

Primjena klaster analize u ekonomskim istraživanjima

The Implementation of Cluster Analysis in Economic Research

ADIS PUŠKA

Predavač visoke škole

Visoka škola eMPIRICA u Brčko distriktu BiH

Bulevara mira 10

Bosna i Hercegovina

adispuska@yahoo.com

ADMIR I. BEGANOVIĆ

Docent

Univerzitet za poslovne studije Banja Luka

Jovana Dučića 23a

Bosna i Hercegovina

admirbeganovic52@gmail.com

Prethodno priopćenje

UDK: 338:303.722.4

Sažetak: Odabir primjena multivarijatne analize je široka i rasprostranjena. U kontekstu multivarijatne analize postoje mnoge metode koje se mogu primjenjivati u statističkim analizama. Klaster analiza je metoda koja uspoređuje jedinice promatranja na temelju njihove povezanosti te ih razvrstava u homogene skupine. Klaster analiza rijetko se primjenjuje samostalno u ekonomskim istraživanjima. Ona se obično primjenjuje s drugim metodama multivarijatne analize. Ovaj rad nastoji objasniti primjenu klaster analize u ekonomskim istraživanjima, teorijski, metodološki i praktičnim segmentima primjene. Osim ovoga daje se i pregled primjene klaster analize u ekonomskim istraživanjima te smjernice za buduću provedbu ove metode. Na kraju rada se kroz praktični primjer predstavlja uporaba klaster analize.

Ključne riječi: Klaster analiza, k-mean analiza, dendrogram

Summary: The application of a multivariate analysis is very broad and wide-spread. Within the context of a multivariate analysis there are many methods that can be used in statistical analysis. Cluster analysis is the method which compares the observation units based on their inter-connection and classifies them into homogeneous groups. Cluster analysis is rarely used independently within economic research. It is usually used in collaboration with other methods belonging to the multivariate analysis. This paper attempts to explain the application of cluster analysis in economic research, as well as the theoretical, methodological and practical segment of application. Additionally it provides an overview of the application of cluster analysis in economic research and it provides guidelines for the future application of this method. The end of the paper presents the application of cluster analysis through a practical example.

Key words: cluster analysis, k-mean analysis, dendrogram

1 Uvod

U istraživanjima se sve više primjenjuju metode multivarijatne analize koja se koristi načelima multivarijatne statistike. Multivarijatna statistika uključuje promatranje i analizu dviju ili više varijabli istodobno zbog čega se i zove multivarijatna analiza. Ove se tehnike u praksi koriste u smislu višedimenzionalnih analiza u kojima se poštuju utjecaji i učinci svih relevantnih varijabli.

Klaster analiza spada u metodu razvrstavanja podataka pa se u nekim knjigama ili radovima rabi naziv analiza razvrstavanja, a u anglosaksonskoj literaturi naziv cluster analysis koji se udomaćio kod nas. Klaster analiza dijeli se na hijerarhijsku i nehijerarhijsku.

Kod hijerarhijske klaster analize provodi se uspoređivanje jedinica promatranja prema njihovim međusobnim odstupanjima pa je ovdje moguće primjenjivati različite formule odstupanja. Osim što je potrebno odrediti mjere odstupanja, treba izabrati i odgovarajuću mjeru razvrstavanja po kojoj će se razvrstavati promatrane jedinice. Nehijerarhijske klaster analize primjenjuju se kod razvrstavanja jedinica promatranja u određen broj homogenih međusobno heterogenih skupina.

Kod klaster analize traži se struktura podataka za razvrstavanje multivarijatnih jedinica promatranja u klaster. Cilj analize je pronalaženje optimalnog razvrstavanja kod kojeg su opažanja unutar svakog klastera slična, ali se različiti klasteri međusobno razlikuju. Kod ove analize pretpostavlja se mogućnost nalaženja prirodnog način razvrstavanja smislenog za istraživača. U klaster analizi ne zna se unaprijed broj skupina niti su one unaprijed poznate. U skladu s navedenim predmet ovoga rada je teorijska i praktična razrada klaster analize te davanje budućih smjernica u primjeni ove analize.

Primjena klaster analize kao značajne statističke metode je široka i konzistentna. Pri predstavljanju klaster analize obrađuju se njezini osnovni koraci i upućuje se na praktičnu primjenu ove analize. Cilj ovoga rada je prikazati primjenu klaster analize u ekonomskim istraživanjima te provesti raspravu o mogućim primjenama klaster analize u budućim istraživanjima.

2 Teorijski i praktični aspekti klaster analize

Prije nego što se definira klaster analiza, potrebno je navesti nazive koji se rabe za ovu metodu: klaster analiza, analiza razvrstavanja, Q-analiza, tipologija, numerika taksonomija, prepoznavanje obrazaca itd.

2.1 Pojam klaster analize

Klaster analiza je metoda koja se primjenjuje za razvrstavanje podataka na osnovi sličnih karakteristika (Vujović et al., 2013, 824). Klaster analiza je objektivna statistička tehnika koja se primjenjuje za identifikaciju prirodnog razvrstavanja u skupu podataka (Prakash i Dagaonkar, 2011, 46). U osnovi svake definicije klaster analize njezina je osnovna svrha, a to je razvrstavanje objekata u homogenu klasu na osnovi svoje sličnosti, a skupine se međusobno razlikuju. Razvrstavanje objekata provodi se na osnovi njihovih zajedničkih obilježja.

S obzirom na to da je cilj klaster analize razvrstati slične objekte zajedno, potrebno je utvrditi koliko su ti objekti slični odnosno različiti (Kurnoga Živadinović, 2007, 4). Osnovni zadatak klaster analize je identifikacija klastera unutar neke populacije objekata. Na početku analize nije poznat konačni broj klastera unutar populacije, niti koji objekt pripada kojem klasteru. Metodologija klaster analize mora odgovoriti na tri osnovna pitanja: kako mjeriti sličnost između objekata, kako formirati klaster te kako utvrditi konačan broj klastera. Na osnovi toga obavlja se razvrstavanje prema stupnju sličnosti podataka unutar skupine koji se maksimiziraju dok međusobno postoji minimalna sličnost s drugim

skupinama. Postupak provedbe klaster analize naziva se klasteriranje. „Klasteriranje je postupak podjele nekog skupa podataka na unaprijed zadani broj klastera tako da je sličnost među elementima klastera najveća a razlika (udaljenost) među klasterima najveća (Markić, 2011, 202).“ No često se ne zna u koliko klastera se može podijeliti skup podataka jer je potrebno odrediti optimalni broj skupina. To je najčešći problem kod primjene klaster analize.

Da bi se odredio broj klastera, koriste se različite mjere udaljenosti po kojima se utvrđuje homogenost unutar skupine odnosno heterogenost između različitih skupina. Obično se koriste neke od sljedećih udaljenosti: euklidska, kvadratno euklidska, Mahalanobisova, Čebiševljeva, Ivanovićeve, Minkovskijeva udaljenost, „a najčešće su korištene euklidska udaljenost i kvadrirana euklidska udaljenost (Kurnoga Živadinović i Sorić, 2008, 197)“.

Odabir po kojem će se mjera udaljenosti primijeniti ovisi o mjernim jedinicama u kojima su izražene prilično osjetljivim varijablama. Zbog toga je prijedno potrebno transformirati originalne podatke putem odgovarajuće normalizacije jer se tada varijable izražene u različitim mjernim jedinicama svode na istu mjernu jedinicu.

Ovisno o tome kako se izračunavaju, klaster metode mogu se podijeliti na:

- hijerarhijske
- nehijerarhijske klaster metode.

