

Klasifikacija promjenjivih zvijezda pomoću metoda strojnog učenja

Božan, Domagoj

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:853647>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-03-29**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

DOMAGOJ BOŽAN

KLASIFIKACIJA PROMJENJIVIH ZVIJEZDA
POMOĆU METODA STROJNOG UČENJA

Diplomski rad

Zagreb, 2020.

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI STUDIJ
FIZIKA; SMJER ISTRAŽIVAČKI

Domagoj Božan

Diplomski rad

**Klasifikacija promjenjivih zvijezda
pomoću metoda strojnog učenja**

Voditelj diplomskog rada: dr. sc. Lovro Palaversa

Ocjena diplomskog rada: _____

Povjerenstvo: 1. _____

2. _____

3. _____

Datum polaganja: _____

Zagreb, 2020.

Zahvaljujem se svome mentoru, doktoru Lovri Palaversi, na danoj prilici za pisanje ovog rada iz područja koje me zanimalo te za vrijeme koje je posvetio kako bi ovaj rad bio što veće kvalitete. Također se zahvaljujem i profesorima Matku Milinu i Ivici Smoliću na vremenu uloženom za ocjenjivanje ovog rada te što su pristali izdvojiti vrijeme da budu članovi povjerenstva.

Zahvaljujem se svojim roditeljima, Ljiljani i Anti, na potpori, ohrabrenju i razumijevanju tijekom ovog studija. Posebno se zahvaljujem svojoj sestri Marini koja je također bila velika podrška, a i uložila je svoje slobodno vrijeme u lektoriranje ovog rada.

Sažetak

Proučavanje promjenjivih događaja u svemiru koji se događaju na relativno kratkim vremenskim skalama od velikog je interesa za astrofizičare jer se iz njih mogu dobiti mnoge korisne informacije. Isto vrijedi i za promjenjive zvijezde, to jest zvijezde čiji sjaj nije konstantan. Promatranjem takvih zvijezda moguće je odrediti svojstva zvijezda, u nekim slučajevima i jednostavnije nego kod promatranja nepromjenjivih zvijezda, kao i udaljenosti u svemiru. Kako bismo bili sigurni da su dobivene informacije ispravne, važno je pravilno odrediti kojoj vrsti pripada promjenjiva zvijezda. Međutim, ručno određivanje kategorije za svaku od stotina tisuća promjenjivih zvijezda koje se opažaju u pregledima neba veoma je dug i težak proces. Kako bi se isti ubrzao i automatizirao, mogu se koristiti metode strojnog učenja. U ovom radu korištene su metode klasifikacije i grupiranja sadržane u Scikit-Learn paketu u programskom jeziku Python kako bi se kategoriziralo oko 100000 promjenjivih zvijezda opaženih u Catalina pregledu neba. Za klasifikaciju korišten je Support Vector Machine (SVM) algoritam nadziranog strojnog učenja, a za grupiranje Gaussian Mixture Model (GMM) algoritam nenadziranog strojnog učenja. Dobiveni rezultati uspoređeni su s kategorizacijom provedenom u Catalina pregledu neba kako bi se dobio uvid u to koja metoda je preciznija. Rezultati pokazuju da je SVM algoritam bolji izbor za klasifikaciju ovakvog tipa podataka, ali i on pokazuje neka odstupanja od Catalina pregleda neba. Zaključeno je da se metode strojnog učenja mogu koristiti u traženu svrhu, ali potrebno je još proučavanja kako bi se proces optimizirao.

Ključne riječi: tranzijenti, promjenjive zvijezde, strojno učenje, klasifikacija, grupiranje

Classification of Variable Stars Using Machine Learning Methods

Abstract

Studying variable cosmic events which happen on relatively short time scales is of great interest for astrophysicists because of a plenitude of useful information one can receive from them. The same is true for variable stars, which are stars whose light is not constant. By observing such stars it is possible to determine properties of stars, in some cases with less difficulty than by observing non-variable stars, and cosmic distances. To assure that the received information is correct, it is of great importance to correctly determine the type of each variable star. However, determining categories of hundreds of thousands of variable stars in sky surveys is a very long and difficult process. In order to accelerate and automatise said process one can use the machine learning methods. In this paper, the methods of classification and clustering contained in the Scikit-Learn package of the Python programming language were used to categorise roughly 100000 variable stars observed in the Catalina Sky Survey. The supervised learning Support Vector Machine (SVM) algorithm was used for classification and the unsupervised learning Gaussian Mixture Model (GMM) algorithm was used for clustering. Obtained results were compared with the categorisation from the Catalina Sky Survey to gain an insight in the precision of each method. The results show that the SVM algorithm is a better choice for classification of this type of data, but it also shows some deviations from the Catalina Sky Survey. It was concluded that the machine learning methods can be used for desired purposes, but further study is needed to optimise the process.

Keywords: transients, variable stars, machine learning, classification, clustering

Sadržaj

| | | |
|----------|----------------------------------------------------------|-----------|
| 1 | Uvod | 1 |
| 1.1 | Promjenjive zvijezde | 1 |
| 1.1.1 | Klasifikacija promjenjivih zvijezda | 2 |
| 1.2 | Motivacija | 5 |
| 2 | Fizika promjenjivih zvijezda i tranzijenata | 6 |
| 2.1 | Mehanizmi pulsacija u promjenjivim zvijezdama | 6 |
| 2.2 | Određivanje parametara iz pomrčinskih zvijezda | 9 |
| 2.3 | Modeli supernova tipa Ia | 13 |
| 3 | Pregledi neba i podatci | 15 |
| 3.1 | Određivanje perioda u pregledima neba | 15 |
| 3.2 | Krivulje sjaja u Catalina pregledima neba | 16 |
| 4 | Metodologija | 20 |
| 4.1 | Scikit-Learn | 20 |
| 4.1.1 | Support Vector Machine | 21 |
| 4.1.2 | Gaussian Mixture Model | 22 |
| 4.2 | Primjena metoda strojnog učenja na podatke | 22 |
| 5 | Rezultati i diskusija | 25 |
| 5.1 | Rezultati SVM algoritma | 25 |
| 5.2 | Rezultati GMM algoritma | 29 |
| 6 | Zaključak | 35 |
| | Literatura | 37 |

1 Uvod

1.1 *Promjenjive zvijezde*

Promjenjive su zvijezde čiji sjaj nije konstantan. Te fluktuacije u sjaju mogu biti uzrokovane na različite načine te mogu varirati od tisućinke magnitude do dvadeset magnituda dok periodi u kojima se fluktuacije odvijaju mogu trajati od djelića sekunde do nekoliko godina [1].

Proučavanje događaja u svemiru koji se događaju u relativno kratkim vremen-
skim skalama, od nekoliko sekunda do nekoliko godina, od velikog je interesa zbog korisnih informacija koje se mogu dobiti njihovim promatranjem. Jedna vrsta takvih događaja su tranzijenti događaji, odnosno događaji koji se za isti objekt ne ponavljaju jer objekt biva uništen u događaju, obično eksplozijom koja se manifestira kroz promjenu sjaja. Dobar primjer tranzijentnih događaja su supernove, pogotovo one tipa Ia. Supernove tipa Ia nastaju u dvojnim sustavima u kojima je jedna zvijezda bijeli patuljak koji prima masu od druge zvijezde u sustavu, koja može također biti bijeli patuljak ili može biti obična zvijezda, ovisno o tome radi li se o dvostruko ili jednostruko degeneriranom modelu, te u konačnici prijeđe Chandrasekharovu granicu što uzrokuje fuzijsku reakciju i bijeli patuljak eksplodira. O ovom procesu više će biti riječ u sljedećem poglavlju. Ono što je značajno kod supernova Ia jest da je zbog ovog procesa njihov luminozitet uvijek u rasponu od 3-5 milijardi luminoziteta Sunca, što ih čini veoma uočljivima. Također je otkriveno da im je maksimalni luminozitet povezan s oblikom krivulje sjaja. Te kvalitete čine ih standardnim svijećama, odnosno objektima s poznatim luminozitetom koji se mogu koristiti za određivanje udaljenosti u svemiru. Uz to, promatranje supernova Ia dovelo je i do otkrića da se svemir širi ubrzano [2].

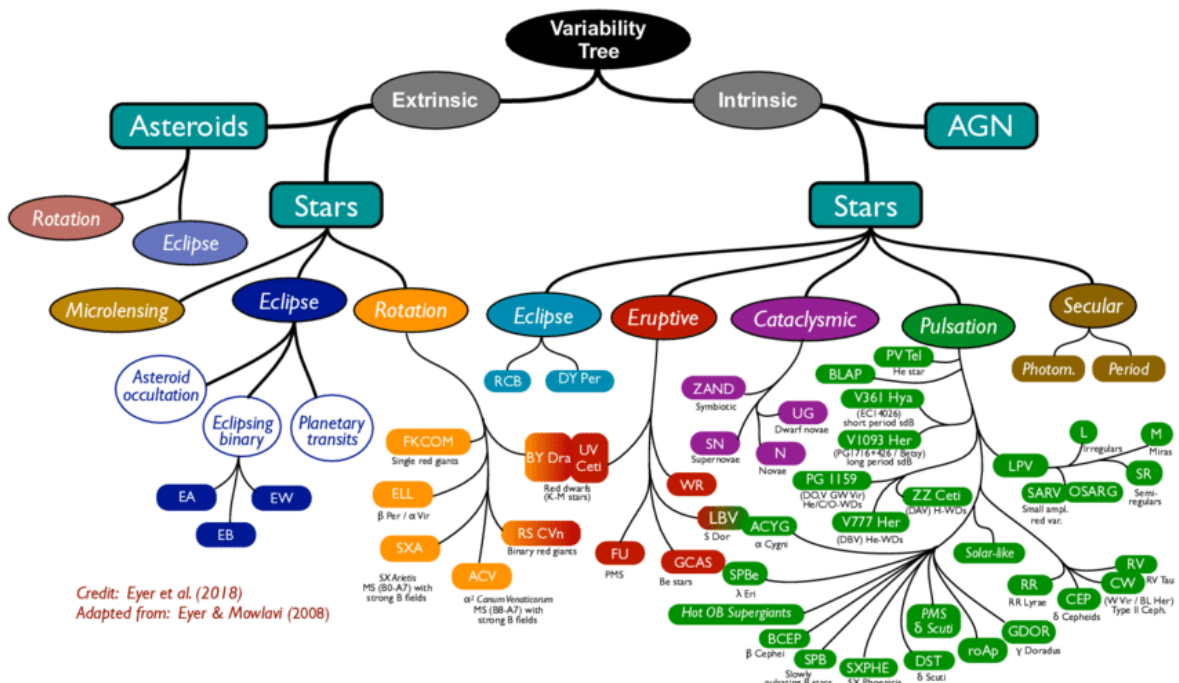
Isto vrijedi i za promjenjive zvijezde. Promatranjem promjenjivih zvijezda moguće je odrediti razna svojstva zvijezda, kao što su masa, radijus, luminozitet, temperatura, struktura, sastav ili evolucijski stadij, u nekim slučajevima jednostavnije nego promatranjem običnih zvijezda. Njihova se korisnost također očituje i kod kozmičke ljestvice udaljenosti. Kozmička ljestvica udaljenosti jest niz mjerenja različitih objekata i pojava koje se nadovezuju jedna na drugu kako bi se odredile sve veće udaljenosti u svemiru, a nekoliko koraka u toj ljestvici ovisi o promjenjivim zvijezdama. Manje udaljenosti moguće je odrediti korištenjem paralakse, ali ona je veoma ma-

lena i za najbliže zvijezde te brzo postaje nepraktična za korištenje kako udaljenosti rastu. Način na koji promjenjive zvijezde doprinose ljestvici uglavnom je preko relacije između perioda i luminoziteta - zvijezde s dužim periodima sjajnije su od zvijezda unutar iste kategorije pulsirajućih promjenjivih zvijezda s kraćim periodima. Pomoću paralakse moguće je kalibrirati ovu relaciju za pulsirajuće zvijezde kao što su RR Lyrae i cefeide koje se zatim koriste kao daljnji koraci u ljestvici pomoću kojih se mogu mjeriti udaljenosti do susjednih galaksija. Za još veće udaljenosti koriste se ranije spomenute supernove tipa Ia [3].

Iz navedenog se vidi važnost ispravnog razaznavanja promjenjivih zvijezda pri promatranjima neba. Promatrajući razlike u veličini fluktuacija i periodu promjenjivih zvijezda te različite načine na koje dolazi do promjena sjaja u zvijezdama, moguće je svrstati promjenjive zvijezde u različite kategorije.

1.1.1 Klasifikacija promjenjivih zvijezda

Podjelu promjenjivih astrofizičkih izvora možemo jednostavno vizualizirati pomoću tzv. "stabla promjenjivosti" (eng. *variability tree*), prikazanog na slici 1.1.



Slika 1.1: Stablo promjenjivosti, metoda za vizualizaciju podjele promjenjivih zvijezda na različite podvrste. Slika je preuzeta iz [4].

Prvotna podjela promjenjivih izvora koje opažamo jest na ekstrinzične i intrinzične. Ekstrinzični su izvori oni čiju promjenjivost uzrokuju vanjski uvjeti koji ne utječu na

svojstva same zvijezde, a intrinzični su oni kod kojih fluktuacije u sjaju uzrokuju fizičke promjene u ili na izvoru. Osim zvijezda, ekstrinzični izvori mogu još biti i asteroidi, a intrinzični mogu biti i aktivne galaktičke jezgre. Asteroidi su vidljivi jer reflektiraju svjetlost te im promjenjivost može uzrokovati njihova rotacija, koja uzrokuje da se količina svjetlosti koju reflektiraju promijeni, ili zasjenjenje. Aktivne galaktičke jezgre snažni su izvori zračenja koji se opažaju u središtima galaksija, a svjetlost koja se opaža s izvora može se brzo mijenjati, od nekoliko sati do nekoliko mjeseci.

Promjenjive zvijezde možemo dalje podijeliti u kategorije ovisno o točnom uzroku promjenjivosti. Ekstrinzične promjenjive zvijezde dalje se dijele na one čije su promjene uzrokovane rotacijom, pomrčinama ili efektom mikroleće. Efekt mikroleće baziran je na efektu gravitacijske leće, a događa se kada jedna zvijezda prođe ispred druge iz perspektive promatrača. Udaljenija zvijezda je sjajnija, a bližu se obično ne može vidjeti sa Zemlje i kada prođe ispred sjanije zvijezde njezina gravitacija uzrokuje prividno povećanje njezinog sjaja [5]. Za razliku od jačih efekata gravitacijske leće, kod efekta mikroleće leća prođe kraj izvora u relativno kratkom vremenskom periodu, od nekoliko sekundi do nekoliko godina, za razliku od milijuna godina kod jakih ili slabih gravitacijskih leća [6]. To omogućuje da se izvor promatra kao promjenjiva zvijezda.

Pomrčine mogu biti uzrokovane prolaskom asteroida, planeta ili kakvog drugog materijala između izvora i promatrača. Kada je riječ o prolasku druge zvijezde, radi se o pomrčinskim dvojnim zvijezdama. Njihove podvrste možemo odrediti iz oblika njihovih krivulja sjaja [7].

Rotacijske su promjenjive zvijezde vrsta promjenjivih zvijezda s nejednolikom raspodjelom površinskog sjaja. Ta nejednolikost može biti uzrokovana prisutnošću pjega ili toplinske ili kemijske nehomogenosti u atmosferi koje su uzrokovane magnetskim poljem čija se os ne podudara s osi rotacije. Rotacijske promjenjive zvijezde često su elipsoidnog oblika. Promjenjivost im uzrokuje osna rotacija u odnosu na promatrača [7].

Intrinzično promjenjive zvijezde dijelimo na eruptivne, kataklizmične, pulsacijske i zvijezde koje sekularno evoluiraju. Također postoji i posebna kategorija koja se može nazvati intrinzične pomrčinske zvijezde, a u koju spadaju zvijezde tipa R

Coronae Borealis (RCB) i DY Persei (DY Per). Zvijezde R Coronae Borealis tipa su zvijezde koje izbacuju materijal koji se kondenzira u oblake prašine kada dosegne dovoljno nisku temperaturu. Ti oblaci prašine zasjene zvijezdu sve dok ih ne otpuše tlak zračenja [8]. Dakle, riječ je o pomrčinskim zvijezdama koje su zasjenjene vlastitim materijalom. Zvijezde tipa DY Persei pokazuju ponašanje slično RCB zvijezdama, ali uz neke razlike. DY Per zvijezde imaju simetrične padove u krivuljama sjaja, oko 10 puta su tamnije od RCB zvijezda, jasno pokazuju linije ^{13}C u spektrima i mogu imati značajne količine vodika u atmosferi [9]. Također je pokazano da su DY Per zvijezde podvrsta AGB zvijezda bogata ugljikom [10].

Eruptivne promjenjive zvijezde pokazuju fluktuacije u svjetlosti uzrokovane burnim procesima i bakljama u kromosferama i koronama. Te fluktuacije često su praćene događajima u ljuskama i izljevima mase u obliku zvjezdanih vjetrova varirajućih intenziteta te interakcijama s međuzvjezdanim medijem [7].

