



## РЕЦЕНЗИЯ

върху дисертационен труд за получаване на образователната и научна степен "доктор" по: област на висше образование: 4. Природни науки, математика и информатика, професионално направление: 4.6. Информатика и компютърни науки, научна специалност: Информатика

**Автор на дисертационния труд:** Никола Йорданов Цайкин – докторант на самостоятелна подготовка към катедра "Математика, информатика и статистика", Аграрен университет

**Тема на дисертационния труд:** "Анализ на многомерна база данни с динамични редове за растениевъдството в България"

**Рецензент:** доц. д-р Тодорка Николова Мокрева,  
област на висше образование: 4. Природни науки, математика и информатика,  
професионално направление: 4.5 Математика, научна специалност: Математическо  
моделиране и приложение на математиката в агрономството.

определенена за член на научното жури със заповед № РД-16-581 от 21.05.2019 г. на Ректора  
на Аграрен Университет

### 1. Кратко представяне на кандидата.

Ас. инж. Никола Йорданов Цайкин е завършил специалност "Изчислителна техника" през 1995 г. във факултета по "Компютърни системи и управление" на Технически университет - София. От септември 1997 г. до юни 1999 г. работи като учител по информатика в ПГХТ - Пловдив. От август 1999 г. работи като асистент в катедра "Информатика" в АУ-Пловдив.

### 2. Актуалност на проблема.

Големи количества информация от различни области се съхраняват в бази данни в Международни и правителствени институции. Обработката на такива бази данни изискава големи изчислителни разходи. Обикновено такива бази данни са представени в Web среда с удобен интерфейс за достъп до отделни съвкупности от данни, но само в редки случаи предлагат онлайн обработката им. Тази съществена особеност налага проектирането и реализирането на модел на база данни, която да съдържа само част от данните на голямата база. Данните в базата да бъдат структурирани според изискванията на потребителя за лесен достъп и визуализация и модела да притежава висока степен на оперативност по отношение на последваща обработка.

### **3. Цел, задачи, хипотези и методи на изследване.**

**Целта**, която авторът си поставя в разработваната от него дисертация е да се създаде информационна среда, подпомагаща специалисти и експерти, които да правят аналитични оценки и вземат адекватни решения за отрасъла растениевъдство. Тази информационна среда трябва да съдържа многомерна база данни с динамични редове за различни аспекти на растениевъдното производство в България и страни от Европейския съюз. Към базата данни да има разработен софтуер за нейното обновяване и анализиране. В същото време, цялостното решение да се съхранява във файловата система на настолен или преносим компютър.

Проектирането и реализирането на многомерна база данни, съхраняваща големи количества данни от растениевъдното производство в България и страни от Европейския съюз е **обект** на дисертационната работа. **Предметът** на изследване е насочен към набор от софтуерни инструменти за автоматично обновяване на база данните, актуализирането ѝ, анализ на данните и решение за съхраняването ѝ във файловата система на РС. Предметът конкретизира обекта и е съсредоточен върху ефекта от използването на информационния модел.

Изследователската **хипотеза** на автора е, че софтуерните инструменти за математико-статистически анализ предоставят на потребителя мощен инструмент за разкриване на нови характеристики на взаимодействията между явленията и процесите и за вземане на решения.

В резултат на проведеното изследване, авторът е систематизирал подходите при изграждане на информационната среда. Чрез интегриране на данни от институционални бази данни е проектиран многомерен модел на БД съхраняващ предметно фокусирана информация с потребителски ориентиран софтуер за обновяване. Предложен е софтуер за извличане на динамични редове от базата данни и графичен интерфейс за тяхното управление. За математико-статистическият анализ на данните е предложен интегриран набор от софтуерни инструменти.

### **4. Онагледеност и представяне на получените резултати.**

Трудът на **Никола Цайкин** съдържа 14 таблици и 38 фигури, чрез които онагледява и обобщава получените резултати. Всички те са в подкрепа на изясняване характера на получените резултати.

## **5. Обсъждане на резултатите и използвана литература.**

Дисертацията е от 125 страници и се състои от списък на таблици и фигури, увод, 3 глави, изводи, приноси на дисертационната работа, списък на публикации по темата на дисертационния труд, литература.

Докторантът познава добре съвременното състояние на проблема, проучил е значителна по обем научна литература, списъка на която обхваща 101 заглавия, от които 89 на латиница и 12 на кирилица. Използваната литература е адекватно подбрана, достатъчна по обем.

**Уводът** обосновава необходимостта от решаването на проблема, в който се прави въведение в проблематиката, подчертават се проблемите и се формулират целите и задачите на дисертационния труд.