Kod hijerarhijskih metoda obavlja se iterativan proces spajanja objekata u klaster tako da se u idućoj etapi spajaju objekti i prethodno formirani klasteri. Kada se jednom formiraju skupine, one se samo proširuju novim objektima prema povezanosti s tim kriterijima, ali tijekom formiranja skupina ne postoji mogućnost prelaska objekata iz jednog klastera u drugi. „Mogućnost prelaska objekata iz jedne u drugu grupu dozvoljavaju tzv. nehijerarhijski metodi grupiranja“ (Kovačić, 1994, 255).

2.2 Moguća primjena klaster analize

Potrebno je reći da klaster analiza nije statistički značajna metoda pa se u ekonomskim istraživanjima obično primjenjuje s drugim multivarijatnim metodama. Osnovni razlog tomu je što klaster metoda provodi samo razvrstavanje podataka na osnovi njihove međusobne povezanosti. Njome se ne može dokazivati hipoteze već je moguće razvrstavanje podataka za daljnju analizu.

Klaster analizu moguće je primjenjivati u makroekonomskim istraživanjima gdje je moguće razvrstavanje određenih zemalja odnosno regija prema povezanosti određenih varijabli, a moguće je korištenje i u mikroekonomskim istraživanjima gdje se kompanije ili potrošači povezuju u određene skupine.

Primjena ove metode najrasprostranjenija je u segmentiranju tržišta gdje se primjenom klaster analize može izvršiti razvrstavanje homogenih promatranih jedinica. Ova metoda može se primjenjivati kao osnova za utvrđivanje uzorka istraživanja. Osnovni skup razvrstava se u homogene klaster čije su jedinice promatranja međusobno povezane tako da istraživač da bi dobio što raznovrsnije podatke može primjenjivati metodu uzrokovanja na klasteru. Na taj se način dobiva sveobuhvatni uzorak, a jedinice promatranja razvrstavaju se prema određenim zajedničkim karakteristikama.

Pri tomu se ne vrši segmentiranje tržišta na osnovi pripadnosti određenoj skupini (prema godinama, prema spolu itd.) već na osnovi zajedničkih povezanosti skupina, npr. sklonost pri kupnji, navike u kupnji, navike u praćenju određenih sredstava informiranja, npr. korištenje interneta, gledanje TV-a, čitanje tiska i sl.

Sve jedinice toga klastera međusobno su povezane tako da svaka od njih predstavlja odlike toga klastera. Na taj je način moguće i reducirati podatke s pomoću klaster analize i to primjenom hijerarhijskog klastera jer on predstavlja podatke prema njihovim udaljenostima. Ako ne postoje velika udaljenost, moguće je reducirati podatke. Međutim, s povećanjem broja varijabli s kojim se skupine promatraju dolazi do proturječnosti između varijabli čime se otežava proces klasterifikacije skupina.

Na osnovi razvrstavanja moguće je sugerirati određeni model, npr. razvitka lokalnih zajednica, primjene određenih modela oglašavanja kada se primjenjuju u marketinškim istraživanjima, određenim reklamnim porukama i sl. Ovdje je interes istraživača na razvrstavanju određenih jedinica promatranja prema određenim njihovim karakteristikama ili varijablama istraživanja. Pa ako istraživača zanima skupina koja se najviše koristi internetom za potrebe internetskog oglašavanja, on reducira osnovni skup i odabire samo skupinu koja je najviše povezana s tom varijablom.

Primjena ove analize moguća je i u generiranju hipoteza (Kovačić, 1994) nakon razvrstavanja jedinica promatranja s pomoću klaster analize. Na taj se način dobije struktura i sastav te skupine pa je na osnovi toga preliminarnog istraživanja ili ad-hoc istraživanja moguće dobiti određenu sliku o osnovnom skupu te je moguće na osnovi njihovih karakteristika definirati hipoteze glavnog istraživanja. Ovdje se klaster analiza provodi u preliminarnom istraživanju gdje se razvrstavaju jedinice promatranja u homogene skupine. Tako se dobivaju preliminarne informacije o njima te se na osnovi ovih informacija mogu generirati hipoteze glavnog istraživanja.

Osim generiranja hipoteza klaster analiza može se primijeniti i u predviđanju. Na osnovi istraživanja provodi se razvrstavanje jedinica promatranja u klastere pa se primjenjuje prognostika, to jest predviđa se mijenjaju li se te skupine i njihov sastav. Naknadnim istraživanjima može se dokazati to predviđanje. Na taj način klaster analiza može poslužiti i u dokazivanju postavljenih hipoteza.

Klaster analize je moguća u svim ekonomskim istraživanjima gdje je potrebno izvršiti razvrstavanje određenih jedinica promatranja ili čak i varijabli ako postoji veći broj njih. Tako se provodi klastiranje osnovnog skupa u nekoliko homogenih skupina te je na taj način moguće dalje provoditi određena istraživanja ili dodatne analize. Primjenom oba načina klaster analize moguće je provesti istraživanje i dati smjernice za buduća istraživanja. Na taj bi se način ova analiza provodila kao samostalna analiza u okviru istraživanja.

3 Metodologija primjene klaster analize

Pri uporabi klaster analize potrebno je primijeniti određenu metodologiju u pripremi i analizi prikupljenih podataka kako bi rezultati koje daje ova metoda bili reprezentativni. Za ovu analizu potrebno je reći da ona razvrstava varijable ili jedinice promatranja u određene međusobno heterogene klastere, a u sebi sadrže homogene podatke.

Način kategorizacije varijabli odnosno oblikovanja klastera ovisi o odabranoj proceduri klasterizacije. Sve ove klasifikacije mogu se podijeliti u tri osnovne skupine i to (Mooi i Sarstedt, 2011, 145):

- hijerarhijske procedure (oblikovanje klastera temelji se na odabranim udaljenostima i odabranom algoritmu klasterizacije)
- procedure dijeljenja (najpoznatiji predstavnik ove skupine je algoritam klasterizacije k-sredine vrijednosti koji vrši minimizaciju varijacija unutar klastera te se na taj način provodi segmentacija varijabli)
- procedure klasterizacije u dva koraka (u prvom koraku provodi se procedura vrlo slična algoritmu klasterizacije, k-sredine vrijednosti, a u drugom koraku modifikacija hijerarhijske klasterizacije).

Kod određivanja sličnosti u klaster analizi primjenjuju se sljedeće metode: mjere udaljenosti, mjere korelacije i mjere razvrstavanja. Zadatak mjera udaljenosti je da se na osnovi sličnosti razvrstavaju varijable unutar klaster analize, a razinom korelacije iskazuje se sličnost među varijablama. Mjere razvrstavanja ili udruživanja služe da se procijeni stupanj podudaranja između varijabli. Na osnovi ovih metoda varijable se razvrstavaju putem klaster analize. Postoji više tipova klaster analize, ali najčešće se primjenjuje hijerarhijska i nehijerarhijska analiza.

Kod hijerarhijskog klastera započinje se s „n“ varijabli (klastera), a zatim se povezuju klasteri s najmanjom udaljenosti. Tako postupak završava s jednim klasterom koji sadrži svih „n“ opažanja. U svakom se koraku opažanje ili već postojeći klaster apsorbira u drugi klaster. Proces se može i obrnuti. Može se krenuti od jednog klastera koji sadrži svih „n“ opažanja i završiti s „n“ klastera pri čemu je u svakom pojedinom klasteru jedno opažanje. Kod podjele opažanja se dijele u „g“ klastera. To se može načiniti polazeći od početne participacije ili od sredine klastera, nakon čega se opažanja smještaju u klastera u skladu s nekim optimalnim kriterijumom.

