The basics of Machine Learning

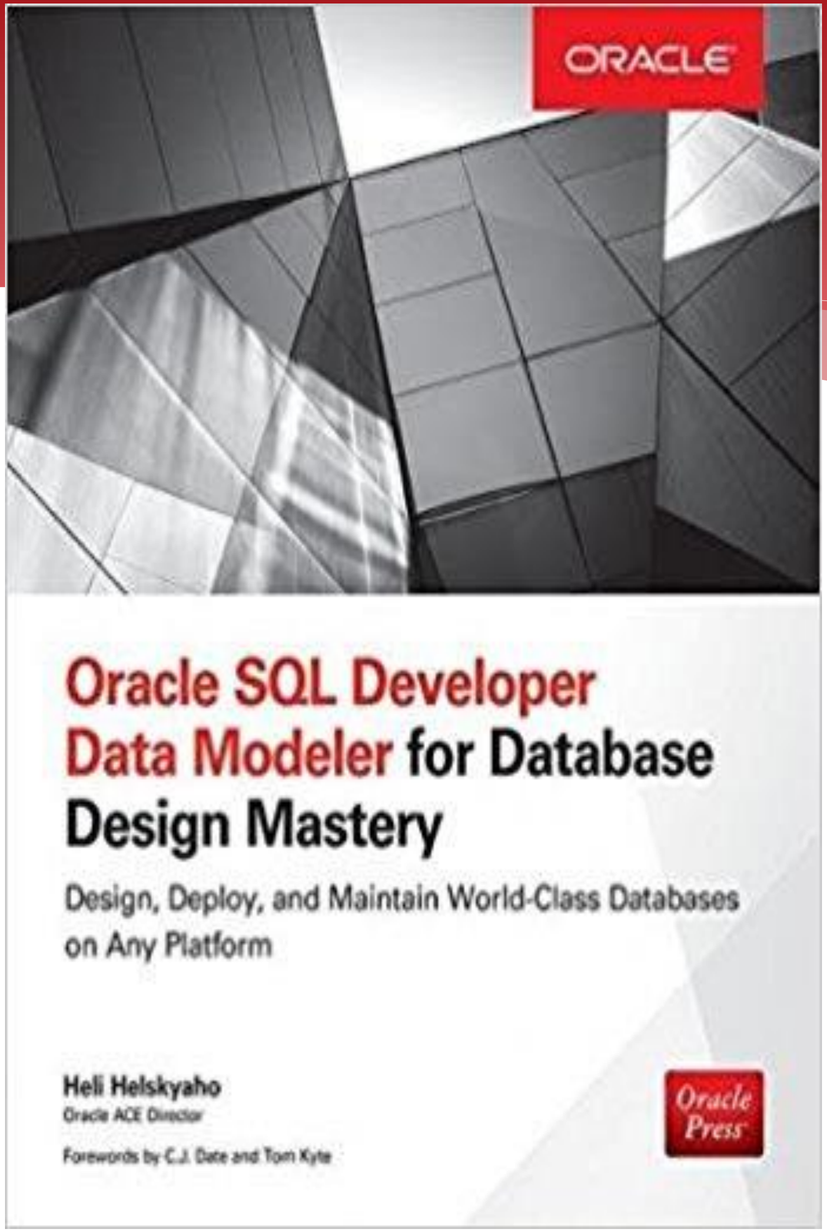
Heli Helskyaho

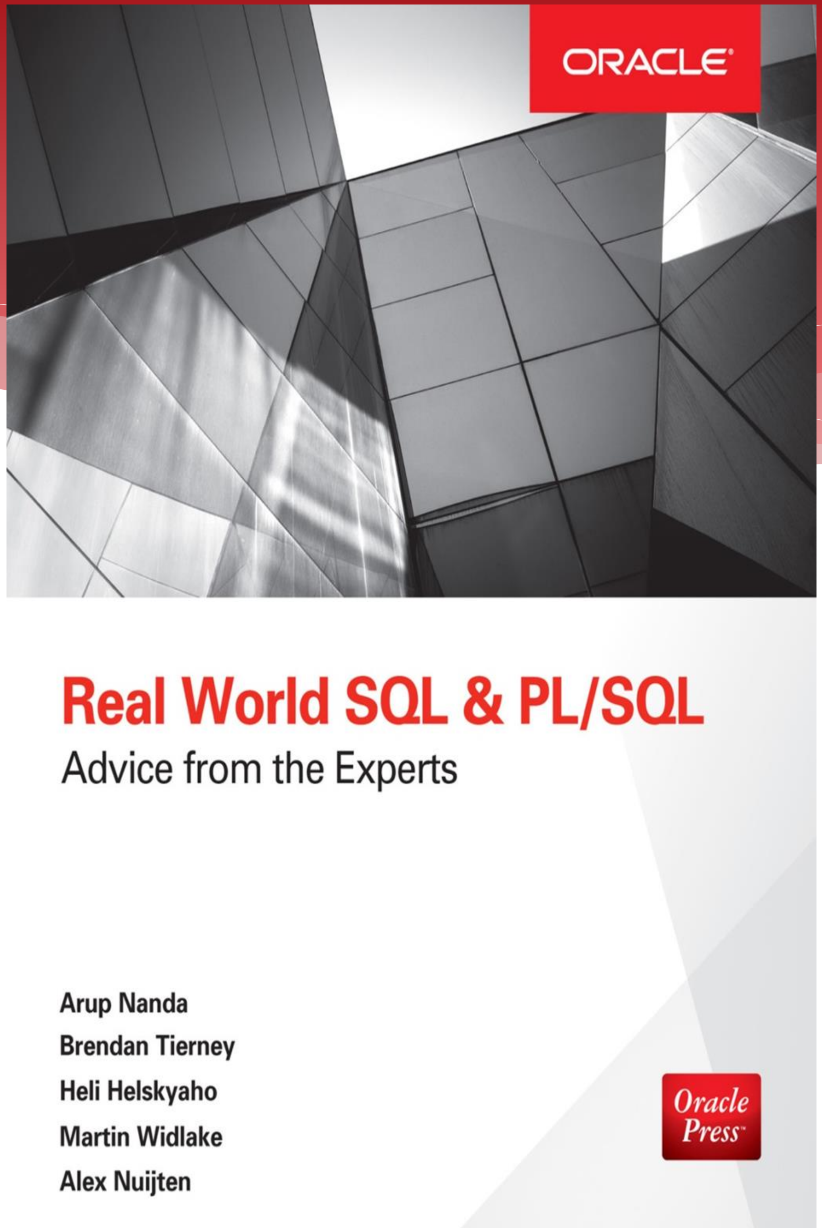
ITOUG Tech Day Rome 2019

Miracle Finland Oy

Introduction, Heli

* Graduated from University of Helsinki (Master of Science, computer science), currently a doctoral student, researcher and lecturer (databases, Big Data, Multi-model Databases, methods and tools for utilizing semi-structured data for decision making) at University of Helsinki
* Worked with Oracle products since 1993, worked for IT since 1990
* Data and Database!
* CEO for Miracle Finland Oy
* Oracle ACE Director
* Ambassador for EOUC (EMEA Oracle Users Group Community)
* Listed as one of the TOP 100 influencers on IT sector in Finland (2015, 2016, 2017, 2018)
* Public speaker and an author
* Winner of Devvy for Database Design Category, 2015
* Author of the book Oracle SQL Developer Data Modeler for Database Design Mastery (Oracle Press, 2015), co-author for Real World SQL and PL/SQL: Advice from the Experts (Oracle Press, 2016)


