



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

III. IKERGAZTE NAZIOARTEKO IKERKETA EUSKARAZ

2019ko maiatzaren 27, 28 eta 29
Baiona, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

GIZARTE ZIENTZIAK ETA ZUZENBIDEA

Desberdin ikusten al dute
informatika eta informatikaria
10-12 urteko neskek eta mutilek?

*Mikele Zurutuza-Renom, Edurne
Larraza-Mendiluze eta
Olatz Arbelaiz-Gallego*

36-43 or.
<https://dx.doi.org/10.26876/ikergazte.iii.02.04>



Desberdin ikusten al dute informatika eta informatikaria 10-12 urteko neskek eta mutilek?

Zurutuza-Renom, Mikele; Larraza-Mendiluze, Edurne eta Arbelaitz-Gallego, Olatz

Donostiako Informatika Fakultatea UPV/EHU

edurne.larraza@ehu.eus

Laburpena

Ezaguna da gaur egun teknologiarekin eta bereziki informatikarekin lotutako ikasketa eta lanbideetan dagoen emakume eskasia baina arrazoiak eta bokazioak noiz aldentzen diren ez dago argi. Gai horri argi printza batzuk emateko asmoz, konputagailuen programazioa gaztetxoaren artean zabaltzeko mundu mailan hedatu den Kode Ordua oinarri hartuta, hiru urtez hainbat ikastetxetan 10-12 urte bitarteko gaztetxoaren artean bildutako galdetegietako erantzun eta marrazkien azterketa egin dugu lan honetan. Datu-meatzaritzako prozesu oso bat egin behar izan dugu horretarako. Lortutako emaitzen artean nabarmendu nahiko genuke badirudiela gure inguruko 10-12 urteko neskek oraindik hein handi batean ikusten dutela Informatikaria neska izan daitekeela, mutilek ordea, oro har gizonezkoak edo identifikatu ezinezko generoa duten gizakiak imajinatzen dituzte.

Hitz gakoak: programazioa, gaztetxoak, datu-meatzaritzak, informatikarien irudiak

Abstract

It is broadly known that the study and work fields of technology and informatics lack women participants. However, the reason causing this is unclear. In order to tackle this issue, based on a worldwide expanded project called The Hour of Code, whose aim is to make kids aware of the basic concepts of computer programming, an analysis of the answers and drawings gathered in a survey conducted among 10-12-years-old youngsters has been performed. With the aim of interpreting the results, a complete data-mining process has been executed. As a conclusion, it seems young girls still can figure a women computer scientists, but young boys don't or at least they represent them as men or as unidentifiable genre.

Keywords: programming, young teenagers, data mining, drawing computer scientists

1. Sarrera eta motibazioa

Bada informatika arloan eztabaidagarri bihurtu den gai bat. Jakina denez, gaur egun informatikaren munduan sartzen direnen kopurua urria da beste arlo batzuekin alderatuta, emakumearen partaidetza are eskasagoa. Ondorioz, kezka handia dago zabalik. Hainbat ikerketek, adibidez, Medel eta Pournaghshband (2017) lanean topa daitezkeenak, agerian utzi dute emakumeak izan dezakeen konfiantza falta arlo honetan. Informatikari baten estereotipoarekin bat ez egitea edota ingurugiro maskulinoa dira hauentzat mehatxu nagusi esparru honetan sartzekoan. Baina, zein unetatik aurrera aldatzen da pertsonen ikuspuntua? 10-12 urteko haurrengan, desberdina al da informatikariari buruz duten ikuspuntua generoaren arabera? Litekeena da egungo gizartearen eraginez sortzea desberdintasunak, are zehatzago, egungo haurrek jasotzen duten hezkuntzak. Mundu-mailan informatika derrigorrezko hezkuntzan txertatzeko mugimendu batzuk badauden arren, adibidez Erresuma Batua (Brown *et al.*, 2014), Lituania (Dagiene, 2002) eta Finlandia (DeRuy, 2017), gutxi dira hori horrela egiten duten herrialdeak. Horrela, informatika gaztetxoei helarazteko mundu mailan ezaguna den eta 180 herrialde baina gehiagotako milioika ikaslerengana iristen den *Hour of Code* proiektuak rol aipagarria izan du (Code.org, a) (Code.org, b).

UPV/EHUko Informatika Fakultateko hainbat ikasle eta irakaslek Kode Ordua aurrera eraman dute zenbait ikastetxetan. Ordubetez modu dibertigarri eta dinamikoan programatzen aritzea eta programazioko oinarriko kontzeptuak barneratzea da Kode Orduaren helburua. Horretarako, code.org (www.code.org) tresna erabiltzen da, zeina 2015-16 ikasturtean euskarara itzuli zen UEUko Informatika Saila eta UPV/EHUko Donostiako Informatika Fakultatearen arteko elkarlanari esker. Kode ordua hiru urtez jarraian gauzatu dugu 10-12 urte bitarteko haurrekin.

2015/2016, 2016/2017 eta 2017/2018 ikasturteetan eraman da aurrera, zortzi, zazpi eta hamabi ikastetxeren parte hartzearekin, hurrenez hurren. Ekimen hau baliatu da haurrei inkesta bat pasa eta hauek informatikari buruz duten ezagutza eta ikuspuntua zein den aztertzeko. Horrez gain, inkestetan jasotako datuen bidez ikasle bakoitzaren inguruak horretan duen eragina ere aztertu nahi izan da.