Katakliksičke promjenjive zvijezde prolaze kroz eksplozivne termonuklearne procese na površini (nove) ili duboko u unutrašnjosti (supernove). Većina ovih zvijezda dvojni su sustavi čije si komponente međusobno utječu na evoluciju. Često se opažaju akrecijski diskovi koji okružuju vrućeg patuljka, a nastali su kao ostatak materije izgubljene iz komponente dvojnog sustava sa slabijom površinskom gravitacijom [7]. Proces nastajanja supernova iz dvojnih sustava bit će detaljnije opisan u sljedećem poglavlju.

Pulsirajuće promjenjive zvijezde pokazuju periodično širenje i skupljanje vanjskih slojeva. Te pulsacije mogu biti radijalne ili neradijalne: radijalno pulsirajuće zvijezde zadržavaju sferičan oblik, a kod neradijalno pulsirajućih zvijezda oblik periodično odstupa od sfernog i susjedna područja na površini mogu imati suprotne faze pulsiranja. U kategoriju pulsirajućih promjenjivih zvijezda spadaju RR Lyrae i cefeide, zvijezde koje od velikog značaja u astrofizici, a koriste se kao standardne svijeće [7].

Posljednja su kategorija zvijezde koje pokazuju sekularnu evoluciju, odnosno evoluiraju u ljudskim vremenskim skalama. Primjer ove kategorije Sakuraijev je objekt, koji prolazi evoluciju u fazi kasnog termalnog pulsa. Kasni termalni puls događa se kada se u zvijezdi koja se nalazi poslije grane asimptotskih divova (AGB, eng. *Asymptotic Giant Branch*) na H-R dijagramu, a čija je jezgra od ugljika i kisika okružena helijem te vanjskom ljuskom od vodika, zapali helij pa nastane termalni puls i zvijezda se vrati u AGB stanje, odnosno postane zvijezda s manjkom vodika koja sagorijeva helij.

Zvijezda nakon ovoga nastavlja evolucijski put po H-R dijagramu, ali ova faza traje vrlo kratko, oko 200 godina, prije nego zvijezda ponovno krene prema fazi bijelog patuljka [11].

1.2 Motivacija

Kako bi se sa sigurnošću utvrdila svojstva zvijezda i udaljenosti iz promjenjivih zvijezda važno ih je ispravno klasificirati. Iako je promjenjive zvijezde moguće klasificirati prema njihovim krivuljama sjaja, izrada krivulja sjaja za svaku od stotina tisuća promjenjivih zvijezda koje se opažaju u pregledima neba dug je i mukotrpan posao. Stoga je potrebno ubrzati i olakšati taj proces kako bi se povećala učinkovitost proučavanja tih podataka. U tu svrhu mogu se koristiti metode strojnog učenja, što je područje koje trenutno prolazi kroz period ekspanzivnog razvoja. Strojno učenje sadrži algoritme koji mogu klasificirati i grupirati velike količine podataka relativno brzo, a uz to je i automatizirano.

U ovom će se radu koristiti metode klasifikacije i grupiranja koje se nalaze u algoritmima strojnog učenja kako bi se klasificiralo oko 100000 promjenjivih zvijezda iz Catalina pregleda neba. Dobiveni rezultati bit će uspoređeni s već klasificiranim zvijezdama iz Catalina pregleda neba kako bi se dobio uvid u preciznost klasifikacije metodama strojnog učenja. Također će različiti algoritmi strojnog učenja biti međusobno uspoređeni kako bi se saznalo koja je metoda strojnog učenja najbolji izbor za ovaj zadatak.

2 Fizika promjenjivih zvijezda i tranzijenata

Kod proučavanja promjenjivih zvijezda i tranzijenata važno je prvo poznavati fiziku iza procesa koji se odvijaju u njima. U ovom će poglavlju biti pojašnjeni mehanizmi pulsacija u promjenjivim zvijezdama, načini na koje se mogu odrediti parametri iz promjenjivih zvijezda te procesi kroz koje nastaju supernove tipa Ia.

2.1 Mehanizmi pulsacija u promjenjivim zvijezdama

Pulsirajuće su zvijezde važne u astrofizici jer se promatranjem svojstava njihovog pulsiranja može steći uvid u njihovu strukturu. Uspoređivanjem opaženih karakteristika pulsiranja, kao što su periodi, amplitude i krivulje sjaja, s istim karakteristikama u numerički izvedenim modelima mogu se testirati teorije o strukturama i evoluciji zvijezda te dobiti detaljan uvid u unutrašnjost zvijezde.

Vjerojatno povijesno najvažnija karakteristika pulsirajućih promjenjivih zvijezda jest njihova relacija između perioda i luminoziteta. Ona zapravo proizlazi iz relacije između perioda i gustoće zvijezde, koja se može jednostavno izvesti. Radijalne oscilacije u pulsirajućoj zvijezdi posljedica su zvučnih valova koji rezoniraju u unutrašnjosti zvijezde, stoga je moguće procijeniti pulsacijski period T ako se uzme u obzir koliko bi trebalo zvučnom valu da prijeđe promjer radijusa R i konstante gustoće ρ . Adijabatska brzina zvuka dana je kao:

$$v_S = \sqrt{\frac{\gamma P}{\rho}}, \quad (2.1)$$

gdje je γ adijabatski indeks, a P tlak. Kako je pretpostavljena konstanta gustoća, moguće je odrediti tlak u hidrostatskoj ravnoteži:

$$\frac{dP}{dr} = -\frac{GM_r \rho}{r^2} = -\frac{G(\frac{4}{3}\pi r^3 \rho)\rho}{r^2} = -\frac{4}{3}\pi G \rho^2 r. \quad (2.2)$$

Integriranjem gornjeg izraza uz rubni uvjet da je $P = 0$ na površini dobije se tlak kao funkcija r :

$$P(r) = \frac{2}{3}\pi G \rho^2 (R^2 - r^2), \quad (2.3)$$

što daje pulsacijski period kao približno:

$$T \approx 2 \int_0^R \frac{dr}{v_S} \approx 2 \int_0^R \frac{dr}{\sqrt{\frac{2}{3}\gamma\pi G\rho(R^2 - r^2)}}, \quad (2.4)$$

odnosno:

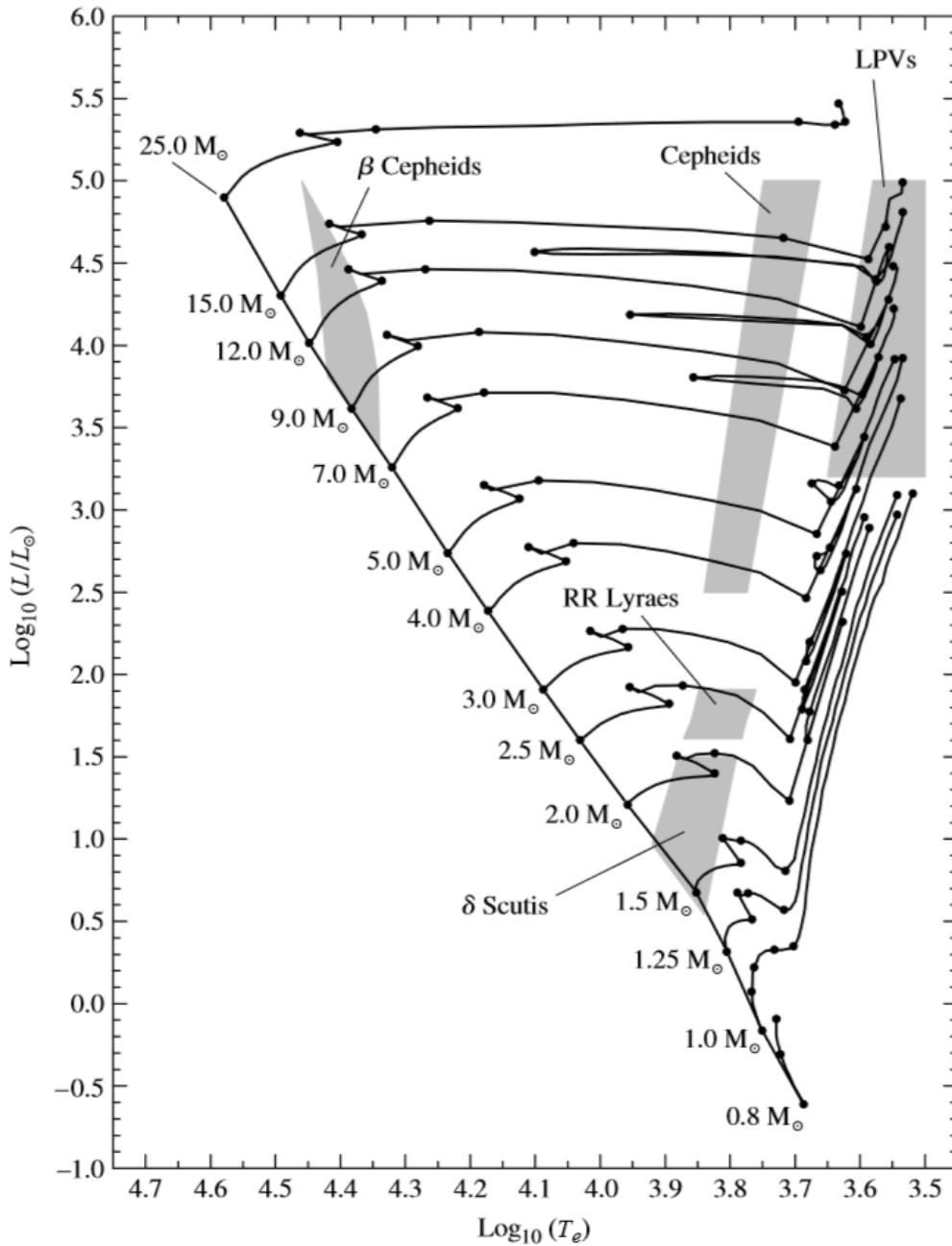
$$T \approx \sqrt{\frac{3\pi}{2\gamma G\rho}}, \quad (2.5)$$

gdje se vidi da je period pulsirajuće zvijezde obrnuto proporcionalan njezinoj gustoći. Ovo je moguće povezati s relacijom između perioda i luminoziteta proučavanjem H-R dijagrama. Relacija 2.5 pokazuje da se pulsacijski period smanjuje od superdivova vrlo malene gustoće do jako gustih bijelih patuljaka. Promatrajući položaj pulsirajućih promjenjivih zvijezda na H-R dijagramu, prikazan na slici 2.1, vidimo da je područje koje one zauzimaju, tzv. "traka nestabilnosti" (eng. *instability strip*), otprilike paralelno s osi luminoziteta, što uzrokuje relaciju između perioda i luminoziteta [12].

Uz relaciju između perioda i luminoziteta, važno je razumjeti i mehanizme pulsiranja zvijezda. Zvučni valovi u radijalnim modovima zvjezdanog pulsiranja u osnovi su stojni valovi. Kako bi opisao mehanizme koji pogone te stojne valove, Arthur Eddington [13] je predložio da su pulsirajuće zvijezde toplinski strojevi. Plinovi koji tvore slojeve zvijezde obavljaju rad dok se šire i skupljaju u pulsacijskom ciklusu. Ako je rezultatni rad u sloju pozitivan, taj sloj pogoni oscilacije, a ako je negativan, guši ih. Također, ako je ukupan rad u svim slojevima pozitivan, amplitude oscilacija će se povećavati, a ako je negativan, smanjivat će se. Promjene u amplitudama pulsacija nastavljaju se dok se ne postigne ravnoteža, odnosno dok ukupan rad u svim slojevima nije jednak nuli. Kako je riječ o toplinskom stroju, rezultatni rad koji vrši svaki sloj zvijezde unutar jednog ciklusa jest razlika topline koja teče u plin i topline koja napušta plin. Da bi sloj pogonio oscilacije, toplina mora ulaziti u sloj tijekom visokotemperaturnog dijela ciklusa i napuštati ga tijekom niskotemperaturnog dijela, odnosno pogonski slojevi pulsirajuće zvijezde moraju apsorbirati toplinu za vrijeme maksimalne kompresije. Maksimalni će tlak nastupiti nakon maksimalne kompresije i oscilacije će biti pojačane [12].

Pitanje koje se sljedeće postavilo jest u kojem se dijelu zvijezde može odvijati taj proces. Eddingtonovo rješenje za taj problem bio je model ventila. Ako sloj zvijezde

Stellar Pulsation



Slika 2.1: Položaj pulsirajućih promjenjivih zvijezda na H-R dijagramu. Slika je preuzeta iz [12].

postane neprozirniji tijekom kompresije, može zapriječiti tok energije prema površini te pogurati površinske slojeve prema gore. Kako šireći sloj postaje sve prozirniji, zarobljena se toplina otpušta i sloj pada natrag u početno stanje i proces počinje iznova. Da bi ovaj model bio valjan, neprozirnost se mora povećavati s kompresijom. Problem je to što se u većini slojeva neprozirnost smanjuje s kompresijom. Kramersov zakon neprozirnosti opisuje ovisnost neprozirnosti κ o gustoći i temperaturi kao $\kappa \propto \frac{\rho}{T^{3.5}}$.

Gustoća i temperatura rastu s kompresijom sloja, ali neprozirnost je osjetljivija na promjene u temperaturi pa se obično smanjuje s kompresijom. Većina zvjezdanih slojeva stoga ima efekt gušenja oscilacija, što objašnjava rijetkost zvjezdanih pulsacija.

Kasnije je otkriveno da Eddingtonov model ventila može uspješno funkcionirati u slojevima zvijezde koji su područja djelomične ionizacije. Razlog tomu je što u takvim zonama dio rada koji se vrši na plinove tijekom kompresije uzrokuje daljnju ionizaciju umjesto da povisuje temperaturu plina, što uzrokuje i veće vrijednost specifičnih toplinskih kapaciteta C_V i C_p u takvim zonama. Zbog manjeg povećanja temperature, član gustoće u Kramersovom zakonu dominira i neprozirnost raste s kompresijom. Tijekom ekspanzije temperatura se ne smanjuje dovoljno zbog rekombinacije iona s elektronima, što otpušta energiju, i neprozirnost opada s gustoćom. Takvi slojevi stoga mogu apsorbirati toplinu tijekom kompresije, biti pogurani prema van i otpustiti toplinu tijekom ekspanzije te pasti natrag u početno stanje i započeti ciklus iznova. Ovakav se mehanizam neprozirnosti naziva κ -mehanizam [12].

2.2 *Određivanje parametara iz pomrčinskih zvijezda*

Mnogi zvjezdani parametri mogu se odrediti poznavanjem zakona zračenja crnog tijela, spektara ili paralakse. Međutim, masu zvijezde direktno se može odrediti samo promatranjem gravitacijskih interakcija zvijezde s drugim objektima, to jest primjenom Keplerovih zakona. Oko polovica svih zvijezda koje se vide na nebu zapravo su dvojni sustavi, što olakšava određivanje mase, ali i drugih zvjezdanih parametara.

Kada je kutna udaljenost komponenti dvojnog sustava dovoljno velika, postaje moguće analizirati orbitalne karakteristike pojedinih zvijezda. Iz orbitalnih podataka može se odrediti orijentacija orbita i položaj centra mase, što omogućuje da se odredi omjer masa zvijezda u sustavu. Ako je i udaljenost do sustava poznata, može se odrediti linearna udaljenost između zvijezda te pomoću toga i pojedinačne mase zvijezda. Za dvije zvijezde koje orbitiraju oko zajedničkog centra mase, i za koje je pretpostavljeno da je ravnina orbite okomita na doglednicu, omjer masa može se pronaći iz omjera kutnih udaljenosti zvijezda od centra mase. Ako uzmemo u obzir

samo duljine vektora \mathbf{r}_1 i \mathbf{r}_2 , vrijedi:

$$\frac{m_1}{m_2} = \frac{r_2}{r_1} = \frac{a_2}{a_1}, \quad (2.6)$$

gdje su a_1 i a_2 velike poluosi elipsa. Za dvojni sustav udaljen za duljinu d od promatrača, kutovi pod kojima se vide velike poluosi mogu se zapisati kao $\alpha_1 = \frac{a_1}{d}$ i $\alpha_2 = \frac{a_2}{d}$, što daje omjer masa:

$$\frac{m_1}{m_2} = \frac{\alpha_2}{\alpha_1}. \quad (2.7)$$

Omjer masa moguće je odrediti i bez poznavanja udaljenosti do sustava. Općeniti zapis Keplerovog trećeg zakona, $P^2 = \frac{4\pi^2}{G(m_1+m_2)}a^3$, daje izraz za zbroj masa komponenti:

$$m_1 + m_2 = \frac{4\pi^2 (\alpha d)^3}{G P^2}. \quad (2.8)$$

U gornjim izrazima a označava veliku poluos orbite reducirane mase sustava i može se izraziti kao $a = a_1 + a_2$, a kut α kao $\alpha = \alpha_1 + \alpha_2$. Uz poznavanje udaljenosti d , jednadžbe 2.7 i 2.8 mogu se kombinirati kako bi se odredile pojedinačne mase zvijezda u sustavu.

