

5.2 DATUAREN BIZITZA ZIKLOA

Sarrera

Egungo gizartean, datuak baliabide estrategiko eta funtsezko bihurtu dira erabakiak hartzeko, prozesuak optimizatzeko eta arazoei aurrea hartzeko. Horien garrantzia ulertzeko, beharrezkoa da jakitea nola kudeatzen diren datuak denboran zehar, sortzen direnetik ezabatzen diren arte. Prozesu horri **datuaren bizi-zikloa** esaten zaio.

Ziklo honek datu bat sortzen denetik behin betiko ezabatzen den arte igarotzen den etapak deskribatzen ditu. Prozesu osoa da, eta barnean hartzen ditu datua biltzea, prozesatzea, biltegitratzea, aztertzea, interpretatzea, erabiltzea eta ezabatzea. Gainera, ekintza horiek kalitate-, segurtasun- eta eraginkortasun-printzipioei jarraiki egitea ahalbidetzen du, legezko arauetara eta arau etikoetara egokituta.



1. Datua atzitzea edo sortzea

Zikloa datua atzematen hasten da, eta eskuz (inkestaren edo formularioen bidez) edo automatikoki (sentsoreen, aplikazioen edo gailu digitalen bidez) gerta daiteke. Informazio horren jatorria pertsonak, makinak edo sistemak izan daitezke.

Adibidez: Norbait web orri batean erregistratzen denean, bere izena, posta elektronikoa eta pasahitza sartzen ditu. Formulario horrek datu pertsonalen multzo bat sortzen du.



2. Biltegitratzea

Sortu ondoren, datuak modu seguruan eta eskuragarrian biltegitratu behar dira. Hau bai *on premise* (lokalean) edo *on cloud* zerbitzarietan egin daiteke. Biltegitratzeak datuen osotasuna eta erabilgarritasuna bermatu behar ditu, datuak babesteko arauak betetzeaz gain.

Fase honetan funtsezkoa da Europar Batasunean DBEO (Datuak Babesteko Erregelamendu Orokorra) edo Espainian DBLO (Datuen Babeserako Lege Organikoa) bezalako araudiak betetzea.

Adibidea: aplikazio batean erregistratutako erabiltzaileen datuak MySQL edo MongoDB bezalako datu-baseetan gordetzen dira, baimendutako kredentzialen bidez soilik eskura daitezkeenak.

3. Prozesamendua

Hemen datuak antolatu, garbitu eta eraldatu egiten dira, errazago aztertzeko. Horren barruan sartzten dira bikoiztuak ezabatzea, balio nuluak edo okerrak tratatzea eta formatuak bihurtzea. Datuen normalizazioaz ari gara. Helburua da datuak informazio erabilgarri eta fidagarri bihurtzea.

Prozesatzeko tresnek analisi estatistikorako softwarea, Python edo R bezalako lengoaiak eta Apache Spark edo Excel bezalako plataformak izan ditzakete.

Adibidez: banaketa-enpresa batek entrega-erregistroak aztertzen ditu. Datu errepikatuak ezabatzen dira, eta oker idatzitako helbide batzuk zuzentzen dira, logistika hobetzeko.

4. Analisia

Analisiak aukera ematen du datuetatik balioa ateratzeko, patroiak, joerak edo anomaliak identifikatzeko. Teknika estatistikoak, iragartzeko ereduak eta bistaratzeak erabiltzen dira informazioa errazago interpretatzeko. Etapa hau funtsezkoa da datuetan oinarritutako erabakiak hartzeko.

Adibidea: Supermerkatu batek aztertzen du zer produktu saltzen diren gehien oporretan, hurrengo urteko inbentarioa planifikatzeko.

5. Interpretazioa eta bistaratzea

Analisiaren emaitzak argi eta garbi komunikatu behar zaizkie erabiltzaileei edo erabakiak hartzeko arduradunei. Hori txostenen, dashboard-en, bistaratze grafikoen edo aurkezpenen bidez egiten da. Datuaren komunikazio eraginkorrak informazioa ezagutza bihurtzeko aukera ematen du.



Adibidea: Marketin sailak hilero egiten duen bilera batean eskualde bakoitzeko erabiltzaileen hazkundera erakusten duen dashboard bat aurkezten du, aztertutako datuetan oinarrituta.

6. Erabakiak hartzea

Datuen interpretazioa jarduteko oinarri gisa erabiltzen da. Erabakiak estrategikoak, operatiboak edo taktikoak izan daitezke, eta analisitik ateratako ebidentzietan oinarritu behar dira. "Data-driven decision making" (datuetan oinarritutako erabakiak hartzea) kontzeptua funtsezkoa da erakunde modernoetan.

