

Informatika Fakultatea

Informatika Ingeniaritzako Gradua

▪ Gradu Amaierako Lana ▪

Konputagailuen Ingeniaritza

Bideo ekintzen sailkapena Point Clouden bidez

Joseba Berasain Garmendia

2020ko ekainaren 20a

Zuzendariak

Basilio Sierra Araujo

Naiara Aginako Bengoa

Laburpena

Proiektu honetan bideotan gertatzen diren ekintzen sailkapena egin da Point Cloudak erabiliz. Ekintza hauek, pertsona batek egindako ekintza sinpleak izango dira (oinez ibili, orraztu, erre, etab.). Ekintza konkretu batzuk deskribatzen dituen Point Clouden datu-base batetik abiatuta eta Weka softwarea paketea erabiliz, sailkapen automatikoa gauzatuko da.

Point Cloud hauek bideoa irudikatzen duten edge puntuak dira, non hauek, XYZ koordenadekin identifikatzen diren. Hauek, Wekan sailkatzeko, lehenengo aurre-prozesaketa bat jasan dute, garbiketa. Ondoren, bi prozesu ezberdin jaso dituzte: M2DP algoritmoaren erabilpena deskriptore globalak lortzeko, eta Point Clouden proiektzioen irudiekin lana. Jarraian, prozesu ezberdin bakoitzarekin lortutako emaitzak alderatuak izan dira. Lehenengo prozesuan puntuak garbitu eta puntuen zorizko aukeraketak egin dira (%20, %50 eta %75) eta M2DP algoritmoa erabiliz deskriptore globalak lortu dira Wekan sailkatzeko.

Amaitzeko, bigarren prozesuan, lehenengo prozesuan lortu diren zorizko %20en puntuak erabili dira. Puntu hauekin Point Cloudaren proiektzioak lortu dira. Azkenik, behin irudiak lortuta, datuak .arff fitxategietan ipini dira eta ondoren fitxategi hauek Wekan sailkatuak izan dira.

Gaien Aurkibidea

Laburpena	1
Gaien aurkibidea	3
Irudi eta taulen zerrenda	5
1. Aurkezpena	7
1.1. Sarrera	8
1.2. Proiektuaren deskribapen orokorra	8
2. Proiektuaren helburuen dokumentua	11
2.1. Irismena	12
2.2. Lanaren deskonposaketa	12
2.3. Atazak	13
2.4. Kronograma	15
2.5. Dedikazio estimazioa	16
2.6. Komunikazio plana	17
2.7. Kalitate plana	17
2.8. Arriskuen plana	18
2.9. Eskuraketen kudeaketa	19
2.10. Lan metodologia	19
2.11. Denbora erreala	20
3. Weka	21
3.1. Sarrera	22
3.2. Instalazioa eta exekuzioa	22
3.3. .arff fitxategiak	22
3.4. Esploratzailea	23
3.4.1. Aurreprozesamendua	24
3.4.2. Saikapena	24
3.4.3. Datuen ikuskaritza	25
3.5. Sailkatzaileak	26
3.5.1. K-NN	26
3.5.2. C.45	27
3.5.3. LMT	27
3.5.4. Random Forest	27
3.5.5. Naive Bayes	28
3.5.6. SMO	28

3.6. Multi-sailkatzaileak	29
3.6.1. Bagging.....	29
3.6.2. Bozketa.....	29
3.7. Aldagai aukeraketa.....	29
4. Proiektuko datu-basea	31
4.1. Sarrera.....	32
4.2. Ezaugarriak.....	33
5. Garapena	35
5.1. Ingurunea	36
5.2. Point Cloudetatik deskriptore globalen .arff-etara	36
5.3. Irudietatik .arff-etara	37
6. Emaitzak	39
6.1. Sarrera.....	40
6.2. Zorizko garbiketaren emaitzak.....	40
6.2.1. Zoriz aukeratutako puntuen %20	40
6.2.2. Zoriz aukeratutako puntuen %50	44
6.2.3. Zoriz aukeratutako puntuen %75	47
6.2.4. Puntu guztiak erabiliz.....	50
6.2.5. Emaitzen analisia	54
6.3. Proiektzioen emaitzak	54
6.3.1. XY ardatzeko proiektzioak.....	55
6.3.2. XZ ardatzeko proiektzioak.....	57
6.3.3. YZ ardatzeko proiektzioak	59
6.3.4. Emaitzen analisia	62
6.4. Bozketaren emaitzak.....	62
7. Ondorioak.....	65
7.1. Ondorioak.....	66
7.2. Etorkizuneko lana.....	67
Bibliografia.....	69
Eranskina. Kodeak.....	71

Irudi eta Taulen zerrenda

IRUDIAK

1.1. irudia	Proiektuaren deskribapena	9
2.1. irudia	LDE diagrama	12
2.2. irudia	Kronograma	16
3.1. irudia	Weka esploratzailea	23
3.2. irudia	Confusion matrix baten adibidea	25
3.3. irudia	Datuen ikuskaritza	25
3.4. irudia	K-NN adibidea k=3 izanik	26
3.5. irudia	Naive Bayes eredu probabilitistikoa	28
4.1. irudia	Gurpilarena egiten	32
4.2. irudia	Erretzen	32
4.3. irudia	Esgriman	32
4.4. irudia	Oinez	32
4.5. irudia	Orrazten	32
4.6. irudia	Ukabilkada ematen	32
4.7. irudia	Bideoen ekintzen Point Clouden zati bat	33
4.8. irudia	BW irudi baten adibidea (orrazten)	33
4.9. irudia	Gray irudi baten adibidea (orrazten)	34
5.1. irudia	Point Cloudetatik deskriptore globalen .arff-etara	37
5.2. irudia	Irudietatik .arff-etara	38

TAULAK

2.1. taula	Dedikazio estimazioa	17
2.2. taula	Benetako dedikazioa	20
6.1. taula	Puntuen %20 filter gabe, asmatze tasak	40
6.2. taula	Puntuen %20 filter CfsSubsetEval, asmatze tasak	41
6.3. taula	Puntuen %20 filter PrincipalCaponents, asmatze tasak	41
6.4. taula	Puntuen %20 filter ReliefAttributeEval, asmatze tasak	42
6.5. taula	Puntuen %20 filter SVMAttributeEval, asmatze tasak	43

6.6. taula	Puntuen %50 filter gabe, asmatze tasak.....	44
6.7. taula	Puntuen %50 filter CfsSubsetEval, asmatze tasak	44
6.8. taula	Puntuen %50 filter PrincipalCopenents, asmatze tasak	45
6.9. taula	Puntuen %50 filter ReliefAttributeEval, asmatze tasak	45
6.10. taula	Puntuen %50 filter SVMAttributeEval, asmatze tasak	46
6.11. taula	Puntuen %75 filter gabe, asmatze tasak.....	47
6.12. taula	Puntuen %75 filter CfsSubsetEval, asmatze tasak	48
6.13. taula	Puntuen %75 filter PrincipalCopenents, asmatze tasak	48
6.14. taula	Puntuen %75 filter ReliefAttributeEval, asmatze tasak	49
6.15. taula	Puntuen %75 filter SVMAttributeEval, asmatze tasak	50
6.16. taula	Puntuen %100 filter gabe, asmatze tasak.....	51
6.17. taula	Puntuen %100 filter CfsSubsetEval, asmatze tasak	51
6.18. taula	Puntuen %100 filter PrincipalCopenents, asmatze tasak	52
6.19. taula	Puntuen %100 filter ReliefAttributeEval, asmatze tasak	52
6.20. taula	Puntuen %100 filter SVMAttributeEval, asmatze tasak.....	53
6.21. taula	XY ardatzeko BW motako irudien asmatze tasak.....	55
6.22. taula	XY ardatzeko Gray motako irudien asmatze tasak.....	56
6.23. taula	XZ ardatzeko BW motako irudien asmatze tasak.....	58
6.24. taula	XZ ardatzeko Gray motako irudien asmatze tasak.....	59
6.25. taula	YZ ardatzeko BW motako irudien asmatze tasak.....	60
6.26. taula	YZ ardatzeko Gray motako irudien asmatze tasak.....	61

1



Aurkezpena

1.1. Sarrera

Gaur egun bideoen ekintzen sailkapena ikertzen ari den arazo bat da. Sailkapen hau hainbat aplikazioetan inplementatu nahi da, baina hau aspalditik datorren zerbait da. Gizakiak betidanik nahi izan du makina batekin jolastu, hitz egin, etab. Gainera, gaur egun, hainbat sektore ezberdinetan aplikatu daitekeela ere ikusi da: kirol errendimenduaren analisisan, zaintza kameretan, auto autonomoetan, eta abar.

Faktore ezberdinen eraginez oraindik bideoen sailkapenak hainbat arazo ditu: eszena desberdinak dira beti, bideoaren grabazioa angelu ezberdinetatik egon daitezke, bideoaren argia, bideoaren kalitatea, pertsonaren ezaugarri fisikoak kontuan hartzea zaila da. Gainera ezin da txantilo bat erabili, beti pertsona distantzia ezberdin batean egongo da. Beraz, mota honetako sailkapenak ongi aztertzea eta egitea zaila da.

Ekintzen azterketa egiteko hainbat teknika ezberdin daude. Proiektu honetan bideoen ekintzak Point Cloud bidez irudikatu dira, zehatzago, ekintzen edge puntuak erabili dira sailkapenak egiteko. Ondoren, puntu horien zorizko aukeraketa egin da informazioa berdina ematen duten puntuak kontuan ez hartzeko eta horrela azterketa azkartzeko. Bi esperimentu desberdin egin dira: Lehenengoan, puntu hauen deskriptore globalak atera dira eta ondoren sailkapena egin da. Bigarrenengan berriz, zorizko puntu horiekin, irudien proiektzioak lortu dira eta berriz ere sailkapena egin da.

1.2. Proiektuaren deskribapen orokorra

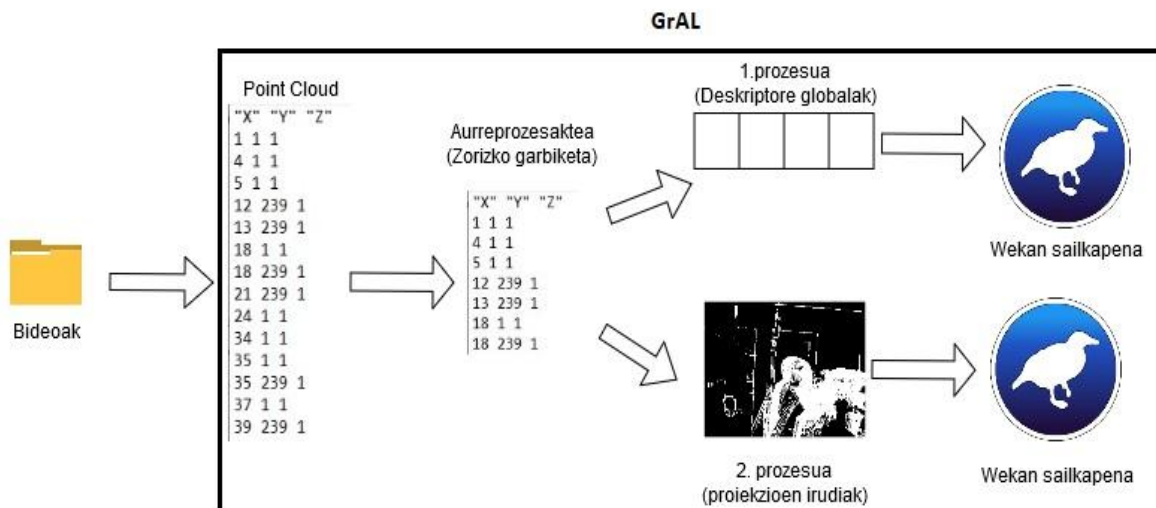
Proiektua EHU-ko Informatika Fakultatean garatu da. Konputazio Zientzia eta Adimen Artifiziala arloko Basilio Sierra Araujo eta Naiara Aginako Bengoa zuzendari bezala izanda. Proiektu honetan bideoen ekintzen sailkapena egiteko Point Cloudak erabili dira datu-base bezala eta hauei prozesaketa desberdinak aplikatu zaizkie. Beraz proiektuaren helburu nagusia bideoen Point Cloudak sailkatzea izan da.

Horretarako jarraitu diren atazak ondorengoak dira:

1. *Puntuen garbiketa eta zorizko aukeraketa:* Puntuak ugariak direnez, hauek ondo erabiltzeko, lehenengo garbiketa bat egin da errepikatutako puntuak ez aukeratzeko. Ondoren, lortutako emaitzarekin, zorizko aukeraketa bat egin da puntuaren %20, %50 eta %75 aukeratuz. Prozesu hau egiteko Garbiketa.sh izeneko programa erabili da.
2. *Deskriptore globalak erabili:* Datu-basea garbitu eta laburtu ondoren, M2DP-a [1] algoritmoa erabili da hauen deskriptore globalak sortzeko. Ondoren, deskriptore globalen .arff-ak lortu dira. Amaitzeko, hauek Weka softwarean datozen algoritmo batzuekin sailkatu dira eta emaitzen azterketa egin da.

3. *Point Clouden proiektzioak erabili:* Behin deskriptore globalak erabilia Point Clouden proiektzioen irudiak erabili dira sailkapena egiteko. Horretarako, garbitutako eta zoriz aukeratutako %20 puntuen datu-basea erabili da. Irudiak perspektiba ezberdinetan daudenez (XY, XZ eta YZ ardatzak) eta haien artean ezberdinak direnez, Wekan ere sailkapena egiterakoan irudi desberdin bezala tratatu dira. Horretarako Aukeratu.sh programa erabili da, honek karpeta bateko irudiak ardatzaren arabera banatuko ditu.
4. *Irudiak banatu ondoren Wekan aztertu:* Behin Point Clouden bisten irudiak edukita, Weka softwarean XY, XZ eta YZ irudiak algoritmo batzuekin sailkatu dira eta emaitzen azterketa egin da.

Proiektu honen ataletan ikusten den bezala bi prozesu ezberdin egin dira. Jarraian, prozesu bakoitzean lortutako emaitzen analisia eta konparaketa egin da. 1.1 irudian ikusi daiteke proiektuaren prozesua.



1.1 irudia: Proiektuaren deskribapena.

Memoria honek proiektuan zehar egin den lana azaltzen du, hainbat ataletan antolatuta. 1. atalean proiektuaren aurkezpena aurkitu daiteke, sarrera eta deskribapen orokor batekin. 2. atalean proiektuaren kudeaketaren azalpena dago. Jarraian, 3. atalean Wekaren azalpenak aurkitu ditzakegu datuen analisirako. 4. atalean datu-basearen deskribapena egin da. 5. eta 6. ataletan proiektuaren garapena ikusi dezakegu, 5. atalean aurre-prozesaketak eta 6. atalean lortutako emaitzak. Amaitzeko, 7. atalean ondorioak jasotzen dira.

Dokumentuaren amaieran bibliografia eta eranskinak aurkitu ditzakegu. Bibliografian proiektuan zehar erabili diren informazio iturriak edo erreferentziak azaltzen dira, eta eranskinetan sortu eta erabili diren kodeak.

2



Proiektuaren helburuen dokumentua

2.1. Irismena

Proiektuaren helburua bideoetan agertzen diren ekintzen Point Cloudak erabiliz analisi automatikoa lantzea izan da. Analisia egiteko Weka softwarea erabili da, ikasketa automatikoan, datu meatzaritzan, Big Datan, eta abarretan erabiltzen den bezala.