Međutim, ravnine većina orbita nisu okomite na doglednicu. Stoga je potrebno uvesti pojam inklinacije i između ravnine orbite i doglednice. Uz pretpostavku da se dvije ravnine sijeku duž linije paralelne malenoj poluosi elipse, promatrač neće mjeriti prave kutove velikih poluosi, već njihove projekcije na doglednicu, $\tilde{\alpha}_1 = \alpha_1 \cos i$ i $\tilde{\alpha}_2 = \alpha_2 \cos i$. Ovo nema nikakvog utjecaja na jednadžbu 2.7, gdje se kosinusi samo pokrate, ali mijenja jednadžbu 2.8, koja se sada može zapisati kao:

$$m_1 + m_2 = \frac{4\pi^2}{G} \left(\frac{d}{\cos i} \right)^3 \frac{\tilde{\alpha}^3}{P^2}, \quad (2.9)$$

gdje je $\tilde{\alpha} = \tilde{\alpha}_1 + \tilde{\alpha}_2$ [12].

Mnoštvo informacija može se dobiti iz dvojnog sustava čak i ako nije moguće razlučiti pojedine zvijezde. Vrlo dobar primjer su pomrčinski, spektroskopski dvojni sustavi u kojima su vidljive spektralne linije obje zvijezde (SB2). U SB2 sustavu je, uz pojedinačne mase zvijezda, moguće odrediti i radijuse zvijezda, omjer tokova zračenja te kroz njega i omjer efektivnih temperatura zvijezda.

Neka je v_1 brzina zvijezde mase m_1 i v_2 brzina zvijezde mase m_2 u nekom tre-

nutku. Opažene radijalne brzine ne mogu premašiti $v_{1r}^{max} = v_1 \sin i$ i $v_{2r}^{max} = v_2 \sin i$, odnosno izmjerene radijalne brzine ovise o položaju zvijezde u tom trenutku. Ako pretpostavimo da je ekscentricitet orbite vrlo malen, brzine zvijezda skoro su konstantne i mogu se izraziti kao $v_1 = \frac{2\pi a_1}{P}$ i $v_2 = \frac{2\pi a_2}{P}$, gdje su a_1 i a_2 velike poluosi, a P period orbita. Uvrštavanjem a_1 i a_2 u jednadžbu 2.6, dobivamo:

$$\frac{m_1}{m_2} = \frac{v_2}{v_1}, \quad (2.10)$$

što se može zapisati i preko opaženih radijalnih brzina:

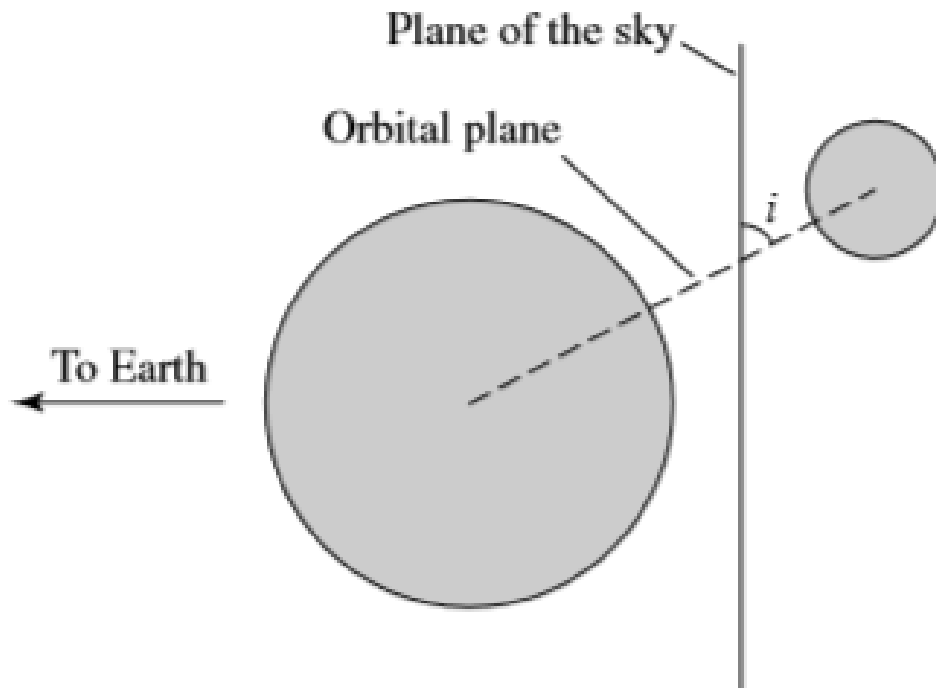
$$\frac{m_1}{m_2} = \frac{v_{2r}/\sin i}{v_{1r}/\sin i} = \frac{v_{2r}}{v_{1r}}. \quad (2.11)$$

Kako bismo odredili sumu masa potrebno je ubaciti $a = a_1 + a_2 = \frac{P}{2\pi}(v_1 + v_2)$ u Keplerov treći zakon, čime se dobiva izraz za sumu masa $m_1 + m_2 = \frac{P}{2\pi G}(v_1 + v_2)^3$. Zapisano preko opaženih radijalnih brzina, ovo se može izraziti kao:

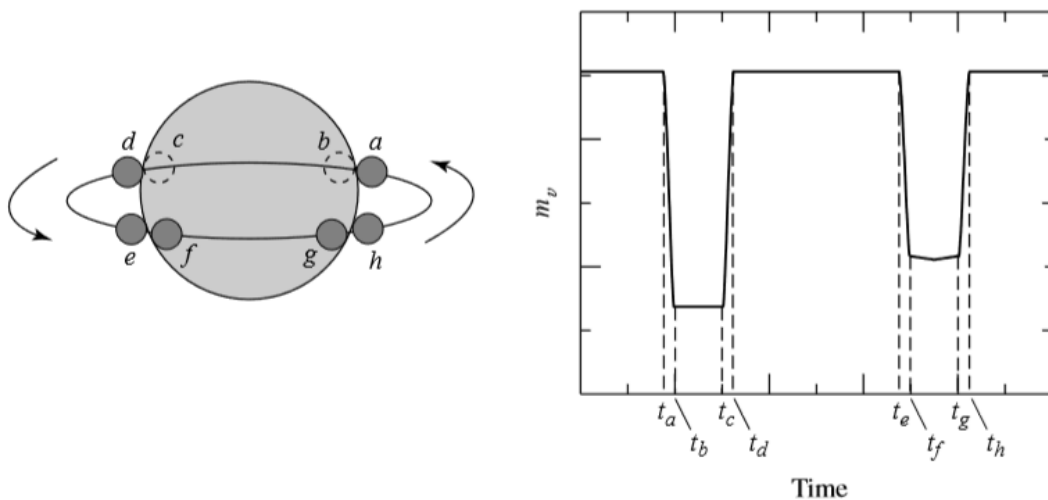
$$m_1 + m_2 = \frac{P}{2\pi G} \frac{(v_{1r} + v_{2r})^3}{\sin^3 i}. \quad (2.12)$$

Čak i uz poznavanje brzina v_{1r} i v_{2r} , nije moguće odrediti mase bez poznavanja inklinacije orbite i . Kako zvijezde mogu biti grupirane prema njihovim efektivnim temperaturama i luminozitetima te uz pretpostavku da postoji relacija između tih vrijednosti i mase, može se dobiti statistička procjena mase za svaku vrstu zvijezda ako se odabere prikladno usrednjena vrijednost za $\sin^3 i$. Za spektroskopske sustave uzima se $\langle \sin^3 i \rangle \approx \frac{2}{3}$ [12].

Kada se u spektroskopskom sustavu opažaju i pomrčine, moguće je dobiti vrlo dobru procjenu kuta i . Ako razmak između komponenti dvojnog sustava nije veći od sume njihovih radijusa, inklinacija sustava u pomrčinskom sustavu mora biti blizu 90° , kao što se vidi sa slike 2.2. Može se pretpostaviti da vrijedi $i = 90^\circ$ jer će odstupanje u toj vrijednosti pridonijeti malenu grešku u izračunu $\sin^3 i$ i određivanju sume masa. Procjenu i moguće je dalje poboljšati promatranjem krivulja sjaja sustava. Primjer je dan na slici 2.3. Sa slike je vidljivo da ako veća zvijezda prekrije manju, dolazi do skoro pa konstantnog minimuma u izmjerenoj svjetlosti sustava za to vrijeme. Ako manja zvijezda prolazi ispred veće, iako ona nije potpuno prekrivena doći će do skoro konstantnog pada u količini opaženog svjetla.



Slika 2.2: Geometrija spektroskopskog dvojnog sustava. Vidljivo je da inklinacija sustava mora biti blizu 90° . Slika je preuzeta iz [12].



Slika 2.3: Krivulja sjaja pomrčinskog dvojnog sustava za koji je $i = 90^\circ$. Vremena označena na krivulji odgovaraju položajima manje zvijezde u odnosu na veću. U prikazanom je slučaju pretpostavljeno da je manja zvijezda toplija. Slika je preuzeta iz [12].

Mjerenjem trajanja pomrčina, moguće je odrediti radijuse pojedinačnih članova u SB2 sustavu. Uz pretpostavku da je $i \approx 90^\circ$, vrijeme između prvog kontakta (t_a na slici 2.3) i minimalne svjetlosti (t_b) uz poznavanje brzina zvijezda v_s (manja zvijezda) i v_l (veća zvijezda) omogućava direktan izračun radijusa manje komponente. Ako je

velika poluos orbite manje zvijezde dovoljno velika u usporedbi s radijusom bilo koje zvijezde i ako je orbita skoro kružna, može se pretpostaviti da se manja zvijezda giba otprilike okomito na doglednicu za vrijeme pomrčine. U tom slučaju radijus manje komponente može se jednostavno izraziti kao:

$$r_s = \frac{v}{2}(t_b - t_a), \quad (2.13)$$

gdje je $v = v_s + v_l$ relativna brzina obaju zvijezda. Slično tomu, ako uzmemo vrijeme između t_b i t_c , može se odrediti radijus veće zvijezde:

$$r_l = \frac{v}{2}(t_c - t_a) = r_s + \frac{v}{2}(t_c - t_b). \quad (2.14)$$

Iz krivulje sjaja moguće je odrediti i omjer efektivnih temperatura zvijezda u sustavu. Sa slike 2.3 vidi se da je minimum dublji kada manja, toplija zvijezda prolazi iza veće zvijezde. Kako bi se ovo pojasnilo, možemo uzeti da je radijacijski površinski tok jednak $F_r = F_{pov} = \sigma T_l^4$, gdje je σ Stefan-Boltzmannova konstanta. Bez obzira na to prolazi li manja zvijezda ispred ili iza veće zvijezde, ista površina je zasjenjena. Uz pretpostavku da je opaženi tok konstantan kroz disk, količina sjaja (eng. *brightness*) koja se detektira kada su obje zvijezde vidljive iznosi $B_0 = k(\pi r_l^2 F_{rl} + \pi r_s^2 F_{rs})$, gdje je k konstanta koja ovisi o udaljenosti od sustava, količini ometajućeg materijala između sustava i detektora i prirodi detektora. U ranije opisanom slučaju, količina sjaja koja je detektirana pri dubljem, ili primarnom, minimumu iznosi $B_p = k\pi r_l^2 F_{rl}$, a ona detektirana pri sekundarnom minimumu $B_s = k\pi(r_l^2 - r_s^2)F_{rl} + k\pi r_s^2 F_{rs}$. Kako se k ne može precizno odrediti, koriste se omjeri. Iz izraza za B_0 , B_p i B_s , dobije se [12]:

$$\frac{B_0 - B_p}{B_0 - B_s} = \frac{F_{rs}}{F_{rl}} = \left(\frac{T_s}{T_l}\right)^4. \quad (2.15)$$

2.3 Modeli supernova tipa Ia

Važnost supernova tipa Ia u astrofizici pojašnjena je u prethodnom poglavlju. S obzirom na njihovu standardiziranost, očekuje se da je i proces kojim nastaju prilično standardiziran. Standardna pretpostavka jest da dolazi do uništenja bijelog patuljka koji se nalazi u dvojnog sustavu. Ako bijeli patuljak primi dovoljnu količinu mase od druge zvijezde u sustavu, njegova vlastita masa će prijeći Chandrasekharovu granicu

(najveća moguća masa stabilnog bijelog patuljka, $1.44 M_{\odot}$), što će uzrokovati eksploziju zvijezde. Trenutno postoje dva glavna modela koji dovode do ovog scenarija: dvostruko degenerirani i jednostruko degenerirani.

U dvostruko degeneriranom modelu radi se o dvojnem sustavu od dva bijela patuljka, jednog veće mase i manjeg radijusa i drugog manje mase i većeg radijusa. Ako u nekom sustavu dođe do promjene u raspodjeli mase, promjene u zakrivljenosti okolnog prostor-vremena mogu se propagirati kao gravitacijski val. Emisija gravitacijskih valova uzrokovat će da počnu orbitirati jedan oko drugog u spiralnoj orbiti. Nakon par orbita, patuljak manje mase će biti potpuno rastrgan te će predati svoju materiju, bogatu s ugljikom i kisikom, masivnijem patuljku. Kako masa tog patuljka raste i približava se Chandrasekharovoj granici, u unutrašnjosti započinju nuklearne reakcije koje u konačnici unište patuljka.

Kod jednostruko degeneriranog modela, zvijezda glavnog niza kruži oko bijelog patuljka te materijal s nje prelazi na patuljka. Helij u plinu sa zvijezde postane degeneriran zbog velikog gravitacijskog pritiska i kada ga se dovoljno nakupi, dolazi do helijevog bljeska, što uzrokuje da helij sagorijeva u ugljik (C) i kisik (O), ali i šalje udarni val u bijelog patuljka koji zapali degenerirani C i O u patuljku. Druga verzija ovog modela ne uključuje helij, već se degenerirani C i O u bijelom patuljku zapale kako se on približava Chandrasekharovoj granici, nakon čega degenerirani plin više ne može održavati masu zvijezde. Model dalje predviđa dva moguća ishoda: sagorijevanje C i O na podzvučnim brzinama ili detonaciju na nadzvučnim brzinama.

Moguće je da se oba modela odvijaju u prirodi, kao i da se odvijaju obje verzije jednostruko degeneriranog modela. U svakom slučaju, standardiziranost krivulja sjaja supernova tipa Ia proizlazi iz erupcije C-O bijelog patuljka blizu Chandrasekharove mase. Moguće varijacije mogu proizaći iz manjih varijacija u masi ili u mehanizmima [12].

3 Pregledi neba i podatci

U ovom radu korišteni su podatci o promjenjivim zvijezdama iz drugog objavljivanja podataka Catalina pregleda neba. Glavna misija Catalina programa jest otkrivanje i praćenje objekata bliskih Zemlji [14]. Podatci iz Catalina pregleda neba također sadrže informacije o oko 110000 promjenjivih zvijezda, kao što su njihov položaj u ekvatorijalnim koordinatama, magnituda, period, amplituda krivulje sjaja i vrsta promjenjive zvijezde. Kada se algoritmu za strojno učenje predaju podatci za učenje, poželjno je da ti podatci imaju što više atributa kako bi algoritam preciznije sortirao promjenjive zvijezde u skupine. U tu svrhu podacima iz Catalina pregleda neba dodani su i podatci iz drugih pregleda neba za odgovarajuće zvijezde. Ti podatci sadrže informacije o sjaju zvijezda kroz različite filtre. Uz Catalina pregled neba, korišteni pregledi neba su Gaia, Pan-STARRS, Skymapper, WISE i NEOWISE i 2MASS. Podatci iz različitih pregleda neba uspoređeni su preko ekvatorijalnih koordinata zvijezda korištenjem SQL (eng. *Structured Query Language*) metode. Detalji o korištenim pregledima neba dani su u tablici 3.1.

| Pregled | Instrumenti | Filteri | Opažano područje | Vremenski period | Broj opaženih objekata | Granične magnitude |
|----------------|------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|-----------------------------------------------|------------------------|--------------------------------|
| Catalina | CSS (Catalina Survey Schmidt) | bez filtera | 20155 kvadratnih stupnjeva u rasponu $0^\circ < \alpha < 360^\circ$, $-22^\circ < \delta < 65^\circ$ | travanj 2005. - svibanj 2011. | ~ 198 milijuna | 12.5 - 19.5 u V |
| Gaia | Gaia satelit | Gaia G, G _{BP} i G _{BP} | cijelo nebo | 25.7.2014. - 23.5.2016. | ~ 1.7 milijardi | 3 - 21 u G |
| Pan-STARRS | teleskop od 1.8 metara s kamerom od 1.4 gigapiksela | <i>g, r, i, z, y</i> | ~ 30000 kvadratnih stupnjeva sjeverno od -30 Dec | lipanj 2010. - ožujak 2014. | ~ 3 milijarde | < 25 |
| Skymapper | teleskop sa zrcalom od 1.35 metara i s mozaikom od 32 CCD kamere | <i>u, v, g, r, i, z</i> | 20200 kvadratnih stupnjeva južnog neba | 15.3.2014. - rujan 2015. | ~ 2.3 milijarde | < 18 |
| WISE i NEOWISE | WISE svemirski teleskop od 40 centimetara | WISE infracrveni filteri W1 (3.4 μm), W2 (4.6 μm), W3 (12 μm), W4 (22 μm) | cijelo nebo | 7.1.2010. - 6.8.2010., 29.9.2010. - 1.2.2011. | ~ 747 milijuna | < 17 u W1 |
| 2MASS | dva teleskopa od 1.3 metara s kamerama s tri kanala | J (1.234 μm), H (1.662 μm), K _s (2.159 μm) | cijelo nebo | svibanj 1997. - veljača 2001. | ~ 471 milijun | > 20, -4 - 16 u K _s |

Tablica 3.1: Detalji o pregledima neba korištenima za podatke u ovom radu. Informacije su preuzete iz [15–23].

