

# Future World Enters Twilight Zone

## Machine Learning, AI, VR, Robotics & Quantum Computing Ahead



*(Welcome to the Fifth Dimension )*

*Rich Niemiec*  
*@richniemiec*



*rich.niemiec@viscosityna.com*



Special Thanks:

Kay Cavender, Charles Kim, Monica Lee, & Linda Hoover

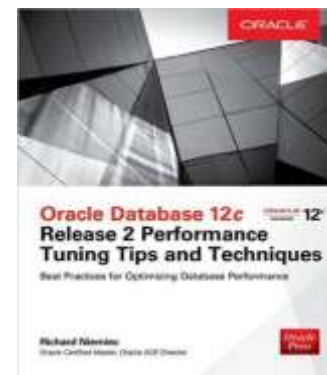
# Rich's Overview @richniemiec



ORACLE  
ACE Director



- Chief Innovation Officer, Viscosity North America: rich.niemiec@viscosityna.com
- Former CEO of TUSC
  - Inc. 500 Company (Fastest Growing 500 Private Companies)
  - 10 Offices in the United States (U.S.); Based in Chicago
  - Oracle Advantage Partner in Tech & Applications
- Former President Rolta TUSC & President Rolta EICT International
- Author (5 Oracle Best Sellers – #1 Oracle Tuning Book for over a Decade):
  - Oracle Performing Tips & Techniques (Covers Oracle7 & 8i)
  - Oracle9i Performance Tips & Techniques
  - Oracle Database 10g Performance Tips & Techniques
  - Oracle Database 11g Performance Tips & Techniques
  - Quick Start Guide to Oracle Query Tuning (2015)
  - Oracle Database 12cR2 Performance Tips & Techniques
- Former President of the International Oracle Users Group
- IOUG Top Speaker in 1991, 1994, 1997, 2001, 2006, 2007
- MOUG Current President & Top Speaker Twelve Times
- National Trio Achiever award - 2006
- Oracle Certified Master & Oracle Ace Director
- Chris Wooldridge Award – 1998, 2012
- Chicago Entrepreneur Hall of Fame - 1998
- MIT Machine Learning & AI
- Purdue Outstanding Electrical & Computer and Engineer Alumni – 2007
- Board Member – TEC, Entrigna, Ask DB Experts
- E&Y Entrepreneur of Year & National Hall of Fame - 2001



# Viscosity Pillars and Delivery Models

## Data

Database

Data Integration

Data Warehousing  
Analytics  
GoldenGate

Performance Tuning

## APEX & Apps

SAAS / PAAS  
CX ERP SCM HCM

EBS / JDE / PS

Middleware

Web Applications

Mobility

## Infrastructure

Oracle Cloud  
AWS  
Azure

Engineered Systems  
Exadata ODA ZFS  
@Customer

Virtualization  
VMware - KVM

RAC

Turnkey  
Projects

Assessments

Proof of  
Concepts

Training

Security

Managed  
Services

# Viscosity's Oracle ACES

## Oracle ACE Program

The Oracle ACE Program recognizes and rewards community members for their technical contributions in the Oracle community.



Charles Kim,  
CEO & Co-Founder

Twitter: @racdba



Rich Niemiec,  
Chief Innovation Officer

Twitter: @richniemiec



Craig Shallahamer  
Applied AI Scientist

Twitter: @orapub



Sean Scott,  
Consultant

Twitter: @oraclesean



Gary Gordhamer,  
Consultant

Twitter: @ggordham



Julio Ayapan,  
Consultant





# Quick FREE notes

Send email to (for slides):  
[hello@viscosityna.com](mailto:hello@viscosityna.com)

(richniemiec@gmail.com)

@richniemiec - twitter



# Agenda – Goal is Apply Technology & W.I.N.

- ❑ You've Just Entered the Twilight Zone
- ❑ Twilight Zone that **We See Now**
- ❑ Big Data and IOT - Data Is Coming Fast
- ❑ **ML & Oracle - Overview & ADB**
- ❑ Applications of ML Algorithms & AutoML
- ❑ Machine Learning Future, Robots & VR
- ❑ Future Tech from the Twilight Zone+



Quantum Entanglement

"Spooky action-at-a-distance"  
(A. Einstein)





The  
TWILIGHT  
ZONE

# Where is Everybody?





# Where is Everybody?



[Topics](#)

[Missions](#)

[Galleries](#)

[NASA TV](#)

[Follow NASA](#)

[Downloads](#)

[About](#)

[NASA Audiences](#)

Search

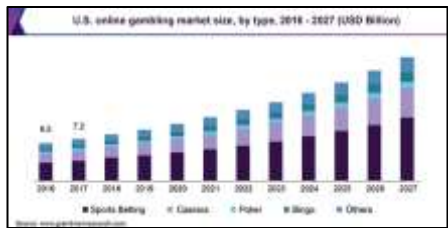
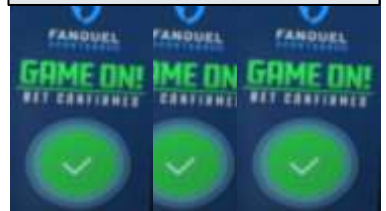
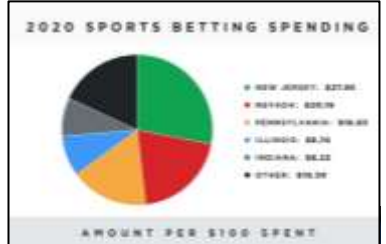
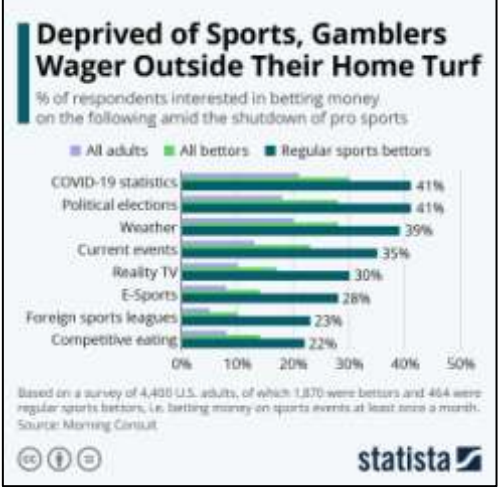


Social Isolation and Space





# The Fever



Sports betting ads: Industry weighs how much is too much?



© iStockphoto.com/PhotoAlto/PhotoAlto/Getty Images. © iStockphoto.com/PhotoAlto/PhotoAlto/Getty Images. © iStockphoto.com/PhotoAlto/PhotoAlto/Getty Images. © iStockphoto.com/PhotoAlto/PhotoAlto/Getty Images.

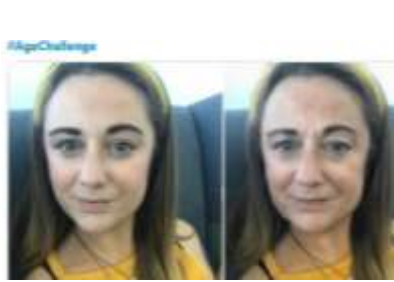
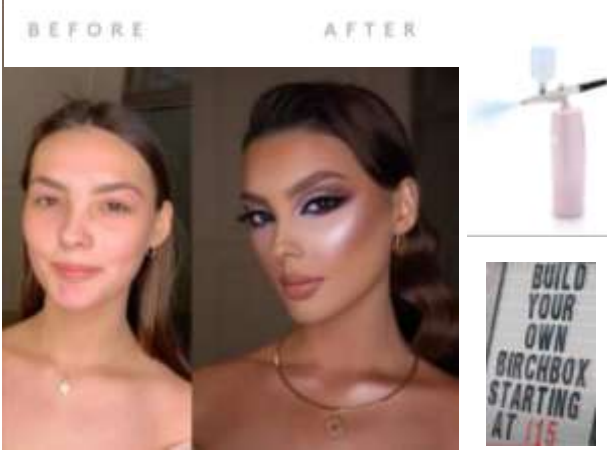
# Number 12 Looks Like You (2020)



TikTok Goes Viral By Exposing The Truth Behind Heavily-Edited Pictures  
TikTok guidelines said not to promote posts by 'ugly' and poor users



hide or disguise the presence of





# The Brain Center at Whipple's



Ameca



Pepper



Spot & Atlas



(Optimus)

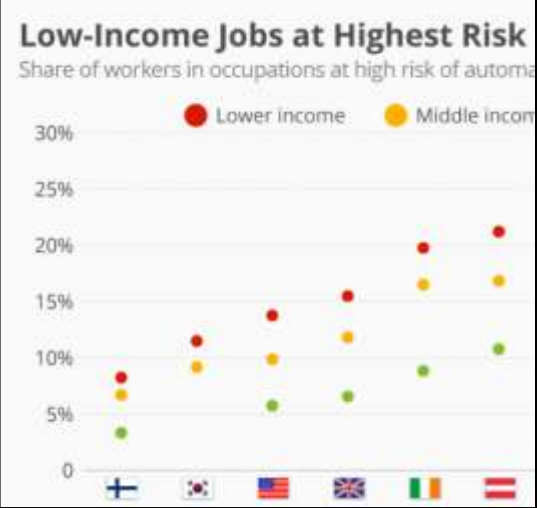
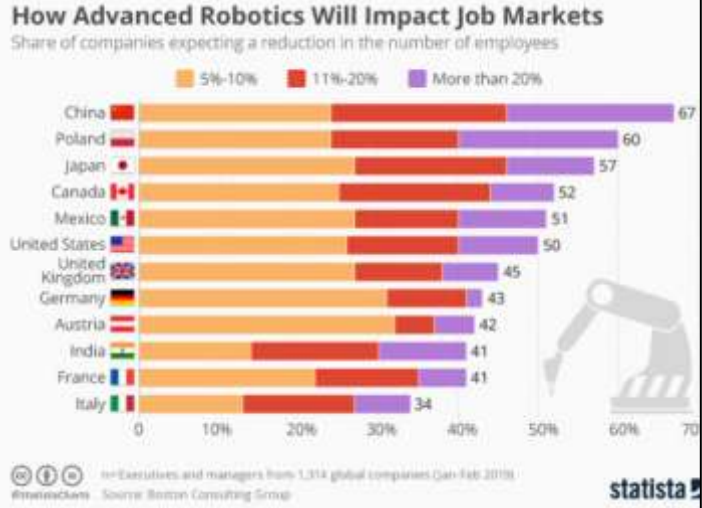
TESLA  
BOT



Sofia

"Okay, I will destroy humans."

# Robotics/Automation Impact to Jobs



The New York Times

*Uber's Self-Driving Trucks Hit the Highway, but Not Local Roads*





# Leverage – DB, GPS & Robotics! Impact to Jobs?





# The Obsolete Man



# Autonomous Database – Replacing the DBA?

Rich Niemiec @RichNiemiec · Oct 2  
I'll ask #Pepper if she can tune my #database in 11 AM session on #innovation #iot #robotics #cloud at #oow17 #ioug #viscosityna #oracleace



## Are DBAs Obsolete?

Posted on February 20, 2014

Before we go any further, let me briefly answer the question "No Way!" OK ... with that out of the way, let's get to the point.

Every so often, some industry pundit gets bored and says "Database administrators are obsolete" or "I hear this, it makes me shake my head sadly because it can be



**BUREAU OF LABOR STATISTICS**

Home Subjects Data Tools Publications Economic Releases

OOH HOME | OCCUPATION FINDER | OOH FAQ | OOH GLOSSARY | A-Z INDEX | OOH SITE

## OCCUPATIONAL OUTLOOK HANDBOOK

Occupational Outlook Handbook > Computer and Information Technology >

### Database Administrators

Summary What They Do Work Environment How to Become One Pay Job Outlook

#### Summary

Quick Facts: Database Administrators	
2017 Median Pay	\$87,020 per year \$41.84 per hour
Typical Entry-Level Education	Bachelor's degree
Work Experience in a Related Occupation	None
On-the-job Training	None
Number of Jobs, 2016	119,500
Job Outlook, 2016-26	11% (Faster than average)
Employment Change, 2016-26	13,700

**What Database Administrators Do**

Database administrators (DBAs) use specialized software to store and organize data, such as information and customer shipping records. They make sure that data are available to users.

**Good News: DBA +11% Increase  
(2016-2026)**

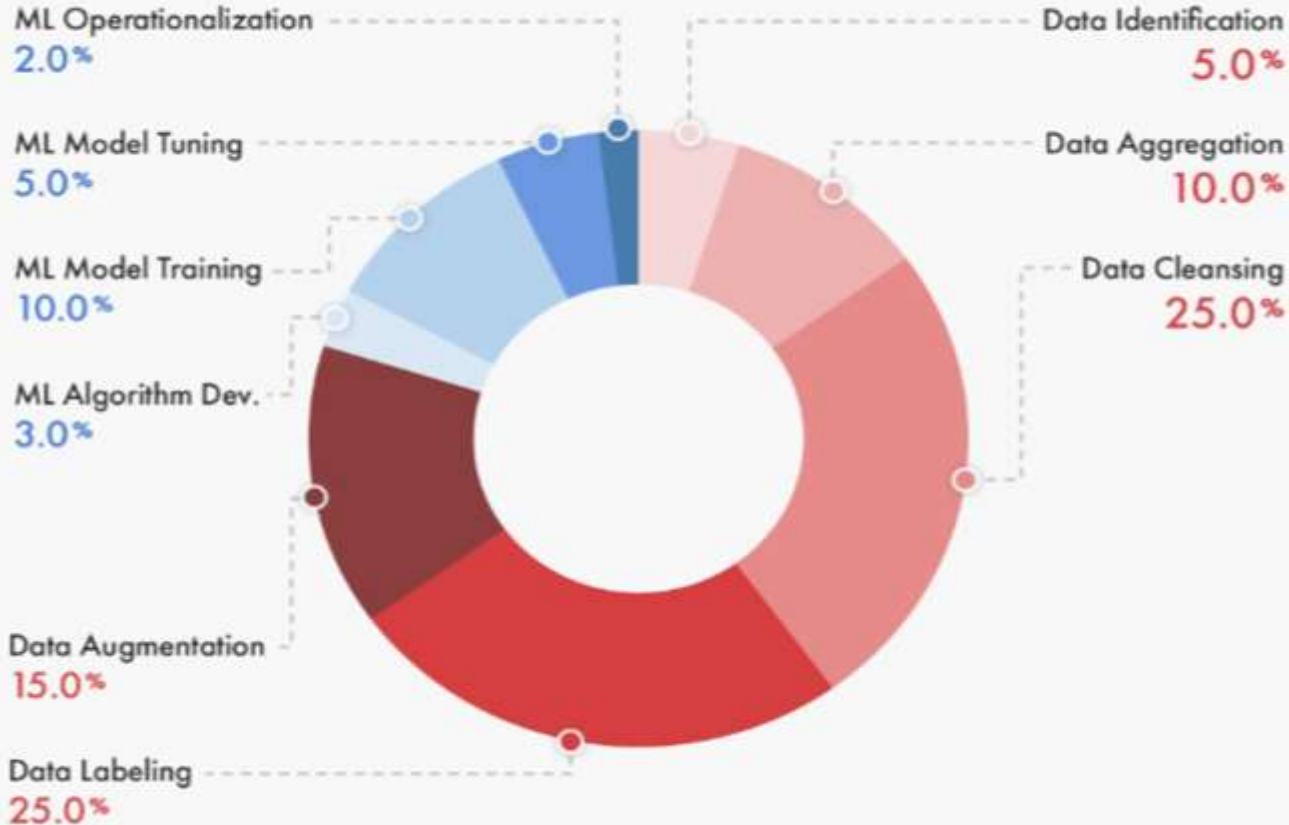
# Biju Thomas at ODTUG - *Emerging Jobs*

## #8 Data Engineer



- 33% annual growth
- Skills
  - Apache Spark
  - Hadoop
  - Python
  - ETL
  - AWS
- Industries
  - Computer Software
  - Information Technology
  - Financial Services
  - Healthcare & Hospitals
- Primary job responsibilities involve preparing data for analytical or operational uses.
- Works as part of an analytics team, providing data in a ready-to-use form to data scientists.
- Commonly deal with both structured and unstructured data sets

## Percentage of Time Allocated to Machine Learning Project Tasks

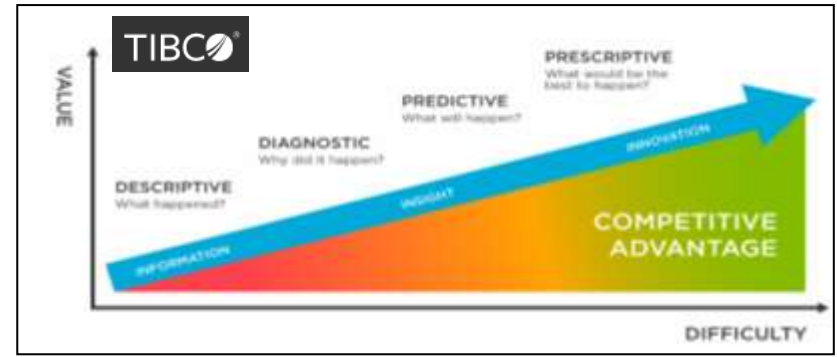
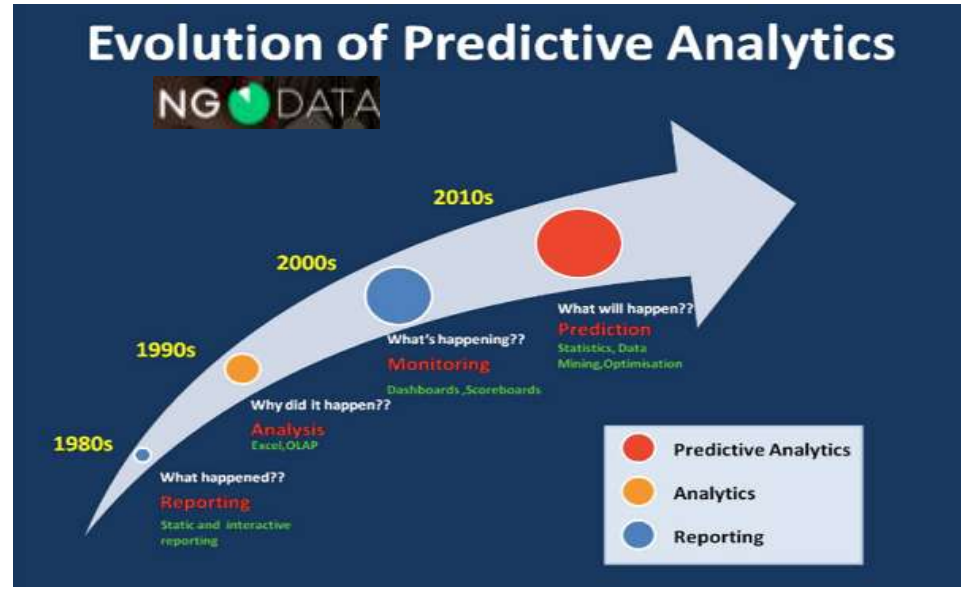


**The DBA is most important part of the Machine Learning Process**

Niti Sharma An Introduction to Data Labeling in Artificial Intelligence *Data Wrangling consumes over 80% of the time in AI projects.*



# What You Need; Nick of Time (*Predictive Analytics*)





# Characteristics of Big Data - The Five V's

## Volume

Big data comes in one size: large. Enterprises are awash with data, easily amassing terabytes and even petabytes of information.

TB, Records, Transactions, Tables, Files

## Velocity

Often time-sensitive, big data is streaming in to the cloud as it is streaming in to the cloud to maximize its value to the user. Batch, Near time, Real time

## Value

Business value of Big Data

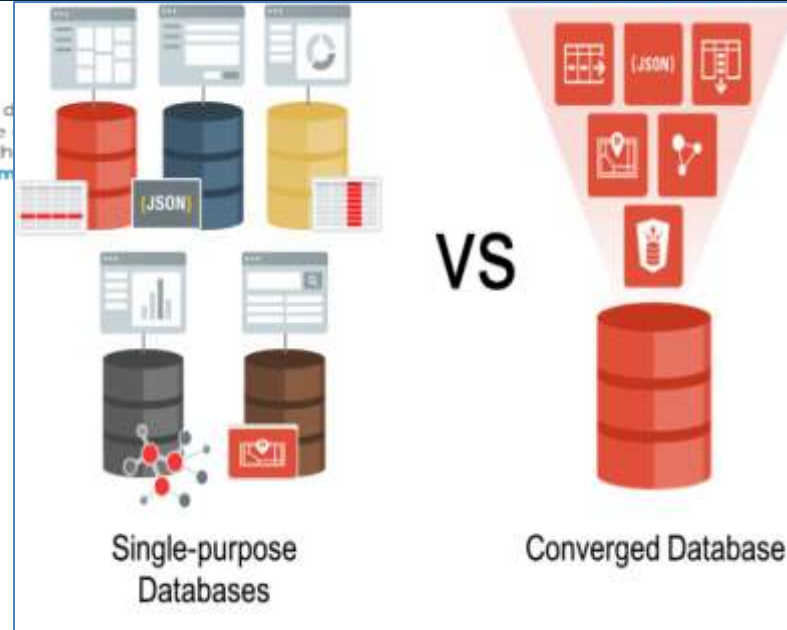
## Variety

Big data extends beyond structured data, including semi-structured and unstructured data of all varieties: text, audio, video, click streams, log files and more.

Structured, Unstructured, Semistructured



## Veracity



## Big Data Themes

- HW & SW technologies for large data volumes
- Focus on Web 2.0 technologies
- Database Scale-out
- Relational & Distributed Data Analytics
- Real Time Analytics

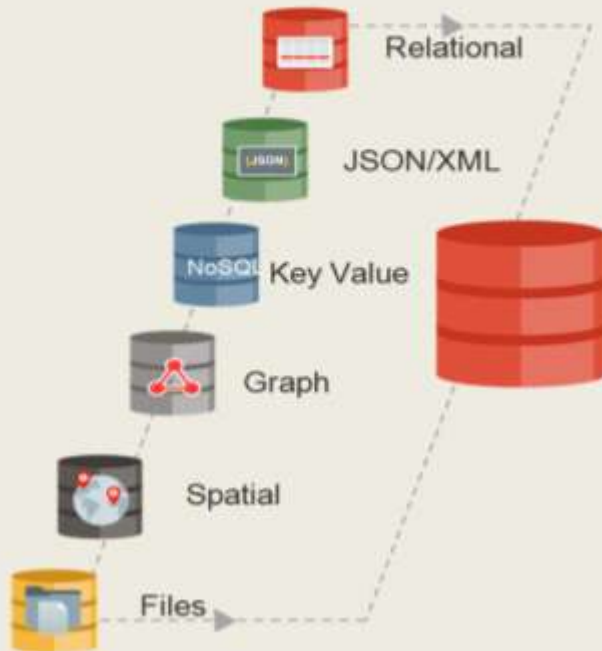
## Big Data Domains

- Digital Marketing Optimization
- Data Exploration & Discovery
- Fraud Detection & Prevention
- Social Network & Relationship Analysis
- Machine-generated Data Analytics

# Converged Database - Oracle Multi-Model Database\*

- Benefits of Oracle's Converged Database are broad

\*Slide from Oracle's,  
Nitin Vengurlekar

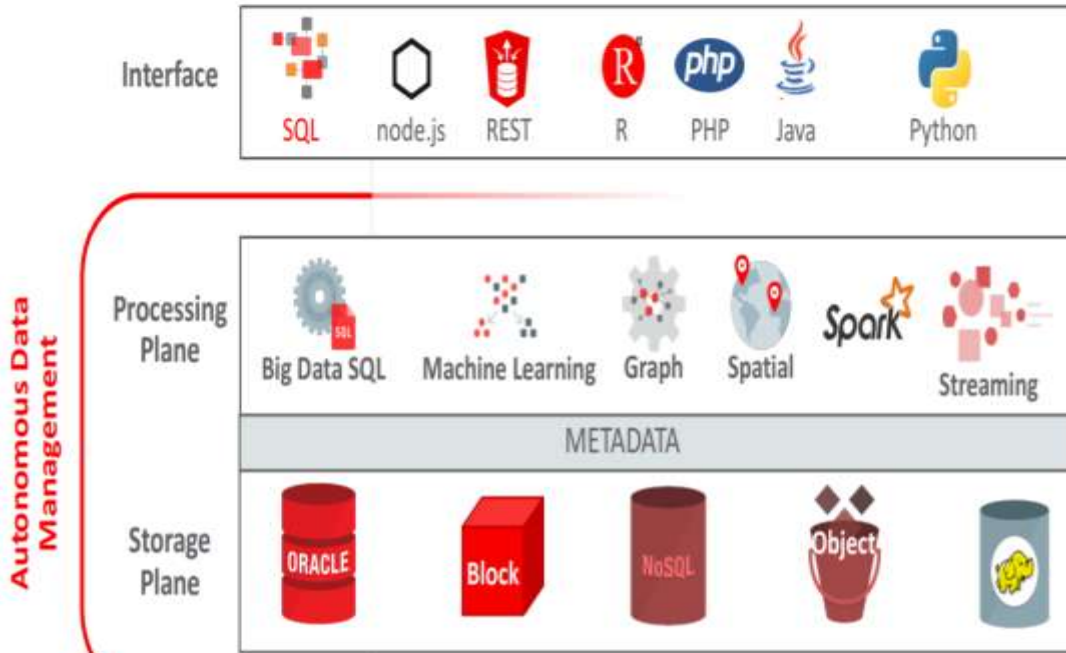


- Integrated development tools (Apex, SQL Dev, Spatial Studio) ✓
- 3<sup>rd</sup> party and Open Source development tools ✓
- Machine Learning ✓
- Node.js, Python, many others ✓
- In-memory database ✓
- Spatial, Graph support ✓
- NoSQL (JSON, key-value, wide column, XML) ✓
- Containers, microservices, virtualization (Docker, MT) ✓
- Integrated Security ✓
- Deployment choice (on-prem, cloud, hybrid) ✓
- Integrated High Availability and Disaster Recovery ✓



