

Oponentský posudek diplomové práce

Téma:	Vytěžování databáze Poradny pro poruchy metabolismu
Diplomant:	Bc. Martin Senft
Vedoucí práce:	prof. RNDr. Jiří Ivánek, CSc.
Oponent:	PhDr. Vladimír Musil
Návrh klasifikace:	nedoporučuji k obhajobě

Předkládaná diplomová práce představuje problematiku data miningu, který je aplikován na databázi Poradny pro poruchy metabolismu Ústavu klinické biochemie a hematologie LF a FN Plzeň.

Práce je rozdělena na 22 stránkovou teoretickou část, tedy kapitoly 1 – 5, a 32 stránkovou část praktickou, kap. 6 – 7 a má, od kapitoly 1 až po závěr včetně 108171 znaků, což odpovídá cca 60 stranám vlastního textu v normostranách. Dále jsou 4 přílohy na 18 stranách.

V teoretické části diplomant vymezuje základní pojmy a přehledně představuje metody data miningu. Ve druhé kapitole je velmi stručně představena metoda GUHA a procedura 4FT-Miner. Zcela úplně mi není jasné, proč je systém Lip-Miner rozdělen do této kapitoly a dále do kapitoly 5. V dalších kapitolách je představeno vyhodnocení data miningu (kap. 3), a jeho nástroje (kap. 4).

V praktické části jsou specifikována data, která sloužila k vytěžení databáze Poradny pro poruchy metabolismu (kap. 6) a prezentovány výsledky práce (kap. 7).

Bohužel, kapitola 6 je opravdu nejslabší součástí této práce a je velmi znát, že ji zpracovával laik v oblasti medicíny. Je zde velké množství chyb, nepřesností až nonsensů. Tak již třeba na začátku, při určení počtu pacientů, se v práci dostáváme do problému, že nevíme přesně, kolik jich je. V kap. 6.2.1 je uvedeno, že bylo získáno 105 záznamů návštěv, z toho 75 záznamů pacientů uvedeného lékaře a zbylých 20 jeho předchůdkyně, ergo 10 záznamů/pacientů chybí ($75 + 20 = 95$). Dále je ale v textu uvedeno, přesněji v kategorizaci (kap. 6.3.4.), že pacienti s kombinací metabolického onemocnění a hypertenze byli rozděleni do dvou záznamů, a tím došlo k navýšení počtu záznamů na 115. V tuto chvíli se ale stává situace naprosto nepřehlednou... Dále je tedy počítáno někdy se 105 (věk, pohlaví, TK), někdy se 115 (OA, RA, abusus, BMI, ALT atd.) „jednotkami“. U TSH a MALB se vyskytují hodnoty 109, resp. 107. V kap. 7.4.5 je uvedeno, že bez léčby farmaky odešlo 33 pacientů ze

115 členného sledovaného souboru a že hlavní roli pro nepředepsání léků hraje věk pod 50 let. Opět se v tuto chvíli dostáváme do diskrepance, jelikož odešlo 33 pacientů ze 115, skupinu „věk“ však tvoří 105 pacientů (věk do 30 (9), 30-40 (22), 40-50 (26), nad 60 (27)).

Výběr vyšetření:

Tato kapitola je také problematická. Vypuštění různých anamnéz je nešťastné. Některé mohou být zásadní pro vznik a vývoj onemocnění. To se týká obzvláště farmakologické anamnézy. Předpoklad, že předchozí farmaceutická léčba, která byla před nasazením nového léčiva, vedla ke stabilizaci hodnot, je úsměvná ve spojení s tím, že byla nasazena nová léčba. Pokud je pacient stabilní, není třeba nasazovat nová farmaka. Samozřejmě se může jednat o farmaka na nové onemocnění, ale i v tomto případě je třeba brát pacientovu předchozí léčbu, včetně alergií na léčebné prostředky, v potaz.

Forma získaných dat:

Diplomant uvádí, že záznamy získal ve formě tištěné lékařské karty, na které začernil údaje: jméno, adresa, datum narození a rodné číslo (dále RČ). V zákoně 372/2011 Sb., o zdravotních službách je v § 51 definováno kdo a kdy má právo nahlížet do zdravotnické dokumentace. Toto je poměrně citlivá záležitost, a jelikož diplomant není ani zdravotník, ve smyslu znění výše zmíněného zákona, a u práce není souhlas Etické komise LF Plzeň či FN Plzeň, nebo podepsaný informovaný souhlas každého z pacientů, domnívám se, že se jedná o poměrně značný problém...

Dále mi také není jasné, proč nezačernil rok narození a proč z RČ začernil vše ostatní, kromě prvních dvou číslic, které jsou duplicitou roku narození. K určení pohlaví pacienta neslouží počátek RČ, ale pouze jeho třetí číslice (0 a 1 pro muže, 5 a 6 pro ženu).

Také nechápu, proč byla data z karet zapisována nejdříve do MS Excel a až poté do MS Access, nebylo by jednodušší a rychlejší mezikrok vynechat?

