **Properties** (Obr. 23) a v ňom zložku **Deployment**. V tejto zložke je potrebné nastaviť parameter **Deployment Mode** na hodnotu **Deploy All**.



Obr. 22: Vytvorenie projektu typu "Analysis Services Project" (vlastné spracovanie)



Obr. 23: Nastavenie módu nasadenia (vlastné spracovanie)

Podobne ako u OLAP analýz, i v tomto prípade je postup budovania projektu naznačený poradím zložiek v okne vývojového prostredia Solution Explorer v pravom hornom rohu. Pracoval som so zložkami: Data Sources, Data Source Views, Mining Structures. Ako dátový zdroj poslúžila databáza „test“ a tabuľka „metrika02B“.

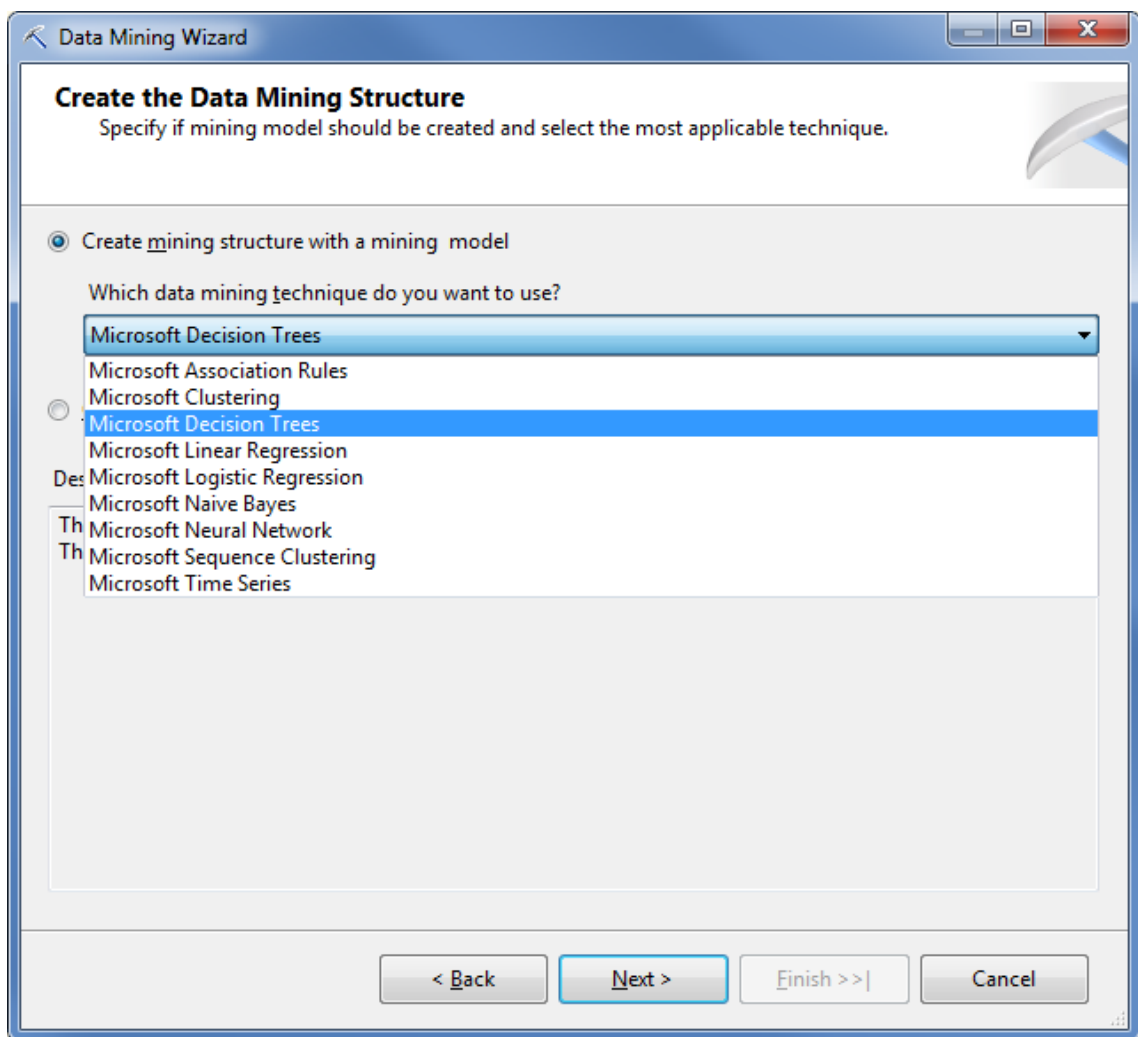
Na záložke **Data Source Views** som vybral tabuľku „Metrika02B“. Pohľad som nazval Test. V prostrednom okne vývojového prostredia sa zobrazil diagram návrhovej štruktúry vybraného pohľadu na dátové zdroje. Pomocou položky **Explore Data** miestnej ponuky som si mohol prehliadnúť vybrané údaje (Obr. 24). Po prípravných fázach definovania zdroja som teda mohol pristúpiť k vytvoreniu samotného data miningového modelu.

idINC	Priority	Emergency	WeightPrior	Response	Workaround	Solution
189833	6	0	30	2396	324501	433881
189834	7	0	20	3797	62387	526067
189835	1	1	50	15	240	15610
189838	5	0	40	938	3365	62196
189840	1	1	50	30	480	102545
189842	6	0	30	2080	21760	132700
189844	6	0	30	1801	52112	154382
189845	5	0	40	1380	2940	56220
189848	3	0	30	30	41119	540585
189849	5	0	40	30	3202	122875
189852	5	0	49	1272	43350	96720
189854	5	0	40	1573	3163	121598
189855	6	0	30	1514	23046	133986
189856	5	0	40	2167	12306	459066
189857	3	0	30	2330	2750	495244
189858	1	0	50	30	480	60480
189860	3	0	30	700	1120	488897
189862	6	0	30	30	360	133753
189864	6	0	30	1318	20998	131938
189865	6	0	30	1038	19518	123228
189869	7	0	20	787	2377	1151700
189870	2	0	40	20	720	4320
189871	3	0	30	119	129917	161477
189872	3	0	30	240	24750	125400
189873	2	0	40	1379	74969	302549
189876	3	0	30	1410	12960	262140

