

Oracle Machine Learning in der öffentlichen Verwaltung



Bewegen Sie die Algorithmen, nicht die Daten

Juli 2020
Copyright © 2020, Oracle and/or its affiliates
Public

INHALTSANGABE

EINFÜHRUNG	3
MACHINE LEARNING	3
ORACLE MACHINE LEARNING – IM DETAIL	5
Bewegen Sie die Algorithmen, nicht die Daten	5
Oracle Database - die Multi-Modell-Datenbank	6
Ermöglichen Sie Benutzern die Verwendung von SQL, R und Python	6
Automatisierung für mehr Produktivität und Modellqualität	7
Produktionsbereitstellung - der entscheidende Schritt	7
Eine Vielzahl von Algorithmen	7
Unterstützung unterschiedlichster Anwender	8
Oracle Big Data und Cloud SQL – Arbeiten in Data Lakes	8
Oracle Machine Learning für Spark	9
Arbeiten mit R auf Hadoop-Clustern	9
Data Science Cloud Service – unternehmensweite Data-Science-Plattform	9
ORACLE MACHINE LEARNING – IM EINSATZ	10
FAZIT	11
WEITERE INFORMATIONEN	12

EINFÜHRUNG

Die Bundesregierung sieht mit dem Einsatz von Künstlicher Intelligenz (KI) im Bereich der öffentlichen Verwaltung die Chance, Informationen und Leistungen zielgerichteter, passgenauer und leichter zugänglich für Bürger und Unternehmen bereitzustellen. Mithilfe der KI-Technologien können die gestiegenen Ansprüche an Verwaltungsprozesse und staatliche Dienstleistungen erfüllt werden. Sie unterstützen Mitarbeiter bei der Erledigung ihrer Aufgaben und verbessern gleichzeitig die Qualität. Das gesamte Verwaltungssystem wird auf diese Weise bürgernäher, effizienter und agiler.

Es gibt vielfältige Anwendungsfälle für diese Technik, so zum Beispiel:

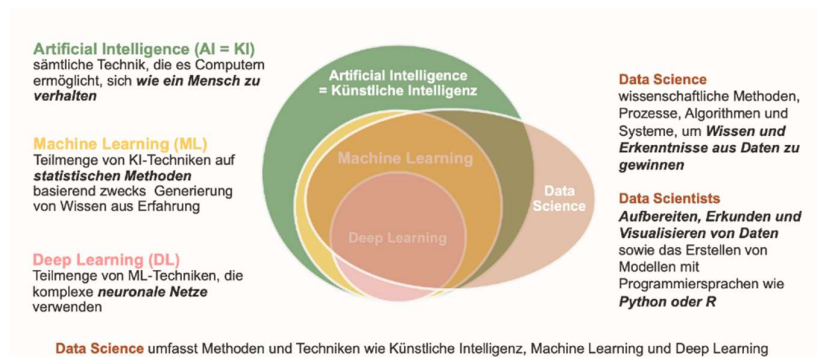
- Automatisierte Vorüberprüfung von Anträgen – zum Beispiel im Zusammenhang mit allen Belangen der Steuerverwaltung oder für jegliche Art von Unterstützungsleistungen im Sozialwesen
- Echtzeit-Monitoring und Trendanalysen in den Bereichen Finanzaufsicht, Verkehrsplanung, öffentliche Sicherheit oder Gesundheitswesen – zum Beispiel Predictive Analytics oder Epidemiologie
- Mustererkennung bei jeglicher Form von Betrug – zum Beispiel in der Finanzverwaltung, dem Sozial- oder Gesundheitswesen u.v.m.

Damit im öffentlichen Bereich in Zukunft KI- und ML-Technologien mit hohem Nutzen eingesetzt werden können, ist es notwendig, die möglichen Anwendungsfälle zu identifizieren und Wissen über diese Technologien aufzubauen.

Für den Einsatz von KI- und ML-Methoden sind die Verfügbarkeit und die Qualität von Daten die zentrale Voraussetzung. Daten sind der bestimmende Faktor für die Verwendbarkeit der Ergebnisse. Gleichzeitig ist der Schutz der Daten, die genutzt werden, von essentieller Bedeutung. Oracle Machine Learning unterstützt dabei, die gesteckten Ziele im Hinblick auf diese Aspekte zu erreichen.

MACHINE LEARNING

Machine Learning ist ein Teilbereich des Themenfeldes KI (oder auch Artificial Intelligence AI). Es basiert auf statistischen Methoden. Sein Zweck ist es, Wissen aus Erfahrung zu generieren.