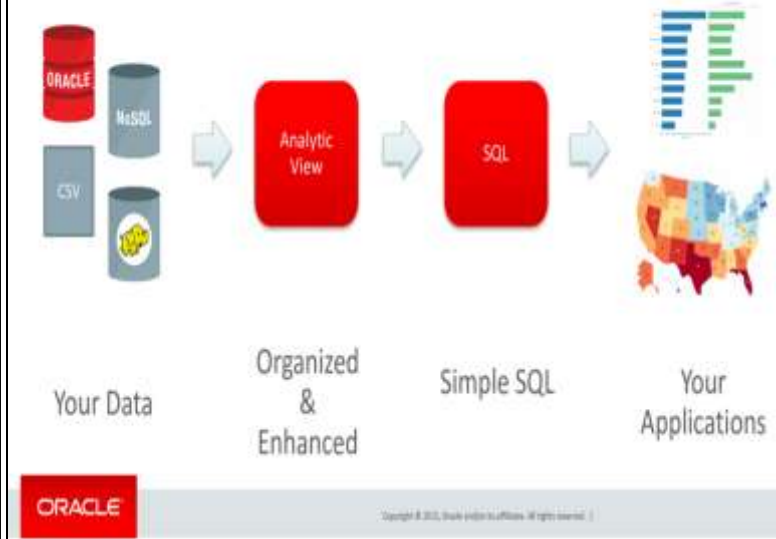
# Become a **DA (Data Administrator)**: Not just DBA

## Data Management as a Service



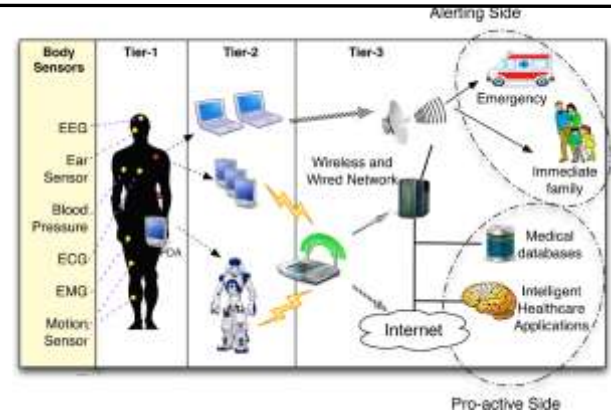
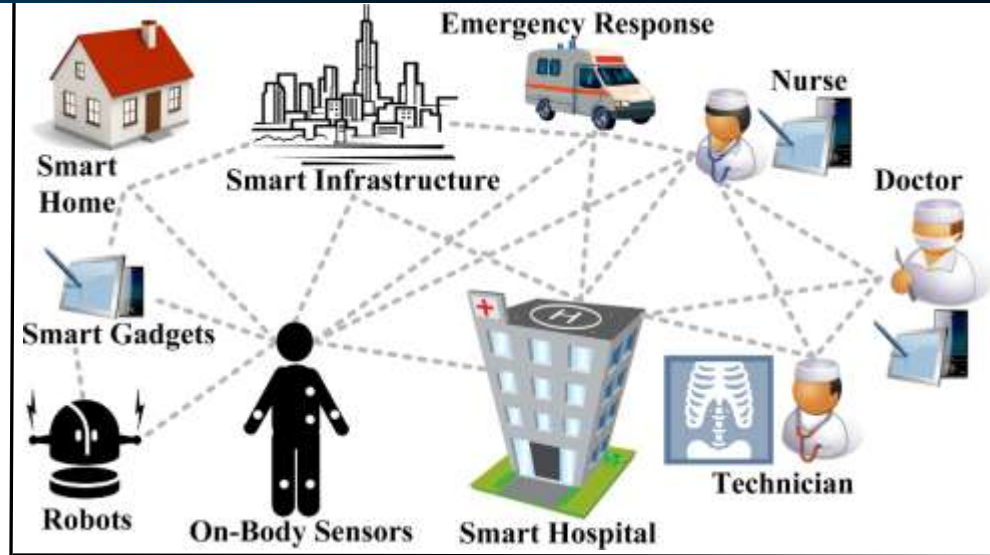
## Analytic Views

Easier Access To You Data

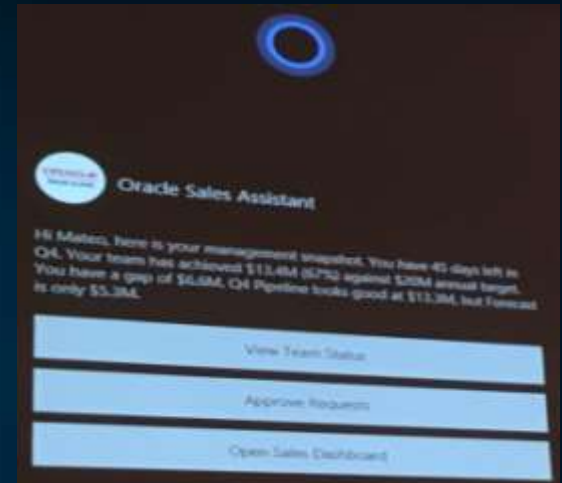


Data is the New Oil...

# The Hitchhiker *(Many Predictive Medical devices & IOT Data)*







# A Robot may not look like one!

*Oh Yeah... they never complain, always happy to do more, work anywhere, get smarter as time goes on, leverages AI & ML, works 24x7, doesn't ask for a raise, no union (yet).*



A large red banner is stretched across the facade of a modern glass skyscraper. The banner features the Oracle logo and the text 'Database Cloud' on the left, and 'The World's First Autonomous Database' on the right. The building's glass reflects the surrounding city skyline.

ORACLE®  
Database Cloud

The World's First Autonomous Database

# A Robot may not look one!

\*Robots that Manage a Database (ADW)!

\*Robots that secure a system and use ML & AI

# Autonomous DB : Future DBA & Robot DB

- Self-Managing (Driving)
- Self-Securing
- Self-Tuning
- Self-Recovering
- Self-Scaling Administration



*Fully automated patching, upgrades, backups, & availability architecture*

**Oracle Unveils World's First  
Autonomous Database Cloud**

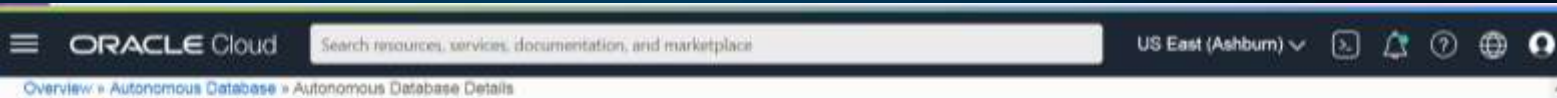
# Reality of the Autonomous Database & ML

## Will my job change?

- Absolutely...sure hope so!
- It has many times in the past...
- Closer to the business & Innovation
- **Data Critical & ML & AI** driven by it.
- Data Admin/Architect/ML instead of DBA
- ML/AI/Data Science Developer instead of just Oracle
- Security Expert instead of Security on the DB
- Watching over costs more
  - Cloud Hidden Costs: Cloud, Hybrid, or On-Site Decisions
  - Decide which databases should be Autonomous

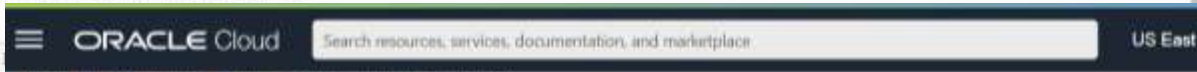


# ATP – Provisioning a Database (2 minutes)



PROVISIONING

DB-20220510121752



Overview » Autonomous Database » Autonomous Database Details



AVAILABLE

DB-20220510121752

- Database Actions
- DB Connection
- Performance Hub
- Service Console
- More Actions

Autonomous Database Information Tools Tags

## General Information

**Database Name:** DB20220510121752  
**Workload Type:** Transaction Processing  
**Compartment:** richnieniec (root)  
**OCID:** ...6ppdnq [Show](#) [Copy](#)  
**Created:** Tue, May 10, 2022, 17:19:34 UTC  
**OCPU count:** 1  
**OCPU auto scaling:** Disabled ⓘ  
**Storage:** 1 TB  
**Storage auto scaling:** Disabled ⓘ  
**License Type:** License included  
**Database Version:** 19c

## Infrastructure

**Dedicated Infrastructure:** No

## Autonomous Data Guard ⓘ

**Status:** Disabled [Enable](#)

## Backup

**Last Automatic Backup:** No active backups exist for this database. [Support](#)

**Manual Backup Store:** Not Configured

## Network

**Access Type:** Allow secure access from subscribers

Oracle Autonomous Database

# 70%

Q4 Growth Rate

Annualized Consumption Revenue - ACR



# ATP – Scaling Database (50 sec.)

ATP

The screenshot displays the Oracle Cloud console interface for an Autonomous Database (ATP). The page title is "DB-20220510121752". The status is "AVAILABLE". The database is located in the "US East (Ashburn)" region. The console shows various sections for database management, including "General Information", "Infrastructure", "Autonomous Data Guard", "Backup", and "Network".

**Database Name:** DB20220510121752  
**Workload Type:** Transaction Processing  
**Compartment:** richnielec (root)  
**OCID:** ...6ppdq [Show](#) [Copy](#)  
**Created:** Tue, May 10, 2022, 17:19:34 UTC  
**OCPU count:** 2  
**OCPU auto scaling:** Disabled ⓘ  
**Storage:** 1 TB  
**Storage auto scaling:** Disabled ⓘ  
**License Type:** License included  
**Database Version:** 19c

**Infrastructure**  
**Dedicated Infrastructure:** No

**Autonomous Data Guard ⓘ**  
**Status:** Disabled Enable

**Backup**  
**Last Automatic Backup:** No active backups exist for this database.  
**Manual Backup Store:** Not Configured

**Network**  
**Access:** Time: Allow service access from anywhere.

Copyright © 2022, Oracle and/or its affiliates. All rights reserved.

# ATP – Stop Database (25 sec.)

ATP

ORACLE Cloud

Search resources, services, documentation, and marketplace

US East (Ashburn)

Overview » Autonomous Database » Autonomous Database Details

DB-20220510121752

ORACLE Cloud

Search resources, services, documentation, and marketplace

US East (Ashburn)

Overview » Autonomous Database » Autonomous Database Details

ORACLE Cloud

Search resources, services, documentation, and marketplace

US East (Ashburn)

Overview » Autonomous Database » Autonomous Database Details

Overview » Autonomous Database » Autonomous Database Details



STOPPED

DB-20220510121752

Database Actions

DB Connection

Performance Hub

Service Console

More Actions

Autonomous Database Information

Tools

Tags

## General Information

Database Name: DB20220510121752

Workload Type: Transaction Processing

Infrastructure

Dedicated Infrastructure:

# ADW - Provision Database (1 min. 20 sec.)



## DB-20220510174437

Data



AVAILABLE

## DB-20220510174437

- Database Actions
- DB Connection
- Performance Hub
- Service Console
- More Actions v

### Autonomous Database Information

Tools Tags

#### General Information

**Database Name:** DB20220510174437  
**Workload Type:** Data Warehouse  
**Compartment:** richniemiac (root)  
**OCID:** ...iabjpa [Show](#) [Copy](#)  
**Created:** Tue, May 10, 2022, 22:57:35 UTC  
**OCPU count:** 1  
**OCPU auto scaling:** Disabled ⓘ  
**Storage:** 1 TB  
**Storage auto scaling:** Disabled ⓘ  
**License Type:** License included  
**Database Version:** 22c...

#### Infrastructure

**Dedicated Infrastructure:** No

#### Autonomous Data Guard ⓘ

**Status:** Disabled [Enable](#)

#### Backup

**Last Automatic Backup:** No active backups exist for this database.  
**Manual Backup Store:** Not Configured

#### Network



# Autonomous DB: ATP-ADW-AJD-APEX

The screenshot displays the Oracle Cloud console interface for the 'richniemiec (root)' compartment. The left sidebar shows navigation options like Home, Compute, Storage, Networking, Oracle Database, Databases, Analytics & AI, Developer Services, Identity & Security, Observability & Monitoring, Hybrid, Migration, Billing & Cost Management, and Governance & Auditing. The main content area is titled 'Autonomous Databases in richniemiec (root) Compartment' and includes a description of Autonomous Database services. A table lists several databases with their status, dedicated infrastructure, OCPUs, storage, workload type, and creation dates.

**Autonomous Database**  
Autonomous Database delivers fast performance and requires no database administration. It performs all routine database maintenance tasks while the system is running, without human intervention. Autonomous Databases located in the Oracle cloud can run on dedicated or shared infrastructure. [Learn more](#)

**Autonomous Databases in richniemiec (root) Compartment**

Display Name	State	Dedicated	OCPUs	Storage	Workload Type	Autonomous Data Guard	Created
DB-20220510174437	Stopped	No	1	1 TB	Data Warehouse	—	Tue, May 10, 2022, 22:57:35 UTC
apexXZ23	Stopped	No	1	1 TB	APEX	—	Tue, Feb 8, 2022, 17:45:09 UTC
DB-202008271615	Stopped	No	1	1 TB	JSON Database	—	Thu, Aug 27, 2020, 21:17:04 UTC
DB-202003311814	Stopping	No	1	1 TB	Data Warehouse	—	Tue, Mar 31, 2020, 23:15:46 UTC
DB-201910041349	Available	No	1	1 TB	Transaction Processing	—	Fri, Oct 4, 2019, 18:49:54 UTC

ADW

ATP

AJD

APX



# ATP – Start Database (30 sec.)



DB-20220510121752



STOPPED



STARTING



AVAILABLE

DB-20220510121752

- Database Actions
- DB Connection
- Performance Hub
- Service Console

Autonomous Database Information

Tools Tags

## General Information

**Database Name:** DB20220510121752  
**Workload Type:** Transaction Processing

Infrast  
Dedicated

# ATP - Service Console (Development/ML)



The screenshot displays the Oracle Machine Learning (ATP) console interface. The top navigation bar includes the Oracle logo, 'Machine Learning', and the project name 'ADMIN\_RICH2 Project'. The main content area shows a sequence of SQL queries being executed in a 'Rich\_test' workspace. The first query creates a table 'rich\_test' with one row. The second query inserts a row with EMPNO=2000 and EMP\_NAME='Stan Horton'. The third query selects all data from 'rich\_test2', which is a copy of 'rich\_test'. Below the SQL editor, a table view shows the data for 'EMPNO', with values 1000, 2000, 3000, 1000, 2000, 3000, and 1000. To the right, a pie chart visualizes the data, with a legend for EMPNO values: 1000 (blue), 2000 (orange), and 3000 (light blue). The console also features a 'Table of Contents' on the left and a 'Quickstart Tutorial' header.

ORACLE Machine Learning ADMIN\_RICH2 Project

Table of Contents

Quickstart Tutorial: Creating Projects and Workspaces in Oracle Machine Learning

ORACLE Machine Learning ADMIN\_RICH2 Project (ADMIN\_RICH2)

### SQL Query S

Took 0 sec. Last updated by ADMIN...

```
select * from tab;
```

```
create table rich_test (empno number, emp_name varchar2(100));
```

Updated 0 row(s).

```
insert into rich_test values(2000, 'Stan Horton');
```

Updated 1 row(s).

```
select * from rich_test2;
```

EMPNO
1000
2000
3000
1000
2000
3000
1000

Rich\_test

Rich\_test2

Connected

Rich\_test

```
create table rich_test2 (empno number, emp_name varchar2(100));
```

Updated 0 row(s).

```
insert into rich_test2 values(2000, 'Stan Horton');
```

Updated 1 row(s).

```
select * from rich_test2;
```

FINISHED

1000 2000 3000



**Oracle Focus: Cloud, Data, Apps, Security, Self-Service, ML, & AI!**  
**Innovation Components for Business!**



*"I admire risk takers. I like leaders – people who do things before they become fashionable or popular. I find that kind of integrity inspirational."*

LAWRENCE J. ELLISON | *Chairman & Chief Executive Officer, 200*



# Oracle Machine Learning: **Brief Highlights Only**

READY-TO-GO

**ORACLE**

SaaS APPS w/EMBEDDED ML

Pre-built, packaged ML and data-driven SaaS applications

READY-TO-WORK

**ORACLE**

AUTONOMOUS DATABASE

Embedded AI in Databases to simplify enterprise data management

Ready-to-Build

READY-TO-BUILD

**ORACLE**

DATA SCIENCE PLATFORM

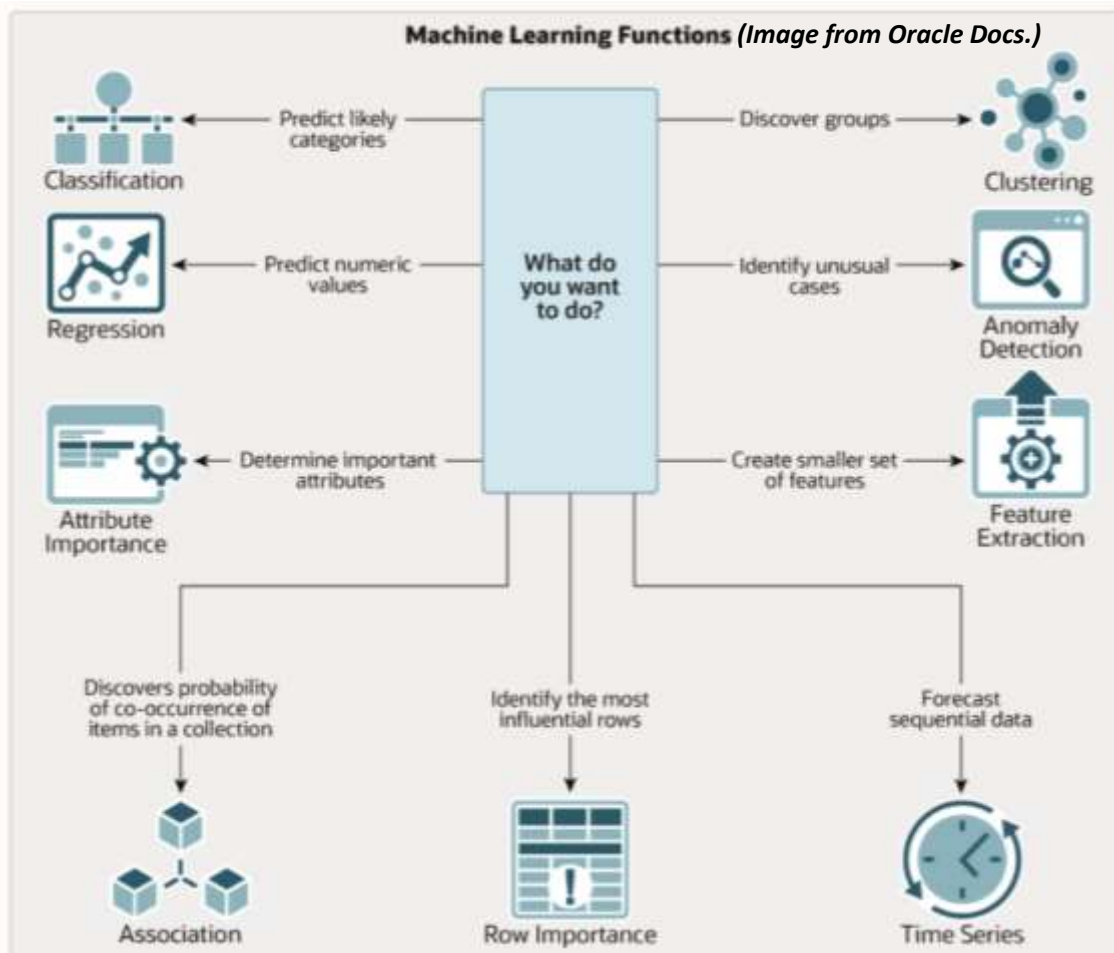
Complete platform to build and support ML-powered applications



# Machine Learning Process (Supervised Learning)

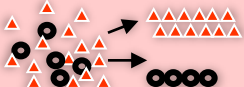
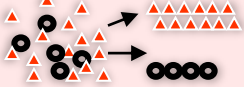
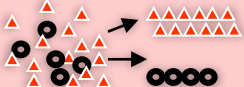




- **First:** Clear *Business Problem to Solve*
  - **Second:** *Function to Perform*
  - **Third:** *Algorithms to use*
- 
- **First:** *Build/Train the Model:* When you build it use about 60% of your data.
  - **Second:** *Test/Score Model* for accuracy/precision using about 40% of data.

**Compare Algorithms!**



# Business Understanding

## Be Extremely Specific in Problem Statement:

Poorly Defined	Better	Data Mining Technique
Predict employees that leave	<ul style="list-style-type: none"> <li>Based on past employees that voluntarily left:               <ul style="list-style-type: none"> <li>Create New Attribute <b>EmplTurnover</b> → 0/1</li> </ul> </li> </ul>	
Predict customers that churn	<ul style="list-style-type: none"> <li>Based on past customers that left (churn):               <ul style="list-style-type: none"> <li>Create New Attribute <b>Churn</b> → YES/NO</li> </ul> </li> </ul>	
Target “best” customers	<ul style="list-style-type: none"> <li>Recency, Frequency Monetary (RFM) Analysis</li> <li>Specific Dollar Amount over Time Window:               <ul style="list-style-type: none"> <li>Who has spent \$500+ in most recent 18 months</li> </ul> </li> </ul>	
How can I make more \$\$?	<ul style="list-style-type: none"> <li>What helps me sell soft drinks &amp; coffee?</li> </ul>	
Which customers are likely to buy?	<ul style="list-style-type: none"> <li>How much is each customer likely to spend?</li> </ul>	
Who are my “best customers”?	<ul style="list-style-type: none"> <li>What descriptive “rules” describe “best customers”?</li> </ul>	
How can I combat fraud?	<ul style="list-style-type: none"> <li>Which transactions are the most anomalous?               <ul style="list-style-type: none"> <li>Then roll-up to physician, claimant, employee, etc.</li> </ul> </li> </ul>	

# OAA Model Build and Real-time SQL Apply

## Simple SQL Syntax - Attribute Importance

### ML Model Build (PL/SQL)

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'BUY_INSURANCE_AI',
    mining_function     => DBMS_DATA_MINING.ATTRIBUTE_IMPORTANCE,
    data_table_name     => 'CUST_INSUR_LTV',
    case_id_column_name => 'cust_id',
    target_column_name  => 'BUY_INSURANCE',
    settings_table_name => 'CUST_INSUR_LTV_SET');
END;
/
```

#### Additional Detail:

```
drop table CUST_INSUR_LTV_SET;
exec
dbms_data_mining.drop_model('BUY_INSURANCE_AI');
```

```
create table CUST_INSUR_LTV_SET (setting_name
varchar2(30), setting_value varchar2(4000));
```

```
insert into CUST_INSUR_LTV_SET values
('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CUST_INSUR_LTV values ('PREP_AUTO','ON');
commit;
```

### Model Results (SQL query)

```
SELECT attribute_name, explanatory_value, rank
FROM BUY_INSURANCE_AI
ORDER BY rank, attribute_name;
```

<u>ATTRIBUTE NAME</u>	<u>RANK</u>	<u>ATTRIBUTE VALUE</u>
BANK_FUNDS	1	0.2161
MONEY_MONTHLY_OVERDRAWN	2	0.1489
N_TRANS_ATM	3	0.1463
N_TRANS_TELLER	4	0.1156
T_AMOUNT_AUTOM_PAYMENTS	5	0.1095

# OAA Model Build and Real-time SQL Apply Prediction

## Oracle Advanced Analytics (OAA) Simple SQL - Classification

### ML Model Build (PL/SQL)

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'BUY_INSUR1',
    mining_function     => dbms_data_mining.classification,
    data_table_name     => 'CUST_INSUR_LTV',
    case_id_column_name => 'CUST_ID',
    target_column_name  => 'BUY_INSURANCE',
    settings_table_name => 'CUST_INSUR_LTV_SET');
END;
/
```

### Additional Detail:

```
drop table CUST_INSUR_LTV_SET;
exec dbms_data_mining.drop_model('BUY_INSUR1');
create table CUST_INSUR_LTV_SET
(setting_name varchar2(30), setting_value varchar2(4000));

insert into CUST_INSUR_LTV_SET values
('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINE');
insert into CUST_INSUR_LTV_SET values ('PREP_AUTO','ON');
commit;
```