ORACLE
Oracle SQL Developer Data Modeler for Database Design Mastery
Design, Deploy, and Maintain World-Class Databases on Any Platform
Heli Helskyaho
Oracle ACE Director
Forewords by C.J. Date and Tom Kyte
Oracle Press


ORACLE
Real World SQL & PL/SQL
Advice from the Experts
Arup Nanda
Brendan Tierney
Heli Helskyaho
Martin Widlake
Alex Nuijten
Oracle Press

500+ Technical Experts Helping Peers Globally

ORACLE ACE Program

ORACLE ACE Director

ORACLE ACE

ORACLE ACE Associate

---

3 Membership Tiers

- Oracle ACE Director
- Oracle ACE
- Oracle ACE Associate

bit.ly/OracleACEProgram

Connect:

oracle-ace_ww@oracle.com

Facebook.com/oracleaces

@oracleace

Oracle Groundbreakers

Nominate yourself or someone you know: acenomination.oracle.com

What is Machine Learning?

* An important part of Artificial Intelligence (AI)
* Machine learning (ML) teaches *computers* to learn from *experience (algorithms)*
  * Learn from data and make predictions
  * Mathematics, statistics,…
* “field of study that gives computers the ability to learn without being explicitly programmed“

-- Arthur Samuel, 1959

* A systematic study of algorithms and systems that improve their *knowledge* or *performance* with *experience*

Miracle Finland Oy

Why ML? Why now?

* Improved technology
* The price for storage solutions
* …
* An environment that NEEDS ML and is finally able to really use it

* Artificial Intelligence (AI)
* BIG DATA

Miracle Finland Oy

What is Big Data?

* There is *no size* that makes a data to be ”Big Data”, it always depends on the capabilities
* The data is ”***Big***” when traditional processing with traditional tools is not possible due to the amount or the complexity of the data
  * You cannot open an attachement in email
  * You cannot edit a photo
  * etc.

Miracle Finland Oy

The three V's

* **Volume**, the size/scale of the data
* **Velocity**, the speed of change, analysis of streaming data
* **Variety**, different formats of data sources, different forms of data; structured, semi-structured, unstructured

The other V's

* **Veracity**, the uncertainty of the data, the data is worthless or harmful if it's not accurate
* **Viability**, validate that hypothesis before taking further action (and, in the process of determining the viability of a variable, we can expand our view to determine other variables)
* **Value**, the potential value
* **Variability,** refers to data whose meaning is constantly changing, in consistency of data; for example words and context
* **Visualization**, a way of presenting the data in a manner that's readable and accessible

Challenges in Big Data

* More and more data (volume)
* Different data models and formats (variety)
* Loading in progress while data exploration going on (velocity)
* Not all data is reliable (veracity)
* We do not know what we are looking for (value, viability, variability)
* Must support also non-technical users (journalists, investors, politicians,...) (visualization)
* All must be done *efficiently and fast and as much as possibly by machines*

When to use ML?

* You have **data**!
  * ML cannot be performed without data
  * part of the data for finding the model, part to prove it (not all for finding the model!)
* Rules and equations are
  * Complex (image recognition)
  * Constantly changing (fraud detection)
* The nature of the data changes and the program must adapt (today's spam is tomorrow's ham) (predicting shopping trends)

Real life use cases for ML

* Spam filters
* Log filters (and alarms)
* Data analytics
* Image recognition
* Speech recognition
* Medical diagnosis
* Robotics
* …

Approximation! A sophisticated guess!

* ML always gives an approximated answer
* Some are better than others, some are useful

* search for patterns and trends
* Prediction accuracy: the higher the number the better it will work on new data
* several models, choose the best, but still: all approximations! There is no correct answer…

Miracle Finland Oy

What do I find the most difficult for a beginner?

* The terminology!
  * So many different terms
  * The same term meaning different things, two (or more) terms for the same thing (sometimes a completely different word, sometimes just a short of the original word)
  * The relationships the terms have

Terms used 1/5

* A Task
  * The problem to be solved with ML
* An Algorithm
  * the “experience” for the computer to learn with, solves the learning problem
  * Produces the Models

Miracle Finland Oy

Terms used 2/5

* A Model
  * The output of ML
  * The Task is Addressed by Models

Terms used 3/5

* Different Models:
  * Predictive model
    * the model output involves the target variable
    * ” forecast what might happen in the future”
  * Descriptive model
    * the model output does not involve the target variable
    * ”what happened”
  * Prescriptive model
    * recommending one or more courses of action and showing the likely outcome of each decision
    * A predictive model + actionable data and a feedback system to track the outcome

Terms used 4/5

* Different models based on the algorithm type:
  * Classification Models
  * Concept learning Models
  * Tree Models
  * Rule Models
  * Linear Models
  * Distance-based Models
  * Probabilistic Models

Terms used 5/5

* Features/Dimensions
  * an individual *measurable property* or *characteristic of a phenomenon* being observed (Bishop, Christopher (2006), Pattern recognition and machine learning)
  * *Deriving features* (feature engineering, feature extraction) is one of the most important parts of machine learning. It turns data into information that a machine learning algorithm can use.
* Methods/Techniques
  * Unsupervised learning
  * Supervised learning

MIRACLE Miracle Finland Oy

The Task

* It is very important to define the Task well
* Machine learning is not only a computational subject, the practical side is very important

It’s all about Algorithms

* Humans learn with *experience*, machines learn with *algorithms*
* It is not easy to find the right Algorithm for the Task
  * usually try with several algorithms to find the best one
  * selecting an algorithm is a process of trial and error

Which algorithm?

* The selection of an algorithm depends on for instance
  * the size and type of data
  * the insights you want to get from the data
  * how those insights will be used
* It's a trade-off between many things
  * Predictive accuracy on new data
  * Speed of training
  * Memory usage
  * Transparency (black box vs "clear-box", how decisions are made)
  * Interpretability (the ability of a human to understand the model)
  * …

Models 1/2

* Geometric models
  * Support vector machines, SVM
  * Notion of distance: Euclidean distance, nearest-neighbour classifier, Manhattan distance
* Probabilistic models
  * Bayesian classifier
* Logical models
  * Decision trees

Models 2/2

* Grouping models, number of groups determined at the training time
  * Tree based models
* Grading models, ”infinite” resolution
  * Geometric classifiers
* …

Features

* A Model is only as good as its Features…
* Interaction between features

* The unnecessary detail can be removed by discretisation (11,1kg vs 10kg)

Miracle Finland Oy

ML in short

* Use the right *Features*
  * with right Algorithms
    * to build the right *Models*
      * that achieve the right *Tasks*

Two types of Methods

* Unsupervised learning
  * finds hidden patterns or intrinsic structures in input data
* Supervised learning
  * trains a model on known input and output data to predict future outputs