В първият раздел на **глава първа** авторът прави обстоен анализ на съществуващите решения за приложение на информационни модели с интегрирани софтуерни инструменти за анализ на динамични редове в аграрната и други сфери на науката. В този раздел се подчертават значението и предимството на отделянето на данни от съответните приложения и съхранението им в информационен модел с определен профил. Представени са решения за осигуряване на логически преглед на данните и достъп само до данни, необходими за конкретна група потребители и софтуерни приложения.

Във вторият раздел на **глава първа** е направен кратък обзор на основните източници на статистически данни за растениевъдната продукция в България и Европейския Съюз - Информационна система на ФАО, Национален Статистически Институт на България (НСИ) и Министерство на земеделието и храните на България.

Във **втора глава** са дефинирани същността и основните концепции на базата данни като специализирана среда за съхранение и управление на голямо количество данни. Посочени са основните критерии за избора на СУБД MS Access като програмна система за разработване на РС база данни и нейното управление. В тази глава е описан начинът на достъп до базата с данни за растениевъдното производство по култури за периода от 1961-2016г. за страни членки на ЕС. Поради отвореният характер и йерархически свързаните съвкупности от данни е възможно нейното разширяване и актуализиране. Представени са основните характеристики на разширена база "Crop\_ProductionBGEC", съдържаща информация за производството и реколтирани площи от земеделски култури в страни от Европейския съюз.

В тази глава са изведени и систематизирани характеристиките на методите за

анализ на динамични зависимости, които позволяват да се установи взаимната обвързаност между динамичните редове, както и да се изследват зависимости. Представена е принципната схема на софтуерната платформа за статистически анализ изградена с развойната среда R за статистическа и математическа обработка на данни.

По същество **трета глава** представлява приложната част на дисертацията в която са предложени и реализирани идеи за анализ на динамични редове за производството на някой култури в България и страни от ЕС. Прави добро впечатление добрата статистическа грамотност на докторанта и компетентното използване на адекватни статистически подходи за решаване на задачите. Кръгът на използваните статистически методи обхваща: базов статистически анализ, корелационен анализ, регресионен анализ, дисперсионен анализ, сравнителни методи, кълстър анализ. За реализирането им са използвани пакета приложни програми SPSS и програмният език R за статистически анализи и графики. Всички получени резултати са илюстрирани с подходящи таблични и графични представления.

При оценката на динамичните редове за производство на слънчоглед за България (1961-2014 г.) по статистически райони по метода Tukey са установени, статистически различия между средния добив по райони, при критично ниво на значимост  $\alpha = 0.05$ , и принадлежността на районите към кълстъри чрез йерархичен кълстърен анализ. Тук се прилага подход за съвместно използване на двата метода, който предоставя по-пълна аналитична информация. Направено е заключение, че районите Северен централен и Североизточен, които имат най-високи средни добиви формират един кълстър. Останалите райони съществено се различават по среден добив от два района и образуват втори кълстър. Чрез показателят на разсейване (стандартно отклонение), авторът прави заключение за устойчивостта на добив от тази култура. Най-слабо разсейване е установено при средният добив от слънчоглед за район Южен централен. Чрез изчислителни процедури са установени процентните дялове на всеки район от общата продукция от слънчоглед за страната.

За представяне на основното направление в изменението на средния добив от слънчоглед по райони е използвано графичното представяне на емпиричните стойности за добив. Целият период е интерпретиран, като е разделен на три периода в зависимост от тенденцията в изменение на добива. Аналитичният модел, който адекватно и обективно да изрази тенденцията не е описан.

Динамиката на производството на ябълки в балканските страни членки на ЕС,

България, Румъния и Гърция за периода 1961 – 2008 година е представена само графично. Аналитичното представяне на изменението на производството на ябълки не е описано. Чрез изчислителни процедури са установени, дялът на ябълките в произведените плодове и регионалното разпределение на производството на ябълки за трите страни.

При проучването извършено за средния добив на ключови за България култури пшеница, соя, ориз и царевица са използвани базови статистически процедури. Изчислени са важни показатели, които дават информация за вариабилността, тенденциите и ръста на добива от тези култури. Проучването показва положителни тенденции на производството им, особено при традиционните за страната пшеница и царевица. Авторът подкрепя виждането, че за запазване на тази тенденция са необходими инвестиции и разработване и прилагане на инженерни решения.

За период от 54 години (1961-2014 г.) е направена статистическа оценка на някой структурни характеристики на лозарството в България. Представени са тенденции графично и аналитично, установени са зависимости и вариации дължащи се на фактора време, направени са оценки на средните по метода на Duncan, определена е екологичната приспособимост към средата на отглеждане на изследваните сортове лози.

Чрез статистически инструментариум са оценени средните добиви от десертни и винени сортове лози, отглеждани в България на ниво статистически райони и са установени достоверни различия между добива от различните райони. Екологичната оценка на добивът от винените сортове лози показва, че в Северозападен, Югозападен и Южен централен райони лозовите насаждения имат най-голяма приспособимост към средата на отглеждане в сравнение с останалите райони. Устойчивост към промените на средата за десертните сортове лози е установена за само за Югоизточен район.

