

Рецензия

от доц. дтн Димитър Грозданов Христозов,
професор по компютърни науки в Американски университет в България,

на дисертационен труд на тема

„Компютърно симулиране и оценка на фирмени рискове“

с автор Слав Емилов Ангелов,

в професионално направление 4.6. Информатика и компютърни науки,
за получаване на образователна и научна степен „доктор“

Обща характеристика на дисертационния труд

Представеният за рецензиране труд е 184 страници и е осигурен достъп до приложения в <https://drive.google.com/drive/folders/1zb7tdwwEdiG8xia5udABP1Tn4wYGDtKo>. Развит е в увод, пет глави и заключение и е приложена библиография от 142 заглавия.

В **увода**¹ са поставени основните задачи на дисертационния труд и е направена оценка на неговата актуалност и приложимост. Уводът включва и обобщение на направения от автора литературен обзор. Разгледани са три групи източници: финансови съотношения и източници, свързани с риска от фалит, статистически методи за моделиране и обзор на източници, свързани с газоразпределителните дружества в България, които представляват предметната област на изследването.

Втора глава „Подбор на данни и методи за изследване“ е посветена на данните, използвани при изграждане на моделите данни, и е направен анализ за приложимостта на подходите при построяване на основани на данни статистически модели. Аргументиран е изборът на многомерен регресионен анализ; коментирани са изборът на независими променливи (предиктори), техниките за оценка на състоятелността на регресионните модели, и е представена процедура за построяване на модел, базиран на данни. Направен е

¹ В номерацията на автора увода е с номер едно.

и кратък коментар за това, как да се интерпретират получените регресионни коефициенти, както точкови, така и интервални.

В трета глава „Моделиране на финансовото представяне на газоразпределителния сектор в България“ е представен секторът от икономиката във фокуса на дисертацията – газоразпределителните дружества. Представена е спецификата на сектора на газоразпределение в българската икономика и как данните от финансовите отчети на фирмите от сектора са използвани за построяване и валидиране на регресионните модели. Посочени са и са коментирани източниците на данни, като е представен и основният инструмент за анализ – финансовите съотношения, изчислявани на базата на данните в отчетните финансови документи. Анализът на финансови документи е широко използван метод, поради задължителния характер на публикуването на отчетите на годишна база и поради необходимостта от обективно представяне на данните, даващ само малки възможности за адаптирането им в зависимост от определена фирмена политика.

Четвърта глава „Регресионна техника за редуциране на влиянието на зависимите предиктори“ разглежда проблема за построяване на регресионни модели при наличие на корелираност между предикторите.

В пета глава „Крос-валидирана постъпково конструирана множествена регресия“ се използва метода на „крос-валидацията“ за повишаване достоверността на прогнозните резултати, чрез валидиране на модела. Методът е широко приложим в условията на малки извадки, при които не е възможно данните да се разделят на обучително (training) и тестово (testing) множество.

В шеста глава „Използване на резултатите при оценка на риска“ е предложен подход за интерпретиране на резултатите от статистическия анализ на финансови отчети за оценка на риска от фалит.

В заключението са обобщени претенциите на автора за приносите на дисертацията и са представени данни за апробацията на резултатите.

Приложенията съдържат два файла с R код, използван за реализация на обработката на данни и изчисляване на необходимите статистики, описани в

дисертацията. В кода изключително се използват пакети, достъпни в CRAN. Налични са и два файла с данни, използвани при построяване на моделите.

Актуалност

Актуалността на проведеното изследване може да се търси в няколко основни насоки:

1. Газоразпределителната индустрия в България се развива динамично в последните години и изследвания, свързани с финансовото здраве на компаниите, ангажирани в сектора, са от изключителна актуалност предвид икономическия и социален ефект от потенциален фалит.
2. Създаването на методология и инструменти, позволяващи, основано на данни оценяване на финансовото здраве и риска от фалит, е преносимо и към други сектори на икономиката.
3. Използването на адекватен статистически апарат повишава надеждността и доверието в получаваните прогнозни резултати.

Значимост на изследвания проблем в научно и научно-приложно отношение.

Значимостта на изследвания проблем се определя не само от неговата актуалност, но преди всичко от подхода на автора към неговото решаване. Дисертацията може да бъде отнесена към кръга изследвания, обозначвани напоследък с термина “Data Science”, изискващи съвместни компетенции в предметната област, в случая – газоразпределителната индустрия; в боравенето с данни, в използването на методи за анализ на данни и при използването на компютърни информационни технологии. От тази гледна точка е и най-същественният, според мен, принос – решаването на проблеми днес изисква разнообразни компетенции, обединени около добре изградена методология. Дисертацията е пример за такъв подход.

Точно формулирани цели и задачи на дисертацията.

Целта на дисертацията – „създаване на инструмент, който да се използва за третиране на неопределеността по отношение на риска от фалит или финансовото здраве“ – е ясно формулирана. Определени са и задачите (стр. 5) за постигането на тази цел.

Определен е обхватът на изследването – газоразпределителните компании в България. Определени са и данните – годишните финансови отчети.

Доказателства, че дисертационния труд е разработен самостоятелно.

Представените към дисертацията сведения за апробация на резултатите, включват четири публикации в списания, една самостоятелна и три с научния ръководител, като авторът на дисертацията е първи автор; и докладвания на 11 форума. Това ми дава основание да смятам, че резултатите, представени в дисертацията, са постигнати от автора.

Степен на познаване на състоянието на проблема и съответствие на използваната литература.

Авторът демонстрира добро познаване на проблемите, свързани с оценка на риска, финансовото здраве и мерките за тяхното оценяване; определено добри са и познанията на проблемите, свързани със статистическата обработка на данни. Остават съмнения по отношение на познаването на състоянието на областта „machine learning”, по-специално развитието на технологиите в областта. От 142 заглавия в библиографията, само 3 са посветени на този основен компонент в дисертацията и това са учебници или техническа документация, свързани с езика R и Data mining.