Hurrengo atalean honaino iristen lagundu diguten hainbat lanei erreferentzia egingo diegu. Hirugarren atalean ikerketaren muina azalduko dugu, jarraitu dugun metodologia eta lortutako emaitzak azalduz. Laugarren atalean, ondorioak eta etorkizunean egin nahiko genukeena kontatuz amaituko dugu.

2. Arloko egoera eta helburuak

Arloko egoera aipatzerakoan, eta programazioa gaztetxoei irakasteaz ari garenez, garrantzitsua deritzogu aipatzea bi lan, programazioa pertsona orori irakasteko arrazoi zuzenak zeintzuk diren azaltzen dituztenak (Lewis, 2017) eta (Guzdial, 2015). Egia esan bi lan horietan ez dira programazioaz ari, "Computer Science" Informatikaz baizik. Hori bai, zehaztu beharra dago Informatika esaten dugunean, zientziaz, matematikaz, ingeniartzaz ari garela eta ez Informatikak sortzen dituen teknologien erabileraz. Bi lan horietan zehazten dute Informatika eta horren barruan programazioa txikitatik irakasteko erabiltzen diren argudioetako asko ez direla baliagarriak, baina bat datoz esaterakoan garrantzitsua dela gaur egungo munduaren parte delako teknologia digitala eta mundu hori ezagutu eta ulertzeko egokia dela. Gainera, oso baliagarria izan daitekeela ere arloaren inguruan dauden estereotipoak ezabatzeko eta ikasleak Informatikatik gertuago ikus dezaten beren burua eta arlo horretako ikasketa sakonagoak egiteko kapazitatea badutela antzemateko. Noski, horretarako irakaskuntza prozesuak helburuekin lerrokaturik egon behar du eta ikasleek ikusi beharra dute zein den egiten dituzten jardueren eta arloaren arteko lotura.

Zaila da jakitea zeintzuk diren eragina duten ekimenak, baina, Estatu Batuetako unibertsitate batzuetan informatika ikasketetan matrikulatzen diren emakumeen portzentajeak igotzea lortu dute (Frieze eta Quesenberry, 2019). Euskal Herriko Unibertsitatean (UPV/EHU) oraindik ez daukagu horrelakoak esaterik, baina arazoaren jatorria bilatzen darraigu Kode Ordua ekimenarekin bat Lehen Hezkuntzako (LH) 5. eta 6. mailako ikasleen iritziak jasoaz eta haiek erabiliz hainbat galdera erantzuten saiatzeko. Hala nola, zein neurritan da desberdina 10-12 urteko neskek eta mutilak informatikari eta informatikariei buruz duten iritzia? Haiei buruz duten irudia bat ote dator *freaky*aren estereotipoarekin? 10-12 urteko haurrek zenbateraino irudikatzen dituzte emakume informatikariak? Galdera horiei erantzunek jarraibideak emango dizkigute etorkizuneko ekintzak erabakitzeko.

3. Ikerketaren muina

Lan honetan, mundu errealeko datu-meatzarritza prozesu oso bat egin da eta horretarako, ohikoak diren urratsak jarraitu dira (Clifton, 2017). Hau da, horretan pentsatu gabe bildutako datuak aurreprozesatu eta bateratu, ikasketa automatikoko tresnekin erabili ahal izateko prestatu, eta jarraian, azalpena eman diezaguketen ikasketa automatikoko metodoak erabili dira hainbat galderei erantzuten saiatzeko.

3.1. Datu-bilketa eta prestatzea

Ikerketa hau egiteko erabili diren datuak, UPV/EHUko Donostiako Informatika Fakultateak antolatuta hiru urtez, 10-12 urte bitarteko hainbat haurri eskaini zaien kode orduan jasotako inkestak dira. 2015/2016, 2016/2017 eta 2017/2018 ikasturteetan, zortzi, zazpi eta hamabi ikastetxetan egin da ekimena. Ikasleei etxean betetzeko inkesta bat eman zaie. Inkesta hori, ikastetxe batzuetan bete dute baina ez guztietan. Hain zuzen ere, 2015/2016 ikasturtean 5 ikastetxe izan dira inkestak bueltatu dizkigutenak (267 inkesta), 2016/2017 ikasturtean hiru (207 inkesta), eta 2017/2018 ikasturtean berriz 7 (393 inkesta). Ikastetxe gehienak Gipuzkoan kokatuta daude, nahiz eta badiren Bizkaia eta Nafarroakoak eta bakarra da hiru urteetan zehar inkestak itzuli dituen ikastetxea. Azken hau, kasu partikular (K.P.) gisa aztertuko da.

Prestatu den galdera-sortaren helburua haurrek informatikari buruz duten ezagutza eta ikuspuntua zein den eta ikasle bakoitzaren inguruak horretan duen eragina aztertzea da. Horren gaztetxoak diren ikasleekin askotan kostatzen da asmatzea nola jaso haien iritzi eta ikuspuntuak. Horretarako sortu ziren hain zuzen ere "Draw A Scientist Test" edo DAST deritzotenak (Chambers, 1983) eta Ingeniarietzako ere erabiliak izan direnak (Knight eta Cunningham, 2004). Hori kontuan izanik, lan honetarako, beste galdera batzuen artean ikasleei eskatu zaie batetik Informatikaria eta bestetik Informatikariaren lantokia marraztea. Irudi horien bidez saiatuko gara aztertzen ikasleek horren adin gaztean dagoeneko informatikariei buruzko genero eta *freaky* estereotipoak barneratuta ote dituzten. Irudiaren generoa, betaurrekoak ote dituzten eta informatikariak lanean, bakarka ala taldean imajinatzen ote dituzten aztertuko dugu horretarako.