3.1 Određivanje perioda u pregledima neba

Uz boje, najvažniji parametar za određivanje vrste promjenjive zvijezde jest period. Stoga je vrlo važno ispravno odrediti period opažene promjenjive zvijezde. Cata-

lina pregled neba u tu svrhu koristi metodu Lomb-Scargle periodograma. Periodogram, pojam kojeg je osmislio Schuster 1898. [24], daje mjeru periodičnosti signala kao funkciju kutne frekvencije ω za jednoliko raspoređene podatke. Za N podataka $\{y_k\}_{k=1}^N$ izmjerene u jednakim intervalima $t_k = t_0 + k\Delta t$ Schusterov periodogram, koji mjeri spektralnu snagu, to jest snagu svjetlosti po jedinici površine i po jedinici valne duljine, kao funkciju kutne frekvencije, dan je kao:

$$C(\omega) = \frac{1}{N} \left| \sum_{k=1}^N y_k e^{i\omega t_k} \right|^2. \quad (3.1)$$

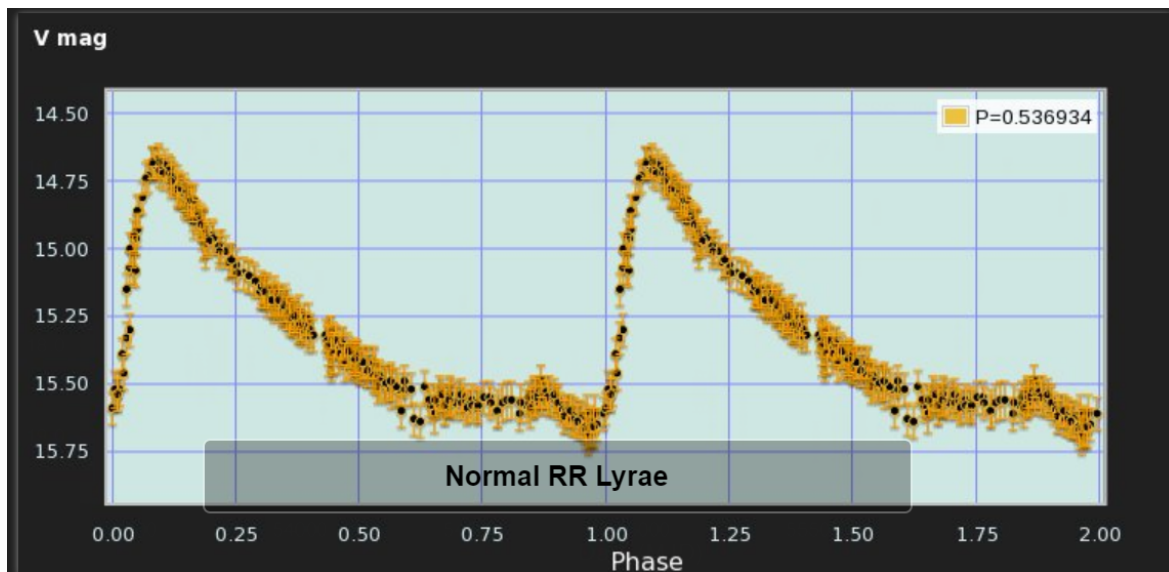
Međutim, astronomska opažanja rijetko su uniformno raspoređena pa je bilo potrebno proširiti ideju periodograma na nejednoliko raspoređena mjerenja. Lomb (1976.) [25] i Scargle (1982.) [26] proširili su ranije ideje i definirali normalizirani periodogram:

$$P_N(\omega) = \frac{1}{2V_y} \left[\frac{[\sum_k (y_k - \bar{y}) \cos \omega(t_k - \tau)]^2}{\sum_k \cos^2 \omega(t_k - \tau)} + \frac{[\sum_k (y_k - \bar{y}) \sin \omega(t_k - \tau)]^2}{\sum_k \sin^2 \omega(t_k - \tau)} \right], \quad (3.2)$$

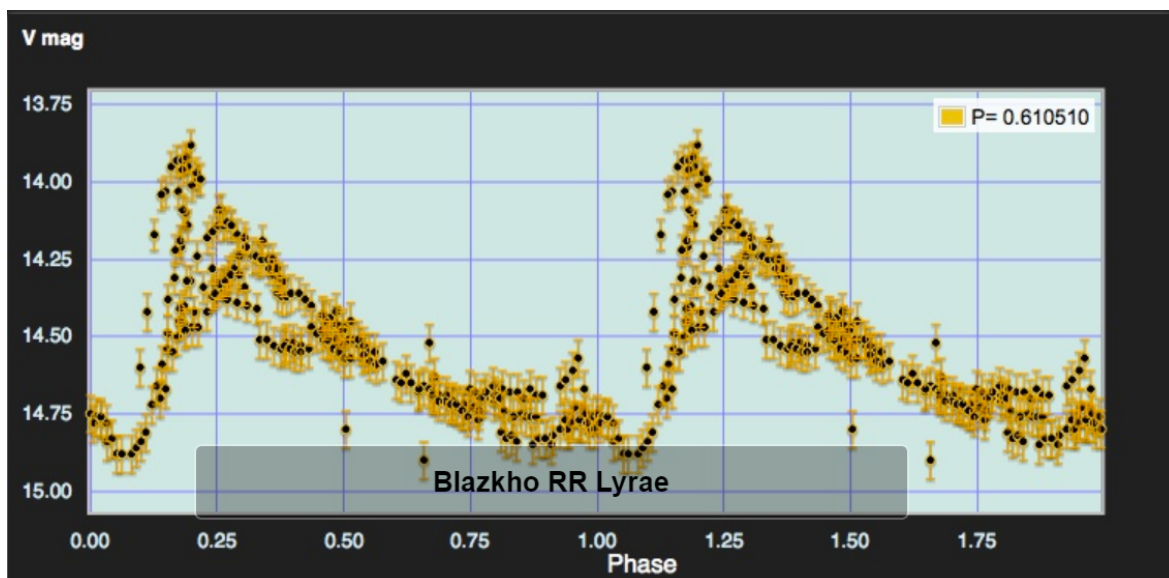
gdje je \bar{y} srednja vrijednost, a V_y odstupanje podataka $\{y_k\}$, τ je vremenski pomak koji čini $P_N(\omega)$ nezavisnim na translacije u vremenu t . Lomb je pokazao da taj pomak ima i dublje značenje: čini P_N identičnim s procjenom harmoničkog sadržaja ako se primijeni metoda najmanjih kvadrata na jednokomponentni sinusoidalni model : $d(t) = A \sin(\omega t + \Phi)$ [27].

3.2 Krivulje sjaja u Catalina pregledima neba

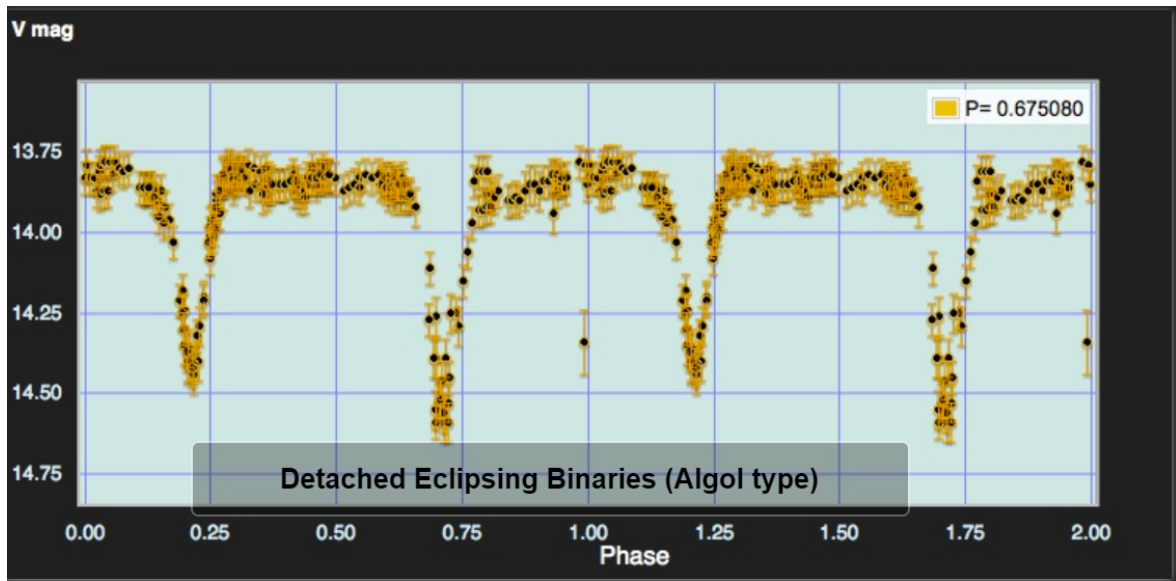
Nakon što je određen period promjenjive zvijezde, moguće je nacrtati njenu faznu krivulju. Krivulje sjaja karakteristične su za svaku vrstu promjenjivih zvijezda pa se preko njih promjenjive zvijezde najlakše mogu klasificirati. Na slikama 3.2, 3.3, 3.4, 3.5 i 3.6 dane su krivulje sjaja za šest kategorija promjenjivih zvijezda. Te su krivulje primjer krivulja sjaja iz Catalina pregleda neba dan na službenoj stranici pregleda [28].



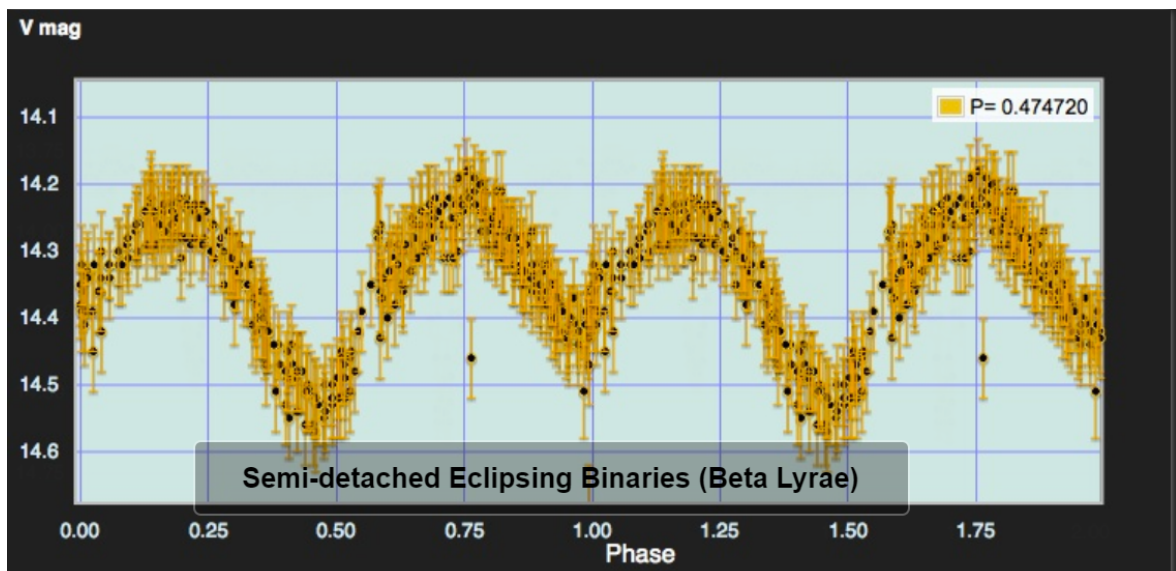
Slika 3.2: Krivulja sjaja za zvijezdu kategorije RR Lyrae. Krivulje sjaja za ovu kategoriju su asimetrične, s naglim skokovima u amplitudi. Slika je preuzeta s [28].



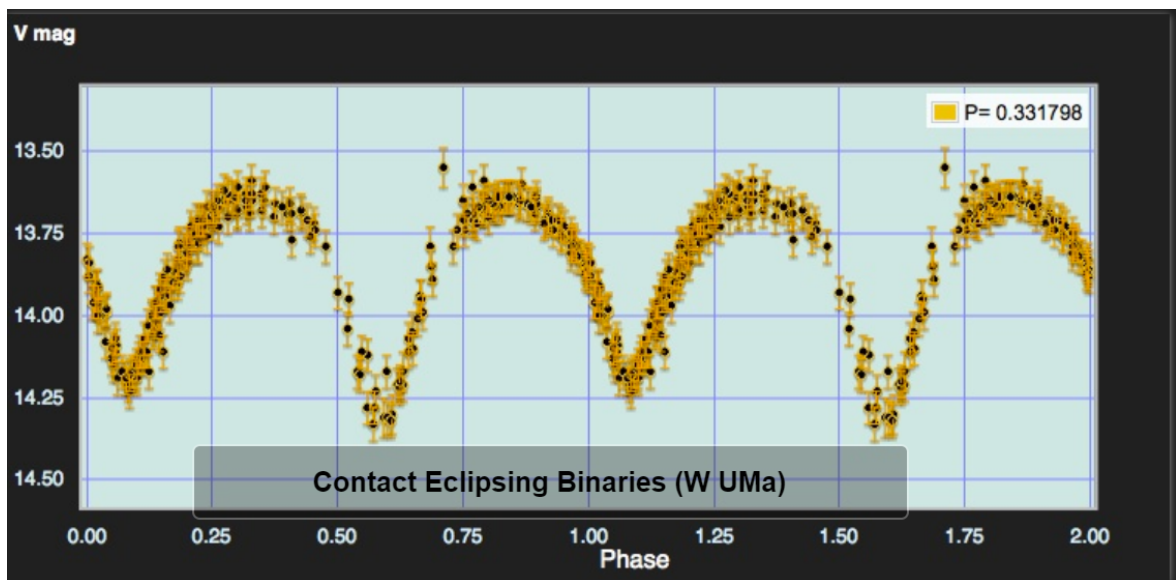
Slika 3.3: Krivulja sjaja za zvijezdu *Blazkho* podvrste kategorije RR Lyrae. Krivulje sjaja za ovu kategoriju slične su kao i za obične RR Lyrae zvijezde, uz periodične varijacije u obliku. Slika je preuzeta s [28].



Slika 3.4: Krivulja sjaja za pomrčinsku binarnu zvijezdu Algol tipa. Krivulje sjaja za ovu kategoriju pokazuju konstantu svjetlost uz dva različita minimuma koji se očituju kao nagli padovi u krivulji. Slika je preuzeta s [28].



Slika 3.5: Krivulja sjaja za pomrčinsku binarnu zvijezdu Beta Lyrae tipa. Krivulje sjaja za ovu kategoriju pokazuju konstante varijacije u svjetlosti i minimume različitih dubina. Slika je preuzeta s [28].



Slika 3.6: Krivulja sjaja za pomrčinsku binarnu zvijezdu W UMa tipa. Krivulje sjaja za ovu kategoriju pokazuju konstante varijacije u svjetlosti i minimume sličnih dubina. Slika je preuzeta s [28].

4 Metodologija

Za precizno određivanje različitih parametara iz promjenjivih zvijezda važno je da su te zvijezde ispravno klasificirane. Kako bi ručno klasificiranje promjenjivih zvijezda preko krivulja sjaja trajalo predugo za velike količine podataka koje je potrebno obraditi u pregledima neba, potrebno je pronaći metodu klasifikacije koja je brza, precizna i automatizirana. U tu svrhu mogu se koristiti metode strojnog učenja, što je područje koje trenutno prolazi kroz period ekspanzivnog razvoja. Strojno učenje sadrži algoritme koji mogu klasificirati i grupirati velike količine podataka relativno brzo, a uz to je i automatizirano.

Strojno učenje, grana umjetne inteligencije, proces je oblikovanja, ili "treniranja", algoritama (modela) kako bi se oni mogli koristiti za predviđanje svojstava nekog skupa temeljem empirijskih opažanja. Problemima strojnog učenja moguće je pristupiti na dva načina: nadziranim učenjem i nenadziranim učenjem.

Kod nadziranog učenja podatci se unose s dodatnim svojstvima koja se žele predvidjeti. Takav problem može biti ili klasifikacija ili regresija. Kod klasifikacije uzorci su svrstani u više klasa te se preko već svrstanih podataka nastoji predvidjeti kojoj klasi pripadaju nesvrstani podatci. Ako umjesto klase želimo predvidjeti jednu ili više kontinuiranih varijabli dijela podataka, tada se radi o regresiji.