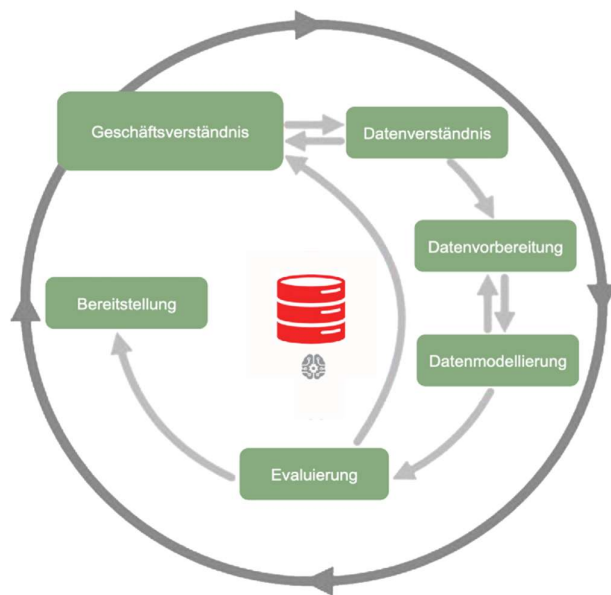


Definition von Data Science, AI, ML und Deep Learning

Die für Machine-Learning-Analysen genutzten Datenmengen nehmen stetig zu. Externe Daten ergänzen die intern verfügbaren Daten. Eine Vielfalt an Datentypen wird für die Analysen herangezogen. Dazu gehören unstrukturierter Text, Bild- und Tondaten, Sensordaten sowie raumbezogene und graphische Daten. Daten werden mit großer Geschwindigkeit neu erzeugt und fließen in die ML-Analysen mit ein.

Hohe Relevanz Künstlicher Intelligenz und Machine Learning in der öffentlichen Verwaltung

Machine Learning: ein Teilbereich der Künstlichen Intelligenz]



Die sechs Phasen des Oracle-Standard-Prozesses für Data Mining

Oracle-Standard-Prozess für Data Mining – angelehnt an den Cross-Industry-Standard für DM (CRISP-DM)

Data Mining als Teil von AI- und ML-Methoden ist ein Prozess, an dessen Anfang eine Fragestellung steht, wie zum Beispiel „Wie kann ich feststellen, ob ein Zustand normal oder anormal ist?“ oder „Wann ist ein Antrag bearbeitbar?“. Zu diesen Fragen werden nun mithilfe von sechs unterschiedlichen Phasen Antworten gesucht. Hierbei wird häufig zwischen den verschiedenen Phasen hin- und hergewechselt.

Folgende Phasen werden unterschieden:

1. *Geschäftsverständnis*: Festlegung der Ziele und Anforderungen; Ableitung der konkreten Fragestellung und der groben Vorgehensweise
2. *Datenverständnis*: Datensammlung bzw. erste Sichtung der zur Verfügung stehenden Daten; Ermittlung möglicher Probleme mit Datenqualität
3. *Datenvorbereitung*: Festlegung des finalen Datensatzes für die Modellierung
4. *Modellierung*: Anwendung der analytischen Methodik, Optimierung der Parameter; in der Regel Ermittlung mehrerer Modelle
5. *Evaluierung*: Interpretation der Ergebnisse und Auswahl des Modells, das die Aufgabenstellung am besten erfüllt. Sorgfältiger Abgleich mit der Fragestellung.
6. *Bereitstellung*: Automatisierung, eventuell Integration des Modells in vorhandene Prozesse

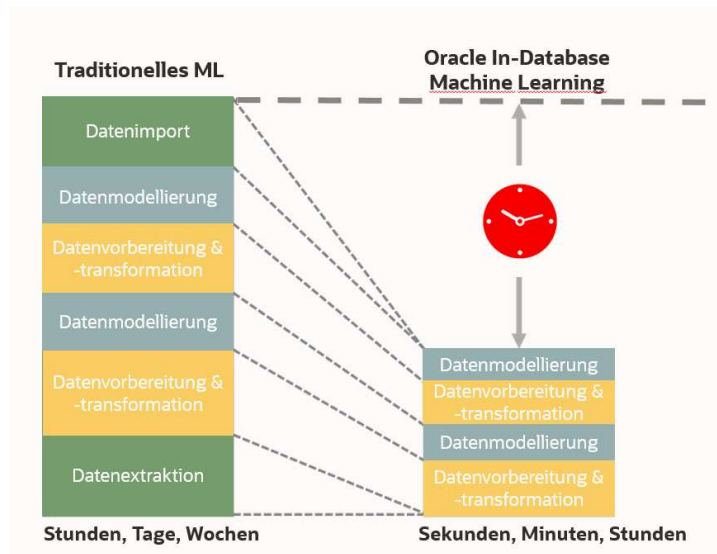
Machine Learning Projekte beginnen für gewöhnlich mit einer repräsentativen Stichprobe von Daten, die auf separate, für Datenanalysten entwickelte Analyseserver exportiert und mit speziellen Werkzeugen untersucht werden. Als statistische Sprache kommen oft SAS, R und Python zum Einsatz. Zeigt sich das verwendete Vorhersagemodell als erfolgversprechend, muss das IT-Referat eine Umgebung bereitstellen, auf der mit der vollständigen Datenbasis gearbeitet werden kann. Dieser Schritt kann eine erhebliche Herausforderung im Hinblick auf die Datenbereitstellung darstellen, inklusive der Integration unterschiedlichster Datenquellen. Zudem müssen Vorhersagemodelle und alle Schritte, die mit der Anwendung verbunden sind, für die umfangreichere Datenbasis verfügbar gemacht und gegebenenfalls angepasst werden: Dies kann entweder dadurch geschehen, dass der anzuwendende Code für die

**Zeit- und ressourcenintensive
herkömmliche ML-
Entwicklungsprozesse**

Datenbank in SQL neu erstellt wird oder indem eine Anwendung in Java entwickelt wird, die die gewünschten Algorithmen ausführt.

