



Centrālā statistikas
pārvalde

Mākslīgā intelekta ietekme uz statistikas ražošanas procesiem

Praktiskais pielietojums

Ruāna Pavasare
Datu zinātnes un metožu atbalsta daļa

Mākslīgais intelekts (AI) statistikas ražošanā ļauj analizēt lielus datu apjomus, automatizēt klasifikāciju un prognozēšanu, kā arī uzlabot datu kvalitāti un procesu efektivitāti

Pielietotās AI apakšnozares statistikas ražošanā un procesos

Mašīnmācīšanās (ML)

Dziļā mašīnmācīšanās
(DL)

Lielo valodu modeļi (LLM)

Mašīnmācīšanās (ML)

modeļi **mācās no datiem**, lai prognozētu vai klasificētu. Klasiska mākslīgā intelekta pieeja, kas jau tiek plaši pielietota statistikas ražošanā.



Pielietojuma veidi

Imputācija

- Trūkstošo datu aizpildīšana
- Piemērs: ienākumi, izglītības līmenis, nozare

Klasterizācija

- grupēšana pēc līdzībām
- Piemērs: iedzīvotāju grupēšana pēc demogrāfiskām pazīmēm, teritoriju sadalījums

Prognozēšana

- rādītāju prognozēšana
- Piemērs: iedzīvotāju skaita prognozēšana

Klasifikācija

- Datu piešķiršana kategorijām
- Piemērs: uzņēmuma NACE koda noteikšana

Datu kvalitāte (anomāliju atklāšana)

- Netipisku vērtību identificēšana
- Piemērs: “izlēcēju” noteikšana

Saimniecisko darbību statistiskā klasifikācija (NACE)

NACE 2 → NACE 2.1 pāreja

Konteksts

Jauna klasifikācijas redakcija, jo parādās jaunas nozares un aktivitātes.

Uzņēmumi jāpārklasificē no NACE 2 uz NACE 2.1.

Izaicinājumi

Daļa kodu pāriet **1:1**.

Daļa kodu pāriet **1:n**

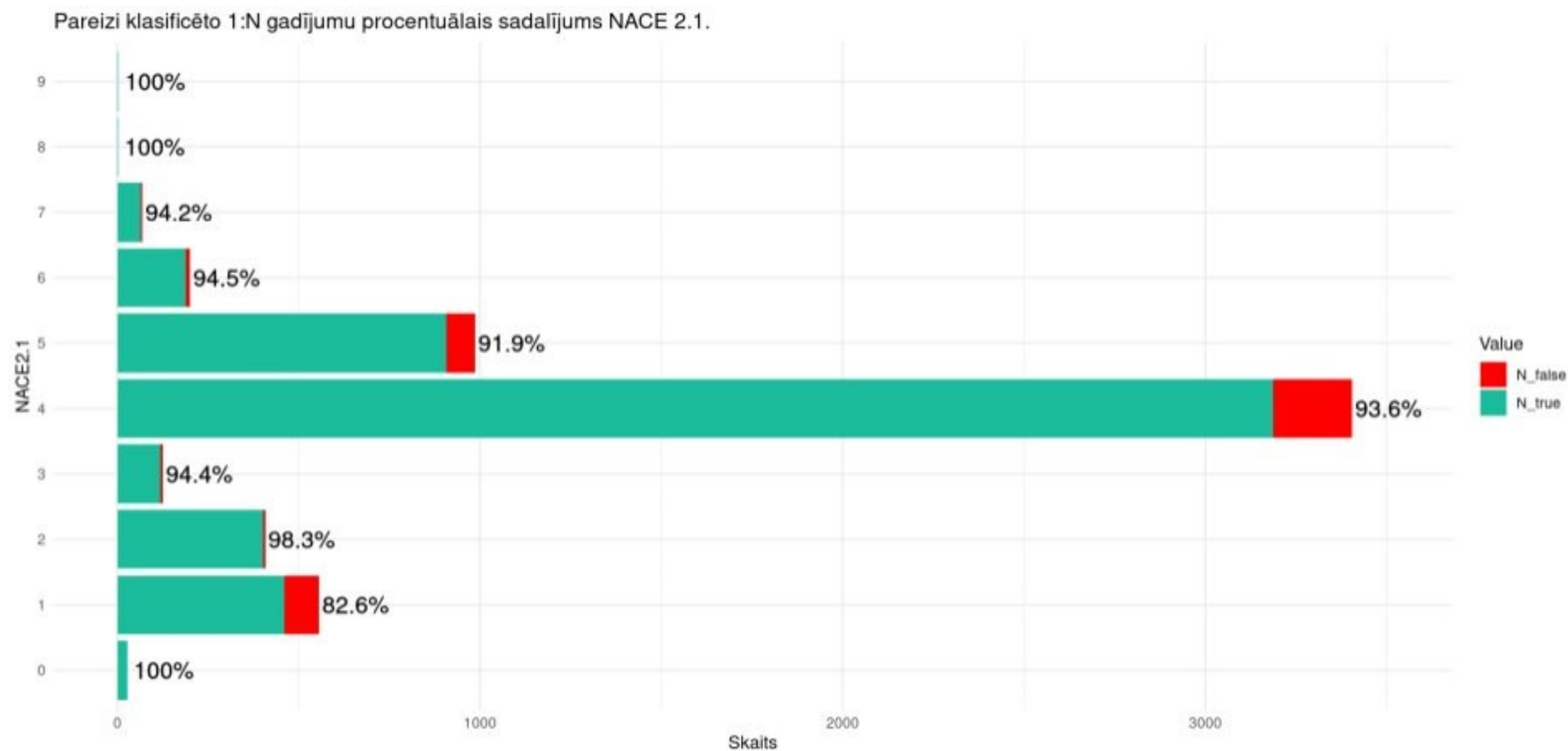
Pieejamā informācija: uzņēmuma saražotās preces + NACE kods pēc vecās redakcijas.

