# Приступни рад

|  |  |
| --- | --- |
| **ПРЕЗИМЕ И ИМЕ** | **Павловић Марија** |
| **БРОЈ ИНДЕКСА** | **2016∕3009** |
| **СТУДИЈСКИ ПРОГРАМ** | **Информациони системи и технологије – Пословна интелигенција** |
| **ПРЕЗИМЕ И ИМЕ МЕНТОРА** | *др Гордана Савић* |
| **НАСЛОВ РАДА** | *Оцењивање и предвиђање ефикасности банака применом анализе обавијања података и метода машинског учења* |
| **АПСТРАКТ РАДА** | *У овом раду анализира се ефикасност банака у Србији за период од 2005-2013. Анализа се састоји из две фазе. У првој фази оцењен је индекс ефикасности банака помоћу технике Анализе обавијања података (ДЕА), док се у другој фази врши предвиђање ефикасности на основу резултата оцене ефикасности. ДЕА је непараметарска техника линеарног програмирања за мерење ефикасности јединица одлучивања. Међутим, за предвиђање ефикасности нове јединице погодније је користити методе машинског учења, јер се тиме постиже решавање оптимизацоиних проблема. Методе машинског учења уче над постојећим резултатим ДЕА методе, откривају законитости на основу којих предвиђају ефикасност нове јединице.* |

**УВОД**

Испитивање ефикасности финансијских система постао је често изучаван проблем. Ефикасност банака директо је везана за продуктивност једне економије. Главна предност повећања ефикасности је смањење разлике између камата на дати кредит и орочени депозит.Ово би могло да доведе до веће потражње за кредитима као и веће количине депозита.

Релативна ефикасност банака у једном систему или на нивоу једне државе се може израчунати коришћењем различитх улазних и излазних параметара. Избор параметара зависи од циља анализе. Они могу бити различити с обзиром на врсту и опсег јединица мере. У том случају се појављује проблем њиховог скалирања и агрегације истворемено са одређивањем производне функције која претвара улазе и излазе Да би се избегло скалирање и претоставке о тежинама и облику производне функције креирана је метода Анализе обавијања података (Data Envelopment Analysis – DEA) (Charnes, Cooper, & Rhodes, 1978). Она омогућава оцену релативне ефикасности сваке јединице у посамтраном скупу узимајући у обзир перформасе осталих јединица (банака). То значи да улазак нове банке на тржиште или процена њене ефикасности захтева нову анализу и решавање *n* линераних модела.

У овом раду се предлаже коришћење метода машинског учења за процену индекса ефикасности нове јединице узимајући у обзир већ израчунате вредности индекса ефикасности свих осталих банака које већ послују на тржишту. Поред тога биће извршена анализа и поређење резултата добијених коришћењем различитих скупова улаза и излаза у циљу одређивања оног који ће омогућити добијање релевантних резултата.

Након увода дата је формулација проблема, где је описана важност испитивања ефикасности пословног система, а посебно банака. Након тога следи лична мотивација и порекло идеје за ово истраживање. Затим је дата организација истраживања и оквирни садржај рада. Након дефинисања дела света у коме се врши истраживање, дат је преглед основних теоријских концепата са прегледом радова релевантних за дати проблем. На крају су описани сврха и циљеви истраживања и дата је листа коришћене литературе.

# ФОРМУЛАЦИЈА ПРОБЛЕМА

Банкарски сектор сматра се веома важним делом једне економије. Развој технологија, доступност информација и јача конкуренција девеле су до многих промена у финанцијским системима. Ефикасност банака постао је важан проблем, којим су се бавиле многе студије. Оценом ефикасности могуће је утврдити своју позицију у односу на конкуренцију, као и уочити одређене трендове на тржишту и благовремено реаговати на њих. Са друге стране, предвиђање ефикасности у великој мери помаже при одлучивању о уласку на ново тржиште. Закључује се да истраживање ефикасности има велику пословну вредности, и значајно утиче на доношење одлука у једном пословном систему.

Резултет истраживања биће утврђивање модела оцене ефикасности као и модела предвиђања ефикасности који имају најбоље перформансе.

С обизиром да резултати оцене ефикасности зависе од атрибута укључених у модел, потребно је одредити комбинацију улаза и излаза који су најрелевантнији за ефикасност банака. Такође, потребно је одретити који алгоритам машинског учења је најпрецизнији у предвиђању ефикасности и који параметри тог алгоритма дају најбоље резултате.