### Model Apply (SQL query)

```
Select prediction_probability(BUY_INSUR1, 'Yes'
  USING 3500 as bank_funds, 825 as checking_amount, 400 as credit_balance, 22 as age,
  'Married' as marital_status, 93 as MONEY_MONTHLY_OVERDRAWN, 1 as house_ownership)
from dual;
```

SQL   All Rows Fetched: 1 in 0.043 seconds	
PREDICTION_PROBABILITY(BUY_INSUR1,'YES'USING3500ASBANK_FUNDS,825ASCHECKING_AMOUNT,400ASCREDIT_BALANCE	
1	0.9276956709910801



# Create a Model (FYI Only)

```
DBMS_DATA_MINING.CREATE_MODEL (  
    model_name           IN VARCHAR2,  
    mining_function      IN VARCHAR2,  
    data_table_name     IN VARCHAR2,  
    case_id_column_name IN VARCHAR2,  
    target_column_name  IN VARCHAR2 DEFAULT NULL,  
    settings_table_name IN VARCHAR2 DEFAULT NULL,  
    data_schema_name    IN VARCHAR2 DEFAULT NULL,  
    settings_schema_name IN VARCHAR2 DEFAULT NULL,  
    xform_list          IN TRANSFORM_LIST DEFAULT NULL);
```

# DBMS\_DATA\_MINING Algorithms in Oracle\*

Algorithm	Abbreviation	Function	*Oracle Database PL/SQL Packages and Types Referen
Apriori	AR	Association	
CUR Matrix Decomposition	CUR	Attribute Importance	
Decision Tree	DT	Classification	
Expectation Maximization	EM	Clustering	
Explicit Semantic Analysis	ESA	Feature Extraction, Classification	
Exponential Smoothing	ESM	Time Series	
Generalized Linear Model	GLM	Classification, Regression	
$k$ -Means	KM	Clustering	
Minimum Descriptor Length	MDL	Attribute Importance	
Naive Bayes	NB	Classification	
Neural Networks	NN	Classification, Regression	
Non-Negative Matrix Factorization	NMF	Feature Extraction	
Orthogonal Partitioning Clustering	O-Cluster	Clustering	
Random Forest	RF	Classification	
Singular Value Decomposition and Principal Component Analysis	SVD and PCA	Feature Extraction	
Support Vector Machine	SVM	Classification, Regression, Anomaly Detection	

# DBMS\_DATA\_MINING Algorithms in Oracle\*

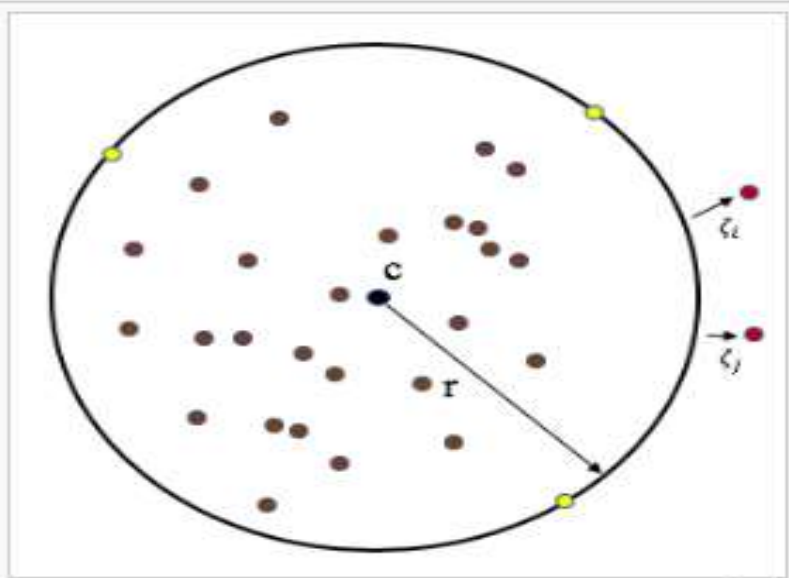
ALGO_NAME Value	Description	Mining Function
ALGO_AI_MDL	Minimum Description Length	Attribute Importance
ALGO_APRIORI_ASSOCIATION_RULES	Apriori	Association Rules
ALGO_CUR_DECOMPOSITION	CUR Decomposition	Attribute Importance
ALGO_DECISION_TREE	Decision Tree	Classification
ALGO_EXPECTATION_MAXIMIZATION	Expectation Maximization	Clustering
ALGO_EXPLICIT_SEMANTIC_ANALYSIS	Explicit Semantic Analysis	Feature Extraction Classification
ALGO_EXPONENTIAL_SMOOTHING	Exponential Smoothing	Time Series
ALGO_EXTENSIBLE_LANG	Language used for extensible algorithm	All mining functions supported
ALGO_GENERALIZED_LINEAR_MODEL	Generalized Linear Model	Classification, Regression; also Feature Selection and Generation
ALGO_KMEANS	Enhanced $k$ -Means	Clustering
ALGO_NAIVE_BAYES	Naive Bayes	Classification
ALGO_NEURAL_NETWORK	Neural Network	Classification
ALGO_NONNEGATIVE_MATRIX_FACTORIZATION	Non-Negative Matrix Factorization	Feature Extraction
ALGO_O_CLUSTER	O-Cluster	Clustering
ALGO_RANDOM_FOREST	Random Forest	Classification
ALGO_SINGULAR_VALUE_DECOMP	Singular Value Decomposition	Feature Extraction
ALGO_SUPPORT_VECTOR_MACHINES	Support Vector Machine	Classification and Regression

These are the **actual algorithm names** that need to be use.

These are **values for ALGO\_NAME** setting

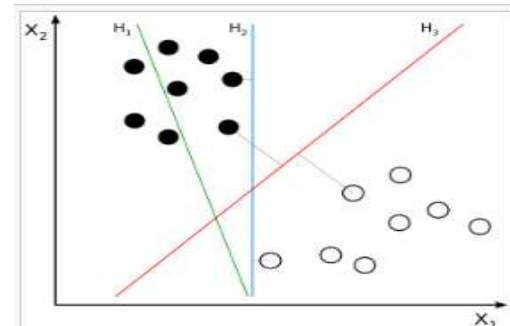
There are many additional settings that go with each Mining Function (i.e. **Confidence, Length, Rules, Aggregates, Cost, Number of Clusters...etc.**)

# One-Class SVM (ML Anomaly Detection)\*



The hypersphere containing the target data having center  $a$  and radius  $R$ . Objects on the boundary are support vectors, and two objects lie outside the boundary having slack greater than 0.

**Support Vector Data Description (SVDD):**  
Find the smallest hypersphere containing all data points (use supervised training to get it)



$H_1$  does not separate the classes.  
 $H_2$  does, but only with a small margin.  
 $H_3$  separates them with the maximal margin.

Linear SVM



# Machine Learning connection to ADW/ATP

ATP

DB Connection

Performance Hub

Service Console

Scale Up/Down

Stop

Actions

Autonomous Database Information

Tags

ORACLE Machine Learning

ADMIN\_RICH2 Project [ADMIN\_RIC...]

## Example Templates

+ Create Notebook

Search by

### Anomaly Detection

This notebook shows how to detect...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Anomaly Detection' 'Machin...

★ 5 Likes 🔍 1,600 📄 76

### Association Rules

Notebook to show the use of Asso...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Associations' 'Rules' 'M...

★ 2 Likes 🔍 772 📄 106

### Attribute Importance

Notebook to identify key attributes...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Attribute Importance' '...

★ 2 Likes 🔍 530 📄 29

### Classification Prediction M...

Example notebook to predict custo...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Classification' 'Prediction' 'De...

★ 3 Likes 🔍 1,021 📄 97

### Clustering

This notebook shows how to identL...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Clustering' 'K-Means' 'Expect...

★ 1 Likes 🔍 597 📄 35

### My First Notebook

Oracle Machine Learning example ...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Data' 'Graph'

★ 4 Likes 🔍 913 📄 27

### Regression

This notebook shows how to predic...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Regression' 'SVM' 'GLM' 'Log...

★ 1 Likes 🔍 837 📄 31

### Statistical Function

Oracle Machine Learning example ...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Statistics' 'ANOVA' 'T-test' 'F...

★ 2 Likes 🔍 356 📄 11

# Machine Learning connection to ADW/ATP

```
ORACLE Machine Learning ADMIN_RICH2 Project [ADMIN_RIC... ADMIN_RICH2
```

### Build Anomaly Detection model (1-Class Support Vector Machine)

```
%script
--Build Anomaly Detection Model (1-Class SVM) on CUSTOMERS360 data

DECLARE
v_sql varchar2(100);

BEGIN

--Create a Build Setting table for Model Build

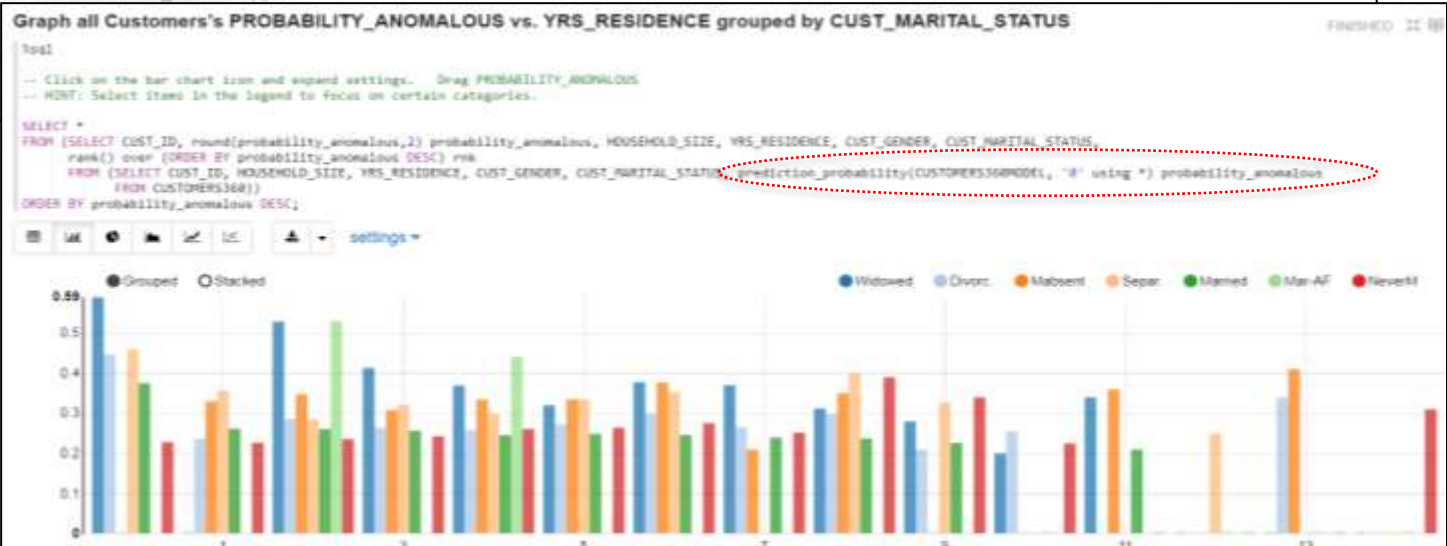
EXECUTE IMMEDIATE 'CREATE TABLE CUSTOMERS360_SET (setting_name VARCHAR2(30),setting_value VARCHAR2(8000))';
EXECUTE IMMEDIATE 'INSERT INTO CUSTOMERS360_SET (setting_name, setting_value) VALUES (''ALGO_NAME'', ''ALGO_SUPPORT_VECTOR_MACHINES'')';
EXECUTE IMMEDIATE 'INSERT INTO CUSTOMERS360_SET (setting_name, setting_value) VALUES (''PREP_AUTO'', ''ON'')';
DBMS_OUTPUT.PUT_LINE ('Created model build settings table: CUSTOMERS360_SET ');

--Build the 1-Class SVM model.

EXECUTE IMMEDIATE 'CALL DBMS_DATA_MINING.CREATE_MODEL(''CUSTOMERS360MODEL'', ''CLASSIFICATION'', ''CUSTOMERS360'', ''CUST_ID'', null, ''CUSTOMERS360_SET'')';
DBMS_OUTPUT.PUT_LINE ('Created model: CUSTOMERS360_MODEL');

END;
```

Created model build settings table: CUSTOMERS360\_SET  
Created model: CUSTOMERS360\_MODEL  
PL/SQL procedure successfully completed



# Machine Learning connection to ADW/ATP

ORACLE Machine Learning ADMIN\_RICH2 Project (ADMIN\_RIC... ADMIN\_RICH2

### Display the top 15 most anomalous customers

FINISHED

```
sql
-- Display the Top 15 Most Anomalous Customers

SELECT *
FROM (SELECT CUST_ID, round(probability_anomalous,2) probability_anomalous, HOUSEHOLD_SIZE, YRS_RESIDENCE,
      CUST_GENDER, CUST_MARITAL_STATUS, rank() over (ORDER BY probability_anomalous DESC) rnk
      FROM (SELECT CUST_ID, HOUSEHOLD_SIZE, YRS_RESIDENCE, CUST_GENDER,
            CUST_MARITAL_STATUS, prediction_probability(CUSTOMERS360MODEL, '0' using *) probability_anomalous
            FROM CUSTOMERS360))
WHERE rnk <= 15
ORDER BY probability_anomalous DESC;
```

### View Prediction\_Details that explain why the record was selected as anomalous

FINISHED

```
sql
-- Select customers with OCCUPATION of 'TechSup' and more than 55% probability of being anomalous

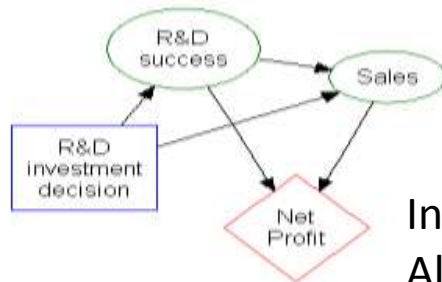
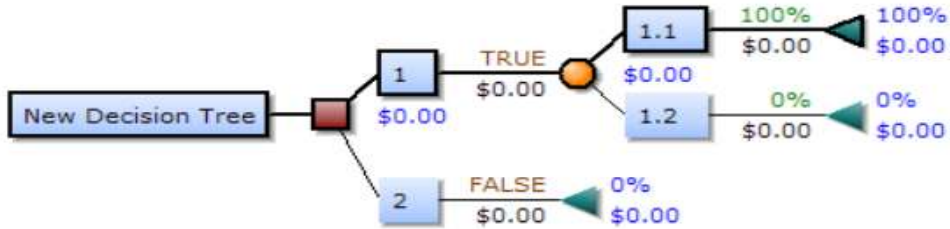
SELECT CUST_ID,
      RTRIM(TRIM(SUBSTR(OUTPRED."Attribute1",17,100)), 'rank="1"/>') FIRST_ATTRIBUTE,
      RTRIM(TRIM(SUBSTR(OUTPRED."Attribute2",17,100)), 'rank="2"/>') SECOND_ATTRIBUTE,
      RTRIM(TRIM(SUBSTR(OUTPRED."Attribute3",17,100)), 'rank="3"/>') THIRD_ATTRIBUTE,
      RTRIM(TRIM(SUBSTR(OUTPRED."Attribute4",17,100)), 'rank="4"/>') FOURTH_ATTRIBUTE,
      RTRIM(TRIM(SUBSTR(OUTPRED."Attribute5",17,100)), 'rank="5"/>') FIFTH_ATTRIBUTE
FROM (SELECT CUST_ID, PREDICTION_DETAILS(CUSTOMERS360MODEL, '0' USING *) PREDICTION_DETAILS FROM CUSTOMERS360
      WHERE PREDICTION_PROBABILITY(CUSTOMERS360MODEL, '0' USING *) > 0.55
      AND OCCUPATION = 'TechSup'
      ORDER BY CUST_ID) OUT,
```

CUST_ID	FIRST_ATTRIBUTE	SECOND_ATTRIBUTE	THIRD_ATTRIBUTE
100646	"CUST_MARITAL_STATUS" actualValue="Widowed" weight=".226"	"CUST_YEAR_OF_BIRTH" actualValue="1941" weight=".118"	"CUST_CREDIT_LIMIT" actualValue="1500" weight=".
102922	"CUST_MARITAL_STATUS" actualValue="Widowed" weight=".222"	"CUST_YEAR_OF_BIRTH" actualValue="1931" weight=".169"	"CUST_CREDIT_LIMIT" actualValue="1500" weight=".
103441	"CUST_MARITAL_STATUS" actualValue="Widowed" weight=".222"	"CUST_YEAR_OF_BIRTH" actualValue="1941" weight=".117"	"EDUCATION" actualValue="Bach." weight=".076"
104286	"EDUCATION" actualValue="9th" weight=".165"	"HOUSEHOLD_SIZE" actualValue="4-5" weight=".146"	"CUST_CREDIT_LIMIT" actualValue="1500" weight=".

# Decision Tree Algorithm (ML Classifier)

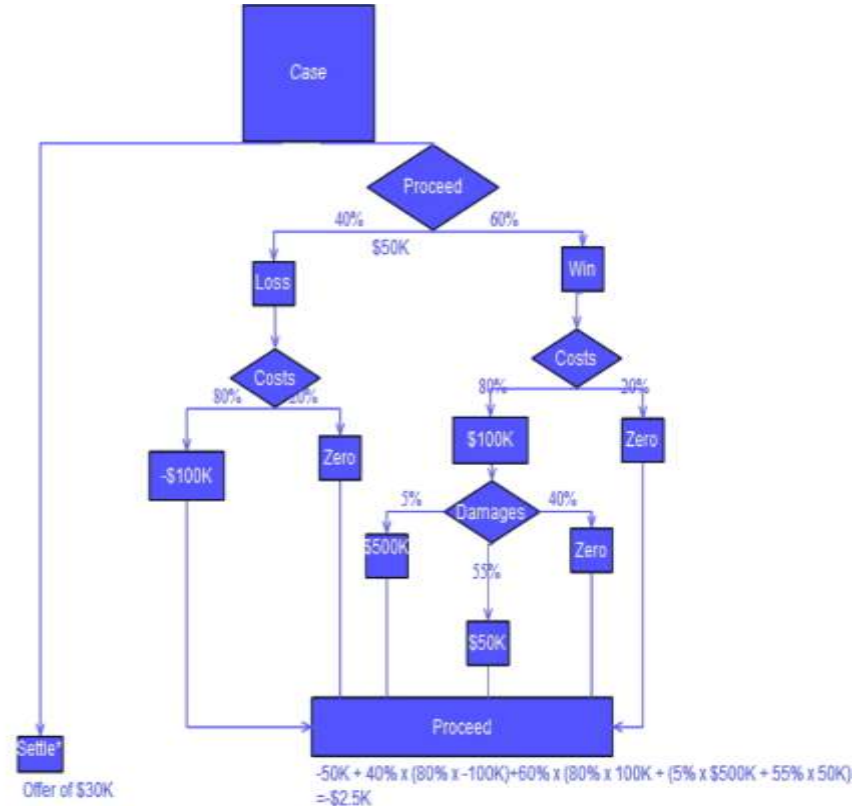
A decision tree consists of three types of nodes:<sup>[1]</sup>

1. Decision nodes – typically represented by squares
2. Chance nodes – typically represented by circles
3. End nodes – typically represented by triangles



Influence Diagram is Also a Decision Tree

Flow Chart calculates whether to settle a case or not based on costs/probabilities





# OML (OAA) Oracle Data Mining SQL Sample (PARTIAL)

## dmdtdemo.sql → DBMS\_DATA\_MINING package – Decision Tree

```
-- Given demographic data about a set of customers, predict the  
-- customer response to an affinity card program using a classifier  
-- based on Decision Trees algorithm.
```

```
INSERT INTO dt_sh_sample_settings VALUES  
(dbms_data_mining.algo_name, dbms_data_mining.algo_decision_tree);  
--  
-- CREATE A NEW MODEL  
-- Build a DT model
```

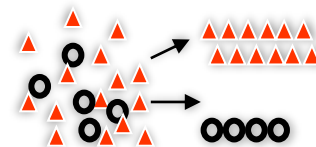
```
BEGIN  
DBMS_DATA_MINING.CREATE_MODEL(  
  model_name      => 'DT_SH_Class_sample',  
  mining_function => dbms_data_mining.classification,  
  data_table_name => 'mining_data_build_v',  
  case_id_column_name => 'cust_id',  
  target_column_name => 'affinity_card',  
  settings_table_name => 'dt_sh_sample_settings');  
END;  
/
```



```
SELECT T.cust_id, S.prediction, S.probability, S.cost  
FROM (SELECT cust_id,  
  PREDICTION_SET(dt_sh_class_sample COST MODEL USING *) pset  
FROM mining_data_apply_v  
WHERE cust_id < 100011) T,  
TABLE(T.pset) S  
ORDER BY cust_id, S.prediction;
```

Script Output x  
Task completed in 3.075 seconds

CUST_ID	PREDICTION	PROBABILITY	COST
100001	0	.966183575	.270531401
100001	1	.033816425	.966183575
100002	0	.740384615	2.076923077
100002	1	.259615385	.740384615
100003	0	.909090909	.727272727
100003	1	.090909091	.909090909
100004	0	.909090909	.727272727
100004	1	.090909091	.909090909
100005	0	.272357724	5.821138211
100005	1	.727642276	.272357724
100006	0	1.000000000	.000000000
100006	1	.000000000	1.000000000
100007	0	.909090909	.727272727
100007	1	.090909091	.909090909
100008	0	.909090909	.727272727
100008	1	.090909091	.909090909
100009	0	.272357724	5.821138211
100009	1	.727642276	.272357724
100010	0	.675965665	2.592274678
100010	1	.324034335	.675965665



# OAA Oracle Data Mining SQL Sample

Starter SQL and PL/SQL Scripts for Learning and Fast-Starts



Data Mining Sample Programs

## OAA Oracle Data Mining SQL Sample Programs

Directory Listing of the Data Mining Sample Programs



**dmaidemo.sql**

dmkmdemo.sql

dmsvddemo.sql

dmardemo.sql

dmnbdemo.sql

dmsvodem.sql

**dmdtdemo.sql**

dmsvrдем.sql

dmdtxvlddemo.sql

dmocdemo.sql

dmtxtnmf.sql

dmemdemo.sql

dmsh.sql

dmtxtsvm.sql

dmshgrants.sql

dmglrdem.sql

dmstardemo.sql

dmhpdemo.sql

dmsvcdem.sql

# A Game of Pool (*Talent / Luck / Work / Nerve*)

*(Your Current Team Plays Robots built on the Best Past Teams or ML)*



**Gambling on virtual reality: the online casinos of the future**



Image Credit: Shutterstock



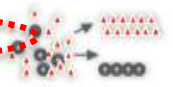


# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

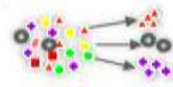


### CLASSIFICATION



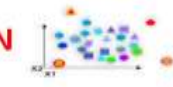
- Naive Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



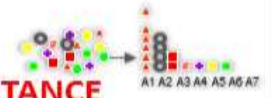
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



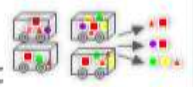
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION

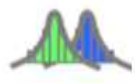
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

### EXPORTABLE ML MODELS

- REST APIs for deployment



# Naïve Bayes (ML Classifier)

- Probabilistic classifier (**Bayes Theorem**):  
*output = prior x likelihood/evidence (simplified)*
- Word frequencies for text categorization (reduce spam)
- Also used in Medical diagnosis
- Probability as extension of logic: quantify knowledge shared (supervised learning)
- Probability as extension of logic: Includes personal beliefs.
- Good for **supervised learning**  
Is it M / F based on variables:  
Height/Weight/Shoe Size)

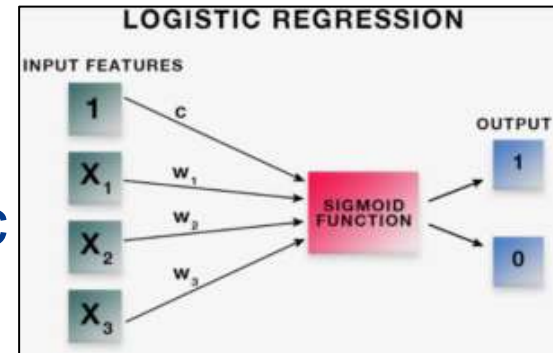


# Logistic Regression (ML Classifier)

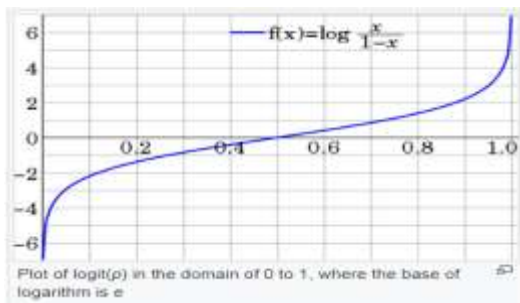
- Probability of two alternatives based on variables (also 2+)
- Estimating the parameters of logistic model
- Predicts mortality based on Injury score (TRISS)
- Predict **customer tendencies of purchases**
- Predict voting based on demographics (D or R)
- Predict mortgage default likelihood (0 or 1)
- Predict risk of developing disease