Die Modellübersetzungsphase beinhaltet aufwändige manuelle Codierungsschritte. Die Fähigkeit, Daten in großem Maßstab für Machine Learning Prozesse zu integrieren und vorzubereiten, ist dabei von entscheidender Bedeutung. Darüber hinaus ist es wichtig, die Zeit von der Definition einer neuen Aufgabenstellung bis zur Bereitstellung der Lösung so kurz wie möglich zu halten. Es ist in der Vergangenheit oft geschehen, dass Machine Learning Projekte nicht zu Ende geführt wurden, weil mit der Modellübersetzungsphase unerwartet große Schwierigkeiten verbunden waren. ORACLE MACHINE LEARNING

Oracle Machine Learning stellt eine Palette leistungsstarker Algorithmen in der Oracle Datenbank zur Verfügung. Auf diese Algorithmen kann über SQL, Python- oder R-APIs zugegriffen werden. Die Oracle Machine Learning Plattform reduziert im Vergleich zu herkömmlichen Machine Learning Produkten die Zeit von der Daten- bis zur Modellbereitstellung von Wochen auf Stunden.



Hohe Zeitersparnis und Sicherheit durch Oracle-Datenbank-basiertes ML

Die Oracle Machine Learning Plattform reduziert im Vergleich die Zeit von der Daten- bis zur Modellbereitstellung von Wochen auf Stunden.

Oracle Machine-Learning-Algorithmen nutzen alle zugehörigen SQL-Funktionen. Sie können Daten in ihrer ursprünglichen Star-Schema-Darstellung verwenden. Verarbeitet werden können strukturierte Standardtabellen und -ansichten, Transaktionsdaten und Aggregate, unstrukturierte Daten, wie sie in CLOB-Datentypen enthalten sind, sowie räumliche und graphische Daten. Die In-Database-Algorithmen von Oracle Machine Learning nutzen die Fähigkeiten der Oracle Datenbank wie zum Beispiel Parallel Execution sowohl für die Modellbildung als auch für die Modellanwendung, das Scoring. Die Sicherheits- und Benutzerberechtigungskonzepte sowie die Revisionskontrolle und das Audit-Tracking der Datenbank werden auch für Oracle ML genutzt.

ORACLE MACHINE LEARNING – IM DETAIL

Bewegen Sie die Algorithmen, nicht die Daten

Mit Oracle Machine Learning bewegt Oracle die Algorithmen zu den Daten. Oracle führt leistungsstarke ML-Algorithmen in der Oracle Datenbank aus, in denen sich die Daten befinden, ob lokal im Rechenzentrum oder in der Cloud. Leistungsstarke Analysen werden direkt im Kernel der Datenbank durchgeführt. Ergebnisse, Erkenntnisse und Modelle für ML sind in Echtzeit in der Oracle Datenbank verfügbar und werden dort verwaltet. Dieser Ansatz minimiert oder

Schnellere Modellentwicklung und mehr Datensicherheit durch weniger Datenbewegungen

eliminiert Datenbewegungen, verbessert Skalierbarkeit, bewahrt Datensicherheit und beschleunigt somit die Bereitstellung von Modellen.

Beim Erstellen von Modellen nutzt Oracle Machine Learning standardmäßig skalierbare Oracle Datenbanktechnologien wie zum Beispiel parallele Ausführung, Bitmap-Indizes und Aggregationstechniken. Auch der hochperformante In-Memory spaltenorientierte Speicher der Datenbank kann für die Modellerstellung und -anwendung verwendet werden. Weitere Technologien wie zum Beispiel parallele Infrastruktur, IEEE-Float, automatische Datenaufbereitung für das Binning, Behandlung fehlender Werte und die Unterstützung für unstrukturierte Textdaten kommen zum Einsatz. Mit diesen Fähigkeiten der Oracle Datenbank erzielt Oracle Machine Learning die Performance, Skalierbarkeit und Automatisierung, die für Data-Science-Projekte erforderlich sind - sowohl On-Premises, in einer Edge Computing Architektur als auch in der Cloud.

Oracle Database - die Multi-Modell-Datenbank

Um auf verschiedene Analysefunktionen zugreifen zu können, sollte man nicht mehrere Datenbanken erstellen und verwalten müssen. Dies erhöht nur die Komplexität und die Kosten. Stattdessen sollten alle Funktionen in einer einzigen Datenbank mit mehreren Datenmodellen zur Verfügung stehen. In der Oracle Datenbank ist eine breite Palette von Algorithmen vorhanden, die Daten mit verschiedenen Datentypen und Datenmodellen verarbeiten können. Dies ist ein wesentliches Unterscheidungsmerkmal und ermöglicht es Oracle, fortschrittliche Entwicklungstools für die unterschiedlichsten Entwickler bereitzustellen.

