

## РЕЦЕНЗИЯ

върху дисертационен труд за получаване на образователната и научна степен

**“Доктор”**

по докторантска програма

*“Автоматизирани системи за обработка на информация и управление”*

професионално направление

5.3. *“Комуникационна и компютърна техника”*

на тема

### **„Изследване на методи и алгоритми за класификация на текст в машинното обучение“**

с автор

инж. Нели Ананиева Арабаджиева - Калчева

от доц. д-р Ирена Въллова, катедра КСТ, РУ “А. Кънчев” Русе

#### **1. Актуалност на разработвания в дисертационния труд проблем в научно и научно-приложно отношение. Степен и мащаб на актуалността на проблема и конкретните задачи, разработени в дисертацията.**

Дисертационният труд е посветен на изследвания в интересно и актуално направление на обработката на информация – алгоритми за класификация на текст в машинното обучение. Днес, с развитието на ИТ сектора и Интернет приложенията почти всичко е достъпно и се обработва онлайн, се събират големи обеми от данни. За да могат да се използват те трябва да бъдат организирани по определени критерии. Търсенето на методи и алгоритми за организацията и бързото търсене в структурирани и неструктурирани данни е много актуален проблем. Машинното обучение се занимава с търсенето и създаването на алгоритми за класификация на данни, което да позволи да се съкрати времето и сложността на процеса на търсене на необходима информация, представена под формата на електронни текстове.

Целта на дисертанта е да се изследва влиянието на обема от данни при класификацията на текст на български и английски език върху различни методи и алгоритми за класификация в машинното обучение. Идеята е да се реализират и

изследват известни методи и алгоритми и на база на сравнителния анализ от получените резултати да се предложат подобряващи ги модификации.

## **2. Познава ли дисертантът състоянието на проблема и оценява ли творчески литературния материал.**

Дисертантът е приложил списък на 132 цитирани литературни източника, включващ статии в престижни специализирани списания и научни форуми, дисертации и книги в предметната област на дисертацията. 34 от тях са на кирилица (български и руски) и 98 на латиница (английски). Включени са както класически и основополагащи, така и най-съвременни публикации по темата. Вижда се, че докторантът е придобил задълбочени познания в разглежданата област и притежава необходимата теоретична основа за успешното разработване на дисертационния труд.

## **3. Избраната методика на изследване може ли да даде отговор на поставената цел и задачи на дисертационния труд.**

Формулираните цел и задачи съответстват на изискванията за дисертационен труд за придобиване на образователната и научна степен "доктор". Избраната методика е класическа: започва с обзор и анализ на съществуващи решения, разработка на концептуален модел и анализ на известните алгоритми в предметната област, описание на предлаганите усъвършенствания или нови решения от докторанта и експериментално доказване и показване на приложимостта на предлаганото решение.

## **4. Кратка аналитична характеристика на естеството и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд.**

Дисертационният труд съдържа 154 страници, включително 38 фигури и 7 таблици. Оформен е като увод, пет глави, заключение, списък с използвана литература от 132 заглавия, списък на публикациите на дисертанта, списък на фигурите и списък на таблиците, две приложения (речник на използваните термини, квантили на разпределението на Стюдънт).

В първа глава на дисертационния труд е представено и анализирано състоянието на проблема. Тя завършва с изводи, които са използвани за формулиране на целта и задачите на изследването. Втора глава разглежда теоретичните изследвания на

методите и алгоритмите за класификация на текст в машинното обучение – описани са 9 алгоритъма. В трета глава се предлага модификация на наивния Бейсов класификатор. Четвърта глава описва методиката на експерименталното изследване и последната – пета глава съдържа описание на самото експериментално изследване и резултатите от него.

От представеното изложение, формулираната цел, поставените във връзка с нея задачи и описанието на тяхното изпълнение, може да се направи заключение, че докторантката добре познава състоянието на разглежданите проблеми. Представените подходи за верифициране чрез експериментални и статистически изследвания и анализи ми дават основание да определя получените резултати и свързаните с тях приноси като достоверни и практически полезни.