Zehazki, honako datuak eskatu zaizkie ikasleei: (a) ikasturtea (aldagai 1), (b) eskola (ald. 1), (c) generoa (ald. 1), (d) adina (ald. 1), (e) guraso bakoitzaren generoa (2 ald.), (f) guraso bakoitzaren ikasketa-maila (2 ald.), (g) irudi bidez adierazitako ogibide multzo bat haietan informatikari batek lan egin ote lezakeen galdetzeko (medikuntza, artea, kimika, autogintza, jantzigintza, etxegintza, bideo-jokoak, musikagintza, zuzenbidea, hizkuntzalaritza, ordenagailu-sareak, hezkuntza) (12 ald.), (h) Irudikatu informatikari bat (2 ald., generoa, betaurrekoak), Irudikatu informatikarien lantokia (ald. 1, bakarrik edo taldean).

Hiru ikasturteetako inkestak, helburu gisa informazio bera eskuratzea duten arren, lehenengo urteko esperientzian oinarrituz egokitu egin dira, eta gainera, digitalizazio-prozesua ere aldatu da urte batzuetatik bestetara. Ondorioz, ezertan hasi aurretik datuak bateratu behar izan dira, aldagai bakoitzarentzat kodeketa egokia erabakiz. Horrela, ondoko ezaugarriak dituen datu-basea lortu dugu: 21 aldagai (zenbakizko bat eta 20 nominal), eta 867 adibide. Zoritxarrez, datu-basearen lehen azterketa batean ikusi ahal izan dugu eremu eta kasu guztiek ez dutela informazioa, hau da, balio hutsak edo *missing values* esaten zaienak ageri direla datu-baseren % 26,66 kasutan. Beraz, besteak beste hauekin zer egin erabaki beharko da. Datu-base orokorra aztertzeaz gain, K.P.n jasotako datuak ere aztertuko ditugu, portaera orokorra ikastetxe bakar batean gertatzen denarekin alderatu ahal izateko. Kasu honetan, adibide kopuru osoa 207koa da eta balio hutsak % 21,74 kasutan azaltzen dira.

3.2. Ikasketa automatikoko teknikak galderak erantzuteko

Ikasketa automatikoko algoritmo asko ez daude prestatuta balio hutsekin lan egiteko. Ondorioz, lehenengo urratsa balio hutsekin zer egingo den erabakitzea izango da. Bestalde, sarreran adierazi den moduan, hainbat galderari erantzuna bilatu nahi zaie. Hori egin ahal izateko, menpeko aldagai desberdinak hautatu beharko dira.

3.2.1. Balio hutsak

Bibliografia aztertuz, hainbat irtenbide daude datu hutsen arazoari aurre egiteko. Horietan onena aukeratzea, datuen domeinuaren eta datu-analisiaren helburuaren arabera da eta hainbat aukera posible dauden arren guk bi sinpleenak probatu ditugu:

1. Datu hutsak dituzten errenkadak ezabatu. Errenkada batean atributu edo aldagairen baten informazioa falta bada, errenkada edo adibide hori baztertzean datza. Hau arazo bat izango da ezabatu beharreko errenkada kopurua oso handia den kasuetan.
2. Batezbesteko edo Moda balioarekin ordezkatu. Falta diren balioak datu-baseko beste adibideetan ageri diren balioak erabiliz egindako kalkuluekin lortutako balioekin ordezkatzeko dira. Zenbakizko aldagaiekin kasuan, aldagai horrek datu-basean hartzen duen batezbestekoagatik ordezkatzeko dira eta aldagai nominalen kasuan aldiz, aldagai horretan datu-base osoan maiztasun handiena duen balioarekin.

Bi hurbilpen hauekin lehen esperimendu batzuk eginez, antzeko asmatze-tasak lortu ditugu eta ondorioz, datuak fidagarriago direlakoan, datu hutsak dituzten errenkadak ezabatzea erabaki dugu. Ondorioz, gure datu-baseak 636 adibide izango ditu eskola guztiak kontutan hartzen diren kasurako eta 162 K.P.rako.

3.2.2. Galderei erantzuteko datu-base desberdinen sorrera

Hemendik aurrerako azterketa guztiak balio hutsak ezabatuta dituen datu-basearekin egin dira. Galdera bakoitza erantzuteko, menpeko aldagai desberdina hautatu eta datu-base desberdin bat sortu dugu. Haien ezaugarriak 1 Taulan ageri dira. Kasu honetan lehenengo erantzun nahiko genukeen galdera da ea 10-12 urteko haurren generoa berezi ote daitekeen inkestetan eman dituzten erantzunetan oinarrituta. Hala nola, neskek eta mutilek adin horretan informatikari buruz duten iritzia eta ezagutza desberdinak dira? Horretarako, menpeko aldagaia generoa izango da eta osatutako datu-basea nahiko orekatua da bi klaseetarako zenbakiek adierazten duten moduan. Erantzun nahi dugun hurrengo galderak, haurrek egindako irudiak kontuan hartzea eskatzen du. 10-12 urteko haurrek informatikaria imajinatzen dute emakume moduan? Eta, zeren arabera da hori? 1 Taulako balioek adierazten diguten moduan, bai datu-base osoan eta baita K.P.koan ere, gehiengoak gizonezkoa irudikatu zuen eta ondorioz, desorekatua dagoen datu-basea izango dugu oraingo honetan.