Grafički prikaz rezultata hijerarhijskog klastera obavlja se s pomoću dendrograma, koji pokazuje način na koji su spojene varijable i kolika je udaljenost između njih. Dendrogram je grafički prikaz postupnog kombiniranja objekata u klastera na kojem se uočavaju udaljenosti između pojedinih razina. Čita se slijeva nadesno. Opažanja su prikazana na vertikalnoj crti, a udaljenosti između klastera, na kojoj su spojeni, na horizontalnoj crti (Kurnoga Živadinović i Sorić, 2008, 197).

K-sredine metoda je algoritam koji svaki element pridružuje klasteru s najbližim centroidom. Potom se uobičajeno udaljenost računa kao euklidska udaljenost standardiziranih ili nestandardiziranih varijabli (Bahovec et al., 2011, 96).

3.1 Osnovni koraci hijerarhijske klaster analize

Pri provedbi hijerarhijske klaster analize koriste se mjere odstupanja, korelacije i razvrstavanja. Najprije se svaka jedinica promatranja definira kao poseban klaster. Zatim se izračunavaju mjere korelacije i odstupanja. Na osnovi udaljenosti jedinica promatranja formira se matrica udaljenosti te se traži najmanja udaljenost između jedinica promatranja i razvrstavaju se podaci. Kada su inicijalni klasteri formirani, provodi se ocjenjivanje tih klastera s preostalim jedinicama promatranja, to jest ponovno se vrši postupak izračunavanja odstupanja tih jedinica s inicijalnim klasterima. Nakon što se formira nova matrica odstupanja, postupak se provodi sve dok ne ostane jedna jedinica promatranja. Kada se na osnovi mjera odstupanja i mjera razvrstavanja odredi pripadnost pojedinih jedinica promatranja, završena je procedura klasterifikacije.

Klasteri se formiraju tako što se unutar njih stvaraju jedinice promatranja koje najmanje odstupaju tako da se formiraju homogene jedinice. Klasteri se međusobno razlikuju na osnovi ovih odstupanja. Razvrstavanje se provodi s pomoću odgovarajućih mjera. „Među mjerama razlike (udaljenosti) najpoznatija je tzv. euklidska mjera udaljenosti na bazi kvantitativnih promjenljivih“ (Kovačić, 1994, 281). U praksi se koriste i neke druge mjere, kao što su:

- Manhattan odstupanje
- Čebiševljeva odstupanje
- Minkowski odstupanje
- Ivanovićevo odstupanje.

Svako od njih koristi specifičnu formulu za izračunavanje odstupanja među promatranim jedinicama. Izračunavanje mjera odstupanja je prvi korak u primjeni klaster analize. Istraživač može primijeniti osim navedenih i druge formule za mjerenje odstupanja te se na osnovi dobivenih rezultata i s pomoću mjera razvrstavanja formiraju klasteri i provodi klaster analiza.

Kada se odabere i proračuna udaljenosti s pomoću odabrane mjere udaljenosti i nakon što se formira matrica udaljenosti potrebno je razvrstati jedinice promatranja na osnovi najmanje udaljenosti između njih. Upravo razvrstavanje jedinica promatranja provodi se putem odgovarajućih mjera razvrstavanja. Treba reći da „različite hijerarhijske metode na različit način određuju udaljenosti između klastera tj. između klastera i negrupiranih jedinica“ (Pivac, 2009, 219). Neke od mjera razvrstavanja su sljedeće:

- Metoda međusobnog razvrstavanja (between-groups linkage), radi na maksimiziranju udaljenosti između svakog para jedinica promatranja iz dva različita klastera. Udaljenost između dva klastera računa se kao prosjek udaljenosti svih kombinacija parova jedinica promatranja iz ta dva klastera.
- Metoda povezivanja unutar skupina (within-groups linkage), radi na minimalnoj udaljenosti svih jedinica promatranja unutar klastera. I ovdje je udaljenost jedinica prosjek udaljenosti svih kombinacija parova jedinica iz toga novonastalog klastera.
- Metoda najbližeg susjeda (nearest neighbor) pretpostavlja da je udaljenost unutar dva klastera jednaka udaljenosti između dviju najbližih jedinica promatranja iz ta dva klastera.
- Metoda najdaljeg susjeda (furthest neighbor) pretpostavlja da je udaljenost između dva klastera jednaka udaljenosti između dviju najudaljenijih jedinica iz ta dva klastera.
- Centroidna metoda (centroid clustering) pretpostavlja da je udaljenost između dva klastera jednaka udaljenosti između aritmetičkih sredina svih jedinica promatranja iz ta dva klastera. „Nedostatak ove metode je što se udaljenost na kojoj se klasteri spajaju može između pojedinih iteracija smanjiti što rezultira time da se klasteri spojeni u kasnijim iteracijama više razlikuju od onih spojenih u ranijim iteracijama“ (Orehovački, 2013, 138).
- Metoda medijana (median clustering) kao i prethodna rabi razlike aritmetičkih sredina samo bez ponderiranja.
- Wardova metoda, poznata kao metoda minimalne varijance, rabi kvadrirane udaljenosti unutar klastera i kvadrirane udaljenosti između klastera (Bahovec i Škrinjarčić, 2013, 16).

Pri provedbi klaster analize istraživač bira koje će mjere udaljenosti i mjera razvrstavanja koristiti. Potrebno je naglasiti da je primjenom različitih mjera moguće dobiti drukčiji konačni rezultat. Zbog toga je potrebno u primjeni klaster analize eksperimentirati s mjerama udaljenosti i razvrstavanja da bi istraživač bio siguran za dobivene rezultate primjenom hijerarhijske klaster analize.

Primjenom ove metode dobivaju se rezultati u vidu tablice i grafikon putem dendrograma. Prva kolona tablice predstavlja udaljenost promatranih jedinica, a druga kolona iduću povezanost prve jedinice s drugim jedinicama. Dendrogram grafički pokazuje povezanost jedinica promatranja te povezanost klastera.

3.2 Osnovni koraci k-sredine klaster analize

Za razliku od hijerarhijske klaster metode koja polazi od razvrstavanja jedinica promatranja na osnovi njihove udaljenosti te putem mjera razvrstavanja formiraju se klasteri, kod nehijerarhijske metode najprije se odredi broj klastera. To jest u koliko će se skupina jedinice promatranja podijeliti pa se dodjeljuju odgovarajuće jedinice promatranju tom klasteru. Od nehijerarhijskih metoda u praksi je najviše u uporabi k-sredine metoda, to jest metoda k-prosjeka koja će se u ovom radu i predstaviti.

K-sredine „klaster analiza ne daje gotove informacije o optimalnom broju klastera, već odluku o tome mora donijeti sam istraživač, rukovodeći se svrhom i ciljevima istraživanja“ (Mihčić i Kursan, 2010, 392).

Definiran broj klastera od istraživača može biti kako prednost tako i nedostatak ove metode. Prednost je ako su jedinice promatranja nehomogene te se one ne mogu razvrstati u manji broj skupina. Nedostatak je što unaprijed definiran broj klastera ne daje stvarnu povezanost jedinica promatranja jer

klaster analize će razvrstati jedinice promatranja bez obzira na njihovu povezanost. Ovdje se događa da zadnje jedinice promatranja koje su dodane klasterima imaju veliku udaljenost u odnosu na druge jedinice unutar istog klastera. „Ako je potrebno, neke jedinice se premještaju iz klastera u klaster sve dok se ne postigne stabilnost sistema“ (Pivac, 2009, 221). To kod hijerarhijske metode nije moguće jer se ne mogu prebacivati već jednom razvrstane jedinice promatranja. „Osnovni argument za primjenu ove metode klasteriranja je da je ovaj način grupiranja objekata u grupe pogodniji ukoliko se radi o grupiranju jedinica na kojima su izmjerena određena obilježja (objektima), a ne o razvrstavanju tih obilježja odnosno varijabli“ (Rašić Bakarić, 2006, 68).