Kod nenadziranog učenja podatci za treniranje algoritma unose se u model bez ikakvih postojećih oznaka. Cilj takvih problema može biti otkriti grupe sličnih primjera u podacima, što se naziva grupiranje, odrediti raspodjelu ulaznih podataka, što se naziva procjena gustoće, ili projicirati podatke iz višedimenzionalnog prostora na samo dvije ili tri dimenzije u svrhu vizualizacije [29]. U ovom radu koristile su se metode klasifikacije i grupiranja u svrhu svrstavanja promjenjivih zvijezda u grupe na osnovu njihovih parametara.

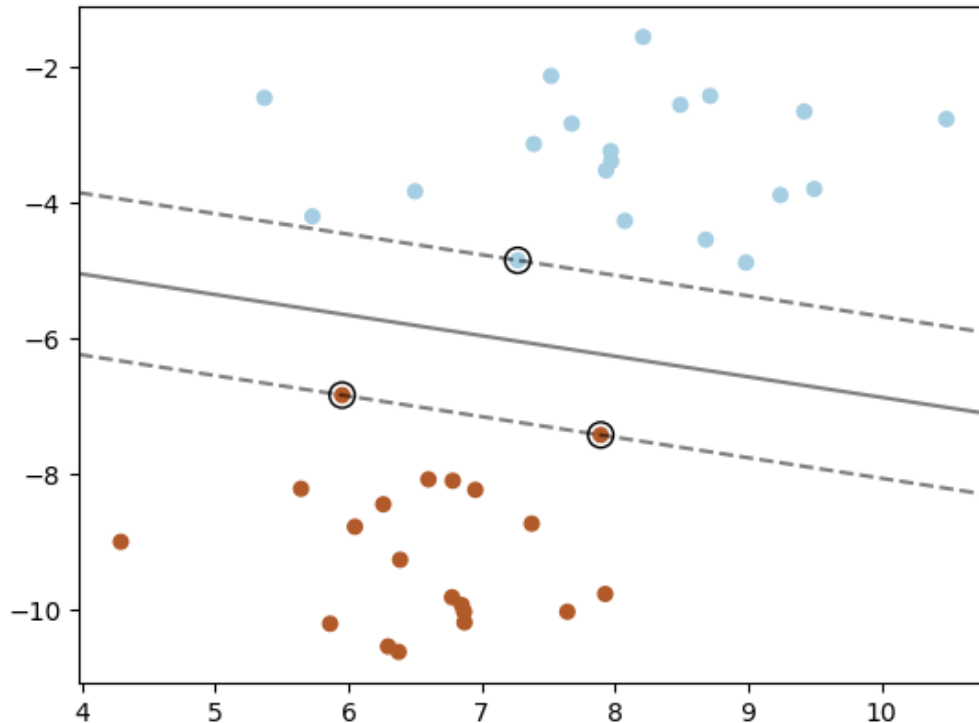
4.1 *Scikit-Learn*

Metode strojnog učenja implementirane su pomoću *Scikit-Learn* paketa u programskom jeziku *Python*. *Scikit-Learn* uključuje široki spektar najnovijih algoritama za strojno učenje za nadzirane i nenadzirane probleme. Naglasak je na jednostavnosti korištenja, učinkovitosti, dokumentaciji i konzistenciji korisničkog sučelja. Projekt je započeo u 2007., a prvi je put javno objavljen 1.2.2010. [30, 31]. U ovom radu

korištene su metode *Support Vector Machine* (SVM) za nadzirano učenje putem klasifikacije i *Gaussian Mixture Model* (GMM) za nenadzirano učenje putem grupiranja.

4.1.1 Support Vector Machine

Support Vector Machine (SVM) algoritmi skup su metoda nadziranog učenja koje se mogu koristiti za klasifikaciju i regresiju. Učinkoviti su u visokodimenzionalnim prostorima parametara, čak i kada je broj dimenzija veći od broja uzoraka. SVM konstruira hiperprostor ili skup hiperprostora u visokodimenzionalnom prostoru koji se tada mogu koristiti za klasifikaciju, regresiju i druge probleme. Dobra se separacija postiže kada je hiperprostor što više udaljen od najbližih točaka podataka za treniranje algoritma bilo koje klase, koji se nazivaju funkcijske margine. Što je margina veća, to je greška pri klasificiranju manja. Na slici 4.1 prikazana je funkcija za linearno separabilni problem sa tri uzorka na rubovima margina, takozvanim pomoćnim vektorima (eng. *support vectors*) [32]. Za problem koji nije linearno separabilan,



Slika 4.1: SVM funkcija za linearno separabilni problem sa tri uzorka na rubovima margina. Slika je preuzeta s [32].

pomoćni vektori su u pravilu uzorci unutar granica margina.

4.1.2 Gaussian Mixture Model

Gaussian Mixture Model (GMM) je probabilistički model koji pretpostavlja da su svi podatci generirani iz konačnog broja Gaussovih raspodjela s nepoznatim parametrima. Glavni problem kod treniranja GMM-ova iz neoznačenih podataka jest taj što nije poznato iz koje latentne komponente dolaze koje točke. Za rješavanje tog problema koristi se statistički algoritam maksimizacije očekivanja (EM, eng. *expectation-maximization*). Problem se rješava iterativnim procesom. Prvo se pretpostave nasumične komponente te se za svaku točku računa vjerojatnost da je generira svaka komponenta modela. Zatim se podešavaju parametri dok se ne postigne maksimum vjerojatnosti podataka u tim uvjetima.

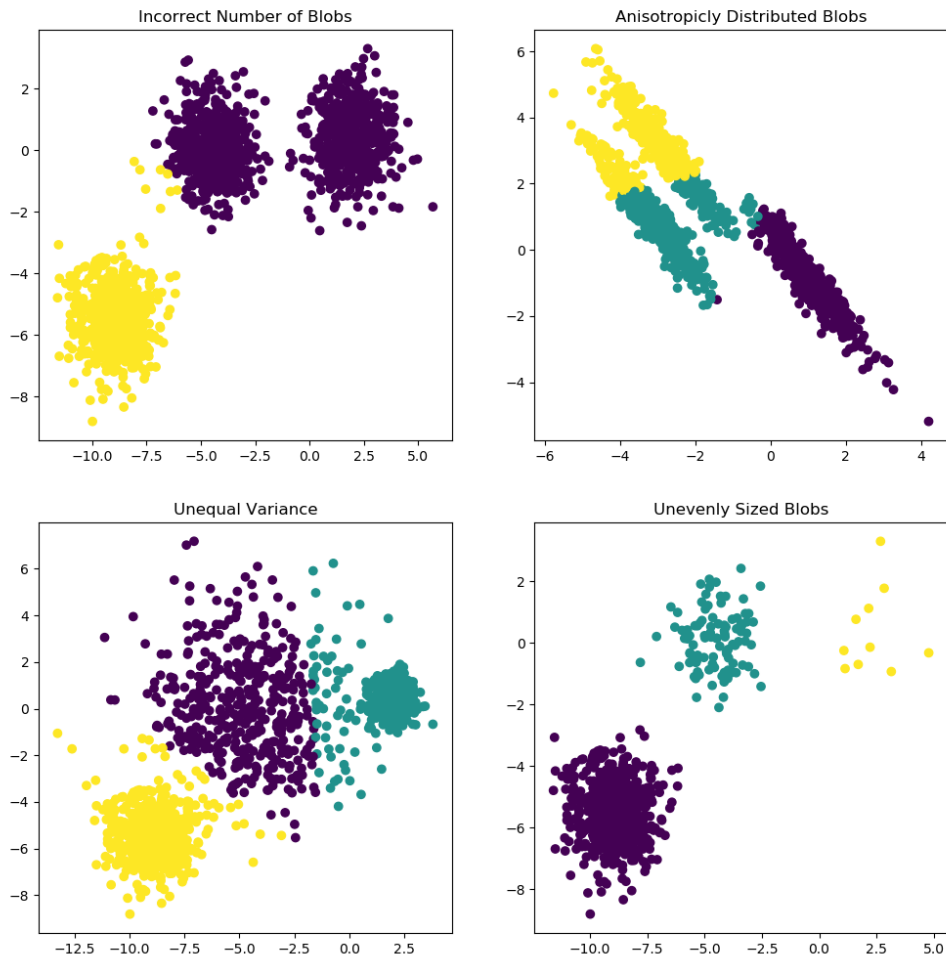
GMM se može zamisliti kao model generaliziranog *K-means* grupiranja koji uključuje informacije o kovarijantnoj strukturi podatka i središtima lokalnih Gaussovih raspodjela. *K-means* algoritam grupira podatke tako dijeli skup od N uzoraka X u K grupa C karakteriziranih srednjom vrijednošću μ_j uzoraka u grupi. Te srednje vrijednosti se nazivaju "centroidi" grupa te je cilj algoritma odabrati centroide koji minimiziraju vrijednost koja se naziva inercija, odnosno suma kvadrata unutar grupe:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2). \quad (4.1)$$

Inercija se može opisati kao mjera unutrašnje koherencije grupa. Problemi kod inercije koji mogu dovesti do problema pri grupiranju su to što nije normalizirana te to što pretpostavlja konveksne i izotropne grupe pa loše reagira na izdužene grupe ili one nepravilnih oblika [33, 34]. Neki slučajevi u kojima *K-means* metoda grupiranja može proizvesti neželjene grupe prikazani su na slici 4.2.

4.2 Primjena metoda strojnog učenja na podatke

Metode SVM-a i GMM-a primijenjene su na podatke iz pregleda neba. Podatci koji su korišteni za treniranje algoritama su periodi i amplitude promjenjivih zvijezda iz Catalina pregleda neba, G-RP i BP-G boje dobivene iz filtera iz Gaia pregleda neba, W1-W2 boja dobivena iz filtera iz WISE i NEOWISE pregleda neba te J-K, J-H i H-K boje dobivene iz filtera iz 2MASS pregleda neba. Periodi i amplitude atributi su krivulja sjaja koje se inače koriste za određivanje vrsta promjenjivih zvijezda, a boje u



Slika 4.2: Primjeri slučajeva u kojima *K-means* metoda grupiranja može proizvesti neželjene grupe. U prva tri slučaja ulazni podatci ne odgovaraju nekima od unutar-njih pretpostavki *K-means* algoritma pa nastaju neželjene grupe. U četvrtom slučaju dobivene su očekivane grupe unatoč nakupinama podataka nejednakih veličina. Slika je preuzeta s [35].

različitim filterima pomažu specificirati o kojoj se podvrsti zvijezde radi. Također, veći broj parametara dat će veću preciznost kod klasifikacije i grupiranja u algoritmima strojnog učenja. Podatci iz Pan-STARRS i Skymapper pregleda nisu korišteni, kao ni W3 i W4 filteri iz WISE i NEOWISE pregleda. Pan-STARRS i Skymapper pregledi nisu korišteni iz razloga što sadrže samo podatke za sjeverni, odnosno južni dio neba te bi kod korištenja njihovih filtera trebalo podijeliti i ostale podatke na sjeverne i južne dijelove neba, što bi rezultiralo smanjenim brojem uzoraka za treniranje algoritma. W3 i W4 filteri nisu korišteni iz razloga što su u infracrvenom području pa previše ne

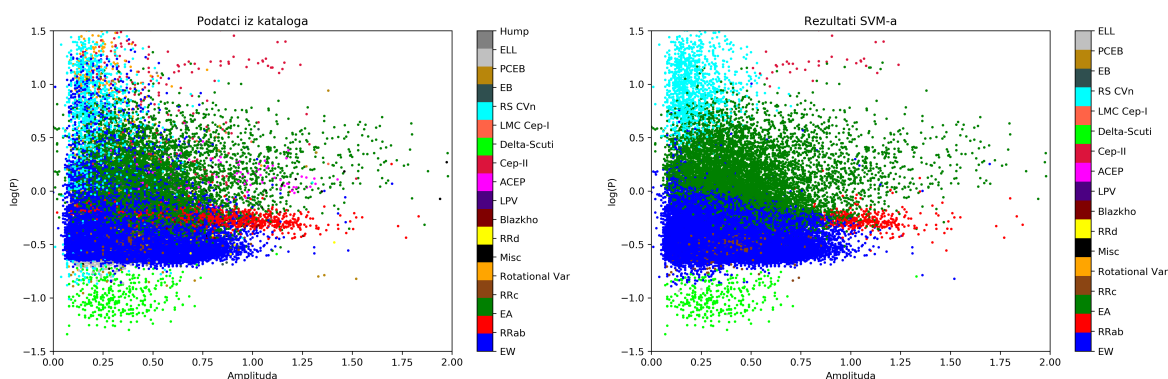
pridonose klasifikaciji. Klasifikacija promjenjivih zvijezda provedena je samo za one zvijezde kod kojih mjerenja postoje u svim katalozima jer bi suprotno uzrokovalo greške u algoritmima. Za nazive klasa promjenjivih zvijezda kod SVM algoritma korišteni su podatci o vrstama promjenjivih zvijezda iz Catalina pregleda neba.

Za treniranje SVM algoritma korišteno je 90% podataka iz kataloga, a zatim je algoritam klasificirao preostalih 10%. Kako bi se mogla obaviti SVM klasifikacija na svim zvijezdama iz kataloga, katalog je podijeljen nasumično deset puta na podatke za treniranje i za testiranje algoritma, s tim da je uzeto u obzir da je 10% podatka na kojima se provodi klasifikacija različito pri svakoj podjeli. Nakon što je obavljena klasifikacija na svih deset uzoraka, oni su opet spojeni u jednu tablicu klasificiranih podataka.

5 Rezultati i diskusija

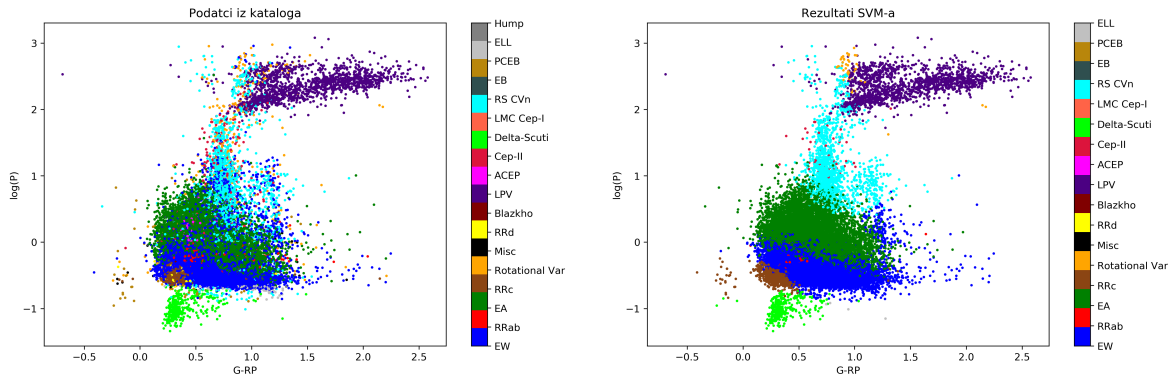
5.1 Rezultati SVM algoritma

Na slikama 5.1, 5.2, 5.3, 5.4 i 5.5 dane su usporedbe klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu. Boje dodijeljene kategorijama promjenjivih zvijezda odabrane su kako bi se različite kategorije dobro razlikovale na slici te ne nose nikakvo posebno značenje. Redoslijed kategorija na skali boja jest redoslijed kojim se pojavljuju u katalogu. Slike 5.6, 5.7 i 5.8 prikazuju usporedbu za samo pet kategorija promjenjivih zvijezda za grafove logaritma perioda i G-RP, W1-W2 i J-K boja. Odabrane kategorije su tri podvrste pomrčinskih dvojnih zvijezda te dvije podvrste RR Lyrae zvijezda. Te su kategorije odabrane zbog sličnosti u njihovim parametrima kako bi se stekao uvid u to kako SVM algoritam rješava takve slučajeve. Kategorije promjenjivih zvijezda na slikama poredane su od one s najviše članova na dnu do one s najmanje članova na površini slike. Na slici 5.9 dana je matrica zabune (eng. *confusion matrix*) za podatke iz kataloga i rezultate SVM-a. Matrica zabune pokazuje koliko se zvijezda svrstanih u pojedinu kategoriju preko SVM algoritma poklapa s istom kategorijom u katalogu. U idealnom slučaju sve vrijednosti bi bile na dijagonali matrice jer vrijednosti van dijagonale znače da je došlo do pogreške u klasifikaciji. Na slici su tamnijim bojama označena područja u kojima je veći broj poklapanja između rezultata SVM-a i kataloga. U zadnjem stupcu matrice dan je ukupan broj zvijezda u svakoj kategoriji u katalogu.

