



4.3 AA eta datuak: erlazio sinbiotiko bat

Adimen artifiziala eta datu-meatzaritza elkarri eragiten dioten eta elkarren osagarri diren bi diziplina dira. **Datu-meatzaritza datu-bolumen handietatik patroiak, ezagutza eta informazio esanguratsua ateratzeko prozesua da.** Patroi eta ezagutza horiek, era berean, AA-ko sistemek, bereziki *Machine Learning*-en oinarritutakoek, ikasteko, erabaki adimendunak hartzeko eta iragarpenak egiteko erabiltzen duten erregaia dira. Teknologia horiek guztiek aukera ematen diete enpresei, datu multzo handiak aztertzeaz gain, prozesu konplexuak automatizatu eta optimizatzeko, erabaki oinarrituak hartzeko, eraginkortasun operatiboa hobetzeko eta balioa sortzeko sektore askotan.

Datuekin lan egiteko prozesua

Datu-meatzaritzako proiektu batean datuekin lan egiteko prozesuak, oro har, funtsezko hainbat etapa izaten ditu:

1. **Datuak biltzea:** Lehen urratsa da eta hainbat iturritatik datuak lortzean datza. Datu horiek **egituratuak** izan daitezke (tauletan antolatuak, datu-baseetan edo kalkulu-orrietan bezala) edo **ez egituratuak** (hala nola, testu librea, irudiak, bideoak, audioa). Bildutako datuen kalitatea eta garrantzia funtsezkoak dira, AA ereduak informazio zehatza eta adierazgarria behar baitute behar bezala ikasteko eta alborapenak saihesteko. Konpondu nahi den arazorako datu egokiak identifikatu eta bildu behar dira. Aurretiko fase erabakigarri bat, batez ere datu-meatzaritzan, **Negozioaren Ulermena** da, non enpresaren beharrak eta helburuak ulertzen diren proiektua estrategiarekin lerrokatzeko.
2. **Datuen aurreprozesamendua:** Bildu ondoren, datuak oso gutxitan daude zuzenean erabiltzeko prest. Aurreprozesamenduak datuak garbitzea eta eraldatzea dakar. Besteak beste, balio nuluak edo falta direnak maneiatzea, datu bikoiztuak edo okerrak ezabatzea, datuak normalizatzea (eskala komun batera egokitzea) eta ezaugarriak eraldatzea (adibidez, datu kategorikoak zenbakizko bihurtzea), algoritmoek modu eraginkorrean lan egin dezaten. Prozesu hori funtsezkoa da ereduak datu okerrak edo partzialak ikas ez ditzan eta bere errendimendua hobetzeko. Hemen ere **Datuen ulermena** egiten da, hasierako patroiak identifikatzeko eta transformazioak erabakitzeko. Datuak prestatzeko, ezaugarri garrantzitsuenak hautatu behar dira.



3. **Modelatzea eta entrenatzea:** Etapa honetan, AA-ren eredua sortu eta entrenatzen da. *Machine Learning* algoritmo egokia hautatzen da arazo motarako eta eskuragarri dauden datuetarako (adibidez, kategoriak iragartzeko sailkapen-algoritmo bat, edo zenbakizko balioak iragartzeko erregresio-algoritmo bat). Ereduaren parametroak doitzen dira aurrez prozesatutako datuak erabiliz (normalean entrenamendu multzo batean eta proba/balidazio multzo batean banatuta), iragarpenetan errorea minimizatzeko helburuarekin. Fase hau prozesu iteratibo bat izan ohi da, algoritmo edo konfigurazio desberdinak probatuz doitasuna hobetzeko.
4. **Ebaluazioa eta balidazioa:** Behin entrenatu ondoren, ereduaren zehaztasuna eta errendimendua egiaztatu behar dira. Horretarako, modeloak entrenamenduan ikusi ez dituen probako datu-multzo bat erabiltzen da. Ereduak datu berriekin nola jokutzen duen ebaluatzen da, eta haren eraginkortasuna neurtzeko metrika egokiak erabiltzen dira (adibidez, zehaztasuna, *recall*, *F1-score* sailkapenean). Ereduak helburuak betetzen ez baditu, modelatze-fasera edo aurreprozesamendu-fasera itzultzen da.
5. **Hedapena (Implementazioa):** Eredua egokitzen jotzen bada, produkzio-ingurunean inplementatzen da, mundu errealean erabakiak hartzeko edo zereginak automatizatzeko erabili ahal izateko. Horrek esan nahi du eredua dauden negozio-prozesuetan integratu behar dela.
6. **Iterazioa eta etengabeko hobekuntza:** AA-ren ereduak ez dira estatikoak. Etengabe monitorizatu, doitu eta birentrenatu behar dira datu berriekin, ingurunean edo datu-ereduetan gertatzen diren aldaketetara egokitzeko eta denboran zehar errendimendu optimoa mantentzeko.

Datu-bolumen handiak (*Big Data*)

Datu-bolumen handiak (*Big Data*) erabiltzea ohikoa da AA-ean. Horrek azpiegitura teknologiko egokiak eskatzen ditu, hala nola hodeiko konputazioa (eskalagarritasuna eta malgutasuna eskaintzen ditu) eta biltegiatze banatuko sistemak, datu masiboak modu eraginkorrean prozesatu eta biltegiatzeko.