Horrez gain, informatikariaren *freaky* fama hori haurrei iritsi ote zaien aztertzeke, informatikariak betaurrekoekin irudikatzen dituzten ala ez alde batetik, eta bestetik, informatikaria lantokian bakarrik irudikatzen ote duten ala taldean aztertu dugu. Betaurrekoekin dagokienean, datu-base osoan ikasleen erdiak baino gutxiagok irudikatzen dute

1. taula. 3 galderetarako datu-basearen banaketa.

Aldagaia	balioa	Eskola guztiak		K.P.	
		Kopurua	Ehunekoa	Kopurua	Ehunekoa
Ikaslearen generoa	Neska	328	% 51.57	89	% 54.93
	Mutila	308	% 48.43	73	% 45.06
	Ez bitarra	0	% 0	0	% 0
Marrazkiaren generoa	Emakumezkoa	186	% 29.25	58	% 35.8
	Gizonezkoa	363	% 57.07	90	% 55.55
	Ezin zehaztu	87	% 13.68	14	% 8.64
Betaurrekoak	Bai	283	% 44.50	88	% 54.32
	Ez	353	% 55.50	74	% 45.68
Ordenagailuak lantokian	Bat	490	% 77.05	124	% 76.545
	Asko	146	% 22.95	38	% 23.46

informatikaria betaurrekodun eta K.P.ren kasuan erdiak baino piska bat gehiagok. Bi kasuetan ikasleek emandako erantzunaren ehunekoa, gizartean betaurrekoak erabiltzen dituen jendearen ehunekoaren azpitik dago eta beraz, ez du atentziorik ematen. Informatikariaren lantokiari dagokionean, bi datu-baseetan gehiengoak konputagailu bakarra marraztu du; gehiengoak informatikariak bakarrik irudikatzen ditu alegia.

1 Taulari begiratu esan liteke adin goiztiarretan hurrek errazago imajinatzen dutela gizonezkoa informatikari gisa, emakumezkoa baino, eta gainera, bakarrik lan egiten irudikatzen dituztela, taldean baino gehiago. Betaurrekoak aldiz, ez dituzte informatikarien ezaugarri bereizgarritzat hartzen.

3.2.3. Ikasketa automatikoa galderei erantzuteko

Azterketaren helburua bikoitza da, egin ditugun galderentzat asmatze-tasa altuak dituzten sailkatzaileak lortzen saiatzea alde batetik, eta, hau posible bada, aukera bakoitzarentzat azalpen bat topatzea. Horretarako, azalpena ematen duten sailkatzaileak erabiliko ditugu, sailkapen zuhaitzak eta erregela-multzoak hain zuzen ere. Hala ere, lehenengo puntua bazterrean ez uzteko, beste sailkatzaile batzuk ere erabiliko ditugu asmatze-tasa hobetzen ote duten aztertzeko. Esperimentuak Witten *et al.* (2011) liburuan aurkezten den Weka software ingurunean egikaritu ditugu. Ondoko sailkatzaileak erabili ditugu:

- Quinlan (1993) lanean aurkeztutako J48 (sailkatze zuhaitzak eraikitze C4.5 algoritmoaren inplementazio librea). Sailkapen zuhaitzak datu-basearen ondoz ondoko banaketa eginez eraikitzen dira, urrats bakoitzean hori egiteko aldagai eta banaketa egokiena hautatzen dira. Hostoek klase bat adierazi ohi dute (bertan erori diren adibide gehienei dagokiena) eta adabegien arteko loturek aldagai eta baldintzak.
- Eibe Frank eta Ian Witten-ek proposatutako PART Algoritmoa (Frank eta Witten, 1998), erregelak sortzeko erregela-algoritmoen bi mota nagusiak konbinatzen dituena: sailkapen zuhaitzetatik erregelak sortzea, eta erregelak ikasteko ohikoa den banatu eta irabazi teknika. Algoritmoak sailkapen zuhaitz partzialak sortzen ditu; erabat garatu gabeak dauden C4.5 zuhaitzak hain zuzen ere.
- Consolidated Tree Construction (CTC) (Pérez *et al.*, 2007). Lagin anitzetan oinarrituta sailkapen zuhaitzak sortzen dituen algoritmoa, klase desoreka dagoen kasutarako egokia eta azalpen egonkorra ematen duena.
- John eta Langley (1995) lanean aurkeztutako Naive Bayes oinarritzko sailkatzailea (NB).
- Sailkapen-zuhaitzetan (J48) oinarritutako meta-sailkatzaile bat ere eraiki dugu: Breiman (1996) lanean aurkeztutako bagging. Bootstrap laginak erabiliz hainbat zuhaitz eraiki eta sailkapenak bozketaz egiten da.