Miracle Finland Oy

Unsupervised Learning

* Learning from unlabeled input data by finding hidden patterns or intrinsic structures in that data
* Machine learning algorithms find natural patterns in data to make better decisions and predictions possible
* used typically when you
  * don't have a specific goal
  * are not sure what information the data contains
  * want to reduce the features of your data as a preprocessing for supervised learning

Miracle Finland Oy

Data for Unsupervised Learning

	A
46529	2007,1,16,2,1712,1715,1810,1815,WN,990,N252,58,60,45,-5,-3,SJC,BUR,296,3,10,0,,0,0,0,0,0,0
46530	2007,1,16,2,1228,1230,1327,1330,WN,1191,N374SW,59,60,46,-3,-2,SJC,BUR,296,2,11,0,,0,0,0,0,0,0
46531	2007,1,16,2,907,905,1003,1005,WN,1445,N409,56,60,46,-2,2,SJC,BUR,296,1,9,0,,0,0,0,0,0,0
46532	2007,1,16,2,1944,1940,2040,2040,WN,1449,N311,56,60,46,0,4,SJC,BUR,296,3,7,0,,0,0,0,0,0,0
46533	2007,1,16,2,650,650,749,750,WN,1650,N364,59,60,46,-1,0,SJC,BUR,296,2,11,0,,0,0,0,0,0,0
46534	2007,1,16,2,2052,2050,2151,2150,WN,2206,N356,59,60,49,1,2,SJC,BUR,296,2,8,0,,0,0,0,0,0,0
46535	2007,1,16,2,2053,2055,2204,2215,WN,889,N234,71,80,59,-11,-2,SJC,LAS,386,4,8,0,,0,0,0,0,0,0
46536	2007,1,16,2,926,925,1047,1045,WN,1088,N340,81,80,67,2,1,SJC,LAS,386,4,10,0,,0,0,0,0,0,0
46537	2007,1,16,2,1748,1750,1902,1910,WN,1113,N423,74,80,63,-8,-2,SJC,LAS,386,2,9,0,,0,0,0,0,0,0
46538	2007,1,16,2,2127,2130,2241,2250,WN,1232,N326,74,80,62,-9,-3,SJC,LAS,386,3,9,0,,0,0,0,0,0,0
46539	2007,1,16,2,700,700,816,820,WN,1325,N725,76,80,61,-4,0,SJC,LAS,386,3,12,0,,0,0,0,0,0,0
46540	2007,1,16,2,1344,1345,1502,1505,WN,2331,N241,78,80,65,-3,-1,SJC,LAS,386,3,10,0,,0,0,0,0,0,0
46541	2007,1,16,2,1552,1555,1709,1715,WN,2583,N236,77,80,64,-6,-3,SJC,LAS,386,4,9,0,,0,0,0,0,0,0
46542	2007,1,16,2,647,635,753,745,WN,123,N659SW,66,70,51,8,12,SJC,LAX,308,7,8,0,,0,0,0,0,0,0
46543	2007,1,16,2,1833,1835,1936,1945,WN,196,N365,63,70,49,-9,-2,SJC,LAX,308,4,10,0,,0,0,0,0,0,0
46544	2007,1,16,2,1420,1325,1531,1435,WN,197,N306SW,71,70,52,56,55,SJC,LAX,308,4,15,0,,0,0,0,1,0,55
46545	2007,1,16,2,1652,1650,1800,1800,WN,756,N631SW,68,70,53,0,2,SJC,LAX,308,7,8,0,,0,0,0,0,0,0
46546	2007,1,16,2,755,755,902,905,WN,1247,N642WN,67,70,52,-3,0,SJC,LAX,308,5,10,0,,0,0,0,0,0,0
46547	2007,1,16,2,1619,1620,1727,1730,WN,1577,N628SW,68,70,52,-3,-1,SJC,LAX,308,5,11,0,,0,0,0,0,0,0
46548	2007,1,16,2,1527,1525,1635,1635,WN,1581,N365,68,70,50,0,2,SJC,LAX,308,5,13,0,,0,0,0,0,0,0
46549	2007,1,16,2,2116,2120,2228,2230,WN,1635,N317SW,72,70,51,-2,-4,SJC,LAX,308,5,16,0,,0,0,0,0,0,0
46550	2007,1,16,2,1429,1430,1535,1540,WN,1664,N619SW,66,70,49,-5,-1,SJC,LAX,308,5,12,0,,0,0,0,0,0,0
46551	2007,1,16,2,1255,1255,1359,1405,WN,1843,N225,64,70,51,-6,0,SJC,LAX,308,3,10,0,,0,0,0,0,0,0
46552	2007,1,16,2,909,910,1040,1025,WN,2087,N684,91,75,50,15,-1,SJC,LAX,308,12,29,0,,0,0,0,15,0,0
46553	2007,1,16,2,1008,955,1116,1105,WN,2164,N601WN,68,70,51,11,13,SJC,LAX,308,6,11,0,,0,0,0,0,0,0
46554	2007,1,16,2,1101,1105,1211,1215,WN,2607,N625SW,70,70,55,-4,-4,SJC,LAX,308,5,10,0,,0,0,0,0,0,0

Miracle Finland Oy

Clustering

* *Clustering* is the most common method for unsupervised learning and used for *exploratory data analysis to find hidden patterns or groupings in data.*
* *Clustering algorithms*
  * *Hard clustering*
    * each data point belongs to *only one* cluster
  * *Soft clustering*
    * each data point can belong to *more than one* cluster

Miracle Finland Oy

Hard clustering algorithms

* each data point belongs to *only one* cluster

Some Hard Clustering Algorithms 1/2

* K-Means (Lloyd's algorithm)
  * Partitions data into k number of mutually exclusive clusters (centroids)
  * Assign each observation to the closest cluster
  * Move the centroids to the true mean of its observations
  * When to use:
    * When the number of clusters is known
    * Fast clustering of large data sets
* K-Medoids
  * Similar to k-means, but with the requirement that the cluster centers coincide with points in the data (chooses datapoints as centers, medoids).
  * Can be more robust to noise and outliers than K-Means
  * When to use:
    * When the number of clusters is known
    * Fast clustering of categorical data

MIRACLE Miracle Finland Oy

Some Hard Clustering, Algorithms 2/2

* Hierarchical Clustering
  * Divisive method, assign all observation to one cluster and the partition that cluster
  * Agglomerative method, each observation to its own cluster and merge those clusters
  * When to use:
    * When you don’t know in advance how many clusters
    * You want visualization to guide your selection

Miracle Finland Oy

Soft clustering algorithms

* each data point can belong to *more than one* cluster

Some Soft clustering algorithms

* Fuzzy C-Means (FCM)
  * Similar to k-means, but data points may belong to more than one cluster.
  * When to use:
    * The number of clusters is known
    * When clusters overlap
    * Typically for pattern recognition
* Gaussian Mixture Model
  * Partition-based clustering where data points come from different multivariate normal distributions with certain probabilities. (example: Prices for a house in different area)
  * When to use:
    * Data point might belong to more than one cluster
    * Clusters have different sizes and correlation structures within them