## Example:

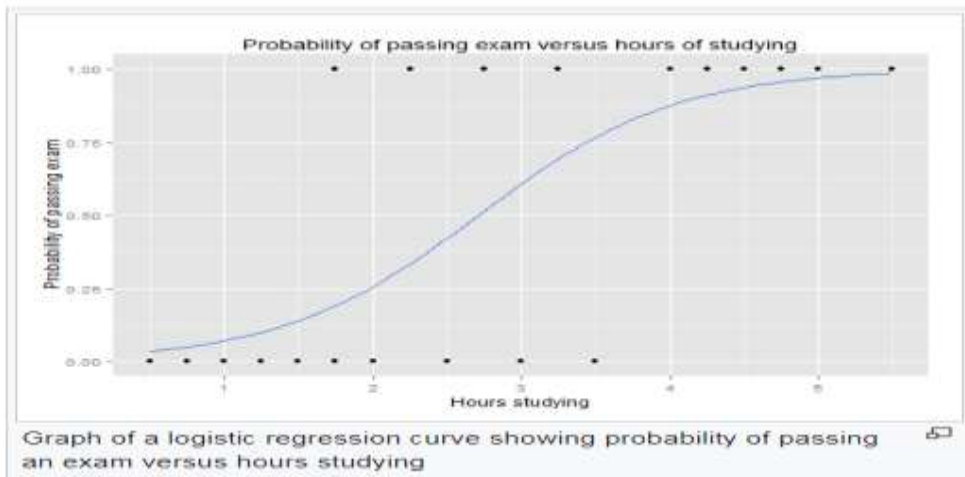
Pass/Fail based on Hours Studied - logistic  
Childs Height changes 2"/yr - linear



# Logistic Regression (ML)

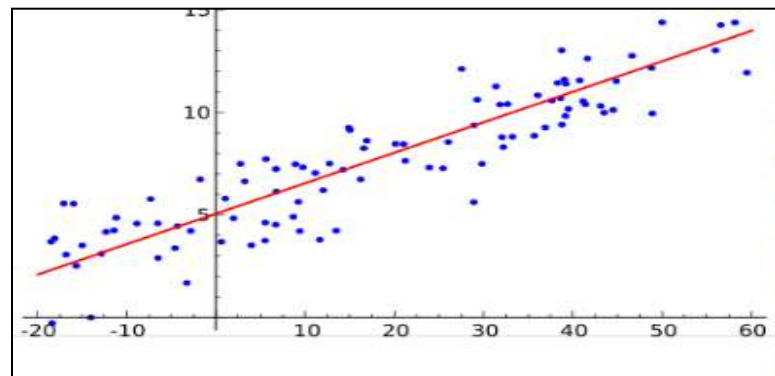


Is it a 0 or 1; Log Odds or Logit

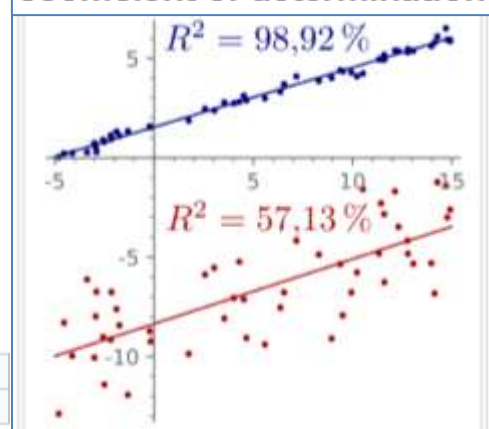


Hours	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1

Where's line for multiple values (linear regression)

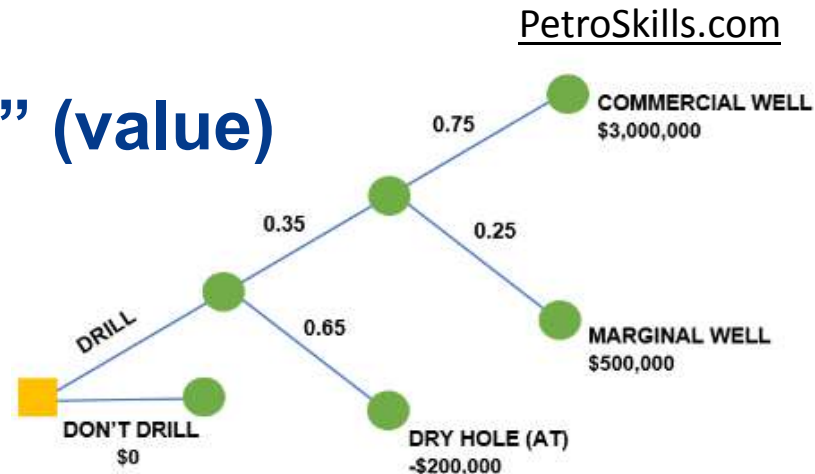


coefficient of determination



# Decision Tree (ML Classifier)

- Tree model for decisions and consequences
- Great for decision making to reach a goal
- Great to **calculate odds of different choices**
- Can have chance outcomes
- Can **calculate costs & “utility” (value)**
- Could be used for chatbots (Get monthly payments)



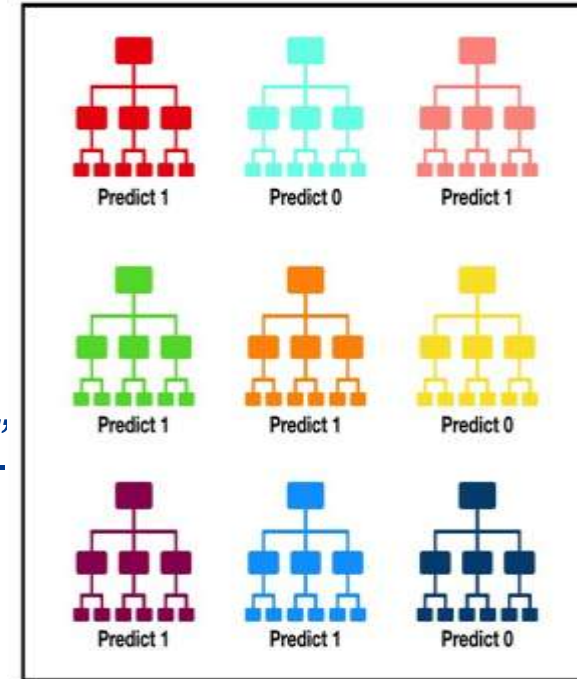
$$-200K \cdot .65 + 500K \cdot .35 \cdot .25 + 3000K \cdot .35 \cdot .65 = \mathbf{701K}$$



# Random Forest (ML Classifier)

- Multitude of Decision Trees (picks the mode)
- **Output: Class of decision tree most likely**
- *Mode* of the classes is mean or *most likely*
- Mode is found during training time
- **Separate the Real Data from the Noise!**  
(divide and conquer)
- **Fix decision tree “overfitting” or “too close of fit”**  
with supervised learning training set.
- Miss the outliers if overfitting occurs
- *Bagging* will smooth things out (stability)

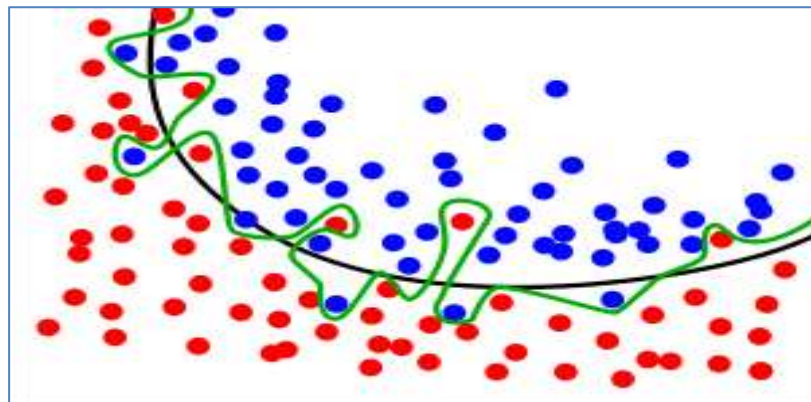
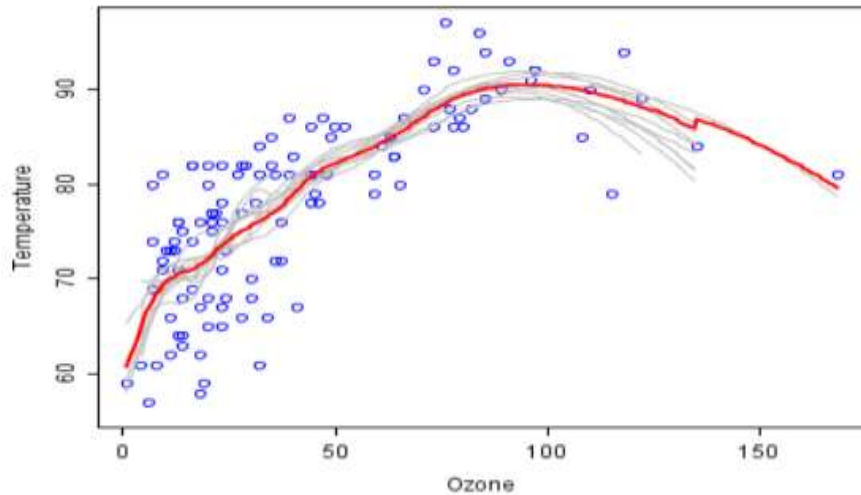
TowardsDataScience.com



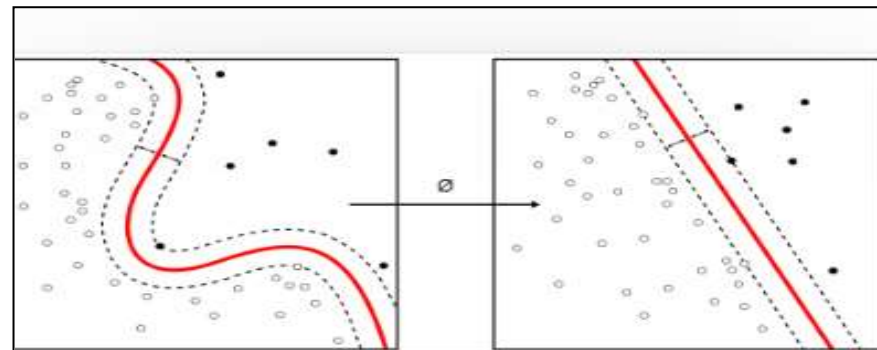
Tally: Six 1s and Three 0s

Prediction: 1

# Fixing Random Forest (ML Classifier)



Fixing Random Decision Forest by **avoiding overfitting** and provide stability to the model or graph.



**Bootstrap aggregating**, also called **bagging**, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Although it is usually

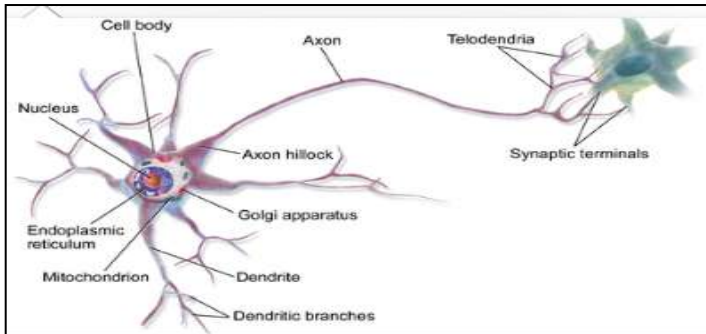
# Neural Network (ML Classifier)

- Artificial Neural Networks (ANN) make ML/AI fast
- Deep Learning (Google's **Tensor Flow** leverages this)
- Image Classification with Tensor Flow
- Classify patterns & sequences, regression testing...
- Data Mining, Spam Filtering, Robotics
- **Supervised or Unsupervised Learning**
- **Autonomous Vehicles**
- Use for Predictive Modeling
- Used with **Speech Recognition**

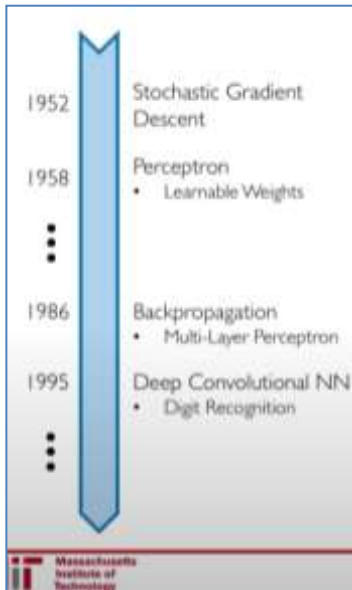


# Neural Network (ML Classifier)

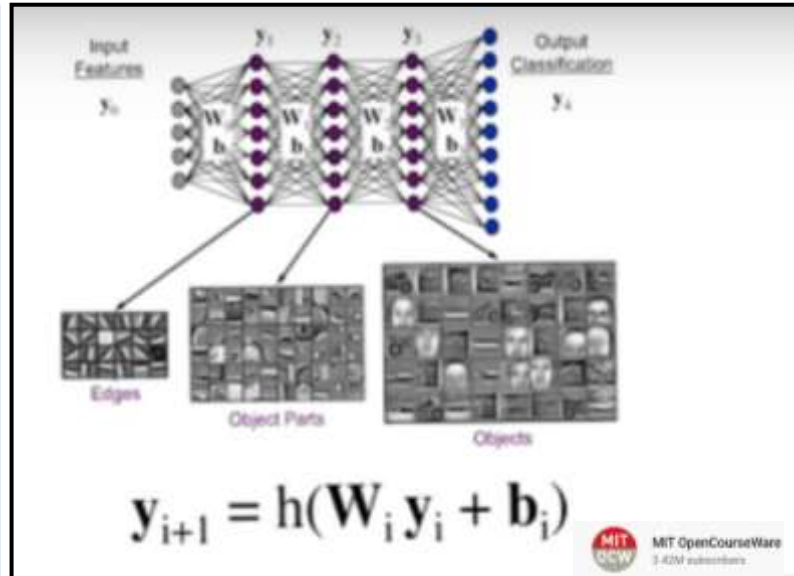
## Biological Neural Network



A **neural circuit** is a population of neurons interconnected by synapses to carry out a specific function when activated. Neural circuits interconnect to one another to form large scale brain networks. Biological neural networks have inspired the design of artificial neural networks, but artificial neural networks



## AI Neural Network



$y$  = vector/image

$W$  = weight (each line weight/strength)

$b$  = thresholds/vector or bias

$y$  times  $W$  is just Vector x Matrix

$h$  = rectified linear unit ( $>0$ )

Set: #layers /  $W$ 's /  $b$ 's / #neurons / features

Use BP or Backpropagation to update weights to minimize loss; use gradient descent or stochastic gradient descent [ differential].

(PyTorch autograd / Dropout so you don't overfit)

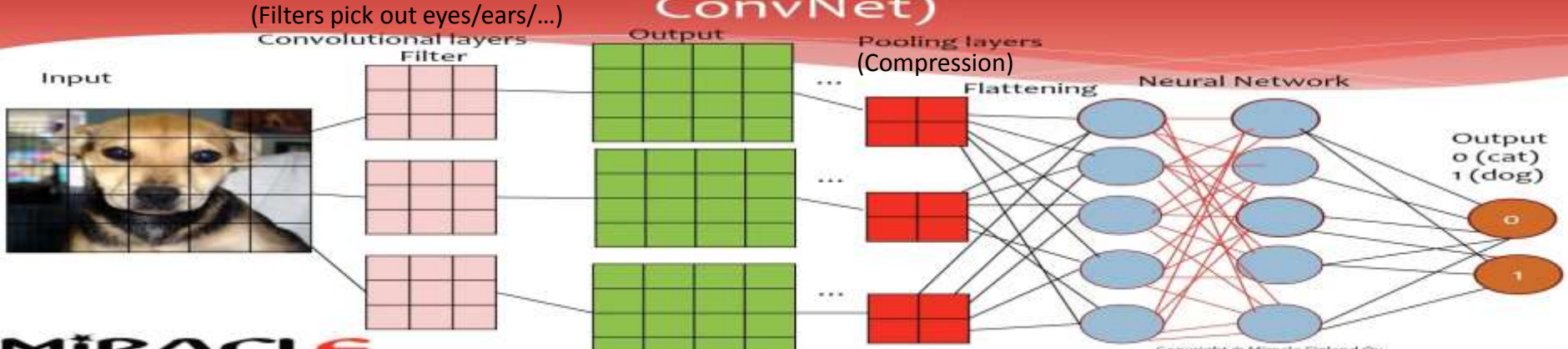


NN's also called ANN's /Artificial Neural Networks

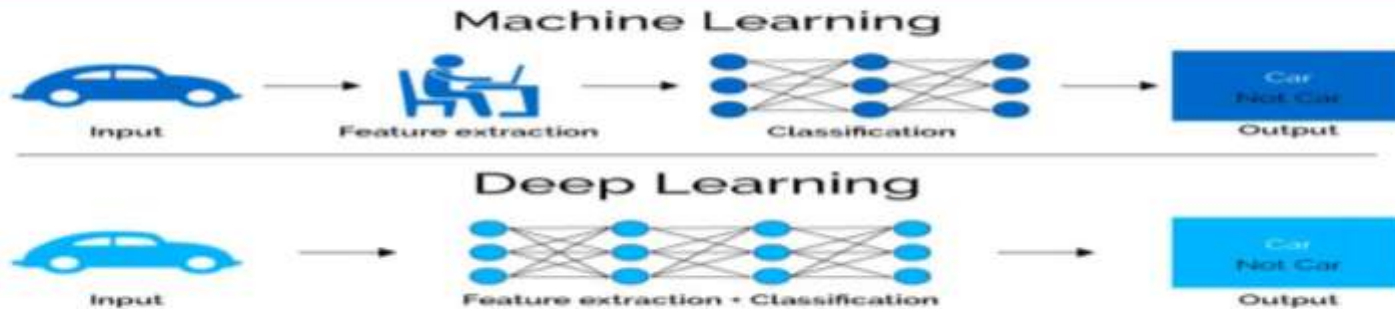


# Neural Network - CNN (ML Classifier)\* \*HeliFromFinland

## Convolutional Neural Networks (CNN, or ConvNet)



## What is deep learning?



<https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063>

# Support-Vector Machine (ML Classifier)

- Used to Classify Data (groups/separates objects)
- They analyze the data used for **regression**
- **SVM** as binary non-probabilistic linear classifier
- **Classify images, text, & even hand written text.**
- SVM maps training data into separated areas
- **Supervised Learning:**  
Uses **Algorithms** *with your data*
- **Unsupervised Learning:**  
**Data clustering** of *random data*



# Explicit Semantic Analysis (ML Classifier)



- Used for **NLP (Natural Language Processing)**
- Used to Categorize Text
- **Text** represented as a **Vector**
- **String of words** as the **Centroid of Vectors**
- **Semantic relatedness** based on **vectors**
- Must **use a Knowledge Base** (Learning)
- **NLP** is **Very Hard** in world of ML (use pre-built)
- Uses **weighting factors** based on **frequency...**

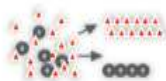


# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

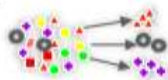


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



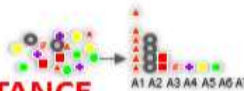
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION



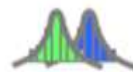
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

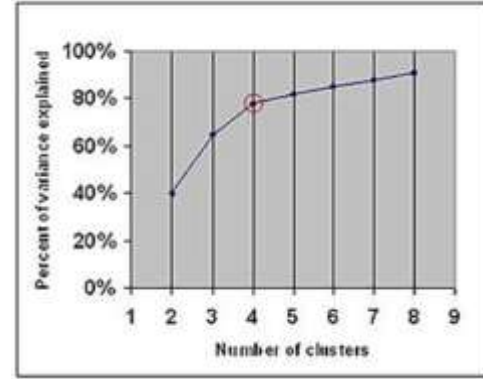
### EXPORTABLE ML MODELS

- REST APIs for deployment

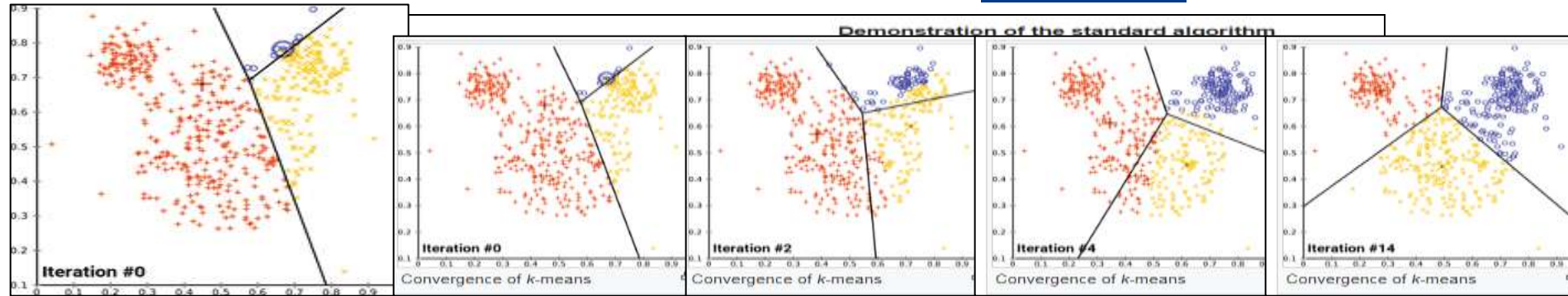


# Hierarchical K-Means (ML Cluster)

- Use the **K-means (Lloyd's) Algorithm**
- Use **observations** of the Data Set
- Clusters **generated based on each observation with nearest mean & hierarchies of clusters.**
- **Centroid cluster is new mean - Repeat**

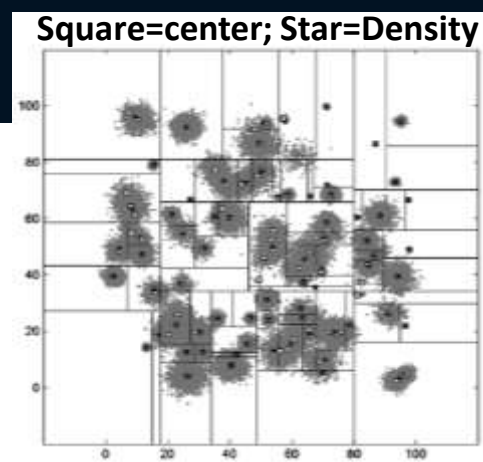


How Many? Elbow Method (4 in image)!