# МОТИВАЦИЈА: ИНТЕЛЕКТУАЛНА И ЛИЧНА МОТИВИСАНОСТ

Мотивација за писање рада на тему испитивања ефикасности долази након иницијалног упознавања са принципима рада ДЕА методе. Уочена је могућност повезивања ове методе са методама машинског учења како би се искористиле позитивне стране ДЕА методе а мане надоместиле.

ДЕА метода је широко коришћена у анализи ефикасности финансијских институција, пре свега због бројних нумеричких показатеља пословања. Отуда идеја за интеграцијом ДЕА методе и метода машинског учења баш у овој области.

**ОРГАНИЗАЦИЈА И МЕТОДЕ ИСТРАЖИВАЊА**

Дефинисани проблем решаваће се у две фазе.

У првој фази одређује се ефикасност банака применом ДЕА методе. Овде ће бити коришћене различите комбинације улаза и узлаза, као и различите варијанте самог ДЕА модела. Биће коришћени модели са константним приносом на обим и модели са варијабилним приносом на обим, као и модели са улазном и излазном оријентацијом. Принос на обим представља однос између промене улаза и излаза. Уколико се са повећањем улаза, излаз пропорционално повећава у питању је константни принос. Ако овај однос није пропорцијалан, у питању је варијабилни принос.

У другој фази, врши се предвиђање ефикасности на основу резултата оцене ефикасности. Биће разматрано неколико алгоритама машинског учења, чије ће се перформасе мерити. Овде је потребно оптимизовати параметре сваког од алгоритама како би резултати били најбољи могући. Атрибут који представља индекс ефикасности биће атрибут који се у овој фази предвиђа. Ови алгоритми имају различите технике учења, па је потребно видети који најбоље учи из датих података.

С обзиром да ДЕА метода даје различите резултате за различите моделе, потребно је за сваки ДЕА модел извршити предвиђање и упоредити перформансе.

ДЕА метода спада у непараметарске методе за оцену ефикасности. За разлику од параметарских метода које захтевају да унапред буде позната функција која повезује улазе и излазе, непараметарске методе не захтевају претпоставке о функционалној форми границе ефикасности нити о тежинама које ће бити додељене појединим улазима и излазима. Ова особина чини ДЕА методу веома популарном и често коришћеном методом. Међутим, из тога следи да резултати зависе од изабраних улаза и излаза, што се може сматрати и маном ДЕА методе (Cinzia & Léopold, 2007).

Како је ДЕА метода дескриптивна а не предиктивна метода, за одређивање ефикасности нове, непознате јединице, било би потребно поново израчунавање индекса ефикасности за цео скуп јединица, што код великог скупа података може да захтева велику количину меморије и доста процесорског времена. Поновна анализа се може избећи коришћењем метода машинског учења. Односно резултати добијени претходним ДЕА анализама би се могли користити да се предвиди индекс ефикасности непознате јединице. Ове методе би училе над постојећим постојећим резултатима ДЕА методе, откривале трендове и шаблоне понашања које би затим примениле на нове јединице и предвиделе њихово понашање. Као излаз из ДЕА методе добија се проценат ефикасности, а ово постаје атрибут чија се вредност предвиђа помоћу различитих алгоритама.

Алгоритми који ће се користити у овом истраживању спадају у групу надгледаних алгоритама. Предвићање ових алгоритама заснива се на скупу случајева за учење и скупу случајева за тестирање. У скупу за учење случајеви имају обележен атрибут чије вредности касније треба предвидети. Сваки алгоритам има другачију технику проналажења патерна у подацима. Случајеви за тестирање немају вредности за означени атрибут, већ се уочени патерни користе да се вредност тог атрибута предвиди.

Другу групу алгоритама чине ненадгледани алгоритми. Они су много мање коришћени од надгледаних. Ови алгоритми немају обележен атрибут чија се вредност предвиђа, већ је њихов циљ да организују податке на неки начин или да помоћу неког шаблона опишу структуру тих података.

Надгледано учење се може поделити у две категорије на основу типа атрибута чија се вредност предвиђа: класификација и регресија.

Код класификације, циљани атрибут је категорички. Циљ алгоритама класификације је да сваки случај додели класи којој је најсличнији. Као типични проблеми касификације могу се издвојити системи препоруке, препознавање слика и говора, као и различита предвиђања у медицини.