Sailkatzaile guztiak defektuzko parametroak erabiliz eraiki ditugu eta haien eraginkortasuna neurtzeko hamar iteraziodun balioztatze gurutzatua (10-fold cross-validation) metodoa erabili dugu.

2 Taula aztertuz hainbat ondorio atera ditzakegu. Lehenik eta behin, esan dezakegu, asmatze-tasei dagokionean, oro har antzekoak direla datu-base osorako eta K.Prako lortzen diren emaitzak. Hau da, problemak eta sailkatzaileak berak eragin handiagoa dutela, lokalizazioak baino. Sailkatzaile desberdinen errendimenduan kokatzen bagara, PART algoritmoa erabiliz lortutako erregela-multzoek lortzen dituzten asmatze-tasak gainerako algoritmoek lortzen dituztenekin alderatuta txikiagoak dira. Bestalde, azalpenik ematen ez duten sailkatzaileek, bagging eta NB, ez dute asmatze-tasa nabarmenki igotzea lortzen. Hori dela eta, hemendik aurrera baztertu egingo ditugu eta oro har, asmatze tasarik handienak lortzen dituen k C4.5ek (J48) sailkatzailean zentratuko gara.

2. taula. Galdera eta algoritmo desberdinentzat asmatze-tasak

DB	Galdera	C4.5	CTC	PART	Bagging	NB
Osoa	Neska, Mutila (ikaslea)	% 66.04	% 65.25	% 63.80	% 64.31	% 66.98
	Emakumea, EzEmak (irudia)	% 72.17	% 70.44	% 69.97	% 74.37	% 68.40
	Betaurrekoa Bai, Ez (irudia)	% 57.07	% 56.60	% 54.09	% 57.86	% 59.90
	Taldean, Bakarrik (irudia)	% 77.04	% 61.48	% 67.29	% 75.47	% 72.96
K.P.	Neska, Mutila (ikaslea)	% 61.73	% 68.52	% 66.05	% 63.58	% 64.81
	Emakumea, EzEmak (irudia)	% 70.37	% 69.75	% 66.67	% 69.14	% 67.28
	Betaurrekoa Bai, Ez (irudia)	% 56.17	% 54.32	% 54.32	% 53.70	% 61.11
	Taldean, Bakarrik (irudia)	% 75.93	% 54.32	% 70.37	% 72.84	% 64.81

Bestalde, galderen erantzunei dagokienean, asmatzen zailena, informatikariak betaurrekoekin ala betaurrekorik gabe irudikatzen ote dituzten da (ausaz adieraziko bagenu betaurrekorik baduten ala ez, asmatze-tasa % 50ekoa izango litzateke eta kasu honetan % 57koa da), eta errazena aldiz, informatikariek taldean ala bakarrik lan egiten duten iragarri nahi duena, % 77,44ko asmatze-tasarekin.

Emaitza hauek adierazten dute, egin diren galdera gehienetan badaudela desberdintasunak aukera bat eta bestearen artean eta gainera, galdeketa lortutako erantzunetan oinarrituta antzeman daitezkeela desberdintasun horiek. Azalpena lortzeko zuhaitzen egitura aztertu aurretik, azalpen sinpleagoak bilatzeko helburuarekin, eta ahal dela asmatze-tasa txikiagotu gabe, Garca *et al.* (2014) lanean gomendatzen den moduan, ikasketa automatikoaren ikuspuntutik esanguratsuenak diren aldagaiak automatikoki hautatzeko algoritmo bat erabili dugu: Hall (2000) laneko *Correlation-based Feature Subset Selection* (CFSS) algoritmoa non korrelazioan oinarriturik aldagai azpimultzo egokia aukeratu den. Asmatze-tasa mantentzea lortzen badugu, emaitza hobea izango da, aldagai gutxiagorekin lortutako sailkatzailea sinpleagoa izango baita.

3. taula. Galdera desberdinentzat asmatze-tasak CFSS aldagai-hautaketa egin ondoren

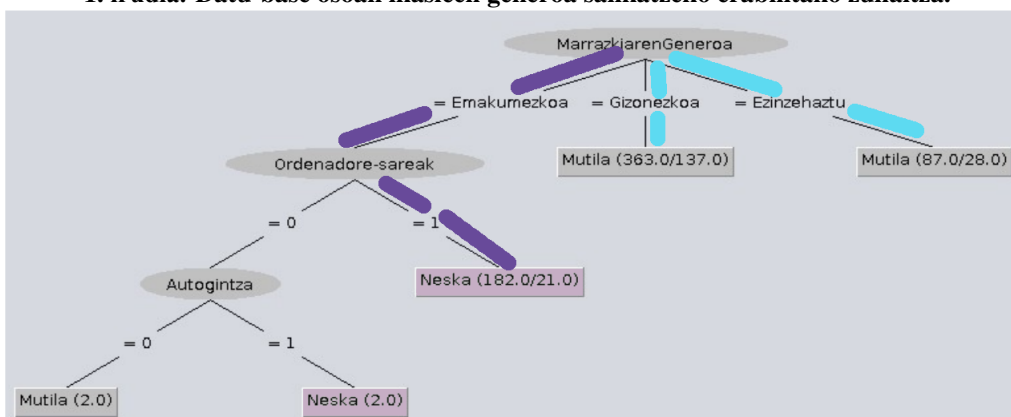
Galdera	C4.5	Hautatutako aldagaiak
		Osoa
Neska, Mutila (ikaslea)	% 70.44	aitaren ikasketa-maila, irudiaren generoa, autogintza, ordenagailu-sareak
Emakumea, EzEmak (irudia)	% 75.62	generoa, betaurrekoak, zuzenbidea
Betaurrekoa Bai, Ez (irudia)	% 56.76	eskola, adina, irudiaren generoa, ordenagailu-sareak
Taldean, Bakarrik (irudia)	% 77.04	eskola, jantzigintza, etxegintza
		K.P.
Neska, Mutila (ikaslea)	% 70.37	aitaren ikasketa-maila, lantokia
Emakumea, EzEmak (irudia)	% 74.70	urtea, generoa, betaurrekoak
Betaurrekoa Bai, Ez (irudia)	% 54.32	irudiaren generoa, artea, kimika, jantzigintza, ordenagailu-sareak
Taldean, Bakarrik (irudia)	% 76.54	generoa, amaren ikasketa-maila, autogintza

