

Kursus i algoritmiske grundprincipper og etik i AI: fra teori til praksis

Værktøjer til synkrone sessioner

CU3 | Algoritmer og deres begrænsninger
Casestudie

Casestudie



Casestudie

Beskrivelse til undervisere

	Beskrivelse
Beskrivelse af opgaven	Deltagerne skal analysere en fiktiv case (se næste slide) om en finansiel teknologivirksomhed (FinTech Innovations), som brugte et algoritmisk system til at vurdere kreditværdighed til godkendelse af lån. Virksomhedens mål var at tilbyde hurtige og effektive lån til en bredere demografisk gruppe. Det blev dog senere opdaget, at algoritmen konsekvent favoriserede ansøgere fra bestemte indkomstgrupper og racemæssige baggrunde, hvilket resulterede i en diskriminerende lånepraksis. Deltagerne skal undersøge de grundlæggende årsager til dette problem, evaluere konsekvenserne for de berørte samfundsgrupper og foreslå effektive afhjælpningsstrategier.
Beskrivelse af, hvordan opgaven skal løses	Gruppearbejde med 3-4 personer. Research og baggrundsanalyse - Gennemgå kursusmaterialet og litteraturen om algoritmisk bias og retfærdighed. - Forstå de specifikke faktorer, der bidrog til FinTech Innovations' diskriminerende resultater, herunder databias, anvendte funktioner og fraværet af fairness-metrikker. Identifikation af grundlæggende årsager - Analyser de måder, hvorpå FinTech Innovations' træningsdata og funktioner kan have indlejret samfundsmæssige bias. - Undersøg fraværet af fairnessmålinger og manglen på input fra interessenter. Anbefalinger til afhjælpning Udvikl nogle anbefalinger til at håndtere de identificerede bias, herunder varierede data, nye valideringsprocedurer, etiske retningslinjer og løbende revision. Udarbejdelse og aflevering Strukturer resultaterne i en rapport med et klart resumé, en analyse af de grundlæggende årsager og detaljerede afhjælpningsstrategier. Inkluder billeder, grafer eller tabeller, hvor det er relevant for at understøtte dine pointer.
Anslået tid til at løse opgaven	30-60 minutter.
Forslag til kilder til at løse opgaven	Kursusmateriale, akademiske informationsdatabaser, GAI-systemer (f.eks. ChatGPT, Copilot osv.) og internetsøgninger.
Detaljeret beskrivelse af, hvordan opgaven skal løses	Skriv en sammenfattende rapport på 500-750 ord.
Information om deadline for aflevering af opgaven	Under den synkrone session.
Kontaktoplysninger eller hvordan man afklarer tvivlsspørgsmål	Læreren skal angive en form for kontaktmulighed (det kan være en e-mailadresse, et telefonnummer eller lign.)

Casestudie

Fiktiv case for studerende

FinTech Innovations er kendt for sin innovative tilgang til at levere hurtige og effektive finansielle services, ofte gennem brug af algoritmiske systemer til vurdering af kreditværdighed og godkendelse af lån. Denne fiktive case drejer sig om en FinTech-virksomhed, der brugte en avanceret kreditvurderingsalgoritme til at vurdere ansøgere og strømline sin lånegodkendelsesproces. Virksomhedens mission var at demokratisere adgangen til finansielle ressourcer og støtte kunder, som fik en dårligere service hos traditionelle banker. Men algoritmen resulterede uventet i en diskriminerende udlånspraksis, der systematisk favoriserede ansøgere fra bestemte indkomstgrupper og racemæssige baggrunde.

Virksomhedens baggrund

Den fiktive FinTech-virksomhed, FinTech Innovations, blev etableret for at disrupte den traditionelle banksektor ved at tilbyde et problemfrit, fuldt online lånegodkendessystem. Systemet brugte en proprietær algoritme til at vurdere kreditrisiko baseret på forskellige faktorer, såsom indkomst, beskæftigelseshistorik, geografisk placering og uddannelsesmæssig baggrund. FinTech Innovations' formål var at komme ind på markedssegmenter, der blev ignoreret af almindelige banker, især yngre personer, minoritetsgrupper og personer med utraditionelle indtægtskilder.

Identifikation af problemer

På trods af sine erklærede mål om inklusivitet favoriserede FinTech Innovations' kredit Scoringsalgoritme konsekvent ansøgere fra højere indkomstgrupper og specifikke racemæssige baggrunde. Det resulterede i uforholdsmæssigt høje afslagsprocenter for ansøgere med lavere indkomst og fra racemæssige minoriteter. Forskelle i lånegodkendelsesprocenter på tværs af demografiske grupper gav anledning til bekymring blandt kundelobbygrupper, hvilket førte til en ekstern revision, der bekræftede algoritmens diskriminerende resultater.

Analyse af grundlæggende årsager

Analysen afslørede flere grundlæggende årsager til bias. For det første afspejlede de historiske data, der blev brugt til at træne algoritmen, eksisterende samfundsmæssige forskelle og indlejrede systemiske bias i modellen. For eksempel brugte den geografisk placering som en funktion, der korrelerede stærkt med indkomst og race. Algoritmen var desuden stærkt afhængig af traditionelle beskæftigelsesdata og overså dem med ukonventionelle indkomster som freelancere eller gig-arbejdere. Endelig gjorde manglen på fairness-metrikker under modelvalideringen det muligt for disse bias at fortsætte uden at blive opdaget.

Konsekvensanalyse

FinTech Innovations' diskriminerende praksis havde betydelige sociale og økonomiske konsekvenser for de berørte grupper. Minoriteter og ansøgere med lav indkomst fik næsten dobbelt så mange afslag på lån som ansøgere med højere indkomst og en privilegeret baggrund. Denne udelukkelse resulterede i mistede muligheder for økonomisk vækst og yderligere forankring af eksisterende uligheder. Virksomheden stod over for skader på sit omdømme, lovgivningsmæssig kontrol og juridiske udfordringer.

Casestudie

Fiktiv case for studerende

Afbødningsstrategier

For at løse disse problemer implementerede FinTech Innovations flere korrigerende foranstaltninger:

- **Diversificering af data**

- De udvidede træningsdataene til at omfatte forskellige ansøgere fra forskellige indkomstniveauer, racer og beskæftigelsestyper.

- **Fairness-metrikker**

- De indarbejdede fairness-metrikker for at sikre, at algoritmens resultater var objektive og repræsentative.

- **Mål for gennemsigtighed**

- Virksomheden forpligtede sig til at dele mere information om, hvordan dens algoritme vurderede kreditværdighed.

- **Etiske retningslinjer**

- De nye retningslinjer krævede løbende overvågning, konsekvensanalyse og inddragelse af interessenter.

Konklusion og anbefalinger

Denne sag illustrerer de utilsigtede konsekvenser af algoritmisk beslutningstagning i FinTech. Den understreger vigtigheden af at bruge forskelligartede træningsdata, validere algoritmer med fairness-metrikker og opretholde gennemsigtighed for at mindske bias og diskrimination. Brancheomfattende etiske standarder og lovgivningsmæssige rammer er afgørende for at fremme fair og inkluderende FinTech-praksis.

TAK

Projektnummer | 2022-1-ES01-KA220-HED-000085257



Europa-Kommissionens støtte til produktionen af denne publikation udgør ikke en del af indholdet, som kun afspejler forfatterens synspunkter, og Kommissionen kan ikke holdes ansvarlig for nogen brug, der måtte blive gjort af oplysningerne heri.