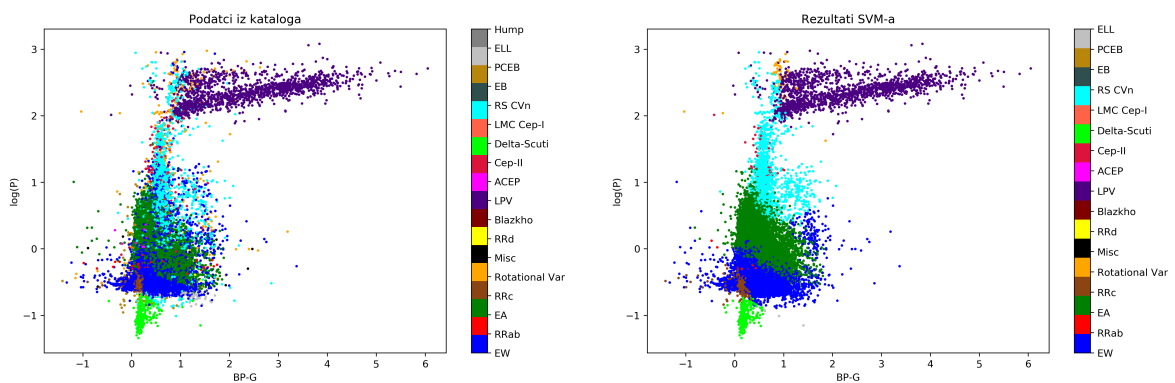


Slika 5.1: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu preko amplituda - logaritama perioda grafa.

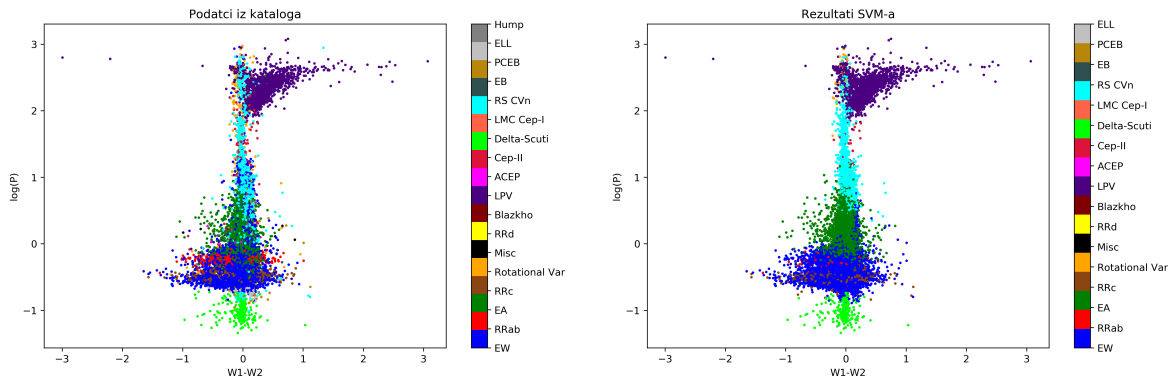
Uočava se dobro podudaranje između vrsta promjenjivih zvijezda iz podataka iz Catalina pregleda neba i vrsta dodijeljenih pomoću SVM algoritma. Primjećuje se da



Slika 5.2: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu preko G-RP boja - logaritam perioda grafa.

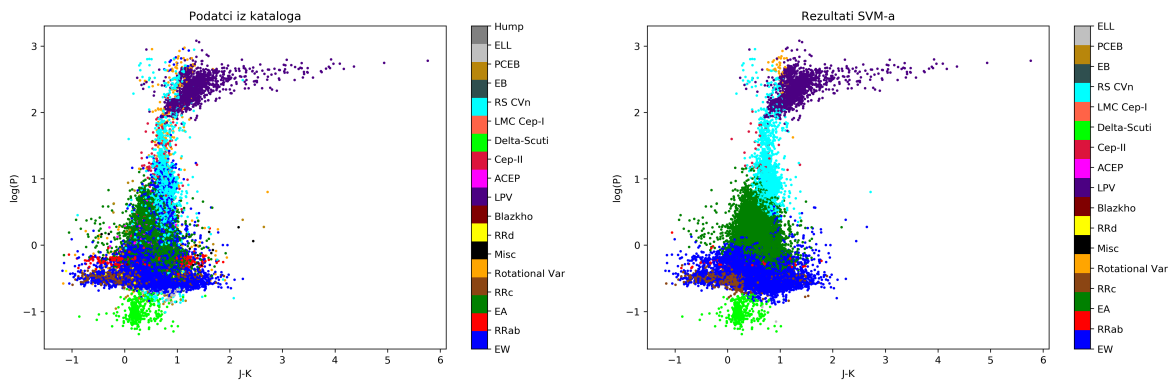


Slika 5.3: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu preko BP-G boja - logaritam perioda grafa.

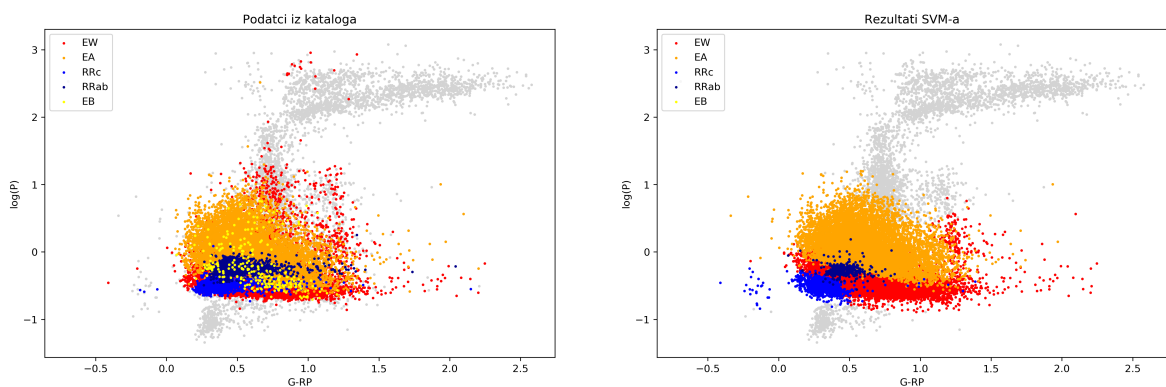


Slika 5.4: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu preko W1-W2 boja - logaritam perioda grafa.

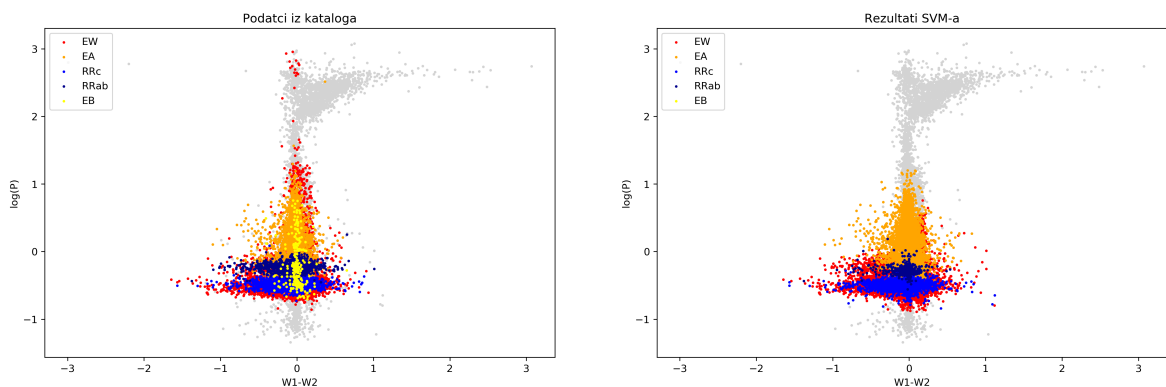
neke vrste promjenjivih zvijezda (npr. cefeide iz Velikog Magellanovog oblaka, kojih ima samo 10 u katalogu) ne postoje kod SVM klasifikacije. Razlog tomu može biti to što je promjenjivih zvijezda tih vrsta znatno manje u pregledima neba nego ostalih zvijezda, kao što su pomrčinske zvijezde ili RR Lyrae, pa su te zvijezde svrstane u neku od drugih kategorija za koju je bilo više uzoraka kod treniranja algoritma.



Slika 5.5: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu preko J-K boja - logaritama perioda grafa.

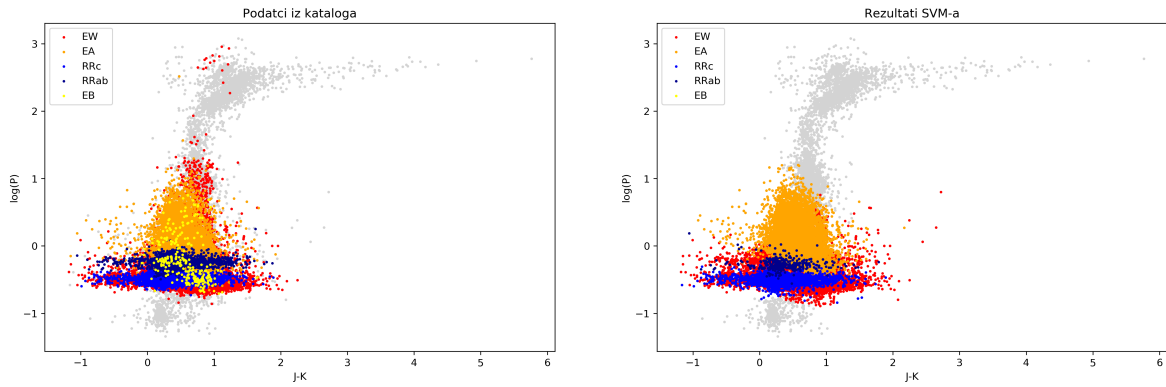


Slika 5.6: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu za pet vrsta promjenjivih zvijezda preko G-RP boja - logaritama perioda grafa.

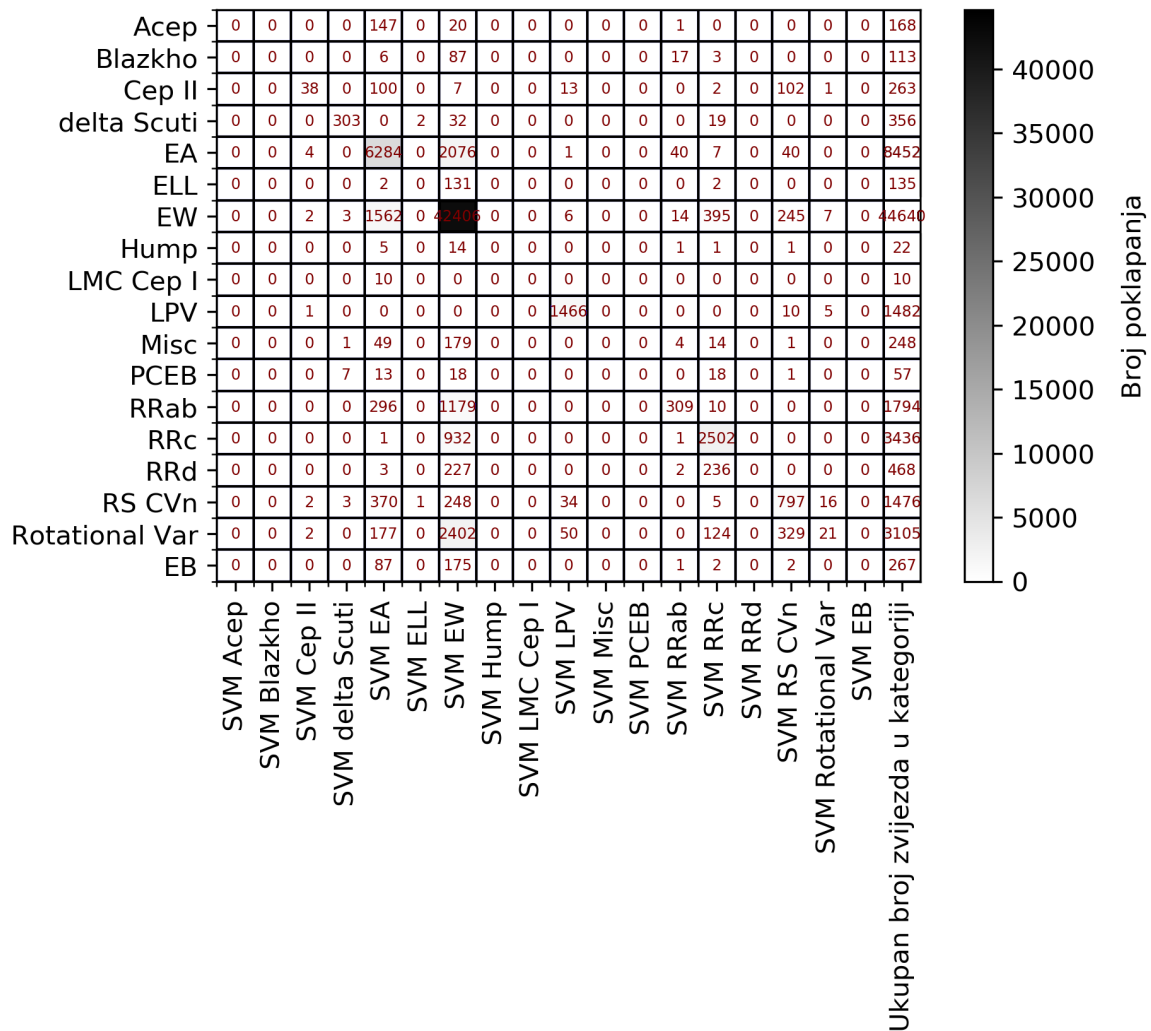


Slika 5.7: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu za pet vrsta promjenjivih zvijezda preko W1-W2 boja - logaritama perioda grafa.

Također se uočava da SVM ne razlučuje dobro različite vrste na područjima grafova gdje dolazi do preklapanja. To ima smisla jer SVM algoritam provodi klasifikaciju prema parametrima koji će na tim područjima biti slični za različite vrste promjenjivih zvijezda.



Slika 5.8: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema SVM algoritmu za pet vrsta promjenjivih zvijezda preko J-K boja - logaritama perioda grafa.



Slika 5.9: Matrica zabune za klasifikaciju promjenjivih zvijezda iz kataloga i klasifikaciju prema SVM algoritmu.

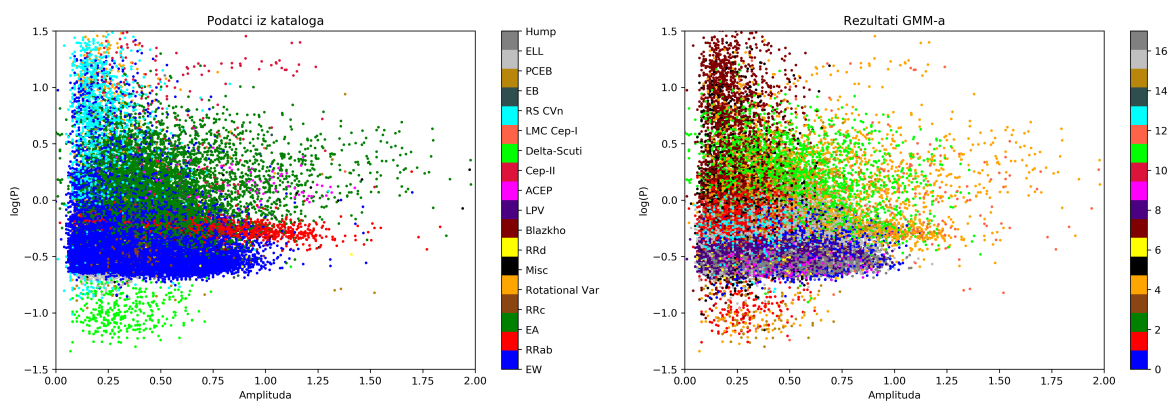
Sa slika 5.6, 5.7 i 5.8 mogu se bolje promotriti slučajevi u kojima dolazi do pogrešne klasifikacije preko SVM algoritma. Primjećuje se da u većini slučajeva u kojima

se parametri više vrsta promjenjivih zvijezda poklapaju algoritam naginje prema onoj vrsti za koju ima veći uzorak. Stoga je većina zvijezda koje pripadaju EB kategoriji, kojih u katalogu ima samo oko 200, svrstana u znatno brojnije EA i EW kategorije. Te tri kategorije imaju i slične krivulje sjaja pa je zamjena razumna.

Matrica zabune pokazuje još jednom da su kategorije zvijezda s najviše članova u katalogu najbolje klasificirane preko SVM algoritma, dok su one s manje članova uglavnom pogrešno pridijeljene nekoj od brojnijih kategorija.