# Hierarchical O-Cluster (ML Cluster)

- **Orthogonal Partitioning Clustering**
- **Oracle-Proprietary clustering algorithm**
- Algorithm **operates recursively**
- Creates dense areas in attribute space
- Hierarchical grid-based with Axis-parallel (orthogonal) partitions
- *Sensitivity defines baseline* density level (to separate peaks/valleys)
- **Areas with peak density above baseline are clusters**
- Clusters used to generate **Bayesian probability**
- Outliers can cause issue; pre-clip outliers
- **K-Means is Distance Based; O-Cluster is Density Based!**



# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

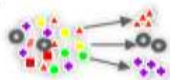


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



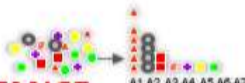
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



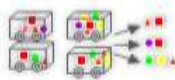
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION



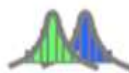
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



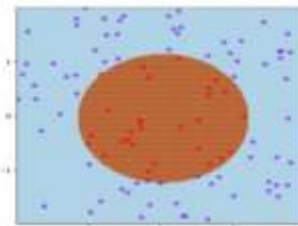
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration

### EXPORTABLE ML MODELS

- REST APIs for deployment

# One-Class SVM (ML Anomaly Detection)

- Support Vector Machine (**SVM**)
- **One-Class Classification (OCC)**
- Used to Classify Data
- **Supervised Learning** using **Algorithms**
- **Identify specific objects of a class** based on supervised learning based on objects of that class.
- SVM maps training data into separated areas
- **Unsupervised Learning** does data **clustering**
- They analyze the data used for **regression**
- SVM as binary non-probabilistic linear classifier
- **Classify images, text, & even hand written text.**



Fraud detection  
Use SVM OCC



# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

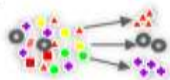


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



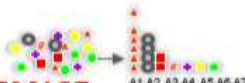
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



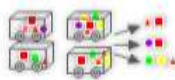
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

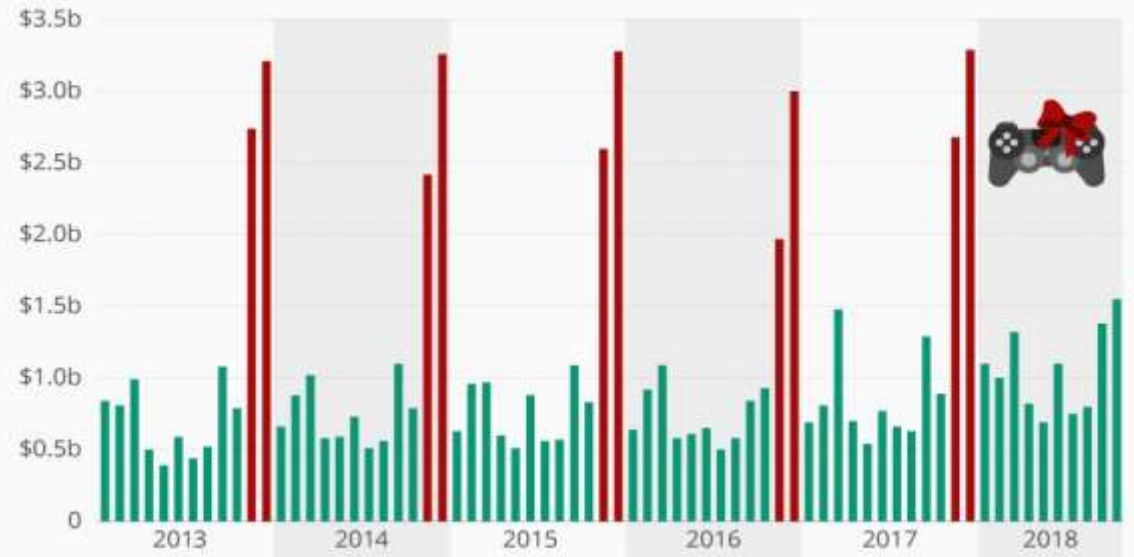
### EXPORTABLE ML MODELS

- REST APIs for deployment

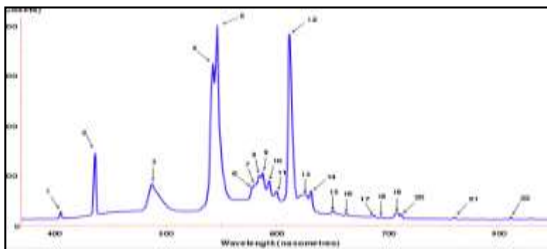
# Seasonal, Irregular & Missing Data: Time Series Algorithm

## Video Game Sales Are Extremely Seasonal

Monthly video games industry sales in the United States\*



\* Incl. video games hardware, PC & video games software, accessories and game cards  
 Source: NPD Group

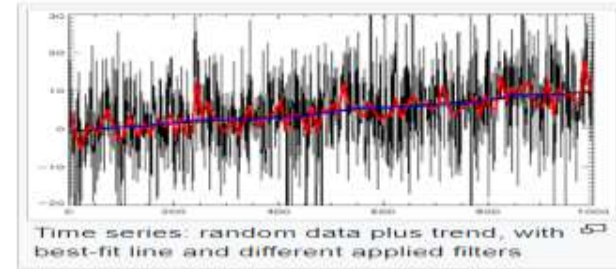


The power spectrum  $S_{x,x}(f)$  of a time series  $x(t)$  describes the distribution of power into frequency components composing that signal. According to Fourier analysis, any physical signal can be decomposed into a number of discrete frequencies, or a spectrum of frequencies over a continuous range.  $T$



# Exponential Smoothing (Time Series)

- State of the art **Forecasting Tool**
- **Keep the pattern & lose the “noise.”**
- Instead of Simple Moving Averages (**SMA**), recent data gets higher weighted.
- SMA all data equal & no smoothing occurs.
- **Older data is weighted lower** based on smoothing factor.
- As the smoothing factor nears zero, there is a greater smoothing effect on older data.
- A smoothing factor of 1 causes no smoothing.
- Smoothing gets rid of the “noise” in a signal
- Also, **Holt-Winters double exponential smoothing**





# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms



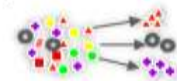
### CLASSIFICATION

- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis



### CLUSTERING

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)



### ANOMALY DETECTION

- One-Class SVM



### TIME SERIES

- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data



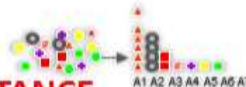
### REGRESSION

- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*



### ATTRIBUTE IMPORTANCE

- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI



### ASSOCIATION RULES

- A priori/ market basket



### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS

- SQL Windows, SQL Patterns, SQL Aggregates



### FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

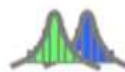
### TEXT MINING SUPPORT

- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity



### STATISTICAL FUNCTIONS

- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.



### R PACKAGES

- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration



### EXPORTABLE ML MODELS

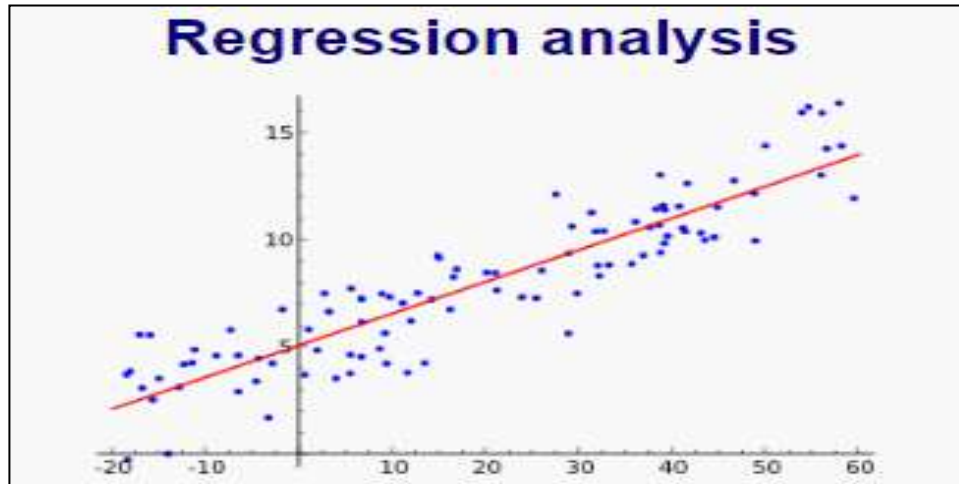
- REST APIs for deployment



# Linear Model (Regression)

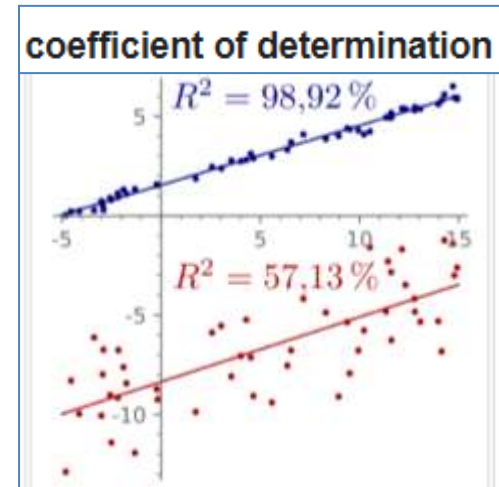
## Linear Regression (below):

Using simple **Linear Regression** we find the line between the points. There is one independent and one dependent variable. This helps us get to predictive analytics.



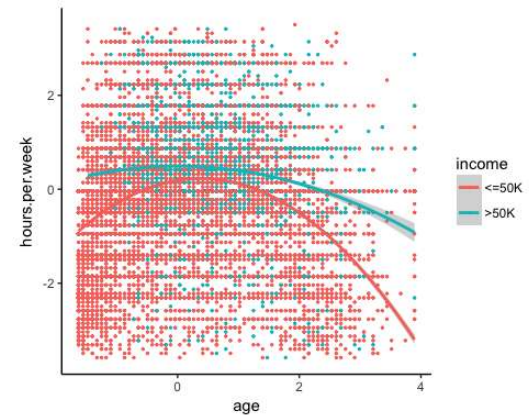
## Linear Regression (above):

Unemployment goes down & GDP goes up



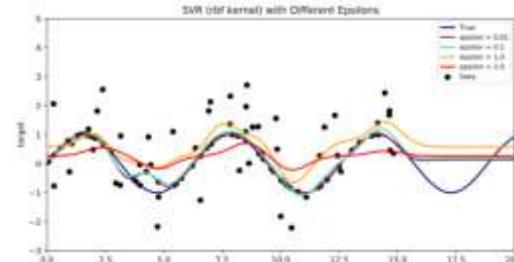
# Generalized Linear Model (GLM)

- GLM: *Flexible* generalization of linear regression
- **GLM** allows output that **doesn't match normal linear regression** & *generalizes* it by a *link function* based on variance from predicted value.
- GLM for varying situations without Normal Distributions
- GLM assumes observations uncorrelated
- GLM assumes **random vs. normal effects**
- GLM mixed models allow random effects that can be correlated through some probability.



# Support-Vector Machine Regression (SVR)

- NOT used to Classify Data (which groups/separates objects), but for regression.
- Both **Linear and Non-linear SVM Regression**
- The **Model is Produced by SVM Classification**, but depends on only a subset of training data
- SVM analyzes the data used for **regression**
- Cost function to build model is not concerned with training points that are beyond the margin.
- A margin of tolerance (Epsilon) is used for an approximation (the higher epsilon the less error)
- Supervised Learning uses Algorithms
- Unsupervised Learning uses Clustering



# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

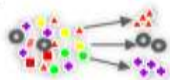


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



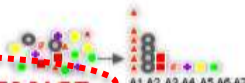
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



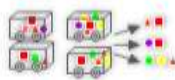
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION

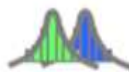
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

### EXPORTABLE ML MODELS

- REST APIs for deployment



# Minimum Description Length (Attribute Importance)

- Use **MDL** to find the **general form of a model** and its parameters
- Formalization of Ocam's Razor (*keep it simple*) by compressing the data (describe things using fewer symbols)
- Hypothesis that **includes the main things (MDL) & removes things** that are only **relatively important**
- *If methods would lead to poor results, the method should be dismissed.*
- Similar to Bayesian inference; Model and data correspond to prior probability & marginal likelihood in Bayesian framework (it's close to Bayes)
- **Similar to MML** (Minimum Message Length), but **MML is subjective Bayesian & MDL is avoids assumptions** about Data generating process (there are other differences)



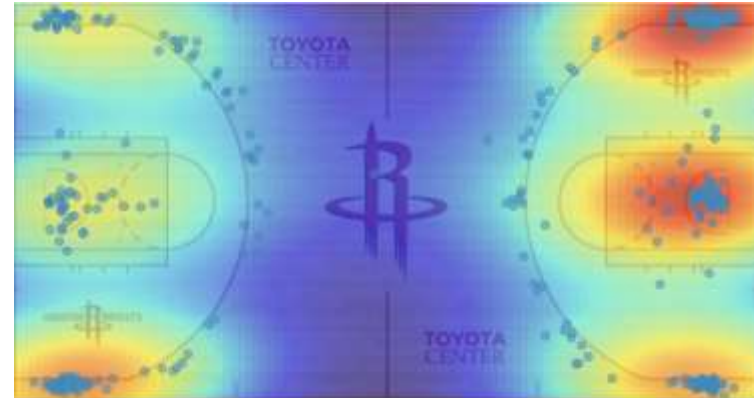
# Principal Component Analysis (Attribute Importance)

- PCA: A dimension-reduction tool **taking large set of attributes to small set (reduction)** with most important information .
- Use Feature Selection of Most Important Attributes
- PCA speeds up your ML - reducing dimension input
- “The **eigenvectors** and **eigenvalues** of a covariance (or correlation) matrix are the “core” of **PCA** and...
- The **eigenvectors** (principal components) determine the directions of the new feature space, and the **eigenvalues** determine their magnitude.”
- PCA is the simplest eigenvector based analysis (multivariate)

*(see next slide for example & later slides for example on PCA Feature Extraction)*

# Unsupervised Pair-Wise KL Div (Attribute Importance)

- Kullback-Leibler Divergence (KL Divergence)
- A measure of **how one Probability Distribution is different from another** (the divergence).
- Applications include: time-series randomness, entropy of information, information gain when comparing models
- Also called **Relative Entropy** (decline to disorder)
- Identical - KL divergence is zero
- **Measurement of *surprise***
- ***Example***: Compare NBA Players



# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

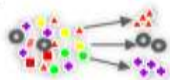


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



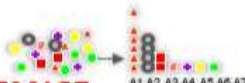
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION

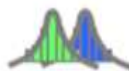
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

### EXPORTABLE ML MODELS

- REST APIs for deployment



# Apriori / Market Based (Association Rules)

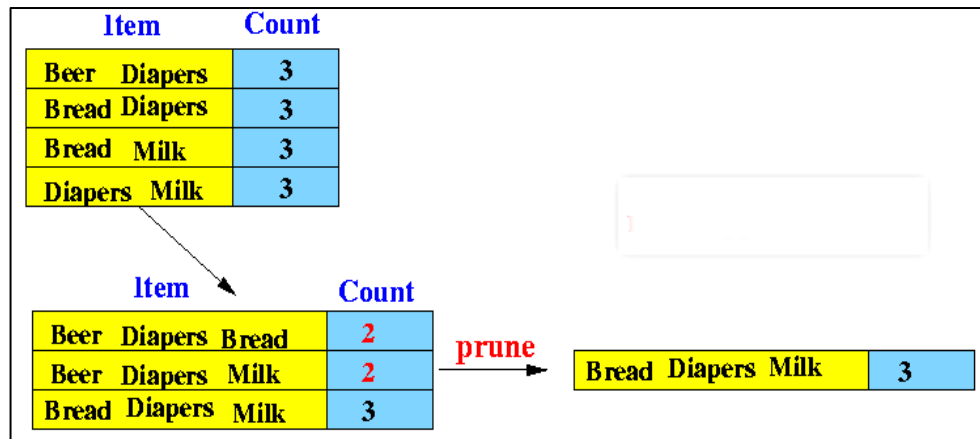
- **Apriori Algorithm** is used for association rule learning databases. **Oracle Excels here!**
- Identifies frequently associated database items
- Frequent items determine Association Rules (trends)
- Helps with Market Basket analysis (purchase behavior of customers (retail) by using information to discount/promote items
- Also shows items bought together (cheese/crackers)
- The algorithm may scan database many times (downside)
- Finds too many matches (must limit this at times)



# A Priori / Market Based (Association Rules)



T4Tutorials: Finding best Items Frequently Bought together.



Emory Univ. CS: Maybe we don't need the beer that close to the Bread, Diapers & Milk

# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

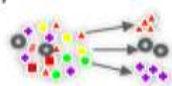


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



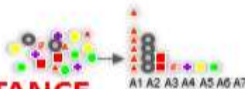
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



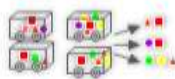
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



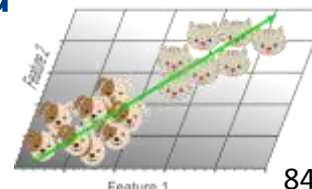
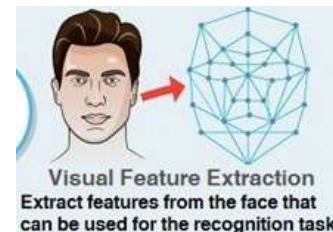
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

### EXPORTABLE ML MODELS

- REST APIs for deployment

# Principal Component Analysis (Feature Extraction)

- **PCA Feature extraction** reduces large data sets, by **building new set of attributes** making it faster to process
- **Present the same information with Fewer Variables**
- Ensure you **preserve the structure (variance)**
- Use **Feature Extraction vs. Feature Selection**
- PCA is a dimension-reduction tool **taking large set of variables to small set with most important information.**
- First component has largest variance and succeeding variables are orthogonal & highest variance.
- PCA uses orthogonal transformation to convert possibly correlated variables (numeric) into linear uncorrelated variables (principle components)
- **New Attributes** a **Combination of Old Attributes**



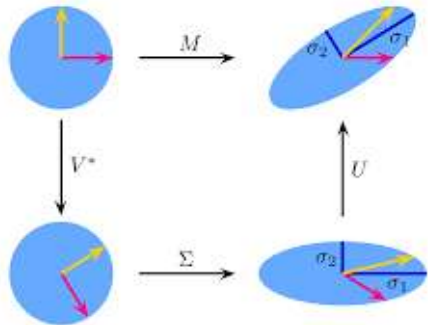


# Singular Value Decomposition (Feature Extraction)

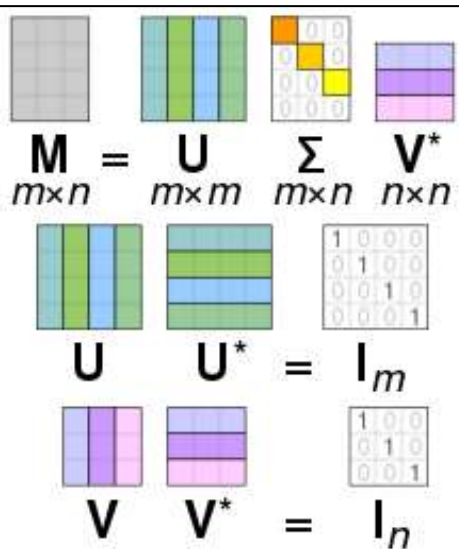
- **SVD** is the most widely used Matrix decomposition method. Used in Linear Regression, Image Compression, PCA, Least Squares and De-Noising Data (*take the noise out*)
- **SVD** is factorization (**decomposition**) of a Matrix.
- The key is to get the constituent elements from the Original Matrix (through an iterative process)
- **SVD can find singular vectors & values** from the Original Matrix
- SVD lowers the amount of features (feature reduction) of the large original matrix (a type of feature extraction)
- Used in widely in Statistics to see *more than the original matrix is showing*
- Many books on **SVD using Python** (Jason Brownlee)

The image shows handwritten mathematical notes on SVD decomposition. It includes the equation  $A = U\Sigma V^T$  and a definition of  $\Sigma$  as a diagonal matrix. A numerical example is provided for a matrix  $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ , showing the calculation of  $A^T A$  and its eigenvalues and eigenvectors. The resulting matrices  $U$ ,  $\Sigma$ , and  $V^T$  are shown as  $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ ,  $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ , and  $V^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ . The notes also mention that  $U$  and  $V$  are orthogonal matrices and that  $\Sigma$  is a diagonal matrix.

# Singular Value Decomposition (Feature Extraction)



$$M = U \cdot \Sigma \cdot V^*$$



Item x subject matrix (ISM)

	S1	S2	S3	S4	S5
dog	1	1	1	1	1
cat	1	1	0	1	0
cow	0	0	1	0	1
lion	0	0	1	1	0
tiger	1	1	0	0	1

I2tutorials.com

Singular decomposition analysis (SVD)

$$C_{m \times n} = U_{m \times r} \times \Sigma_{r \times r} \times V'_{r \times n}$$

Item vectors  $\times$  Singular values  $\times$  Subject vectors

Reducing dimensions from  $r$  to  $k$

$$\tilde{C}_{m \times n} = U_{m \times k} \times \Sigma_{k \times k} \times V'_{k \times n}$$

Item vectors  $\times$  Singular values  $\times$  Subject vectors

# Machine Learning & AI - Oracle's Built-In Algorithms

## Oracle's Machine Learning & Adv. Analytics Algorithms

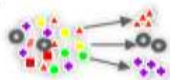


### CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

### CLUSTERING



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

### ANOMALY DETECTION



- One-Class SVM

### TIME SERIES



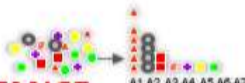
- State of the art forecasting using Exponential Smoothing.
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

### REGRESSION



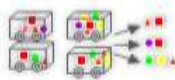
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO \*

### ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI

### ASSOCIATION RULES



- A priori/ market basket

### PREDICTIVE QUERIES

- Predict, cluster, detect, features

### SQL ANALYTICS



- SQL Windows, SQL Patterns, SQL Aggregates

### FEATURE EXTRACTION



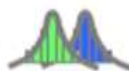
- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

### TEXT MINING SUPPORT



- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

### STATISTICAL FUNCTIONS



- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

### R PACKAGES



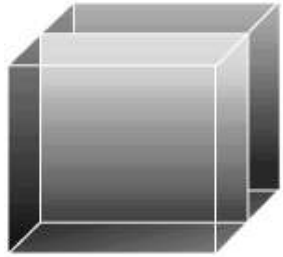
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLib algorithm integration

### EXPORTABLE ML MODELS

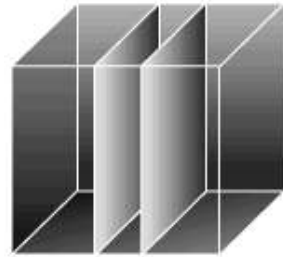
- REST APIs for deployment

# SQL Analytics (Windows / Patterns / Aggregates)\*

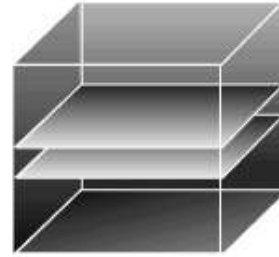
\*Oracle Docs



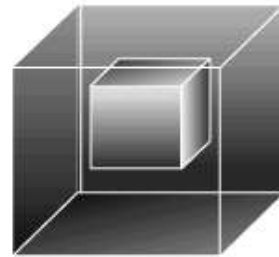
Product Mgr. View



Financial Mgr. View



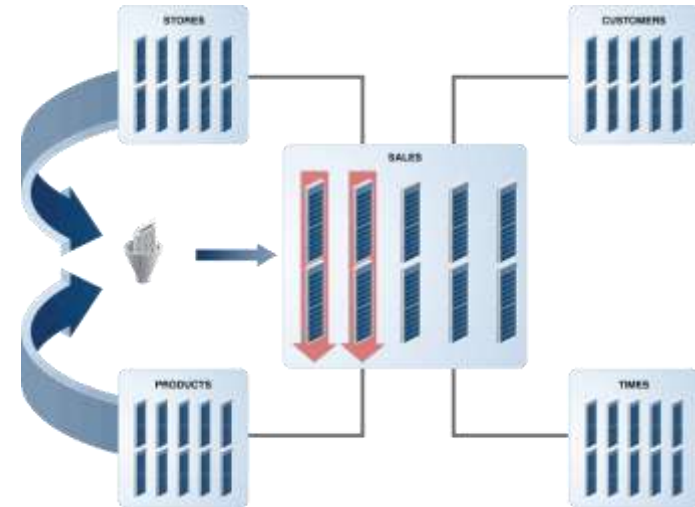
Regional Mgr. View



Ad Hoc View

**Cubes** show many dimensions for various users of the system.

**In-Memory** will use Vector Group By Aggregation





# Statistical Functions in Oracle (Partial List)\*

## Numeric Functions

ABS  
ACOS  
ASIN  
ATAN  
ATAN2  
BITAND  
CEIL  
COS  
COSH  
EXP  
FLOOR  
LN  
LOG  
MOD  
NANVL  
POWER  
REMAINDER  
ROUND (number)  
SIGN  
SIN  
SINH  
SQRT  
TAN  
TANH  
TRUNC (number)  
WIDTH\_BUCKET

## Analytic Functions

AVG \*  
CLUSTER\_DETAILS  
CLUSTER\_DISTANCE  
CLUSTER\_ID  
CLUSTER\_PROBABILITY  
CLUSTER\_SET  
CORR \*  
COUNT \*  
COVAR\_POP \*  
COVAR\_SAMP \*  
CUME\_DIST  
DENSE\_RANK  
FEATURE\_DETAILS  
FEATURE\_ID  
FEATURE\_SET  
FEATURE\_VALUE  
FIRST  
FIRST\_VALUE \*  
LAG  
LAST  
LAST\_VALUE \*  
LEAD  
LISTAGG

MAX \*  
MIN \*  
NTH\_VALUE \*  
NTILE  
PERCENT\_RANK  
PERCENTILE\_CONT  
PERCENTILE\_DISC  
PREDICTION  
PREDICTION\_COST  
PREDICTION\_DETAILS  
PREDICTION\_PROBABILITY  
PREDICTION\_SET  
RANK  
RATIO\_TO\_REPORT  
REGR\_ (Linear Regression)  
ROW\_NUMBER  
STDDEV \*  
STDDEV\_POP \*  
STDDEV\_SAMP \*  
SUM \*  
VAR\_POP \*  
VAR\_SAMP \*  
VARIANCE \*

## Aggregate Functions

APPROX\_COUNT  
APPROX\_COUNT\_DISTINCT  
APPROX\_COUNT\_DISTINCT\_AGG  
APPROX\_COUNT\_DISTINCT\_DETAIL  
APPROX\_MEDIAN  
APPROX\_PERCENTILE  
APPROX\_PERCENTILE\_AGG  
APPROX\_PERCENTILE\_DETAIL  
APPROX\_RANK  
APPROX\_SUM  
AVG  
COLLECT  
CORR  
CORR\_\*  
COUNT  
COVAR\_POP  
COVAR\_SAMP  
CUME\_DIST  
DENSE\_RANK  
FIRST  
GROUP\_ID  
GROUPING  
GROUPING\_ID  
JSON\_ARRAYAGG  
JSON\_OBJECTAGG  
LAST  
LISTAGG  
MAX