Miracle Finland Oy

Supervised Learning

* Learning from known, labelled data
* Training a model on known input and output data to predict future outputs (remember that uncertainty is always involved)

Miracle Finland Oy

Data for Supervised Learning

1	Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,C
2	2007,1,1,1,1232,1225,1341,1340,WN,2891,N351,69,75,54,1,7,SMF,ONT,389,4,11,0,,0,0,0,0,0,0
3	2007,1,1,1,1918,1905,2043,2035,WN,462,N370,85,90,74,8,13,SMF,PDX,479,5,6,0,,0,0,0,0,0,0
4	2007,1,1,1,2206,2130,2334,2300,WN,1229,N685,88,90,73,34,36,SMF,PDX,479,6,9,0,,0,3,0,0,0,31
5	2007,1,1,1,1230,1200,1356,1330,WN,1355,N364,86,90,75,26,30,SMF,PDX,479,3,8,0,,0,23,0,0,0,3
6	2007,1,1,1,831,830,957,1000,WN,2278,N480,86,90,74,-3,1,SMF,PDX,479,3,9,0,,0,0,0,0,0,0
7	2007,1,1,1,1430,1420,1553,1550,WN,2386,N611SW,83,90,74,3,10,SMF,PDX,479,2,7,0,,0,0,0,0,0,0
8	2007,1,1,1,1936,1840,2217,2130,WN,409,N482,101,110,89,47,56,SMF,PHX,647,5,7,0,,0,46,0,0,0,1
9	2007,1,1,1,944,935,1223,1225,WN,1131,N749SW,99,110,86,-2,9,SMF,PHX,647,4,9,0,,0,0,0,0,0,0
10	2007,1,1,1,1537,1450,1819,1735,WN,1212,N451,102,105,90,44,47,SMF,PHX,647,5,7,0,,0,20,0,0,0,24
11	2007,1,1,1,1318,1315,1603,1610,WN,2456,N630WN,105,115,92,-7,3,SMF,PHX,647,5,8,0,,0,0,0,0,0,0
12	2007,1,1,1,836,835,1119,1130,WN,2575,N493,103,115,88,-11,1,SMF,PHX,647,7,8,0,,0,0,0,0,0,0
13	2007,1,1,1,2047,1955,2332,2240,WN,2608,N733SW,105,105,89,52,52,SMF,PHX,647,7,9,0,,0,49,0,0,0,3
14	2007,1,1,1,2128,2035,2245,2200,WN,139,N348,77,85,66,45,53,SMF,SAN,480,3,8,0,,0,0,0,3,0,42
15	2007,1,1,1,935,940,1048,1105,WN,747,N358,73,85,63,-17,-5,SMF,SAN,480,2,8,0,,0,0,0,0,0,0
16	2007,1,1,1,1251,1245,1405,1410,WN,933,N413,74,85,65,-5,6,SMF,SAN,480,2,7,0,,0,0,0,0,0,0
17	2007,1,1,1,1729,1645,1843,1810,WN,1054,N416,74,85,64,33,44,SMF,SAN,480,3,7,0,,0,3,0,0,0,30
18	2007,1,1,1,825,825,941,950,WN,1106,N383SW,76,85,63,-9,0,SMF,SAN,480,3,10,0,,0,0,0,0,0,0
19	2007,1,1,1,1042,1040,1158,1205,WN,1554,N316SW,76,85,66,-7,2,SMF,SAN,480,2,8,0,,0,0,0,0,0,0
20	2007,1,1,1,1726,1725,1839,1850,WN,1604,N691WN,73,85,63,-11,1,SMF,SAN,480,3,7,0,,0,0,0,0,0,0
21	2007,1,1,1,1849,1820,2016,1940,WN,1975,N308SW,87,80,69,36,29,SMF,SAN,480,3,15,0,,0,20,0,7,0,9
22	2007,1,1,1,2219,2105,2332,2225,WN,2083,N205,73,80,62,67,74,SMF,SAN,480,3,8,0,,0,0,0,0,0,67
23	2007,1,1,1,2012,1940,2131,2105,WN,2577,N603SW,79,85,66,26,32,SMF,SAN,480,3,10,0,,0,9,0,0,0,17

A process of supervised learning 1/2

1. Train
   1. Load data
   2. Pre-process data
   3. Learn using a method and an algorithm
   4. Create a model
   * iterate until you find the best model

MIRACLE Miracle Finland Oy

A process of supervised learning 2/2

2. Predict (use the model with new data)

1. New data
2. Pre-process data
3. Use the model
4. Get predictions
5. Integrate the models into applications

MIRACLE
Miracle Finland Oy

Supervised Learning, methods/techniques

* Predictive models
  * Classification
  * Regression

Supervised Learning, Classification

* Classification models are trained to *classify* data into *categories*.
* They predict discrete responses
  * an email is genuine or spam
  * a tumor is small, medium size, or large
  * a tumor is cancerous or benign
  * a person is creditworthy or not
* For example applications like medical imaging, speech recognition, and credit scoring

Miracle Finland Oy

Supervised Learning, Classification

* Can the data be tagged or categorized? Can it be separated into specific groups or classes?
  * Classification might be the right answer

Classification, Some Algorithms

* k Nearest Neighbor (kNN)
  * kNN categorizes objects based on the classes of their nearest neighbors all ready categorized
  * kNN predictions assume that objects near each other are similar
  * When to use:
    * need a simple algorithm to establish benchmark learning rules
    * memory usage of the trained model is a lesser concern (can be very memory consuming)
    * prediction speed of the trained model is a lesser concern (can be slow if the amount of data is large or several dimensions are used)

Classification, Some Algorithms

- Naïve Bayes
  - assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature when the class is defined
  - classifies new data based on the highest probability of its belonging to a particular class (a fruit is red -> an apple, a fruit is round -> an apple, together a stronger probability to be an apple)
  - When to use:
    - For a dataset containing many parameters (dimensionality of the inputs is high)
    - Simple to implement, easy to interpret

Classification, Some Algorithms

- Discriminant Analysis
  - The classes are known a prio, an observation is classified to into one class based on the measured characteristics.
    - Example, bank notes:
    - two populations of bank notes, genuine, and counterfeit
    - Six measures: length, right-hand width, left-hand width, top margin, bottom margin, diagonal across the printed area
    - Take a bank note of unknown origin and determine using these six measurements whether or not it is real or counterfeit.
  - When to use:
    - need a simple model that is easy to interpret
    - memory usage during training is a concern
    - need a model that is fast to predict

MiRACLE Miracle Finland Oy

Classification, Some Algorithms

* Neural Network
  * Imitates how biological nervous systems, the brain, process information
  * A large number of highly interconnected processing elements (neurones) work together to solve specific problems
  * When to use:
    * For modeling highly nonlinear systems
    * When data is available incrementally and you wish to constantly update the model
    * Unexpected changes in your input data may occur
    * Model interpretability is not a key concern

MIRACLE Miracle Finland Oy

Classification, Some Algorithms

* Decision Trees, Bagged and Boosted Decision Trees
  * A tree consists of branching conditions, predict responses to data by following the decisions in the tree from the root down to a leaf node
  * A bagged decision tree consists of several trees that are trained independently on data. Boosting involves reweighting of misclassified events and building a new tree with reweighted events.
  * When to use:
    * Need an algorithm that is easy to interpret and fast to fit
    * To minimize memory usage
    * High predictive accuracy is not a requirement
    * The time taken to train a model is less of a concern

MiRACLE
Miracle Finland Oy

Classification, Some Algorithms

- Logistic Regression
  - Predict the probability of a binary response belonging to one class or the other
    - For example how does hours spent studing affect the probability for a student to pass the exam (yes/no)
  - When to use:
    - When data can be clearly separated by a single, linear boundary
    - Logistic regression is commonly used as a starting point for binary classification problems
    - As a baseline for evaluating more complex classification methods

Supervised Learning, Regression

* To predict continuous responses
  * changes in temperature
  * fluctuations in electricity demand
* For example applications like forecasting stock prices, handwriting recognition, acoustic signal processing, failure prediction in hardware, and electricity load forecasting.