Obr. 24: Pohľad na dátový zdroj (vlastné spracovanie)

3.3. Výber algoritmu a modelu

Na záložke **Mining Structures** je potrebné aktivovať Data Mining Wizard. Tento sprievodca mi umožnil vytvoriť data miningový model z existujúcej relačnej databáze. V tomto prípade som čerpal dáta z databáze „Test“. V sprievodcovi som teda označil voľbu **From existing relational database or data warehouse**. Nasledoval výber algoritmu. Tu som sa rozhodol pre nevyvážené rozhodovacie stromy, teda algoritmus **Microsoft Decision Trees** (Obr. 25).



Obr. 25: Výber algoritmu (vlastné spracovanie)

V ďalšom kroku som vybral tabuľku, z ktorej som zdrojové dáta čerpal. V tomto prípade to bola „Metrika02B“. Tu je potrebné určiť primárny kľúč, tzn. taký, pomocou ktorého bude jednoznačne identifikovaná každá entita. Tu som teda vybral idINC.

Zostáva označiť hodnoty, ktoré použijem ako vstupy a hodnotu, ktorú budem predikovať. Čiže vstupy budú Emergency, Response, Solution, WeightPrior a Workaround, predikovaná hodnota bude Priority (Obr. 26).

Data Mining Wizard

Specify the Training Data
Specify the columns used in your analysis.

Mining model structure:

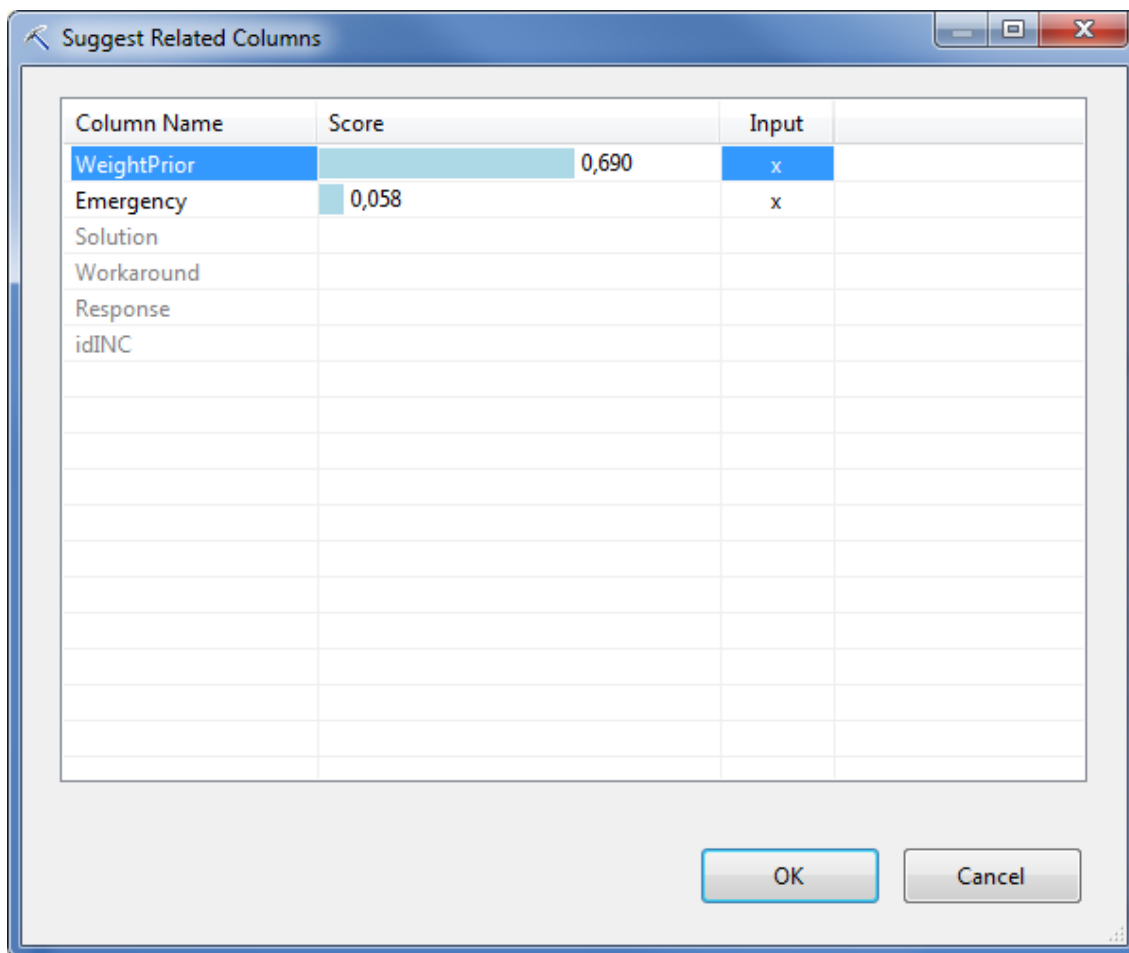
☒	Tables/Columns	Key	☐ Input	☐ Predict...
☒	metrika02B			
☒	Emergency	☐	☒	☐
☒	idINC	☒	☐	☐
☒	Priority	☐	☐	☒
☒	Response	☐	☒	☐
☒	Solution	☐	☒	☐
☒	WeightPrior	☐	☒	☐
☒	Workaround	☐	☒	☐

Recommend inputs for currently selected predictable:

< Back Next > Finish >> | Cancel

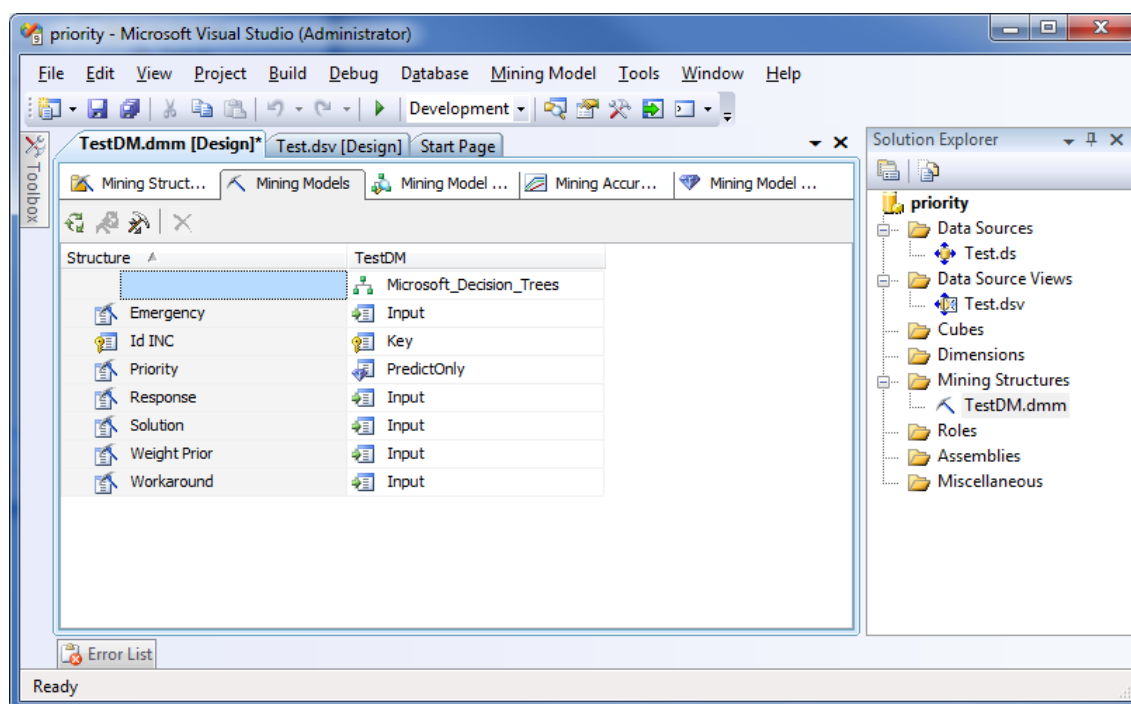
Obr. 26: Výber vstupných a predikovaných atribútov (vlastné spracovanie)

Tento sprievodca takisto dokáže určiť, pomocou možnosti **Suggest**, ktorá z vybraných vstupných hodnôt bude mať najväčší vplyv na hodnotu, ktorá bude predikovaná. Ukázalo sa, že najväčší vplyv na výsledok predikcie/výpočtu bude mať atribút WeightPrior (Obr. 27).



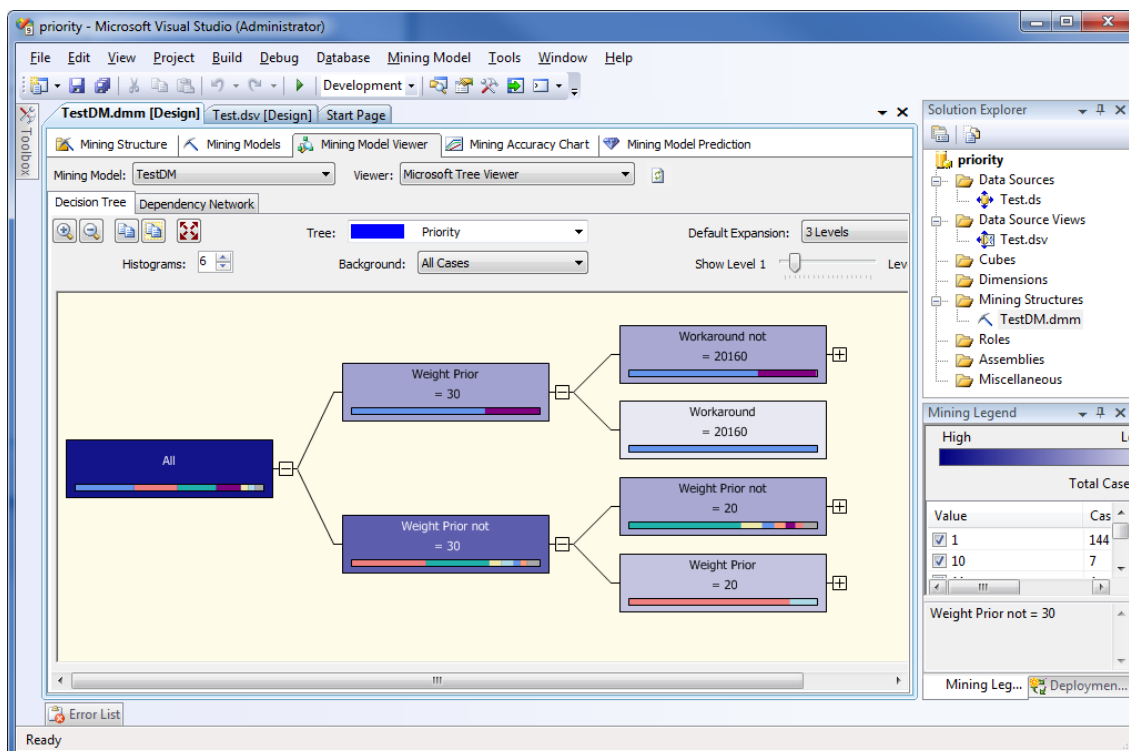
3.4. Učiaci fáza aplikovaná na dosiaľ existujúcich prípadoch

Okno pracovnej obrazovky sa rozdelilo na záložky: **Mining Structure, Mining Models, Mining Model Viewer, Mining Accuracy Chart, Mining Model Prediction**. Záložka Mining Structure obsahuje zoznam atribútov, záložka Mining Models (Obr. 28) obsahuje vybrané vstupné stĺpce, predikovaný a kľúčový stĺpec pre každý data miningový model. Po nastavení všetkých parametrov som mohol data miningový model zostaviť a zaviesť na server cez položku **Build → Deploy Solution**.