Резултатите от изследването на тенденциите на средните добиви от десертни и винени сортове лози отглеждани в България на национално ниво, са представени графично и аналитично чрез уравнение от втора степен. За добива от винените сортове лози и фактора време е установена умерена корелация ( $r = 0,46$ ) и детерминация ( $R^2 = 21\%$ ). Значителна корелация ( $r = 0,61$ ) и детерминация ( $R^2 = 37\%$ ) са отчетени за добива от десертните сортове лози и фактора време. Направено е заключени, че на национално ниво тенденциите на средният добив от десертни и винени сортове лози са намаляващи, особено ярко изразени през годините на преход.

При изследване динамиката на данните за сировина (грозде) и краен хранителен продукт (вино) за България, са отчетени силни корелационни връзки между

производството на грозде и вино и фактора време. Установено е, че по-голяма част от дисперсията на производството на грозде и вино се дължи на фактора време. Регресионните модели за изменение на производството на грозде и вино във времето, са описани графично и аналитично. Направена е оценка за адекватност на моделите.

Графично и емпирично е представено изследване за изменението в стойностите на износа и вноса на вино във времето. Адекватно е избрано уравнението от втора степен за представяне на тенденцията на износа и вноса на вино. Установено е, че коефициентите на регресионният модел описващ изменението на вноса на вино е неадекватен и не представя обективно процеса.

Направена е сравнителна оценка по площи засяти с лозя, производство на грозде, производство на вино, износ на грозде (изключение прави Франция) и износ на вино и процедура за клъстерилизация по тези показатели на 21 страни членки на ЕС за периода 1961 г. – 2016 . По изследваните показатели Франция, Италия, Испания съществено се отличават от останалите европейски страни, като образуват отделен клъстер. В клъстера обединяващ останалите 18 страни попада и България. По износ на грозде и вино България съществено се отличава с една голяма част от европейските страни.

Високо оценявам способността на докторанта да обединява данни и да провежда мащабни статистически анализи в растениевъдството. Получените методологични и емпирични резултати могат да бъдат от интерес за изследователите и практическите дейности, както и основание за други теоретични и емпирични проучвания в областта на растениевъдството.

## **6. Приноси на дисертационния труд.**

Приемам приносите в дисертационното изследване така, както дисертанта ги е формулирал в съответната справката. Те имат научно-приложен и практико-приложен характер и могат да бъдат класифицирани по следния начин:

### **Научно-приложни приноси**

Разработена е специализирана среда за съхранение на многомерна база данни с динамични редове за различни аспекти на растениевъдното производство в България и страни от Европейския съюз.

Създаден е потребителски ориентиран софтуер за управление на динамични редове от базата данни (достъп, извлечане, наблюдаване) и софтуерни техники за автоматизирано актуализиране, което гарантира нейния технологичен живот.

Изградена е софтуерна платформа за математико-статистически анализ, чрез използването на пакети приложни програми и развойната среда R.

### **Практико-приложни приноси**

Аналитичните резултати получени при анализа на динамични редове за производството на някой култури, могат да бъдат от интерес за изследователите и практическите дейности в сектора растениевъдство, както и основание за други научни проучвания в областта на растениевъдството и други научни и практически дейности.

Може да се каже, че са изпълнени всички поставени задачи и са получени важни и оригинални резултати, които напълно отговарят на изискванията за една дисертация.

### **7. Критични бележки и въпроси.**

Нямам други критични бележки, освен вече споменатите в текста на рецензията.

### **8. Публикувани статии и цитирания.**

По темата на дисертацията Никола Йорданов Цайкин е представил три публикации в съавторство. Две от тях, една от които е реферирана в SCOPUS (*SJR=0.211 това осигурява 30 точки на докторанта според минималните изисквания в Постановление №26 от 19.02.2019 г.*), са представени на престижни международни конференции, а третата е публикувана в списание “Селскостопански науки”. Удовлетворено е изискването най-съществените резултати от дисертационното изследване да бъдат публикувани.

## **ЗАКЛЮЧЕНИЕ:**

В резултат на всичко гореизложено считам, че представеният дисертационен труд отговаря на изискванията на ЗРАСРБ, Правилника за неговото приложения и Правилника на Аграрния университет, което ми дава основание да го оценя **ПОЛОЖИТЕЛНО**.

Позволявам си да предложа на почитаемото Научно жури също да гласува положително и да присъди на **Никола Йорданов Цайкин** образователната и научна степен **“доктор”** по научната специалност **Информатика**.

**Дата: 08.06.2019 г.**

гр. Пловдив

**РЕЦЕНЗЕНТ: ... ...**

(доц. д-р Тодорка Мокрева)