Algoritam k-sredine klasteriranja čini niz sljedećih koraka:

- slučajno odabrati k-klastera
- odrediti središta (centroide) za svaki klaster
- ponavljati dok se ne dobiju uređene skupine.

Prvi korak je odabir broja klastera, to jest skupina u koje će se jedinice promatranja razvrstati. Taj korak može se korigirati kako bi se dobila što bolja povezanost jedinica promatranja unutar klastera. Drugi korak ove metode je pronalaženje k-centar klastera prema unaprijed zadanom broju klastera. Nakon toga se jedinice promatranja razvrstavaju na osnovi njihove udaljenosti od k-centra. Najviše korištena mjera razvrstavanja je Wardova metoda, iako se mogu primjenjivati i druge mjere razvrstavanja.

Da bi se maksimizirala povezanost jedinica promatranja unutar klastera kod primjene k-sredine metode primjenjuje se analiza varijance (ANOVA). Primjena ANOVA-e u klaster analizi je u tome da se razvrstavanje vrši u određen broj skupina koje istraživač mora sam definirati. Ponekad je nemoguće znati u koliko skupina se promatrana pojava treba razvrstati pa se pri tome koristi ANOVA tako da se krene od manjeg broja skupina i provede razvrstavanje putem k-sredine klaster analize. Nakon toga se izračunava ANOVA, to jest povezanost unutar skupine odnosno nepovezanost skupina. Ako je značajnost veća od 0,05 (5%), potrebno je povećati broj skupina dok ova vrijednost ne bude manja od 0,05. Međutim, kada postoji veći broj kriterija po kojima se provodi klaster analiza, nije moguće uvijek imati odgovarajuću razinu značajnosti. „Ako se neka jedinica nikako ne može klasterirati ni u višim fazama klasteriranja, ona se smatra netipičnom vrijednošću (outlier)“ (Pivac, 2009, 221) i isključuje se iz analize. Isključivanje jedinica analiza povlači sa sobom da se ne razvrstavaju sve jedinice promatranja već samo neke te se sam postupak klasteriranja dovodi u pitanje.

3.3 Prednosti i nedostaci klaster analize

Osnovna prednost klaster analize je ta što se s pomoću nje osnovni skup može razvrstati na nekoliko srodnih klastera čije su jedinice promatranja međusobno povezane. Na taj se način formira homogeni skup gdje su pojedine jedinice toga skupa međusobno povezane. Hijerarhijskom klaster analizom dobiju se podaci o tome koje su jedinice promatranja najviše povezane. Ta se povezanost prikazuje dendrogramom.

Klaster analiza ne može dati statističke zaključke o populaciji. Na njezine rezultate utječu mnogi elementi pa analiza daje rezultate bez obzira na to kakvi se podaci koriste. Na taj način varijabilitet jedinica promatranja utječe na to da rješenja nisu jedinstvena. Ponekad ne postoji statističko značajno razvrstavanje već sam istraživač mora odrediti broj skupina prilikom primjene k-sredine klaster analize.

Prilikom primjene klaster analize moguće je koristiti različite mjere odstupanja i razvrstavanja te u skladu s korištenim mjerama konačni rezultat može varirati. Zbog toga je u klaster analizi potrebno

upotrijebiti više različitih mjera odstupanja i razvrstavanja kako bi rezultati bili bolji za istraživača. Ako više različitih mjera daje isti rezultat, istraživač može biti siguran u rezultate klaster analize.

Kada postoji više jedinica promatranja koje su heterogene, ponekad nije moguće dobiti određen broj skupina koje su statistički povezane, a čija povezanost se izračunava s pomoću ANOVA-e, već sam istraživač treba odrediti taj broj. Na taj način razvrstavanje nije naučno već intuitivno. Klaster analiza će u tom slučaju razvrstati promatrane jedinice u točno određeni broj skupina tako da postoji mogućnost veoma malog broja jedinica u skupinama, pa čak da skupina ima jednog člana. Kada se jednočlani klaster (ili kada ima mali broj jedinica promatranja u jednom klasteru) pojavi, istraživač mora odlučiti je li važeća strukturalna komponenta u uzorku ili je odbačena kao nereprezentativna. Kada je neko promatranje odbačeno, istraživač bi se trebao vratiti na klaster analizu i početi je ponovno.

Rezultat klaster analize ovisi o varijabilitetu promatranih jedinica. pa ako se mijenjaju te jedinice dodavanjem ili oduzimanjem istih iz promatranja moguće je dobivati različite rezultate klaster analize. Istraživač tada mora voditi računa koje sve jedinice promatranja uključuje u analizu i varijabilitet među tim jedinicama jer sve to utječe na konačan rezultat.

Klaster analize u praksi se prijenjuju s drugim metodama da bi se dobili određeni značajno statistički rezultati. Njezina najveća zamjerka je da ona ne može na znanstveni način odrediti važnost klastera, jer se njezinom primjenom uvijek postiže neko razvrstavanje bez obzira na to kakvi se koriste podaci. „Krajnji rezultat klaster analize u potpunosti ovisi o varijablama koje su korištene kao osnova za mjerenje sličnosti pa istraživač mora voditi računa o procjeni utjecaja svake odluke prilikom izbora varijabli“ (Devčić et al., 2012, 19). Potrebno je naglasiti da je klaster metoda razvrstavanja nepristrana i transparentna. Uvažava konkretne matematičke izračune i za rezultat ima nepristrano razvrstavanje promatranih jedinica, naravno, ako su podaci koji se koriste u analizi također nepristrani (Pivac, 2009, 221).

4 Dosadašnja istraživanja primjene klaster analize u ekonomskim istraživanjima

Klaster analiza ima više svojih mogućnosti. Klaster analiza može se primijeniti u sljedećim slučajevima (Zahirović, 2005, 116):

- kada iz jednog skupa objekata treba formirati klase sa što homogenijim obilježjima
- kada na jednom skupu objekata treba utvrditi postojanje određenih kategorija (tipova)
- kada postoji mogućnost izvođenja određenih tretmana te je prijeko potrebno iz skupa objekata odrediti one klase na kojima se mogu primijeniti određeni tretmani
- kada je na jednom skupu objekata, na osnovi izmjerenih obilježja nužno izvesti odgovarajuće korektivne mjere u cilju poboljšanja kvaliteta mjere.

Najveću primjenu klaster analiza je pronašla u području ponašanja potrošača i razumijevanja tržišta gdje predstavlja najčešće korištenu metodu za identificiranje homogenih skupina potrošača, tzv. segmentaciju tržišta. Naime, razumijevanje tržišta često podrazumijeva segmentiranje potrošača u homogene skupine koje imaju slične karakteristike ili se slično ponašaju (Devčić et al., 2012, 18). Klaster analiza se može provoditi u ekonomskim istraživanjem odnosno konkretno u bankarstvu kako bi se razvrstali korisnici banaka u određene klasterne (Kazem Zadeh et al., 2011; Kumar et al., 2012) odnosno razvrstavanje gradova prema korisnicima banaka (Kowal et al., 2014).