Azterketa egiterakoan Point Cloudak erabili dira bideoen ekintzen ordezkari bezala. Hauek ekintzaren *edge* puntuak dira eta prozesu ezberdinak aplikatu ondoren sailkatuak izan dira.

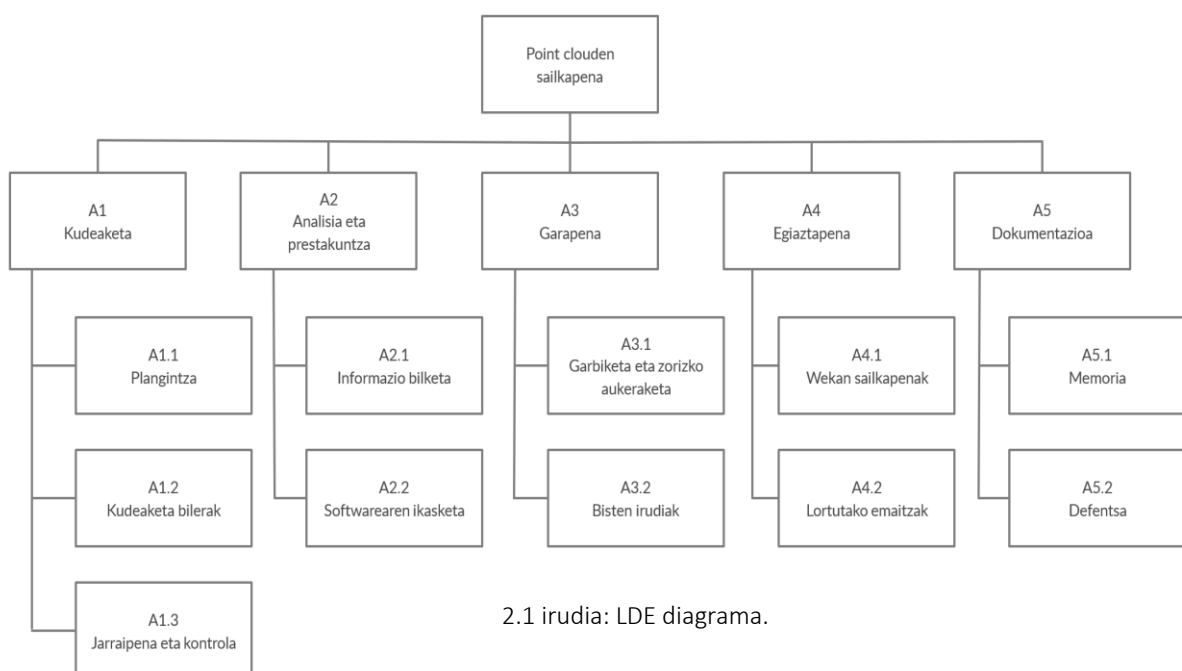
Hau burutzeko, lehenik eta behin Point Clouden garbiketa bat egin da, askotan errepikatzen ziren puntuak alde batera uzteko. Garbiketa honekin informazio berdina ematen diguten puntuak alde batera utzi dira, puntuen kopurua txikituz eta hauen moldaketa erraztuz. Puntu hauetatik zoriz %20, %50 eta %75 aukeratu dira eta M2DP-a [1] algoritmoaz baliatuz, hauen deskriptore globalak sailkatu ahal izan dira Wekan.

Behin sailkapen horiek eginda eta emaitzak lortuta, beste esperimentu bat gauzatu da: zorizko %20ko puntuen datu-basea erabiliz proiektzioen irudiak lortu dira. Irudi hauek XY, XZ eta YZ ardatzetan daudenez eta beraien artean ezberdinak direnez, bakoitza bere aldetik tratatu da Wekan sailkatzeko eta emaitzak alderatzeko.

Weka plataforman sailkapenak egiteko hainbat filtro eta sailkatzaile erabili dira. Amaitzeko, emaitza guzti hauen artean konparaketak egin dira eta ondorioak atera.

2.2. Lanaren deskonposaketa

2.1 irudiko diagraman proiektuko lan pakete ezberdinak ikusi daitezke.



2.3. Atazak

Ondoren azaltzen diren atazak proiektua garatzeko beharrezkoak izan direnak dira.

A1. Kudeaketa

A1.1 Plangintza

A1.1.1 Irismena

A1.1.2 LDE diagrama

A1.1.3 Atazak

A1.1.4 Kronograma

A1.1.5 Dedikazio estimazioa

A1.1.6 Kalitate plana

A1.1.7 Arriskuen plana

A1.1.8 Eskuraketen Kudeaketa

A1.1.9 Lan metodologia

A1.2 Kudeaketa bilerak

A1.2.1 Lehenengo bilera

A1.2.2 Bigarrenengo bilera

A1.2.3 Hirugarren bilera

A1.2.4 Itxiera bilera

A1.3 Jarraipena eta kontrola

A1.3.1 Desbiderapenak

A2. Analisia eta prestakuntza

A2.1 Informazio bilketa

A2.1.1 Artikuluak

A2.1.2 Datu-basea

A2.1.3 Weka

A2.2 Softwarearen ikasketa

A2.2.1 Instalazioa

A2.2.2 Wekaren funtzionamendua ulertu

A3. Garapena

A3.1 Garbiketa eta zorizko aukeraketa

A3.1.1 Puntuen Garbiketa egin

A3.1.2 Puntuen zorizko aukeraketa egin

A3.1.3 Puntuak garbitu eta zorizko aukeraketa egin

A3.1.4 Deskriptore globalak atera

A3.2 Bisten irudiak

A3.2.1 Ardatzen arabera banatu

A4. Egiaztapena

A4.1 Wekan sailkapenak

A4.1.1 M2DP zorizko deskriptoreekin

A4.1.1.1 Oinarrizko sailkatzaileak

A4.1.1.2 Vote

A4.1.1.3 Bagging

A4.1.1.4 Aldagai aukeraketa

A4.1.2 Bistak erabiliz

A4.1.2.1 Oinarrizko sailkatzaileak

A4.1.2.2 Vote

A4.1.2.3 Bagging

A4.1.2.4 Bozketa sinplea

A4.2 Lortutako Emaitzak

A4.2.1 M2DP zorizko deskriptoreekin

A4.2.1.1 Oinarrizko sailkatzaileak

A4.2.1.2 Vote

A4.2.1.3 Bagging

A4.2.1.4 Aldagai aukeraketa

A4.2.2 Bistak erabiliz

A4.2.2.1 Oinarrizko sailkatzaileak

A4.2.2.2 Vote

A4.2.2.3 Bagging

A4.2.2.4 Bozketa simplea

A5. Dokumentazioa

A5.1 Memoria

A5.1.1 Informazio bilketa

A5.2 Defentsa

A5.2.1 Informazio bilketa

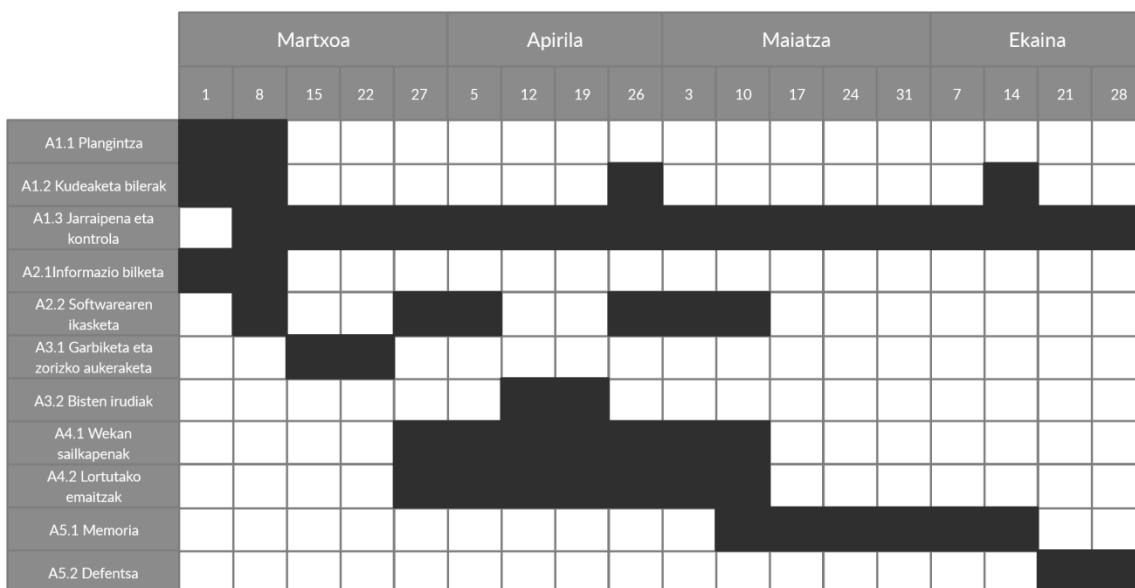
2.4. Kronograma

2.2 irudiaren kronograman proiektuaren plangintza ikus daiteke.

Bertan ikusten den bezala guztira 18 aste iraun ditu proiektuak. Lehenengo bi asteak bilera, plangintza eta informazio bilketarako izan dira batez ere.

Martxoan eta apirilean proiektuaren garapena landu da. Lehenengo garbiketa eta zorizko puntuak lortu dira. Jarraian, hauen deskriptore globalak lortu dira M2DP algoritmoarekin. Horiek sailkatu eta emaitzak lortu. Ondoren bisten irudiak landu dira. Amaitzeko irudien sailkapena egin da eta emaitzak lortu dira.

Maiatza eta ekaineko lehenengo bi asteak memoria garatzeko erabili da. Azkenik, ekaineko azken bi asteak defentsa prestatzeko erabili da.



2.2 irudia: Kronograma.

2.5. Dedikazio estimazioa

2.1 taulan proiektuaren dedikazio estimazioa ikus daiteke, hau da, proiektua garatzeko lan-pakete bakoitzak behar duen denboraren estimazioa.

Bertan ikusten den bezala guztira 300 ordu iraungo ditu proiektuaren garapenak. Lan-paketeka aztertzen badugu 2.1 taula, garbi ikusten da dedikazio gehien memoria egiteko eta defentsa prestatzeko direla.

Ala ere, garapena eta egiaztapena paketeak ere garrantzitsuenetarikoak dira. Beraz, biak batera hartuta, esan dezakegu denbora gehien bi hauei eskainiko zaiela, hauek elkarrekin lotuta baitoaz.

Azkenik, dedikazio gutxien behar duena kudeaketa lan-paketea izatea espero da.

Lan-Paketea	Dedikazio estimazioa (ordutan)
A1. Kudeaketa	20
A1.1 Plangintza	5
A1.2 Kudeaketa bilerak	4
A1.3 Jarraipena eta kontrola	11
A2. Analisia eta prestakuntza	40
A2.1 Informazio bilketa	6
A2.2 Softwarearen ikasketa	34
A3. Garapena	60
A3.1 Garbiketa eta zorizko aukeraketa	30
A3.2 Bisten irudiak	30
A4. Egiaztapena	70
A4.1 Wekan sailkapenak	50
A4.2 Lortutako emaitzak	20
A5. Dokumentazioa	110
A5.1 Memoria	90
A5.2 Defentsa	20
GUZTIRA	300

2.1 taula: Dedikazio estimazioa.

2.6. Komunikazio plana

Proiektuaren garapenean zehar zalantzak edo arazoak sortu badira, EHUko posta elektronikoa erabili da zuzendariekin kontaktuan jartzeko.

Bilerak egiteko zuzendarien bulegoak erabili dira hasiera batean. Ondoren, Covid-19aren arazoa izan dela eta, bilerak eGelako BlackBoard Collaborate bitartez egin dira. Bilerak orokorrean garapenari eta egiaztapenari buruz hitz egiteko ezin dira. Azkenik, itxiera bilera egingo da proiektuari amaiera emateko.

2.7. Kalitate plana

Proiektuaren kalitatea bermatzeko, ondorengo lerroetan kalitate plana aurkezten da.

Proiektuaren oinarritzko betekizunak:

- Point Clouden garbiketa eta zorizko aukeraketa egin.

- M2DP zorizko deskriptoreak lortu.
- Garbitu eta zoriz aukeratutako Point Clouden proiektzioen irudiak lortu.
- Wekaren bidez azterketa egin emaitzak lortzeko.
- Lortutako emaitzetatik ondorioak atera.

Proiektuaren kalitate dimentsioak:

- Point Clouden garbiketa ondo egitea eta zorizko aukeraketa ondo aplikatzea.
- Deskriptore globalak zuzen lortzea.
- Garbitu eta zoriz aukeratutako Point Clouden proiektzioen irudiak ondo banatzea ardatzaren arabera.
- Wekaren bidez azterketa egiteko erabilitako sailkatzaileak egokiak izatea.
- Lortutako emaitzak onak izatea.

2.8. Arriskuen plana

Proiektuaren garapenean ahalik eta arazo gutxien edukitzeko arriskuen plan bat egin da.

Arrisku bakoitzak izan dezakeen eragina ezberdina izango da. Horregatik, eragin handia duten arriskuei irtenbide azkarra eta egokia bilatzen saiatuko gara.

Jarraian, proiektuan zehar gerta daitezkeen arriskuen identifikazioa egin da, eta hauei irtenbide posibleetako bat eman zaie.

1. **Ordenagailuaren matxura:** Proiekturako erabiltzen den ordenagailua matxuratzea gerta daiteke.

-Arazo honen irtenbidea, lanean lortutako aurrerapen guztiak Google Driven gordetzea izango da. Honek uneoro segurtasun kopiak egiten ditu, gainera, aldaketak desegiteko eta aurreko bertsioak berreskuratzeko aukera ere dago.

2. **Atzerapenak:** Proiektuan zehar garatzaileak beste hainbat lan eta ikasketa ezberdin dituen, atzerapenak sortuko dira.

-Honi aurre egiteko, astero plangintza begiratu eta aldaketa posibleak aztertuko dira. Horrela, atzerapen bat badago, birplanteamendu berri batekin arazoa konponduko da.

3. Garapenean zailtasunak: Proiektuaren garapenean hainbat zailtasun aurkitu ditzakegu. Beraz, hauei denbora gehiago eskainiko zaie.

-Arazo hau ekiditeko, denbora gehiago behar duten atazei denbora gehiago eskainiko zaie eta plangintza berri egingo da lana moldatzeko.

Zerrenda honetan identifikatu gabeko beste arazo batzuk gertatuko dira. Beraz, hori gertatzen denerako prest egon beharra dago ahalik eta azkarren soluzioak emateko.

2.9. Eskuraketen kudeaketa

Proiektuaren garapenerako bost eskuraketa izango dira garrantzitsuenak.

Lehenik eta behin, Linux sistema instalatu beharko da bertan hainbat programekin lan egiteko. Horretarako, internetetik deskargatutako VMware makina birtuala instalatuko da.

Jarraian, proiektuan erabiliko diren bideon Point Cloudak lortu beharko dira. Datu-base hau zuzendarien eskutik lortuko da. Ondoren, Point Clouden bisten irudiak lortu beharko dira. Honetarako ere zuzendariekin ipiniko da harremanetan.

Amaitzeko, sailkapenak eta azterketak egiteko Weka programa instalatu beharko da. Hau Internetetik zuzenean deskargatuko da.

2.10. Lan metodologia

Proiektuaren garapenerako metodologia sinplea da. Garatzaileak beharrezko gauzekin prestatuko du ordenagailua (programak, makina birtuala, etab.) aurre-prozesaketak eta garapena egin ahal izateko. Zalantzak sortuko balira, zuzendariekin kontaktuan mezu bidez ipiniko da. Fase ezberdinak amaitu ondoren zuzendariekin bilerak egingo dira gauzak argitzeko.

