

Qlik AutoML

Dirbtinis intelektas Qlik profesionalo kasdienybei

Milda Vitalytė Masaitė, Qlik architektė

m.vitalyte@theinfotrust.com

The Infotrust, UAB

Dirbtinis intelektas vs Mašininis mokymasis INFOTRUST

Dirbtinis intelektas
(Artificial intelligence – AI)

Technologijos kurios leidžia simuliuoti
žmogaus elgseną

Mašininis mokymasis
(Machine learning – ML)

Dirbtinio intelekto šaka, kuri leidžia
mašinai mokintis iš praeities duomenų ir
prognozuoti ateitį

ML modelio kūrimo būdai

ML modelio kūrimas viską programuojant

- 1 duomenų mokslininkas
- 5 dienos
- Kodo rašymas

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split

# Load dataset
df = pd.read_csv('training_data.csv')
train, test = train_test_split(df, test_size=0.2)

# X, y split
X_train = train.drop('target_column', inplace=False, axis=1)
y_train = train['target_column']
X_test = test.drop('target_column', inplace=False, axis=1)
y_test = test['target_column']

# Logistic Regression
lr = LogisticRegression()
lr.fit(X_train, y_train)

# Predict
y_pred = lr.predict(X_test)

# Performance metrics (Linear Regression)
print('R2: ' + str(metrics.r2_score(y_test, y_pred)))
print('Mean Squared Error: ' + str(metrics.mean_squared_error(y_test, y_pred)))
```

Python paketų skriptai

- 1 duomenų mokslininkas
- 1 diena
- Kodo rašymas

```
In [ ]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LogisticRegression

In [ ]: # Load dataset we need:
df = pd.read_csv('training_data.csv')
train, test = train_test_split(df, test_size=0.2)

In [ ]: # X, y split:
y = train['target_column']
X = train.drop('target_column', inplace=False, axis=1)
y_test = test['target_column']
X_test = test.drop('target_column', inplace=False, axis=1)

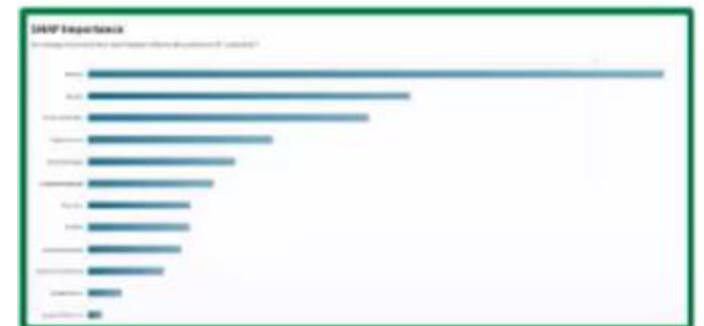
In [ ]: # Logistic Regression:
lr = LogisticRegression()
lr.fit(X, y)
y_pred = lr.predict(X_test)

