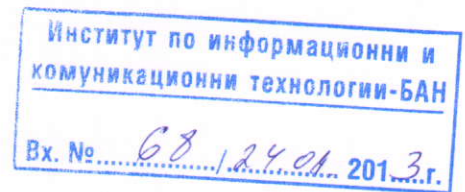


Становище



от акад. Иван Петков Попчев

за дисертационния труд за образователната и научна степен „доктор” по специалност 01.01.12 „Информатика” на Дорина Петрова Кабакчиева на тема „Изследване на Data Mining модели за класификация”

Със заповед № 143 / 27.12.2012 г. на проф. дмн Св. Маргенов – директор на ИИКТ, на основание на чл. 30, ал. 3 от Правилника за прилагане на Закона за развитие на академичния състав в Република България и Решение на Научния съвет на ИИКТ (Протокол № 13 / 21.12.2012 г.), във връзка с процедурата за придобиване на образователната и научна степен „доктор” по специалност 01.01.12 „Информатика” в професионално направление 4.6. „Информатика и компютърни науки” от **Дорина Петрова Кабакчиева** с дисертация на тема „Изследване на Data Mining модели за класификация”, съм утвърден за член на Научното жури.

Като член на Научното жури съм получил:

1. Заповед № 143 / 27.12.2012 г.;
2. Дисертация за присъждане на образователната и научна степен „доктор” (хартиен и електронен вариант);
3. Автореферат (хартиен и електронен вариант);
4. Пълен текст на публикациите по дисертацията в електронен вид.

Дисертационният труд съдържа 194 страници, 75 фигури и 36 таблици и е структуриран в три глави, заключение, списъци и библиография.

При оценката на дисертационния труд определящи са изискванията на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за неговото прилагане (ППЗ) и Правилника за условията и реда на придобиване на научни степени и за заемане на академични длъжности в БАН. Поради това ще бъдат точно предадени следните основни норми:

1. Съгласно чл. 6(3) от ЗРАСРБ „дисертационният труд трябва да съдържа **научни или научно-приложни резултати, които представляват оригинален принос в науката. Дисертационният труд трябва да показва, че кандидатът притежава задълбочени теоретични знания по съответната специалност и способности за самостоятелни научни изследвания**”.
2. Според чл. 27(2) от ППЗ дисертационният труд трябва да се представи във вид и обем, съответстващи на специфичните изисквания на първичното звено. **Дисертационният труд трябва да съдържа: заглавна страница; съдържание; увод; изложение; заключение – резюме на получените резултати с декларация за оригиналност; библиография.**

Целта на дисертационния труд, според текста на стр. 57, е „*да се генерират и изследват Data Mining модели за класификация (обучени класификатори, получени чрез прилагане на различни методи за класификация при извличане на знания от големи обеми данни), за предсказване успеваемостта на студентите в универси-*

мета на базата на техните персонални характеристики преди и след приемане в университета, и за класификация на морски обекти”.

За постигане на посочената цел са поставени следните **пет основни задачи** (стр. 57–58), в които се включват:

- разработване на методика;
- избор на подходящи софтуерни средства;
- генериране на Data Mining модели за класификация;
- оценяване на генерираните Data Mining модели;
- сравняване на изследваните Data Mining модели.

Като се следват целта и определените пет задачи, в дисертацията са включени:

- Списък на използваните термини (4–5).
- Увод (6–8).
- Обзор и актуално състояние в областта на извличане на знания от големи обеми данни (Data Minin) и методите за класификация (**Глава 1**, 9–59).
- Методика за провеждане на изследването. Подготовка на данните (**Глава 2**, 60–99).
- Резултати от изследването на Data Mining моделите за класификация (обучените класификатори) (**Глава 3**, 100–170).
- Заключение (171–181).
- Списък на фигурите (182–185).
- Списък на таблиците (186–187).
- Библиография (188–193) с включени 71 източника и 6 публикации по дисертацията.
- Декларация за оригиналност на резултатите.

Списъкът на **публикациите по дисертацията** включва 6 заглавия – всички на английски език, които могат да се класифицират така:

- **1 публикация е в електронно списание, издавано в чужбина (№ 1);**
- **1 публикация е приета за печат в списание, издавано у нас (№ 5);**
- 2 публикации са в научно-тематични поредици, публикувани в чужбина (№№ 2 и 3);
- 2 публикации са в научно-тематични поредици, публикувани у нас (№№ 4 и 6);
- 2 публикации са написани **самостоятелно** (№№ 1 и 5).

На стр. 180 в дисертацията, за труд № 2 е отбелязано цитиране в списание *IET Radar, Sonar and Navigation*.

Изследванията в дисертационнията са **непосредствено свързани** с:

- Проект № НИД НИ 2-1/2012 на УНСС – НИД (2012-2013).
- Проект № ДДВУ 02/50/20.12.2010 г. на СУ „Св. Кл. Охридски” – НИС, финансиран от Фонд „Научни изследвания” (2010-2013).
- Проект № ДТК 02/28/2009 г. на ИИКТ-БАН, финансиран от Фонд „Научни изследвания” (2009-2013).