Proiektua fase ezberdinetan garatu da. Lehenengo, proiektuan egin beharrekoa ondo identifikatu eta ulertuko da. Ondoren, datu-basea aztertu eta garbiketa eta zorizko aukeraketen programa egingo da jarraian aplikatzeko. Segidan, lortutako puntu horien deskriptore globalak erabiliz Wekan sailkatuko dira.

Sailkapenak egin ondoren, zoriz aukeratutako %20ko Point Clouden bisten irudiak lortuko dira. Jarraian, irudi hauek ardatzaren arabera banatzeko inplementazioa garatuko da eta Wekan sailkatuko dira.

Amaitzeko, lortutako emaitzak aztertu eta ondorioak aterako dira.

2.11. Denbora erreala

Proiektua amaituta, benetan dedikatu zaion denbora ikusi dezakegu 2.2 taulan.

Lan-Paketea	Benetako dedikazio (ordutan)
A1. Kudeaketa	20
A1.1 Plangintza	7
A1.2 Kudeaketa bilerak	4
A1.3 Jarraipena eta kontrola	9
A2. Analisia eta prestakuntza	38
A2.1 Informazio bilketa	8
A2.2 Softwarearen ikasketa	30
A3. Garapena	74
A3.1 Garbiketa eta zorizko aukeraketa	51
A3.2 Bisten irudiak	23
A4. Egiaztapena	82
A4.1 Wekan sailkapenak	52
A4.2 Lortutako emaitzak	30
A5. Dokumentazioa	116
A5.1 Memoria	96
A5.2 Defentsa	20
GUZTIRA	330

2.2 taula: Benetako dedikazioa.

Proiektuaren dedikazio estimazioarekin konparatuta (2.1 taula) aldaketa txikiak izan dira. A1 eta A2 atazetan ondo bete dira estimazioak. A3 eta A4 atazak garatzeko, esperotakoa baino denbora gehiago behar izan da. Amaitzeko, gainontzeko denbora A5 ataza egiten pasa da.

3



Weka

3.1. Sarrera

Weka [2] ikasketa automatikoan eta datu-meatzaritzan erabiltzen den softwarea da. Zeelanda Berriko Waikatoko unibertsitateak garatu zuen, Java lengoaian inplementatuta dago eta GNU-GPL lizentziapean dagoen software librea da.

Wekak [3] datuen aurre-prozesaketa egiteko eta ikasketa automatikorako teknika ezberdin eta ugariak eskaintzen ditu. Gainera, erabilerraza da bere interfazeari esker. Wekaren bidez datu-meatzaritzako ataza asko ebatzi daitezke: clustering-a, erregresioa, sailkapen-problemak, datuen aurreprozesamenduak, etab.

3.2. Instalazioa eta exekuzioa

Weka [3] software librea denez Wekaren orrialdetik deskargatzea nahikoa da. Weka deskargatzean .exe fitxategi bat deskargatuko da, ondoren, hau exekutatu eta Weka instalatuko da. Programa exekutatzeke nahikoa da instalazioan mahaigainean sortutako ikonoan sartzea.

3.3. .arff fitxategiak

Weka programak datuekin lan egin ahal izateko eta irakurketan arazorik ez izateko fitxategiak formatu berezi bat jarraitu behar dute. Formatu hau .arff da eta ondorengo egitura izan behar du:

1. *Goiburua*: Erlazioa definitzeko erabiltzen da, klasearen izena.

```
@relation <erlazio-izena>
```

2. *Atributuen Erazagupena*: Gune honetan atributuak erazagutuko dira motaz jarraituta.

```
@attribute <atributu-izena> <mota>
```

Motak ondorengoak izan daitezke:

- *Numeric*. Zenbaki errealak.
- *Integer*. Zenbaki osoak.
- *Date*. Data formatua, komatxoaren artean eta gidoiez edo hutsunez banatuta.
- *String*. Karaktere kateak.

- *Enumerate*. Atributuak har ditzakeen balio posibleak.
3. *Datuak*: Gune honetan azertu nahi ditugun datuak egongo dira. Datu hauek koma edo hutsune bitartez banatzen dira eta erlazio desberdinak lerro jauzi bitartez.

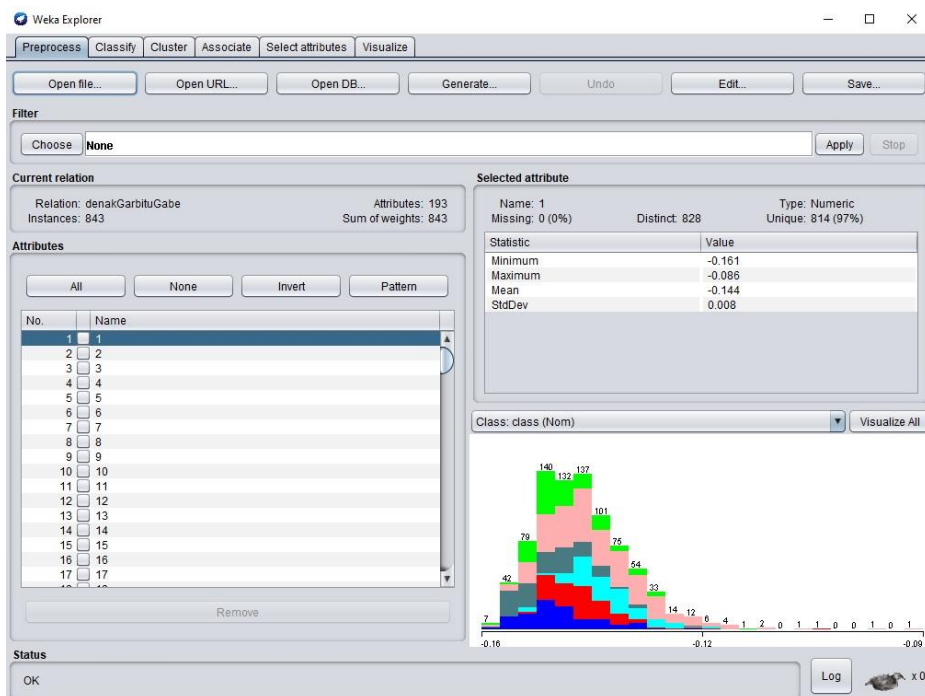
```
@data
1,2,3...
1 2 3...
```

3.4. Esploratzailea

Wekaren esploratzailea (Explorer) datuak aztertzeko eta ikusteko interfazea da. .arff fitxategi bat kargatu ondoren hainbat gauza ikus ditzakegu 3.1 irudian bezala, eta ondorengo zeregin ezberdinetan erabil dezakegu:

- Datuen aurreprozesamendua egiteko.
- Saillapenak egiteko.
- Atributuen aukeraketa egiteko.
- Datuen ikuskaritzarako.

Zeregin hauez gain, beste hainbat daude. Horiek direnez garrantzitsuenak eta proiektuan erabili direnak, horiek bakarrik azalduko dira jarraian.



3.1 irudia: Weka esploratzailea.

3.4.1 Aurreprozesamendua

Lanean hasteko datuak kargatu behar dira eta horretarako hainbat aukera daude. Proiektu honen garapenean bi soluzio daudenez bakoitza modu ezberdinean egin da. Garbituta eta zoriz aukeratutako .arff-ekin nahiko da hauek atzitzea, aldiz bisten irudiekin .arff-a atzitu eta ondoren irudien karpetaen helbidea adierazi da.

Datuen fitxategia kargatu ondoren iragazkiak (filter) aukeratzeko aukera dago. Iragazkiak aplikatuta sailkapenerako atributu batzuk hartzen dira kontuan. Bi taldetan banatzen dira iragazkiak: gainbegiratuak (supervised) eta gainbegiratu gabeak (unsupervised).

3.4.2 Sailkapena

Datuen fitxategia kargatuta, sailkapenera igaroko gara. Interfazean *classify* botoia sakatuko da horretarako.

Sailkatzeko aukera ugaria izango da. Bertan sailkatzaile asko daude ordenatuta karpeta motaren arabera. Sailkatzaile bakoitzak aukera ezberdinak eskaintzen ditu hau moldatua izateko. Sailkapena egin ondoren, lortutako emaitzei erreparatzean datu ugari aztertzen dira, non garrantzitsuenak edo esanguratsuenak ondorengoak diren:

- **Correctly Classified Instances:** Zuzen sailkatutako instantzien balioa ehunekotan.
- **Incorrectly Classified Instances:** Oker sailkatutako instantzien balioa ehunekotan.
- **TP rate:** Zuzen sailkatutako instantzien tasa klasearen arabera.
- **FP rate:** Oker sailkatutako instantzien tasa klasearen arabera.
- **Precision:** Zuzen sailkatutako instantzien proportzioa klaseko, klase hori bezala sailkatutako instantzien totalarekin zatituta.
- **Recall:** Iragarritako klase bezala sailkatutako instantzien proportzioa, klase horretakoak diren instantzien totalarekin zatituta.
- **ROC eremua:** 1 balioaren gertu ibiliko da sailkatzailea optimoa bada emandako datu-basearentzat. 0.5 baliotik gertu badago sailkatzailea ez dela fidagarria erakusten du.
- **F-measure:** Emaitzen zehaztasunaren neurria da. Horretarako doitasuna (Precision) eta estalduran (Recall) oinarritutako formula bat erabiltzen da.

$$F = (2 * Precision * Recall) / (Precision + Recall)$$

- **Confusion matrix:** Datuak klase bakoitzeko ondo sailkatu diren edo ez ikus daiteke. Gainera non sailkatu dituen eta zenbat ere agertzen da 3.2 irudiko adibidean bezala.

```

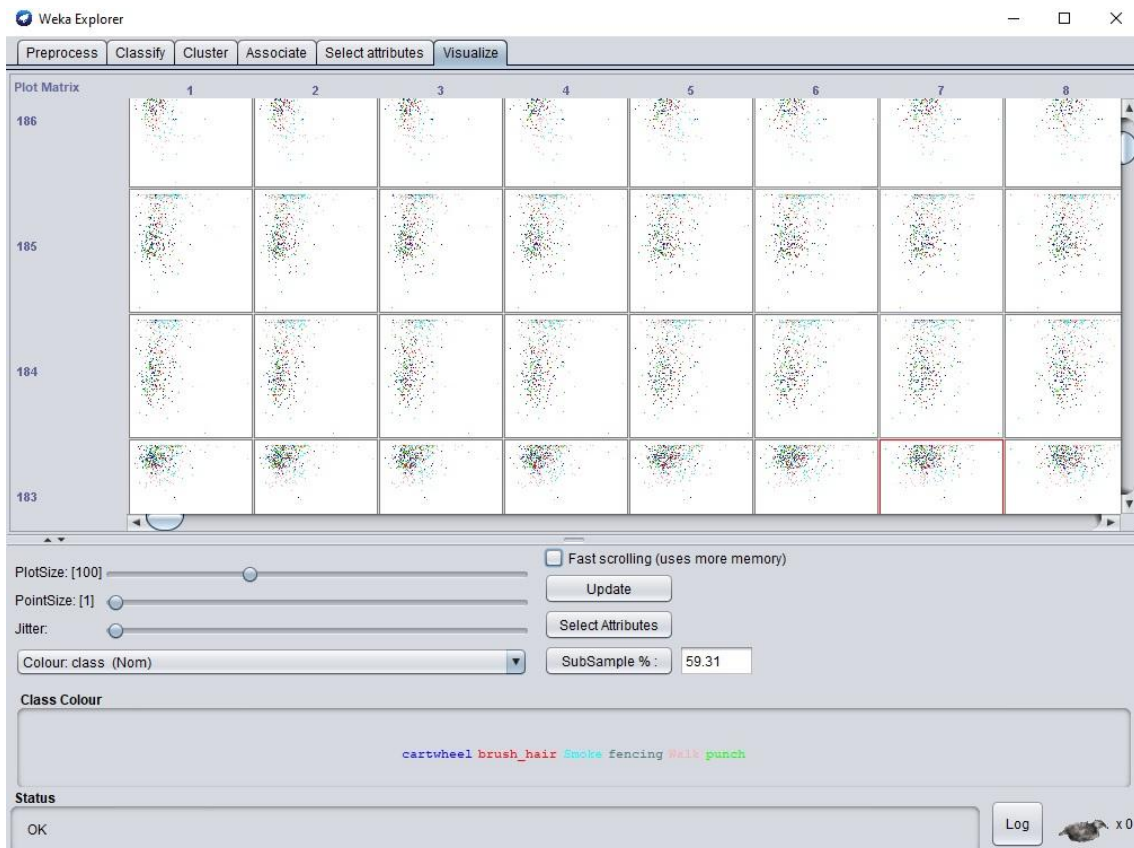
a   b   c   d   e   f   <-- classified as
35  16  0   26  2   24 |   a = cartwheel
12  55  0   14  5   21 |   b = brush_hair
0   2  69   2   34  2 |   c = Smoke
18  15  1   57  4   21 |   d = fencing
8   17  47   9  191  10 |   e = Walk
8   20  2   19  9   68 |   f = punch

```

3.2 irudia: Confusion matrix baten adibidea.

3.4.3 Datuen ikuskaritza

Behin fitxategia kargatuta, datuak ikusteko *visualize* leihora jo beharko da. 3.3 irudian ikus daitekeen bezala atributu guztien datuak ikusi ditzakegu, gainera horietako edozeinetan klikatuz gero xehetasun handiagoarekin ikusiko dira erlazioak.



3.3 irudia: Datuen ikuskaritza.

3.5. Sailkatzaileak

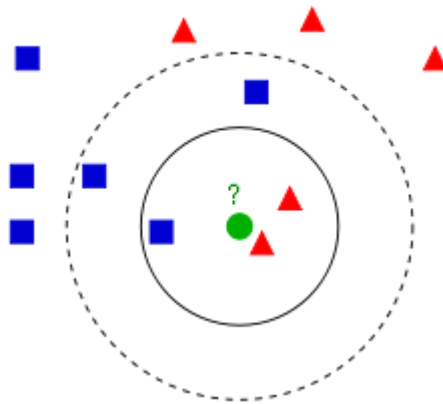
Proiektu honetan datuen analisia egiteko Weka softwareak eskaintzen dituen hainbat algoritmo erabili dira. Jarraian, proiektuan erabili diren sailkatzaileak azalduko dira:

3.5.1 K-NN

K-NN metodoa k auzokide hurbilenetan oinarritzen da kasu berrien sailkapena egiteko. Hurbileneko k auzokideen balioa aztertu eta sarrien agertzen den klasea esleitzen zaio kasu berriari.

Datu-baseko N kasu hartuko dira. Kasu horietako bakoitzak n aldagai iragarle bidez X_1, \dots, X_n eta iragarri beharreko aldagaiaz C klaseko aldagaiaz karakterizaturik dago.

Sailkatze fasean X bektorearen bidez deskribatutako kasu berri bat iristen denean, datu-baseko X_i kasu guztietara duen distantzia kalkulatu behar da, normalean distantzia euklidearra. Ondoren kasuak distantzia txikienetik handienara ordenatu eta erabiltzaileak definitutako k balioaren arabera hurbilen dauden k kasuak hartuko dira. Auzokide hurbilenak diren k kasu horien Y_i klasea begiratu eta gehien agertzen den klasea esleituko zaio X kasu berriari.



3.4 irudia: K-NN adibidea $k=3$ izanik.