2 eta 3 tauletako emaitzak konparatzen baditugu zera ondoriozta dezakegu: asmatze-tasetan ez dago diferentzia handirik; izan ere, hainbat kasutan hobetu ere egiten da aldagai hautaketarekin, eta bestalde, eraiki ditugun sailkatzaileak askoz sinpleagoak izango dira. Menpeko aldagaia eta beste 20 aldagai erabiltzetik, hiru edo lau erabiltzera pasatu baikara. Hautatutako aldagaiak goi-mailan aztertuz hainbat ondorio orokor atera ditzakegu. Lehenik eta behin, esan dezakegu adin horietan haurren generoak dagoeneko baduela eragina gaztetxoek informatika eta informatikarion inguruan duten ikuspuntuarengan. Izan ere, kasu askotan ageri da aldagai hori hautatua. Estereotipoak guk uste baino lehenago iristen dira gaztetxoengana. Datu-base osoan egindako aldagai hautaketari begiratu esan dezakegu, haur hori hezi den inguruak ere eragiten duela bere ikuspuntuan. Izan ere, eskola aldagai gisa agertzen da informatikariaren lantokia deskribatzerakoan eta baita betaurrekoak jarri ala ez erabakitzerakoan.

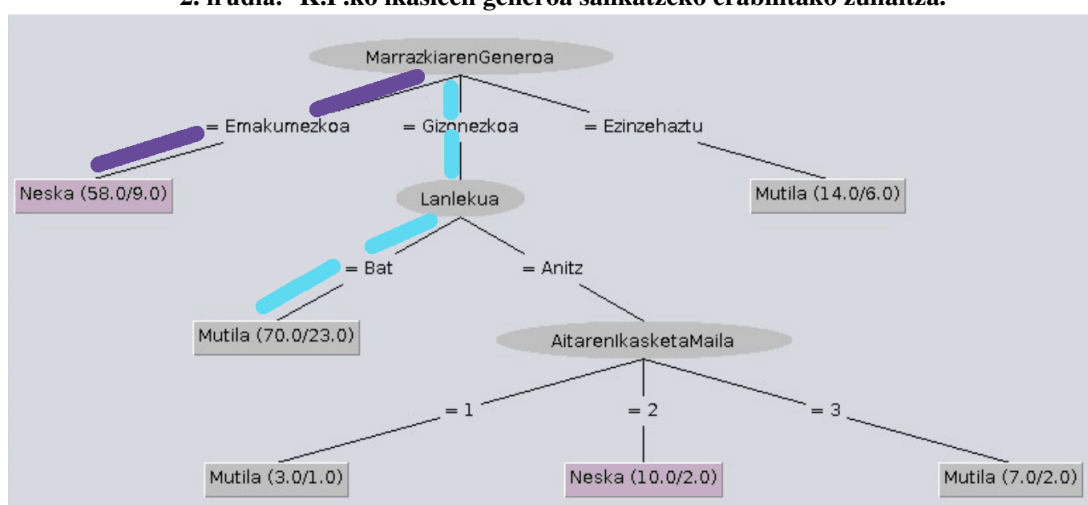
Aldagai horien eta menpeko aldagaiaren arteko erlazioari buruz gehiago jakin dezakegu eraiki ditugun zuhaitzen egiturari erreparatu. Adibide gisa, datu-base osorako eta K.P.rako sortutako bina sailkapen zuhaitz aztertuko ditugu: 1 eta 2 Irudietakoak ikasleen generoak bereizteko eraikitakoak, eta 3 eta 4 Irudietakoak, irudiaren generoa marrazteko eraikitakoak. Sailkapen zuhaitz horien adabegi bakoitzak klase bat adierazten du eta errotik adabegietarainoko bideak aldiz, egindako sailkapenari dagokion azalpena.

1 eta 2 Irudiak aztertuz, esan liteke datu-base osoan zein K.P.n, ikasleen generoa determinatzeko garrantzia gehien duen aldagaia, hauek marrazkiari ematen dioten generoa dela. Nola nahi ere, neskek (morez markatuak)

1. irudia. Datu-base osoan ikasleen generoa sailkatzeko erabilitako zuhaitza.



2. irudia. K.P.ko ikasleen generoa sailkatzeko erabilitako zuhaitza.



