

СТАНОВИЩЕ

от доц. д-р Христина Николова Кулина,

ПУ „Паисий Хилендарски“

на дисертационен труд за присъждане на образователната и научна степен „доктор“ в област на висше образование 4. *Природни науки, математика и информатика*; професионално направление 4.5. *Математика*; докторска програма *Математическо моделиране и приложение на математиката*

Автор: *Антоанета Петрова Йорданова*, докторант към Факултет по математика и информатика на ПУ „Паисий Хилендарски“

Тема: „*Приложение на дейта майнинг методи за статистическо моделиране*“

Научен ръководител: проф. д.м.н. Снежана Георгиева Гочева-Илиева, ПУ „Паисий Хилендарски“

1. Общо представяне на процедурата и докторанта

Със заповед Р33-1451/27.04.2021 г. на Ректора на Пловдивския университет „Паисий Хилендарски“ (ПУ) съм определена за член на научното жури по процедурата за защита на дисертационен труд на тема „Приложение на дейта майнинг методи за статистическо моделиране“ за придобиване на образователната и научна степен „доктор“ в област на висше образование 4. *Природни науки, математика и информатика*, професионално направление 4.5. *Математика*, докторска програма *Математическо моделиране и приложение на математиката*. Автор на дисертационния труд е Антоанета Петрова Йорданова – докторант в редовна форма на обучение към катедра „Математически анализ“ с научен ръководител проф. д.м.н. Снежана Георгиева Гочева-Илиева от ФМИ при ПУ „Паисий Хилендарски“.

За изготвяне на настоящото становище, като член на научното жури, съм получила всички необходими документи, приложени към молбата от докторанта до Ректора на ПУ „Паисий Хилендарски“ за откриване на процедура за придобиване на ОНС „доктор“. Антоанета Йорданова завършва семестриално специалност „Математика“ в СУ „Климент Охридски“ през 1992 г. През 1999 г. завършва специалност „Застраховане и социално дело“ (бакалавър и магистър) към СА „Димитър А. Ценов“, гр. Свищов. През 2015 г. придобива магистърска степен по специалност „Приложна математика“ към ПУ „Паисий Хилендарски“. През август 2017 г. е зачислена в редовна докторантура към ФМИ на ПУ. Антоанета Йорданова е участник в три проекта към НПД на ПУ, изнасяла е доклади на международни конференции, участвала е активно в работата на редовния научен семинар „Математическо моделиране и приложения“ към катедра „Математически анализ“. Владее добре английски и руски език. По настоящем Антоанета Йорданова работи като асистент към Медицинския колеж на Тракийски университет – гр. Стара Загора.

2. Актуалност на тематиката и познаване на проблема

Темата на дисертационния труд е пряко свързана с динамичното развитие на едно от най-съвременните направления в статистиката – машинното обучение. В работата се прилагат дейта майнинг методи с машинно обучение (МО) за моделиране и анализ на многомерни данни от животновъдството – основна област от аграрните науки. Такъв тип методи за емпирични данни в България досега не са използвани. Работата е посветена на изследването на 305-дневната млечна продуктивност на крави от породата Холщайн-Фрезийска в зависимост от фермата, където живеят животните и редица линейни екстериорни признаци. Тук следва да отбележим, че за разглежданата порода крави е постигната най-голяма степен на стандартизация на линейните признаци поради

контрола и организацията от страна на Световната Холщайн Федерация. Тематиката е актуална, както като с предмета си на изследване, така и с използваните за целта методи.

Представеният труд е в пълно съответствие с направлението на докторската програма и поставената тема.

3. Обща характеристика и оценка на резултатите и приносите на дисертационния труд

Дисертационният труд е с общ обем от 119 печатни страници и включва: увод, четири глави, заключение и библиография от 105 заглавия.

Глава 1 е представен предмета на изследванията, описани са използваните методи: многомерен линеен регресионен анализ, факторен анализ, регресионен анализ с главните компоненти и дейта майнинг методите: класификационни и регресионни дървета (CART), ансамблови методи: CART Ensemble and Bagging (CART EBag) и Случайни гори (RF). Направен е литературен обзор за изследванията и актуалните задачи в двете области.

Глава 2 представя резултатите от приложението на метода на класификационните и регресионни дървета (CART). В изследването се използват данни от лактации на 97 Холщайн крави, събрани в периода 2013-2014 г. от четири ферми, разположени в различни региони в България. Описано е приложението на метода CART за зависимата променлива 305-дневна млечност и 12 екстериорни признака от ординален тип. Оценен е относителният принос на на тези фактори върху млечността. Изграденият модел описва около 48% от изследваните данни.