Také nechápu, proč by měl být výjimečným pacientem pacient jiné národnosti? Domnívám se, že měla být zmíněna spíše rasová klasifikace, která má v medicíně široké spektrum využití, např. ve farmakoterapii, některé léky nezabírají u pacientů europoidní rasy, ale u pacientů negroidní rasy mají vynikající efekt. Národnost zde nehraje roli.

Věta: „těžké poruchy metabolismu tuků jsou léčeny řadu let, většinou pak celý život“, zní velmi úsměvně. Stejně je to s větou týkající se vstupního a kontrolního vyšetření, kdy dochází

k nasazení léčby. Nedomnívám se, že by několik let byl pacient ponechán v dekompenzovaném stavu.

Několik vybraných dalších připomínek:

- Není mi úplně jasné, jak může k určení hypertenze stačit pouze údaj o systolickém tlaku. O hypertenzi se jedná i v případě tlaku 120/100 či 115/100. Jde o tzv. diastolickou hypertenzi, která může svědčit o vysoké cévní rezistenci a je třeba ji léčit (viz Guidelines European Society of Cardiology či České kardiologické společnosti pro diagnostiku a léčbu hypertenze).
- Dále je uváděn systolický tlak stejně jako diastolický a nevyplývá z textu, o kterém se zrovna mluví.
- Chybí například endokrinologické vyšetření, které může hrát velkou roli při hypertenzi.
- Diplomant někdy uvádí rodinou anamnézu jako nedůležitou a v následujícím odstavci jako důležitou.
- Je také uvedeno, že lékař zařadil výsledky do pěti skupin: obecně známé, odborně známé, překvapivé a nesouvisející. Jeden chybí (s. 25).
- V některých větách není úplně jasné, o čem se píše (např. s. 26 – druhý a třetí bod, s. 22, s. 28...).
- Často se v odstavcích, či větách, objevují stejná slova, nejsou používána synonyma či zájmena. Také je zmatek v časech. V jedné větě je čas minulý, následován přítomným a hned zase minulým (např. s. 21).
- Na s. 9 a 23 by měl být uveden v citaci Cios et al.
- V seznamu citované literatury chybí práce Dolejší, 2002 ze s. 21.
- V seznamu zkratk jich spousta chybí.
- V práci je velké množství typografických chyb – chybí tečky za odstavci či za větami, nejsou oddělené informace u odrážek atd.
- V práci je často chybně uváděn střední rod: např. jsou nějaké laboratorní vyšetření, řešení jsou drahé atd.
- Je uváděno několik způsobů zápisu stejného termínu: např. abusus, abúzus. Na s. 22 je uvedeno rozdělení pacientů z projektu STULONG do 3 skupin: obecné charakteristiky, vyšetření a neřesti s tím, že neřesti jsou např. alkohol, kouření,

kofeinové nápoje – právě tyto věci jsou uváděny pod termínem abusus, zde to však uvedeno není.

- S. 63: „míru rizikovosti aterosklerózy mají ženy obecně posunutou o deset let“. Ano to je pravdivá informace, ale zde není uvedeno kterým směrem, zda nahoru, či dolů, takže z této informace nevyplývá, jestli je toto onemocnění pro ženy méně či více rizikové.
- Překvapuje mě používání zkratky umol v textu, když může být používána oficiální zkratka μmol .
- V práci je špatně používán zápis procent, je rozdíl mezi 70% = sedmdesátiprocentní a 70 % = sedmdesát procent.
- 4. příloha, Osnova řízeného rozhovoru je označena jako příl. č. 5.

Závěr je poměrně stručný a opět se zde vyskytuje problém s počtem pacientů - 115.

Celkově na mě práce na mě působí tak, že byla medicínská data získána z karet, ale dále nebyly výsledky konzultovány s lékařem a také, že ji, ve finální verzi, lékař neviděl, protože mnoho z výše uvedených údajů by určitě nepustil dále.

Předložená diplomová práce je zpracována poměrně přehledně a má i poměrně logickou strukturu. Klíčová praktická část je ale zpracována velmi chaoticky, jak bylo zmíněno výše, a proto kazí celkový dobrý dojem z jinak zajímavé práce, a proto práci **nedoporučuji k obhajobě v současné podobě**, ale doporučuji autorovi, aby odstranil nejen formální chyby a nedostatky, ale hlavně medicínské nonsensy a práci podal znovu. Práce v této podobě s těmito vybranými daty nemá reálné praktické využití. Pokud by se jednalo o práci s imaginárním souborem dat, byly by výsledky akceptovatelné. Dále doporučuji dát práci ve finální formě ke kontrole lékaři z ÚKBH LF Plzeň, aby byly odstraněny medicínské nedostatky. Celkově doporučuji práci znovu přečíst, opravit gramatické a typografické chyby a sjednotit sloh jednotlivých kapitol. Také doporučuji dodržet zákon a opatřit si souhlas Etické komise či informované souhlasy pacientů, popř. to vyřešit způsobem, který by byl v souladu se zákonem.

V Praze, 14. 5. 2014

.....
PhDr. Vladimír Musil