

РЕЦЕНЗИЯ

Институт по информационни и
комуникационни технологии-БАН

Вх. № 110 М. 02. 2013 г.

на дисертационен труд за придъждане на образователна и научна степен "ДОКТОР"
по професионално направление 4.6. "Информатика и компютърни науки",
научна специалност 01.01.12. "Информатика",

ТЕМА: "Изследване на Data Mining модели за класификация"

АВТОР НА ДИСЕРТАЦИОННИЯ ТРУД: асист. маг. Дорина Петрова Кабакчиева

ИЗГОТВИЛ СТАНОВИЩЕТО: доц. д-р Силвия Христова Илиева, ИИКТ-БАН

1. Актуалност на проблематика, върху която работи кандидата

Дисертационният труд е в актуалната област на извличане на знания от големи обеми данни. В съвременното общество ежедневно се натрупват огромни количества данни при това от различни типове, като е все по-важно да се извлече съществената информация за специфичните нужди на потребителите. „Educational Data Mining” пък е едно от най-новите направления в Data Mining, оформило се през последните 10 години в световен мащаб и представения дисертационен труд е първи в това направление в България. Като член на екипа по качество в европейски проект IDEALIST следя всички заявки за търсене на партньори за формиране на консорциуми за участие в конкурсите на 7 Рамкова програма. За последния конкурс постъпи заявка именно в областта на „Educational Data Mining” и изграждането на компетенции в тази област в България е както предпоставка за бъдещото участие на български колектив в европейски научни проекти, така и потенциал за подпомагане на българските висши учебни заведения при предсказване на успеха на студентите им.

2. Анализ на научните постижения на кандидата по съдържанието на дисертационния труд

Дисертацията е с обем 193 страници и е структурирана в 3 глави. Прави впечатление прецизното оформление – фигурите са грижливо направени и коректно реферирани. Логическо структуриране на дисертационния труд е подходящото, като прави добро впечатление наличието на систематизирани изводи в края на всяка глава. Стилът на изложение е лесен за четене, въпреки сложността на предметната област.

Глава 1 е обзорна. Тя въвежда в областта на Data Mining, както и в двете приложни области на изследване – Educational Data Mining и радарни сигнали. Представеният обзор на Data Mining методите за класификация - откриване на знания в данни чрез обучение на класификатори, показва познаване на терминологията и задълбочени знания за съществуващите методи и подходи в областта, както и критерии за оценка и сравнение на генерираните модели за класификация.

Глава 2 съдържа избраната методика за провеждане на изследванията в дисертацията. Предложената методика за изследване от дисертанката съответства на поставените задачи. След проучване на съвременното състояние в областта се предлага подход за решаване на поставената задача базираща се на Cross-Industry Standard Process for Data Mining (CRISP-DM) методологията, избира се подходящ инструментариум, който се прилага в две различни предметни области и се анализират получените резултати. Представени са и важните дейности от предложената методика – подготовка на данни и получаване на крайна съвкупност от данни в двете предметни области.

Глава 3 представлява същностната част от дисертационния труд, където са представени и анализирани резултатите от изследването на обучението на класификатори за извличане на знания от данни, чрез избраните класификационни алгоритми, т.е. приложимостта на Data Mining моделите за класификация, в двете предметни области. Резултатите показват, че Data Mining методите и средствата могат да бъдат ефективно прилагани върху различни видове данни от различни научни области.

Цитираната литература обхваща 71 източника. Всичките са на английски език и болшинството са публикувани през последните 10 години.

3. Научни и научно-приложни приноси реализирани в дисертацията

Приемам авторските претенции за научно-приложни приноси, които са представени в следните три групи - получаване и доказване на нови факти, потвърдителни факти и приноси за внедряване, въпреки че все още биха могли да бъдат формулирани по-ясно в стегнат стил.