Regression, Some Algorithms

* Linear Regression
  * used to describe a continuous response variable as a linear function of one or more predictor variables
  * When to use:
    * need an algorithm that is easy to interpret and fast to fit, often the first model to be fitted to a new dataset
    * As a baseline for evaluating other, more complex, regression models

Miracle Finland Oy

Regression, Some Algorithms

- Nonlinear Regression
  - describe nonlinear relationships in experimental data

  - When to use:
    - When data has nonlinear trends and cannot be easily transformed into a linear space
    - For fitting custom models to data

Regression, Some Algorithms

* Generalized Linear Model (GLM)
  * A special case of nonlinear models that uses linear methods: it fits a linear combination of the inputs to a nonlinear function (the link function) of the outputs
  * When to use:
    * When the response variables have non-normal distributions

Miracle Finland Oy

Regression, Some Algorithms

* Gaussian Process Regression Model (GPR)
  * nonparametric models that are used for predicting the value of a continuous response variable
  * When to use:
    * For interpolating spatial data
    * As a surrogate model to facilitate optimization of complex designs such as automotive engines
    * Can be used for example forecasting of mortality rates

Regression, Some Algorithms

* Regression Tree
  * Decision trees for regression are similar to decision trees for classification, but they are modified to be able to predict continuous responses
  * When to use:
    * When predictors are categorical (discrete) or behave nonlinearly

Improving Models

* Why to improve
  * To increase the accuracy and predictive power of the model
  * To increase the ability to recognize data from noise
  * To increase the performance
  * To improve the Measures wanted
  * …

Improving Models

* Model improvement involves
  * Feature engineering
    * Feature selection
    * Feature transformation/extraction
  * Hyperparameter tuning

Feature selection

* Also called variable selection or attribute selection
    * Identifying the most relevant features that provide the best predictive model for the data
    * *Adding* variables to the model to improve the accuracy or *removing* variables that do not improve model performance

Feature selection techniques

* **Stepwise regression:**
  * adding or removing features sequentially until there is no improvement in prediction accuracy
* **Sequential feature selection:**
  * adding or removing predictor variables iteratively and evaluating the effect of each change on the performance of the model
* **Regularization:**
  * Using shrinkage estimators to remove redundant features by reducing their weights (coefficients) to zero
* **Neighborhood component analysis (NCA):**
  * Finding the weight each feature has in predicting the output, so that features with lower weights can be discarded

Feature transformation

* Feature transformation is a form of *dimensionality reduction*
* Used when
  * want to reduce the dimensions/features of your data as a preprocessing for supervised learning
  * As datasets get bigger, you frequently need to reduce the number of features, or dimensionality.

Feature transformation

* Techniques:
  * Principal component analysis (PCA)
  * Factor analysis
  * Non-negative matrix factorization

Principal component analysis (PCA)

* Converts a set of observations of possibly correlated variables into a smaller set of values of linearly uncorrelated variables called *principal components*
* The first principal component will capture the most variance, followed by the second principal component, and so on.

Miracle Finland Oy

Factor analysis

* identifies underlying correlations between variables in a dataset to provide a representation in terms of a smaller number of unobserved variables, factors

Miracle Finland Oy

Non-negative matrix factorization (NNMF)

* Also called non-negative matrix approximation
* used when model elements must represent *non-negative* quantities, such as physical quantities

Hyperparameter tuning

* Also called as Hyperparameter optimization
* Choosing an optimal set of hyperparameters for a learning algorithm
  * Hyperparameters are parameters whose values are set *prior* to the commencement of the learning process (the value of other parameters is derived via training)
    * Number of clusters in a clustering, number of leaves or depth of a tree,…
  * Hyperparameters control how a machine learning algorithm fits the model to the data.

MIRACLE
Miracle Finland Oy

Hyperparameter Tuning

* Tuning is an iterative process
  * Set parameters based on a best guess
  * Aim to find the best possible values to yield the best model
  * As you adjust hyperparameters and the performance of the model begins to improve, you see which settings are effective and which still require tuning
* Some examples of optimization algorithms:
  * Grid search
  * Bayesian optimization
  * Gradient-based optimization
  * Random Search
* A simple algorithm with well-tuned parameters is often better than an inadequately tuned complex algorithm, in many ways.

MIRACLE Miracle Finland Oy

How do I know when to tune?

* How does the model perform on the data?
* Which of the models is the best?
* Which of the learning algorithms gives the best model for the data?
* …
* To be able to answer questions like these we need to have **measuring**

What to measure?

* Number of positives, number of negatives, number of true positives, number of false positives, number of true negatives, number of false negatives
* Portion of positives, portion of negatives
* Class ratio
* Accuracy, Error rate
* ROC curve, coverage curve,
* …
* It all depends on how we define the performance for the answer to our question (experiment): *experimental objective*

How to measure?