Geringere Komplexität und Kosten durch die Multi-Modell-Datenbank

Ermöglichen Sie Benutzern die Verwendung von SQL, R und Python

R und Python bieten eine Vielzahl statistischer und graphischer Anwendungen. Zusätzliche Open Source Bibliotheken können herangezogen werden. Typische R- und Python-Anwender stehen jedoch vor Herausforderungen hinsichtlich Speicherbeschränkungen und Single-Thread-Ausführung. Zudem lassen sich in R und Python geschriebene Programmpakete nur mit erheblichem Aufwand in Behördenanwendungen oder Dashboards integrieren.

Erlaubt R- und Python-Anwendern ohne dedizierte SQL-Kenntnisse Zugriff auf Datenbankinhalte und -funktionalitäten

Oracle Machine Learning bietet an dieser Stelle native Schnittstellen für die wichtigsten Data-Science-Sprachen wie SQL, R und Python. Durch die Unterstützung von R und Python greifen Data Scientists ohne SQL-Kenntnisse auf Oracle Datenbanken als leistungsstarke KI-/ML-Plattform für skalierbare Datenexploration und -vorbereitung zu. Die Oracle Datenbank wird als Modell- und Prozess-Verwaltung integriert, mit allen Sicherheitskonzepten der Oracle Datenbank.

Oracle Machine Learning for Python* (OML4Py) integriert Python in die Oracle Datenbank und stellt ausgewählte Python-Funktionen als äquivalente SQL- und datenbankinterne Algorithmen zur Verfügung. Dadurch können Python-Benutzer Daten, die in der Datenbank liegen, ohne Verwendung von SQL bearbeiten. OML4Py wird über Oracle Machine Learning Notebooks aufgerufen. Auf die gleiche Weise werden R-Funktionen für R-Benutzer mit Oracle Machine Learning for R (OML4R) bereitgestellt. OML4R steht auf lokalen Instanzen der Oracle Datenbank und den Oracle Database Cloud Services zur Verfügung. Mit der eingebetteten R-Ausführung von OML4R können Benutzer die native Funktionalität von Oracle Machine Learning erweitern. Es können benutzerdefinierte Funktionen erstellt werden, die die auf dem Datenbankserver installierten Open-Source-R-Pakete von Drittanbietern verwenden.

Automatisierung für mehr Produktivität und Modellqualität

Ein wesentlicher Teil des Machine-Learning-Prozesses besteht in der Iteration. In gängigen Methoden wie CRISP-DM sind diese iterativen Schleifen integriert. Dabei werden verschiedene Algorithmen und Variationen in den Einstellungen ausprobiert, um den besten Algorithmus zu finden und dessen Hyperparameter zu optimieren. Auf diese Art wird das beste Modell für ein bestimmtes Geschäftsproblem gefunden und ausgewählt.

Die in Oracle Machine Learning bereitgestellte Automatisierung umfasst die automatische Datenaufbereitung (ADP), die automatisierte Textverarbeitung, automatisierte partitionierte Modelle und das automatisierte Machine Learning (AutoML) in OML4Py. Letzteres konzentriert sich auf die Auswahl der passenden Algorithmen, der aussagekräftigen Datenmerkmale (Features) und die Optimierung von Hyperparametern. All diese Funktionen tragen dazu bei, den Aufwand und in einigen Fällen das nötige Wissen zu reduzieren, das erforderlich ist, um mithilfe von Machine Learning von der Fragestellung zur Lösung zu gelangen. So unterstützt AutoML auch Nicht-Experten dabei, von der in die Datenbank integrierten Intelligenz zu profitieren, um bessere ML-Modelle zu erstellen.

Schnelle und effiziente Modellbereitstellung durch hohen Automatisierungsgrad

Produktionsbereitstellung - der entscheidende Schritt

Man könnte die Produktionsbereitstellung als Achillesferse von Data-Science-Projekten bezeichnen. Wenn Anwendungsentwickler oder IT-Mitarbeiter versuchen, ML-Modelle oder Open-Source-R- und Python-Skripte in die Produktion zu überführen, müssen sie die Themen Sicherheit, Wiederherstellung, Sicherheit, Skalierbarkeit und Versionsabhängigkeiten berücksichtigen. Mit Oracle Machine Learning erfolgt diese Bereitstellung implizit. Von SQL abgeleitete Modelle sind in der Datenbank vorhanden und können über SQL-Abfragen aufgerufen werden. R- und Python-Skripte können in der Oracle-Datenbank gespeichert, automatisch erzeugt und gesteuert ausgeführt werden.