Obr. 28: Záložka Mining Models (vlastné spracovanie)

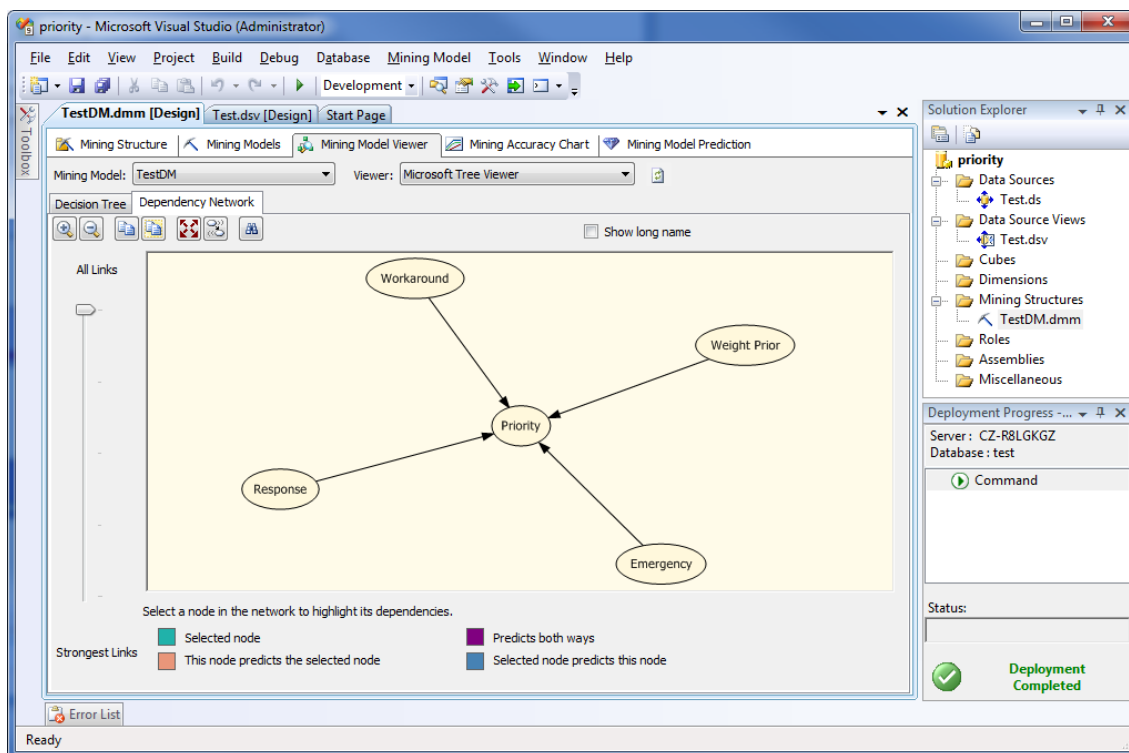
Po zostavení modelu sa dej presunie na záložku Mining Model Viewer. Záložka je riešená univerzálne pre prehliadanie výsledkov činnosti všetkých algoritmov. V mojom prípade som použil Microsoft Decision Trees. Na zobrazenom diagrame nevyváženého rozpadového stromu (Obr. 29) je možné analyzovať hĺbku jednotlivých závislostí a vzťahy vo vnútri nájdennej hierarchie. Čím bola farba políčka tmavšia, tým bola pravdepodobnosť výskytu tejto vlastnosti vyššia. Každý obdĺžnik diagramu stromovej štruktúry obsahuje i pomerový indikátor. Pomocou symbolov + a – je možné rozbaľovať a zbaľovať ďalšie úrovne hierarchie. Pomer závislosti výskytu predikovanej veličiny sa dá približne určiť z pomerového stĺpcového diagramu.



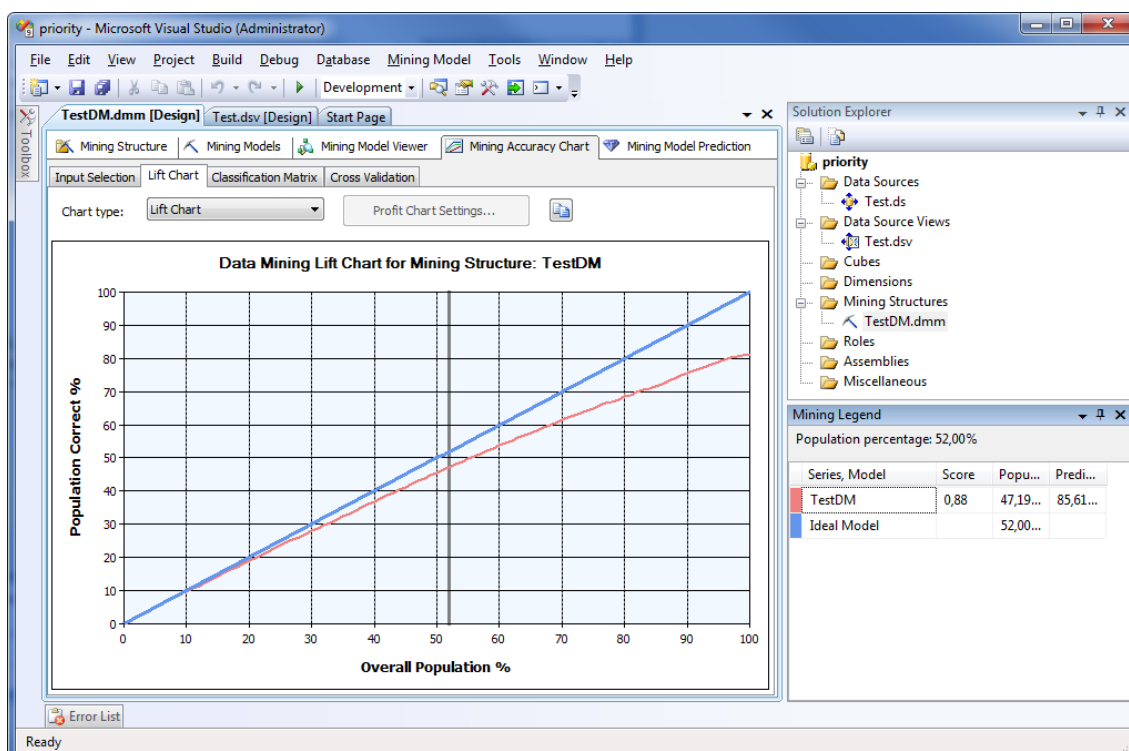
Obr. 29: Diagram rozpadového stromu (vlastné spracovanie)