В библиографията липсват „Elements of Statistical Learning“ на Hastie, Tibshirani и Friedman, и „An Introduction to Statistical Learning: with Applications in R“ на James, Witten, Hastie и Tibshirani, които представляват общопризнати източници при решаване на задачите, които си е поставил авторът.

Коректност при цитирането на представителен брой автори.

Цитираните източници, като обхват и разнообразие, са представителни за областта на газоразпределение в България, оценка на финансови рискове и статистически методи. Не е застъпена в необходимата степен литературата, свързана с технологиите за машинно обучение.

Цитиранията не са според БДС I690:2011 Информация и документация, а стандартът APA (American Psychological Association) не е приложен прецизно и систематично.

В текста на дисертацията има позовавания, които не са посочени. Например на стр. 29 „В статия е описан...“, без да се посочва източника; по-долу „статия на Ел-Кеиб и

колектив [?]“ авторът е на кирилица, а в библиографията на латиница; стр. 66, таблица 2.1. – не е ясно дали е взета от цитираната статия или е адаптирана от автора.

Наличие на обоснован и разработен теоретичен модел на изследването.

Разработката следва теоретично добре обоснован подход. Разработената и следвана методология е издържана в теоретично и приложно отношение. Изборът на многомерна регресия като модел за описание на динамиката в явленията, представляващи финансов риск, е добре теоретично обоснован, независимо че не са разгледани някои възможни алтернативни подходи, например модели с използването на Марковски процеси и Бейсови статистики. Подходено е професионално при оценка на параметрите на модела и доказване на надеждността на получените резултати. Използваните техники, като крос-валиране, са надеждни и доказали се в множество изследвания.

Съответствие на избраната методология и методика на изследване с поставената цел и задачи на дисертационния труд.

Изследването е осъществено на базата на добре обоснована методология, съобразена с целите на изследването, предметната област и наличните данни.

Наличие на собствен принос при събирането и анализирането на емпиричните данни.

Използваните данни са публично достъпни. За целите на изследването те се явяват първични (стр. 40), но по същество са вторични, обобщени годишни счетоводни отчети. Използвани са данни за отчети от 2006-2014 г. (стр. 84), като към днешна дата са публикувани и отчетите за 2015-2017 г. При определено малкия обем наблюдения, показващи динамиката в сектора, използването на всички налични достъпни данни е задължително. По тази причина не мога да приема и подхода с отстраняване на непълни записи вместо използването на някои от методите за тяхното прогнозиране.

Критични бележки

1. Номерацията на формулите. В глави 4 и 5 формулите са номерирани, а в останалите не са. Това прави текста труден за проследяване. Например на стр. 115 в условието на Лема 4.1. се цитира „регресия от вида 5.4“, като липсва такава

означение.

2. Индикатор за риск от фалит е и снижаването във времето на отошението „приходи от дейност/други приходи“, което не е разгледано в дисертацията. Това съотношение зависи от сектора на икономиката, но за случая с газоразпределителните дружества може би е информативно.
3. Използвана терминология: (1) „хранилища на данни“ е термин, използван като превод на „data warehouse“, което е структура от данни с определени характеристики. Таблица на Excel не съответства на такава структура; (2) преводът „машинно самообучение с учител“ на „supervised machine learning“ не е подходящ; (3) FL се определя веднъж като „задължнялост (стр. 18) и втори път като „задължения към активи“ (стр. 13). Не е включен в списъка със съкращения; (4) използва се „кеш“ (стр.18) във формулата, без да се уточнява дали става дума за наличност в каса, в текуща сметка, депозит или в друга форма на финансови средства с висока ликвидност; (5) на стр. 54 се разглеждат категоррите данни като числови и категорийни, като категорийните са разграничени на номинални и ординални, но числовите също са в две категории – интервални и относителни, които имат различни характеристики.
4. Към днешна дата има две основни направления в областта на machine learning – statistical learning и deep learning. Авторът се спрял на първия от тях, без да аргументира избора си.
5. Използването на езика R и библиотеките към него е добър избор при разработване на прототип, но последващият път до изграждане на пълноценно компютърно приложение е много дълъг.

Коментари по отношение на приносите

Приемам постигнаите в дисертацията резултати, а именно:

1. Направено е изследване на газоразпределителния сектор на икономиката от гледна точка на финансовия риск. Не зная за друго такова изследване и приемам твърдението на автора, че е първо по рода си.
2. Разработен е прототип на приложение за машинно обучение, позволяващо

прогнозирането на риска от фалит. Подходът, методологията и използвания инструментариум са приложими за различни сектори на икономиката.

3. Разработеният метод за конструиране на регресионна функция при наличие на значима корелираност между предикторите с използване на ортогонална трансформация заслужава висока оценка.
4. Опитът на автора за минимизиране на средно-квадратичната грешка при прогнозиране с използване на техниката „крос-валидиране“ е приложим в множество реални задачи и също заслужава висока оценка.

Заклучение

Независимо от критиките, отправени към дисертационния труд, най-същественният положителен резултат е търсената и получена се синергия за решаване на проблем с използване на резултати от различни научни направления – икономика, статистика, и информатика. От тази гледна точка докторантът заслужава подкрепа и адмирации.

Дисертацията отговаря на изискванията на закона и правилника за прилагането му и ще гласувам за присъждане на образователната и научна степен „доктор“ в професионално направление 4.6. „Информатика и компютърни науки“ на Слав Емилов Ангелов.

Рецензент:

(Димитър Христозов)

Дата 18 март 2019 г.