3.4 irudian agertzen den bezala puntu berdea ze klasetikoa den jakiteko, gertuko bizilagunen klaseak begiratuko dira. Kasu honetan $k=3$ denez, 3 gertuko bizilagunen klaseak begiratu dira eta horietatik 2 triangeluak dira. Beraz puntu berdea triangelua dela esango da.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *lazy* motako sailkatzaileen barruan IBk aukera dago. IBk-ren kasuan k parametroa definituko da nahi den balioarekin, normalean zenbaki bakoitia izango da, horrela ezinezkoa baita berdinketa egotea. Ondoren, algoritmoak k auzokide erabiliko ditu sailkapena egiteko.

3.5.2 C.45

Zuhaitzen motako sailkatzailea da, non zuhaitza sortzeko informazio handieneko aldagaiaren aukeraketan oinarritzen da.

Zuhaitza sortzeko pausoak ondorengoak dira:

1. Oinarrizko kasuak egiaztatu.
2. A atributu bakoitzerako, A atalaren informazio normalizatuaren irabazia aurkitu.
3. Informazio normalizatu handien duen atributua A_{best} izan dadin utzi.
4. A_{best} banatzen duen erabaki-nodo bat sortu.
5. Errepikatu A_{best} banaketaren bidez lortutako azpelistetan, eta gehitu nodo horiek nodoaren seme-alaba gisa.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *trees* motako sailkatzaileen barruan J48 aukeratuko da.

3.5.3 LMT

LMT (Logistic Model Tree) [4] eredu logistikako zuhaitzak sortzen dituen sailkatzaile bat da. Honek erregresio logistikoa eta erabaki-zuhaitzen ikaskuntza konbinatzen ditu.

Zuhaitz logistikokoak zuhaitz eredu baten ideian oinarritzen dira. Erabaki-zuhaitz batek erregresio linealeko ereduak ditu hostoetan. LMT algoritmoak zuhaitzeko nodo bakoitzean erregresio logistikako eredu bat egingo du, eta ondoren, hau banatzeko C4.5. irizpidea erabiliko da. Azkenik, zuhaitza inausi egiten da.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *trees* motako sailkatzaileen barruan LMT aukeratuko da.

3.5.4 Random Forest

Random forest [5] zuhaitzak sortzen dituen sailkatzaile bat da. Bagging-aren moldaketa bat da, korrelaziorik gabeko zuhaitzen bilduma eraikitzen duena, eta ondoren, batezbestekoak kalkulatzeko dituen.

Baggingaren funtsezko ideia eredu zaratatsu baina gutxi gorabehera inpartzial askoren batezbestekoa egitea da, eta, beraz, aldakuntza murriztea. Zuhaitzak nabarmen zaratatsuak direnez, batezbestekoak egiteak mesede handia egiten die.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *trees* motako sailkatzaileen barruan RandomForest aukeratuko da.

3.5.5 Naive Bayes

Naive Bayes [6] sailkatzailea Bayesen teoreman eta aldagaien arteko independentziaren hipotesian oinarritzen da.

Naive Bayes sailkatzailean suposatzen da ezaugarriak elkarren artean independenteak direla klasearen balioa ezaguna denean. Adibidez fruta bat marrubia dela esan genezake gorria bada, kono formakoa eta 5cm ingurukoa. Ezaugarri horietako bakoitzak modu independentean eragiten du marrubia izateko probabilitatean.

Sailkatzailea eralkitzeko eredu probabilistikoa erabakitze-erregela batekin konbinatzen da. Normalean probabilitate handieneko hipotesia aukeratzen da. Horren arabera, Naive Bayes sailkatzaileak $y = C_k$ klasea esleituko dio C klase-aldagaiari, non (3.3 irudia)

$$\hat{y} = \underset{k \in \{1, \dots, K\}}{\operatorname{argmax}} p(C_k) \prod_{i=1}^n p(x_i | C_k).$$

3.5 irudia: Naive Bayes eredu probabilistikoa.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *Bayes* motako sailkatzaileen barruan NaiveBayes aukeratuko da.

3.5.6 SMO

SMO algoritmoa iteratibo bat da. Algoritmo honek problemak azpi problema txikiagoetan banatzen ditu eta ondoren analitikoki konpondu.

Funtzio koadratikoaren minimoa aurkituko da problema analitikoki ebazteko. K emaitzaren balioa negatiboa izango da eta iterazio bakoitzean ebartziko da.

Algoritmoa honela ebazten da:

1. Lagrangen biderkatzailea aurkitu (α_1).
2. Bigarren biderkatzailea hartu eta lehenengoarekin batera optimizatu (α_1, α_2).
3. Konbergitu arte prozesua errepikatu.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *functions* motako sailkatzaileen barruan SMO aukeratuko da.

3.6. Multi-sailkatzaileak

3.6.1 Bagging

Bagging [7] algoritmoen egonkortasuna eta zehaztasuna hobetzeko diseinatua den meta algoritmo bat da.

Algoritmo honek bariantza murrizten du eta gehiegizko doikuntza saihesten laguntzen du. Normalean erabaki-zuhaitzetako metodoei aplikatzen zaien arren, edozein metodo motarekin erabil daiteke.

Wekan datu-basea aztertu nahi bada sailkatzaile hau erabiliz, *meta* motako sailkatzaileen barruan Bagging aukeratuko da. Ondoren bertan, zein sailkatzaileekin erabiliko den aukeratu beharko da.

3.6.2 Bozketa

Sailkapen hau egiteko programa bat garatu da Google Colaboratory plataforman. Honetarako beste sailkapen batzuen emaitzak gordetzen dira .csv fitxategi batean, non bertan azaltzen den irudi bakoitzerako egin den iragarpena eta benetan ze klasekoa den.

Fitxategiak lortutakoan, sailkatzaile bakoitzak irudi bakoitzari egin dion iragarpena begiratzen da, eta egindako iragarpenen klaseen batura egiten da. Beraz, bozketaren emaitza irudi horretarako bozka gehien dituen klasea izango da.

Adibidez:

- 1.sailkatzaileko emaitzen 1. Irudia → Klasea1
- 2.sailkatzaileko emaitzen 1. Irudia → Klasea2
- 3.sailkatzaileko emaitzen 1. Irudia → Klasea2

Adibidean Klasea2-k bi boto dituen eta Klasea1-ek boto bakarra, emaitza Klasea2 izango da.

3.7. Aldagai aukeraketa

Aldagai aukeraketaren helburua aldagai iragarle esanguratsuenak identifikatzea da. Honekin sailkatzailea hobetuko da eta orokortzeko erraztasun handiagoa izango da.

Aldagai aukeraketarako metodoak erabiltzeko *filter* ataleko *supervised* metodoen barruan *attribute* karpetan sartu eta bertan *AttributeSelection* aukeratuko da. Azken hau aukeratu

ondoren, ebaluatzaile desberdinak erabiltzeko aukera dago. Proiektu honetan ebaluatzaile bakoitza parametro desberdinekin definitu da ondoren emaitzak aztertzeko.

Proiektu honetan erabili diren ebaluatzaileak ondorengoak izan dira:

- **CfsSubsetEval:** Atributuen azpimultzo baten balioa ebaluatzen du, ezaugarri bakoitzaren banakako iragarpen gaitasuna eta haien arteko erredundantzia-maila kontuan hartuta. Nahiago dira klasearekin lotura handia duten ezaugarrien azpimultzoak.
- **PrincipalComponents:** Osagai nagusiak eta datuen eraldaketa aztertzen ditu, Ranker search batekin batera erabiltzen da. Dimentsionaltasunaren murrizketa lortzeko bektore propio batzuk aukeratzen dira (%95). Atributuen zarata iragazteko, transformazio batzuk egiten dira bektore txarrenak kenduz.
- **ReliefAttributeEval:** Atributu baten balioa ebaluatzen du, instantzia bateko lagin errepikatuak hartuz eta klase bereko eta desberdineko instantzia hurbilenerako emandako atributuaren balioa kontuan hartuz. Datu mota diskretuekin nahiz jarraituekin funtziona dezake.
- **SVMAttributeEval:** Atributu baten balioa ebaluatzen du SVM sailkatzaile bat erabiliz. Atributuak SVMk esleitutako pisuaren karratuaren arabera sailkatzen dira.

4



Proiektuko datu-basea

4.1. Sarrera

Proiektu honetan erabili den datu-basea bideon ekintzen Point Cloudak izan dira. Point Cloud hauek lortzeko, bideo bakoitzaren fotogramari *edge* iragazki bat pasatu zaio. Modu honetan fotograma bakoitzeko irudi horren ertzak identifikatzen duen irudi txuri-beltz bat lortzen da. Irudi hauen bilduma izango da ekintzen Point Cloud-a eta ertza kontsideratzen den pixel bakoitza XYZ koordinatuarekin identifikatuta egongo da.

Proiektu honetan azertu diren bideoak iraupen laburrekoak dira. Bideotako ekintzen Point Cloudak erabili dira lanerako. Bideotan adierazten diren ekintzak sei klaseetan banatuta daude eta ondorengoak dira: Orrazten, esgriman, erretzen, oinez, ukabilkada ematen, gurpilarena egiten.

Jarraian, erabili diren klase ezberdinetako bideoen irudiak ikusi daitezke:



4.1 irudia: Gurpilarena egiten.



4.2 irudia: Erretzen.



4.3 irudia: Esgriman.



4.4 irudia: Oinez.



4.5 irudia: Orrazten.



4.6 irudia: Ukabilkada ematen.

4.2. Ezaugarriak

Esan bezala Point Cloud hauek bideoa irudikatzen duten edge puntuak dira, non hauek, XYZ koordenadekin identifikatzen diren. Beraz, gure proiektuaren datu-basea ardatzen bidez adierazten diren puntuak izango dira 4.7 irudiko adibidean bezala.

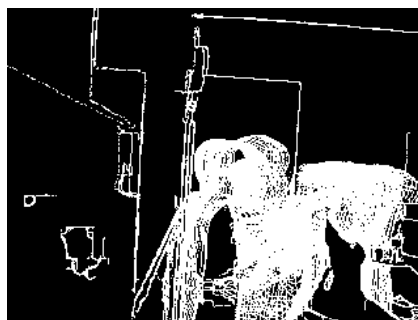
```
"X" "Y" "Z"  
1 1 1  
4 1 1  
5 1 1  
12 239 1  
13 239 1  
18 1 1  
18 239 1  
21 239 1  
24 1 1  
34 1 1  
35 1 1  
35 239 1  
37 1 1  
39 239 1
```

4.7 irudia: Bideoen ekintzen Point Clouden zati bat.

Puntu horiekin, lehenengo prozesamendua egin ondoren emaitza batzuk lortu dira. Ondoren, bigarren prozesamendua egiteko datu-base berri bat lortu da. Datu base hori Point Clouden proiektzioen irudiak dira. Irudi hauek sortzeko zoriz aukeratutako Point Clouden %20a erabili da eta 3 perspektibatatik daude lortuta: XY, XZ eta YZ.

Bi irudi mota lortu dira:

- **BW irudiak (4.8 irudia):** Ardatz bikote bakoitza aztertzen da eta pixel bat zuria badago nahikoa da perspektiban pixel hori zuria agertzeko. Adibidez XY perspektiban edozein z baliorako x,y balio konkretu batetan pixel bat zuria bada (x,y) koordinatu horretan pixelak 1 balio edukiko du.



4.8 irudia: BW irudi baten adibidea (orrazten).

- **Gray irudiak (4.9 irudia):** Ardatz batean dauden pixel txuri guztiak kontatzen dira koordinatu bikote bakoitzerako. Hau da, XY perspektibarako (x,y) konkretu batentzat z ardatzean dauden pixel txuri guztiak kontatzen dira. Beraz XY perspektibako irudiak z ardatzean dauden pixel txuri kopurua adieraziko luke. Baina irudi bat denez balioak [0,255] artean normalizatzen dira.



4.9 irudia: Gray irudi baten adibidea (orrazten).

4.8 eta 4.9 irudiak berdinak dira, perspektiba berdinekoak, baina proiektzioa lortzeko prozesua desberdina da.

5



Garapena

5.1. Ingurunea

Proiektu hau garatzeko bi sistema eragile erabili dira: Linux eta Windows. Linux sistema eragilea Windows sistematik exekutatu da VMware makina birtualaren bitartez.

Linux bidez Point Clouden garbiketa eta zorizko programak aplikatu dira, jarraian puntu berri hauek fitxategi berrietan gorde eta beraien deskriptore globalen .arff-ak lortu dira. Puntu berriekin ondoren, proiektzioak lortu dira, irudi horien banaketa egin da eta .arff-ak lortu dira.

Windows bidez Weka programa exekutatu da .arff fitxategien datuen analisia egiteko.

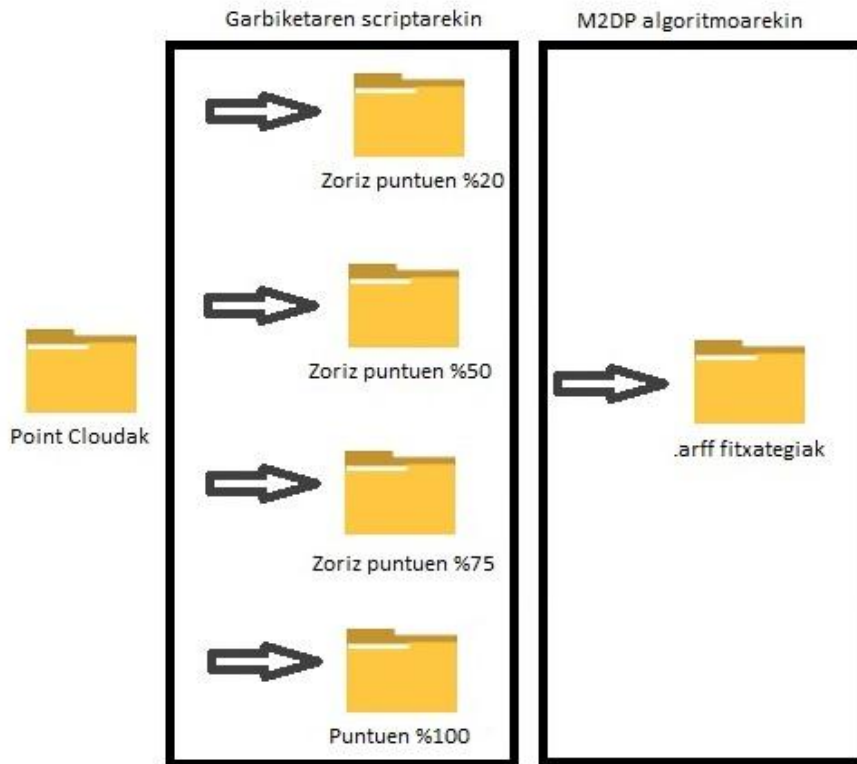
Ondorengo orrietan Point Cloudetatik abiatuta Wekan sailkapenak egin arteko prozesu ezberdinak azalduko dira.

5.2. Point Cloudetatik deskriptore globalen .arff-etara

Proiektu honen lehenengo prozesua Point Clouden garbiketa egitea eta zoriz horietako batzuk aukeratzea izan da (5.1 irudia). Script honetan ondorengo pausuak eman dira:

1. Bideoetako ekintza desberdin bakoitzarentzat karpeta berri bat sortu da. Adibidez, smoke puntuentzako smokeGarbia izeneko karpeta sortu da.
2. Behin karpeta sortuta Errpikapenak.c izeneko programa aplikatu zaie karpetako Point Clouden fitxategi bakoitzari.
3. Errepikatutako puntuak kenduta, ZorizAukeratu.c izeneko programa aplikatu zaie aurreko irteeran lortutako emaitzei. Emaitza berri hauek hasieran sortutako karpeta berrian gordeko dira. ZorizAukeratu.c programan ehuneko desberdinekin aukeratu dira zoriz puntuak: %20, %50 eta %75.
4. Karpeta berrian puntuen garbiketa guztiak edukita, M2DP-a algoritmoa erabili da Point Cloud-aren deskriptore globalak lortzeko eta jarraian .arff fitxategiak lortzeko.