Práca s grafom závislostí na záložke **Dependency Network** (Obr. 30) bola úplne intuitívna. Vo forme prehľadného diagramu sa zobrazí, na čom a v akej miere závisí predikovaná veličina. Posúvaním potenciometru v ľavej časti obrazovky sa dajú slabšie väzby postupne eliminovať, takže nakoniec zostanú len tie najsilnejšie závislosti.

Záložka **Mining Accuracy Chart** (Obr. 31) poskytuje možnosť zobrazovania prehľadných Lift Chart grafov pre testovanie data miningových modelov. Princíp grafu je jednoduchý. Najprv sa zoradia výsledky podľa klesajúcej pravdepodobnosti predpovedí a porovnávajú sa s náhodným výberom, čo je teoreticky najhorší možný prípad s najlepším teoretickým možným výberom. Krivka pre správne navrhnutý model by samozrejme mala byť čo najbližšie k ideálnej krivke a jej sklon by sa mal plynule znižovať. V opačnom prípade je nesprávne zvolený algoritmus, alebo zle navrhnutý model.



Obr. 30: Diagram závislostí atribútov (vlastné spracovanie)



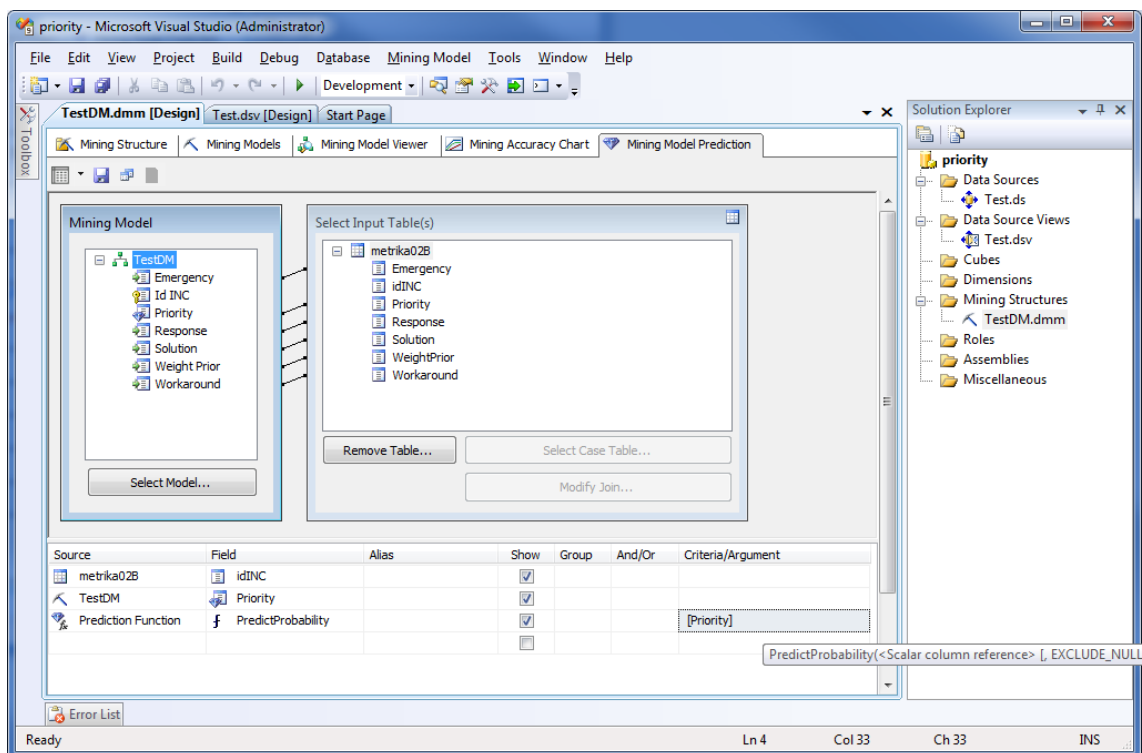
Obr. 31: Mining Accuracy Chart (vlastné spracovanie)