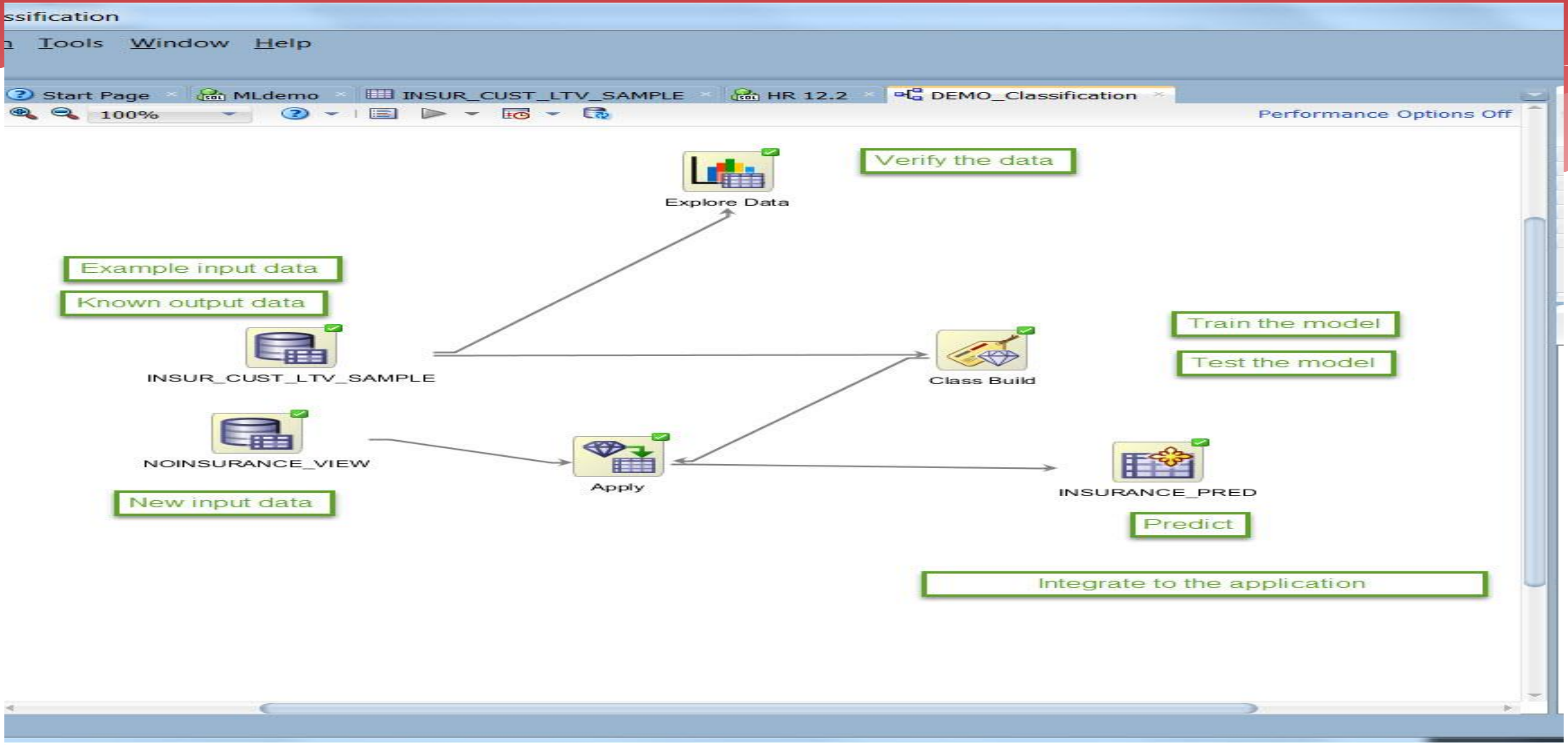
* And how to interpret?
* It all depends what we are measuring…
* Example: Testing the model accuracy
  * Tool: Cross validation

Cross validation

* Sometimes called Rotation Estimation
* Divide the data in n parts of equal size
* Use n-1 parts for training and 1 for testing
* Repeat n times so that each of the sets will be used for testing

Miracle Finland Oy

Oracle SQL Developer demo



Oracle SQL Developer, Data Miner

* Oracle SQL Developer is a free tool from Oracle
* Has an add-on called Data Miner
* *Advanced analytics* (Data Miner uses that) is a **licensed product** (in the EE database separately licensed, in the Cloud: Database Service either High Performace Package or Extreme Performance Package)
* Oracle Data Miner GUI Installation Instructions

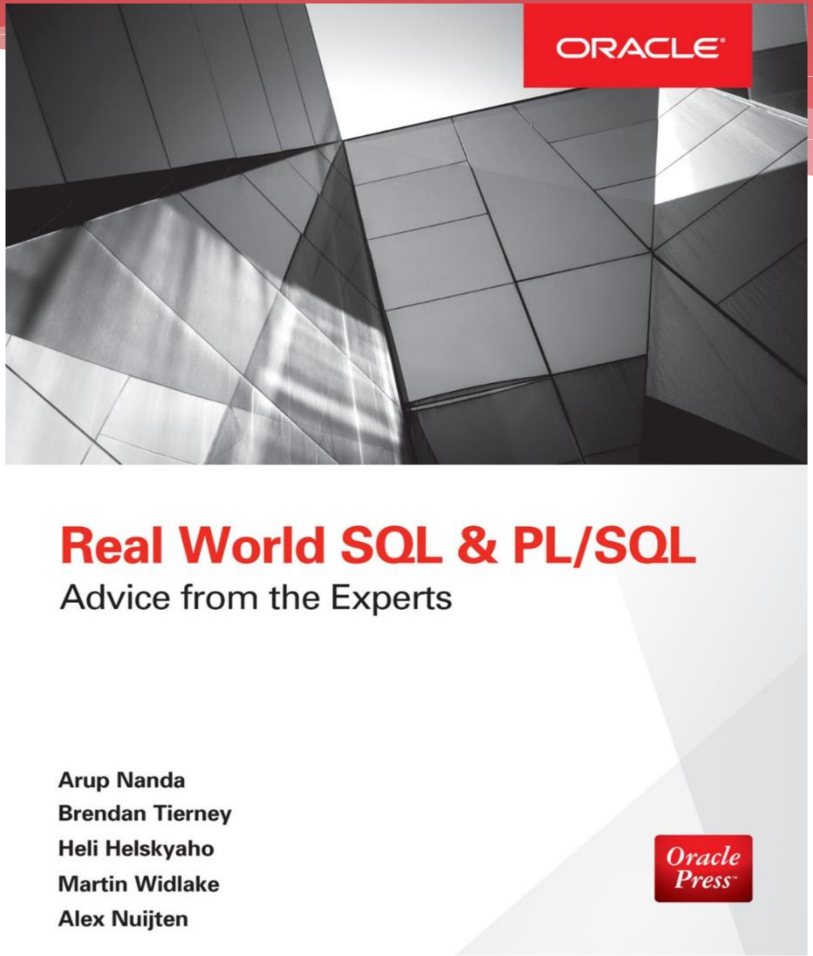
http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odmrinstallation-2080768.html