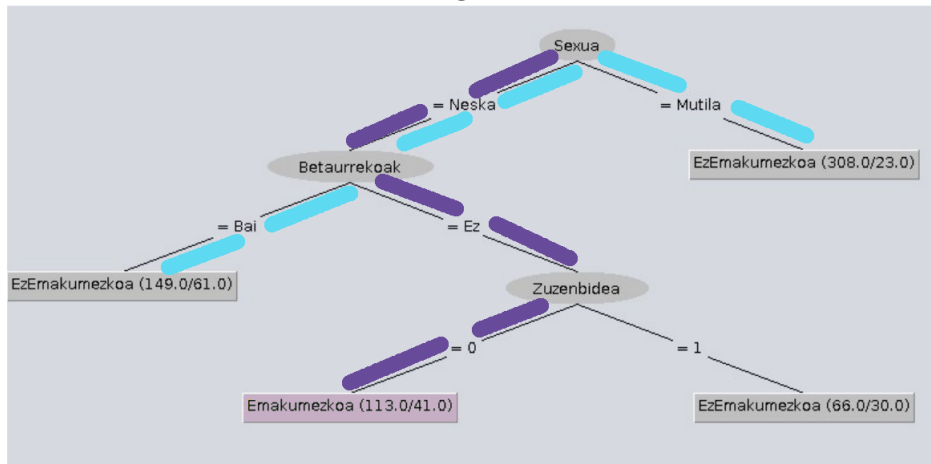
eta mutilak (urdinez markatuak) bereizteko lortzen diren erregela nagusiak, edo kasu gehien biltzen dituztenak, ez dira berdinak bi kasuetan. Datu-base osoaren kasuan ondoko bi erregela nagusia ondoriozta genitzake: (1) neskek dira informatikaria emakumezko marraztu eta ordenagailu-sareetan informatika badagoela diotenak (2) mutilak dira informatikaria gizonetzko edo definitu gabe marrazten dutenak. K.P.ko kasuan aldiz, lehenengo erregela berdin errepikatzen da baina bigarrenak, aldaketa batzuk ditu, (2) mutilak dira informatikaria definitu gabe marrazten dutenak eta gizonetzko marrazteaz gain, lantokia pertsona bakarrek marrazten dutenak. Badago aukera gizonetzko marrazten dutenen artean neskek identifikatzeko.

3 eta 4 Irudietako zuhaitzak aztertzen baditugu, informatikaria emakumezko gisa zeren arabera marrazten duten ondorioztatu ahal izango dugu. Datu-base osoa kontutan hartuz, (1) informatikaria emakume imajinatzen dute neskek baldin eta betaurrekorik gabe imajinatzen badute eta zuzenbidean ez dela informatikarik erabiltzen pentsatzen badute. Bestalde, (2) informatikaria ez dute emakume imajinatzen mutilek eta betaurrekoekin irudikatzen duten neskek. K.P.n aldiz, badirudi taldeak edo urteak eragina duela. Kasu honetan, neskek 2016 eta 2018. urtean informatikaria emakume irudikatu zuten baina ez aldiz, 2017. urtean. Mutilek aldiz, ez dute emakumezkoa imajinatzen informatikaria.

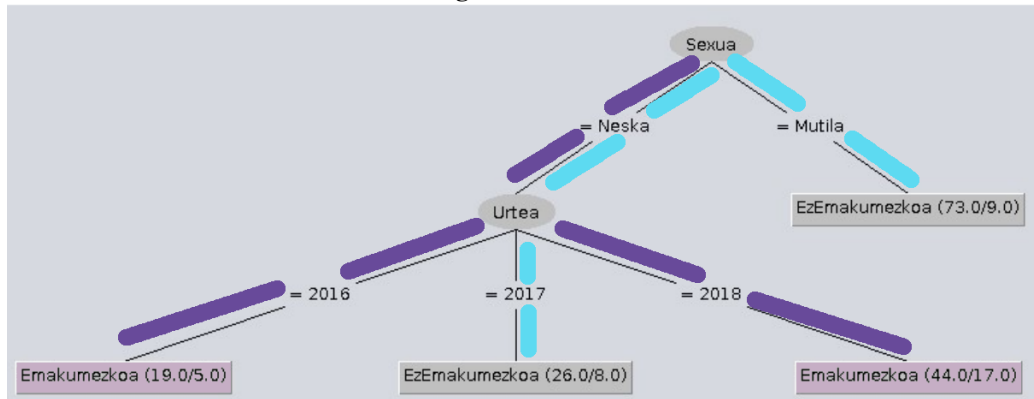
4. Ondorioak eta etorkizuneko lana

Motibazioa informatika-ikasketetara iristen diren emakumeen kopuru urriaren jatorria aztertzea izanik, lan honetan hiru urtetan zehar Euskal Herriko hainbat eskoletan 10-12 urte arteko gaztetxoei egindako inkesta batzuk baliatu ditugu, adin horretan haiek duten iritzia aztertzeko. Datu-meatzaritzako prozesu oso bat egin behar izan

3. irudia. Datu-base osoan irudiaren generoa sailkatzeko erabilitako zuhaitza.