3.5. Testovanie modelu, analýza a predikcia nových prípadov

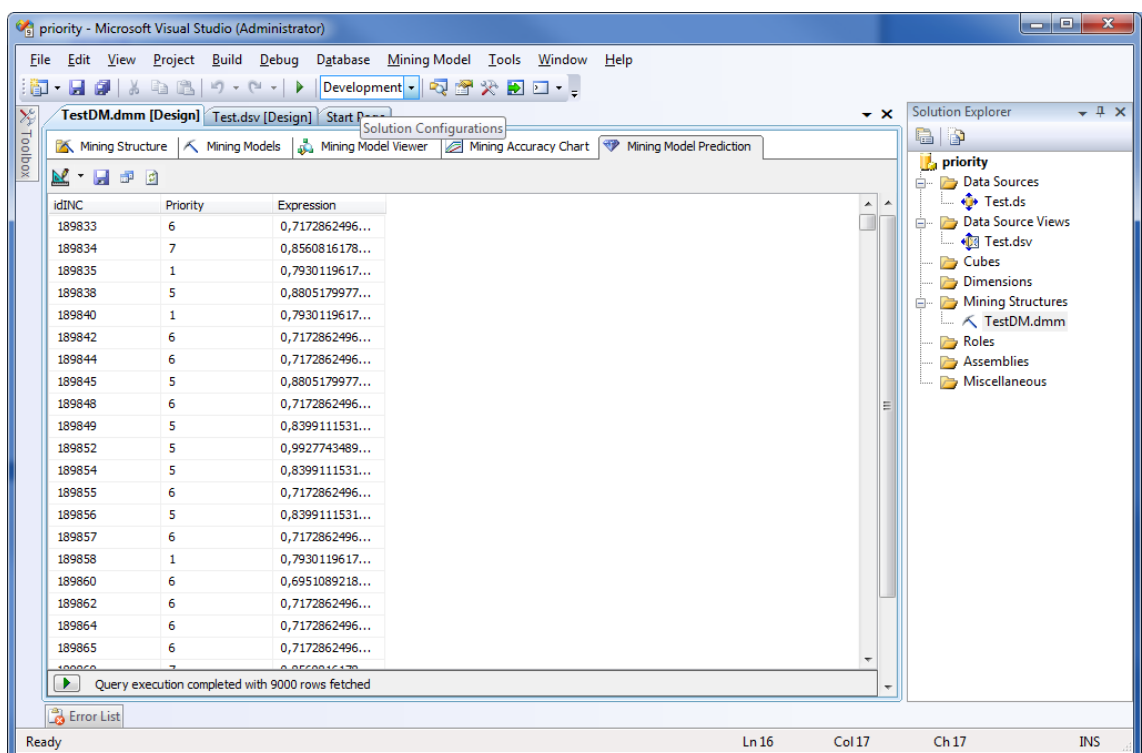
Návrh SQL príkazu pre predikciu sa odohrávala v záložke Mining Model Prediction. Ako zdroj dát pre predikciu som použil rovnakú tabuľku ako pre tréning modelu. Na obrazovke sa zobrazili dve tabuľky, v jednej sú stĺpce z modelu a v tej druhej sú stĺpce z predikčnej tabuľky. Pre návrh predikčného SQL príkazu slúži spodná časť obrazovky, kam sa dá presúvať pole z modelu, prípadne predikčnej tabuľky. Z predikčnej tabuľky som ako prvú hodnotu presunul kľúčový stĺpec, ktorý popisuje jednotlivé prípady, teda idINC. Z modelu som použil stĺpec Priority, ktorého hodnotu som potreboval predikovať. Najprv som predikoval pravdepodobnosť každého prípadu. V treťom riadku som nastavil Prediction Function, field PredictProbability. Do stĺpca Kritéria/Argument som zasal hodnotu Priority (Obr. 32). Na základe tohoto postupu mi sprievodca vlastne vytvoril nasledujúci SQL príkaz.

```
SELECT
    t.[idINC],
    [TestDM].[Priority],
    PredictProbability([Priority])
From
    [TestDM]
PREDICTION JOIN
    OPENQUERY([Test2],
        'SELECT
            [idINC],
            [Priority],
            [Emergency],
            [WeightPrior],
            [Response],
            [Workaround],
            [Solution]
        FROM
            [dbo].[metrika02B]
        ') AS t
ON
    [TestDM].[Priority] = t.[Priority] AND
    [TestDM].[Emergency] = t.[Emergency] AND
    [TestDM].[Weight Prior] = t.[WeightPrior] AND
    [TestDM].[Response] = t.[Response] AND
    [TestDM].[Workaround] = t.[Workaround] AND
    [TestDM].[Solution] = t.[Solution]
```

Po prepnutí do okna **Results** som si následne mohol prehliadnuť výsledky predikcie (Obr. 33). Pre každý konkrétny prípad je v tabuľke jednak výsledok predikcie a jednak koeficient zhody a tak som si mohol overiť hodnovernosť predikcie každého konkrétného prípadu.



Obr. 32: Návrh predikcie – Design (vlastné spracovanie)



Obr. 33: Návrh predikcie - Result (vlastné spracovanie)

Keď som nahradil funkciu Predict Probability funkciou Predict, následne sa vo výsledkoch zobrazili údaje pre každý prípad, tieto som mohol následne vyexportovať do databázovej tabuľky (VysledokPredikcie), pretože ma logicky zaujímalo, na koľko bola predikcia úspešná, nasledovalo posúdenie úspešnosti predikcie. Tabuľka je uložená v rovnakej databázi ako zdrojová tabuľka metrika02B, takže som jednoducho mohol porovnať výsledky reality a predikcie. Jednoduchými SQL príkazmi som najprv vytvoril tabuľku PorovnaniePredikcie a následne som do nej uložil výsledky „zosypaním“ údajov z oboch tabuliek a vypísal obsah tejto tabuľky.

```
CREATE TABLE PorovnaniePredikcie
(
    idINC float,
    realita float,
    predikcia float,
    pravdepodobnost float
);

INSERT INTO PorovnaniePredikcie (idINC, realita, predikcia,
pravdepodobnost)
SELECT
m.idINC,
m.Priority,
p.Priority,
p.Expression
FROM metrika02B AS m, VysledokPredikcie AS p
WHERE
m.idINC=p.idINC;

SELECT * FROM [test].[dbo].[PorovnaniePredikcie];
```

Nasledujúcim príkazom zistil, v koľkých prípadoch sa algoritmus zmýlil.

```
SELECT COUNT(*) AS Chybna_Predikcia
FROM [test].[dbo].[PorovnaniePredikcie]
WHERE realita <> predikcia;
```

Výsledkom tohoto príkazu bolo číslo 1768, toľko bolo nesprávne určených prípadov z 9000 záznamov. Percento správne predikovaných prípadov môžeme teda v percentách vyjadriť ako $100 - (100 * 1768) / 9000 = 80.36$, takže to je i celková úspešnosť predikovaných prípadov. Výhodou prostredia SQL BI Development studio je, že je tu možnosť na danej množine vyskúšať si i iné algoritmy a zmenou jedinej položky vidieť

výsledky toho istého modelu za použitia iného algoritmu. V tomto prípade som v podstate rozradľoval incidenty podľa priorít do 14 skupín, takže výber algoritmu nevyvážených rozhodovacích stromov sa osvedčil. Taktiež výber algoritmu neurónových sietí dával dobré výsledky vzhľadom na predchádzajúci algoritmus, ale rozhodovacie stromy si stále držali svoju pozíciu.