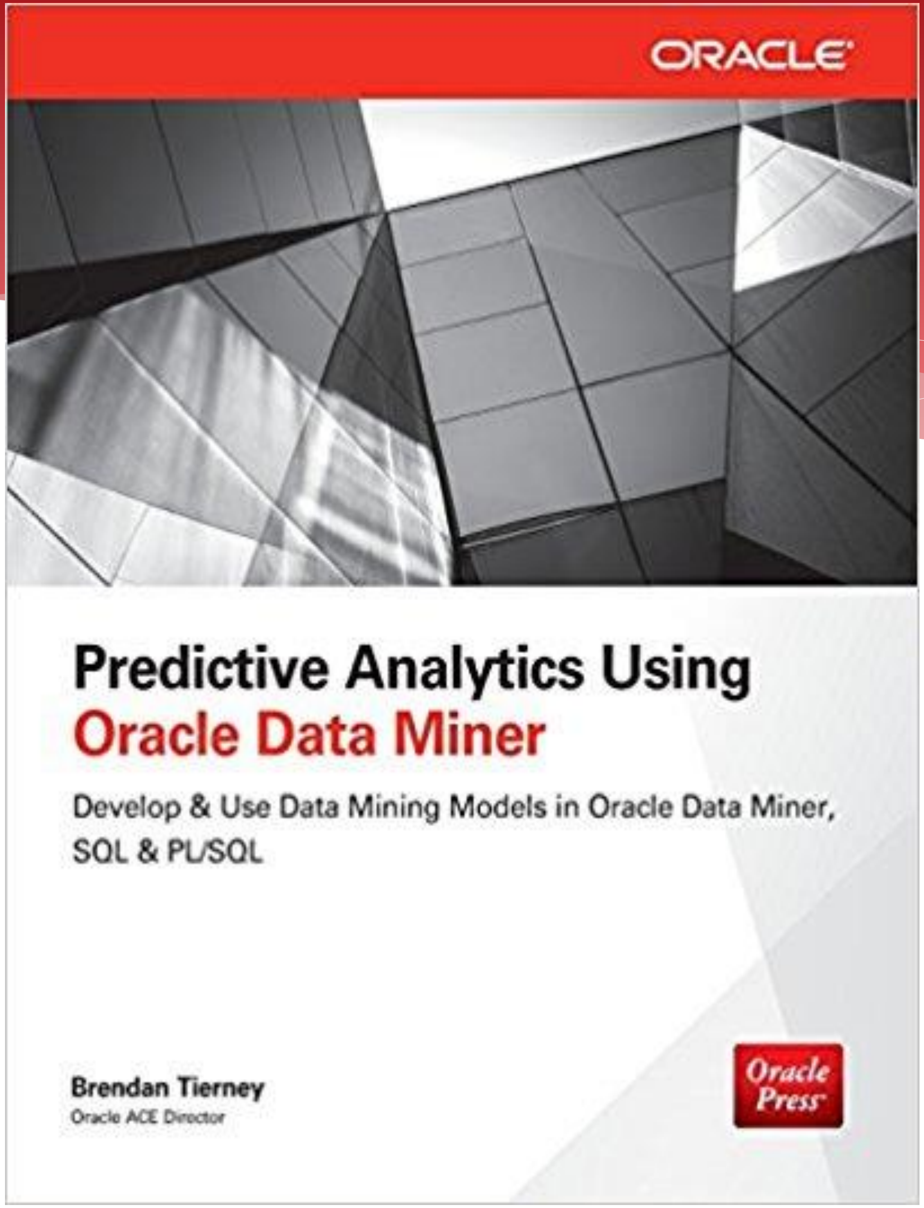
* Tutorial

http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/12c/BigDataDM/ODM12c-BDL4.html

MiRACLE Miracle Finland Oy

Chapter 10


ORACLE
Real World SQL & PL/SQL
Advice from the Experts
Arup Nanda
Brendan Tierney
Heli Helskyaho
Martin Widlake
Alex Nuijten
Oracle Press

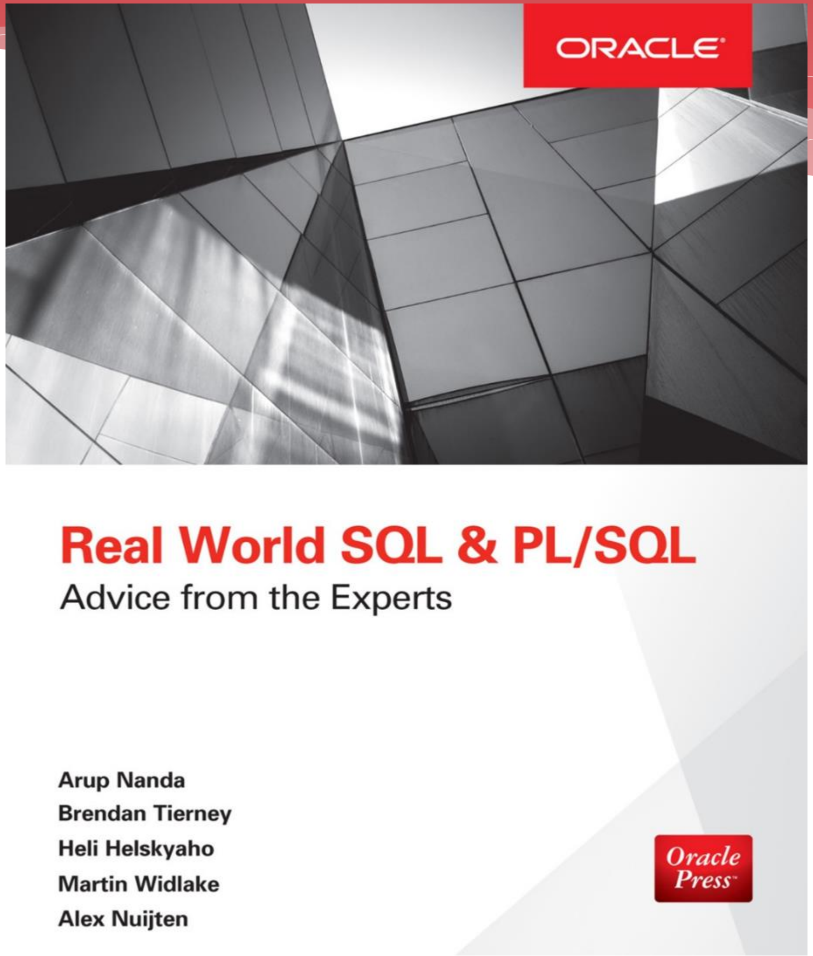


Miracle Finland Oy

Predictive Queries in Oracle 12c

* Predictive Queries enable you to build and score data quickly using the in-database data mining algorithms
* Predictive Queries can be
  * built using Oracle Data Miner
  * written using SQL