Najraniji primjeri ove metode upravo su u segmentiranju tržišta. U radu Green et al. (1967) korištena je klaster analiza za klasifikaciju gradova u male skupine gdje je na osnovi četrnaest varijabli, uključujući veličinu grada, cirkulaciju novinama i dohotka po stanovniku i sl. izvršena segmentacija te je izabran jedan grad iz svake skupine koji je korišten kao sredstvo odabira test tržišta. Lawrence i

Jagdish (1975) su u svojem radu predstavili kako se klaster analiza može primjenjivati u marketinškom istraživanju. Kiel i Layton (1981) razvijaju taksonomiju potrošača za ponašanje vezano uz traženje informacija na tržištu automobila u Australiji.

Chakrapani (2004) je u svojem istraživanju pokušao dokazati da kada se kupuje sportski automobil, to nije isključivo na osnovi nečijih sredstava ili nečije dobi nego je to više lifestyle odluka. Uz sportski auto kupac nam predstavlja svoj način života koji je drukčiji od onih koji ne kupuju sportske automobile. Tako se putem klaster analize pokušavaju identificirati ljudi s odgovarajućim načinom života koji kupuju sportske automobile i na taj način kreirati ciljnu marketinšku kampanju.

Rašić Bakarić (2006) je u svojem istraživanju primijenila metodu nehijerarhijskog klasteriranja, „k-sredine“ metodu. Osnovni argument za primjenu ove metode klasteriranja je da se na ovaj način razvrstaju objekti istraživanja u pogodnije skupine. Ovdje se primijenilo razvrstavanje jedinica promatranja a ne razvrstavanje varijabli. Ona je svoje istraživanje provela u cilju mjerenja regionalnih nejednakosti.

Mihić je u svoja dva rada provela klaster analizu u cilju segmentacije tržišta potrošnje, gdje je u prvom radu (2006a) koristila multivarijatne tehnike izdvajanja različitih tipova utjecaja referentnih skupina te utvrdila mogu li oni poslužiti kao osnovica za segmentiranje tržišta potrošača. Klaster analizom izdvojena su tri segmenta tržišta potrošača. U drugom radu (2006b) koristeći se isto multivarijatnim tehnikama na osnovi stavova potrošača o prodajnom osoblju i o izgledu trgovine odredila je skupine faktora te utvrdila mogu li one poslužiti kao osnova za segmentiranje tržišta. S pomoću klaster analize ustanovljena su tri segmenta tržišta.

Kurnoga Živadinović (2007) je s pomoću klaster analize i drugih metoda višekriterijske analize klasificirala županije Hrvatske u veće skupine sličnih socioekonomskih obilježja. Pri tome se koristila hijerarhijskom i nehijerarhijskom klaster analizom. Primjenjujući različite metode hijerarhijske klaster analize, dobila je da je najbolje interpretabilno rješenje dobiveno Wardovom metodom s kvadriranim euklidskim udaljenostima. Ovom metodom odabrano je rješenje s četiri klastera te je provedena nehijerarhijska klaster analiza tih klastera primjenom k-sredine metoda.

Bevanda (2008) je pri odabiru financijskih institucija pokazala da su osnovne preferencije klijenata s obzirom na faktore imidža dovoljno različite da mogu poslužiti za diferenciranje značajnih tržišnih segmenata klijenata. Klaster analizom izdvojena su dva segmenta nazvana tradicionalisti i vizualisti.

Vasić i suradnici (2008) u svojem su istraživanju s pomoću dvostupnje klaster analize identificirali skupine klijenata banke. Oni su primjenom klaster metode s drugim metodama multivarijatne analize pokušali povećati prodaju bankarskih proizvoda. Zatim su putem klaster analize pokušali umjesto na dosadašnje klasifikacije bankarskih klijenata izvršiti segmentiranje klijenata na nekoliko prirodnih i jasno razdvojenih klastera, gdje bi banka bila u mogućnosti da zasebno za svaki klaster izradi poslovnu i marketinšku strategiju.

Kennedy i suradnici (2008) primijenili su klaster analizu u segmentiranju američkih kupaca hrane što se tiče njezine sigurnosti. U svojem radu primijenili su faktorsku i hijerarhijsku klaster analizu. Anić i suradnici (2010) u svojem su istraživanju s pomoću faktorske analize identificirali oblike orijentacije žena pri kupnji odjeće te primjenom k-sredine klaster analize segmentirali tržište i izdvojili pet različitih segmenata pri kupnji ženske odjeće.

Bahovec i Škrinjarić (2013) istražili su mogućnost primjene klaster analize s drugim metodama multivarijatne analize kroz optimizaciju portfelja u okviru Markowitzeva modela. Pokušali su povezati rezultate faktorske analize glavnih komponenti i klaster analize sa standardnim pristupom optimizacije portfelja. U klaster analizi primijenili su Wardovu metodu. Řezanková (2014) je rabila različit pristup korištenja klaster analize te skrenula pozornost na ovu vrstu multivarijatne analize i njezinu primjenu u ekonomskim podacima.

Iz ovih svih radova moguće je vidjeti da je upravo najveća primjena klaster analize u cilju segmentiranja pojedinih ciljnih tržišta u skladu s povezanostima promatranih jedinica. Ono što autori naglašavaju u svojim radovima je da klaster analiza nije revolucionarna metoda već da se obično prijenjuje uz druge metode multivarijatne analize i to prije svega metode faktorske analize i diskriminacijske analize.

5 Primjena klaster analize u bankarskom sektoru na prostoru Brčko distrikta BiH

Za potrebe ovoga rada razvrstat će se banke s područja Brčko distrikta BiH na osnovi njihovih rezultata poslovanja kroz pokazatelje profitabilnosti. Pokazatelji profitabilnosti odnosno rentabilnosti za razliku od drugih financijskih pokazatelja uzimaju u obzir kako podatke iz bilance stanja tako i podatke iz bilance uspjeha, za razliku od npr. pokazatelja likvidnosti koji uzimaju samo podatke unutar bilance stanja. Na taj način u isti se rang stavljaju sve banke bez obzira na iznos kapitala s kojima one raspolažu. Za mjerenje pokazatelja profitabilnosti banaka koristit će se sljedeći pokazatelji:

- povrat na aktivu (engl. Return On Assets – ROA)
- utjecaj na prinos kapitala (engl. Return On Equity – ROE)
- profitna marža (engl. Return On Sell – ROS)

Osnova za izračunavanje ovih pokazatelja su financijski izvještaji banaka za 2013. godinu koji su dostupni na njihovim web stranicama.

Tablica 1. Rezultati profitabilnosti banaka na prostoru Brčko distrikta BiH
Izvor: Istraživanja autora, 2015.

Rb.	Naziv banke	ROA	ROE	ROS
1.	Bosna bank international	0,3072	3,4113	9,9813
2.	Hypo Alpe-Adria-Bank	-5,8810	-34,3956	-94,6054
3.	Intesa Sanpaolo Banka	1,0102	7,1828	13,4641
4.	NLB Banka Tuzla	0,5280	5,2290	6,7593
5.	Privredna banka	-11,9272	-86,4049	-148,8828
6.	Raiffeisen Bank	1,1248	8,3333	16,0645
7.	Sberbank	0,3162	2,2741	4,6018
8.	Sparkasse Bank	0,9637	8,1014	11,7778
9.	UniCredit Bank	1,4901	9,6006	22,1141
10.	Bobar banka	0,4315	3,0982	4,6317
11.	Komercijalna banka	0,0478	0,3969	1,4767
12.	MF banka	0,1937	1,4263	2,4216
13.	Nova banka	0,5823	8,8819	9,8135
14.	NLB Razvojna banka	1,0071	9,9679	16,6011
15.	Pavlović International Bank	0,2494	1,7889	2,9367

Najprije će se izračunati hijerarhijska klaster metoda te će se zatim primijeniti k-mean metoda. Kod primjene svih klaster analiza kao mjera udaljenosti primijenit će se kvadratna Euklidova udaljenost, a kao mjera razvrstavanja metoda međusobnog razvrstavanja. Na osnovi matrice udaljenosti banke Hypo Alpe-Adria-Bank i Privredna banka najviše su udaljene od drugih banaka. Razlog tome je da su one poslovale s gubitkom u 2013. godini. Zbog toga ove banke utječu na same rezultate klaster analize.