Risinājums

Pielietota **ML pieeja** (klasifikācijas modelis).

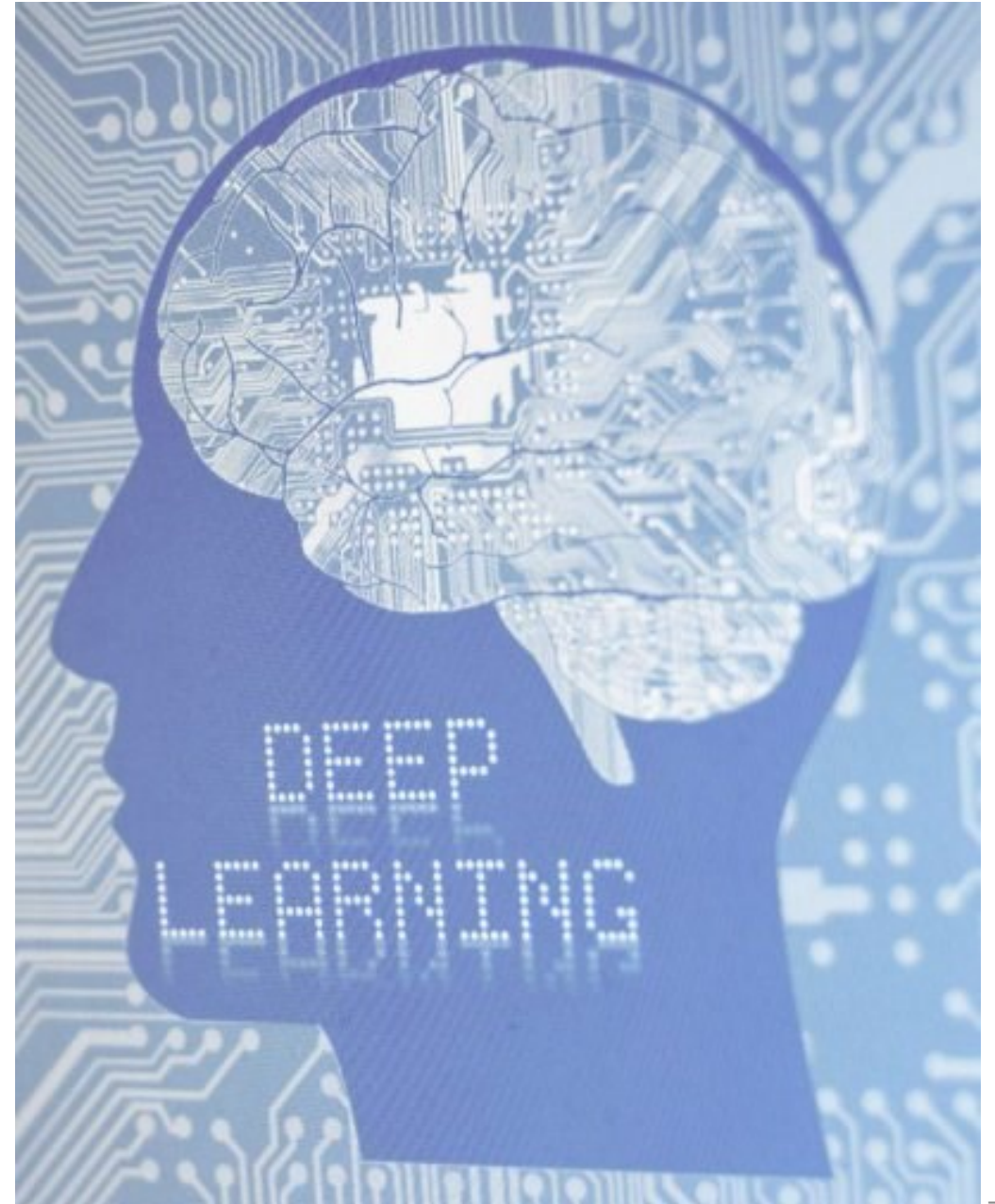
Tika nodrošināta uzņēmumu pāreja uz jauno klasifikāciju.