Sicherheit durch In-Database-Bereitstellung

Eine Vielzahl von Algorithmen

Oracle Machine Learning bietet eine Bibliothek von skalierbaren und parallelisierten Algorithmen und Analysefunktionen in Oracle-Datenbank- und Big-Data-Umgebungen. Dazu gehören Clustering, Regression, Klassifizierung, Zuordnungsregeln, Merkmalsextraktion, Zeitreihen und Erkennung von Anomalien. Oracle Machine Learning stellt diese Algorithmen als SQL-Funktionen zur Verfügung. Darauf zugegriffen werden kann mit SQL (Oracle Machine Learning für SQL), per Drag-and-Drop über die Benutzeroberfläche des Oracle Data Miner, auch mit Open Source R und Python und in einer Big-Data-Umgebung unter Apache.

Komfortable Flexibilität durch leistungsstarke Algorithmen

CLASSIFICATION <ul style="list-style-type: none">• Naive Bayes• Logistic Regression (GLM)• Decision Tree• Random Forest• Neural Network• Support Vector Machine (SVM)• Explicit Semantic Analysis• XGBoost	CLUSTERING <ul style="list-style-type: none">• Hierarchical K-Means• Hierarchical O-Cluster• Expectation Maximization (EM) TEXT MINING SUPPORT <ul style="list-style-type: none">• Algorithms support text columns• Tokenization and theme extraction• Explicit Semantic Analysis (ESA) for document similarity	REGRESSION <ul style="list-style-type: none">• Linear Model• Generalized Linear Model (GLM)• Support Vector Machine (SVM)• Stepwise Linear regression• Neural Network• LASSO• XGBoost ATTRIBUTE IMPORTANCE <ul style="list-style-type: none">• Minimum Description Length• Principal Component Analysis (PCA)• Unsupervised Pair-wise KL Div• CUR decomposition for row & AI
TIME SERIES <ul style="list-style-type: none">• Forecasting - Exponential Smoothing• Includes popular models e.g. Holt-Winters with trends, seasonality, irregularity, missing data	STATISTICAL FUNCTIONS <ul style="list-style-type: none">• Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.	PREDICTIVE QUERIES <ul style="list-style-type: none">• Predict, cluster, detect, features
FEATURE EXTRACTION <ul style="list-style-type: none">• Principal Comp Analysis (PCA)• Non-negative Matrix Factorization• Singular Value Decomposition (SVD)• Explicit Semantic Analysis (ESA)	R AND PYTHON PACKAGES <ul style="list-style-type: none">• Third-party R and Python Packages through Embedded Execution• Spark MLlib algorithm integration ANOMALY DETECTION <ul style="list-style-type: none">• One-Class SVM• MSET-SPRT	SQL ANALYTICS <ul style="list-style-type: none">• SQL Windows• SQL Patterns• SQL Aggregates
ASSOCIATION RULES <ul style="list-style-type: none">• A priori/ market basket		

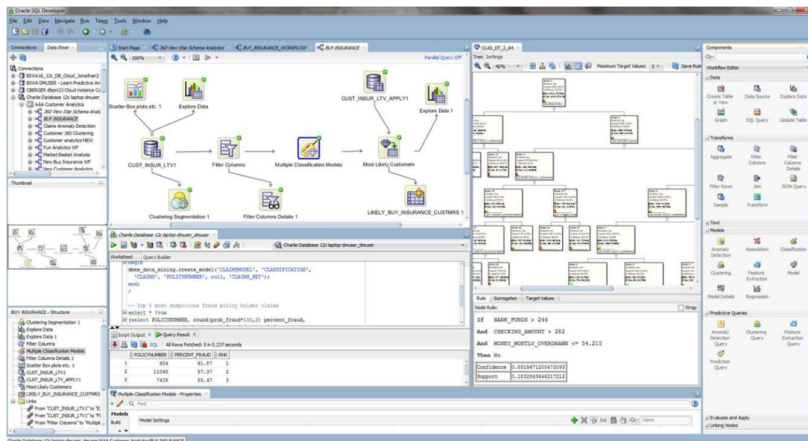
Oracle Machine Learning bietet eine Vielzahl von Algorithmen und Analysefunktionen.

Unterstützung unterschiedlichster Anwender

Data-Science-Projekte erfreuen sich in der öffentlichen Verwaltung zunehmender Beliebtheit. Der Erfolg dieser Projekte basiert nicht nur auf geeigneten Daten, einer entsprechenden Datenaufbereitung, Visualisierung und der Anwendung von ML-Algorithmen. Der Erfolg basiert auch auf der Fähigkeit, Lösungen schnell und einfach mit dem notwendigen Fachwissen bereitzustellen. Dies führt zu einem zunehmend breiteren ML-Anwenderspektrum über die Rolle des Data Scientists hinaus. Zukünftig werden auch Fach- und Datenanalysten, Datenbankadministratoren und IT-Experten, Anwendungs- und Dashboard-Entwickler sowie Führungskräfte von Data-Science-Projekten mit ML-Technologien arbeiten.