Tablica 2. Matrica udaljenosti
Izvor: Istraživanja autora, 2015.

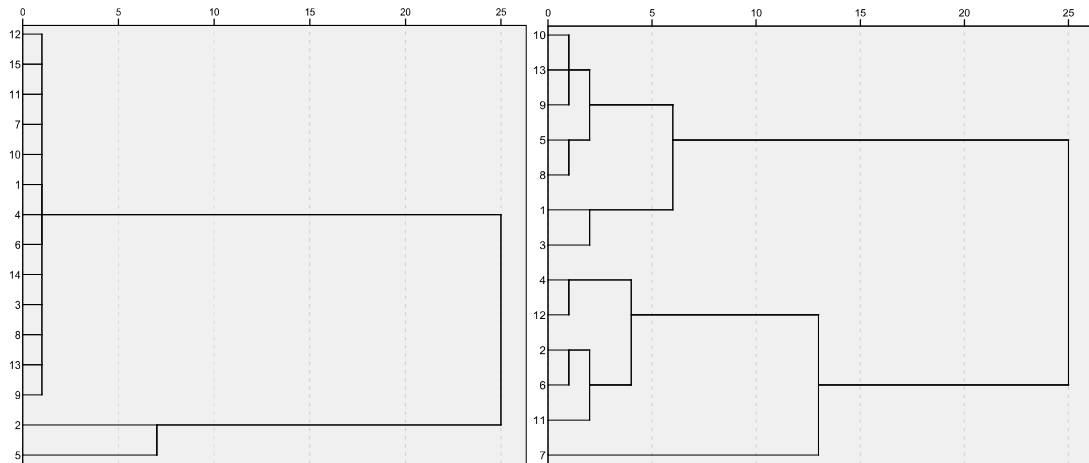
	1	2	3	4	5	...	14	15
1	.00	12406.03	26.84	13.73	33454.43	...	87.30	52.26
2	12406.03	.00	13455.26	11885.98	5687.56	...	14382.45	10861.36
3	26.84	13455.26	.00	49.00	35282.55	...	17.59	140.49
4	13.73	11885.98	49.00	.00	32776.36	...	119.54	26.52
5	33454.43	5687.56	35282.55	32776.36	.00	...	36839.93	30975.57
6	61.90	14122.66	8.09	96.58	36353.29	...	2.97	215.93
7	30.23	11225.14	103.11	13.43	31571.38	...	203.65	3.01
8	25.65	13170.23	3.69	33.62	34909.44	...	26.75	118.52
9	186.91	15613.44	80.89	255.80	38637.02	...	30.76	430.33
10	28.73	11293.63	95.03	9.07	31730.24	...	190.79	4.62
11	81.48	10477.43	190.67	51.48	30285.93	...	321.27	4.11
12	61.10	10734.34	155.74	33.38	30754.25	...	274.67	.40
13	30.03	12818.02	16.39	22.67	34420.57	...	47.43	97.71
14	87.30	14382.45	17.59	119.54	36839.9300	254.18
15	52.26	10861.36	140.49	26.52	30975.57	...	254.18	.00

Klaster analiza prvo pronalazi jedinice promatranja koje imaju najmanju udaljenost, to jest koje su povezane i formira se prvi par. Nakon toga ponovno se gleda koje su nakon tih jedinica najviše povezane pa se formira par itd. dok se ne spoje sve jedinice promatranja u završni par. Upravo anglomeracijski raspored u razini 1 pokazuje koje jedinice promatranja najmanje međusobno odstupaju. To su banke dvanaest i petnaest, a nakon toga banke sedam i deset te banke šest i četrnaest pa da bi se zatim prvi par povezao s bankom jedanaest. Postupak se ponavlja sve dok se u ovom slučaju sve banke ne povežu.

Tablica 3. Anglomeracijski raspored
Izvor: Istraživanja autora, 2015.

Razina	Klaster kombinacija		Koeficijenti	Razina klastera		Sljedeća razina
	Klaster 1	Klaster 2		Klaster 1	Klaster 2	
1	12	15	.400	0	0	4
2	7	10	.693	0	0	6
3	6	14	2.974	0	0	9
4	11	12	3.042	0	1	6
5	3	8	3.690	0	0	7
6	7	11	8.603	2	4	10
7	3	13	10.505	5	0	9
8	1	4	13.734	0	0	10
9	3	6	26.334	7	3	11
10	1	7	38.772	8	6	12
11	3	9	82.401	9	0	12
12	1	3	170.272	10	11	14
13	2	5	5687.560	0	0	14
14	1	2	23093.732	12	13	0

Grafički prikaz ove tablice daje se dendrogramom.



Slika 1. Rezultati predstavljeni dendrogramom
Izvor: Istraživanja autora, 2015.

Iz lijevog dijela slike dendrograma vidljivo je da su trinaest banaka spojene dok su banke dva i pet izdvojene od njih. Njihovi rezultati poslovanja utiču na sam rezultat hijerarhijske klaster analize. Na osnovi dendrograma ne može se dalje rasporediti preostalih trinaest banaka u skupine jer je prevelika razlika između navedenih banaka s ostalim bankama. Zbog toga je na desnoj strani slike prikazan dendrogram bez ove dvije banke kako bi se vidjela stvarna povezanost ostalih banaka. Izostavljanjem banaka koje su poslovale s gubitkom iz analize dobija se bolja slika povezanosti ostalih trinaest banaka. Na osnovi ove slike može se reći da se ostale banke povezuju u dva klastera. To pokazuje ovaj veliki prijeelaz između banke pet i šest. Hijerarhijskom klaster analizom je pokazano da se banke mogu razvrstati u tri klastera. Iako se i iz ove analize može vidjeti kojom klasteru pripada pojedina banka k-sredine, klaster analiza to prikazuje u tablici gdje je naznačeno kojom klasteru pripada koja banka. Primjenom k-sredine klaster analize razvrstat će se banke u tri klastera.

Tablica 4. Pripadnost klasteru
Izvor: Istraživanja autora, 2015.

Rb.	Klaster	Udaljenost
1.	3	2.051
2.	2	.000
3.	3	4.439
4.	3	2.680
5.	1	.000
6.	3	7.283
7.	3	5.743
8.	3	3.621
9.	3	13.397
10.	3	5.313
11.	3	9.397
12.	3	8.053
13.	3	3.542
14.	3	8.528
15.	3	7.425

Korištenjem k-sredine klaster analize kod svrstavanja u tri klastera dobio se slučaj jednočlanih klastera, to jest trinaest banaka je svrstano u jedan klaster, a po jedna banka u drugi i treći klaster.

Razlog tome je što u provedbi algoritma k-sredine klaster analize on razvrstava jedinice promatranja na osnovi mjere odstupanja i primjenom mjera razvrstavanja.

Tablica 5. Rezultati ANOVA-e
Izvor: Istraživanja autora, 2015.

	Klaster		Pogreška		F	Sig.
	Prosječni kvadrat	df	Prosječni kvadrat	df		
ROA	87.997	2	.193	12	456.863	.000
ROE	4424.175	2	11.884	12	372.275	.000
ROS	15649.882	2	40.547	12	385.966	.000

Budući da je i odstupanje između banke dva i pet dosta veća od međusobnog odstupanja ostalih banaka dobiven je ovakav rezultat. Rezultati ANOVA-e pokazuju da ne postoji značajno statističko odstupanje između korištenih varijabli za razvrstavanje podataka, to jest ovo razvrstavanje je znanstveno utemeljeno. Da bi se dobio bolji rezultat, potrebno je povećati broj klastera na četiri.