NACE pārejas rezultāti



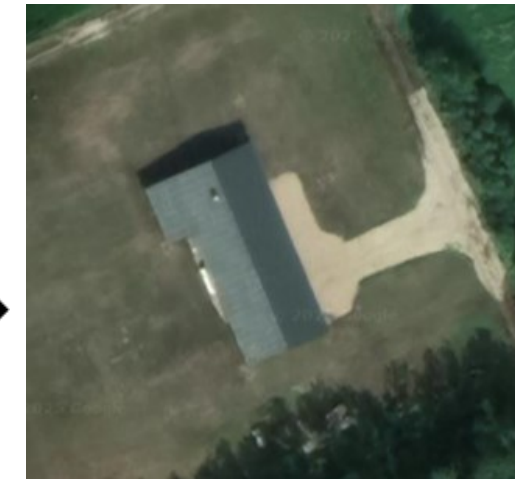
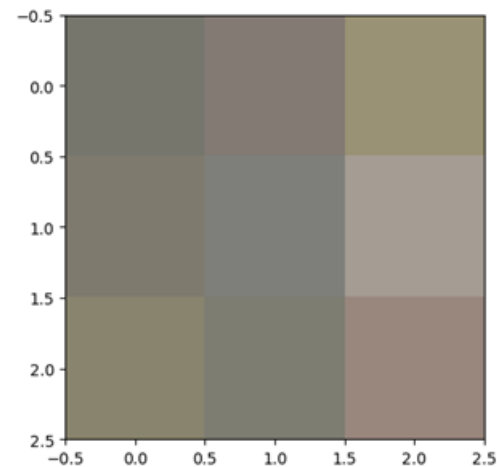
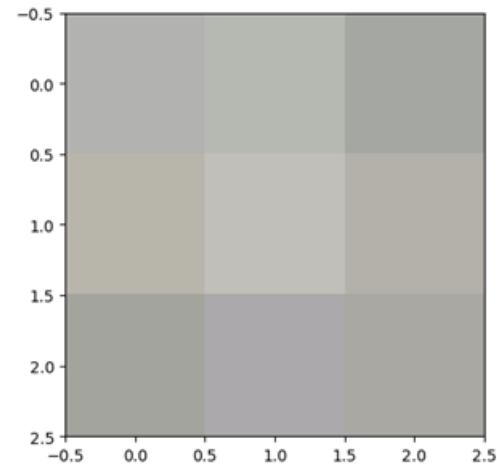
Dziļā mašīnmācīšanās