Важен принос на дисертационния труд е детайлното изследване на налични данни, предоставени от Български университет, като са предложени редица оригинални решения на въпроси, като: подготовка на данните; анализът им; формиране на база емпиричните умения на докторанта, формиране на крайна съвкупност от данни за студентите за изследване; генериране на Data Mining модели за класификация (обучение на класификатори) и изследване на избраните Data Mining алгоритми за класификация чрез софтуер WEKA, всички те са свързани с практическото използване на избраните data mining алгоритми. В този смисъл изследването е ново и отразява спецификата на този тип данни за нашата страна и като подкрепа на тази теза има забелязано цитиране на

публикация на докторантката по тези въпроси в международна конференция. Предложеното от дисертанката решение може да бъде разглеждано и като основа за бъдещи анализи и изследвания в тази област.

За момента не са известни изследвания, свързани с използването на data mining методи за класификацията на радарни морски цели, което може да се отчете като принос за бъдещото развитие в областта. Заслужава да се отбележи, че има забелязано цитиране на публикация на докторантката по тези въпроси в международно списание с висок импакт фактор.

Принос е и обстойното изследване и получените резултати в областта на Educational data mining , като анализът им е препоръчително да се редактира в посока по-високо ниво на абстракция и обобщение.

Изводите и резултатите от дисертационния труд могат да се използват и в други висши учебни заведения за подпомагане вземането на управленски решения.

4. Публикации по дисертационния труд

Публикациите към дисертацията са шест на брой. В тях са представени основните резултати, получени в дисертацията. Всички публикации са на английски език, като две са в списания (едната под печат) и четири в трудове на специализирани международни конференции. Четири от публикациите са в съавторство, което показва уменията за работа в екип на дисертанката. Заслужава да се отбележи, че двете публикации в списания са самостоятелни. Представена е справка за 2 цитирания през 2012 г на публикации 2 и 3 на дисертанката съответно в списания International Journal of Advanced Research in Computer Science and Software Engineering и IET Radar, Sonar and Navigation.

5. Забележки и препоръки

Със задоволство отбелязвам, че болшинството от отправените от мен (в качеството ми на предварителен рецензент) забележки по време на предварителното обсъждане на дисертационния труд в секция "Интелигентни системи" на ИИКТ са взети предвид. В предишните секции бяха отбелязани някои препоръки.

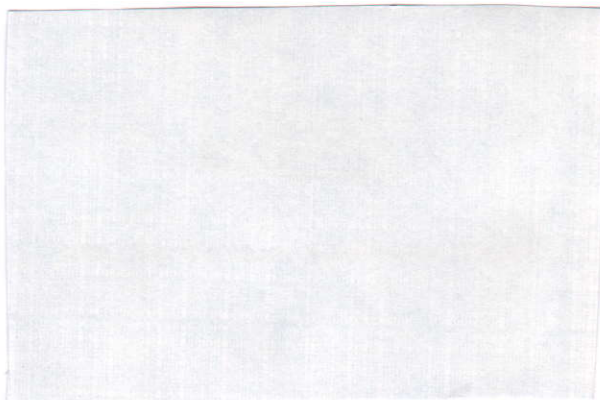
Авторефератът е изготвен според изискванията, въпреки че обем от 53 стр е малко по-голям от очакваното.

Не беше необходимо да се показват част от избраните класификатори за извличане на знания от радарни данни при изследванията, т.к. не са ефективни, а предварително да се отхвърлят.

Бих препоръчала на дисертанката да публикува изследванията си повече в списания с оглед бъдещо ѝ развитие.

6. Заключение

Представеният дисертационен труд отговаря на изискванията и има необходимите качества, установени от „Закона за развитието на академичния състав в Република България“, както и на правилника на ИИКТ-БАН за получаване на исканата образователна и научна степен ”Доктор”. В заключение, давам **положителна оценка** на представената дисертация на тема “Изследване на Data Mining модели за класификация” и препоръчвам на научно жури по обявената процедура за защита на докторска дисертация в ИИКТ-БАН, да присъди образователната и научна степен „доктор” в професионално направление 4.6. „Информатика и компютърни науки“ на автора на дисертационния труд ас. Дорина Кабакчиева.



гр. София
10.02.2013 г.