



**LATVIJAS
UNIVERSITĀTE**

Statistiskā mašīnmācīšanās un mākslīgais intelekts

Rīga, 2022

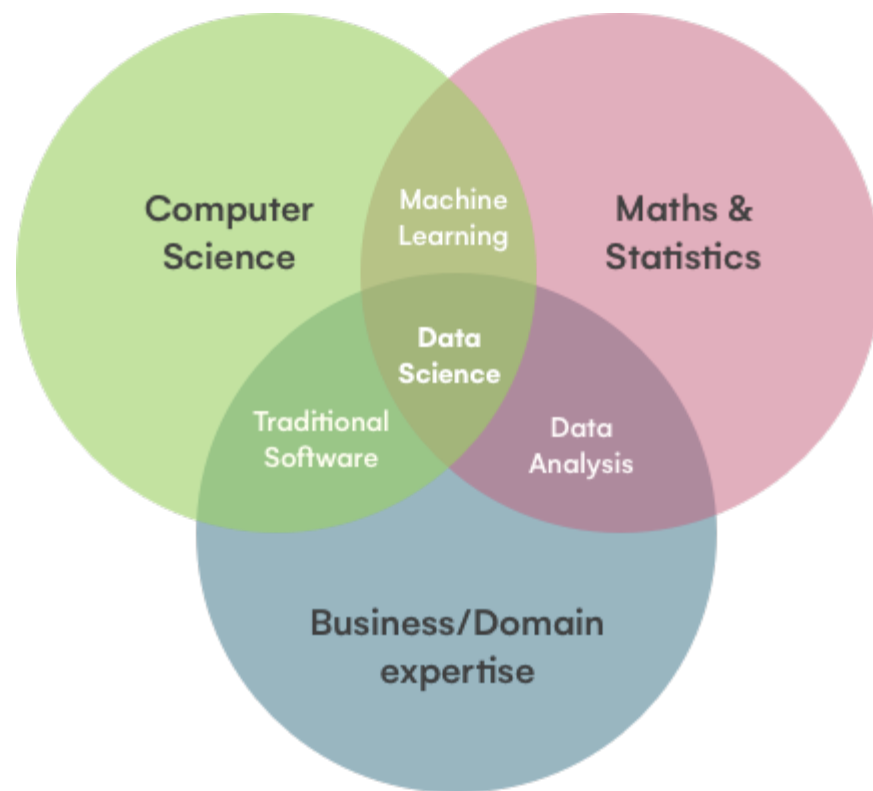
ARTIS ALKSNIS

Prezentācijas saturs

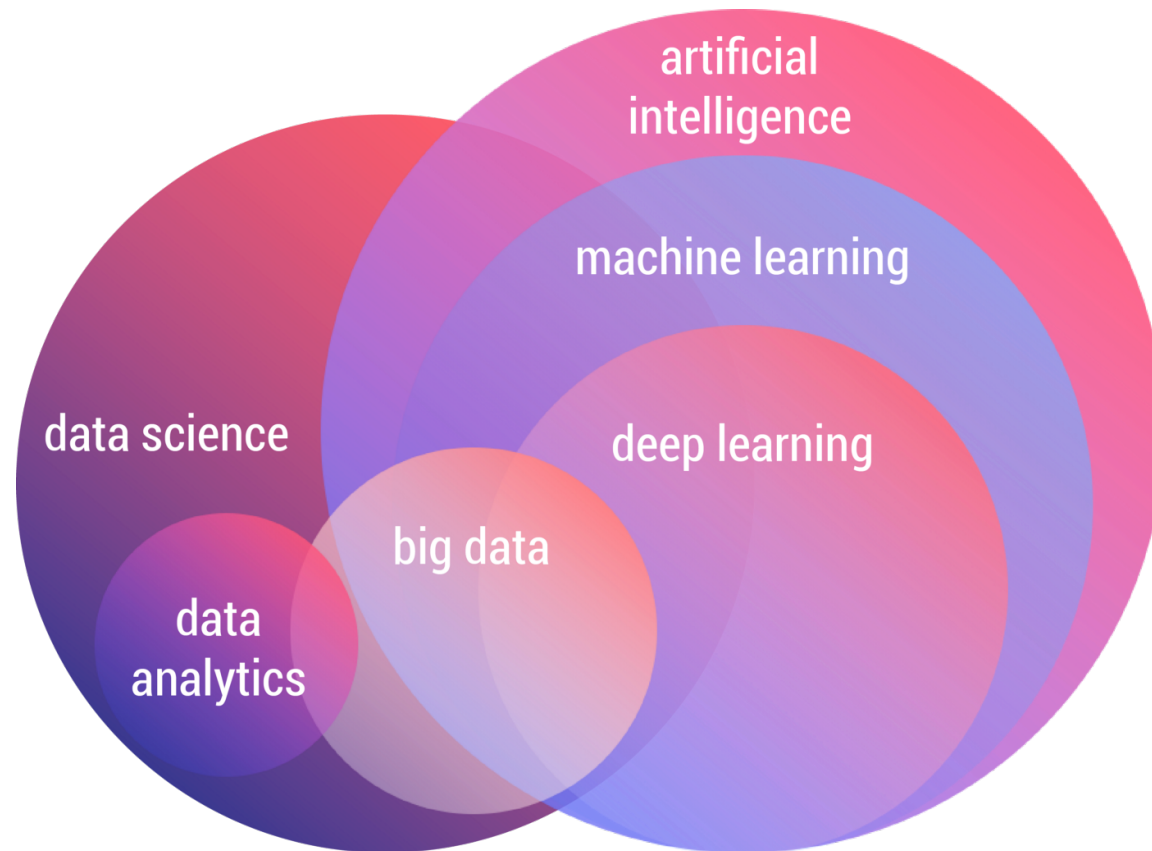
1. Kas ir datu zinātne, mašīnmācīšanās un mākslīgais intelekts?
2. Pamatprincipi
3. Interaktīvi piemēri
4. Dati un risināmās problēmas raksturošana
5. Datu vispārēja analīze
6. Datu iegūšana un sagatavošana
7. Mašīnmācīšanās un mākslīgā intelekta metodes un to izvēle
8. Modeļa trenēšana
9. Modeļu izmantošana un uzturēšana
10. Praktisks piemērs