- izmanto dziļos neironu tīklus, kas sastāv no daudziem slāņiem.
- lieliem un sarežģītiem datiem
 - Attēli
 - Satelītdati



Satelīta dati

- 2026.gada sākumā Granta projekts par satelītdatu izmantošanu, lai noteiktu būvniecības uzsākšanu
- Metožu un datu pirms izpēte
 - Sentinel (1,2)

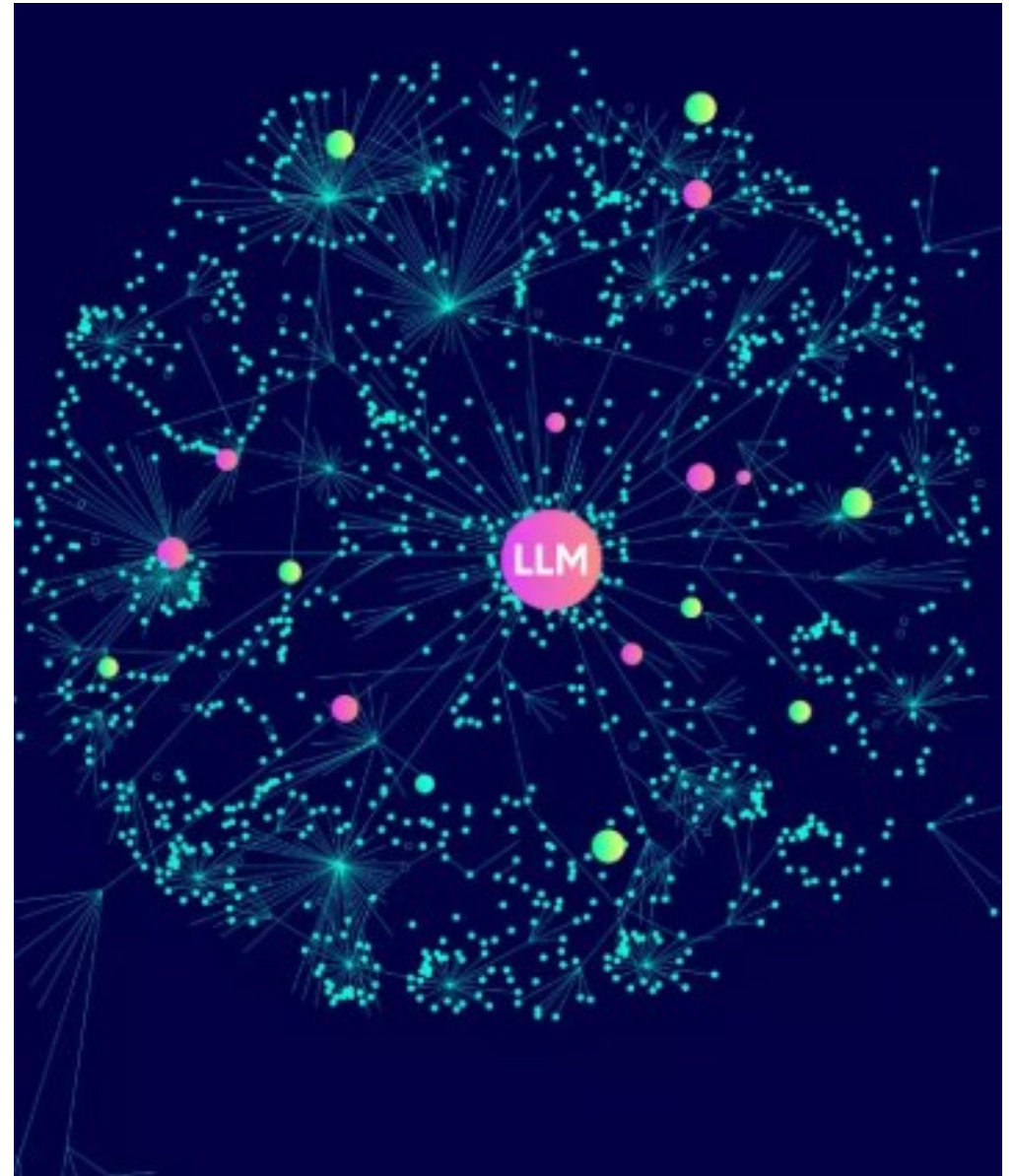


Klasifikācijas
modelis

Lielo valodu modeļi (LLM)

specializēti **darbam ar valodu**.