.arff fitxategia sortuta Weka softwarea erabil daiteke datuen analisia egiteko bertako algoritmo ezberdinak probatuz.



5.1 irudia: Point Cloudetatik deskriptore globalen .arff-etara.

5.3. Irudietatik .arff-etara

Point Clouden proiektzioak edukita (zoriz %20 aukeratutako puntuak) bi script erabili dira proiektuko bigarren prozesua garatzeko 5.2 irudian agertzen den bezala. Lehenengoa, proiektzioak ardatzaren arabera banatzeko izan da, irudi desberdinak direlako eta desberdin tratatzeko. Aldiz bigarren scripta irudien .arff fitxategiak lortzeko erabili da. Jarraian lehenengo scriptaren pausoak:

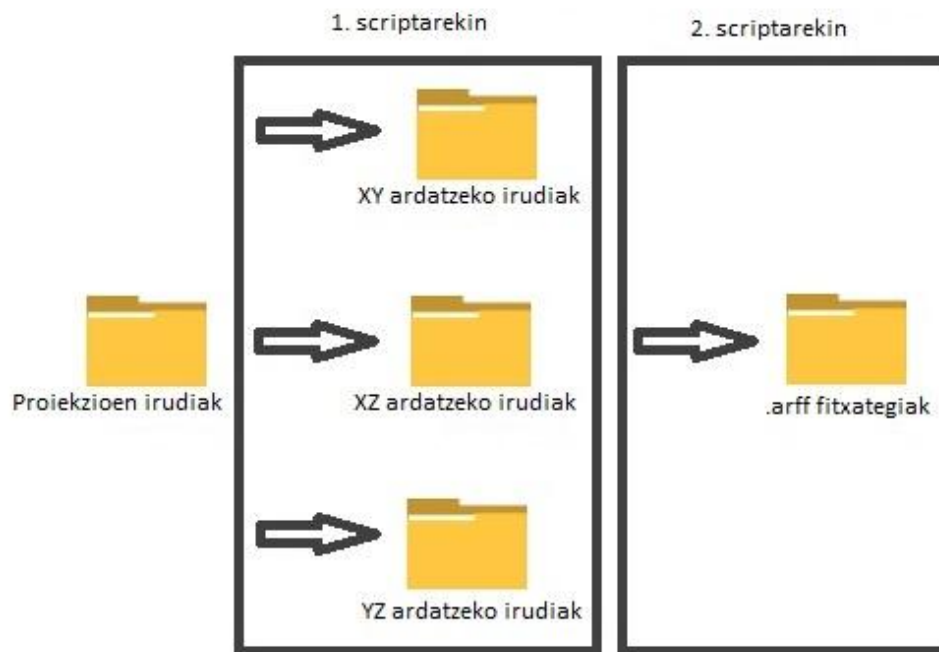
1. Hasteko irudi mota eta ardatz bakoitzeko karpeta berri bat sortuko da. Adibidez Gray motako eta XZ ardatzeko irudientzat XZ_Gray izeneko karpeta bat sortuko da.
2. Karpetak sortuta, irudi guztiak karpeta berrietan kopiatuko dira baina mota eta ardatzaren arabera. Adibidez Gray motako eta XZ ardatzeko proiektzioa hasieran sortu dugun karpetan kopiatuko da, XZ_Gray karpetan.

Behin irudiak motaren eta ardatzaren arabera banatuta edukitakoan bigarren scripta erabili da. Honek jarraitutako pausoak ondorengoan izan dira:

1. Karpeta berri bat sortuko da irteera izenekoa.

2. Behin karpeta berria sortuta, karpetan bertan ordenatuak izeneko direktorioan argazkiak gordeko dira, izen bezala klasea eta zenbakia jarrita. Adibidez k1 klaseko lehenengo argazkia bada, k1-1.png bezala gordeko da ordenatuak karpetan.
3. Irteera izenako karpetan .arff fitxategia sortuko da eta bertan .arff fitxategi batek bete behar dituen baldintzak idatziko dira ondoren datuak aztertzeke.

.arff fitxategia sortuta Weka softwarea erabil daiteke datuen analisia egiteko bertako algoritmo ezberdinak probatuz.



5.2 irudia: Irudietatik .arff-etara.

6



Emaitzak

6.1. Sarrera

Point Cloude-tatik .arff fitxategiak sortu ondoren, Weka softwarea erabili da algoritmo desberdinekin datuen analisia egiteko.

Sailkapena bi esperimentuetan banatzen da, lehenengoa deskriptore globalekin egindako sailkapenak eta jarraian proiektzioekin egindakoa. Kasu desberdinak direnez, sailkatzaile desberdinekin eta esperimentu desberdinekin probatu dira. Ondoren, lortutako emaitzak tauletan idatzi dira. Amaitzeko, taula bakoitzetik eta emaitzak konparatuz ondorioak atera dira eta hauek azaldu.

6.2. Zorizko garbiketaren emaitzak

Prozesu honetan errepikatutako puntuak kendu eta zoriz aukeratutako puntuak erabili dira. Zoriz aukeratutako puntu horien deskriptore globaletatik lortu dira .arff fitxategiak.

Zoriz aukeratu diren puntuak %20, %50 eta %75 izan dira, horrez gain zorizko puntuen aukeraketa egin gabeko emaitzak ere lortu dira. Beraz, atal honetan 4 kasu aztertuko ditugu sailkatzaile eta aldagai aukeraketa desberdinekin: puntuen %20a aukeratzen duena, %50, %75 eta %100.

6.2.1 Zoriz aukeratutako puntuen %20

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote = Vote, sailkatzaile bezala IBk3, Random Forest eta Naive Bayes erabiliz.
- Rand = Random Forest.

6.2.1.1 Filter gabe

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
40.93	53.02	56.46	52.31	54.69	58.24	48.16	43.77	51.72	58.48

6.1 taula: Puntuen %20 filter gabe, asmatze tasak.

6.1 taulan agertzen den bezala, lortutako emaitzak ez dira oso onak. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da, %58.48 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %40.93 sailkatu ditu ondo.

IBk1 algoritmoa (%56.46) eta Bagging algoritmoa Random Forest erabiliz (%58.25), emaitza onenarengatik hurbil geratu dira. Emaitza txarrenarengatik hurbilen J48 geratu da, %43.77ko asmatze tasarekin.

Emaitza guztien batezbestekoa %51.78 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%40.93), Vote (%48.16), J48 (%43.77) eta LMT (%51.72) izan dira.

6.2.1.2 Filter CfsSubsetEval

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
47.81	50.89	53.38	52.55	53.02	58.96	50.42	48.40	50.77	60.38

6.2 taula: Puntuen %20 filter CfsSubsetEval, asmatze tasak.

6.2 taulan agertzen den bezala, lortutako emaitzak hobetu egin dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da, %60.38 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %47.81 sailkatu ditu ondo.

Bagging algoritmoa Random Forest erabiliz (%60.38), emaitza onenarengatik hurbil geratu da. Emaitza txarrenarengatik hurbilen J48 geratu da, %48.40ko asmatze tasarekin.

Emaitza guztien batezbestekoa %52.66 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%47.81), SMO (%50.89), IBk3 (%52.55), Vote (%50.42), J48 (%48.40) eta LMT (%50.77) izan dira.

6.2.1.3 Filter PrincipalComponents

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
47.57	50.30	56.82	53.86	52.08	59.43	54.80	41.52	51.36	58.72

6.3 taula: Puntuen %20 filter PrincipalComponents, asmatze tasak.

6.3 taulan agertzen den bezala, lortutako emaitzak 6.1 taularen antzekoak dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest sailkatzailea erabiliz, %59.43

sailkatu ditu ondo. Sailkatzaile txarrena berriz, J48 sailkatzailea izan da, %41.52 sailkatu ditu ondo.

IBk1 algoritmoa (%56.82) eta Random Forest (%58.72), emaitza onenarengatik hurbil geratu dira. Emaitza txarrenarengatik hurbilen BayesNet geratu da, %47.57ko asmatze tasarekin.

Emaitza guztien batezbestekoa %52.65 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%47.57), SMO (%50.30), IBk5 (%52.08), J48 (%41.52) eta LMT (%51.36) izan dira.

6.2.1.4 Filter ReliefAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	43.06	52.08	58.60	55.52	53.74	57.77	49.23	46.50	49.70	59.07
125	43.18	51.96	57.89	56.58	55.87	58.60	48.99	46.74	50.53	58.24
Guztiak	40.93	52.91	56.46	52.31	54.69	58.24	48.16	44.48	51.72	59.67

6.4 taula: Puntuen %20 filter ReliefAttributeEval, asmatze tasak.

6.4 taulan agertzen den bezala, lortutako emaitzak 6.2 taularen antzekoak dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da aldagai guztaik erabiliz, %59.67 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %40.93 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.53, 125 aldagaiekin %52.86 eta aldagai guztiak %51.96. Beraz, emaitza hoberenak 125 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

6.2.1.5 Filter SVMAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	44.25	53.62	55.63	55.87	55.52	58.96	49.70	46.03	51.13	59.07
125	41.52	53.38	57.53	53.14	53.50	58.24	48.28	44.84	52.55	59.07
Guztiak	40.93	53.02	56.46	52.31	54.69	58.84	47.45	43.65	51.72	59.19

6.5 taula: Puntuen %20 filter SVMAttributeEval, asmatze tasak.

6.5 taulan agertzen den bezala, lortutako emaitzak orain arteko onenak dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da aldagai guztiak erabiliz, %59.19 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %40.93 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.98, 125 aldagaiekin %52.21 eta aldagai guztiekin %51.83. Beraz , emaitza hoberenak 75 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

Zorizko %20 puntuak erabiliz orokorrean lortu diren emaitzak %52 inguruan dabilta. Emaitza guztiak kontuan hartuz sailkatzaile onena Random Forest dela garbia da, txarrena aldiz BayesNet. Emaitza onena filter ReliefAttributeEval aldagai guztiak erabiliz eta Random Forest sailkatzailearekin lortu da, %60.38 sailkatu ditu ondo. Aldiz txarrena filter gabeko BayesNet sailkatzailearekin lortu da, %40.93 sailkatu ditu ondo.

Emaitza guztiak kontuan hartuz batezbesteko hobereana filter SVMAttributeEval 75 aldagaiekin erabiliz lortu da, batezbeste %52.98 sailkatu ditu ondo. Aldiz, txarrena filter gabe lortu da, batezbeste %51.78 sailkatu ditu ondo.

Amaitzeko, emaitza guztien batezbestekoa kalkulatu da. Zoriz %20 puntuak erabilita emaitzen batezbestekoa %52.38 da.

6.2.2 Zoriz aukeratutako puntuen %50

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote = Vote, sailkatzaile bezala IBk3, Random Forest eta Naive Bayes erabiliz.
- Rand = Random Forest.

6.2.2.1 Filter gabe

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
40.57	52.08	56.46	53.14	54.92	59.19	47.81	41.28	52.79	58.36

6.6 taula: Puntuen %50 filter gabe, asmatze tasak.

6.6 taulan agertzen den bezala, lortutako emaitzak ez dira oso onak. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest erabiliz sailkatzaile bezala, %59.19 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %40.57 sailkatu ditu ondo.

IBk1 algoritmoa (%56.46) eta Random Forest algoritmoa (%58.36), emaitza onenarengatik hurbil geratu dira. Emaitza txarrenarengatik hurbilen J48 geratu da, %41.28ko asmatze tasarekin.

Emaitza guztien batezbestekoa %51.66 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%40.57), Vote (%47.81) eta J48 (%41.28) izan dira.

6.2.2.2 Filter CfsSubsetEval

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
45.79	50.89	55.28	52.43	52.67	59.91	49.23	47.81	50.18	59.43

6.7 taula: Puntuen %50 filter CfsSubsetEval, asmatze tasak.

6.7 taulan agertzen den bezala, lortutako emaitzak hobetu egin dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest erabiliz sailkatzaile bezala, %59.91 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %45.79 sailkatu ditu ondo.

Random Forest algoritmoa (%59.43), emaitza onenarengatik hurbil geratu da. Emaitza txarrenarengatik hurbilen J48 geratu da, %47.81eko asmatze tasarekin.

Emaitza guztien batezbestekoa %52.36 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%45.79), SMO (%50.89), Vote (%49.23), J48 (%47.81) eta LMT (%50.18) izan dira.

6.2.2.3 Filter PrincipalComponents

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
46.50	51.01	57.18	53.26	51.72	58.96	54.69	45.20	50.65	57.77

6.8 taula: Puntuen %50 filter PrincipalComponents, asmatze tasak.

6.8 taulan agertzen den bezala, lortutako emaitzak berriz ere hobetu dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest sailkatzailea erabiliz, %58.96 sailkatu ditu ondo. Sailkatzaile txarrena berriz, J48 sailkatzailea izan da, %45.20 sailkatu ditu ondo.

IBk1 algoritmoa (%57.18) eta Random Forest (%57.77), emaitza onenarengatik hurbil geratu dira. Emaitza txarrenarengatik hurbilen BayesNet geratu da, %46.50ko asmatze tasarekin.

Emaitza guztien batezbestekoa %52.69 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%46.50), SMO (%51.01), IBk5 (%51.72), J48 (%45.20) eta LMT (%50.65) izan dira.

6.2.2.4 Filter ReliefAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	44.25	50.89	58.48	55.04	52.55	59.07	49.94	43.65	50.53	59.67
125	41.87	52.19	57.89	55.87	54.57	59.31	49.23	44.72	50.89	59.19
Guztiak	40.57	52.31	56.46	53.14	54.92	58.84	47.92	42.94	52.79	59.43

6.9 taula: Puntuen %50 filter ReliefAttributeEval, asmatze tasak.

6.9 taulan agertzen den bezala, lortutako emaitzak 6.7 taularen antzekoak dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da 75 aldagai erabiliz, %59.67 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %40.57 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.41, 125 aldagaiekin %52.57 eta aldagai guztiak %51.93. Beraz ,emaitza hoberenak 125 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

6.2.2.5 Filter SVMAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	43.42	52.19	56.35	53.74	53.62	58.48	50.18	44.48	53.14	59.91
125	40.81	51.84	56.46	53.86	52.19	58.01	48.28	45.55	52.19	58.24
Guztiak	40.57	52.55	56.46	53.14	54.92	59.19	47.45	42.23	52.79	59.55

6.10 taula: Puntuen %50 filter SVMAttributeEval, asmatze tasak.

6.10 taulan agertzen den bezala, lortutako emaitzak 6.9 taularekin konparatuz okertu egin dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da 75 aldagai erabiliz, %59.91 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %40.57 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.55, 125 aldagaiekin %51.74 eta aldagai guztiak %51.89. Beraz , emaitza hoberenak 75 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

Zorizko %50 puntuak erabiliz orokorrean lortu diren emaitzak %52 inguruan dabilta. Emaitza guztiak kontuan hartuz sailkatzaile onena Random Forest dela garbia da, txarrena aldiz BayesNet. Emaitza onena filter SVMAttributeEval 75 aldagai erabiliz eta Random Forest sailkatzailearekin lortu da, %59.91 sailkatu ditu ondo. Aldiz txarrena filter gabeko BayesNet sailkatzailearekin lortu da, %40.57 sailkatu ditu ondo.