Hier unterstützt der Oracle Data Miner, eine Erweiterung des Entwicklungswerkzeugs Oracle SQL Developer. Der Oracle Data Miner ermöglicht das Arbeiten mit einer benutzerfreundlichen Oberfläche und setzt keine SQL-Kenntnisse voraus. Gearbeitet wird mithilfe von Drag-and-Drop, um die Workflows zu erstellen. Über diese wird dann direkt mit Daten in der Datenbank gearbeitet. Datenanalysten können ihre Daten schnell graphisch visualisieren und untersuchen, aber auch Daten aufbereiten und transformieren. Mehrere ML-Modelle können erstellt und anhand statistischer Kenngrößen, wie zum Beispiel ROC und Lift, bewertet werden. Umfangreiche Viewer für Modelle und Modellbewertungen können verwendet werden. Anschließend können Oracle-Machine-Learning-Modelle auf neue Daten angewendet oder SQL- und PL/SQL-Skripte generiert werden, um den analytischen Workflow bereitzustellen. Diese Skripts können dann zur Anwendung in der Oracle Datenbank an Datenbankadministratoren übergeben werden. Die Oracle Data Miner PL/SQL Workflow API erlaubt die einfache Integration von Vorhersagemethoden in Anwendungen für eine breite Verwendung in der gesamten Organisation einer Behörde.

Drag-and-Drop-Funktionalitäten für einen breiten Anwenderkreis



Oracle Data Miner bietet per Drag-&-Drop-Workflow eine Benutzeroberfläche, mit der Sie Daten untersuchen, Vorhersagemodelle erstellen, auswerten, anwenden und erweiterte Analysemethoden als SQL- und PL/SQL-Skripte bereitstellen können.

Oracle Big Data und Cloud SQL – Arbeiten in Data Lakes

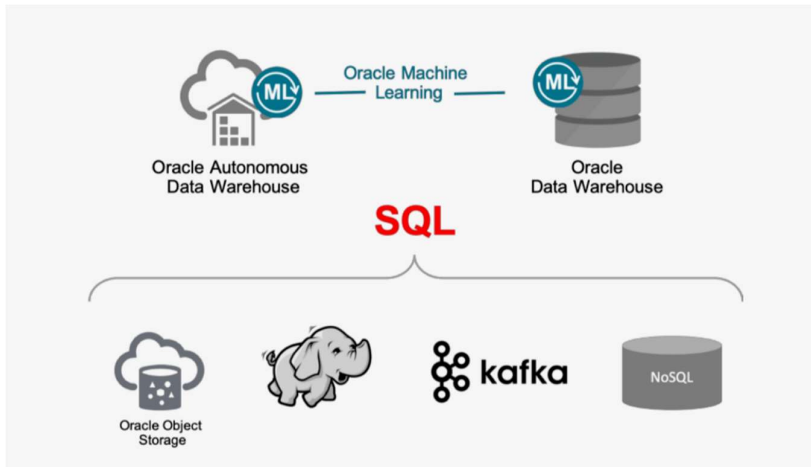
Daten von Sensoren, dem Internet der Dinge, „Tweets“ aus sozialen Netzwerken und aus anderen Quellen mit einem hohen Volumen werden heutzutage oft in einer Hadoop-Umgebung gespeichert und als Data Lake oder Data Reservoir bezeichnet. Diese Datenumgebung außerhalb der Oracle Datenbank bringt neue Herausforderungen bei der Datenverwaltung und Datenanalyse mit sich.

Oracle bietet mit Big Data SQL oder Cloud SQL die Möglichkeit, diese Massendaten mit SQL zu verarbeiten und somit die Datenbank-Daten mit den Data-Lake-Daten direkt zu verknüpfen. Benutzer können per Cloud SQL Massendaten aus Data Lakes auf verschiedenen Ebenen filtern. Die Data-Lake-

Einbindung von Massendaten aus Hadoop-Quellen durch erweiterte, SQL-basierte Vorselektion

Daten werden in Hadoop bereits verarbeitet. Dadurch können Daten aus den dortigen Speicherformaten wie Dateien, Tabellen, Datenströmen oder NoSQL-Datenbanken transparent für Abfragen zur Verfügung gestellt werden.

Somit können Daten aus Data Lakes auch für alle Analysen und ML-Prozesse verwendet werden und ergänzen somit die Daten der Datenbank und des Data Warehouses. Diese All-Data-Architektur erweitert erheblich den Nutzen des vorhandenen Datenbestandes.



Mit SQL und Cloud-SQL können Datenanalysten auf Daten von Hadoop-Servern und der Datenbank zugreifen, diese zusammenfassen, filtern und aggregieren. Diese wiederum werden für eine umfassendere 360-Grad-Ansicht kombiniert. Mithilfe von Oracle Machine Learning werden dann Vorhersagemodelle erstellt.

Oracle Machine Learning für Spark

Arbeiten mit R auf Hadoop-Clustern

R stellt in einigen großen Projekten die Datenschnittstelle für die unterschiedlichsten Datenquellen als Drehscheibe zur Verfügung. Darunter fallen sowohl Daten aus Dateien, Datenbanken und JDBC-Quellen als auch Hadoop-Quellen wie HDFS, HIVE, Impala und Spark Data Frames.