Chapter 12

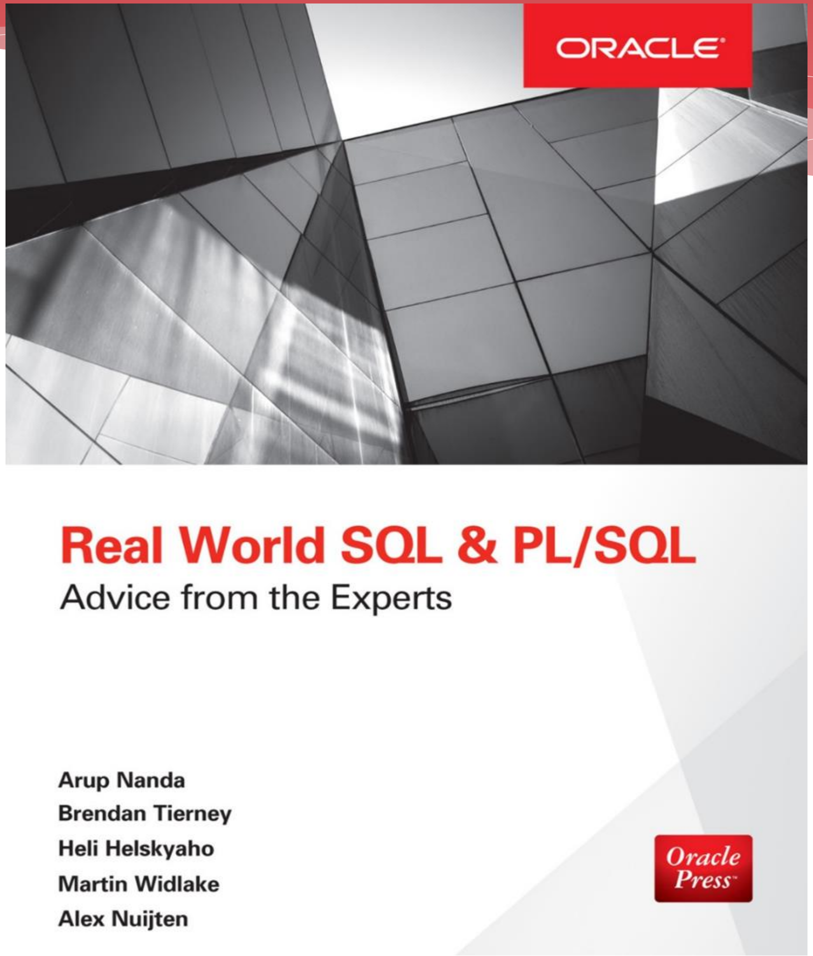


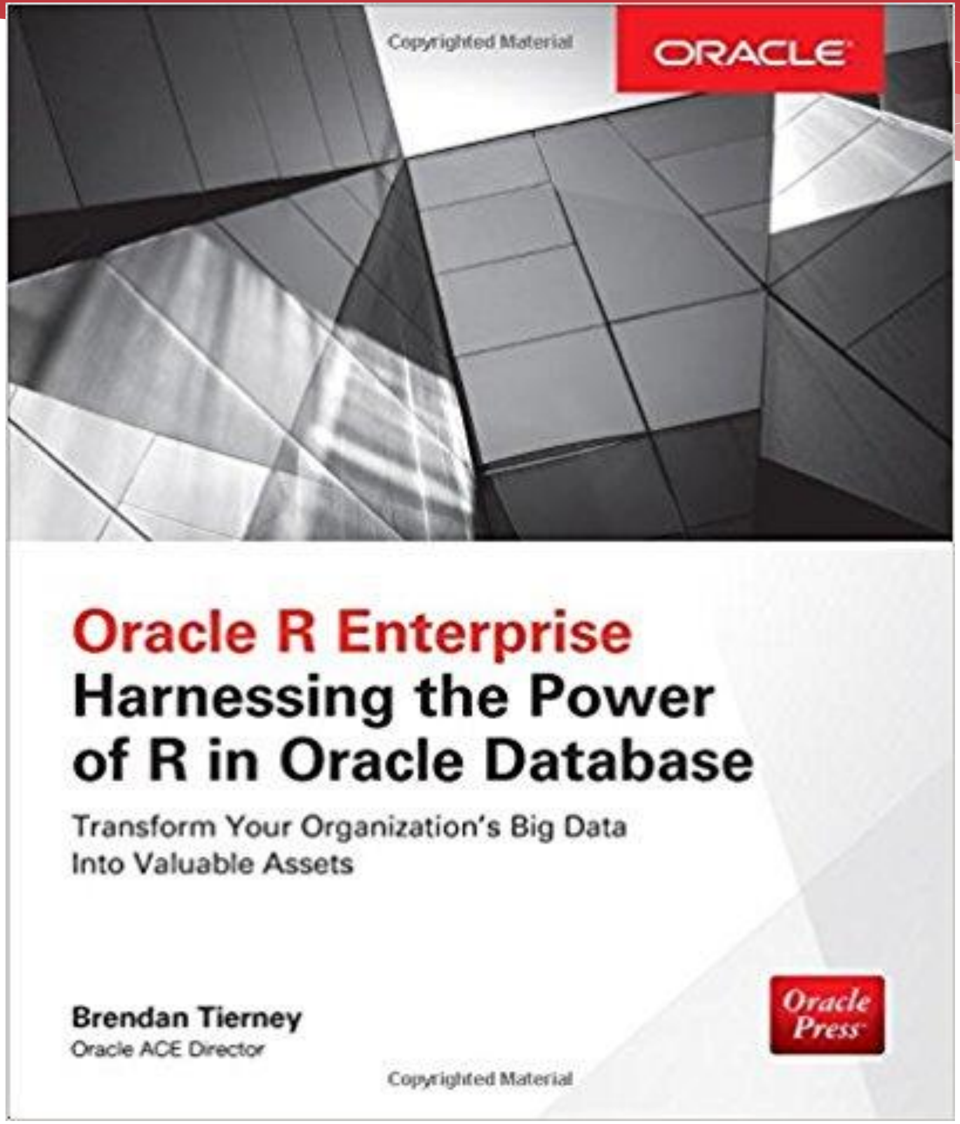
Miracle Finland Oy

Oracle R Enterprise

* a component of the Oracle Advanced Analytics Option (payable option)
* open source R statistical programming language in an Oracle database

Chapter 11


ORACLE
Real World SQL & PL/SQL
Advice from the Experts
Arup Nanda
Brendan Tierney
Heli Helskyaho
Martin Widlake
Alex Nuijten
Oracle Press



MIRACLE
Miracle Finland Oy

R demo

Python demo

And so many more languages to learn…

* Python
* C/C++
* Java
* JavaScript
* Julia, Scala, Ruby, Octave, MATLAB, SAS

* https://medium.com/towards-data-science/what-is-the-best-programming-language-for-machine-learning-a745c156d6b7

Miracle Finland Oy

What's next to learn?

* There is still so much more about ML…
* Reinforcement learning
  * the machine or software agent learns based on feedback from the environment
* Preference learning
  * inducing predictive preference models from empirical data
* Multi-task learning
  * multiple learning tasks are solved at the same time, while exploiting commonalities and differences across tasks
* Online machine learning
  * data becomes available in a sequential order and is used to update our best predictor for future data at each step

What's next to learn?

* Active learning
  * A learning algorithm is able to interactively query the user (or some other information source) to obtain the desired outputs at new data points
* Deep learning
  * Images and anything that is in ”several layers”
* Adaptive Intelligence
  * People and machines

The future and now!

* AI and machine learning is here and it's the future
* So many interesting areas to learn
* Pick your area and START LEARNING!

Conclusions

* The time for Machine Learning is now because we technically able to use it and because of Big Data

Conclusion

* Several V's related to Big Data…
  * Volume
  * Velocity
  * Variety
  * Veracity
  * Viability
  * Value
  * Variability
  * Visualization
  * …

Conclusion

* ML can be used ”everywhere”:
  * Spam filters
  * Log filters (and alarms)
  * Data analytics
  * Image recognition
  * Speech recognition
  * Medical diagnosis
  * Robotics
  * …

Conclusion

* Machine learning is all about approximation and educated guess
* Unsupervised Learning vs supervised Learning
  * Unsupervised Learning
    * Clustering: hard or soft
  * Supervised Learning
    * Train, Predict
* Predictive Models:
    * classification, regression

Conclusion

* Improving Models
  * Feature engineering
  * Hyperparameter tuning

* What to measure? How to interpret the measures?

* There is so much more to learn in ML…

MiRACLE
Miracle Finland Oy

THANK YOU!

QUESTIONS?

Email: heli@miracleoy.fi
Twitter: @HeliFromFinland
Blog: Helifromfinland.com

Miracle Finland Oy