1. Kas ir datu zinātne, mašīnmācīšanās
un mākslīgais intelekts?

Datu zinātne



Statistiskā mašīnmācīšanās un mākslīgais intelekts

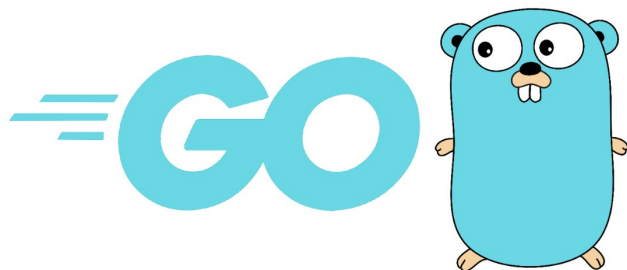


2. Pamatprincipi

Populārākās programmēšanas valodas



pythonTM



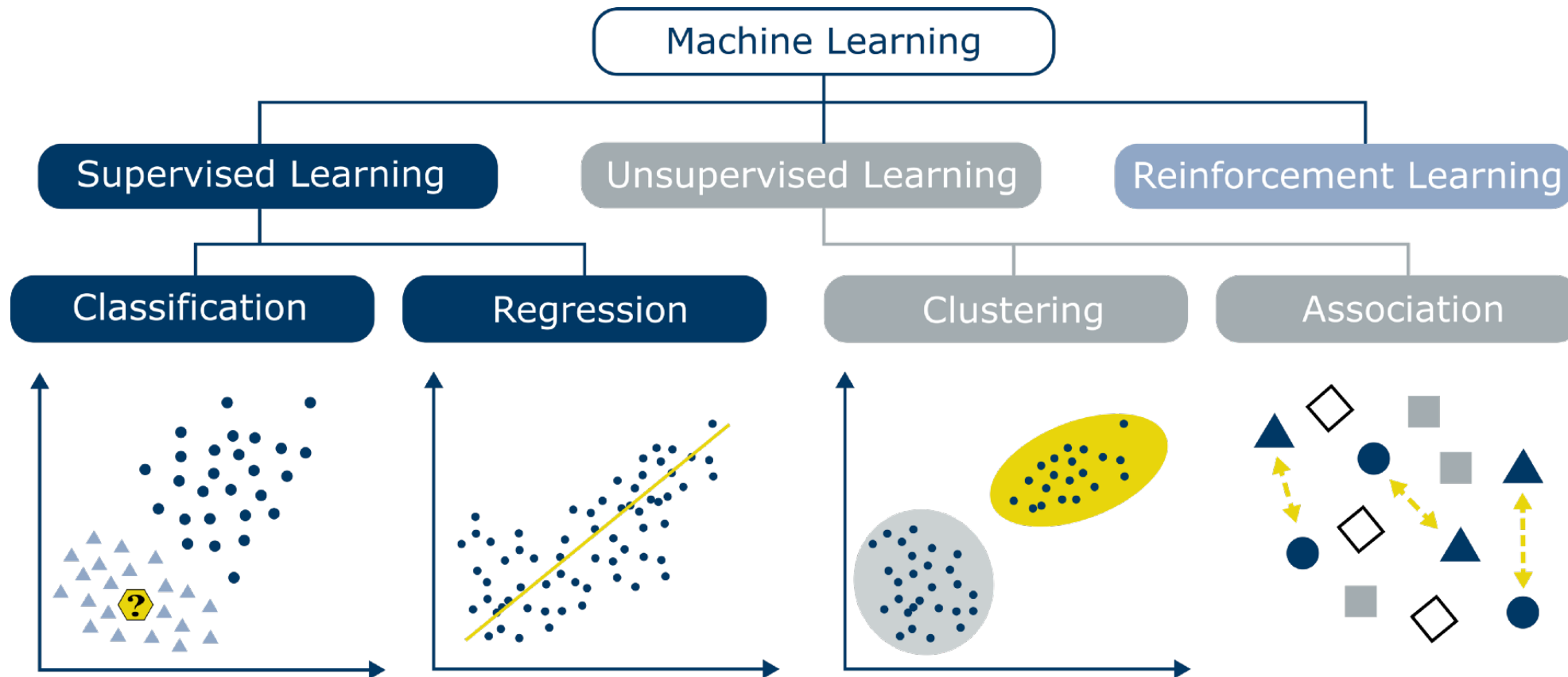
Datu zinātnes un mākslīgā intelekta pamatmērķi