Adibidea: Unibertsitate batek ikasle askok ikasgai jakin bat uzten dutela detektatzen du, eta irakaskuntza-plana berrikustea eta laguntza akademiko gehigarria eskaintzea erabakitzen du.

7. Mantentzea eta eguneratzea

Datuak antolatuta eta eguneratuta mantendu behar dira, etorkizuneko neurketekin alderatu nahi badira.

8. Artxibatzea eta kontserbatzea

Erabili ondoren, datuak gorde egin daitezke, lege-arrazoiengatik, arrazoi historikoengatik edo etorkizuneko azterketa-arrazoiengatik. Fase honetan, atxikipen-politikak ezarri behar dira, eta datuek zenbat denboraz egon behar duten erabilgarri zehaztu behar da. Era berean, garrantzitsua da hondamendien aurrean babes-sistemak (*backup*-ak) eta berreskuratze-sistemak ezartzea, datu garrantzitsuetarako sarbidearen jarraitutasuna ziurtatzeko.

Adibidea: Ospitale batek hainbat urtez gordetzen ditu bere pazienteen historia medikoak, indarrean dagoen osasun-araudiaren arabera.

9. Ezabatzea edo araztea

Datuak jada ez direnean erabilgarriak edo lege ezin badira gordeta mantendu, modu seguruan ezabatu behar dira. Horrek pertsonen pribatutasuna babesten du, eta baimenik gabe sartzeko arriskua murrizten du. Ezabatzeko, ezabatze logikoko edo suntsitze fisikoko teknikak erabil daitezke, euskarri motaren arabera.

Adibidea: Enpresa batek sei hilabeteren ondoren lanpostu baterako baztertutako hautagaien curriculumak ezabatzen ditu, datuak babesteko duen politikaren arabera.

Datuaren bizi-zikloa ez da lineala: erakunde askotan, datuak aurreko etapetara itzul daitezke, berrerabili edo informazio-multzo berri bihurtu. Horregatik, funtsezkoa da **datua kudeatzeko estrategia integral bat** ezartzea, arau etiko, legal eta teknikoak jasoko dituen, haren kalitatea eta segurtasuna bermatzeko.



Datuen analisia eta datu-zientzia

Datuen analisiak zer gertatu den eta zergatik ulertzea du helburu. Eskuragarri dauden datuetatik abiatuta, egoerak azaltzen edo galdera espezifikoei erantzuten saiatuko gara. Adibidez, fabrika batek azter dezake zenbat unitate ekoizten diren egunero, zenbat denboran egon diren geldirik makinak edo zenbat produktuk izan dituzten akatsak. Informazio horri esker, prozesuaren egungo egoera jakin daiteke, baina ez da beharrezkoa etorkizunean zer gertatuko den aurreratzea.

Datuen zientziak, aldiz, urrats bat harago egiten du. Diziplina arteko diziplina da, estatistika, informatika eta matematika uztartzen baititu. Horiekin, aztertu ez ezik, gertatuko dena iragarri eta ekintzak iradokitzen ditu. Datuen zientzia hainbat teknikatan oinarritzen da, hala nola, **datu-meatzaritzan, ikaskuntza automatikoan edo adimen artifizialean,** eta, horrela, erabaki konplexuak hartzen laguntzen duten eredu prediktiboak garatzen dira.

Adibidez, aurreko adibidean aipatutako fabrikan, datu-zientzia ez da gertatzen dena deskribatzera mugatuko; aitzitik, makinen geldialdiak noiz eta zergatik gertatzen diren aztertuko du, aldagai kritikoak identifikatuko ditu (esaterako, makinen tenperatura) eta akatsak noiz errepika daitezkeen aurreikusteko ereduak garatuko ditu. Informazio horrekin, mantentze-lanetan aldaketak egitea edo prozesuak doitzea iradoki ahal izango da, etorkizunean arazorik ez izateko. Sistemak zenbait arriskuren aurrean eman behar duen erantzuna automatizatzea ere ahalbidetzen da, eta hori pentsaezina izango litzateke analisi tradizional batekin.

Azkenik, azpimarratu behar da datuen analisia datu-zientiaren parte dela, baina azken horrek ikuspegi zabalagoa ematen du, inplementazioa, monitorizazioa eta eredu automatizatuen etengabeko hobekuntza bezalako etapak barne hartzen baititu.

Ondorioa

Bai datuen analisia bai datuen zientzia funtsezkoak dira informazioaren erabilera modernoan. Lehenengoak zer gertatu den ulertzeko aukera ematen duen bitartean, bigarrenak gerta daitekeenerako prestatzen gaitu. Bi kasuetan, beharrezkoa da datuaren bizi-zikloa behar bezala kudeatzea; izan ere, hori gabe, diziplina horietako bat ere ezingo litzateke eraginkortasunez aplikatu.

(azken eguneraketa: 2025/05/21)

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