4. irudia. K.P.ko irudiaren generoa sailkatzeko erabilitako zuhaitza.



dugu horretarako: datuak batu, aurreprozesatu, erantzun nahi genituen galderak erantzuteko prestatu, sailkatzailerekin erabiltzeko egokitu, sailkatzailerekin eraiki eta aztertu. Prozesu horri esker, jasotako inkestetan parte hartu duten ikasleei buruz hainbat ondorio atera ditugu. Lehenik eta behin, neskek eta mutilek informatika eta informatikariei buruz duten ikuspuntua ez da berdina adin horretan eta % 70eko asmatze-tasarekin bereizteko gai gara. Asmatze-tasa hori zerbait igotzen da haurrek informatikariari irudikatzen dioten generoa eta lanean bakarrik edo taldean imajinatzen ote duten asmatzen saiatzen garenean, eta jaitsi aldiz, betaurrekoekin irudikatzen ote dituzten asmatzen saiatzen garenean. Horrez gain, aldagai-hautaketa eta sailkapen zuhaitzak aztertuz, esan dezakegu haurren bizi-inguruneak (eskolak) eta taldeak baduela eragina haurrek gai honen inguruan duten iritzia gainean. Azkenik, bereziki nabarmentzekoa iruditzen zaigu lortutako emaitzen artean, badirudiela gure inguruko 10-12 urteko mutilek oro har gizonezkoak edo identifikatu ezinezko generoa duten gizakiak irudikatzen dituzten arren, neskek oraindik hein handi batean ikusten dutela informatikaria neska izan daitekeela. Emaitza honek adierazten digu, adin-tarte horretarako prestatutako ekintzek eragina izan dezaketela etorkizunean informatika ikasiko duten emakume kopuruarengan.

Datuen beste azterketa bat egitea aurreikusten dugu lehenengo etorkizuneko lan gisa. Gainbegiratutako ikasketa automatikoko prozesua hasieran planteatu ditugun galderak baldintzatu dute. Baliteke, hori horrela izanik, datuetan benetan existitzen ez zen egitura bat bilatzen aritu izana. Horregatik, aurreikusten dugu clustering algoritmoen bidez datuen egitura naturala topatu eta talde bakoitzaren azterketa egitea, taldeko adibideen antzekotasun nagusiak zeintzuk diren ondorioztatzeko.

Erreferentziak

- Breiman, Leo. 1996. Bagging predictors. *Mach. Learn.* 24.123–140.
- Brown, Neil, Sue Sentance, Tom Crick, eta Simon Humphreys. 2014. Restart: The resurgence of computer science in uk schools. *ACM Transactions on Computing Education (TOCE)* 14.9.
- Chambers, David W. 1983. Stereotypic images of the scientist: The draw-a-scientist test. *Science education* 67.255–265.
- Clifton, Christopher, 2017. Data mining.
- Code.org. Code.org: About us. <https://code.org/about>. [Online; 2019ko Otsailaren 13an eskuratua].
- . The hour of code. <https://hourofcode.com/es/en>. [Online; 2019ko Otsailaren 13an eskuratua].
- Dagiene, Valentina. 2002. The model of teaching informatics in lithuanian comprehensive schools. *Journal of Research on Computing in Education* 35.176–185.
- DeRuy, Emily, 2017. In finland, kids learn computer science without computers. <https://www.theatlantic.com/education/archive/2017/02/teaching-computer-science-without-computers/517548/>. [Online; 2019ko Otsailaren 13an eskuratua].
- Frank, Eibe, eta Ian H. Witten. 1998. Generating accurate rule sets without global optimization. *Generating Accurate Rule Sets Without Global Optimization* 144–151. cited By 46.
- Frieze, Carol, eta Jeria L. Quesenberry. 2019. How computer science at cmu is attracting and retaining women. *Communications of the ACM* 62.23–26.
- Garca, Salvador, Julian Luengo, eta Francisco Herrera. 2014. *Data Preprocessing in Data Mining*. Springer Publishing Company, Incorporated.
- Guzdial, Mark. 2015. Learner-centered design of computing education: Research on computing for everyone. *Synthesis Lectures on Human-Centered Informatics* 8.1–165.
- Hall, Mark A. 2000. Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, 359–366, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- John, George H., eta Pat Langley. 1995. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI'95*, 338–345, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Knight, Meredith, eta Christine Cunningham. 2004. Draw an engineer test (daet): Development of a tool to investigate students' ideas about engineers and engineering. In *ASEE Annual Conference and Exposition*, volume 2004.
- Lewis, Colleen M. 2017. Good (and bad) reasons to teach all students computer science. In *New Directions for Computing Education*, 15–34. Springer.
- Medel, Paola, eta Vahab Pournaghshband. 2017. Eliminating gender bias in computer science education materials. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education*, 411–416. ACM.
- Pérez, Jesús M., Javier Muguerza, Olatz Arbelaitz, Ibai Gurrutxaga, eta José I. Martín. 2007. Combining multiple class distribution modified subsamples in a single tree. *Pattern Recogn. Lett.* 28.414–422.
- Quinlan, J. Ross. 1993. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Witten, Ian H., Eibe Frank, eta Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 3rd edition.

5. Eskerrak eta oharrak

Lan honek finantzaketa jaso du Eusko Jaurlaritzako hezkuntza saileko unibertsitateak eta ikerketa atalak sustenga-tzen duen ADIAN ikerketa taldetik (erreferentzia IT980-16).