Datu izpēte un analīze

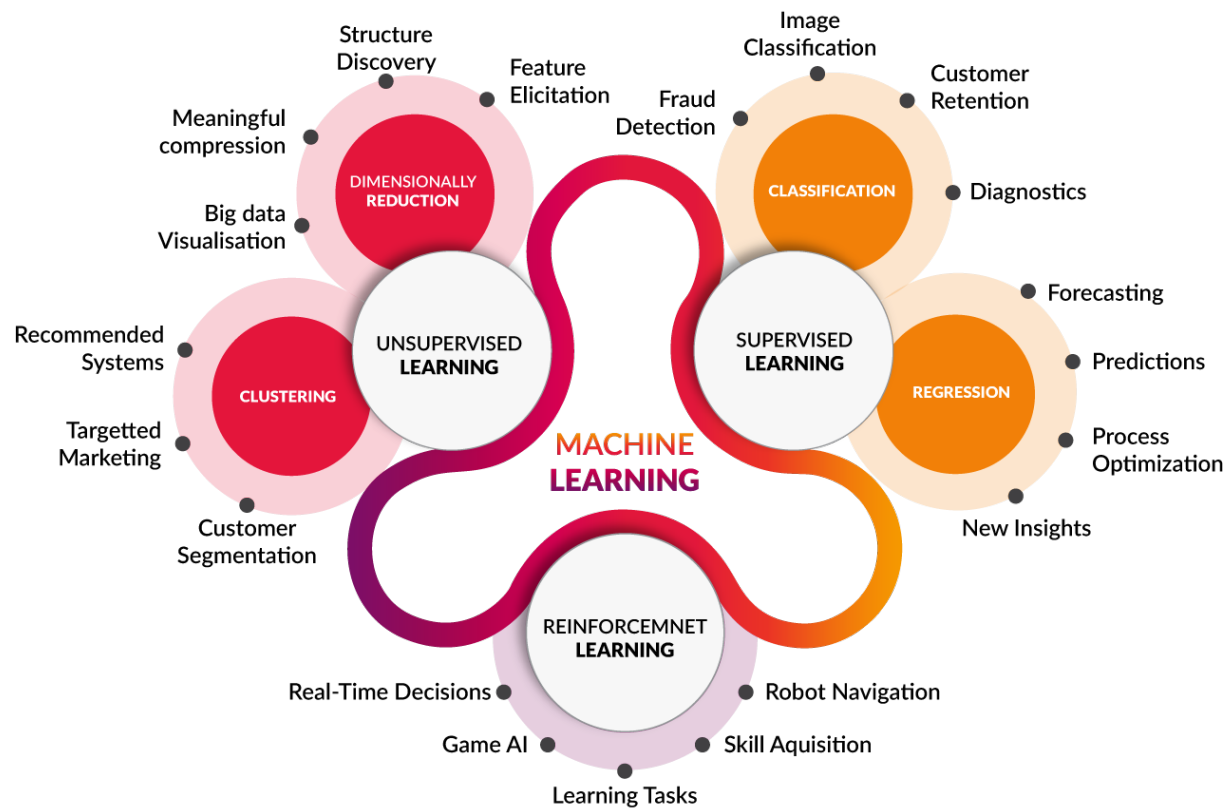
Datu vizualizācija

Modeļu veidošana

Mašīnmācīšanās un mākslīgā intelekta modeļu tipi



Mašīnmācīšanās un mākslīgā intelekta modeļu tipi



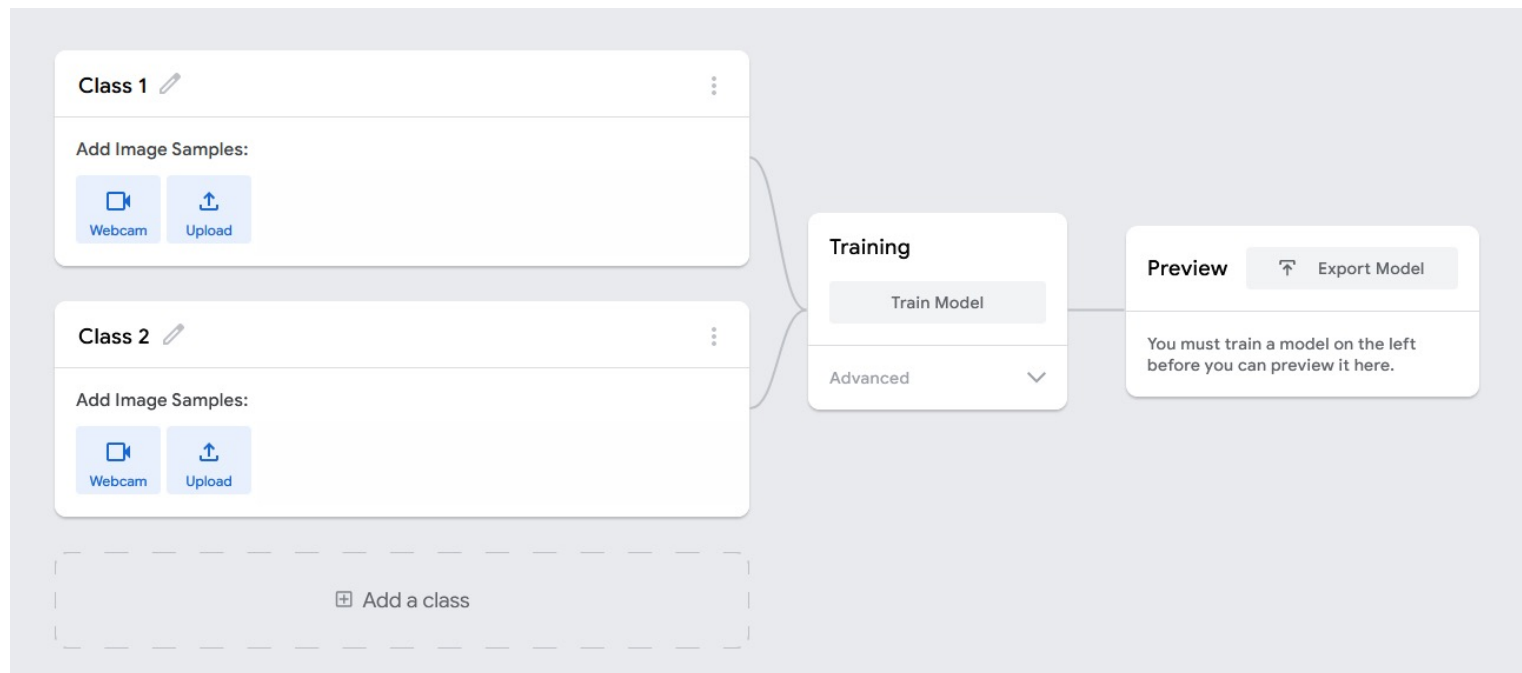
3. Interaktīvi piemēri

“Šī persona neeksitē”



<https://this-person-does-not-exist.com/en>

Attēla atpazīšanas projekts



<https://teachablemachine.withgoogle.com/train>

4. Dati un risināmās problēmas raksturošana

Dati un risināmās problēmas raksturošana

Izpētīt pieejamos datus un definēt risināmo problēmu.