Tablica 6. Pripadnost klasteru
Izvor: Istraživanja autora, 2015.

Rb.	Klaster	Udaljenost
1.	1	5.369
2.	2	.000
3.	3	2.124
4.	1	3.420
5.	4	.000
6.	3	1.149
7.	1	.259
8.	3	3.247
9.	3	7.216
10.	1	.598
11.	1	3.856
12.	1	2.517
13.	3	5.182
14.	3	2.078
15.	1	1.897

Kao što su rezultati pokazali, i dalje su dva klastera jednočlana, ali je zato treći klaster podijeljen na dva dijela koji sada daje bolje rezultate. Rezultati ANOVA-e pokazuju da su povezane varijable i kod ovakvog razvrstavanja.

I daljim povećanjem broja klastera sig. kod ANOVA-e je i dalje .000. No i dalje se javljaju jednočlani klasteri zbog čega će se ostati na ovom broju klastera. Razlog zbog čega su rezultati ANOVA-e takvi treba tražiti u povezanosti varijabli u ovom istraživanju.

Tablica 7. Rezultati ANOVA-e
Izvor: Istraživanja autora, 2015.

	Klaster		Pogreška		F	Sig.
	Prosječni kvadrat	df	Prosječni kvadrat	df		
ROA	59.244	3	.052	11	1136.561	.000
ROE	2990.317	3	1.819	11	1643.885	.000
ROS	10547.184	3	13.162	11	801.348	.000

Velika povezanost varijabli promatranja i manjeg broja korištenih varijabli utječe da se ne može ANOVA-om odrediti broj klastera. Zbog toga sam istraživač mora reći u koliko klastera će razvrstati promatrane banke.

Tablica 8. Rezultati korelacije varijabli
Izvor: Istraživanja autora, 2015.

	ROA	ROE	ROS
ROA	1	0.996	0.993
ROE	0.996	1	0.982
ROS	0.993	0.982	1

Na kraju ovoga primjera primjene klaster analize prikazat će se podjela banaka u četiri klastera. Banke će se poredati prema rezultatima ROA-e.

Tablica 9. Rezultati klaster analize
Izvor: Istraživanja autora, 2015.

Klaster	Ime banke	ROA	ROE	ROS
1. Visoko profitabilne banke	UniCredit Bank	1,4901	9,6006	22,1141
	Raiffeisen Bank	1,1248	8,3333	16,0645
	NLB Razvojna banka	1,0071	9,9679	16,6011
	Intesa Sanpaolo Banka	1,0102	7,1828	13,4641
	Sparkasse Bank	0,9637	8,1014	11,7778
	Nova banka	0,5823	8,8819	9,8135
2. Profitabilne banke	NLB Banka Tuzla	0,5280	5,2290	6,7593
	Bobar banka	0,4315	3,0982	4,6317
	Sberbank	0,3162	2,2741	4,6018
	Bosna bank international	0,3072	3,4113	9,9813
	Pavlović International Bank	0,2494	1,7889	2,9367
	MF banka	0,1937	1,4263	2,4216
	Komercijalna banka	0,0478	0,3969	1,4767
3. Neprofitabilne banke	Hypo Alpe-Adria-Bank	-5,8810	-34,3956	-94,6054
4. Neprofitabilne banke	Privredna banka	-11,9272	-86,4049	-148,8828

Budući da ovaj primjer predstavlja prvu krajnju alternativu za korištenje klaster analize, postoji manji broj varijabli po kojima se jedinice promatranja klasterificiraju. Osim toga postoji velika povezanost između tih varijabli. Zbog toga ANOVA ne može pomoći u određivanju broja skupina. Druga krajnja alternativa je da postoji velik broj međusobnih varijabli koje nisu međusobno povezane.

6 Rasprava

Ovim istraživanjem želio se pokazati način na koji se klaster analiza može primjenjivati u ekonomskim istraživanjem. Na praktičnom primjeru banaka na prostoru Brčko distrikta BiH banke su razvrstane u četiri klastera i to u visoko profitabilne banke, profitabilne banke te u dvije skupine gdje su uvrštene neprofitabilne banke. Međutim, dobiveni klasteri banaka trebaju se uzeti sa zadržkom jer su banke ispitivane samo s tri pokazatelja i to: povrat na aktivu, utjecaj na prinos kapitala i profitna marža. Svi pokazatelji tiču se isključivo profitabilnosti, stoga su banke razvrstane na ovaj način. Glavni nedostatak ovoga rada je što su korištena samo tri pokazatelja i razvrstavanje je provedeno na osnovi njih.

Rezultati ove analize pokazuju da je UniCredit banka najprofitabilnija i pokazuje najbolje rezultate kod pokazatelja ROA-e i ROS-a dok NLB Razvojna banka pokazuje najbolje rezultate kod pokazatelja ROA-e. Rezultati klaster analize razvrstali su i napravili jaz između visoko profitabilnih i profitabilnih banaka. Kod ROA pokazatelja taj jaz je između 0,5823 i 0,5280 te su banke koje imaju rezultate ovoga pokazatelja veći od 0,55 visoko profitabilne, a banke koje imaju manji vrijednost ovoga pokazatelja su profitabilne banke. Kod pokazatelja ROA-e taj jaz je između 8,8819 i 5,2290. Upoređujući ovaj jaz s prethodnim može se zaključiti da postoji daleko veći jaz kod ovoga pokazatelja koje razvrstava banke u visoko profitabilne i profitabilne banke u odnosu na polazatelj ROA-e. Kod pokazatelja ROS-a ovaj jaz je narušen kod banke Bosna bank international jer ima bolju vrijednost ROS pokazatelja u odnosu na Novu banku, iako je ova banka svrstana u profitabilne banke, a ne u visoko profitabilne banke. Razlog tome treba tražiti u druga dva pokazatelja s kojima su svrstane banke.

Primjena klaster analize kod razvrstavanja banaka donosi rezultate koji su logični slijed rezultata provedenih analiza profitabilnosti banke, te je i bez ove analize moguće razvrstati banke po pokazatelju profitabilnosti. Ovim razvrstavanjem klaster analiza napravila je jaz po kojem se banke razvrstavaju. Kao što se može vidjeti, neprofitabilne banke imaju sve negativne vrijednosti korištenih pokazatelja, ali klaster analiza je upravo te banke podijelila u dvije skupine pa se može reći da Privredna banka pokazuje najlošije rezultate pokazatelja profitabilnosti. Ako bi se nastavio ovakav trend ove banke, njezini gubici bi premašili njezin osnovi kapital čime bi bila narušena njezina financijska sigurnost. Malo manje je pogođena Hypo Alpe-Adria-Bank, ali i kod nje je ostvaren gubitak u poslovanju. Ove banke koje su neprofitabilne i svrstane u treću i četvrtu skupinu trebale bi poraditi na svojem poslovanju te ostvariti veću profitabilnost kako bi poboljšali svoju financijsku sigurnost.

Banke na osnovi razvrstavanja s pomoću klaster analize mogu vidjeti gdje se nalaze u odnosu na konkurenciju te kojoj skupini pripadaju. Korisnicima rezultati ove analize mogu poslužiti da odaberu visoko profitabilne banke jer su na taj način njihovi mogući depoziti sigurniji nego u odnosu na ostale skupine banaka.