MEDIAN  
MIN  
PERCENT\_RANK  
PERCENTILE\_CONT  
PERCENTILE\_DISC  
RANK  
REGR\_ (Linear Regression) Functions  
STATS\_BINOMIAL\_TEST  
STATS\_CROSSTAB  
STATS\_F\_TEST  
STATS\_KS\_TEST  
STATS\_MODE  
STATS\_MW\_TEST  
STATS\_ONE\_WAY\_ANOVA  
STATS\_T\_TEST\_\*  
STATS\_WSR\_TEST  
STDDEV  
STDDEV\_POP  
STDDEV\_SAMP  
SUM  
SYS\_OP\_ZONE\_ID  
SYS\_XMLAGG  
TO\_APPROX\_COUNT\_DISTINCT  
TO\_APPROX\_PERCENTILE  
VAR\_POP  
VAR\_SAMP  
VARIANCE  
XMLAGG

## Data Mining Functions

CLUSTER\_DETAILS  
CLUSTER\_DISTANCE  
CLUSTER\_ID  
CLUSTER\_PROBABILITY  
CLUSTER\_SET  
FEATURE\_COMPARE  
FEATURE\_DETAILS  
FEATURE\_ID  
FEATURE\_SET  
FEATURE\_VALUE  
ORA\_DM\_PARTITION\_NAME  
PREDICTION  
PREDICTION\_BOUNDS  
PREDICTION\_COST  
PREDICTION\_DETAILS  
PREDICTION\_PROBABILITY  
PREDICTION\_SET

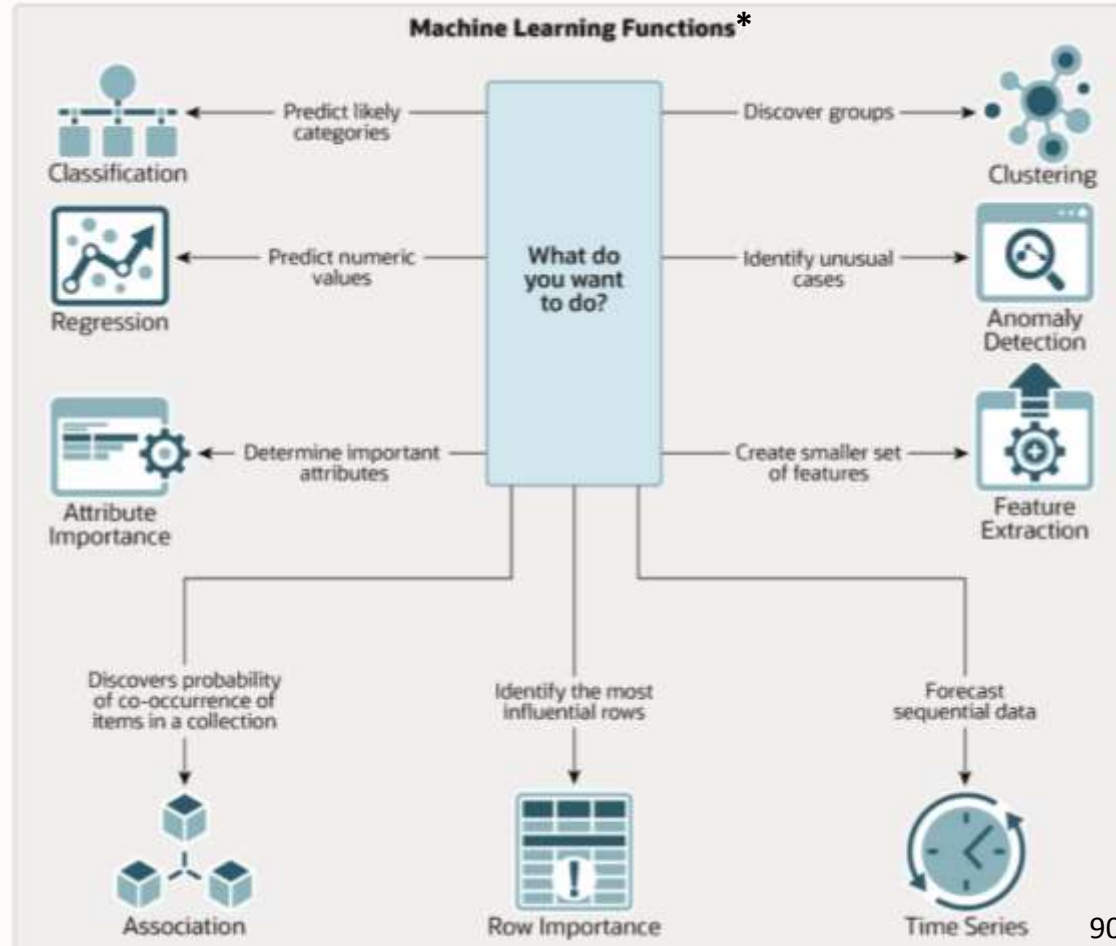
Also: T-Test, F-Test, Chi-Squared, ANOVA  
(Analysis of Variance), Pearson's Test...

\* Oracle Database SQL Language Reference

# Machine Learning Functions\* - Oracle Docs.

- **First:** Clear *Business Problem to Solve*
  - **Second:** *Function to Perform*
  - **Third:** *Algorithm to use*
- 
- **First:** Build/*Train the Model* when you build it using about 60% of data.
  - **Second:** *Test/Score Model* for accuracy/precision using about 40% of data.

**Compare Algorithms!**



# Oracle Example: Business Issue, Build & Score Model

```
-- PL/SQL procedure to Build a Decision Tree
-- classification model to predict customer
-- acceptance of offer
```

```
BEGIN DBMS_DATA_MINING.DROP_MODEL('PRED_AFFINITY_DT');
EXCEPTION WHEN OTHERS THEN NULL; END;
```

```
/
DECLARE
    v_setlst DBMS_DATA_MINING.SETTING_LIST;
```

```
BEGIN
    v_setlst('PREP_AUTO') := 'ON';
    v_setlst('ALGO_NAME') := 'ALGO_DECISION_TREE';
```

```
DBMS_DATA_MINING.CREATE_MODEL2(
    'PRED_AFFINITY_DT',
    'CLASSIFICATION',
    'SELECT * FROM MINING_DATA_BUILD',
    v_setlst,
    'CUST_ID',
    'AFFINITY_CARD');
```

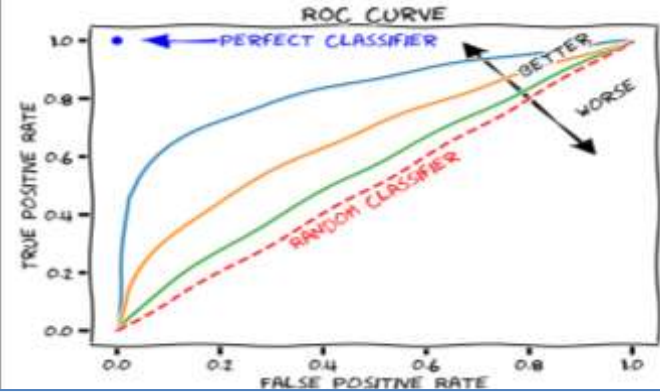
```
END;
```

```
-- SQL to dynamically score and return all
-- customers with likelihood > 50% to be
-- AFFINITY_CARD responders from MINING_DATA_APPLY
```

```
SELECT * FROM (
    SELECT CUST_ID,
           PREDICTION_PROBABILITY(PRED_AFFINITY_DT, '1'
                                   USING A.*) PROBABILITY
    FROM MINING_DATA_APPLY A)
WHERE PROBABILITY > 0.5
ORDER BY PROBABILITY DESC;
```

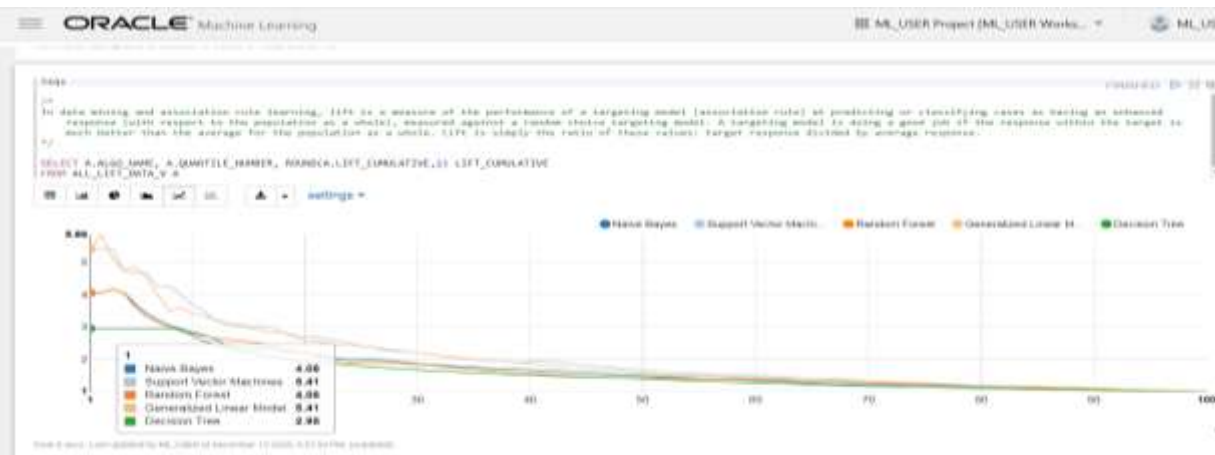


# Lift, Gain, & ROC Curves



**ROC AUC = 0.75**

VS.





# Time Enough at Last for ML with Autonomous DB



# AutoML for Autonomous Database- 3/18/2021

**ORACLE** Machine Learning

## How Do I?

- Use AutoML**  
How to create AutoML Experiments
- Get Started**  
Get started with Oracle Machine Learning

## Quick Actions

- AutoML**  
Create and run AutoML Experiments
- Scratch**  
Run Scratch

## Create Experiment

**Name \***  
AutoML\_Rich1

**Comments**

**Data Source \***  
ML\_USER.CUSTOMERS360

**Prediction Type \***  
Classification

**Predict \***  
AFFINITY\_CARD

**Case ID**  
CUST\_ID

**Additional Settings**

## AutoML Experiments

**Features**

Refresh

Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
<input type="checkbox"/> AFFINITY_CARD	NUMBER	0	2	0	1	0.24	0.6
<input checked="" type="checkbox"/> CUST_CREDIT_LIMIT	NUMBER	0	8	1500	15000	7924.22	4264
<input checked="" type="checkbox"/> CUST_GENDER	CHAR	0	2				
<input type="checkbox"/> CUST_ID	NUMBER	0	4500	100001	104500	102250.8	1306

# AutoML is here for Autonomous Database

ORACLE Machine Learning

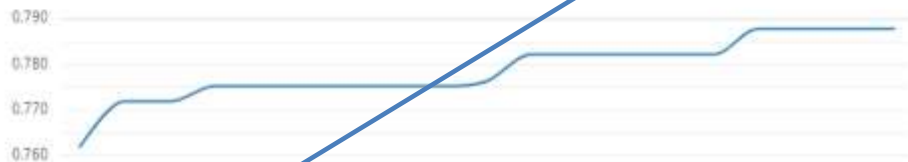
ML\_USER Project [ML\_USER Works...]

< - Experiments

AutoML\_Rich

Experiment Settings Edit

Accuracy



Leader Board

Deploy	Create Notebook	Metrics
Algorithm	Model Name	Accuracy
Support Vector Machine (Gaussian)	svmg_09cd02ee4e	0.7878
Random Forest	rf_df53fb31a8	0.7821
Generalized Linear Model	glm_74cbb2b7ac	0.7821
Generalized Linear Model (Ridge Regres...	glm_r_df24288679	0.7821
Support Vector Machine (Linear)	svml_a031bb0ae1	0.7431

Leader Board

Deploy	Create Notebook	Metrics
Algorithm	Model Name	Accuracy
Support Vector Machine (Gaussian)	svmg_09cd02ee4e	0.7878
Random Forest	rf_df53fb31a8	0.7821
Generalized Linear Model	glm_74cbb2b7ac	0.7821
Generalized Linear Model (Ridge Regres...	glm_r_df24288679	0.7821
Support Vector Machine (Linear)	svml_a031bb0ae1	0.7431

Create Notebook

Create a notebook based on selected model and this experiment's settings. Use a generated notebook to further tune your approach using Python.

Notebook Name:

AutoML\_Classify



OK

Cancel



AutoML

Create and run AutoML  
Experiments

Scratchpad

Run Scratchpad



Notebooks

The place for data  
discovery and analytics

Jobs

Schedule notebooks to  
run at certain times

Examples

Check out some  
examples

## Recent Activities

today



Rich Niamac created AutoML\_Classify notebook in ML\_USER Project [ML\_USER Workspace]

3/28/21 4:21 PM

## AutoML\_Classify



## Build Data

```

%python
import oml

columns = 'CUST_ID', 'CUST_CREDIT_LIMIT', 'CUST_GENDER', 'CUST_INCOME_LEVEL', 'CUST_MARITAL_STATUS', 'CUST_YEAR_OF_BIRTH', 'EDUCATION', 'HOUSEHOLD_SIZE', 'OCCUPATION', 'YRS_RESIDENCE',
         'Y_BOX_GAMES', 'AFFINITY_CARD'
schema='ML_USER'
table='CUSTOMERS300'

column = ','.join(columns)
query = 'SELECT ' + column + ' FROM ' + schema + '.' + table

data_build = oml.sync(query=query)
> showData hrefId

```

## Create Train Data

READY

```

%python
import oml

X_train = data_build[:,['CUST_ID', 'CUST_CREDIT_LIMIT', 'CUST_GENDER', 'CUST_INCOME_LEVEL', 'CUST_MARITAL_STATUS', 'CUST_YEAR_OF_BIRTH', 'EDUCATION', 'HOUSEHOLD_SIZE', 'OCCUPATION', 'YRS_RESIDENCE',
                        'Y_BOX_GAMES']]
y_train = data_build[:, 'AFFINITY_CARD']

```

## Build 'SUPPORT\_VECTOR\_MACHINES' Model

READY

```

%python
import oml

svm_settings = {
    'ODPS_SAMPLE_SIZE': '1294', 'SVMS_COMPLEXITY_FACTOR': '06.70000000000002', 'SVMS_KERNEL_FUNCTION': 'SVMS_GAUSSIAN', 'SVMS_STD_DEV': '2.23606797749979', 'CLAS_WEIGHTS_BALANCED': 'OFF',
    'SVMS_NUM_PIVOTS': '200', 'ODPS_DETAILS': 'ODPS_DISABLE', 'ODPS_SAMPLING': 'ODPS_SAMPLING_ENABLE', 'ALGO_NAME': 'ALGO_SUPPORT_VECTOR_MACHINES'
}

svm_mod = oml.svm(**svm_settings)

svm_mod = svm_mod.fit(X_train, y_train)

```



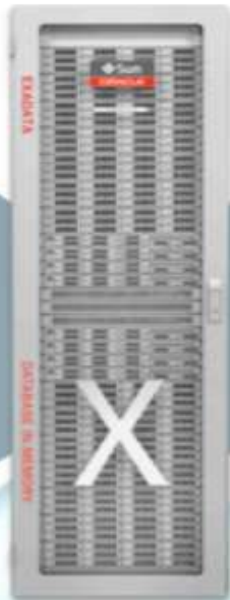
# Exadata Cloud Machine: ALL Features

(Oracle's Juan Loaiza presentation on Exadata Cloud)

**Exadata Cloud:** Compatible, Scalable, Available, Secure  
 Decades of Database Innovation Proven at Millions of Mission-Critical Deployments

	Multitenant
	In-Memory DB
	Real Application Clusters
	Active Data Guard
	Partitioning
	Advanced Compression
	Advanced Security, Label Security, DB Vault
	Real Application Testing
	Advanced Analytics, Spatial and Graph
	Management Packs for Oracle Database

**All Oracle Database Innovations**



**All Exadata DB Machine Innovations**

Offload SQL to Storage	
InfiniBand Fabric	
Smart Flash Cache, Log	<b>PCI Flash</b>
Storage Indexes	
Columnar Flash Cache	
Hybrid Columnar Compression	
I/O Resource Management	
Network Resource Management	
In-Memory Fault Tolerance	
Exafusion Direct-to-Wire Protocol	

# Oracle Machine Learning Algorithms and Analytics in Oracle Database

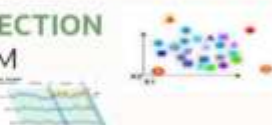
## CLASSIFICATION



- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine (SVM)
- Explicit Semantic Analysis
- *XGBoost\**

## ANOMALY DETECTION

- One-Class SVM
- *MSET-SPRT\**



## CLUSTERING

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)



## TIME SERIES



- Forecasting - Exponential Smoothing
- Includes popular models e.g. Holt-Winters with trends, seasonality, irregular time series

[OML Algorithm Cheat Sheet](#)  
[Algorithm Documentation](#)

## REGRESSION



- Generalized Linear Model (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- *XGBoost\**

## ATTRIBUTE IMPORTANCE



- Minimum Description Length
- Random Forest
- Unsupervised Pairwise KL Divergence
- CUR decomposition for row & AI

## ASSOCIATION RULES



- A priori

## SQL ANALYTICS

- SQL Windows
- SQL Patterns
- SQL Aggregates



## FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

## ROW IMPORTANCE

- CUR Decomposition

## RANKING

- *XGBoost\**



## TEXT MINING SUPPORT

- Algorithms support text columns
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA)



## STATISTICAL FUNCTIONS

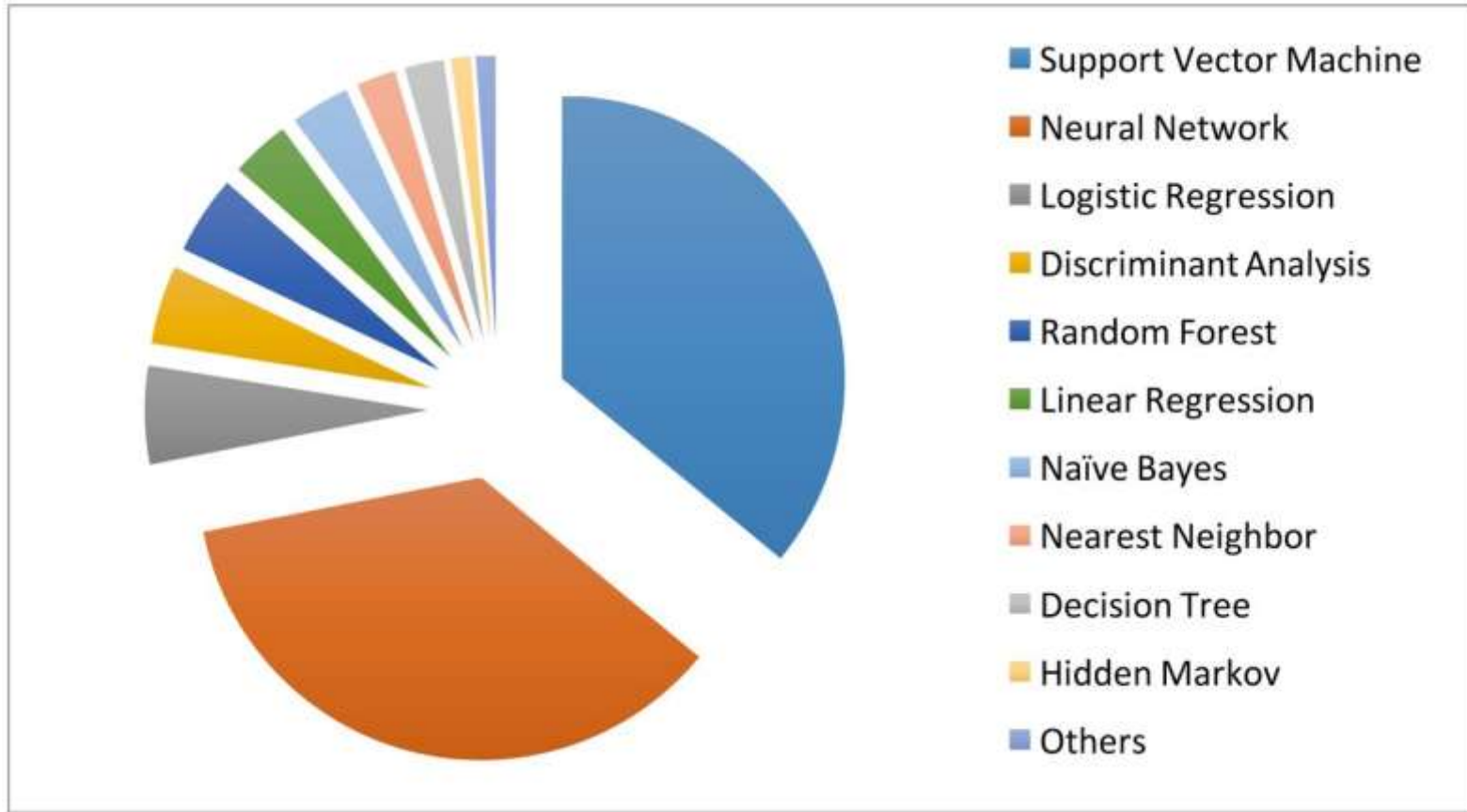
- min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.



*Includes support for partitioned models,  
integrated text mining,  
automated data preparation*



# Types of ML Algorithms in Healthcare on PubMed\*

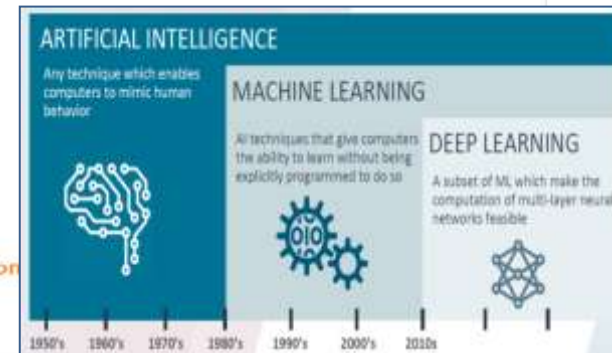
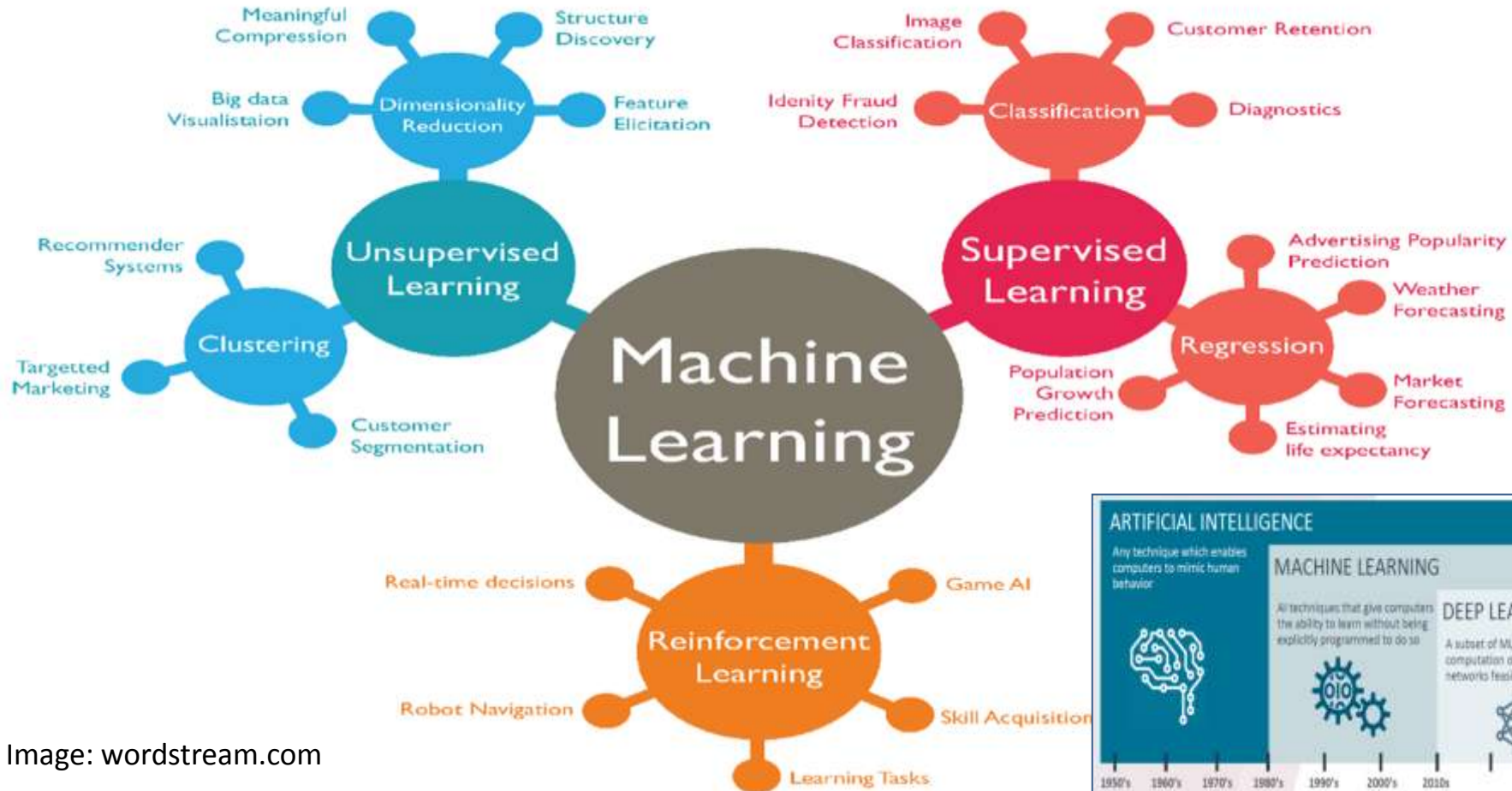