Piemēri

- ❑ Reklāmas (Facebook)
- ❑ Google meklēšana algoritms
- ❑ Rekomendācijas (TikTok, YouTube, Netflix utt.)
- ❑ Tulkošana (google translate, deepl utt.)
- ❑ Automātiskā teksta pabeigšana

5. Datu vispārējā analīze

Datu vispārējā analīze

- Datu pieejamība (vai visi dati ir pieejami, par kādu period ir pieejami dati)
- Datu kvalitātes novērtējums (vai trūkst vērtības, vai trūkst ieraksti utt.)
- Datu vizualizācija (veikt piemērotu datu vizualizāciju)
- Datu struktūra (bilde, tabulas utt.)
- Datu tipi
 - Skaitliskas vērtības
 - Kategoriālas vērtības
 - Laikrindu dati
 - Teksts

6. Datu iegūšana un sagatavošana

Datu iegūšana

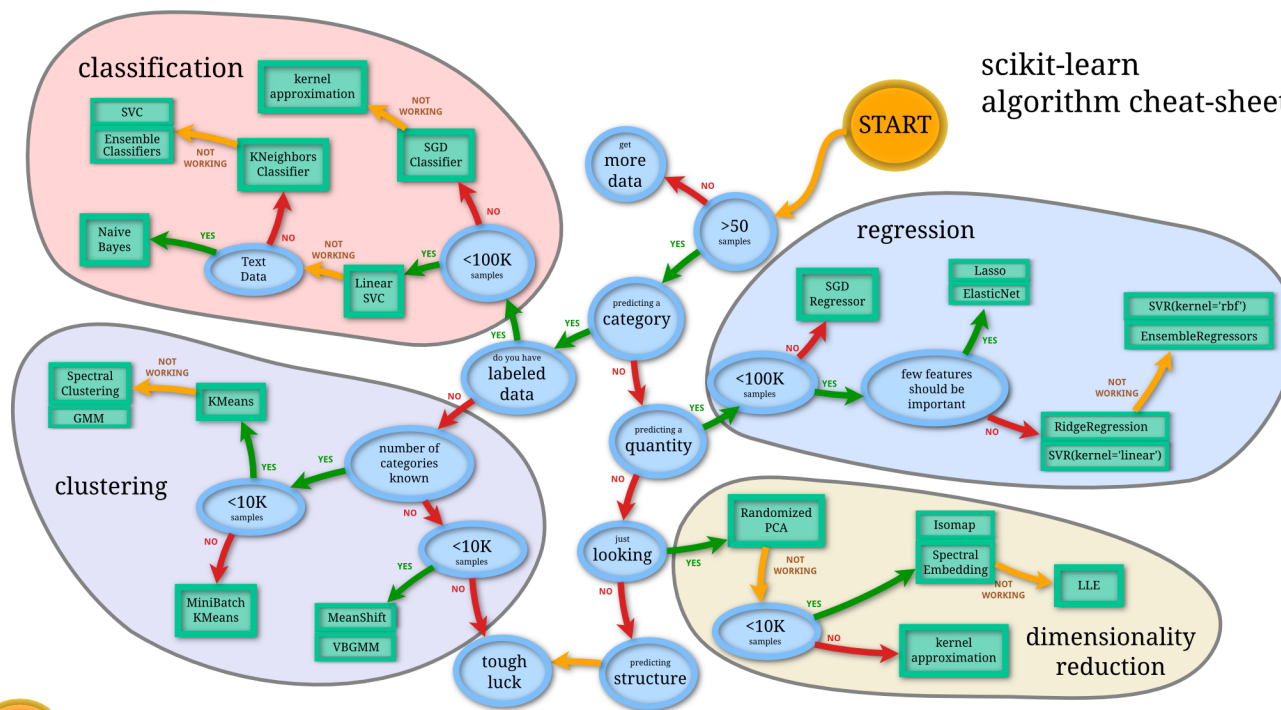
- Dažāda veida datubāzes
- Tīmekļa vietnes
- Faili (xlsx, csv, json utt.)
- un citi iespējamie datu avoti.

Datu sagatavošana

- Datu formatēšana
- Datu tīrīšana
- Trūkstošo vērtību / ierakstu aizvietošana vai dzēšana
- Datu konsolidācija / apvienošana
- Jaunu mainīgo veidošana (pēc nepieciešamības)
- Cita veida manipulācijas un transformācijas ar datu tipu vai struktūru