Построен е и втори CART модел при същите фактори, като се включва допълнителен предиктор фермата. Изграденият модел описва 70% от данните. Проведеният анализ показва значението на фермата като най-съществен фактор за млечността.

Глава 3 описва моделиране на същите данни от Глава 2 с метода Случайни гори (RF). Отново са построени два модела – първия с 12^{-те} екстериорни признака като предиктори, втория – с 13, включвайки и фермата. Първият модел описва 95% от данните като подрежда факторите по относителен принос (в низходящ ред), както следва: ширина на виме, ширина на гърди, скакателни стави и походка. Построеният втори модел описва отново 95% от данните като извежда по значение на преден план фермата, а останалите предиктори подрежда в същия ред. Направено е сравнение на съответстващите CART и RF модели. Направени са изводи, относно по-добрите предсказващи свойства на методите RF в случаите на неголеми извадки. Методите са описани подробно и пълно.

Глава 4 в първата си част представя прилагането на регресия с главните компоненти за анализ на разширена извадка на данните от Глава 2 за 160 крави. Тук данните за млечността са трансформирани до усреднена млечност по ферми. След установена мултиколинеарност на предикторите, чрез факторен анализ, са получени 11 взаимноортогонални компонента. Построен линеен регресионен модел, в който извлечените главни компоненти участват като независими променливи. Полученият модел обяснява 53% от данните. Във втората част от Глава 4 с извлечените 11 главни компоненти са построени 16 модела с фактора ферма и без него с прилагане на метода CART Ensembles and Bagging. Направен е сравнителен анализ на моделите по следните показатели: R^2 , RMSE, MAE, MSE, MAPE и е избран е най-добър модел.

Всички модели са описани изключително детайлно и пълно. Резултатите са илюстрирани с таблици и графики.

Авторефератът е в обем от 32 страници и е изготвен съгласно изискванията и правилно отразява съдържанието, резултатите и приносите на дисертационния труд.

4. Публикации по дисертационния труд

Авторът на дисертационния труд е приложил списък с 3 публикации, всички в рецензирани издания (две от тях са на английски език). Публикациите са както следва – една в списание, една в сборник на международна конференция (реферирани в БД на Scopus, с SJR), една в трудове на национална конференция.

Не е приложен списък с цитиранията, но следва да отбележа, че публикацията в списание има две забелязани цитирания с импакт фактор.

Считам, че участието на докторантката в публикациите е **равностойно** на останалите съавтори. Приемам описанието на научно-приложните приноси за адекватно и коректно представени в дисертационния труд.

Публикациите по дисертацията покриват специфичните изисквания на ФМИ при ПУ съгласно чл. 36. (1), т. 9 от ПРАСПУ за придобиване на образователната и научна степен „доктор“.

5. Критични забележки и препоръки

По същество нямам критични забележки и препоръки. Впечатлението ми от представения дисертационен труд е, че Антоанета Йорданова е овладяла детайлно и в същина едни от най-съвременните методи на МО и ще може да ги прилага в много и различни сфери. Изложението в дисертационния труд е последователно, структурирано и обосновано.

Познавам Антоанета Йорданова от времето на магистърската ѝ програма в ПУ. В съвместната ни работа тя винаги е била добър и коректен партньор. Пожелавам ѝ този успешно финализиран дисертационен труд да е поредната от много успешни стъпки в бъдещата ѝ научна кариера.

ЗАКЛЮЧЕНИЕ

Дисертационният труд съдържа научно-приложни резултати, които **отговарят на всички** изисквания на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за прилагане на ЗРАСРБ и съответния Правилник на ПУ „Паисий Хилендарски“. Представените материали и дисертационни резултати **напълно** съответстват на специфичните изисквания на Факултета по математика и информатика, приети във връзка с Правилника на ПУ за приложение на ЗРАСРБ.

Дисертационният труд показва, че Антоанета Йорданова **притежава** задълбочени теоретични знания и професионални умения по темата на дисертационния труд, като **демонстрира** качества и умения за самостоятелно провеждане на научно изследване.

Поради гореизложеното, убедено давам своята **положителна оценка** за проведеното изследване, представено в дисертационен труд, автореферат, постигнати резултати и приноси, и **предлагам на почитаемото научно жури да присъди образователната и научна степен „доктор“** на Антоанета Петрова Йорданова в област на висше образование: 4. *Природни науки, математика и информатика*, професионално направление 4.5. *Математика*, докторска програма *Математическо моделиране и приложение на математиката*.

02.06. 2021 г.

Изготвил становището:

(доц. д-р Христина Кулина)