Emaitza guztiak kontuan hartuz batezbesteko hobereana filter PrincipalComponents erabiliz lortu da, batezbeste %52.69 sailkatu ditu ondo. Aldiz, txarrena filter gabe lortu da, batezbeste %51.66 sailkatu ditu ondo.

Amaitzeko, emaitza guztien batezbestekoa kalkulatu da. Zoriz %50 puntuak erabilita emaitzen batezbestekoa %52.20 da.

6.2.3 Zoriz aukeratutako puntuen %75

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote = Vote, sailkatzaile bezala IBk3, Random Forest eta Naive Bayes erabiliz.
- Rand = Random Forest.

6.2.3.1 Filter gabe

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
41.04	52.91	56.58	53.02	54.57	58.71	47.45	44.84	51.36	58.36

6.11 taula: Puntuen %75 filter gabe, asmatze tasak.

6.11 taulan agertzen den bezala, lortutako emaitzak ez dira oso onak. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest erabiliz sailkatzaile bezala, %58.71 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %41.04 sailkatu ditu ondo.

IBk1 algoritmoa (%56.58) eta Random Forest algoritmoa (%58.36), emaitza onenarengatik hurbil geratu dira. Emaitza txarrenarengatik hurbilen J48 geratu da, %44.84ko asmatze tasarekin.

Emaizta guztien batezbestekoa %51.89 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%41.04), Vote (%47.45), J48 (%44.84) eta LMT (51.36) izan dira.

6.2.3.2 Filter CfsSubsetEval

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
45.91	50.06	54.57	52.67	52.67	59.19	51.36	44.48	50.53	59.79

6.12 taula: Puntuen %75 filter CfsSubsetEval, asmatze tasak.

6.12 taulan agertzen den bezala, lortutako emaitzak hobetu egin dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da, %59.78 sailkatu ditu ondo. Sailkatzaile txarrena berriz, J48 sailkatzailea izan da, %44.48 sailkatu ditu ondo.

Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz (%59.19), emaitza onenarengatik hurbil geratu da. Emaizta txarrenarengatik hurbilen BayesNet geratu da, %45.91eko asmatze tasarekin.

Emaizta guztien batezbestekoa %52.12 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%45.91), SMO (%50.06), Vote (%51.36), J48 (%44.48) eta LMT (%50.53) izan dira.

6.2.3.3 Filter PrincipalComponents

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
47.33	50.89	57.30	53.14	51.48	59.19	55.40	44.13	51.48	59.07

6.13 taula: Puntuen %75 filter PrincipalComponents, asmatze tasak.

6.13 taulan agertzen den bezala, lortutako emaitzak berriz ere hobetu dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest sailkatzailea erabiliz, %59.19 sailkatu ditu ondo. Sailkatzaile txarrena berriz, J48 sailkatzailea izan da, %44.13 sailkatu ditu ondo.

IBk1 algoritmoa (%57.30) eta Random Forest (%59.07), emaitza onenarengatik hurbil geratu dira. Emaizta txarrenarengatik hurbilen BayesNet geratu da, %47.33ko asmatze tasarekin.

Emitza guztien batezbestekoa %52.94 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%47.33), SMO (%50.89), IBk5 (%51.48), J48 (%44.13) eta LMT (%51.48) izan dira.

6.2.3.4 Filter ReliefAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	43.53	51.60	58.13	54.33	53.14	58.36	49.23	44.37	51.48	59.67
125	42.70	52.5	57.41	55.28	54.80	59.43	48.64	44.37	50.42	59.55
Guztiak	41.04	52.79	56.58	53.02	54.57	59.43	47.33	44.48	51.36	58.01

6.14 taula: Puntuen %75 filter ReliefAttributeEval, asmatze tasak.

6.14 taulan agertzen den bezala, lortutako emaitzak 6.12 taularen antzekoak dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da 75 aldagai erabiliz, %59.67 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %41.04 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.38, 125 aldagaiekin %52.51 eta aldagai guztiekin %51.86. Beraz ,emaitza hoberenak 125 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

6.2.3.5 Filter SVMAttributeEval

6.15 taulan agertzen den bezala, lortutako emaitzak 6.14 taularekin konparatuz okertu egin dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da sailkatzaile bezala Random Forest erabiliz eta aldagai guztiekin, %59.55 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %41.04 sailkatu ditu ondo.

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	43.89	52.55	56.82	54.21	53.50	58.24	50.18	43.42	52.79	59.43
125	41.40	52.43	57.53	53.50	52.91	58.36	48.75	45.20	51.36	58.48
Guztiak	41.04	52.79	56.58	53.02	54.57	59.55	47.69	44.84	51.36	58.48

6.15 taula: Puntuen %75 filter SVMAttributeEval, asmatze tasak.

IBk1 algoritmoa eta Random Forest algoritmoa izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %52.50, 125 aldagaiekin %51.99 eta aldagai guztiekin %51.99. Beraz, emaitza hoberenak 75 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

Zorizko %75 puntuak erabiliz orokorrean lortu diren emaitzak %52 inguruan dabilta. Emaitza guztiak kontuan hartuz sailkatzaile onena Random Forest dela garbia da, txarrena aldiz BayesNet. Emaitza onena filter CfsSubsetEval erabiliz eta Random Forest sailkatzailearekin lortu da, %59.79 sailkatu ditu ondo. Aldiz txarrena filter gabeko BayesNet sailkatzailearekin lortu da, %41.04 sailkatu ditu ondo.

Emaitza guztiak kontuan hartuz batezbesteko hobereana filter PrincipalComponents erabiliz lortu da, batezbeste %52.94 sailkatu ditu ondo. Aldiz, txarrena filter ReliefAttributeEval aldagai guztiak erabiliz lortu da, batezbeste %51.86 sailkatu ditu ondo.

Amaitzeko, emaitza guztien batezbestekoa kalkulatu da. Zoriz %75 puntuak erabilia emaitzen batezbestekoa %52.24 da.

6.2.4 Puntu guztiak erabiliz

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.

- Vote = Vote, sailkatzaile bezala IBk3, Random Forest eta Naive Bayes erabiliz.
- Rand = Random Forest.

6.2.4.1 Filter gabe

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
38.79	52.19	56.35	52.55	53.86	60.14	51.36	51.72	55.63	60.02

6.16 taula: Puntuen %100 filter gabe, asmatze tasak.

6.16 taulan agertzen den bezala, lortutako emaitzak orain arteko onenak dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest erabiliz sailkatzaile bezala, %60.14 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %38.79 sailkatu ditu ondo.

Random Forest algoritmoa (%60.02), emaitza onenarengatik hurbil geratu da. Emaitza txarrenarengatik hurbilen Vote geratu da, %51.36ko asmatze tasarekin. %51ko tasatik jaisten den sailkatzaile bakarria BayesNet da, beste sailkatzaile guztiek nahiko emaitza txukunak ematen dituzte.

Emaitza guztien batezbestekoa %53.26 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%38.79), SMO (%52.19), IBk3 (%52.55), Vote (%51.36) eta J48 (%51.72) izan dira.

6.2.4.2 Filter CfsSubsetEval

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
49.11	48.87	54.33	51.72	49.70	61.80	52.67	50.18	50.30	61.09

6.17 taula: Puntuen %100 filter CfsSubsetEval, asmatze tasak.

6.17 taulan agertzen den bezala, lortutako emaitzak 6.16 taulakoak baino okerragoak dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da Random Forest erabiliz sailkatzaile bezala, %61.80 sailkatu ditu ondo. Sailkatzaile txarrena berriz, SMO sailkatzailea izan da, %48.87 sailkatu ditu ondo.

Random Forest algoritmoa (%61.09), emaitza onenarengatik hurbil geratu da. Emaitza txarrenarengatik hurbilen BayesNet geratu da, %49.11eko asmatze tasarekin.

Emaizta guztien batezbestekoa %52.98 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%49.11), SMO (%48.87), IBk3 (%51.72), IBk5 (%49.70), Vote (%52.67), J48 (%50.18) eta LMT (%50.30) izan dira.

6.2.4.3 Filter PrincipalComponents

BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
41.87	50.89	57.06	53.14	50.77	56.23	55.52	47.33	54.33	56.35

6.18 taula: Puntuen %100 filter PrincipalComponents, asmatze tasak.

6.18 taulan agertzen den bezala, lortutako emaitzak berriz ere okertu dira. Asmatze tasa altuena IBk1 algoritmoaren bidez lortu da, %57.06 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da, %41.87 sailkatu ditu ondo.

Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz (%56.23) eta Random Forest (%56.35), emaitza onenarengatik hurbil geratu dira. Emaizta txarrenarengatik hurbilen J48 geratu da, %47.33ko asmatze tasarekin.

Emaizta guztien batezbestekoa %52.35 izan da. Batezbesteko honetatik bera geratu diren algoritmoak BayesNet (%41.87), SMO (%50.89), IBk5 (%50.77) eta J48 (%47.33) izan dira.

6.2.4.4 Filter ReliefAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	39.50	53.39	57.65	55.75	53.02	60.14	51.48	47.57	53.50	60.02
125	39.03	53.14	58.48	51.96	54.21	60.97	49.70	48.64	55.87	61.21
Guztiak	38.79	52.19	56.35	52.55	53.86	59.31	51.72	50.89	55.99	61.57

6.19 taula: Puntuen %100 filter ReliefAttributeEval, asmatze tasak.

6.19 taulan agertzen den bezala, lortutako emaitzak 6.16 taularen antzekoak dira. Asmatze tasa altuena Random Forest algoritmoaren bidez lortu da aldagai guztiak erabiliz, %61.57 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %38.79 sailkatu ditu ondo.

IBk1 algoritmoa eta Bagging algoritmoa sailkatzaile bezala Random Forest erabiliz izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da nahiz eta BayesNet algoritmoaren emaitzekiko tarte handia izan.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %53.20, 125 aldagaiekin %53.32 eta aldagai guztiekin %53.32. Beraz ,emaitza hoberenak 125 aldagai eta aldagai guztiak erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

6.2.4.5 Filter SVMAttributeEval

Aldagai kopurua	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote	J48	LMT	Rand.
75	43.42	51.72	56.94	53.26	53.26	60.97	51.13	51.01	57.30	60.50
125	40.33	53.50	56.23	54.92	54.69	59.07	52.79	50.18	54.69	60.02
Guztiak	38.79	52.19	56.35	52.55	53.86	59.19	51.48	51.25	55.75	59.79

6.20 taula: Puntuen %100 filter SVMAttributeEval, asmatze tasak.

6.20 taulan agertzen den bezala, lortutako emaitzak orain arteko hoberenak dira. Asmatze tasa altuena Bagging algoritmoaren bidez lortu da sailkatzaile bezala Random Forest eta 75 aldagai erabiliz, %60.97 sailkatu ditu ondo. Sailkatzaile txarrena berriz, BayesNet sailkatzailea izan da aldagai guztiak erabiliz, %38.79 sailkatu ditu ondo.

IBk1 algoritmoa eta Random Forest algoritmoa izan dira emaitza onenarengatik hurbil geratu diren algoritmoak. Emaitza txarrenarengatik hurbilen J48 algoritmoa geratu da nahiz eta BayesNet algoritmoaren emaitzekiko tarte handia izan.

Aldagai kopuru desberdinekin lortu diren batezbestekoak ondorengoak dira: 75 aldagaiekin %53.95, 125 aldagaiekin %53.64 eta aldagai guztiekin %53.12. Beraz , emaitza hoberenak 75 aldagai erabiliz lortu dira.

Aldagai kopuru ezberdin guztiak kontuan hartuz, emaitza hoberenak ematen dituen algoritmoa Random Forest da. Txarrenak aldiz, BayesNet algoritmoak.

Zorizko %100 puntuak erabiliz orokorrean lortu diren emaitzak %53 inguruan dabilta. Emaitza guztiak kontuan hartuz sailkatzaile onena Random Forest dela garbia da, txarrena aldiz

BayesNet. Emaidza onena filter CfsSubsetEval erabiliz eta Bagging sailkatzailearekin lortu da sailkatzaile bezala Random Forest erabiliz, %61.80 sailkatu ditu ondo. Aldiz txarrena filter gabeko BayesNet sailkatzailearekin lortu da, %38.79 sailkatu ditu ondo.

Emaidza guztiak kontuan hartuz batezbesteko hoberena filter SVMAttributeEval eta 75 aldagai erabiliz lortu da, batezbeste %53.95 sailkatu ditu ondo. Aldiz, txarrena filter PrincipalComponents erabiliz lortu da, batezbeste %52.35 sailkatu ditu ondo.

Amaitzeko, emaitza guztien batezbestekoa kalkulatu da. Zoriz %100 puntuak erabilita emaitzen batezbestekoa %53.24 da.

6.2.5 Emaidzen analisia

Lehenengo prozesaketa honetan lortu diren emaitzak %52 inguruan dabilta. Sailkatzaile batzuk emaitza txarrak lortu dituzte, beste batzuk aldiz batezbestekoaren gainetik egon dira. Emaidza onenak Random Forest algoritmoarekin lortu dira, Bagging eta IBk algoritmoekin jarraituz. Emaidza txarrenak alde handiarekin BayesNet sailkatzailearekin lortu dira, eta honen pareko baina emaitza hobekin J48 sailkatzailearekin lortu dira.

Emaidza guztiak kontuan hartuta, emaitza onena eman diguna zorizko puntuen aukeraketa egin gabekoarena izan da. Bertan, filter CfsSubsetEval erabiliz eta Bagging sailkatzailearekin sailkatzaile bezala Random Forest erabiliz lortu da emaitzarik onena, %61.80 sailkatu ditu ondo. Txarrena ere zorizko puntuen aukeraketa egin gabekoarekin lortu da. Kasu honetan filter gabeko BayesNet sailkatzailearekin lortu da, %38.79 sailkatu ditu ondo.

Batezbesteko emaitza onenak zorizko garbiketa egin gabe lortu dira, batezbeste %53.24 sailkatu ditu ondo. Emaidza txarrenak zorizko %50 puntuen aukeraketak eman dizkigu, batezbeste %52.20 sailkatu ditu ondo.

Zorizko garbiketen emaitzen analisisan nabaria da ez dela ezer hobetu, honen arrazoi nagusia Point Cloudak zoriz aukeratzen direlako izan da. Aukeratutako Point Clouden garbiketa irizpide bat jarraituz egingo balira seguraski emaitza hobegoak lortuko liratekeela.

6.3. Proiekzioen emaitzak

Prozesu honetan garbiketan zoriz aukeratu diren puntuen %20ko datu-basea erabili da. Point Cloud horietatik beraien 3D proiekzioak lortu dira eta hauekin egin da lan.

Proiekzioen irudiak perspektiba (XY ardatza, XZ ardatza eta YZ ardatza) eta mota (BW eta Gray) ezberdinekoak dira. Lehenengo irudi hauek banatu dira beraien artean ezberdinak direlako eta bakoitza ezberdin tratatzeko. Ondoren ardatz bakoitzaren eta motaren .arff fitxategiak lortu dira eta Weka softwarean sailkatu dira. Wekan sailkatzako irudi mota bakoitzari 3 filtro aplikatu zaizkie: Edge Histogram, Phog eta Color Layout.

6.3.1 XY ardatzeko proiektzioak

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote1 = Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz.
- Rand = Random Forest.

Vote2 sailkatzailean .arff fitxategi konkretu baterako lortu diren emaitza onenen sailkatzaileak erabiltzen dira. Beraz, atal bakoitzean adierazten da Vote2 zer den.

6.3.1.1 BW motako irudiak

6.21 taulan Vote2 sailkatzailea ondorengo da Color Layout eta Edge Histogram filtroen kasuan:

- Vote2 = Vote, sailkatzaile bezala IBk1, Random Forest eta Bagging (Random Forest) erabiliz.