- teksta klasifikācija, semantisko tekstu meklēšana, informācijas apkopošana
- Veikala preču datu klasifikācija, vakanču datu apstrāde - informācijas izgūšana un profesiju klasifikācija



Veikala preču datu klasifikācija

Klasificēšana uz individuālā patēriņa veidu klasifikāciju (COICOP)

- Ieejošie dati
 - sarežģīti, liels datu apjoms

Augu pulv. kafijas baltin. COFFEETA 200g

DayUp Blue 120g

S. dz. DON SIMON DISFR. trop. augļu 50%1,5L D

Sald. RŪJIENAS piparm. šokolāde piedevu 65g

Šprotes Jūra tomāti mērc. stik. bur. 250g/168g

produkts	veikala plaukts	coicop
Augu pulv. kafijas baltin. COFFEETA 200g	Sausais piena pulveris	Sausais piens
DayUp Blue 120g	Functional yogurt	Jogurts un līdzīgi produkti
S. dz. DON SIMON DISFR. trop. augļu 50%1,5L D	Multiaug.sula > 0,5l	Augļu un dārzeņu sulas
Sald. RŪJIENAS piparm. šokolāde piedevu 65g	Ice cream cones	Ledus, saldējums un sorbets
Šprotes Jūra tomāti mērc. stik. bur. 250g/168g	Konservētas šprotes	Citas pelaģiskās zivis – sagatavotas vai konservētas

LLM tiek galā ar brīviem, nepilnīgiem tekstiem
ML pieejai nepieciešami papildus resursi, lai iegūtu tīrus, kvalitatīvus datus
apmācību modelim

Neizslēdz papildus pārbaudes LLM rezultātu validēšana ar LLM

Produkts	Coicop 1.rezultāts	validācija	Ticamības līmenis	Coicop 2.rezultāts
FRENCH FRIES KG	Sviestmaizes, picas, sāļie pīrāgi, gaļas vai zivju pīrāgi, saldēti un nesaldēti	French fries are not sandwiches,pizzas,savory pies,or meat or fish pies.	0.3	Sagriezti kartupeļi, saldēti
Sautēta brieža gaļa 250g	Konservēta gaļa	Stewed venison is a canned meat product,fitting the category.	0.9	Konservēta gaļa

Vakanču datu apstrāde

- Dati tiek izskrāpēti no vakanču portāliem
- Teksta nolasīšana no sludinājumiem, kas publicēti kā attēli
 - OCR (Optical Character Recognition)

```
{
  "id": "pardosanas-komandas-vaditaju-agentu-koordinatoru/68c8167983da4",
  "link": "https://www.visidarbi.lv/darba-sludinajums/pardosanas-komandas-vaditaju-agentu-koordinatoru/68c8167983da4",
  "time": "Pilna slodze",
  "added": "pirms 8 stundām",
  "title": "Pārdošanas komandas vadītāju, aģentu koordinatoru.",
  "salary": "1400 - 2100 (bruto)",
  "company": "Ažiņa komercfirma Markets SIA",
  "duedate": "Terminš 29.09",
  "location": "Cēsu novads",
  "description": "Pienākumi: vadīt un motivēt aģentu komandu; jaunu klientu piesaiste, ilgtermiņa sadarbības uzturēšana; pieredze pārdošanā; nodrošināt pārdošanas plānu izpildi; pārdošanas rezultātu analīze, atskaišu sagatavošana; darbs ar lielajiem klientiem. No Jums sagaidām: B kategorijas autovadītāja apliecība; latviešu valodas zināšanas; vēlamas angļu valodas zināšanas sarunvalodas līmenī; labas iemaņas darbā ar MS Office datorprogrammām, e-pasts, u.c.; labas komunikācijas un organizatoriskās prasmes, orientēts uz rezultātu; prasme plānot un organizēt darbu. Mēs piedāvājam: interesantu un dinamisku darbu stabilā uzņēmumā; sociālās garantijas; veselības apdrošināšanu; darba automašīna. atalgojumu no 1400.00-2100.00 EUR bruto (fiksētā daļa + bonusi par sasniegtiem rezultātiem). Sūtiet savu pieteikumu CV uz e-pastu: [email protected] līdz 2025.gada 28.septembrim. Sazināsimies ar kandidātiem, kas tiks aicināti uz interviju klātienē."
}
```