7. Mašīnmācīšanās un mākslīgā intelekta metodes un to izvēle

Mašīnmācīšanās metožu karte



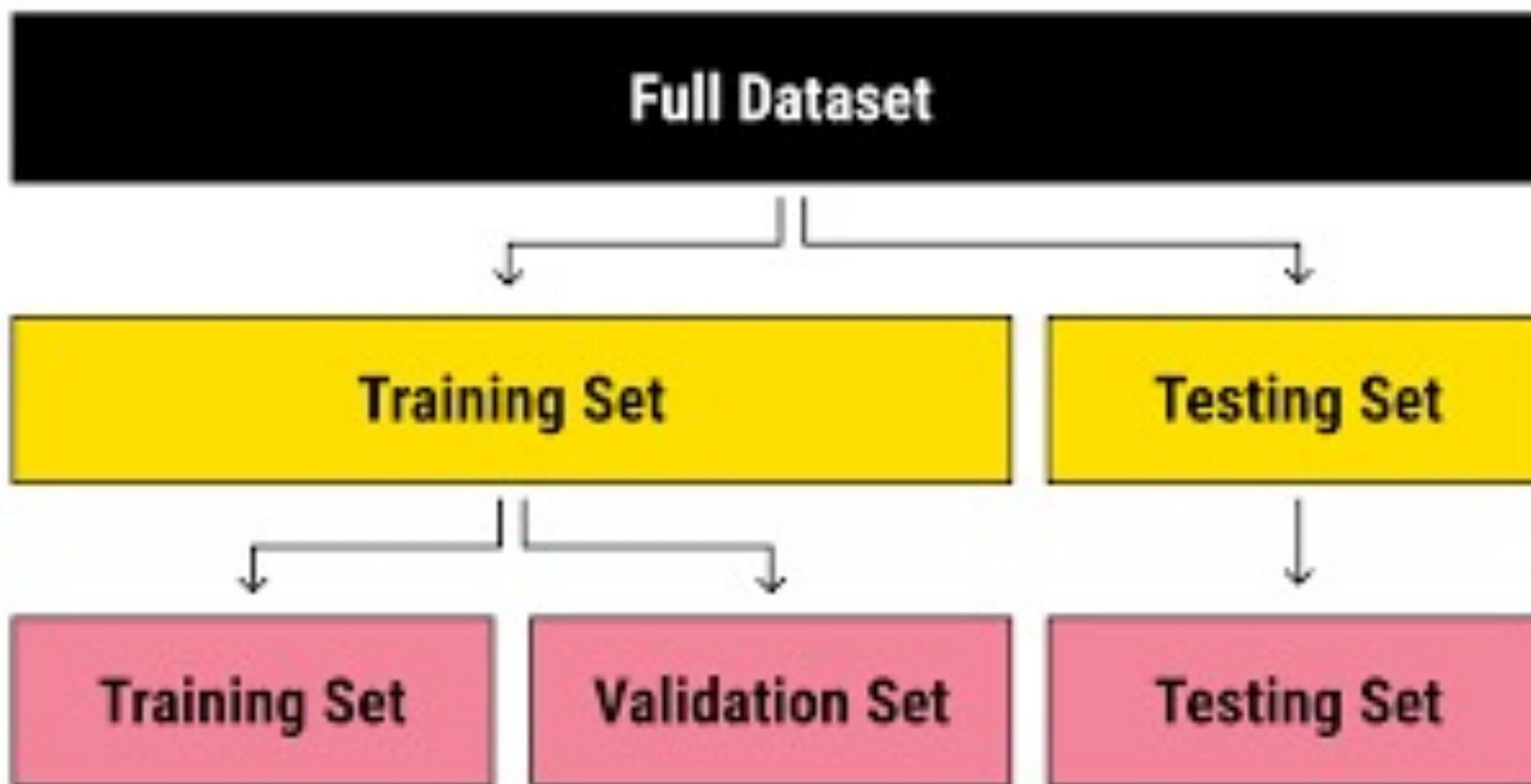
Kā izvēlēties “pareizo” metodi?

Atkarīgs no:

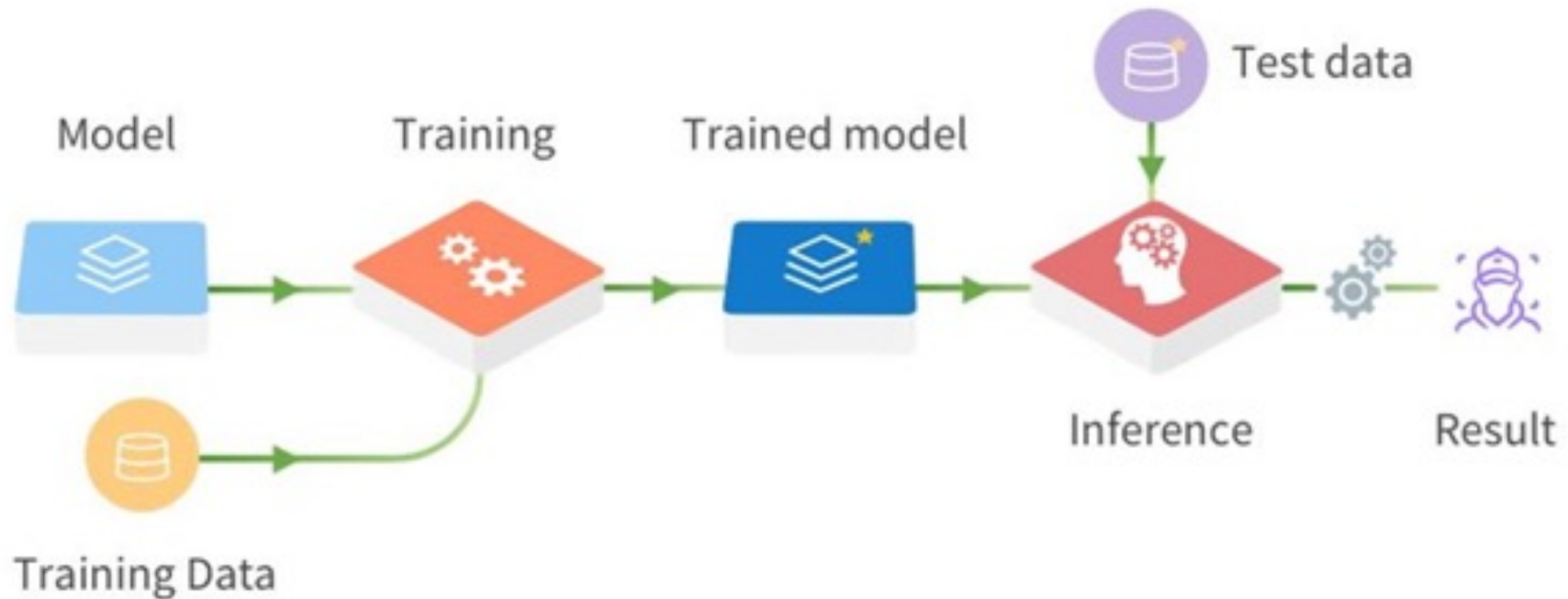
- Risināmās problēmas veida
- Datu apjoma
- Citas attiecīgās problēmas specifikas (piemēram, izpildes laika vai modeļa trenēšanas/pārtrenēšanas laika)