Presnejšie výsledky by sa dali dosiahnuť doplnením ďalších kritérií, ako napr. dopad na zákazníka, urgencia na vyriešenie problému zo strany zákazníka napr. počet telefónnych hovorov, emailov, počet eskalácií zo strany zákazníka. To by vyžadovalo ďalšiu analýzu nad dátami, tvorbu komplexnejších SQL pohľadov, ale vo všeobecnosti to znamená nájsť všetky relevantné kritéria, ktoré ovplyvnia výslednú hodnotu. Potenciálny model by sa teda približoval ku skutočnosti.

Značnou prekážkou pri aplikovaní data miningu vo firme boli problémy získať dáta také, ktoré by pomohli k vyriešeniu nejakého problému. Spomením pár príkladov: Vo firme máme školiť, školiť príde do do nevyškoleného tímu, učí 2 týždne, oddíde. Na základe týchto poznatkov som sa pokúšal hľadať dáta incidentov v databázi, avšak márne. Buď dáta v počte vyriešených incidentov v dobe pred a po školení neboli správne vyplnené, neexistovali, atp. Taktiež samotné zistenie, že školiaci tím školí iba nových zamestancov mi znemožnilo vykonávať takéto merania a následný data mining.

Najväčšou prekážkou však ostáva fakt, že dáta pochádzajú z operatívneho prostredia, tzn. záleží, ako pravdivo support, sustaining alebo i R&D údaje zadáva, od toho závisí kvalita údajov. Tu sa vždy potvrdilo pravidlo, že nekvalitné (nehodnoverné) dáta budú produkovať i nekvalitnú analýzu.

Záver

Cieľom tejto práce bolo popísať firemné procesy v spoločnosti Acision a zhodnotiť aktuálny stav vybranej skupiny špecialistov s využitím data miningu. V prvej časti boli tieto procesy detailne vysvetlené ako i náplň práce jednotlivých špecialistov podľa zaradenia. Procesy sú navzájom previazané a preto je spolupráca medzi jednotlivými tímami nevyhnutná. Taktiež bola predstavená aplikácia, ktorá tieto procesy zaznamenáva a zároveň podporuje a interakcia jednotlivých špecialistov napomáha k jednoduchšiemu organizovaniu práce. Tento cieľ bol splnený.

V druhej časti tejto práce bol popísaný spôsob, ako je aktuálne riešená analýza firmy nad týmito procesmi a za použitia akých prostriedkov. Tu bol očividný problém, firma sa totižto rozhodla postupovať cestou čo najmenších nákladov, čo sa taktiež odráža v jednoduchosti ale neefektívnosti daného riešenia. Tu by som jednoznačne doporučil projekt dátového skladu, ktorý nemôže byť neúspešný. Investícia do komerčného, ale kvalitného riešenia je na mieste. V závere kapitoly som aktuálne riešenie zhodnotil a navrhol, akým smerom by sa mala firma uberať, pokiaľ chce produkovať kvalitné analýzy a reálne čísla.

V poslednej časti som sa pokúsil aplikovať techniky a algoritmy používané v procese data miningu nad databázou MySQL, ktorá momentálne slúži k produkovaniu štatistík, je prístupná za účelom tvorby cielených SQL pohľadov manažmentom. Pri definícií meraní, nad ktorými som chcel samotný data mining vykonávať, som však narazil na problémy, ktoré mi nedovolili vypracovať model so značným prínosom. Preto som na teoretickom príklade vysvetlil, akým spôsobom by som postupoval a využil tak možnosti a algoritmy používané v data miningovom procese.

Firme by som preto navrhol investovať do kvalitného, moderného a komerčného riešenia, nad ktorým sa dá vykonávať kvalitná OLAP analýza. Ďalej založiť BI tím ľudí, alebo investovať do preškolenia aktuálneho tímu BI konzultantom. Najdôležitejšie sú vstupné dáta, ktoré do systému zadávajú zamestnanci, preto je potrebné zjednotenie všetkých procesov a rázne dodržiavanie pravidiel pri zadávaní údajov do systému, pretože tie slúžia ako podklad k analýze. Taktiež prehodnotiť funkciu tímu školiťov.

Zoznam použitej literatúry

Monografické zdroje:

- (1) DOSTÁL, Petr. Pokročilé metody analýz a modelování v podnikatelství a veřejné správě, Brno : CERM, 2008. 340 s. ISBN 978-80-7204-605-8.
- (2) DOSTÁL, Petr; RAIS, Karel; SOJKA, Zdeněk. Pokročilé metody manažerského rozhodování, Praha : Grada, 2005. 166 s. ISBN 80-247-1338-1.
- (3) LUBOSLAV, Lacko. Business Intelligence v SQL Serveru 2008, Brno : Computer Press, 2009. 456 s. ISBN 978-80-251-2887-9.
- (4) LUBOSLAV, Lacko. Business Intelligence v SQL Serveru 2005, Brno : Computer Press, 2006. 391 s. ISBN 80-251-1110-5.