6.21 taulan Vote2 sailkatzailea ondorengo da Phog filtroaren kasuan:

- Vote2 = Vote, sailkatzaile bezala SMO, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	41.76	41.64	54.57	47.57	49.35	55.16	54.69	54.57	43.18	55.04
Edge Histogram	43.53	51.25	56.46	53.26	53.02	53.62	59.55	56.58	47.69	56.11
Phog	49.23	59.55	59.07	56.70	55.75	60.38	62.75	63.23	54.69	61.92

6.21 taula: XY ardatzeko BW motako irudien asmatze tasak.

6.21 taulako emaitzarik onenak Phog filtroarekin lortu dira. Txarrenak aldiz Color Laout filtroarekin lortu dira.

Emaitzarik onena irudietan Phog filtroa aplikatuz lortu da, Vote2 sailkatzailearekin, hau da, Vote sailkatzaile bezala SMO, Random Forest eta Bagging (Random Forest) erabiliz, %63.23 sailkatu ditu ondo. Emaitzarik txarrena Color Layout filtroarekin lortu da, SMO sailkatzailea erabiliz, %41.64 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %49.75, Edge Histogram filtroarekin %53.11 eta Phog filtroarekin %58.33. Beraz, emaitza onenak ematen dituen filtroa Phog da.

Filtro guztiak kontuan hartuz, hau da, XY ardatzeko eta BW motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 53.73 da.

6.3.1.2 Gray motako irudiak

6.22 taulan Vote2 sailkatzailea ondorengoa da 3 filtroentzat:

- Vote2 = Vote, sailkatzaile bezala IBk1, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	36.30	35.47	44.48	38.79	40.57	55.28	46.62	44.60	39.38	56.35
Edge Histogram	41.40	48.64	62.51	55.75	55.28	59.55	65.12	62.75	49.82	60.14
Phog	45.31	58.72	64.53	56.70	57.30	61.68	68.21	65.60	60.02	62.40

6.22 taula: XY ardatzeko Gray motako irudien asmatze tasak.

6.22 taulako emaitzak 6.21 taulan lortu diren antzekoak dira. Emaitzarik onenak Phog filtroarekin lortu dira. Txarrenak aldiz Color Laout filtroarekin lortu dira.

Emaitzarik onena irudietan Phog filtroa aplikatuz lortu da, Vote1 sailkatzailearekin, hau da, Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz., %68.21 sailkatu ditu ondo.

Emaitzarik txarrena Color Layout filtroarekin lortu da, SMO sailkatzailea erabiliz, %35.47 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %43.78, Edge Histogram filtroarekin %56.10 eta Phog filtroarekin %60.05. Beraz, emaitza onenak ematen dituen filtroa Phog da.

Filtro guztiak kontuan hartuz, hau da, XY ardatzeko eta Gray motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 53.31 da.

Orokorrean Gray motako irudiekin emaitza hobekak lortu dira. Color Layout filtroak emaitza oso txarrak lortu ditu, hori dela eta batezbestekoa asko jaitsi du, baina Phog eta Edge Histogram filtroekin emaitza hobekak lortu dira BW motako irudiekin alderatuz.

6.3.2 XZ ardatzeko proiektzioak

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote1 = Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz.
- Rand = Random Forest.

Vote2 sailkatzailean .arff fitxategi konkretu baterako lortu diren emaitza onenen sailkatzaileak erabiltzen dira. Beraz, atal bakoitzean adierazten da Vote2 zer den.

6.3.2.1 BW motako irudiak

6.23 taulan Vote2 sailkatzailea ondorengoa da Color Layout filtroaren kasuan:

- Vote2 = Vote, sailkatzaile bezala IBk1, Random Forest eta Bagging (Random Forest) erabiliz.

6.23 taulan Vote2 sailkatzailea ondorengoa da Phog eta Edge Histogram filtroen kasuan:

- Vote2 = Vote, sailkatzaile bezala LMT, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	32.03	33.45	39.26	35.47	36.18	45.67	40.33	39.26	30.60	44.96
Edge Histogram	48.87	48.16	45.20	43.90	45.55	52.19	48.04	53.02	51.13	51.48
Phog	42.82	41.28	39.62	38.67	38.32	52.43	44.01	48.64	41.40	50.53

6.23 taula: XZ ardatzeko BW motako irudien asmatze tasak.

6.23 taulako emaitzarik onenak Edge Histogram filtroarekin lortu dira. Txarrenak aldiz Color Layout filtroarekin lortu dira.

Emaitzarik onena irudietan Edge Histogram filtroa aplikatuz lortu da, Vote2 sailkatzailearekin, hau da, Vote sailkatzaile bezala LMT, Random Forest eta Bagging (Random Forest) erabiliz, %53.02 sailkatu ditu ondo. Emaitzarik txarrena Color Layout filtroarekin lortu da, LMT sailkatzailea erabiliz, %30.60 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %37.72, Edge Histogram filtroarekin %48.75 eta Phog filtroarekin %43.77. Beraz, emaitza onenak ematen dituen filtroa Edge Histogram da.

Filtro guztiak kontuan hartuz, hau da, XZ ardatzeko eta BW motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 43.41 da.

6.3.2.2 Gray motako irudiak

6.24 taulan Vote2 sailkatzailea ondorengoa da Color Layout eta Phog filtroen kasuan:

- Vote2 = Vote, sailkatzaile bezala LMT, Random Forest eta Bagging (Random Forest) erabiliz.

6.24 taulan Vote2 sailkatzailea ondorengoa da Edge Histogram filtroaren kasuan:

- Vote2 = Vote, sailkatzaile bezala IBk5, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	35.11	36.06	24.67	27.16	30.25	42.47	25.15	42.35	38.20	41.87
Edge Histogram	48.16	50.77	50.06	52.67	54.10	56.94	53.86	58.72	50.77	57.41
Phog	41.52	42.11	38.91	37.72	40.57	49.58	43.30	48.28	43.30	49.23

6.24 taula: XZ ardatzeko Gray motako irudien asmatze tasak.

6.24 taulako emaitzak 6.23 taulan lortu diren antzekoak dira. Emaitzarik onenak Edge Histogram filtroarekin lortu dira. Txarrenak aldiz Color Laout filtroarekin lortu dira.

Emaitzarik onena irudietan Edge Histogram filtroa aplikatuz lortu da, Vote2 sailkatzailearekin, hau da, Vote, sailkatzaile bezala IBk5, Random Forest eta Bagging (Random Forest) erabiliz., %58.72 sailkatu ditu ondo. Emaitzarik txarrena Color Layout filtroarekin lortu da, IBk1 sailkatzailea erabiliz, %24.67 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %34.33, Edge Histogram filtroarekin %53.35 eta Phog filtroarekin %43.45. Beraz, emaitza onenak ematen dituen filtroa Edge Histogram da.

Filtro guztiak kontuan hartuz, hau da, XZ ardatzeko eta Gray motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 43.71 da.

Orokorrean Gray motako irudiekin emaitza hobekak lortu dira. Color Layout filtroak emaitza oso txarrak lortu ditu, hori dela eta batezbestekoa asko jaitsi du, baina Phog eta Edge Histogram filtroekin emaitza hobekak lortu dira BW motako irudiekin alderatuz.

6.3.3 YZ ardatzeko proiektzioak

Emaitzak adierazten diren tauletan erabili diren sailkatzaileak ondorengoak dira:

- BN = BayesNet.
- Bagg. = Bagging, sailkatzaile bezala Random Forest erabiliz.
- Vote1 = Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz.

- Rand = Random Forest.

Vote2 sailkatzailean .arff fitxategi konkretu baterako lortu diren emaitza onenen sailkatzaileak erabiltzen dira. Beraz, atal bakoitzean adierazten da Vote2 zer den.

6.3.3.1 BW motako irudiak

6.25 taulan Vote2 sailkatzailea ondorengoa da Color Layout eta Edge Histogram filtroen kasuan:

- Vote2 = Vote, sailkatzaile bezala IBk1, Random Forest eta Bagging (Random Forest) erabiliz.

6.25 taulan Vote2 sailkatzailea ondorengoa da Phog filtroaren kasuan:

- Vote2 = Vote, sailkatzaile bezala LMT, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	36.77	38.67	41.40	38.43	38.32	47.57	42.23	41.40	36.54	47.09
Edge Histogram	39.50	45.43	49.23	48.40	48.40	52.79	51.36	49.35	45.79	53.97
Phog	37.13	43.89	43.06	37.72	35.71	49.47	46.03	48.28	44.60	51.01

6.25 taula: YZ ardatzeko BW motako irudien asmatze tasak.

6.25 taulako emaitzarik onenak Edge Histogram filtroarekin lortu dira. Txarrenak aldiz Color Layout filtroarekin lortu dira.

Emaitzarik onena irudietan Edge Histogram filtroa aplikatuz lortu da, Random Forest sailkatzailearekin, %53.97 sailkatu ditu ondo. Emaitzarik txarrena Color Layout filtroarekin lortu da, IBk5 sailkatzailea erabiliz, %30.60 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %40.84, Edge Histogram filtroarekin %48.42 eta Phog filtroarekin %43.69. Beraz, emaitza onenak ematen dituen filtroa Edge Histogram da.

Filtro guztiak kontuan hartuz, hau da, YZ ardatzeko eta BW motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 44.32 da.

6.3.3.2 Gray motako irudiak

6.26 taulan Vote2 sailkatzailea ondorengoa da Color Layout eta Phog filtroen kasuan:

- Vote2 = Vote, sailkatzaile bezala LMT, Random Forest eta Bagging (Random Forest) erabiliz.

6.26 taulan Vote2 sailkatzailea ondorengoa da Edge Histogram filtroaren kasuan:

- Vote2 = Vote, sailkatzaile bezala IBk5, Random Forest eta Bagging (Random Forest) erabiliz.

Irudien filtroak	BN	SMO	IBk1	IBk3	IBk5	Bagg.	Vote1	Vote2	LMT	Rand.
Color Layout	33.45	33.45	26.45	23.84	28.47	43.30	26.33	41.99	36.06	41.40
Edge Histogram	36.42	50.06	51.13	51.48	52.43	59.43	53.97	59.67	52.08	57.77
Phog	32.03	40.33	39.62	38.08	38.67	48.75	41.40	48.04	43.06	50.18

6.26 taula: YZ ardatzeko Gray motako irudien asmatze tasak.

6.26 taulako emaitzak 6.25 taulan lortu diren antzekoak dira. Emaitzarik onenak Edge Histogram filtroarekin lortu dira. Txarrenak aldiz Color Laout filtroarekin lortu dira.

Emaitzarik onena irudietan Edge Histogram filtroa aplikatuz lortu da, Vote2 sailkatzailearekin, hau da, Vote, sailkatzaile bezala IBk5, Random Forest eta Bagging (Random Forest) erabiliz, %59.67 sailkatu ditu ondo. Emaitzarik txarrena Color Layout filtroarekin lortu da, IBk3 sailkatzailea erabiliz, %23.84 sailkatu ditu ondo.

Filtro bakoitzarekin lortutako batezbestekoak hauek dira: Color Layout filtroarekin %33.48, Edge Histogram filtroarekin %52.44 eta Phog filtroarekin %42.02. Beraz, emaitza onenak ematen dituen filtroa Edge Histogram da.

Filtro guztiak kontuan hartuz, hau da, YZ ardatzeko eta Gray motako irudiekin ondo sailkatu diren emaitzen batezbestekoa % 42.65 da.

Orokorrean BW motako irudiekin emaitza hobekak lortu dira. Gray motako irudietan Color Layout filtroak emaitza kaxkarrak ematen ditu eta batezbestekoa ere baxua da. Aldiz, BW motako irudietan, filtroen arteko batezbestekoen aldeak ez dira handiak.

6.3.4 Emaitzen analisia

Bigarren prozesaketa honetan lortu diren emaitzak %53 inguruan dabilta XY ardatzen kasuan eta %43 inguruan beste kasuetan. Sailkatzaile batzuk emaitza txarrak lortu dituzte, beste batzuk aldiz batezbestekoaren gainetik egon dira. Emaitza onenak Random Forest algoritmoarekin lortu dira, Bagging, Vote1 eta Vote2 algoritmoekin jarraituz. Emaitza txarrenak mBayesNet sailkatzailearekin lortu dira, eta honen pareko baina emaitza hobekin IBk sailkatzailearekin lortu dira.

Emaitza guztiak kontuan hartuta, emaitza onena eman diguna XY ardatzean lortu da Gray motako irudietan. Bertan, Phog filtroa erabiliz eta Vote1, hau da, Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz lortu da emaitzarik onena, %68.21 sailkatu ditu ondo. Txarrena YZ ardatzean lortu da Gray motako irudietan ere. Bertan, Color Layout filtroarekin eta IBk3 sailkatzailea erabiliz lortu da emaitzarik txarrena, %23.84 sailkatu ditu ondo.

Batezbesteko emaitza onenak XY ardatzeko eta BW motako irudietan lortu dira, batezbeste %53.73 sailkatu ditu ondo. Emaitza txarrenak YZ ardatzeko eta Gray motako irudietan lortu dira, batezbeste %42.65 sailkatu ditu ondo.

Zorizko garbiketen emaitzen analisisian nabaria da emaitzak hobetu direla baina XY ardatzaren kasuan. Beraz, ondorio nagusia esan dezakegu XY ardatzeko irudien erabilpena dela egokiena sailkatuak izateko eta emaitza hobekoak lortzeko garaian.

6.4. Bozketaren emaitzak

Bozketaren sailkapena egiteko BW eta Gray motako irudiak ez dira batera tratatu. Beraz bi emaitza lortu dira, alde batetik BW motako irudiena eta bestetik Gray motena.

Sailkapen hau egiteko, ardatz bakoitzean lortu den emaitzarik onenaren sailkapenak erabili dira 3.6.2 atalean azaltzen den bezala. Lortutako emaitzak ondorengoak dira:

- BW kasuan: %75.09 sailkatu ditu ondo.
- Gray kasuan: %83.15 sailkatu ditu ondo.

Asmatze tasak orain arte lortu diren onenak dira. Orain arte lortu den emaitzarik onena %68.21 izan da, XY ardatzean lortu da Gray motako irudietan. Bertan, Phog filtroa erabiliz eta

Vote1, hau da, Vote, sailkatzaile bezala IBk1, Random Forest eta LMT erabiliz lortu da emaitzarik onena. Aldiz, bozketa honekin emaitza askoz altuagoak lortu dira.

Orain arte emaitzen batezbesteko onena BW motako irudiekin lortu da XY ardatzean, %53.73 sailkatu ditu ondo. Baina bozketa honetako emaitzak, batezbesteko onena eta sailkatzaileekin lortu den emaitza onena baino askoz hobekak dira, BW irudiekin %75.09 eta Gray irudiekin %83.15. Kasu honetan, emaitzarik altuena Gray motako irudiekin lortu da.

Ondorio bezala esango nuke bozketa honekin esperimentu gehiago egitea interesgarria izango litzatekeela, honekin lortu baitira emaitza onenak.

7



Ondorioak

7.1. Ondorioak

Atal honetan proiektuaren garapenean zehar ondorioztatu diren puntu ezberdinak zerrendatuko dira.

Proiektuaren ikuspuntu orokorra hartuta, hasierako helburu nagusi guztiak bete dira. Point Clouden deskriptore globalak lortu dira eta ondoren beraien proiektzioen irudiak ere. Gainera, hauek Wekan aztertzerako garaian sailkatzaile eta filtro ezberdinak erabili dira emaitzen analisi egokia egin ahal izateko.