8. Modeļu trenēšana

Treniņa, testa un validācijas datu kopu izdale

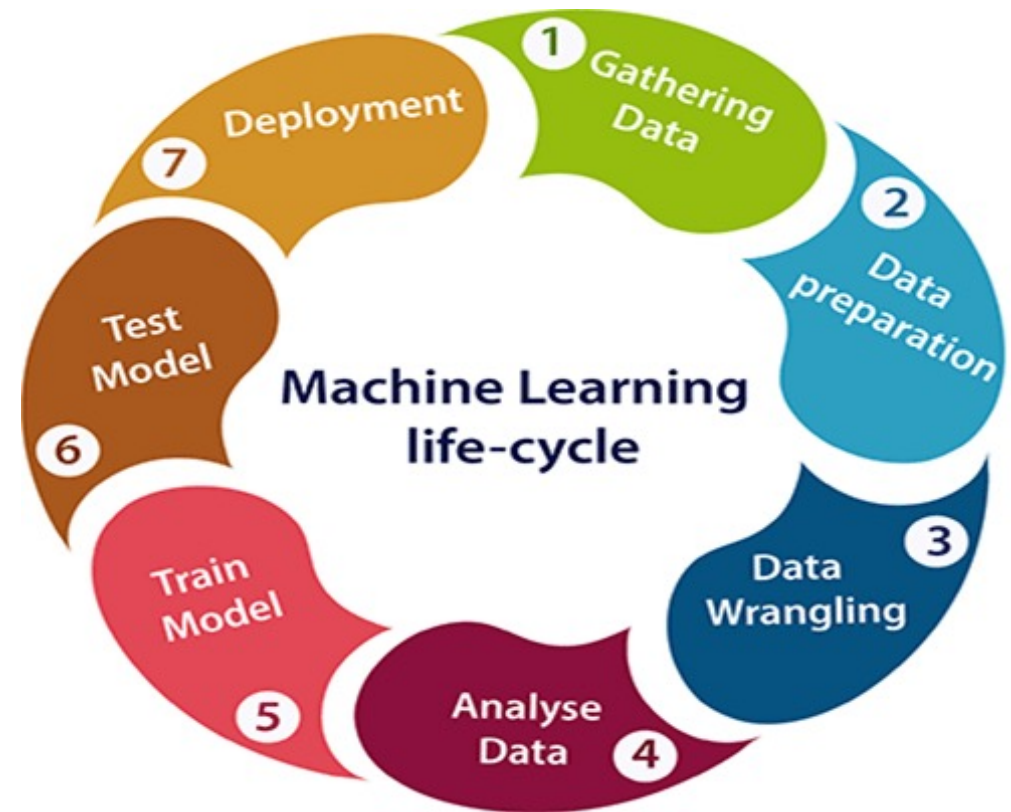
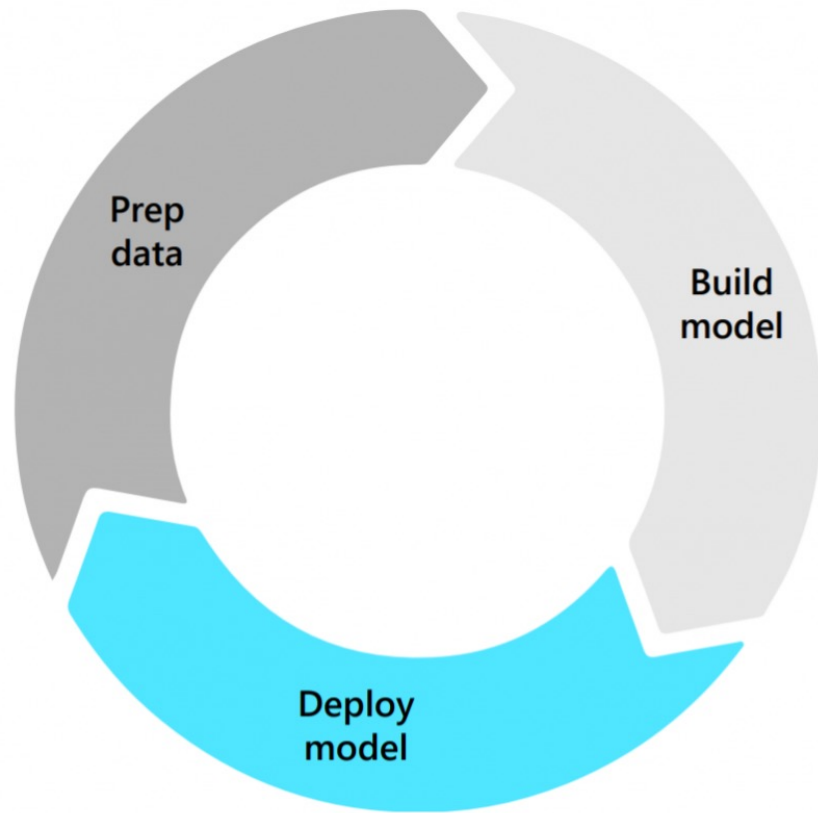


Modeļu trenēšana process



9. Modeļu izmantošana un uzturēšana

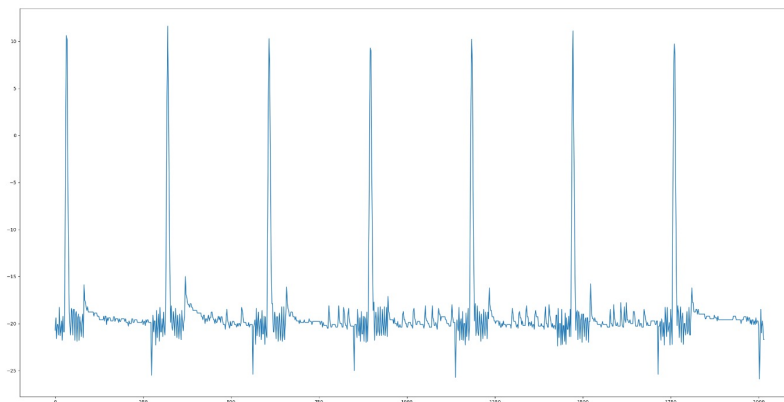
Modeļa dzīves cikls



10. Praktisks piemērs

Par pētījumu!