Pieprasam sev
nepieciešamo struktūru,
lai iegūtu strukturētus
un apstrādājamus datus

```
"adrese": {
  "type": "object",
  "properties": {
    "valsts": {
      "type": "string",
      "description": "Valsts kurā ir paredzama darba vieta.",
      "instrukcijas": "Jāizgūst faktiskās darba vietas valsts."
    },
    "pilseta_pagasts_novads": {
      "type": "string",
      "description": "Pilsēta, pagasts, novads kurā ir paredzama darba vieta.",
      "instrukcijas": "No sludinājuma teksta jāizgūst pilsētu, pagastu, novadu. Piemēri: \"Ķīšupe divi 151, Ķīšupe 2, Saulkrastu pag.\"."
    },
    "iela": {
      "type": "string",
      "description": "Darba vietas iela",
      "instrukcijas": "No sludinājuma teksta jāizgūst iela kurā ir darba vieta. Piemēri: Zemītes iela 6A, Ķīšupe divi 151."
    }
  }
},
"nozare": {
  "type": "object",
  "properties": {
    "nace_kods": {
      "type": "string",
      "description": "Nozares kods atbilstoši NACE klasifikācijai",
      "instrukcijas": "Izgūsti iespējamo nozares raksturojumu no sludinājumā."
    },
    "nozares_nosaukums": {
      "type": "string",
      "description": "Informācija par nozares nosaukumu atbilstoši NACE 2.1",
      "instrukcijas": "Noskaidro nozares nosaukumu atbilstoši NACE 2.1 klasifikācijai"
    }
  }
},
"profesija": {
  "type": "object",
  "properties": {
    "kods": {
      "type": "string",
      "description": "Profesijas kods atbilstoši profesiju klasifikatoram",
      "instrukcijas": "Analizē sludinājumu saturu, veicamos darbus."
    }
  }
},
```

Tālāka datu apstrāde

Profesiju klasifikācija

- Izvilktais nestrukturizētais profesijas nosaukums klasificējams uz ISCO standarta kodiem.

LLM pielietojums kā klasifikācijas uzdevums

Programmētājs – back-end izstrādātājs
head of payments
Konsultants darījumu konsultāciju nodaļā
guidance, navigation, and controls engineer
vecākais finanšu sistēmu analītiķis (dynamics 365fo)
Tehnikas operators
vecākais java programmētājs
tāmētājs (elektrotīkli un vājstrāvas tīkli)
QA Automation Engineer (Ruby)
Personāla procesu koordinators
Komandas vadītājs
uzņēmuma vadītājs
export sales director (german speaking)
Radošo projektu vadītājs
Operatīvā vadības centra darbinieks
user support & access management specialist
hotel guest experience host/hostess

Nākotne statistikā ar mākslīgo intelektu

- Mākslīgā intelekta loma statistikas ražošanā noteikti tikai pieaugs
- Pielietojamība visiem datiem - izlašu apsekojumi, administratīvie un alternatīvie datu avoti (teksti, attēli, satelītdati....)
- datu vākšana un apstrāde, analīze, datu bagātināšana, kvalitātes pārbaude un validācija
 - Nemainīgi - nepieciešama neatkarīga, precīzi veidota validācijas kopa rezultātu izvērtēšanai
- Rezultātu izmantošana un analīze

Statistikas portāls – kā sarunu biedrs

Dabiskā valoda: lietotājs var uzdot jautājumus tā, kā sarunātos ar cilvēku.

Personalizēta pieredze: atbildes pielāgojas lietotāja vajadzībām un kontekstam.

Ātra piekļuve: nav jāmeklē datu tabulās – LLM sameklē un pasniedz informāciju.

Interaktivitāte: iespēja precizēt jautājumus un veidot dialogu ar portālu, vizualizēt.

Vai tas ir iespējams?



Centrālā statistikas
pārvalde

 @CSP_Latvija

 csplatvija

www.csp.gov.lv