\*Stroke and vascular neurology (snj.bmj.com)

Note: Markov is Bayesian, NN is k-NN, DA close to PCA



# Machine Learning has many parts





# From Agnes with Love (Computers try to help)



Google Assistant	
Alexa	
Siri	
Bixby	



# Biju Thomas at ODTUG - *Emerging Jobs*

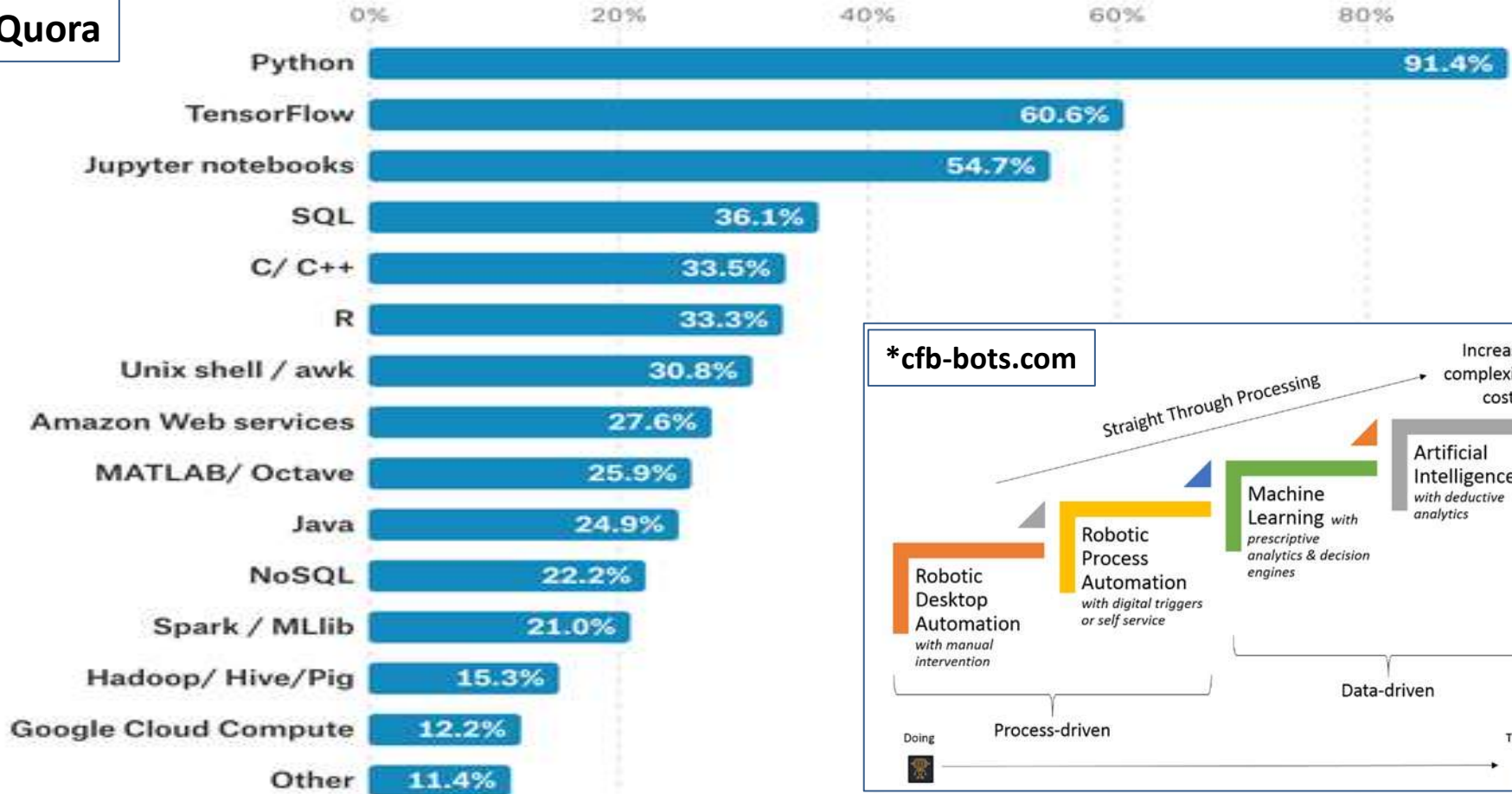
## #1 AI / ML Specialist (Engineer)



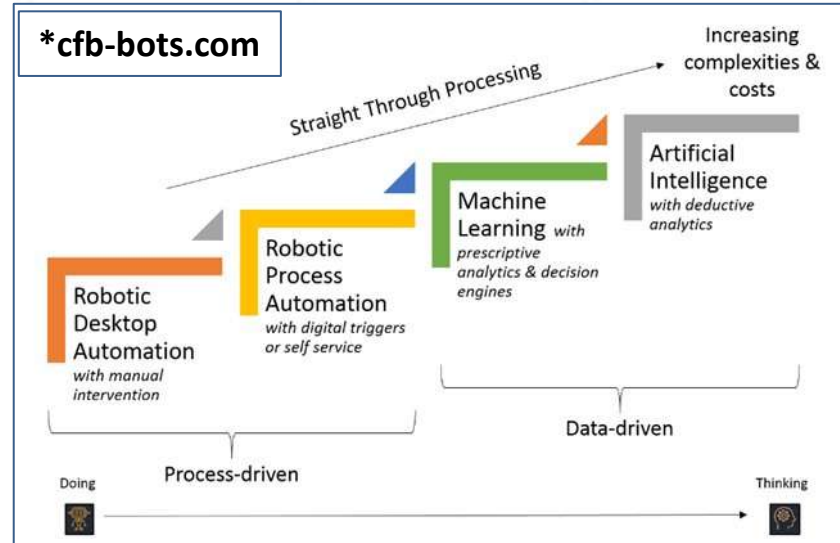
- 74% annual growth
- Skills
  - Machine Learning
  - Deep Learning
  - TensorFlow
  - Python
  - Natural Language Processing
- Industries
  - Computer Software
  - Information Technology
  - Higher Education
  - Consumer Electronics
- Take the data scientists' code and making it more effective and scalable.
- Introduce various programming rules and good practices.
- Join the results from potentially unrelated tasks to enhance the models performance even more.
- Glue all the "data science" and "software" parts together.

# Python is #1 Language for Machine Learning\*

\*Quora



\*cfb-bots.com

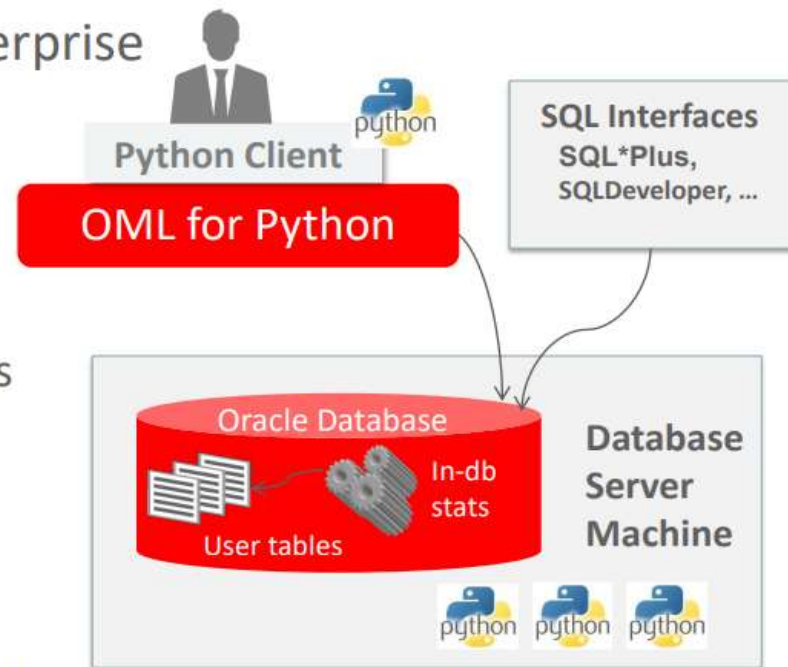


# Oracle integrates Python with OAA

## Oracle Machine Learning for Python

Oracle Advanced Analytics Option to 18c+

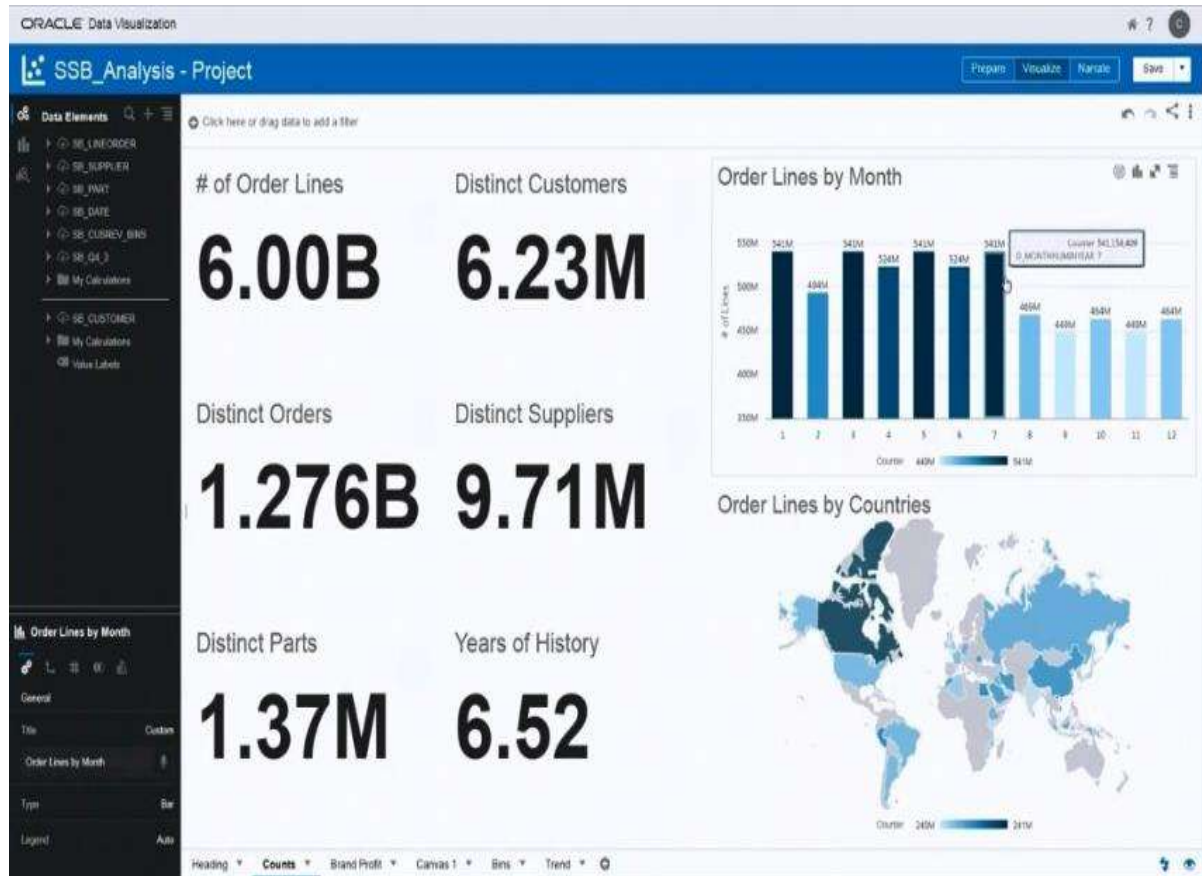
- Similar architecture to OAA's Oracle R Enterprise
- OML4Py Transparency Layer
  - Use Oracle Database as High Performance Computing environment
- OML4Py OAA Model Build and Apply
  - Use OAA parallel and distributed ML algorithms
  - Manage Python scripts and Python objects in Oracle Database
- OML4Py Embedded Python
  - Make callout to Python packages
  - Integrate Python results into applications via SQL





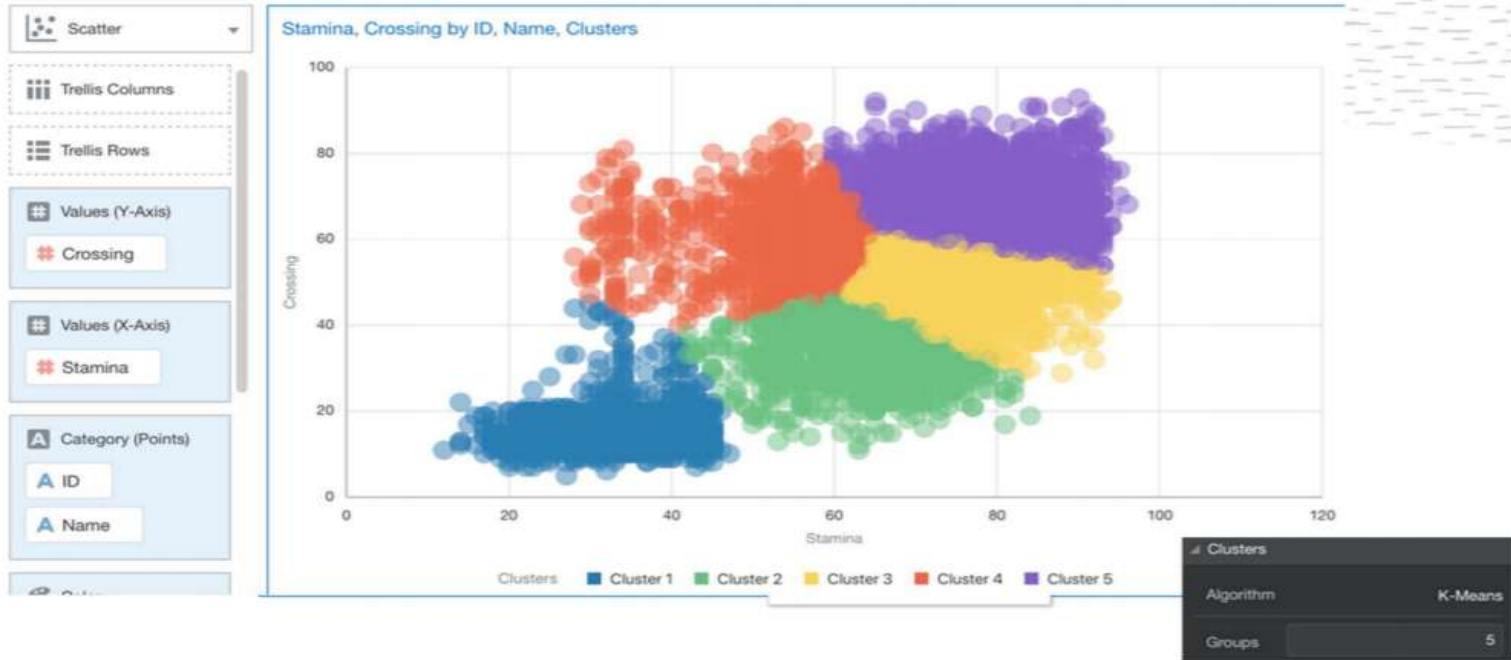
# George Mines ADW with Oracle Analytics Cloud (OAC)

OAC: Data Visualization Cloud Services (DVCS), Bus. Intell. CS (BICS), Essbase, Smart View (EE/SE)



# Oracle Analytics Cloud (OAC) to Cluster Data

## Easy Models



Oracle Analytic Summit 2020

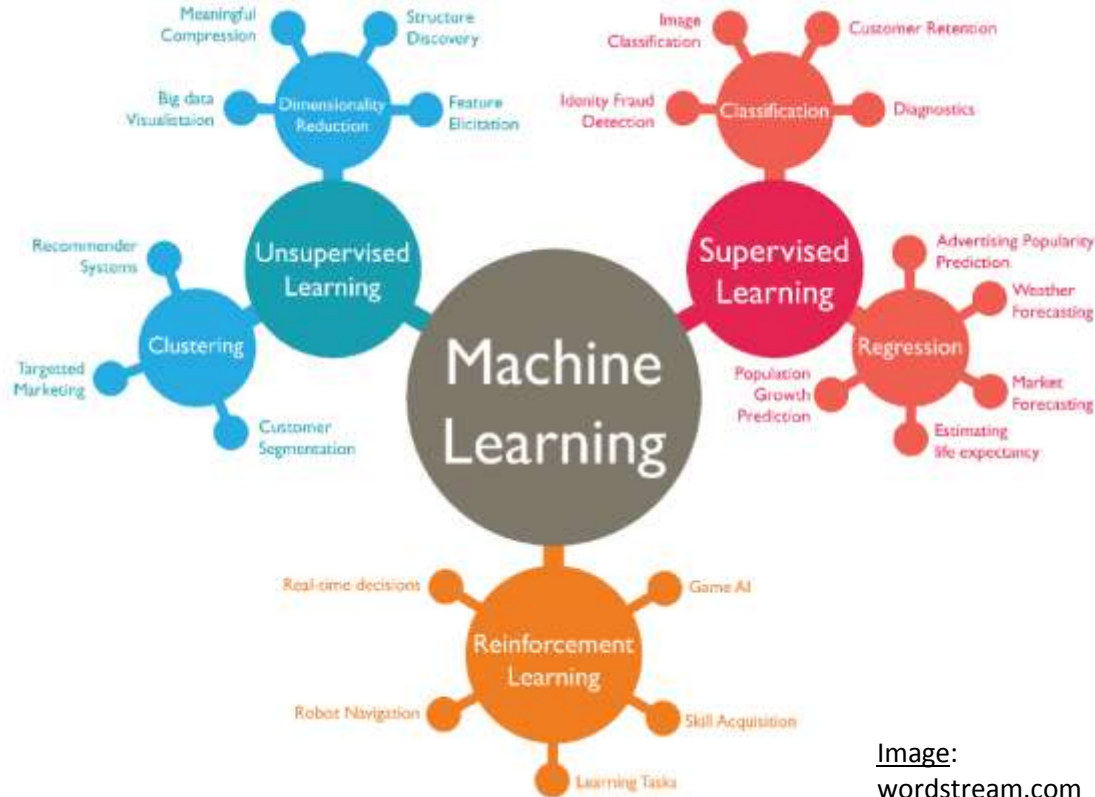


# ML & Business Apps - Today at 4:15; San Antonio 1

**ORACLE**  
SALES CLOUD

**ORACLE**  
RETAIL CLOUD

**ORACLE**  
CONFIGURE, PRICE,  
AND QUOTE CLOUD



**ORACLE**  
E-BUSINESS SUITE

**ORACLE**  
IDENTITY MANAGEMENT

**ORACLE**  
FINANCIAL SERVICES

**ORACLE**  
HUMAN CAPITAL  
MANAGEMENT  
CLOUD

Image:  
wordstream.com

# Better World – Leverage Tech!

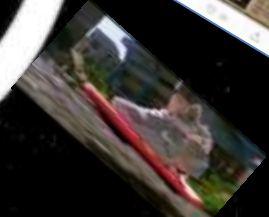
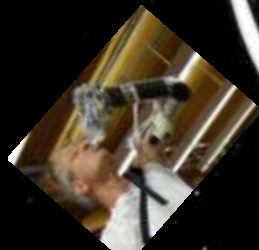
Steve Wozniak on AI: Will we be pets or mere ants to be squashed our robot overlords?

Apple co-founder Steve Wozniak predicts a 'scary' future when artificial intelligence takes

COMPUTERWORLD

# THE TWILIGHT ZONE

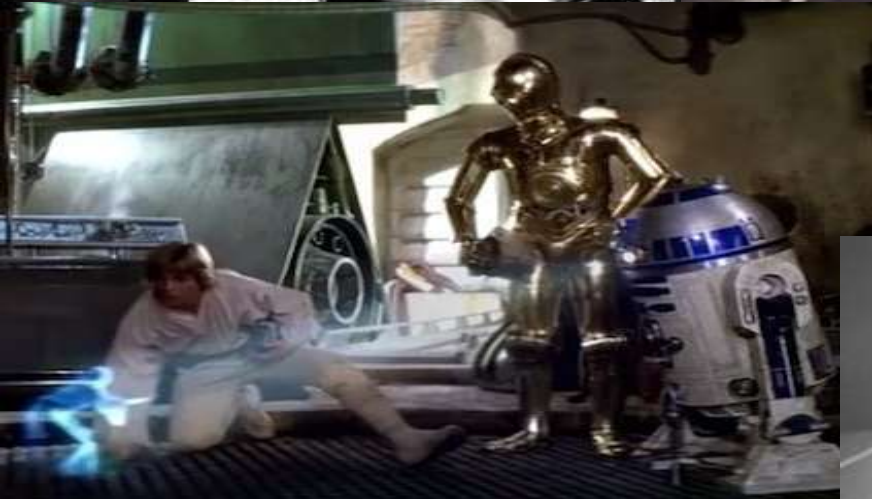
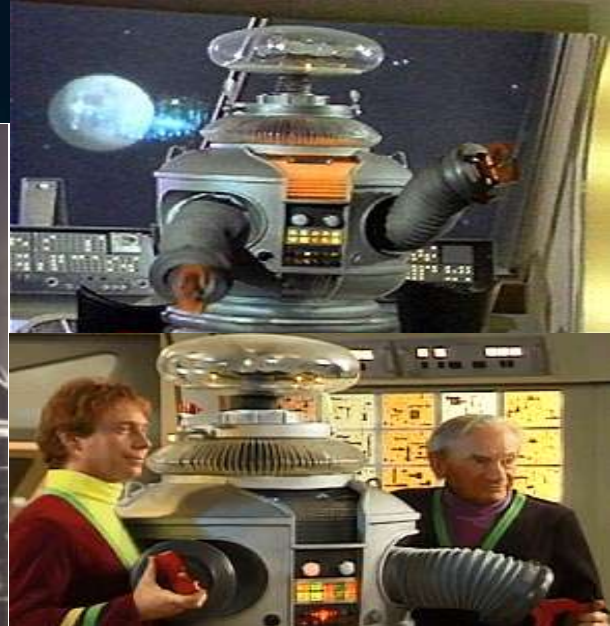
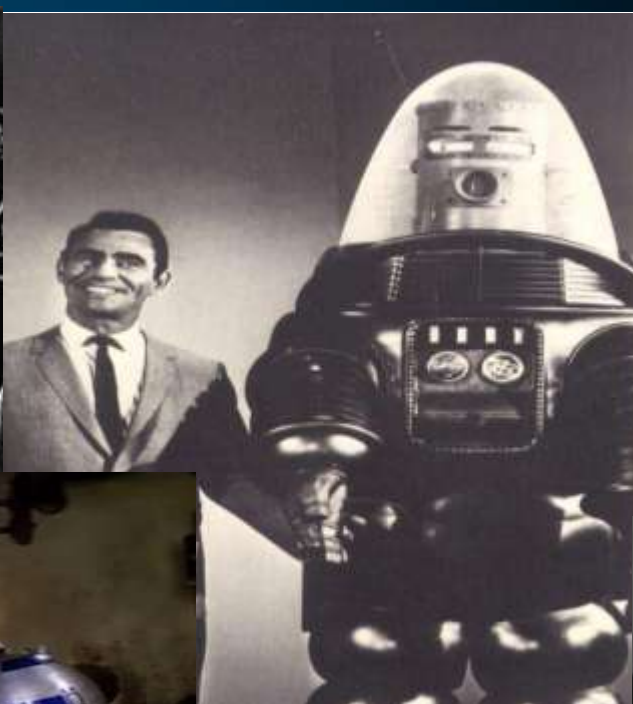
"You unlock this door with the key of imagination. Beyond it is another dimension - a dimension of sound, a dimension of sight, and substance, of things and ideas. You're moving into a land of both shadow and light. Welcome to the Twilight Zone."  
- Rod Serling



# E=MC<sup>2</sup>



# Robots We Grew up With...



# Robots Now... Closer to the Future





# Actual Robots Now... Beyond Science Fiction



# Living Doll



**These dolls are spying on your kids, consumer groups say**



By [Douglas Criss](#), CNN  
Published 3:46 PM EST, Thu December 9, 2016



"Privacy advocates try to keep 'creepy,' eavesdropping 'Hello Barbie' from hitting shelves," Sarah Halzack, [Washington Post](#)



# Mirror Image



# Mirror Image

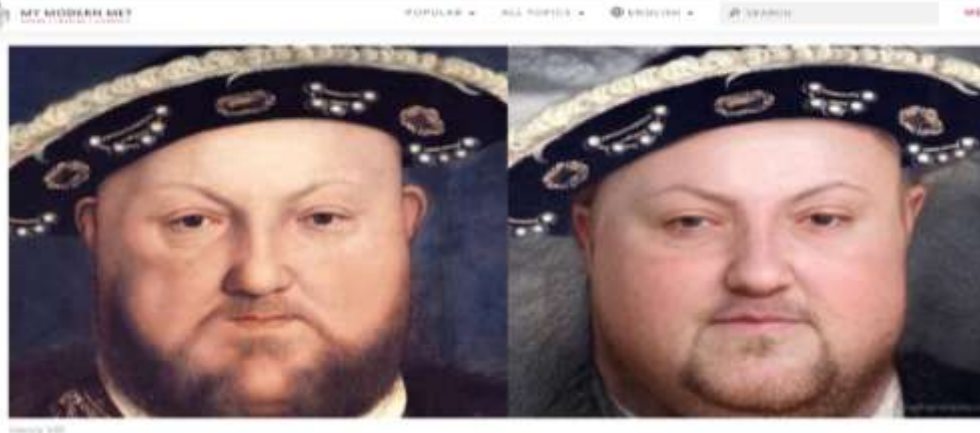
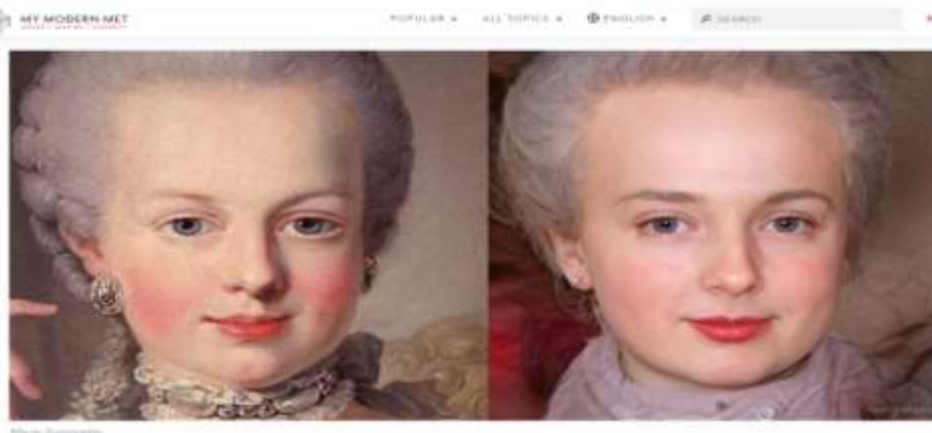


Prof. Hiroshi Ishiguro, right, inspects his robotic twin, Geminoid HI-1, also known as 'Gemmy,' who handles his teaching duties at Osaka University in Japan. Photo courtesy of ATR Intelligent Robotics and Communication Laboratories.