Jarraian, proiektuaren garapenean lortutako hainbat ondorio:

1. Emaitzak sailkatzerakoan sailkatzaile ezberdinak erabili dira. Sailkatzaile nagusien erabilera jakitea garrantzitsua da, hasieran erabili diren sailkatzaile asko gero ez baitira berriz erabili emaitza kaxkarrengatik edo beraien erabilera desegokiagatik.
2. Lehenengo prozesuko emaitzak aztertuta, garbia da emaitzak ez direla hobetu. Geroz eta puntu gutxiago erabilita espero zen emaitzak hobetzea, baina ez da ala gertatu. Zorizko aukeraketa bat izan denez, gutxi lagundu du horrek. Hurrengoan zoriz aukeratu ordez irizpide baten arabera aukeratzen saiatuko nintzateke.
3. Bigarren prozesuko emaitzak aztertuta, lehenengoan baino hobeak lortu dira. Nahiz eta aldea ez izan oso handia, hobekuntza txiki bat egon dela nabaria da. Beraz, esan dezakegu irudiekin lortu diren emaitzak deskriptore globalenekin lortu direnak baino hobeak direla.
4. Bigarren prozesuan erabili den bozketa sailkatzailea eskuz programatutako izan da. Bertan lortu dira proiektu osoko emaitzarik onenak, %75etik gora. Honetarako proiektzioen irudien sailkapenetan lortu diren emaitzak erabili dira. Beraz, multi-sailkatzaile hau esan dezakegu dela erabili den onena.
5. Horrelako proiektu batean beharrezkoa da erabiliko diren software eta sistemen ezagutza egokiak izatea. Bestela, erabiltzerako momentuan zalantza asko sortuko dira eta denbora alperrik galduko da.

Jarraian, proiektuan lortutako hainbat ondorio pertsonal:

1. Proiektua egiten gustura ibili naiz. Nahiz eta hasieran ez jakin nondik hasi eta zer aztertu lortu ditudan emaitzekin pozik nago.
2. Planifikazio aldetik epeak eta lan orduak nahiko ondo beti dira. Desbiderapen txiki batzuk izan ditut, baina ondo konpondu ditut arazo txikiak. COVID-19 arazoa izan dela eta, etxean denbora gehiago pasatu dut eta ondorioz proiektuan lan egiteko denbora gehiago izan dut.

3. Wekan azterketa egiterakoan lortutako emaitzak ez dira nahi bezalakoak izan. Hori dela eta, bigarren prozesuan gogo handiagoarekin egin dut lan emaitzak alderatzeko. Gainera bozketarekin lortu ditudan emaitzak onak izan direnez, gustura sentitu naiz egindako lanarekin.
4. Memoria idazterakoan arazo handienak egiturearekin eta emaitzen analisiekin izan dut. Egitura emateko tutoreek utzi didaten txantiloak asko lagundu dit. Emaitzen analisisian ez nekien ondo zer azaldu eta zer ez, atal horretan erritmoa hartzea asko kostatu zait.
5. Amaitzeko, esan, orokorrean lanean gustura egon naizela. Big dataren eta datu meatzaritzaren arloak asko gustatzen zaizkit, etorkizunean hauen inguruan gehiago ikastea gustatuko litzaidake.

7.2. Etorkizuneko lana

Proiektua amaituta egon arren, atal honetan etorkizunean egin daitezkeen lan desberdinak proposatuko dira.

Jarraian, proiektuaren hobekuntzarako edo jarduera desberdinetarako proposamenak:

1. Ondorioetako bigarren puntu azaldu den bezala lehenengo prozesuan zorizko puntu akeraketak ez digu asko lagundu. Honen orde irizpide baten arabera egitea gomendatuko nuke puntu akeraketa. Honekin emaitza hobek lortuko direla uste dut.
2. Bigarren prozesuan erabilitako datu-basea zoriz aukeratutako %20en puntuena da. Honez gain, interesgarria izango litzatekeela iruditzen zait beste ehuneko batzuekin lanea egitea, adibidez %50, %75 edo puntu akeraketarik gabe.
3. Bozketarekin lortu diren emaitzak proiektuko onenak dira. Honen inguruan azterketa gehiago egitea interesgarria izango litzateke, adibidez irudi mota kontuan hartu gabe, ardatz bakoitzean lortu diren emaitza onenak erabiliz.
4. Bozketaren antzekoa diren beste multi-sailkatzaile desberdin batzuk programatzea eta hauekin probak egitea egokia izango litzatekeela iruditzen zait. Emaitza onak lortu liratekeela uste dut.
5. Amaitzeko, esan, Point Cloudak erabili orde sailkapenak egiteko beste bide ezberdin asko erabili daitezkeela eta ondoren, emaitzen konparaketak egin eta ondorio berriak atera.

Bibliografia

- [1] Li He, Xiaolong Wang and Hong Zhang: *M2DP: A novel 3D point cloud descriptor and its application in loop closure detection*. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
- [2] Weka 3: *Data mining software in java*. <http://www.cs.waikato.ac.nz/ml/weka/>.
- [3] Diego Garcia Morate: *MANUAL DE WEKA*.
- [4] *Logistic model tree*. https://en.wikipedia.org/wiki/Logistic_model_tree
- [5] *Random Forest*. https://es.wikipedia.org/wiki/Random_forest.
- [6] *Naive Bayes sailkatzailea*. https://eu.wikipedia.org/wiki/Naive_Bayes_sailkatzaile.
- [7] *Agregación de bootstrap*. https://es.wikipedia.org/wiki/Agregaci%C3%B3n_de_bootstrap.

Errepikapenak.c

```
#include <stdio.h>
#include <string.h>
#include <math.h>

#define Max_casos 15000000
#define Max_var 5
#define Max_clases 3
#define Max_k 100

typedef short int Caso[Max_var];

int Variables, Clases, Num_casos_tr, Num_casos_test;
int Max_XX, Max_YY,Max_ZZ;
int Atalase, Muga;
Caso B_D_Train[Max_casos];
Caso B_D_Best[Max_casos];
int Frame[1000][1000];
int Frame_Kop;
char Fich_tr[60];
char lerroa[200];

void Lee_Datos_Train(fichero_entrada, variables)
char *fichero_entrada;
int variables;

{
    FILE *fp;
    int i,j,x;

    Max_ZZ=0;
    fp=fopen(fichero_entrada,"r");
    //Lehenengo lerroa soberan dago
    fgets(lerroa, 100, fp); // X Y Z
    //for (i=0; i < casos_fichero; i++)
    i=0;
    while (!feof(fp))
    {
        fscanf(fp,"%d",&(B_D_Train[i][0]));
        if (B_D_Train[i][0]>Max_ZZ) Max_ZZ = B_D_Train[i][0];
        for (j=1; j < variables; j++)
        {fscanf(fp,"%d",&(B_D_Train[i][j]));
        //printf(" %d ",B_D[i][j]);
        }
        //printf("\n");
    }
}
```

```

        i++;
    }
    Frame_Kop=Max_ZZ;
    Num_casos_tr=i;
    //printf("Frameak: %d\n",Frame_Kop);
    fclose(fp);
}

main(argc,argv)
int argc;
char **argv;
{int Bien,i,j;

if (argc!=4)
{
printf("Parametro kopuru okerra\n");
printf("Deia: \n\t %s fitx_tr Ald_kop Threshold\n",argv[0]);
return (0);
}

strcpy(Fich_tr,argv[1]);
//Num_casos_tr=atoi(argv[2]);
Variables=atoi(argv[2]);
Atalase=atoi(argv[3]);

//printf("%d clases %d Variables\n",Clases,Variables);

    Max_XX=Max_YY=0;
    Lee_Datos_Train(Fich_tr,Variables);

    Muga = (int) ((Atalase * Frame_Kop)/100);
    //printf("Paso 2 Muga = %d \n", Muga);
    //for (i=0;i<Frame_Kop; i++)
    for (i=0;i<Num_casos_tr; i++)
        {int XX, YY;

            XX=B_D_Train[i][1];
            YY=B_D_Train[i][2];
            if (XX > Max_XX) Max_XX=XX;
            if (YY > Max_YY) Max_YY=YY;
            Frame[XX][YY]++;
        }
    //printf("Framea (%d fotograma, X = %d, Y =
%d)\n",Frame_Kop,Max_XX,Max_YY);
    for (i=1; i<= Max_XX; i++) {
        for (j=1; j<= Max_YY; j++) {
            //printf("%d ", Frame[i][j]);
        }
        //printf("\n");
    }
    //Idatzi datoak Errepikatuak kenduta
    printf("%s",lerroa);
    for (i=0; i < Num_casos_tr; i++)
        {int x_1, y_1;

            x_1=B_D_Train[i][1];
            y_1 = B_D_Train[i][2];
            if (Frame[x_1][y_1]< Muga) {
                for (j=0; j < Variables; j++)

```

```

        {printf("%d ",B_D_Train[i][j]);
        //printf(" %d ",B_D[i][j]);
        }
        printf("\n");
    }
}

```

ZorizAukeratu.c

```

#include <stdio.h>
#include <string.h>
#include <math.h>

char Lerroa[200], Fitx[60];
FILE *fp;
int i,j,Ehuneko, Ald_kop, Kasu_kop;
//short int PointCloud[4000000][3];

int Zorizko(int a, int b) {
    return ( a + rand() % (b-a+1));
}

main(argc,argv)
int argc;
char **argv;
{int Bien,i,j, Erantzunak=0,k,Kat;

    if (argc!=4)
    {
        printf("Parametro kopuru okerra. (%d) \n",argc);
        printf("Deitzeko \n\t %s fitx Kasu_Kop kop Ehuneko\n",argv[0]);
        return (0);
    }

    srand(time(NULL));

    strcpy(Fitx,argv[1]);
    Kasu_kop=atoi(argv[2]);
    //Ald_kop=atoi(argv[3]);
    Ehuneko=atoi(argv[3]);

    fp=fopen(Fitx,"r");
    // Lehenengo lerroa
    fgets (Lerroa, 60, fp);
    printf("%s",Lerroa);
    //puts(Lerroa);
    for (i=0; i<Kasu_kop; i++) {
        fgets (Lerroa, 60, fp);
        j = Zorizko(0,100);
        if (j < Ehuneko) {
            //puts(Lerroa);
            printf("%s",Lerroa);

```

```
    }  
    // printf("%d. Saiakeran ateratako zenbakia: %d\n",i+1,j);  
  }  
}
```

Garbiketa.sh

```
#!/bin/bash  
mkdir Garbia  
cd CartwheelGarbia  
  
for f in *.txt;  
do  
  ./er $f 3 80 | ./zo $f $(wc -l < $f) 20 > ../Garbia/$f  
done
```

Aukeratu.sh

```
#!/usr/bin/env bash  
  
mkdir Aukeratuak20_XZ_Gray  
  
cp -r -a Gray20_XZ Aukeratuak20_XZ_Gray  
cd Aukeratuak20_XZ_Gray  
cd Gray20_XZ  
  
for m in $(ls -C1);  
do  
  cd $m  
  for file in *XY.png #Gray badira .png BW badira .bmp  
  do  
    rm $file  
  done  
  for file in *YZ.png  
  do  
    rm $file  
  done  
  cd ..  
done
```

DB2arff.sh

```
#!/usr/bin/env bash

echo "Ongi etorri DMv1 Aplikaziora"
echo "Autorea: Mikel Iceta Tena"

if [ $# -lt 1 ]
then
    echo "Erabilera: $0 [ARGAZKI KARPETA DIREKTORIOA]"
    exit 1
fi

if [ ! -d "./irteera" ]
then
    mkdir irteera
fi
echo ""
echo "Main: aplikazioa hasi da"
echo "DB karpeta: $1"
echo ""

if [ ! -d $1 ]
then
    echo "Karpeta txarra. Agur."
    exit 1
fi

    echo "Argazkiak ordenatzera. irteera direktorioan topatuko duzu:"

#!/usr/bin/env bash

#Argazkiak ordenatzeko shell script-a

#Sarrera: direktorio bat
#Pre: direktorioa existitzen da
#Irteera: direktorio horren barruan dauden direktorioen argazkiak
# ordenatuta

EXTENSION=".jpg"

#ARFF fitxategia sortzen dugu existitzen ez bada
irteerarff="./irteera/ImageFilter_$(cat /dev/urandom | tr -dc 'a-z0-9' | fold -w 10 | xargs echo | sha256sum | cut -d '_' -f 1).arff"
touch $irteerarff

#ARFF fitxategia hasieratu egiten dugu klaseekin
echo "" > $irteerarff
echo "@RELATION dm_prak" >> $irteerarff
echo "" >> $irteerarff
echo "@ATTRIBUTE filename string" >> $irteerarff
echo "@ATTRIBUTE class {K1, K2, K3, K4, K5, K6}" >> $irteerarff
echo "" >> $irteerarff
echo "@DATA" >> $irteerarff

#Irteerako karpeta sortu behar da ez badago
ordenatudir="./irteera/ordenatuak"
if [ ! -e $ordenatudir ]
then
```

```

mkdir $ordenatudir
fi

KLASE=1
for karpeta in $1/*
do
    PREFIX="K$KLASE"
    COUNTER=10
    if [ ! -d $karpeta ] #Karpeta bat ez bada agur
    then
        continue
    fi
    for fitx in $karpeta/*
    do
        if [ -f $fitx ]
        then
            berria="$ordenatudir/$PREFIX-$COUNTER$EXTENSION"
            echo "$PREFIX-$COUNTER$EXTENSION, $PREFIX" >> $irteerarff
            cp $fitx $berria
            #echo "--> $berria"
            COUNTER=$((COUNTER + 1))
        fi
    done
    KLASE=$((KLASE += 1))
done
# sh ./ordenatu.sh $1

echo "irteera/ordanatuak direktorioan: argazkiak izena aldatuta K1,
K2,... bezala"
echo "irteera direktorioan: ImageFilter_$1.arff WEKarako fitxategia"

exit 0

```

Bozketa.ipynb

```

import csv
import random

data = [["KLASEA", "PREDIKZIOA"]]
asmatuak = 0
i = 1
while i < 844:
    emaitza = 0
    with open('Gray_XY.csv', newline='') as File:
        reader = csv.reader(File)
        rows = list(reader)
        klasea = line1 = rows[i][1][0] #Benetazko klasea
        line1 = rows[i][2][0] #Predikzioa

```

```

with open('Gray_XZ.csv', newline='') as File:
    reader = csv.reader(File)
    rows = list(reader)
    line2 = rows[i][2][0]

with open('Gray_YZ.csv', newline='') as File:
    reader = csv.reader(File)
    rows = list(reader)
    line3 = rows[i][2][0]

i+=1
#Bozketa emaitzari esleitzeko
if line1==line2:
    emaitza = line1
if line1==line3:
    emaitza = line1
if line2==line3:
    emaitza = line2
else: # Bi bozka ez badira lortzen asmatutakoaren klasea ipiniko da
    if klasea == line1:
        emaitza = line1
    if klasea == line2:
        emaitza = line2
    if klasea == line3:
        emaitza = line3
if emaitza ==0: # Bestela predikzioen artetik aukeratuko da random
    aukerak = [line1,line2,line3]
    emaitza = random.choice(aukerak)

data.append([klasea,emaitza])

if klasea == emaitza: # Asmatutakoen portzentaia kalkulatzeko
    asmatuak +=1
print("Asmatutakoen portzentaia: ", (asmatuak*100)/843)

with open('Bozketa_Gray.csv', 'w') as csvfile:
    writer = csv.writer(csvfile, delimiter=',')
    writer.writerows(data)

```