**5. В какво се заключават научните или научно-приложните приноси на дисертационния труд? Формулиране и обосноваване на нов научен проблем (област). Формулиране и обосноваване на нова теория (хипотеза). Доказване с нови средства на съществуващи нови страни в съществуващи научни проблеми и теории. Създаване на нови класификации, методи на изследване, нови конструкции, технологии, препарати. Получаване и доказване на нови факти. Получаване на потвърдителни факти. Характер на приносите за внедряване: методи, конструкции, технологии, препарати, схеми и т.н. Каква е, конкретно, значимостта на тези приноси за науката и практиката?**

Дисертантът е класифицирал петте свои приноса като 3 научни и 2 научно-приложни. Имам известни колебания за принос 3, който смятам за по-скоро научно-приложен, но нямам съмнение като цяло за приносите на дисертанта. Основен и най-важен принос е предложеният модифициран алгоритъм на Наивния Бейсов класификатор, използващ разпределението на Лаплас. Неоспорими са реализацията, сравнителните анализи и статистически доказателства за поведението на различните известни алгоритми в предметната област.

**6. Може ли да се оцени и в каква степен дисертационния труд и приносите представляват лично дело на дисертанта?**

Публикациите на автора и представеното съдържание и структура на дисертационния труд дават основание да се заключи, че приносите в труда несъмнено са лично дело на дисертанта.

**7. Преценка на публикациите по дисертационния труд: брой, характер на изданията, в които са отпечатани. Какво е отражението им в науката – използване и цитиране от други автори, в други лаборатории, страни и пр.?**

Общият брой на публикациите по дисертацията е 4. Публикувани са в периода 2016-2018 година. Има три публикации на български език в сборници с доклади в конференции в България и една - на английски език в международна конференция в чужбина (на нея дисертантът е без съавтори). Две са в съавторство с ръководителя, една - в съавторство с друг колега и една самостоятелна. Не са ми известни и не са показани цитирания на публикациите на автора от други автори.

**8. Резултатите от дисертационния труд използвани ли са вече в научната и социалната практика? Има ли постигнат пряк икономически ефект и пр.? Документи, на които се основава твърдението.**

Няма данни и не са приложени документи към дисертацията, доказващи това, както и липсват цитирания на трудовете на дисертанта.

**9. Мотивирани препоръки за бъдещо използване на научните и научно-приложните приноси: какво и къде да се внедри.**

Работата по използването на научните и научно-приложните приноси на труда би трябвало да продължи. Интересно е другите подобни решения (у нас) и проекти в същата област какви алгоритми използват и дали биха използвали предложеното от дисертанта.

**10. Авторефератът направен ли е съгласно изискванията, правилно ли отразява основните положения и научните приноси на дисертационния труд?**

Авторефератът е направен съгласно изискванията и отразява основните положения и приносите на дисертационния труд.

**11. Критични бележки по дисертацията, включително и по литературната осведоменост на кандидата.**

Нямам съществени забележки към дисертационния труд. Оформен е старателно и показва голям обем на извършените научни изследвания и експерименти. Високото ниво на представяне на резултатите свидетелства за доброто познаване на предметната област от авторката.

Имам две дребни препоръки:

- В 5 глава се използват много често термините recall, precision, F-measure и други директно на английски, а във фигури 5.12 и 5.13 има и посочен превод. Отделно има речник на термините в края на дисертационния труд, който е добра идея, но няма никаква индикация при среща на термина на английски, че той е включен в този речник.
- Би било добре да се посочи в какъв тип софтуерни приложения биха намерили място предложените от дисертанта модифицирани алгоритми.

**12. Други въпроси, по които рецензентът счита, че следва да вземе отношение.**

От дисертацията и приложените към нея документи се вижда, че дисертантът е положил много труд и е работил самостоятелно над проблемите.

**13. Заключение с ясно становище да се даде ли или не научна степен.**

Въз основа на предоставените ми материали, постигнатите резултати и приноси, считам, че дисертационния труд отговаря на изискванията на Закона за развитието на академичния състав в Република България и Правилника за научни степени на ТУ-Варна. Предлагам на членовете на научното жури да присъдят на инж. Нели Ананиева Арабаджиева – Калчева образователната и научна степен “доктор” в Професионално направление 5.3. „Комуникационна и компютърна техника“.

21.07.2020 г.

.....

/доц. д-р Ирена Вълва /