Код регресије, циљани атрибут је дискретан. Циљ је конструисати регресиону функцију чији су аргументи независни атрибути а чија је вредност заправо циљани атрибут. Пример проблема регресије може бити предвиђање цена на берзи и потрошње енергије.

С обзиром да је атрибут који се предвиђа у овом истраживању индекс ефикасности који је нумеричког типа, биће коришћени алгоритми који решавају проблем регресије.

У наставку је приказан оквирни садржај рада.

1. Увод
2. Методологија
   1. Анализа обавијања података
      1. Основни модел
      2. Оријентација модела
      3. Принос на обим
      4. Потпуно рангирање
   2. Машинско учење
      1. Линеарна регресија
      2. Метод подржавајућих вектора
      3. Неуронске мреже
      4. Најближи суседи
      5. Гаусов процес
3. Опис проблема и анализа података
   1. Опис проблема
   2. Дескриптивна анализа података
   3. Модели података
4. Анализа резултата
   1. Анализа ефикасности
   2. Анализа модела предвиђања
   3. Дискусија
5. Закључак
6. Литература

# ОПИС ДЕЛА СВЕТА КОЈИ ЋЕ БИТИ ИЗУЧАВАН

У овом истраживању биће анализирана ефикасност банака у Републици Србији за период од 2005. до 2013. године. На основу скупа показатеља пословања банака биће израчуната њихова релативна ефикасност. Затим ће се ти резултати користити за конструисање модела предвиђања ефикасности нове банке која би желела да уђе на ово тржиште.

# РЕФЕРЕНТНА ЛИТЕРАТУРА, КОНЦЕПТИ И ТЕОРИЈСКИ ФОКУС

Уовом раду је за оцену ефикасности коришћена Анализа обавијања података (Charnes, Cooper, & Rhodes, 1978). Она је уведена као непараметарска метода за оцену ефикасности на основу више разнородних улаза и излаза. Ефикасност описује у којој мери су ресурси (улази) добро искоришћени за реализацију постављених задатака. Анализа обавијања података рачуна ефикасности као однос између тежинске суме излаза и тежинске суме улаза, где се тежине одређују за сваку јединицу одлучивања. Основни модел са претпоставком константног приноса дат је у наставку (1) (Savić & Martić, 2017).

p.o. (1)

С обзиром да дати модел није линеаран, ради лакшег рачунања потребно је линеаризивати овај модел. Линеарни модел дат је формулама (2) (Savić & Martić, 2017).

p.o. (2)

ДЕА модели могу имати улазну и излазну оријентацију. Улазна оријентација означава да се тежи минимизацији улаза за постојећи ниво излаза. Код излазне оријентације циљ је максимизирати излаз за постојећи ниво улаза.

Модел (2) има улазну оријентацију. Општи модел са излазном оријентацијом дат је формулама (3) (Savić & Martić, 2017).

p.o. (3)

У овом раду ће, такође, бити рађено рангирање јединица према њиховој ефикасности. Постоји неколико метода рангирања (Savić & Martić, 2017):

* Матрица унакрсне ефикасности
* Број појављивања у скупу референтних тачака
* Увођење идеалне тачке
* Мерењем суперефикасности

Класични модели за све ефикасне јединице дају ефикасност од 100%, па је због тога погодно користити последњу методу – модел за оцену суперефикасности (*Andersen P, Petersen NC. 1993).* Овај метод за све ефикасне јединице даје ефикасност већу од 100% код улазне оријентације, односно мању од 100% код излазне оријентације што омогућава рангирање јединица на основу индекса ефикасности.

Улазни модел са оценом суперефикасности дат је формулама (4) (Savić & Martić, 2017).

p.o. (4)

ДЕА метода се често користи у банкарском сектору. У Индији је спроведено истраживање ефикасности страних и домаћих банака за период од 2000.-2010. године у циљу упоређивања ефикасности банака у односу на власништво и величину, као и упоређивање кретања ефикасности кроз дати период (Karimzadeh, 2012).

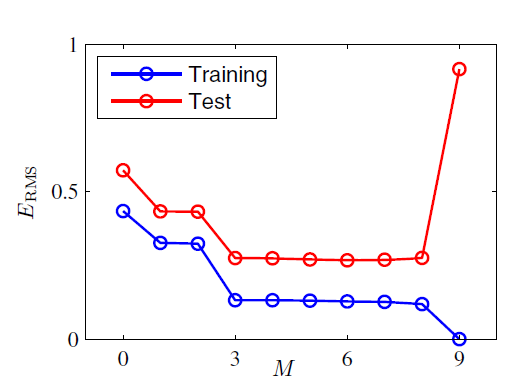
Рад на тему примене ДЕА методе за оцену ефикасности банака у Румунији идентификује најефикасније и најнеефикасније банке на посматраном тржишту. Такође се врши поређење ефикасности банака у односу на тип власништва и величину банке (NIŢOI, 2009).