Приносите в дисертацията могат накратко да се систематизират както следва:

1. Решена и изследвана е задача за откриване на знания в данни чрез обучение на класификатори за две съвкупности от данни за предсказване:
 - успеваемостта на студентите за два варианта на предсказваната променлива (с пет и с две стойности). Разпределението на студентите в избрани класове в най-голяма степен зависи от предсказващите променливи;
 - на класа на морски обекти, представени като номинална променлива с три стойности. Точността на предсказване на получените модели за класификация на морските цели е сравнима с резултатите, получени от изследователи в Бирмингамския университет.
2. Апробирана е методика на изследването, която включва предварителна подготовка на данните, обучение на класификатори за откриване на знания в данни и оценка и сравнение на получените модели. Потвърдена е възможността за успешно прилагане на методиката върху съвкупности от данни в различни предметни области.
3. Формулирани са насоки за бъдещите изследвания, свързани с анализирането и извличането на знания от данните за:
 - студенти чрез приложение на методите за клъстерен анализ, регресионен анализ и използване на предсказващи променливи, включващи само най-информативни променливи;
 - областта на класификацията на радарни цели чрез приложението на методите за клъстерен анализ, избор на най-подходящи променливи и методи, свързани с малкия размер на съвкупността от данни и неравномерното разпределение на обектите в класовете.

Авторефератът е в обем от 53 страници и включва основните резултати.

Като научен ръководител на ас. маг. Дорина Петрова Кабакчиева мога да потвърдя, че от 03.01.2006 г., когато тя бе зачислена за редовен докторант в ИИТ-БАН, до публичната ѝ защита докторантката премина един начален период на запознаване и обучение в новите научни направления на Business Intelligence и Data Mining, за да достигне състоянието на самостоятелен изследовател в тези направления със забележителни резултати.

Когато се оценява развитието на ас. Дорина Кабакчиева може да се маркира, че тя е завършила през 1995 г. във ВМЕИ-София специалността „Микроелектроника и технология на микроелектронното производство“. **Дисертационният труд е едно доказателство, че докторантката има задълбочени теоретични познания по съответната специалност с подчертан интерес към практически приложения и с изявени способности за самостоятелни научни изследвания.**

Получените приноси в дисертацията са изключително лично дело на ас. Дорина Кабакчиева.

По насоките на бъдещите изследвания, които обхващат едно изключително широко и дългосрочно във времето поле на изследвания, може да се **препоръча** заедно с приложението на методите на клъстерния анализ върху задължително големи обеми от данни, да се разширяват и предметните области. Това е от една страна. От друга страна, работата с по-малки съвкупности от данни, които да включват само най-информативните променливи, може да доведе до йерархично (в определен смисъл: последователно) включване на обучени класификатори, което най-вероятно представлява и определен теоретичен интерес.

Като своеобразен „резонанс“ от последната публична дискусия по дисертационния труд, в която дискусия присъстваха и терминологични нюанси, то значително внимание заслужава броят на публикациите по години и по ключови думи в Web of Knowledge и Science Direct, които към 10.01.2013 г. се представят в следните две таблици.

Ключови думи	2007	2008	2009	2010	2011	2012
data mining	937	1049	1272	886	804	570
data mining for prediction	9	11	18	19	11	6
data mining prediction	24	21	42	32	27	18
data mining for classification	15	11	15	8	12	10
data mining classification	24	27	41	20	22	18
data mining for regression	2	1	0	2	4	1
data mining regression	5	4	6	5	10	2
educational data mining	6	2	2	6	5	3
education data mining	6	3	4	4	5	5
business intelligence	61	92	106	86	66	32

Таблица 1: Резултати в Web of Knowledge по ключови думи към 10.01.2013 г.

Ключови думи	2007	2008	2009	2010	2011	2012
data mining	1043	1155	1461	1409	1401	1086
data mining for prediction	22	28	48	44	41	39
data mining prediction	22	28	48	44	41	39
data mining for classification	30	30	43	39	38	46
data mining classification	30	30	43	39	38	46
data mining for regression	4	4	10	5	12	4
data mining regression	4	4	10	5	12	4
educational data mining	5	3	5	8	12	11
education data mining	2	7	7	8	4	7
business intelligence	76	106	124	152	152	106

Таблица 2: Резултати в Science Direct по ключови думи към 10.01.2013 г.

Коментар не е нужен.

Заклучение

Може да се констатира, че са изпълнени всички изисквания на Закона за развитие на академичния състав (ЗРАСРБ), Правилника за неговото прилагане (ППЗ) и Правилника за условията и реда за придобиване на научни степени и за заемане на академични длъжности в БАН.

Всичко това ми дава достатъчно основание за положителна оценка и предлагам на почитаемото Научно жури да присъди образователната и научна степен „доктор“ по специалостта 01.01.12 „Информатика“ на ас. маг. Дорина Петрова Кабакчиева.

14.01.2013
София