AA-rako datuen tratamenduan, zenbait alderdi kritiko hartu behar dira kontuan:



- **Datu pertsonalen pribatutasuna eta babesa:** Funtsezkoa da erabiltzaileen datuak, bereziki sentikorrak badira (hala nola, datu medikoak edo finantzarioak), babestuta daudela eta etikoki eta indarrean dauden araudien arabera erabiltzen direla bermatzea (adibidez, DBEO).
- **Ekitatea eta alborapena:** Ereduak entrenatzeko erabiltzen diren datuek joera historiko edo sozialak izan ditzakete. Ez badira jorratzen, ereduak ikasi eta anplifikatu egin ditzake joera horiek, eta zenbait talderen aurkako erabaki bidegabeak edo diskriminatzaileak har daitezke. Funtsezkoa da joera horiek identifikatzea, arintzea edo ezabatzea.
- **Interpretazioa eta azalpena (gardentasuna):** AA-ren eredu askok, batez ere *Deep Learning*-ekoek, "kutxa beltz" gisa funtziona dezakete, eta zaila da ulertzea nola iristen diren erabaki jakin batera. Interpretagarritasuna eta azalgarritasuna garrantzitsuak dira konfiantza sortzeko eta ereduak arazteko eta hobetzeko, bereziki sektore kritikoetan, hala nola medikuntzan edo finantzetan, erabakiak justifikagarriak izan behar baitute.
- **Erantzukizuna eta kontuak ematea:** Argi eta garbi zehaztu behar da nor den AA-ko sistema baten emaitzen eta erabakien erantzulea, bereziki kalterik eragiten badu. Modeloen errendimendua ikuskatzeko eta erabaki automatizatuak zuzenak eta bidezkoak direla ziurtatzeko mekanismoak behar dira.

Datu-meatzaritza

Azkenik, datu-meatzaritza hainbat motatan sailka daiteke, helburuaren arabera:

- **Deskribatzailea:** Bere helburua datuak aztertzea da, datuen ezaugarri nagusiak laburbiltzen dituzten patroiak, joerak eta deskribapenak aurkitzeko. Zer gertatu den ulertzen laguntzen du.
- **Iragarlea:** Datu historikoak erabiltzen ditu etorkizuneko emaitzak edo portaerak aurreikus ditzaketen ereduak eraikitzeko. Adibidez, bezero batek zerbitzu bat uzteko probabilitatea aurreikustea edo etorkizuneko salmentak aurreikustea.
- **Asoziatiokoa:** Datu-multzo batean aldagai edo item desberdinen arteko loturak edo lotura-arauak aurkitu nahi ditu. Adibide klasiko bat erosketa-saskiaren analisisa da, elkarrekin erosi ohi diren produktuak identifikatzeko (adibidez, "pixoihalak erosten dituzten bezeroek ere garagardoa erosten dute").
- **Sekuentziala:** Gertaeren edo aldi baterako datuen sekuentziak aztertzen ditu, denboran zehar ordena espezifiko batean gertatzen diren patroiak identifikatzeko. Adibidez, web bateko klik sekuentziak edo banku-transakzioetako patroiak aztertzea.



- **Espezializatua:** Datu mota espezifikoei aplikatutako teknikak barne hartzen ditu:
 - **Testu-datuak:** Testu-bolumen handien analisia, informazioa ateratzeko, gaiak identifikatzeko, sentimenduak edo iritziak aztertzeke.
 - **Datu espazialak:** Datu geografiko edo espazialen analisia, eskualdeko joerak, banaketa-ereduak edo kokapenean oinarritutako harremanak identifikatzeko.
 - **Sare sozialak:** Sare sozialetako plataformetako interakzio sozialen, komunitateen eta eragin-fluxuen informazioa eta patroiak ateratzea.
 - **Denborazko datuak:** Sekuentzialaren antzekoa, baina iragarpenetarako, anomalien detekziorako eta abarretarako denbora-serieak aztertzeraz bideratua.

Prozesu eta alderdi horiek ulertzea ezinbestekoa da sistema arduratsuak eta eraginkorrak garatzeko eta ezartzeko.

Ondorioa

Azken batean, adimen artifiziala eta datu-meatzaritza ez dira soilik elkarri lotutako eremuak, baizik eta sinbiosi ahaltsua eta binomio banaezina, erabakiak hartzea eta enpresaren eraginkortasuna birdefinitzen ari dena. Datu-meatzaritzak patroiak eta ezagutza esanguratsua ateratzeko duen gaitasuna, prozesu zorrotz baten bidez, bilketa eta aurreprozesamendutik hasi eta modelatu eta hedatzeraino, funtsezko erregaia da AA-ko sistemei ikasteko, modu adimentsuan erabakitzeko eta sektore askotan berrikuntza eta eraginkortasuna bultzatzeko aukera ematen diena. Hala ere, potentzial hori modu jasangarrian eta benetan onuragarrian gauzatu dadin, funtsezkoa da berezko erronka etiko eta teknikoei seriotasunez heltzea, hala nola, datuen pribatutasunari, ekitate algoritmikoari eta gardentasunaren beharrari. Elementu horiek ulertzea eta kudeatzea, beraz, ezinbestekoa da ahaltsuak eta eraginkorrak ez ezik, bidezkoak, arduratsuak eta gizartearen aurrerabidearekin bat datozenak ere garatu ahal izateko.

(azken eguneratzea 2025/06/04)

Eusko Jaurlaritzaren Lanbide Heziketako Sailburuordetza. Lan honek Creative Commons Aitortu-EzKomertziala-PartekatuBerdin 4.0 Nazioarteko Baimena dauka (CC BY-NC-SA 4.0).