Што се тиче Србије, такође постоје радови на тему испитивања ефикасности банкарског сектора. Једно истраживање пореди ефикасност банака у Србији добијену помоћу два ДЕА модела са различитим улазима и излазима и издваја лидере на тржишту добијене овим методама (Maletić, Kreća, & Maletić, 2013).

Осим тога, потребно је поменути рад који пореди рангове банака добијене ДЕА методом и *I-distance* статистичком методом за рангирање. Овде се разматрају предности и мане ових метода као и могућност њихове интеграције (Mihailovic, Bulajić, & Savić, 2009).

Машинско учење је скуп метода које аутоматски уочавају патерне у подацима, а затим те патерне користе за предвиђање непознатих података или доношење различитих одлука. Резултат извршавања алгоритма је функција која на основу атрибута даје вредност циљаној варијабли. Облик ове функције одређује се у фази тренирања, односно учења. Након тога, добијени модел се може применити у фази тестирања. Битно је нагласити да подаци за тренирање садрже вредности циљане варијабле, док подаци за тестирање не садрже ове вредности, већ се управо оне предвиђају помоћу добијеног модела. Фази учења често претходи фаза предпроцесирања где се врши трансформација и селекција атрибута, како би се задати проблем лакше решио (Murphy, 2012).

При конструисању функције тежи се што бољој апроксимацији , тако да грешка буде што мања. Ако се грешка дефинише као одступање између предвиђених и стварних вредности, циљ је да то одступање буде што мање како би модел био што прецизнији. Са повећањем сложености функције, она све боље апроксимира стварно стање. Готово идеалну апроксимацију представља функција која пролази кроз сваку тачку у подацима за учење. Међутим, показало се да такве функције дају врло лоше резултате у фази тестирања. Ова појава претераног прилагођавања функције тако да над непознатим подацима она даје доста лошију апроксимацију назива се *over-fitting.* Са друге стране, треба водити рачуна да функција не буде превише проста, јер се онда такође неће извршити добра апроксимација. Ова појава је назива *under-fitting.* На слици 1 приказан је однос грешке (ЕRMS) и сложености функције (М) над подацима за учење, приказано плавом линијом, и над подацима за тестирање, што је приказано црвеном линијом (Bishop, 2006).



Слика 1 Грешка у односу на сложеност функције (Bishop, 2006)

У овом раду биће коришћени следећи алгоритми:

* **Линеарна регресија**:

Алгоритам који покушава да успостави линеарну комбинацију објашњавајућих променљивих, тако да добијена линеарна једначина најбоље одговара датим подацима (Murphy, 2012)

* **Метод подржавајућих вектора**:

Алгоритам који се заснива на конструкцији једне или више хиперравни у *n*-димензионом простору, применљив је на проблеме и регресије и класификације (Murphy, 2012)

* **Гаусов процес:**

Широко примењен алгоритам који се ослања на Нормалну (Гаусову) расподелу, применљив је на проблеме и регресије и класификације (Davis, 2001)

* **Неуронске мреже:**

Алгоритам који симулира рад људских неуронских мрежа, састоји се од великог броја међусобно повезаних чворова који у циклусима подешавају тежине и уче на грешкама из претходних циклуса (Bishop, 2006)

* ***k*-најближих суседа**:

алгоритам који вредност посматраном случају додељује на основу *k* најбижих суседа, где *k* може бити произвољан цео број (Bishop, 2006)

Евалуација перформанси алгоритама биће вршена помоћу крос валидације. Крос валидација је метод валидације који дели скуп података на *n* једнаких делова, у овом случају 10, узима један део за тестирање а остале делове за учење. Затим се поступак понавља за сваки део.

Један од првих радова који је иницирао ово истраживање је рад који говори о интеграцији ДЕА методе и алгоритама машинског учења на примеру ефикасности кошаркаша. Овде се истичу две фазе истраживања. У првој фази се помоћу ДЕА методе одређује ефикасност корашкаша. Затим разултати прве фазе постају улаз у другу фазу где се врши предвиђање ефикасности кошаркаша са посебним освртом на метод подржавајућих вектора (Radovanović, Radojičić, & Savić, 2014).