7 Zaključak

Primjena klaster analize u ekonomskim istraživanjima je vrlo široka, ali rijetko se primjenjuje sama. Za razliku od faktorske analize koja se isto bavi razvrstavanjem podataka kod klaster analize nema uvjeta. Primjena ove analize usmjerena je samo na razvrstavanje jedinica promatranja putem određenih odstupanja odnosno njihove međusobne udaljenosti.

Primjenom hijerarhijske klaster analize uspoređuju se varijable istraživanja s pomoću određenih mjera odstupanja i mjera razvrstavanja. Korištenjem različitih mjera odstupanja i razvrstavanja moguće je dobiti drukčiji konačan rezultat. Primjenom nehijerarhijskih klaster analiza može razvrstati promatrane jedinice u određene skupine koje su povezane prema određenim varijablama promatranja. Na taj se način dobiju homogene skupine koje su međusobno heterogene.

Klaster analiza većinom se koristi kao pomoćna analiza u istraživanjima. Uglavnom je vezana za preliminarna istraživanja kada je potrebno izvršiti redukciju podataka ili dobivanje homogenih skupina za potrebe istraživanja. Klaster analizu moguće je koristiti u prognoziranju kretanja nekih jedinica prema skupnoj pripadnosti kada se ova analize može koristiti kao osnova za prognostiku.

Kao i sve druge metode, klaster analiza ima brojne prednosti i nedostatke koje istraživaču moraju biti na umu kada se primjenjuje. Osnovni problemi su da klaster analiza uvijek razvrstava jedinice promatranja bez obzira na njihovu međusobnu homogenost te mogućnost nastanka jednočlanog klastera koji istraživaču postavlja pitanje da li da tu jedinicu promatranja izostavi iz daljnje analize i da ponovno provede postupak klasterizacije ili da ostavi taj rezultat.

Važno je znati rezultate dobivene ovom metodom interpretirati na pravi način što je moguće samo ako se dovoljno poznaje teorijska podloga istraživanja. Primjena klaster analize je raznovrsna pa se može koristiti u svim istraživanjima gdje postoji ili velik broj jedinica promatranja ili velik broj varijabli, a potrebno je razvrstati podatke radi bolje analize.

Literatura

Anić, I.-D., Mihić, M., Jurić, I. (2010), Kupovna orijentacija žena na tržištu odjeće, *Privredna kretanja i ekonomska politika*, 123, str. 27-55.

Bahovec, V., Škrinjarić, T., (2013), Mogućnosti optimizacije portfelja na Zagrebačkoj burzi uz pomoć odabranih metoda multivarijatne analize, *Ekonomski pregled*, 64 (1), str. 3-29.

Bahovec, V., Dumičić, K., Palić, I. (2011), Multivarijatna analiza pokazatelja društveno-ekonomskog razvoja u odabranim europskim zemljama, *Zbornik Ekonomskog fakulteta u Zagrebu*, 9 (1), str. 89-107.

Bevanda, A. (2008), Segmentacija finansijskog tržišta u Federaciji Bosne i Hercegovine na temelju čimbenika imidža, *Market*, 20 (2), str. 179-193.

Chakrapani, C., (2004), *Statistics in Market Research*, Arnold, London.

Devčić, K., Tonković Pražić, I., Župan, Ž., (2012), Klaster analiza: primjena u marketinškim istraživanjima, *Zbornik radova Međimurskog veleučilišta u Čakovcu*, 3 (1), str. 15-22.

Green, P. E., Frank, R. E., Robinson, P. J., (1967), Cluster analysis in test market selection, *Management Science*, 13, str. 387-400.

Kazem Zadeh, R. B., Faraahi, A., Mastali, A., (2011), Profiling bank customers behaviour using cluster analysis for profitability, *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management*, str. 458-467.

Kiel, G. C., Layton, R. A., (1981), Dimensions of Consumer Information Seeking Behavior. *Journal of Marketing Research*, 18, str. 233-239.

Kennedy E. et al. (2008), Segmentation of US consumers based on food safety attitudes, *British Food Journal*, 110 (7), str. 691-705.

Kovačić, Z., (1994), *Multivarijaciona analiza*, Ekonomski fakultet, Beograd.

- Kowal, R., Yeleyko, Y. I., Kharkhalis V. I., (2014), Application of cluster analysis to assess the regional development of foreign banking in Ukraine, *Studia i Materiały. Miscellanea Oeconomicae*, 18 (4), str. 255-261.
- Kumar, M. V., Chaitanya, M. V., Madhavan, M. (2012), Segmenting the Banking Market Strategy by Clustering, *International Journal of Computer Applications*, 45 (17), str. 10-15.
- Kurnoga Živadinović, N., (2007), Multivarijantna klasifikacija županija Hrvatske, *Zbornik Ekonomskog fakulteta u Zagrebu*, 5 (1), str. 1-15.
- Kurnoga Živadinović, N., Sorić, P., (2008), Klaster analiza županija Hrvatske prema sredstvima dobivenim iz programa Europske unije, *Zbornik Ekonomskog fakulteta u Zagrebu*, 6 (1), str. 193-207.
- Lawrence, S., Jagdish, S. N., (1975), Cluster analysis and its applications in marketing research, Faculty working papers, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign.
- Markić, B., (2011), Neuronska mrežna klasifikacija u menadžerskom računovodstvu, *Informatol*, 44 (3), str. 200-206.
- Mihić, M., (2006a), Segmentacija tržišta temeljem osjetljivosti potrošača na vrstu utjecaja referentnim grupa: multivarijantna analiza, *Tržište*, 18 (1-2), str. 15-28.
- Mihić, M., (2006b), Segmentacija kupaca u supermarketima na osnovi njihova zadovoljstva prodajnim osobljem i izgledom prodavaonice: multivarijantna analiza *Ekonomski pregled*, 57 (12), str. 919-938.
- Mihić, M., Kursan I., (2010), Segmentacija tržišta na temelju impulzivnosti u kupovnom ponašanju potrošača, *Ekonomski vjesnik*, 23 (2), str. 386-398.
- Mooi, E., Sarstedt, M., (2011), *A Concise Guide to Market Research*, Springer-Verlag, Berlin
- Orehovački, T., (2013) Metodologija vrjednovanja kvalitete u korištenju aplikacijama web 2.0, Doktorski rad, Fakultet organizacije i informatike, Varaždin.
- Pivac, S. (2009), Statističke metode – integrirana predavanja, Ekonomski fakultet, Split.
- Prakash, M. M., Dagaonkar, A., (2011), Application of cluster analysis to physicochemical parameters of Munj Sagar Talab, Dhar (Madhya Pradesh, India), *Recent Res. Sci. Technol.* 3, str. 41-50.
- Rašić Bakarić, I. (2006), Primjena faktorske i klaster analize u otkrivanju regionalnih nejednakosti, *Privredna kretanja i ekonomska politika*, 15 (105), str. 53-76.
- Řezanková, H., (2014), Cluster Analysis of Economic Data, *Statistika*, 94 (1), str. 73-86.
- Vasić V., Banićević, D., Vojvodičan, M., (2008), Algoritam dvostupnje klaster analize u identifikaciji grupa klijenata banke, *Zbornik radova YuInfo2008*, 6 pages, CD edition.
- Vujović, S., Kolaković, S., Bečelić-Tomin, M., (2013), Procena kvaliteta vode značajno izmenjenih vodnih tela na teritoriji Vojvodine primenom multivarijacionih statističkih metoda, *Hemijska industrija*, 67 (5), str. 823-833.
- Zahirović, S., (2005), Multivarijaciona analiza, osnove teorije i primjene u marketinškim istraživanjima, Infograf, Tuzla.