Internetové zdroje:

- (5) Acision [online]. 2011 [cit. 2011-04-16]. Products and Solutions. Dostupné z WWW: <<http://www.acision.com/Products-and-Solutions.aspx>>
- (6) BMC [online]. 2010 [cit. 2011-04-16]. BMC Remedy Action Request System. Dostupné z WWW: <<http://www.bmc.com/products/product-listing/22735072-106757-2391.html>>
- (7) Total Service [online]. 2008 [cit. 2011-04-16]. SLA smlouva. Dostupné z WWW: <<http://www.totalservice.cz/cesky/sluzby-a-reseni/ICT-Outsourcing/service-level.html>>
- (8) SystemOnline [online]. 2007 [cit. 2011-04-16]. Business intelligence staví na datovém skladu. Dostupné z WWW: <<http://www.systemonline.cz/business-intelligence/business-intelligence-stavi-na-datovem-skladu.htm>>
- (9) Kurzy.cz [online]. 2011 [cit. 2011-04-24]. Acision Czech Republic s.r.o. data ze statistického úřadu. Dostupné z WWW: <<http://rejstrik-firem.kurzy.cz/26772825/acision-czech-republic-sro/statisticky-urad/>>

Internetové zdroje dostupné z intranetu společnosti Acision:

- (10) Acision Intranet [online]. 2009 [cit. 2011-04-16]. Sustaining slides JAR Sep 2008 v0.3 short version. Dostupné z WWW: <<http://portal.intinfra.com/workspace/Sustaining/Sustaining%20Documents/1%20-%20Sustaining%20Mission/Sustaining%20slides%20JAR%20Sep%202008%20v0.3%20-%20short%20version.ppt>>
- (11) Acision Intranet [online]. 2009 [cit. 2011-04-16]. Customer Service Engineer Job Description. Dostupné z WWW: <<http://portal.intinfra.com/workspace/000771/00%20Shared%20Documents/Job%20Descriptions/Custom%20Services%20Engineer.doc>>
- (12) Acision Intranet [online]. 2009 [cit. 2011-04-16]. Terms of Reference Sustaining Engineer. Dostupné z WWW: <<http://portal.intinfra.com/workspace/Sustaining/Quality/Shared%20Documents/>>

8%20-%20Team%20Management/Terms%20of%20Reference/Acision-ToR-SustainingEng-2009.doc>

- (13) Acision Intranet [online]. 2005 [cit. 2011-04-16]. GTP CZ SW Developer Role. Dostupné z WWW:
<<http://portal.intinfra.com/workspace/000476/ToR/00%20Shared%20Documents/2006/CC-429-Payments/GBG/GTPCZ-SG-TOR-SWDeveloper-2005.doc>>

Zoznam obrázkov, grafov, tabuliek

Zoznam obrázkov:

Obr. 1: Hierarchia informačných úrovní (3).....	12
Obr. 2: Procesná schéma data miningu na platforme SQL Server 2008 (3).....	18
Obr. 3: Jednoduchý rozpadový strom posudzujúci, či si zákazník kúpi bicykel (3)	20
Obr. 4: Princíp odhaľovania zhlukov prípadov (3).....	21
Obr. 5: Časové rady (3).....	23
Obr. 6: Predikovaná hodnota na základe priebehu veličiny v minulosti (3).....	23
Obr. 7: Schéma neurónovej siete s jednou skrytou vrstvou (3).....	25
Obr. 8: Použitie algoritmov pre typické okruhy úloh (3)	28
Obr. 9: Portfólio real-time mobilných dátových riešení spoločnosti Acision (5).....	31
Obr. 10: Zákazníci spoločnosti Acision vo svete (10).....	32
Obr. 11: Oblasti, vzťahy, a komunikácia jednotlivých tímov (10).....	39
Obr. 12: Riešenie problému – UML diagram aktivít (vlastné spracovanie).....	43
Obr. 13: NiKiTa, záložka Incident (vlastné spracovanie)	45
Obr. 14: NiKiTa Incident, záložka Details (vlastné spracovanie)	46
Obr. 15: NiKiTa Incident, záložka History (vlastné spracovanie)	47
Obr. 16: NiKiTa Incident, záložka Resolution (vlastné spracovanie)	48
Obr. 17: NiKiTa Development Ticket, záložka Details (vlastné spracovanie)	50
Obr. 18: Databáze v Acision s.r.o. (vlastné spracovanie).....	53
Obr. 19: Statistical Tool Reporter, výber SLA (vlastné spracovanie).....	55
Obr. 20: Statistical Tool Reporter, výber Workload (vlastné spracovanie).....	56
Obr. 21: Ideálne riešenie dátového skladu v Acision (vlastné spracovanie)	60
Obr. 22: Vytvorenie projektu typu "Analysis Services Project" (vlastné spracovanie) .	64
Obr. 23: Nastavenie módu nasadenia (vlastné spracovanie)	64
Obr. 24: Pohľad na dátový zdroj (vlastné spracovanie).....	65
Obr. 25: Výber algoritmu (vlastné spracovanie)	66
Obr. 26: Výber vstupných a predikovaných atribútov (vlastné spracovanie).....	67
Obr. 27: Doporučenie pre výber vstupných atribútov (vlastné spracovanie)	68
Obr. 28: Záložka Mining Models (vlastné spracovanie)	69
Obr. 29: Diagram rozpadového stromu (vlastné spracovanie)	70
Obr. 30: Diagram závislostí atribútov (vlastné spracovanie)	71
Obr. 31: Mining Accuracy Chart (vlastné spracovanie).....	71
Obr. 32: Návrh predikcie – Design (vlastné spracovanie).....	73
Obr. 33: Návrh predikcie - Result (vlastné spracovanie).....	73

Zoznam grafov:

Graf 1: Statistical Tool Reporter, zobrazenie SLA (vlastné spracovanie).....	56
Graf 2: Statistical Tool Reporter, výstup Workload 1 (vlastné spracovanie).....	57
Graf 3: Statistical Tool Reporter, výstup Workload 2 (vlastné spracovanie).....	58

Zoznam tabuliek:

Tab. 1: Rozdiely medzi produkčnou databázou a dátovým skladoom (3).....	15
Tab. 2: Základné údaje o spoločnosti Acision Czech Republic s.r.o. (9).....	30