Комплементарност ове две методе може се видети и на примеру избора добавњача. У чланку на ову тему поређени су резултати ДЕА методе са константним и варијабилним приносом, а затим и резултати предвиђања помоћу неуронских мрежа и методе подржавајућих вектора (Farahmanda, Desaa, Nilashi, & Wibowo, 2015).

У још једном од релевантних радова истиче се предност комбиновања ДЕА методе и метода машинског учења при анализи над великим скупом података. Ово може захтевати велику количину меморије и процесорског времена. У овом раду се, такође, предлаже коришћење методе подржавајућих вектора у другој фази анализе ефикасности, као методе која је у поређењу са неуронским мрежама дала боље резултате (Farahmand, 2014).

Осим тога, показано је да повезивање ове две методе може да буде и у супротном смеру. Један рад даје опис такве методологије. Овде се у првој фази врши кластеризација јединица одлучивања, где се оне групишу тако да међусобно имају најсличије вредности улаза и излаза. Затим се за нову јединицу врши класификација и она се додељује најсличнијој групи, док се у трећој фази врши оцена ефикасности само те групе јединице где је нова јединица додата (Hong, Ha, Shin, Park, & Kim, 1999).

СВРХА И ЦИЉЕВИ ИСТРАЖИВАЊА

Сврха рада је стварање основе за унапређење ефикасности банака, као и основе за испитивање исплативости уласка на тржиште. У складу са тиме, постављени су следећи циљеви:

* Прикупити релевантна знања из досадашњих истраживања
* Испитати ефикасност банака из датог сета података
* Конструисати модел за предвижање ефикасности банака
* Анализирати добијене резултате и дати предлоге за даље кораке примене модела

# РЕФЕРЕНЦЕ

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer Science+Business Media, LLC.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of Decision Makin Units. *European Journal of Operational Research, 2*, 429-444.

Cinzia, D., & Léopold, S. (2007). The Measurement of Efficiency. У S. L. Daraio Cinzia, *Advanced Robust and Nonparametric Methods in Efficiency Analysis.* Springer US.

Davis, R. A. (2001). Gaussian Processes. У A. H. El-Shaarawi, & W. W. Piegorsch, *Encyclopedia of Environmetrics.* Wiley. Преузето са http://www.stat.columbia.edu/~rdavis/papers/VAG002.pdf

Farahmand, M. (2014). A Combined Data Envelopment Analysis and Support Vector Regression for Efficiency Evaluation of Large Decision Making Units. *International Journal of Engineering and Technology*, 2310-2321.

Farahmanda, M., Desaa, M. I., Nilashi, M., & Wibowo, A. (2015). An Improved Method for Predicting and Ranking Suppliers Efficiency Using Data Envelopment Analysis. *Jurnal Teknologi*, 91-97.

Hong, H. K., Ha, S. H., Shin, C. K., Park, S. C., & Kim, S. H. (1999). Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning. *Expert Systems with Applications*, 283–296.

Karimzadeh, M. (2012). Efficiency Analysis by using Data Envelop Analysis Model: Evidence from Indian Banks. *International Journal of Latest Trends in Finance and Economic Sciences*, 228-237.

Maletić, R., Kreća, M., & Maletić, P. (2013). APPLICATION OF DEA METHODOLOGY IN MEASURING EFFICIENCY IN THE BANKING SECTOR. *Economics of Agriculture*, 843-855.

Mihailovic, N., Bulajić, M., & Savić, G. (2009). RANKING OF BANKS IN SERBIA. *Yugoslav Journal of Operations Research* , 323-334.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective.* Massachusetts Institute of Technology.

NIŢOI, M. (2009). Efficiency in the Romanian Banking System: An Application of Data Envelopment Analysis. *Romanian Journal of Economics*, 162-176.

Radovanović, S., Radojičić, M., & Savić, G. (2014). Two-Phased DEA-MLA Approach for Predicting Efficiency of NBA Players. *Yugoslav Journal of Operations Research*, 347-358.

Savić, G., & Martić, M. (.). *MERENJE EFIKASNOSTI POSLOVNIH SISTEMA.* Преузето са Poslovna Analitika, Katedra za Operaciona istraživanja i Statistiku, Fakultet Organizacionih nauka: http://pa.fon.bg.ac.rs/wp-content/uploads/2017/03/MEPS-II-dan1.pdf