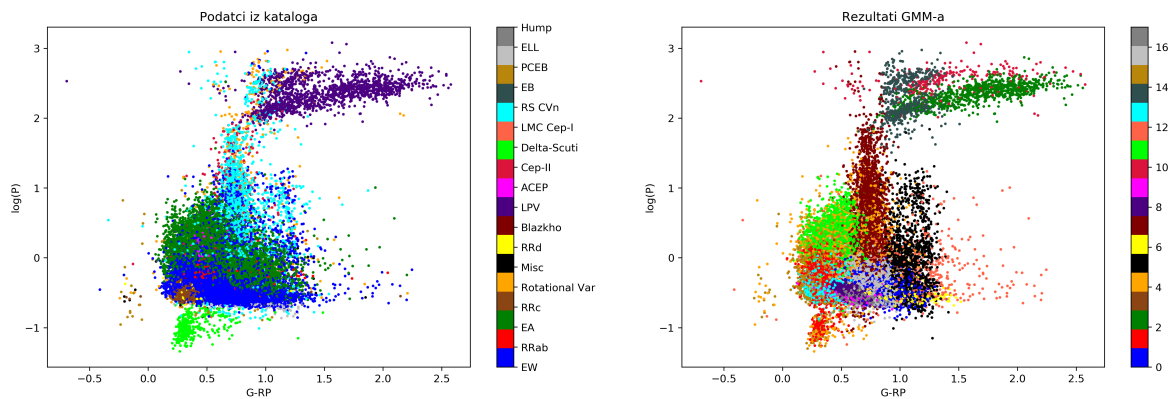
5.2 Rezultati GMM algoritma

Pri pokretanju GMM algoritma, potrebno je odrediti broj komponenti, odnosno broj Gaussovih raspodjeli od kojih su sastavljeni podatci, što će odgovarati i broju grupa na koje će GMM podijeliti podatke. Intuitivno je da se u tom slučaju za broj komponenti stavi i broj vrsta promjenjivih zvijezda u katalogu, koji je 18, ali prevelik broj komponenti može uzrokovati greške kod GMM algoritma. Primjeri rezultata GMM-a za 18 komponenti dani su na slikama 5.10 i 5.11. Boje dodijeljene kategorijama promjenjivih zvijezda odabrane su kako bi se različite kategorije dobro razlikovale na slici te ne nose nikakvo posebno značenje. Redoslijed kategorija na skali boja jest redoslijed kojim se pojavljuju u katalogu. Nazivi kategorija za rezultate GMM-a dodijeljeni su automatski jer je GMM nenadzirana metoda. Odmah je jasno da algoritam dijeli podatke na previše grupa, odnosno pronalazi grupe i tamo gdje ih nema.



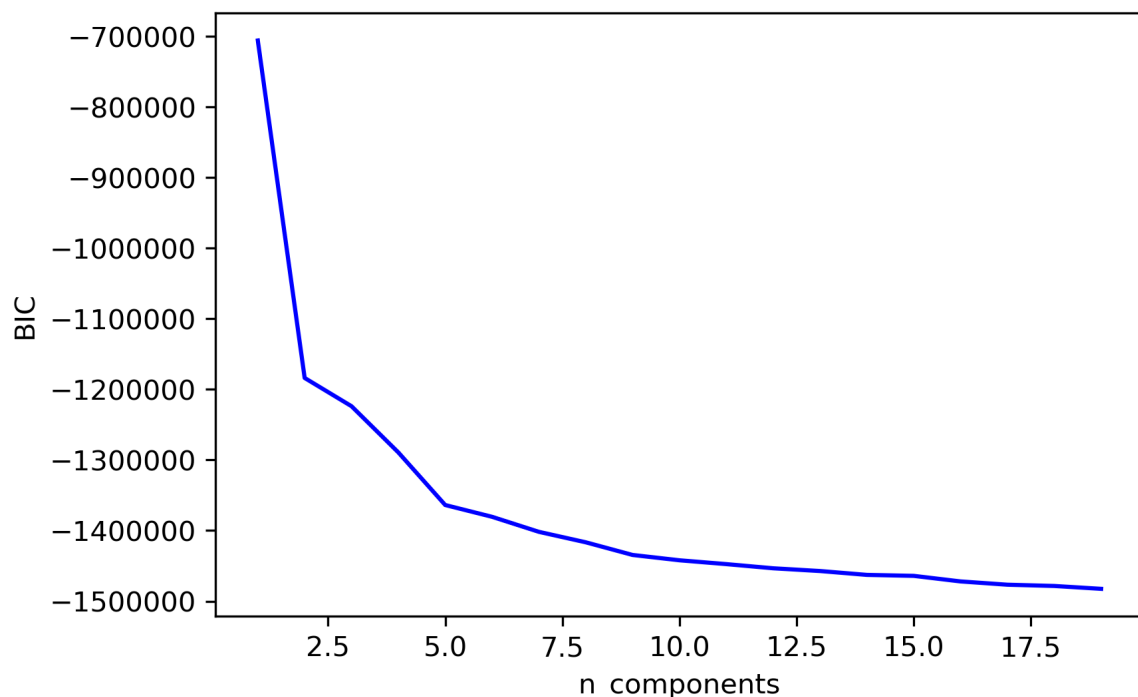
Slika 5.10: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko amplituda - logaritama perioda grafa za 18 komponenti.

Metoda za određivanje broja komponenti kod GMM modela jest tzv. *Bayesian Information Criterion* (BIC), statistički kriterij koja se koristi za izbore između dva ili više različitih modela. Što je vrijednost BIC-a niža, to je model bolji izbor. Kod



Slika 5.11: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko G-RP boja - logaritam perioda grafa za 18 komponenti.

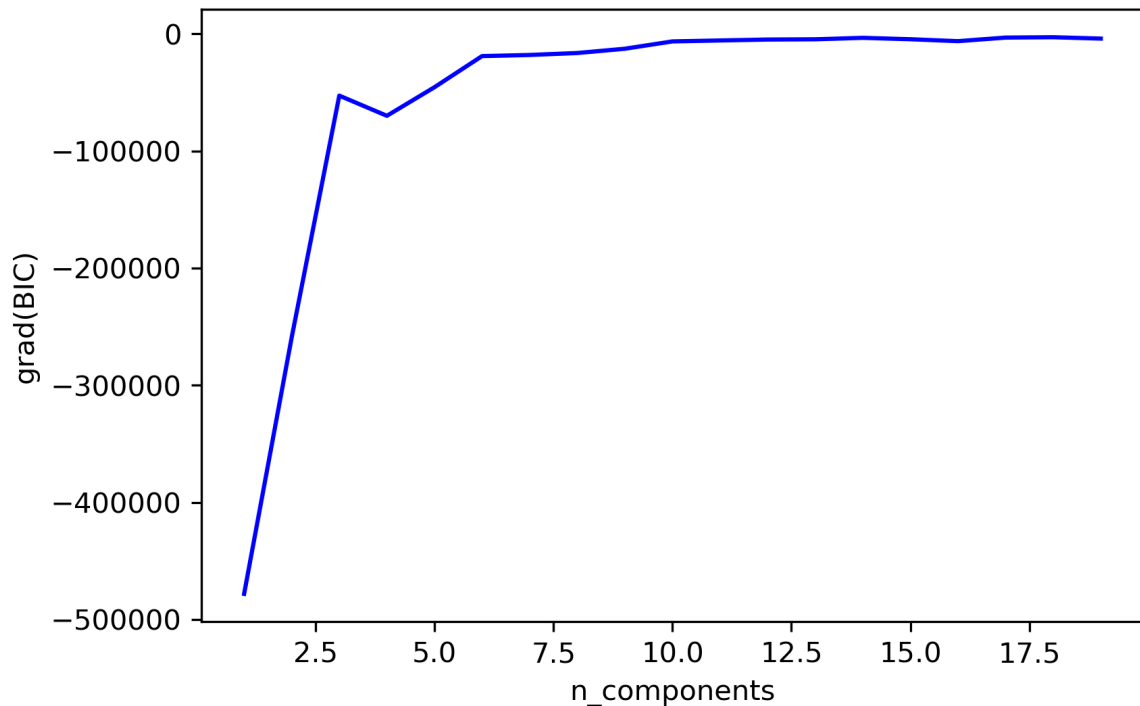
GMM-a to će značiti da je broj komponenti koji odgovara najnižem BIC-u najbolji izbor. Računanje vrijednosti BIC-a za GMM je ugrađeno u *Scikit-Learn* paket pa je ta vrijednost određena za brojeve komponenti od 1 do 20 te je nacrtan graf koji prikazuje ovisnost BIC-a o broju komponenti, dan na slici 5.12.



Slika 5.12: Graf ovisnosti vrijednosti BIC-a o broju komponenti za GMM model.

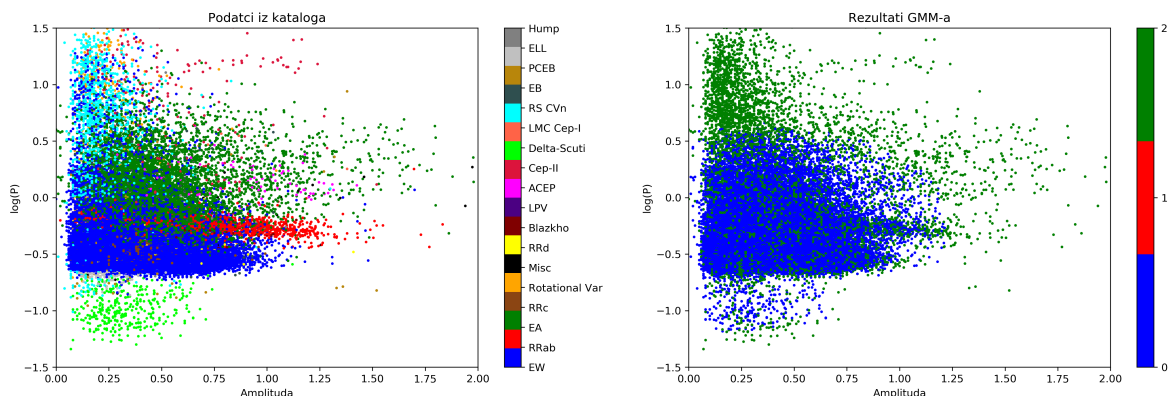
Sa slike izgleda kao da će model biti bolji za veći broj komponenti, ali već znamo da to nije točno. Problem je u tome što BIC kriterij ne djeluje dovoljno dobro na kompleksnije modele. Međutim, sa slike 5.12 vidi se da krivulja ima različite nagibe na različitim dijelovima. U takvom slučaju, može se promatrati promjena nagiba

krivulje, odnosno ovisnost gradijenta BIC-a o broju komponenti [36]. Taj graf dan je na slici 5.13.



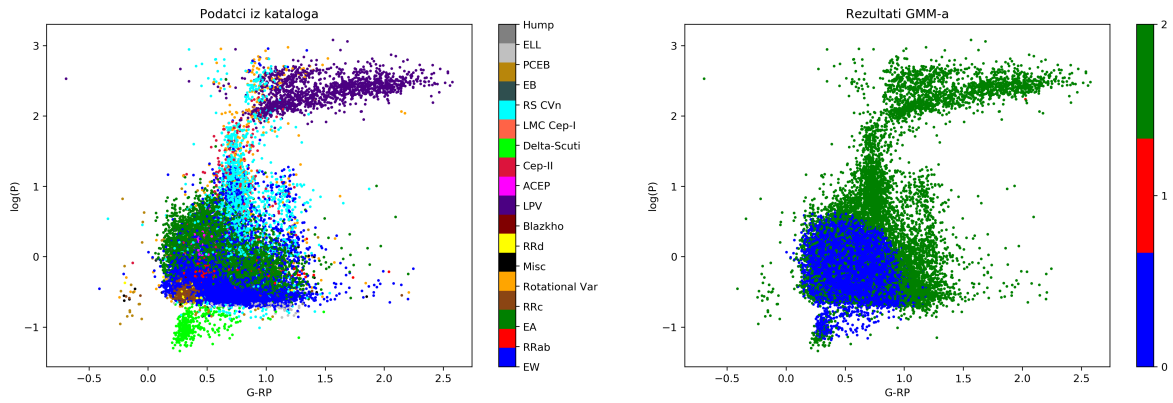
Slika 5.13: Graf ovisnosti vrijednosti gradijenta BIC-a o broju komponenti za GMM model.

S grafa je vidljivo da postoje točke infleksije na mjestima koja otprilike odgovaraju brojevima komponenti 3, 7 i 10. Međutim, kao što se vidi sa slika 5.14, 5.15, 5.16 i 5.17, modeli s 3 i 7 komponenti ne stvaraju dovoljno grupa za dobru podjelu podataka. Iz tih razloga je uzet broj komponenti 10 za grupiranje podataka.

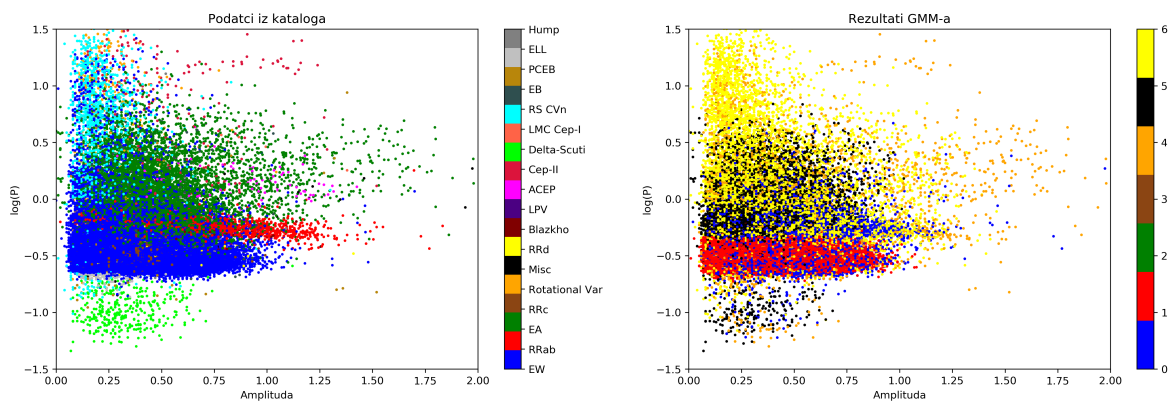


Slika 5.14: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko amplituda - logaritam perioda grafa za 3 komponente.

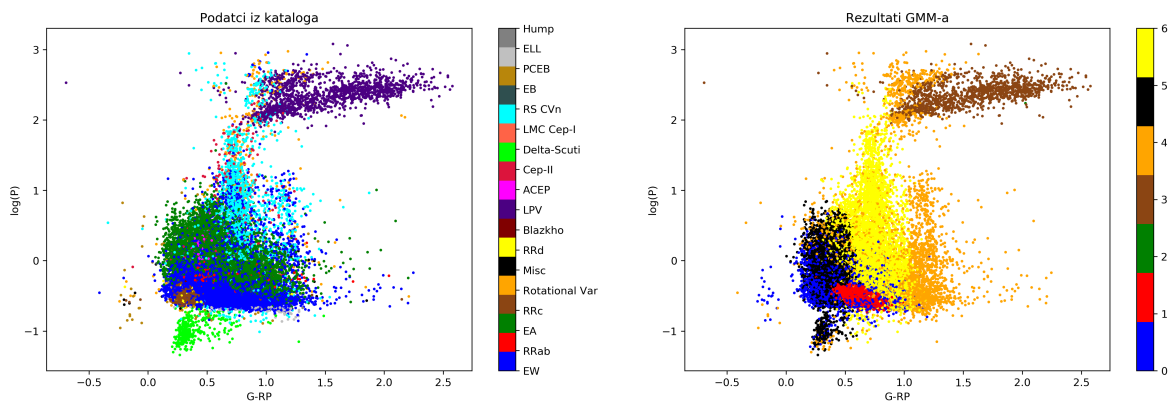
Na slikama 5.18, 5.19, 5.20, 5.21 i 5.22 dane su usporedbe klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu.



Slika 5.15: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko G-RP boja - logaritam perioda grafa za 3 komponente.

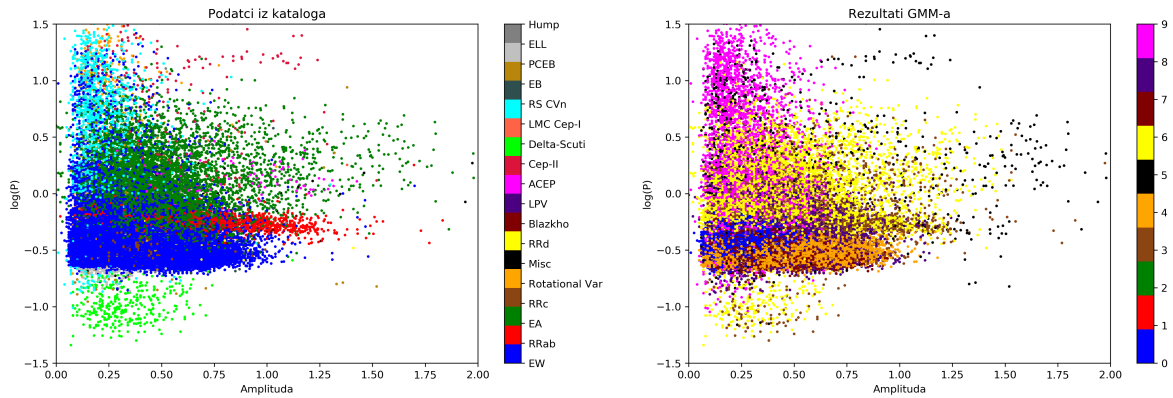


Slika 5.16: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko amplituda - logaritam perioda grafa za 7 komponenti.

