

**Naiara AGINAKO BENGEOA**

Donostiako UPV-EHUko Informatika Fakultateko irakasle eta ikertzailea

Aro digitalean, datuak dira urre berria

Munduan zehar gertatzen ari den digitalizazio prozesu honetan, datuak dira egunero erabiltzen dugun teknologiak

bultzatzen dituen erregaiak. Datuak egungo gizarteko urre berria bihurtu dira, negozio eta erakundeentzako aktibo baliotsua, alegia. Azken urteetan hain gora egin duen adimen artifizialak duen datuen beharrak bultzatu du datuen nagusitasun hau. Adimen artifizialeko sistema batek irensten dituen datuen kalitatea eta kantitatea funtsezkoak dira bere eraginkortasunerako. Funtsean, datuak dira adimen artifizialeko algoritmoen elikagai; hau da, sistemei ikasteko, egokitze eta erabakiak hartzeko aukera ematen dietenak.

Horregatik bilakatu da hain garrantzitsua datuak bildu eta gordetzea. Baina batzuetan errealitatean jaso daitezkeen datu horiek guztiak ez dira nahikoa. Zer gertatzen da, adibidez, osasun arloari buruz hitz egiten dugunean? Posible al da datu horiek guztiak erabiltzea gaixoen pribatutasunaren gaitetik? Eredu batek sortu dituen datuak erabiltzeak sortutako arazo honi datu sintetikoek erantzun bat emango liokete. Datu sintetikoak datu errealean ordez erabiltzeko helburuarekin sortu ziren, datuak sortzeko prozesua kontrolatuz, datu errealean antz handiena izatea bermatzeko intentzioarekin.

Osasun arloan jarraituz, datu sintetikoek, pribatutasun-kezkei aurre egiten laguntzeaz gain, badituzte beste erabilera batzuk. Demagun, munduan oso arraroa den gaixotasun bati bu-

ruz ari garela. Zaila da gaixotasun horri buruzko datu erreal asko jasotzea, eta ondorioz, gaixotasun horren eredu izan daitezkeen datu sintetikoak sortzeak gaixotasuna hobeto ezagutzeko aukera emango du.

Baina datu sintetikoek domeinu desberdin askotan sortzen diren beste arazoetan ere lagundu dezakete. Adibidez, gidatze autonomoan jartzen badugu arreta, gaur egun oso zaila da egunerokoan ager daitezkeen egoera guztien datu errealak berreskuratzea. Alde batetik, gidatze autonomoa ikertzen ari diren taldeek ez dute normalean kotxe autonomo bat baino gehiago izaten, garestia delako, eta, gainera, ez da oso bideragarria trafikoaren agertoki desberdinak modu errealean martxan jartzea. Horregatik, simulatzaileen bidez, eman daitezkeen agertoki hipotetiko sin-esgarriak sor daitezke, eta bertan sortutako datu sintetikoak erabili ikerketarako. Are gehiago, datu sin-

tetiko horiek kontrolpeko egoera batean sortu direnez, datu horiek zuzenean etiketatuta agertzen dira, eta honek asko errazten du teknologia berrien garapena.

Goian aipatutako arrazoiez gain, badaude beste arrazoi batzuk ere datu sintetikoak sortzeko. Baina kasu guztietan, datuen sorrera prozesu horretan hauen erabilgarritasuna kontuan hartzea ezinbestekoa da. Datu sintetikoak benetako datu multzoen kalitatezko ordezkoak izan behar dira. Kalitate hori bermatzerik ez badago, datu sintetiko horiek edukitzeak egin dezakeen kaltea handia izango baita.

Datuak aro honetako urre izanik, datu sintetikoen inguruan ere negozio handia sortu da. Gero eta enpresa gehiago agertzen ari dira datu sintetikoak sortzen dituztenak edo horretarako aplikazioak garatzen ari direnak. Merkatu honek hurrengo hamarkadan izango dituen diru sarrerek abiadura bizian egingo dute gora. •



Funtsean, datuak dira adimen artifizialeko algoritmoen elikagai. GETTY IMAGES