Hier ermöglicht Oracle Machine Learning for Spark (OML4Spark) einen transparenten Zugriff und eine transparente Datenverwendung über diese Grenzen in R. Die R-Bibliothek erlaubt die Entwicklung von Modellen, Funktionen und R-Paketen über alle Datenquellen und bietet damit auch für R-Entwickler eine All-Data-Architektur. Dazu werden die Algorithmen und Datenobjekte transparent in Spark übertragen und auf den Hadoop-Knoten ausgeführt. Die OML4Spark-Funktionen sind für parallele Ausführungen optimiert und nutzen alle Knoten eines Hadoop-Clusters für skalierbares, leistungsstarkes ML in einer Big-Data-Umgebung. Damit können große, komplexe und datengesteuerte Aufgabenstellungen gelöst werden, bei denen die zu entdeckenden Quelldaten und Muster in Big Data, relationalen Daten oder einer Kombination aus beidem liegen können.

Ermöglicht R-Entwicklern das Lösen von komplexen Aufgabestellungen innerhalb von Big-Data-Clustern

Data Science Cloud Service – unternehmensweite Data-Science-Plattform

Mit dem Oracle Data Science Cloud Service können Data-Science-Teams unternehmensweit ihre Arbeit einfach organisieren, auf Daten und Computerressourcen zugreifen sowie in Python erstellte Modelle in der Oracle Cloud erstellen, trainieren, bereitstellen und verwalten. Die projektbasierte Benutzeroberfläche ermöglicht das gemeinsame Arbeiten an End-to-End-Modellierungsworkflows, unterstützt durch eine integrierte Versionskontrolle und High Performance Computing Instanzen (HPC).

Die Cloud-Plattform für unternehmensweite, integrierte und produktive Zusammenarbeit

ORACLE MACHINE LEARNING – IM EINSATZ

Viele Oracle-Kunden haben mit Oracle Machine Learning beeindruckende Ergebnisse erzielt, darunter:

NHS Business Services Authority

Der britische National Health Service (NHS), das System der öffentlichen Gesundheitsdienstleister im Vereinigten Königreich, optimiert mit Oracle Machine Learning die Patientenversorgung. NHS sammelt und analysiert Milliarden von Daten zu Rezepten, Medikamenten, Arztbeziehungen und Call-Center-Diensten aus der gesamten Organisation. Auf dieser Grundlage deckt NHS zum Beispiel neue positive Wirkungsgrade von Behandlungen auf und erreicht so eine schnellere und damit bessere Heilung. Des Weiteren werden auf diese Art betrügerische Handlungsmuster aufgedeckt.

Optimierung der Patientenversorgung dank Oracle Machine Learning



Mithilfe von Vorhersagen visualisiert das italienische Ministerium für Inneres die Regionen und Bereiche mit dem höchsten Kriminalitätsrisiko. Oracle Spatial Analytics ermöglicht die Mustererkennung und Ermittlung von Wahrscheinlichkeiten für mögliche Tatorte und Tatzeiten auf Basis des aktuellen Kriminalitätslagebild. Auf dieser Basis wird eine ortsbezogene Kriminalprognose erstellt. Die daraus gewonnen Erkenntnisse werden auf die mobilen Endgeräte der Polizei übermittelt und steuern so die Einsatzrouten der Polizei.

Kriminalprognose mit Predictive Analysis



Mit Oracle-Technologie gewonnene ortsbezogene Kriminalprognosen steuern die Einsätze der Polizei.



StubHub, der weltweit größte Ticketmarktplatz mit Sitz in USA, verwendet datenbankinterne Oracle-Machine-Learning-Modelle und integrierte R-Funktionen, um Betrugserkennungsmodelle in Echtzeit in seiner Datenbank auszuführen. Mit dieser Lösung konnte StubHub den Online-Betrug um 90% reduzieren. Mit Oracle Machine Learning kann StubHub sehr schnell auf die sich immer schneller ändernden Betrugsmuster reagieren. Während die Anwendung und die Datenbank noch ausgeführt werden, erfolgt bereits die Aktualisierung

Schadensvermeidung durch schnelle Betrugserkennung

des Vorhersagemodells. Somit kann bereits eine Stunde nach Entdeckung eines Betrugsfalls das neue Modell ohne Ausfallzeiten bereitgestellt werden.



CaixaBank ist Spaniens führendes Finanzdienstleistungsunternehmen für Banken und Versicherungen mit mehr als 13 Millionen Kunden und 5.600 Filialen. Es ist die größte Privatkundenbank des Landes und bietet innovative Dienstleistungen, mit denen Kunden jederzeit und von überall aus Finanztransaktionen durchführen können. Um den zukünftigen geschäftlichen Herausforderungen und den sich schnell verändernden Kundenanforderungen gerecht zu werden, investierte die spanische CaixaBank in Big-Data- und Oracle-Lösungen. Mit dem Projekt Datenpool, verkürzt CaixaBank nun die Zeiträume für die Projektentwicklung und beschleunigt die Verwaltung von Daten auf ihrem alten Mainframe. Dadurch lassen sich viel größere Datenmengen effizient analysieren und gleichzeitig sicher verwalten. Dies führt zu mehr Effizienz, verbesserter Rentabilität, erhöhter struktureller Flexibilität und größerer Transparenz.