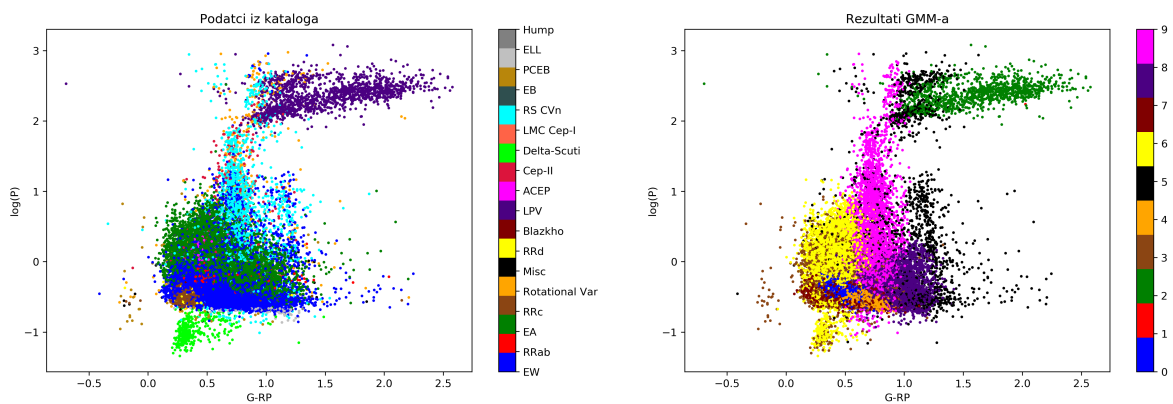


Slika 5.17: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko G-RP boja - logaritam perioda grafa za 7 komponenti.

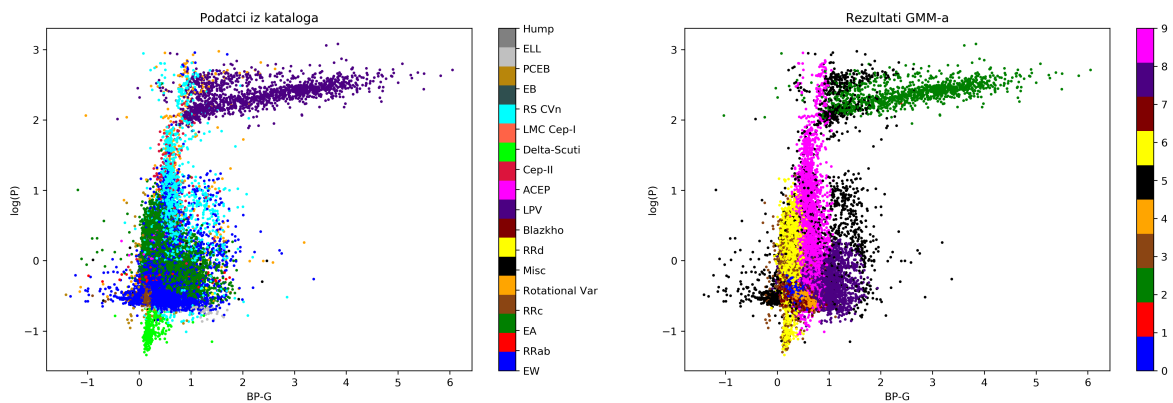
S grafova je vidljivo da su rezultati znatno lošiji od rezultata dobivenih klasificiranjem preko SVM-a. Broj grupa na koje GMM dijeli podatke manji je od broja vrsta promjenjivih zvijezda u katalogu, ali to je očekivano jer je uzet broj komponenti 10. Uz to, algoritam svrstava podatke koji su u katalogu iste vrste u različite grupe. Ta



Slika 5.18: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko amplituda - logaritam perioda grafa.

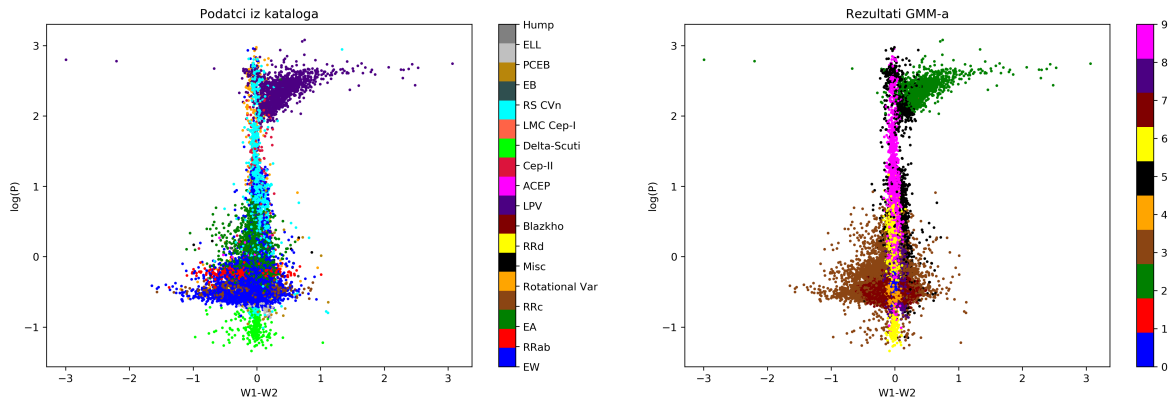


Slika 5.19: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko G-RP boja - logaritam perioda grafa.

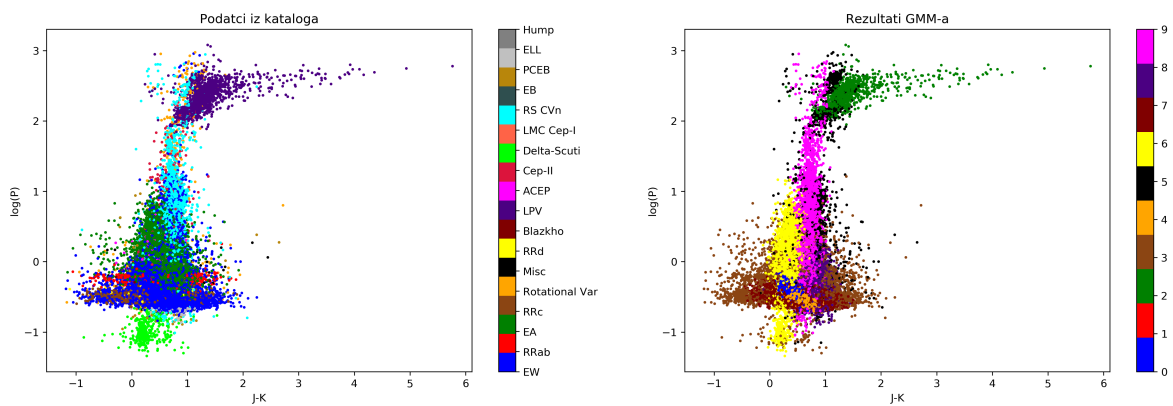


Slika 5.20: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko BP-G boja - logaritam perioda grafa.

pogreška ima smisla jer, za razliku od SVM-a, algoritam sam definira grupe u koje će svrstati podatke pa nije čudno da su podatci koji se na grafovima nalaze na bliskim područjima svrstani u iste grupe. Također, podatci nisu jednoliko raspoređeni već im je raspodjela nepravilna, a to stvara probleme kod algoritma, kao što je već spome-



Slika 5.21: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko $W1-W2$ boja - logaritam perioda grafa.



Slika 5.22: Usporedba klasifikacija promjenjivih zvijezda iz kataloga i klasifikacija prema GMM algoritmu preko $J-K$ boja - logaritam perioda grafa.

nuto u poglavlju 4.1.2. Važno je i napomenuti da EM algoritam kojeg GMM koristi pronalazi lokalne, a ne globalne optimalne točke tako da rezultati modela mogu biti različiti pri različitim pokretanjima programa čak i za iste početne postavke, što znači da rezultati GMM modela ponekad mogu ispasti bolje, a ponekad lošije [36].

6 Zaključak

U ovom radu korištene su metode strojnog učenja za klasifikaciju promjenjivih zvijezda kako bi se stekao uvid u mogućnost automatizacije tog procesa. Podatci o periodima, amplitudama te bojama preko različitih filtera od oko 70000 različitih promjenjivih zvijezda iz Catalina, Gaia, WISE i 2MASS pregleda neba dani su SVM i GMM algoritmima strojnog učenja kako bi se promjenjive zvijezde svrstale u različite vrste preko metoda klasifikacije i grupiranja. Dobiveni rezultati zatim su uspoređeni s klasifikacijom promjenjivih zvijezda iz Catalina pregleda neba kako bi se moglo promotriti koliko su precizni te koja metoda daje bolje rezultate. Rezultati SVM modela pokazuju dobro poklapanje s klasifikacijom iz Catalina pregleda neba uz odstupanja kod vrsta zvijezda za koje postoji malen uzorak u katalogu te za zvijezde različitih vrsta sa sličnim parametrima pomoću kojih je treniran algoritam. Pri pokretanju GMM algoritma odlučeno je da će algoritam podatke dijeliti na manji broj grupa nego što je kategorija promjenjivih zvijezda u katalogu, što očekivano daje manje precizne rezultate nego SVM algoritam. To je odlučeno jer opažane veličine ne definiraju potpuno odijeljene skupove u prostoru parametara pa za prevelik broj grupa algoritam daje nejasne rezultate. Osim toga, odstupanja uzrokuje i oblik raspodjele podataka. Uz navedene probleme, za razliku od klasifikacije preko SVM metode koja ima fizikalnu osnovu jer su klase u koje svrstaje podatke određene prema stvarnim vrstama promjenjivih zvijezda, grupiranje preko GMM metode samo promatra raspodjelu u prostoru određenih parametara bez prethodnog znanja o kategorijama primjenjivih zvijezda. Zbog toga je i očekivano da će GMM metoda grupiranja dati veća odstupanja od kategorizacije dane u katalogu nego klasifikacija SVM metodom. Također, zbog EM algoritma koji je jedna od ključnih stavki GMM modela, rezultati mogu ispasti različiti pri svakom pokretanju programa zato što se algoritam može pokrenuti iz različitih početnih točaka. Stoga se za iste postavke mogu dobiti različite grupe te je moguće da se mogu dobiti i rezultati koji bolje odgovaraju katalogu od onih prikazanih u ovom radu.

Kod primjene SVM metode, a možemo pretpostaviti i kod primjene drugih sličnih metoda nadziranog strojnog učenja, relativna zastupljenost pojedine klase u uzorku za treniranje algoritma utječe na ishod klasifikacije. Kako bi se u budućnosti dobili bolji rezultati moguće je smanjiti broj kategorija promjenjivih zvijezda u koje se svr-

staju podatci. To bi se moglo učiniti tako da se vrste promjenjivih zvijezda za koje postoji malen uzorak u katalogu sve svrstaju u jednu kategoriju. Kategorija *miscellaneous* (ostalo) već jest jedna od 18 kategorija u katalogu pa bi se sve nisko zastupljene vrste mogle svrstati u tu kategoriju. Također, dodavanje više parametara u modele moglo bi smanjiti problem miješanja kategorija sa sličnim parametrima kod SVM modela kao i omogućiti GMM modelu veću raznovrstnost pri stvaranju grupa. Primjer parametra koji bi mogao poslužiti takvoj svrsi bio bi položaj zvijezde u H-R dijagramu jer je to parametar koji se jasno razlikuje za različite vrste promjenjivih zvijezda.

Bibliography

- [1] Variables: What Are They and Why Observe Them?, (24.08.2017), American Association of Variable Star Observers, <https://www.aavso.org/variables-what-are-they-why-observe-them>, 22.04.2020.
- [2] Ryden, B. Introduction to Cosmology: Cambridge University Press, 2006
- [3] The Cosmic Distance Ladder, (06.05.2010), American Association of Variable Star Observers, <https://www.aavso.org/cosmic-distance-ladder>, 22.04.2020.
- [4] Eyer, L., Rimoldini, L., Audard, M., Anderson, R. Nienartowicz, K., et al.: Gaia Data Release 2. Variable stars in the colour-absolute magnitude diagram. // Astronomy and Astrophysics - A&A, EDP Sciences, Vol. 623, A110 (2019), str. 4.
- [5] Gravitational Microlensing, Las Cumbres Observatory, <https://lco.global/spacebook/exoplanets/gravitational-microlensing/>, 6.8.2020.
- [6] Paczynski, B.: Gravitational Microlensing by the Galactic Halo. // The Astrophysical Journal, Vol. 304, 1986, str. 1.
- [7] Samus N.N., Kazarovets E.V., Durlevich O.V., Kireeva N.N., Pastukhova E.N., General Catalogue of Variable Stars: Version GCVS 5 // Astronomy Reports, Vol. 61, 1 (2017), str. 80-88
- [8] Clayton, Geoffrey C.: The R Coronae Borealis Stars. // Publications of the Astronomical Society of the Pacific, Vol. 108, 721 (1996), str. 225-241.
- [9] Miller, A. A., et al.: Discovery Of Bright Galactic R Coronae Borealis And Dy Persei Variables: Rare Gems Mined From Acvs. // The Astrophysical Journal, Vol. 755, 2, (siječanj 2012), str. 98.
- [10] Soszyński, I., et al.: The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of Variable Stars. V. R Coronae Borealis Stars in the Large Magellanic Cloud. // Acta Astronomica, Vol. 59, 4 (prosinac 2009), str. 335-347.

- [11] Aerts, C., Christensen-Dalsgaard, J., Kurtz, D. W. Asteroseismology: Springer, str. 37–38, 2010
- [12] Carroll, B.W., Ostlie D.A. An Introduction to Modern Astrophysics: Pearson, 2014
- [13] Eddington, A.: The Pulsation Theory of Cepheid Variables // The Observatory, Vol. 40, (kolovoz 1917), str. 290-293.
- [14] About CSS, The University of Arizona, <https://catalina.lpl.arizona.edu/about>, 30.5.2020.
- [15] Drake, A.J., et al.: The Catalina Surveys Periodic Variable Star Catalog // The Astrophysical Journal Supplement Series, Vol. 213, 1 (srpanj 2014).
- [16] Gaia Collaboration: The Gaia mission // Astronomy & Astrophysics, Volume.595, (studeni 2016).
- [17] Gaia Collaboration: Gaia Data Release 2. Summary of the contents and survey properties // Astronomy & Astrophysics, Vol. 616, (kolovoz 2018).
- [18] Magnier, E.A., et al.: Pan-STARRS Pixel Analysis : Source Detection and Characterization // eprint arXiv:1612.05244, (prosinac 2016).
- [19] Wolf, C., et al.: SkyMapper Southern Survey: First Data Release (DR1) // Publications of the Astronomical Society of Australia, Vol. 35, (veljača 2018).
- [20] The AllWISE Data Release, (13.11.2013), The Wide-field Infrared Survey Explorer at IPAC, <http://wise2.ipac.caltech.edu/docs/release/allwise/>, 9.7.2020.
- [21] Wide-field Infrared Survey Explorer (WISE) and NEOWISE, Infrared Processing and Analysis Center, <https://irsa.ipac.caltech.edu/Missions/wise.html>, 9.7.2020.
- [22] Wright, E.L., et al.: The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance // The Astronomical Journal, Vol. 140, 6 (prosinac 2010), str. 1868-1881.

- [23] Skrutskie, M.F., et al.: The Two Micron All Sky Survey (2MASS) // The Astronomical Journal, Vol. 131, 2 (veljača 2006), str. 1163-1183.
- [24] Schuster, A.: On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena // Terrestrial Magnetism, Vol. 3, 1 (1898), str. 13-41.
- [25] Lomb, N.R.: Least-Squares Frequency Analysis of Unequally Spaced Data // Astrophysics and Space Science, Vo. 39, 2 (1976), str. 447-462.
- [26] Scargle, J.D.: Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. // Astrophysical Journal, Vol. 263, (prosinac 1982), str. 835-853.
- [27] Vanderplas, J. T., Ivezić, Ž.: Periodograms For Multiband Astronomical Time Series. // The Astrophysical Journal, Vol. 812, 1 (2015), str. 18.
- [28] The Catalina Surveys Data Release 2 (CSDR2), Catalina Sky Survey, <http://nesssi.cacr.caltech.edu/DataRelease/>, 11.8.2020.
- [29] An introduction to machine learning with Scikit-Learn, Scikit-Learn, <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>, 9.7.2020.
- [30] About us, Scikit-Learn, <https://scikit-learn.org/stable/about.html>, 9.7.2020.
- [31] Pedregosa et al.: Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research, Vol. 12, (2011), str. 2825-2830.
- [32] 1.4. Support Vector Machines, Scikit-Learn, <https://scikit-learn.org/stable/modules/svm.html>, 9.7.2020.
- [33] 2.1. Gaussian mixture models, Scikit-Learn, <https://scikit-learn.org/stable/modules/mixture.html>, 9.7.2020.
- [34] 2.3.2. K-means, Scikit-Learn, <https://scikit-learn.org/stable/modules/clustering.html#k-means>, 9.7.2020.

- [35] Demonstration of k-means assumptions, Scikit-Learn, https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_assumptions.html, 9.7.2020.
- [36] Gaussian Mixture Model clustering: how to select the number of components (clusters), (21.11.2018), Towards Data Science, <https://towardsdatascience.com/gaussian-mixture-model-clusterization-how-to-select-the-number-of-components-clusters-553bef45f6e4>, 22.7.2020.