

Classification Evaluation

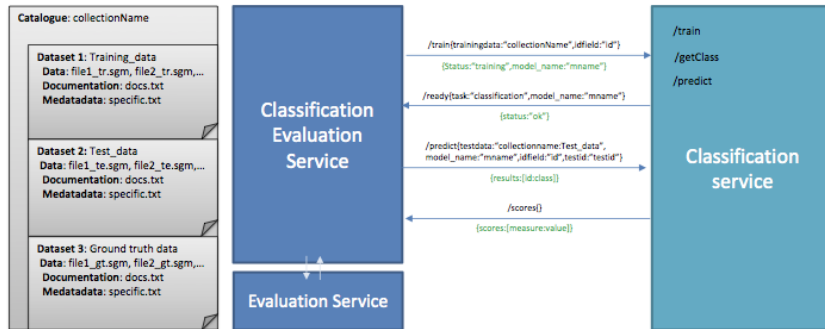


Abb. 1: Service Evaluation als automatisierter Prozess

Hintergrund

Das Data Market Austria (DMA) Projekt etabliert ein Daten-Services Ökosystem in Österreich durch die Schaffung einer deutlich verbesserten Technologiebasis für sichere Datenmärkte und Cloud-Interoperabilität und die Etablierung eines Daten-Innovationsumfeldes.

Diese Plattform bietet die Möglichkeit verschiedenste Service zu integrieren. Um die Qualität von diesen beurteilen zu können ist es hilfreich eine Infrastruktur zur Evaluierung zu haben.

Methode

Dieser Prototyp kann als Basis für den Evaluierungsprozess verstanden werden. In diesem konkreten Fall wird ein Service, welcher mittels Machine Learning eine Klassifizierung eines Datenbestands vornimmt, evaluiert. Die Evaluierung wird gestartet, indem eine beliebige Test Collection und einige Metadaten des Services and die Schnittstelle gesendet wird. Bei einem Klassifizierungsservice

besteht die Test Collection aus den Trainings und Testdaten sowie der Ground Truth Tabelle. Nachdem der Service mit dem Training fertig ist und eine Prediction geliefert hat, kann der Evaluation Service die Daten mit der Ground Truth Tabelle abgleichen und das Resultat berechnen (Confusion Matrix, TP, TN, FP, FN)

Ziel

Das Resultat der Evaluierung soll Einblicke in die Performance des Service bieten und Entwicklern die Möglichkeit geben ihre Resultate mit ähnlichen schon verfügbaren DMA Services auf der DMA Plattform zu vergleichen.

Innovation

- Abstrahierung der Evaluierung für verschiedenste Services (Klassifizierung, Suche, ...)

Nutzen

- Basis für weitere Evaluierungen
- Ermöglicht Vergleich zwischen bestehenden Services der selben Kategorie

Demonstration

- Data Market Austria – “Mini-Plenary”

Publikation

- Matthias Traub, Heimo Gursch, Elisabeth Lex, Roman Kern; Data Market Austria: Austria's First Digital Ecosystem for Data, Businesses, and Innovation; Contributions to the 2nd International Conference on New Business Models

Classification Evaluation Test Driver

[Start Dummy Evaluation](#) [Start Animal Evaluation](#)
[Check latest evaluation](#)
[Show latest evaluation result](#)

```
"classes": [
  {
    "name": "rabbit",
    "truePositives": 9,
    "trueNegatives": 48,
    "falsePositives": 23,
    "falseNegatives": 20,
    "accuracy": 0.57,
    "misclassificationRate": 0.43,
    "truePositiveRate": 0.31,
    "falsePositiveRate": 0.32,
    "falseNegativeRate": 0.69,
    "trueNegativeRate": 0.68,
    "precision": 0.28,
    "prevalence": 0.29
  },
  {
    "name": "cat",
    "truePositives": 8,
    "trueNegatives": 41,
    "falsePositives": 23,
    "falseNegatives": 28,
    "accuracy": 0.49,
  }
]
```

Abb. 2: Einfaches Demonstrator Interface zur Service Evaluierung

Abb. 3: (formatiertes) Ergebnis einer Evaluation eines zufallsbasierten Klassifikations-Services