Weitere Informationen zu diesen und anderen Kundengeschichten finden Sie unter <https://www.oracle.com/database/technologies/datawarehouse-bigdata/oml-customers.html>.

FAZIT

Machine Learning bietet neue Möglichkeiten zur Analyse von Daten mit Hilfe statistischer Methoden. Insbesondere bei sehr großen Datenmengen, komplexen Datenstrukturen und vielfältigen Datenformaten können zusätzliche Erkenntnisse gewonnen werden, was ohne Einsatz dieser Methoden nicht oder nur mit einem unverhältnismäßig hohen Aufwand möglich gewesen wäre. Entscheidungsprozesse können dadurch qualitativ verbessert und beschleunigt werden. Organisatorische Prozesse können leichter automatisiert oder teilautomatisiert werden.

Für Behörden, die die neuen ML-Technologien für sich nutzen möchten, sind die schnelle Erfassung und das Sammeln von Daten, die ordnungsgemäße Datenverwaltung, die Zusammenstellung relevanter Daten sowie die Analyse, Automatisierung und Bereitstellung der umsetzbaren Erkenntnisse der Schlüssel zum Erfolg.

Oracle Machine Learning reduziert das zeitaufwendige Modellieren von geeigneten Algorithmen durch die multimodale Datenplattform erheblich. Es bietet eine Bibliothek skalierbarer, parallelisierter ML-Algorithmen in der Oracle Datenbank. Diese Bibliotheken bieten ML-Techniken einschließlich Clustering, Regression, Klassifizierung, Zuordnungsregeln, Merkmalsextraktion, Zeitreihen und Erkennung von Anomalien. Oracle Machine Learning stellt diese Algorithmen als SQL-Funktionen zur Verfügung. Auf diese kann über SQL (Oracle Machine Learning for SQL), über die Workflow-Benutzeroberfläche des Oracle Data Miner per Drag & Drop, dank der engen Integration mit Open Source R (Oracle Machine Learning for R) und innerhalb einer Big Data Umgebung (Oracle Machine Learning for Spark) zugegriffen werden.

Die Algorithmen von Oracle Machine Learning werden in der Oracle-Datenbank ausgeführt. Sie nutzen damit die Skalierbarkeit, Sicherheit und Integrationsfähigkeit der Oracle Datenbank sowohl für strukturierte als auch für unstrukturierte Daten voll aus. Dies macht Oracle zur idealen Plattform für Big-Data- und Analyseverfahren, sowohl im eigenen Rechenzentrum als auch in der Oracle Cloud. Oracle Machine Learning baut auf der jahrzehntelangen Erfahrung von Oracle im Bereich Datenmanagement auf. Es kombiniert diese Erfahrungen mit der Strategie der Verschiebung der Algorithmen zu den Daten.

Effizientere Analyse und größere Transparenz mithilfe von Oracle-Big-Data-Lösung

Bessere Entscheidungsfindung durch neue Analysemöglichkeiten

Sichere ML-Projektumsetzung durch Multi-Modell-Datenbank

Schnelle Ergebnisse durch erprobte In-Database-Funktionalitäten

Durch die Integration von Big Data Management und Analyse in eine einzige einheitliche Oracle-Plattform reduziert Oracle die Gesamtbetriebskosten, eliminiert Datenbewegungen und bietet den schnellsten Weg, um verwaltungsweit ML-Verfahren bereitzustellen.

Öffentliche Verwaltung und Wirtschaft haben gemeinsam sowohl staatliche als auch private Datensilos aufgebrochen und als Open Data in vollem Umfang nutzbar gemacht. Auch das Internet bietet unzählige Möglichkeiten, zusätzliches Wissen aus großen Datenmengen zu generieren. Jetzt gilt es, diese Daten in die vorhandenen behördeneigenen Datenbestände zu integrieren, mithilfe der neuen ML-Technologien immer wieder aufs Neue und mit großer Geschwindigkeit erweiterte und neue Erkenntnisse zu erlangen, diese in die Verwaltungsprozesse einfließen zu lassen und damit gleichzeitig die Mitarbeiter in der Verwaltung zu entlasten. Auf diese Weise kann mit ML für alle Beteiligte ein hoher Nutzen durch den Einsatz von KI und ML erzielt werden.

Reduzierung der Gesamtbetriebskosten durch Eliminierung von Datenbewegungen

Fokussierung der Arbeitsschwerpunkte durch Einsatz von Oracle Machine Learning

WEITERE INFORMATIONEN

Finden Sie unter <https://oracle.com/machine-learning>

* in Kürze verfügbar

Kontaktieren Sie Oracle Deutschland für weitere Fragen: 0180 2 672253 oder info_de@oracle.com

Oracle Niederlassungen: <https://www.oracle.com/de/corporate/contact/field-offices.html>

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

Copyright © 2020, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0120

WHITE PAPER
Oracle Machine Learning in der öffentlichen Verwaltung
June 2020