In [ ]: # Performance metrics (Linear Regression):
print('R2: ' + str(metrics.r2_score(y_test, y_pred)))
print('Mean Squared Error: ' + str(metrics.mean_squared_error(y_test, y_pred)))
```

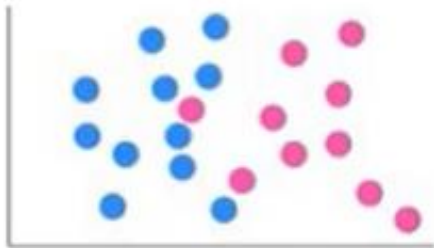
Qlik AutoML

- 1 analitikas
- 1 valanda
- Be kodo rašymo

Model	Accuracy	Precision	Recall	F1 Score	AUC
LogisticRegression	0.85	0.85	0.85	0.85	0.95
SupportVectorMachine	0.80	0.80	0.80	0.80	0.90
DecisionTreeClassifier	0.75	0.75	0.75	0.75	0.85
RandomForestClassifier	0.82	0.82	0.82	0.82	0.92
XGBoost	0.88	0.88	0.88	0.88	0.98
AdaBoost	0.78	0.78	0.78	0.78	0.88
MLP	0.83	0.83	0.83	0.83	0.93



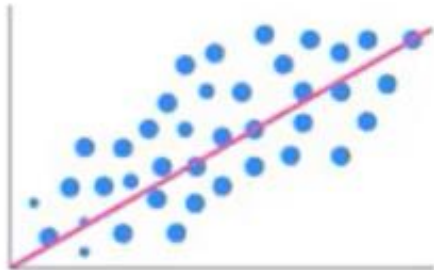
Qlik AutoML sprendžiamos problemos



Klasifikacija

Prognozės tipas: kategorinis

Pvz: Ar klientas sustabdys savo narystę (Taip/Ne)?



Regresija

Prognozės tipas: skaitinis

Pvz: Kokia žmogaus prognozuojama gyvenimo trukmė?



Laiko eilutės

Prognozės tipas: skaitinis + data

Pvz: kokie prognozuojami pardavimai ateinantį ketvirtį?

AutoML reikalavimai duomenims

Duomenų rinkinio sudedamosios dalys:



Įvykis



Prognozė



Atributai, požymiai



Laiko intervalas

DEMO

AutoML verslo sprendimų sritys

INFOTRUST

Pardavimai

- Laimėjimų / pralaimėjimų prognozė
- Klientų išlaikymas ir praradimas
- Klientų paieška

Marketingas

- Paklausos / pajamų prognozavimas
- Kliento vertė
- Kitas geriausias pasiūlymas klientui

Finansai

- Kapitalo investicijų optimizavimas
- Išlaidų valdymas
- Rizikos mažinimas

Operacijos

- Darbo jėgos paklausos prognozavimas
- Pajėgumų paskirstymas



Žmogiškieji ištekliai

- Darbuotojų išėjimo prognozavimas
- Darbuotojų pasitenkinimas
- Įdarbinimas / kandidatų profiliavimas

IT

- Programinės įrangos / licencijavimo naudojimas
- Infrastruktūros našumo prognozavimas

Tiekimo grandinė

- Atsargų sumažėjimo numatymas
- Tiekimo grandinės veikimas ir kliūtis
- Transporto optimizavimas

Paslaugos ir Aptarnavimas

- Gedimų numatymas
- Priežiūros ir nusidėvėjimo prognozavimas

AutoML verslo klausimų pavyzdžiai

- Kurie potencialūs sandoriai gali būti užbaigti šį ketvirtį ? Kuriuos sandorius užbaigsime sėkmingai? Kokie veiksniai turės tam įtakos?
- Kurie klientai paslaugų atsisakys?
- Kokios pajamos prognozuojamos artimiausiais mėnesiais?
- Kurie klientai atneš daugiausiai pajamų per ateinančius 12 mėn.?
- Į kuriuos klientų segmentus ir demografinius rodiklius turėtume žiūrėti ateityje? Kurie klientų segmentai pirks per ateinančius 3 mėnesius?
- Kokios kapitalo investicijos atneš didžiausią gražą ateityje?
- Kokias išlaidas pagal kategorijas numatome kitą ketvirtį?
- Koks mūsų darbuotojų skaičius turėtų būti ateinančiais metais? Kuriose pozicijose darbuotojų trūks ir reiks samdyti daugiau?

AutoML verslo klausimų pavyzdžiai

- Kaip turėtume paskirstyti savo gamybos pajėgumus ateinančiais metais, kad užtikrintume paklausos poreikį?
- Kurie įmonei vertingi darbuotojai gali išeiti iš darbo?
- Koks darbuotojų pasitenkinimas bus ateinančiais metais?
- Koks numatomas licencijų sunaudojimas ateinančiais metais? Kurios licencijos ar prenumeratos gali būti pilnai neišnaudotos?
- Kaip turėtume planuoti savo infrastruktūros pajėgumus ateinančiais metais? Kur gali kilti pajėgumų trūkumų ar gedimų?
- Kokių atsargų pritrūksime ateinančiais metais?

Ačiū

Milda Vitalytė Masaitė, Qlik architektė

m.vitalyte@theinfotrust.com

The Infotrust, UAB