- ❑ Izpētes ietvaros tika aplūkotas **temperatūras mērījumu** sensoru laikrindas.
- ❑ Izpētes mērķis – izstrādāt reāllaika **anomālijas noteikšanas** modeli, kas spētu identificēt potenciālās problēmas ar sensoru vai iekārtu.
- ❑ Izpētes uzdevums – izpētīt un izveidot piemērotāko dziļās apmācības (**LSTM**) modeli.
- ❑ Izstrādāt modeļa anomāliju noteikšanas modeļa **prototipu**.



Laikrindu vispārēja analīze

- Vizuāla analīze
- Stacionaritātes pārbaude
- Autokorelācijas funkcijas attēlojums
- Parciālās autokorelācijas funkcijas attēlojums
- Laikrindu dekompozīcija
- Laikrindas struktūras sarežģītības novērtējums (*Sample entropy*)

Datu sagatavošana un priekšapstrāde

- ❑ Datu trūkstošo vērtību aizpildīšanas metodes
- ❑ Laikrindu transformācijas
- ❑ Gludināšanas metodes
- ❑ Mērogošanas metodes
- ❑ Anomāliju likvidēšana

Prognozēšana!

Tika apskatīti un salīdzināti vairāki laicrindu prognozējošie algoritmi, ieskaitot sekojošos:

- Holt-Winters (HW) - trīskāršās eksponenciālās gludināšanas modelis
- LSTM – dziļo neironu tīklu modelis
- LightGBM - gradientu pastiprināšanas algoritms
- SARIMA – sezonālā ARIMA
- SVM – atbalsta vektoru mašīnas
- Citi gradient pastiprināšanas algoritmi (CatBoost & XGBoost)
- FBProphet – Facebook prognozējošais algoritms

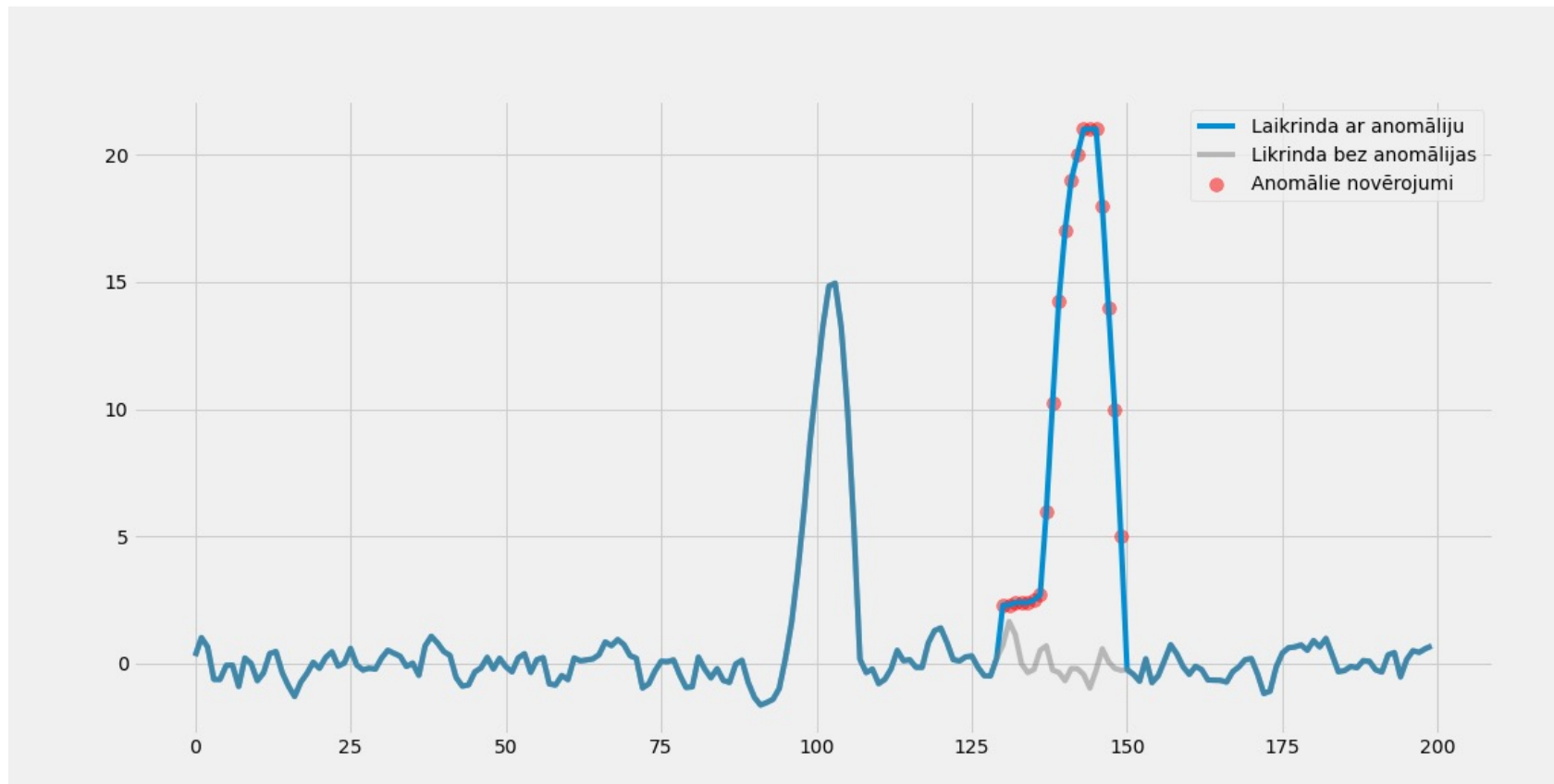
Trīskāršās eksponenciālās gludināšanas (*Holt-Winters*) modelis

Trīskāršās eksponenciālās gludināšanas (HW) aditīvais modelis sastāv no sekojošām komponentēm:

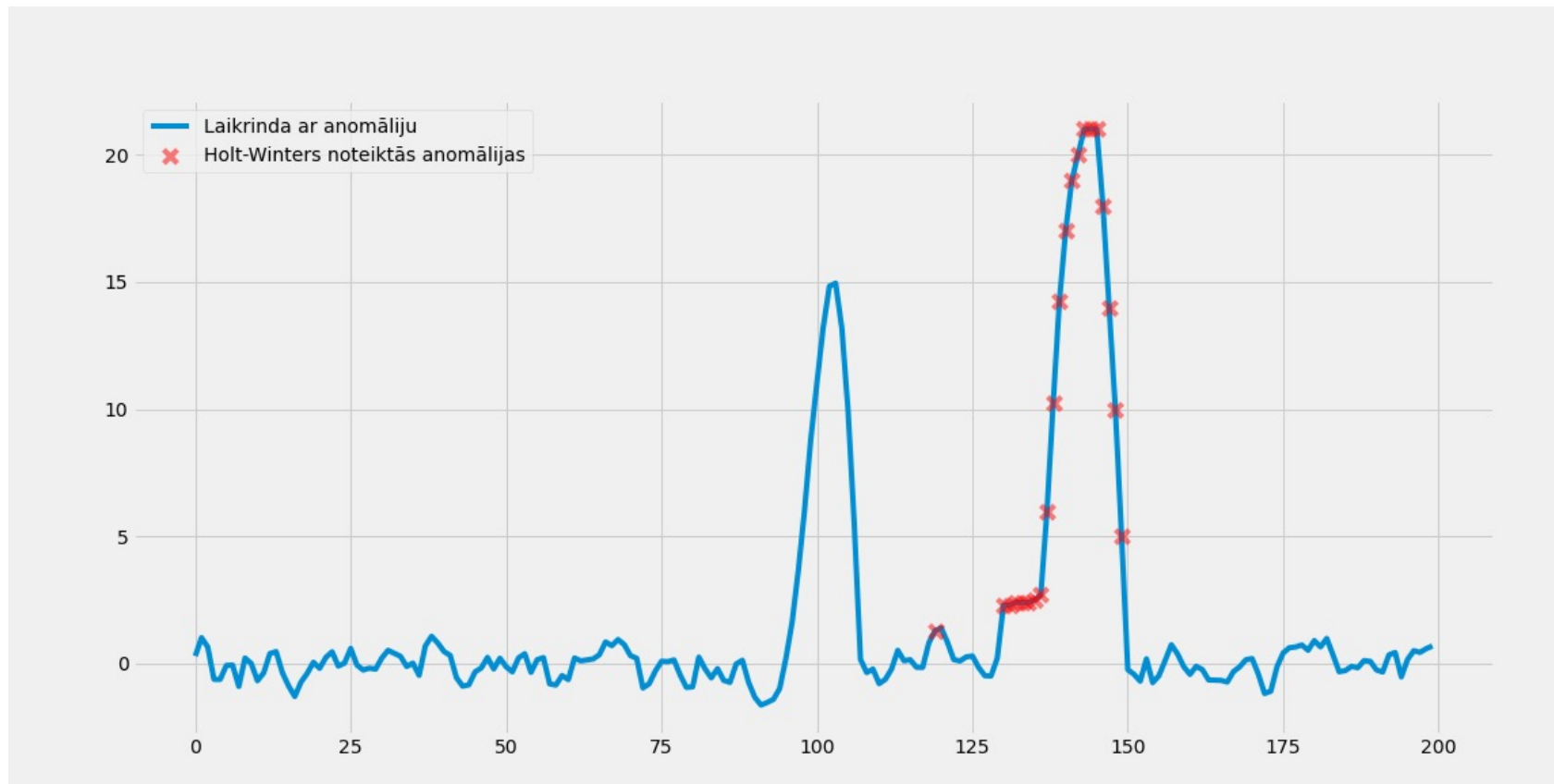
Modeļa vienādojums	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$
Līmeņa komponente	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
Trenda komponente	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
Sezonālā komponente	$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$

kur h ir prognozējamās vērtības laika solis, m ir sezonas garums un k ir veselā daļa no $(h-1)/m$. ℓ_t ir vidējais svērtais novērtējums starp sezonāli pielāgoto novērojumu ($y_t - s_{t-m}$) un nesezonālo ($\ell_{t-1} + b_{t-1}$) prognozi laika momentā t un $0 \leq \alpha \leq 1$. b_t ir līmeņa starpības - $\ell_t - \ell_{t-1}$ un pagājušā laika momenta trenda - b_{t-1} vidējā svērtā vērtība, kur $0 \leq \beta^* \leq 1$. s_t ir pašreizējās sezonālas indeksa - ($y_t - \ell_{t-1} - b_{t-1}$) un pagājušās sezonas (m laika vienības atpakaļ) indeksa vidējais svērtais novērtējums, kur $0 \leq \gamma \leq 1 - \alpha$.

Anomāliju noteikšana!



Anomāliju noteikšana – HW modelis



Modeļu salīdzinājums

Modelis	Modeļa tips	Izpildes laiks, sekundes	R ²	MAE	MedAE	MSE	MAPE
Holt-Winters	Stat	2.12	0.96	0.37	0.31	0.22	158.75
LightGBM	ML	3.13	0.95	0.45	0.36	0.32	113.75
SARIMA	Stat	33.76	0.95	0.44	0.38	0.31	169.50
Sezonālās vidējās vērtības prognoze	Naivais	0.01	0.95	0.44	0.36	0.31	195.67
ETS + ARIMA	Stat	1.25	0.95	0.45	0.34	0.33	213.73
LSTM	DL	77.65	0.95	0.46	0.40	0.31	580.15
SVM	ML	0.03	0.94	0.47	0.40	0.34	122.29
CatBoost	ML	12.82	0.94	0.47	0.43	0.34	155.01
XGBoost	ML	2.96	0.94	0.46	0.39	0.34	224.62
Theta	Stat	0.04	0.81	0.78	0.64	1.16	468.61
Vidējās vērtības prognoze	Naivais	0.00	-0.02	1.34	0.97	6.10	603.96
Pēdējās vērtības prognoze	Naivais	0.01	-0.23	1.24	0.72	7.32	493.00

Paldies par uzmanību!