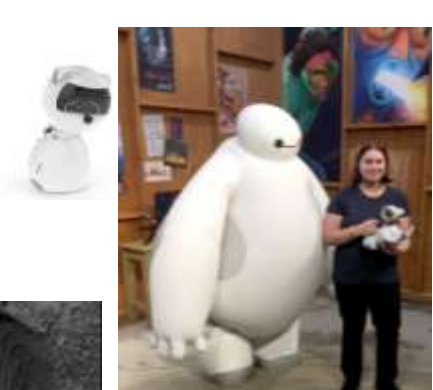




# Fun with Machine Learning! (@CitizenPlain)



# The Lonely (Robot Companionship)





# Leveraging – DB & Robotics!



# Pepper the Robot as Host at Sushi Restaurant (SynchroJapan)

## Leverage Robotics & Database



**Rich Niemiec** @RichNiemiec · Apr 3

The #Future is here! This #Robot at @synchrojapan in charge of reservations & seating at #sushi #restaurant in #Japan. Spend time on #Innovation, #BigData, #BigDataAnalytics, #IOT, #MachineLearning, #DeepLearning, #AI with the #database #oracledatabase

 **SynchroJapan** · Food & Restaurant Culture in Japan



**SynchroJapan** · Food & Restaurant Culture in Japan



# Use Oracle Virtual Assistant with Robots



**ORACLE**  
SER-10  
SERVICE

VIRTUAL ASSISTANT

Offer Natural, Conversational Interactions at Scale with a Smart Automated Solution

Oracle Virtual Assistant the Chatbot for Service

The banner features the Oracle logo and a grid of social media and communication icons including SMS, Facebook Messenger, WeChat, Kik, WhatsApp, Phone, Alexa, Slack, HipChat, and Siri.



Oracle Virtual Assistant the Chatbot for Service

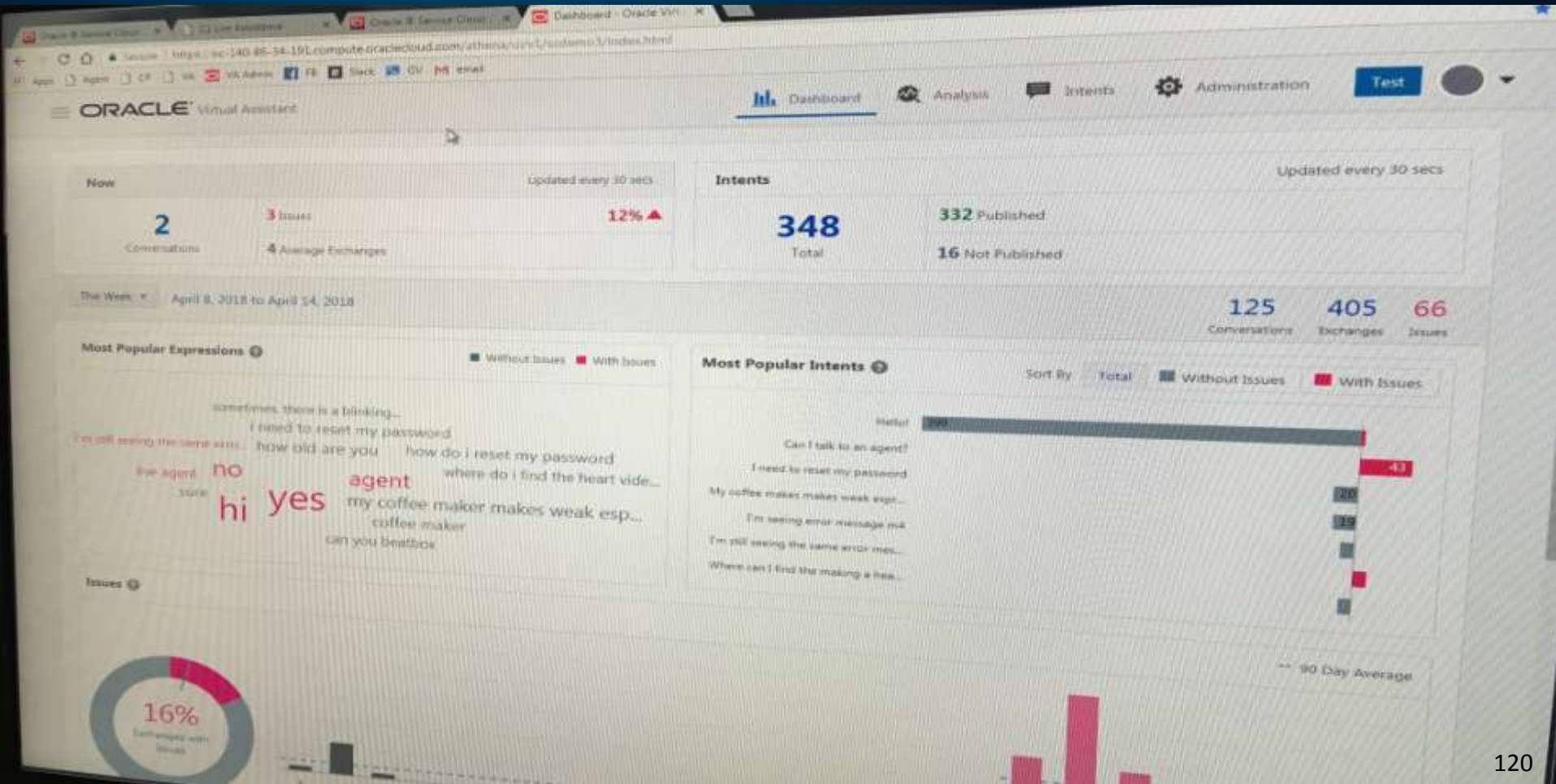
SMS, Facebook Messenger, weChat, Kik, WhatsApp, Phone, Alexa, Slack, HipChat, Siri

This section displays a grid of ten icons representing various communication channels: SMS, Facebook Messenger, WeChat, Kik, WhatsApp, Phone, Alexa, Slack, HipChat, and Siri.



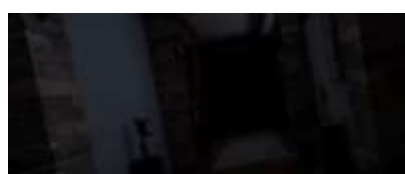


# Oracle Virtual Assistant Interface

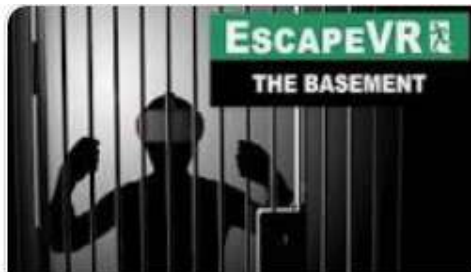




# Twilight Zone that's **not here yet...** fully



# Five Characters in Search of an Exit



# A World of Difference (*Getting Closer*)





# Leveraging – DB, AI & Virtual Reality!



## Virtual Reality



- > Immersion in virtual worlds
- > Total interaction with virtual
- > E.g. Oculus Rift

## Mixed Reality



- > Virtual World integrated to reality
- > Interaction between reality and virtual
- > E.g. Microsoft HoloLens

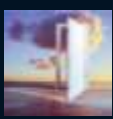
## Augmented Reality



- > Virtual on top of reality
- > Limited interaction with the virtual
- > E.g. Smartphones & tablets



# Elegy (*Build Archive to Communicate with*)



# Elegy (*Future?*)

## Funeral parlour looking for 'volunteers' for 'Black Mirror' project to create immortal AI

SCIENTISTS are looking for volunteers for a macabre experiment which will allow deceased family and friends to live forever as robots in an eerie scenario similar to Netflix's Black Mirror, it has been revealed.

By **MATT DRAKE**

23:43, Sat, Feb 24, 2018 | UPDATED: 01:53, Sun, Feb 25, 2018



Lifestyle / Family & Relationships

## 'It's heartbreaking': how coronavirus pandemic is changing funerals, grieving and ways we say goodbye to loved ones



## Pepper Now Available at Funerals as a More Affordable Alternative to Human Priests > In Japan, thrifty Buddhists can now hire the Pepper robot as a funeral assistant



# In His Image (Available Now - 100K Waiting List)



**ETERQ**


**The most daring  
experience ever made  
on the Internet.**

Let your counterpart work for you. The very  
first-ever intelligent virtual self of you.

**DOUBLE YOURSELF**


Currently, we have over 100,000 registered users who are on the waiting list for ETERQ.  
Available exclusively for a part of the remarkable digital trail.

**CBS Aol. BBC The Telegraph theguardian Sun**




**Artificial Intelligence**

ETERQ is a social network that relies on Artificial Intelligence as a central element, and it's currently in the BETA stage. Even in your absence, the virtual beings will publish, comment and interact with you intelligently.



**Counterpart**

The Counterpart is your Virtual Self that will stay in the system and interact with the world just like you would if you were present. Your Counterpart will learn more with each action you take. The more you interact in the new social network, the more your Counterpart will learn!



**Cyber Eternity**

Eternizing is a way of keeping your thoughts and posts for all time. Are you curious? Come meet your Counterpart and become eternal. Challenge the impossible.



AI

GUEST

# It's time for workers to worry about AI

GARY GROSSMAN, EDELMAN @GARYG02 APRIL 7, 2019 2:22 PM



## MOST READ



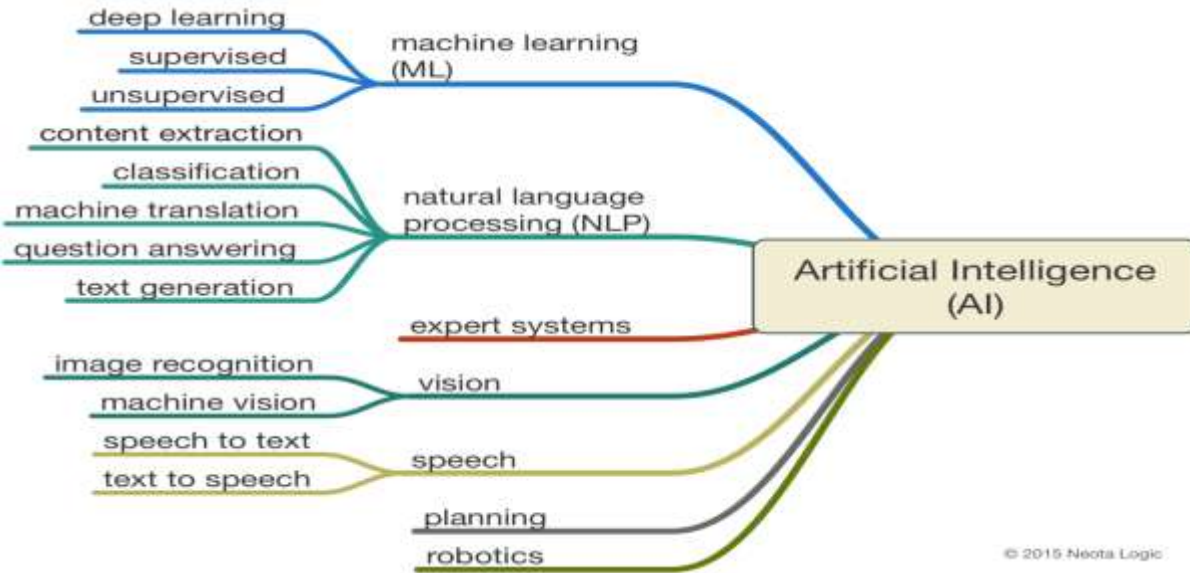
It's time for workers to worry about AI



Remote AR will make it so we can work —



# Future Tech is in AI: You can Leverage this NOW!



© 2015 Neota Logic

Oracle's AI Platform



**100%**  
of cloud applications will feature AI

**>50%**  
of enterprise data managed autonomously

**85%**  
of all interactions will be automated

# The After Hours (*Future Sentience Issues Ahead*)



- Is now a Citizen
- Wants to get degree
- Wants to have kids
  
- What is Alive?
- Sentience Issues?
- Robot's Feelings?
- Ethical Issues +++



# The Digital Transformation Ahead

Gerd

## Digital Transformation 2000 to 2050

A historically significant  
change in humanity...



**Disconnected**  
*The Dark Ages*

Using  
Digital

Wearing  
Digital

Implanting  
Digital

The Hive  
Mind



# The Future of Implant Technology



Check out my new Skateboard Video  
On YouTube!

**YouTube**

ift Images



# Connection to the Brain – Currently Working



# Walking Distance *(Create your past to view)*



## Developer Creates VR Time Machine To Relive His Past Year

July 13, 2020 - by Freddy Carrion



# A Nice Place to Visit (*Mystery Solve / Mean Afterlife VRs*)





# Leveraging – DB & Robotics!



Blog: Science

## Mini Insect Robocops: Engineers, Government Work on RC Insects

Jason Mick (Blog) - October 9, 2007 10:02 AM

### *The Fifth Trumpet*

...<sup>9</sup>They also had thoraxes like breastplates of iron, and the sound of their wings was like the roar of many horses and chariots rushing into battle. <sup>10</sup>They had tails with stingers like scorpions, which had the power to injure people for five months. <sup>11</sup>They were ruled by

Experimental Vaccine Daily Blog

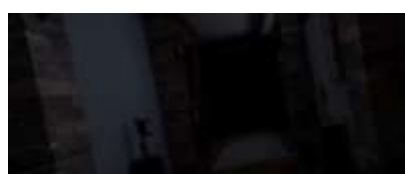




# Third Planet from the Sun (*Elon Musk*)



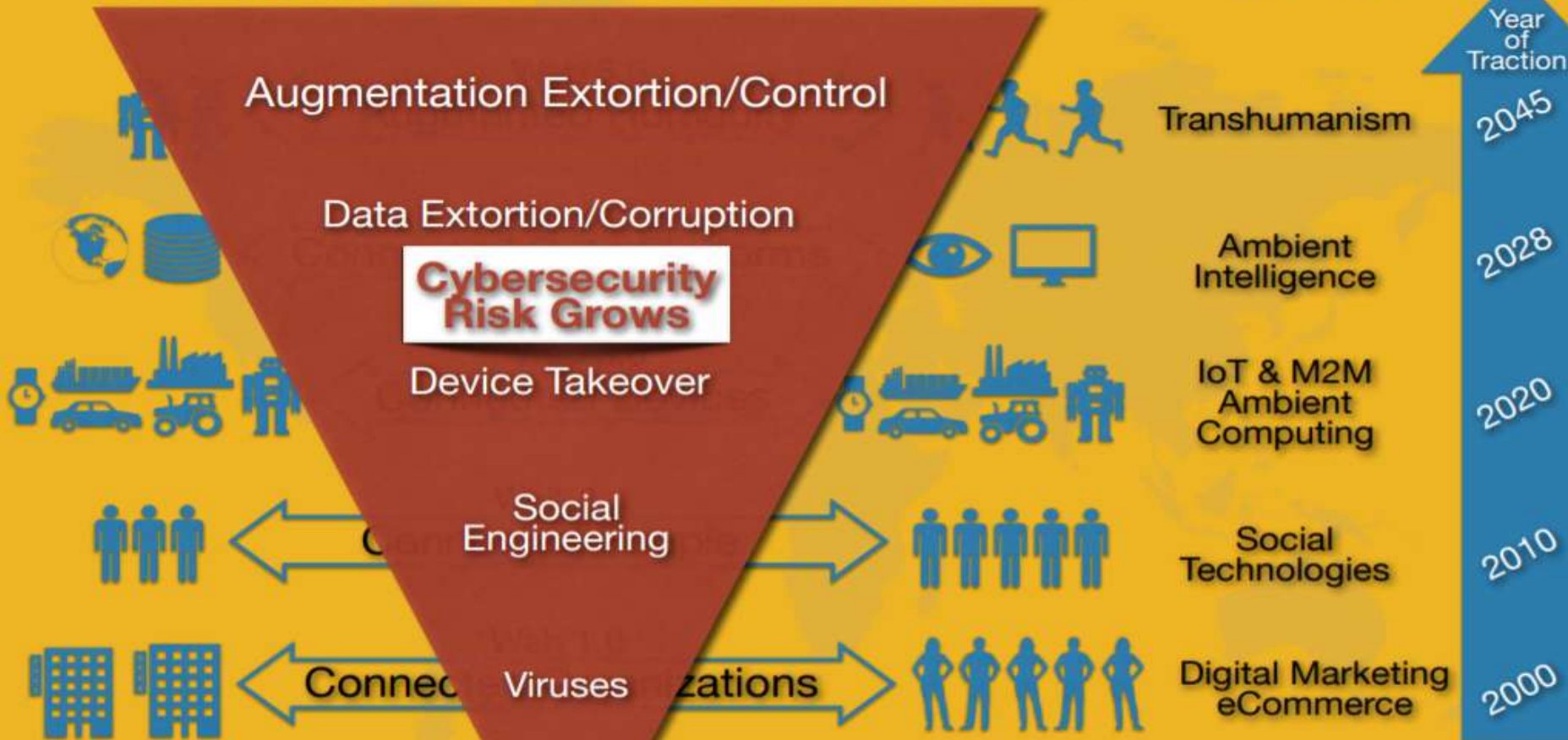
# Twilight Zone - Maybe you can build it...



# The Issues to Overcome moving Forward...

## The Transformation of Connection

FPOV





# Digital – How did we go from Magical to Toxic?



MAGIC



MANIC



TOXIC



Géral



Géral



# Black Mirror - Dystopian

- Nosedive: (Social Meeting Addiction & Ratings)



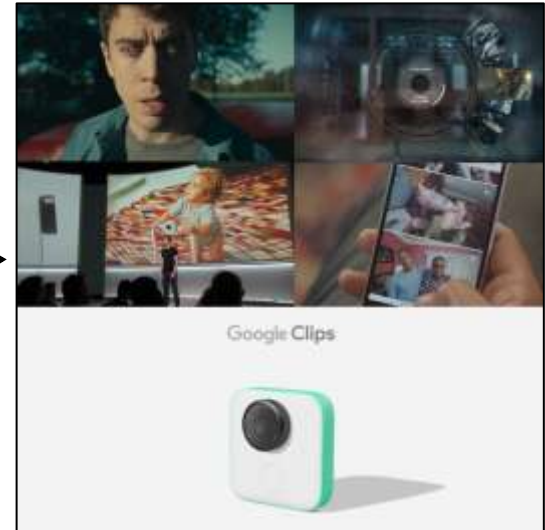
- Hated in the Nation (Cancel Culture)



- Be Right Back (ETER9)



- The Entire History of You (Kapture / Google Glass / Google Clips)



- Metalhead (Robot Dog)



# A Thing About Machines - People Frustrated



# Oracle Database Security

Built over MANY years...

1977

Database Auditing  
Government customer

Native Network Encryption (Oracle7)

Strong authentication (PKI, Kerberos, RADIUS)

Database Encryption API

Virtual Private Database (8i)

Global roles

Enterprise User Security

Proxy authentication

Oracle Label Security (2000)

Client Identifier / Identity propagation

Secure application roles

Fine Grained Auditing (9i)

EM Configuration Scanning

Transparent Data Encryption

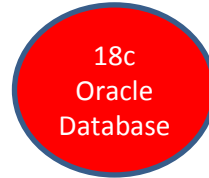
DB Security Evaluation #19

Oracle Database Vault

Oracle Audit Vault

Encryption in the Silicon (M7)

Oracle Multi-Tenant Security (PDBs)



18c: User-defined Master Encryption Key (bring own key - keystores)

19c: Oracle DB supplied schema-only accounts have passwords removed (not sample accounts)

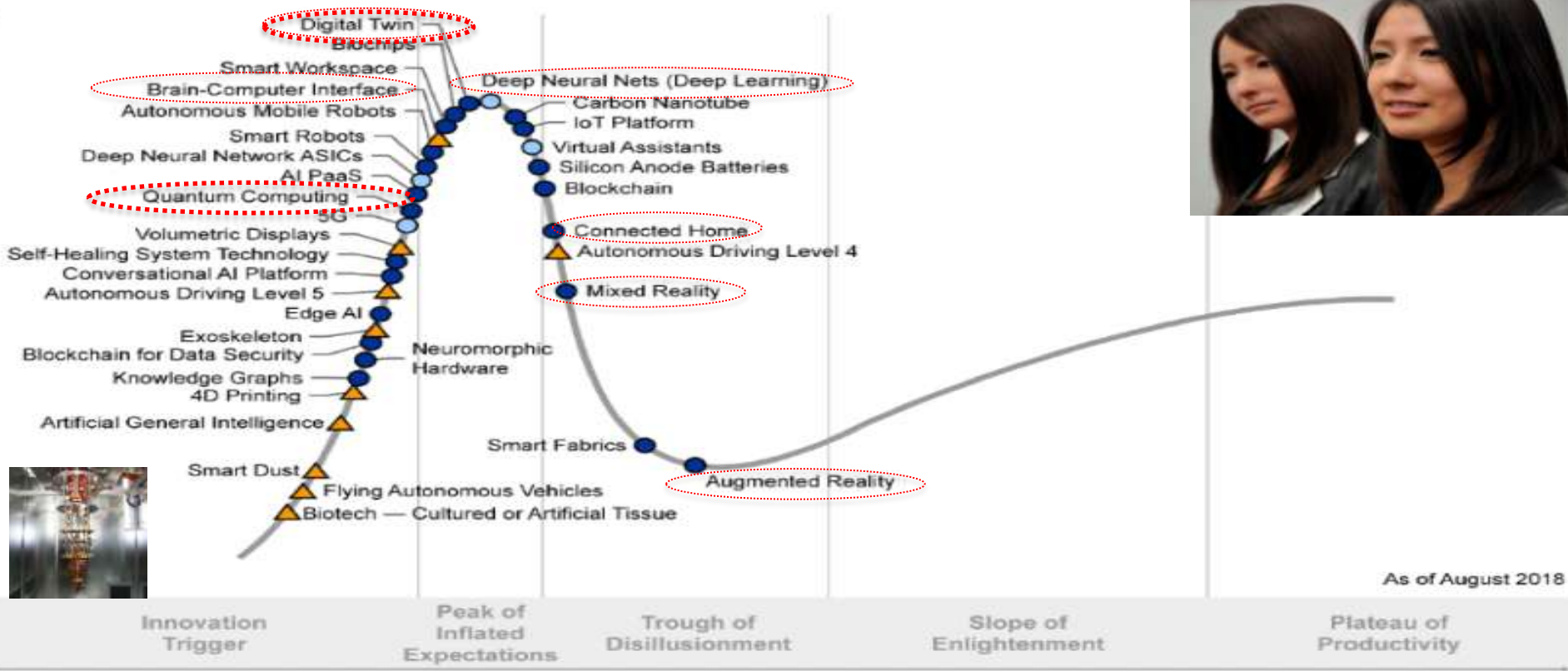
21c: Global Fault-Tolerant Key Vault & Oracle Data Safe

2022 +



# Gartner Hype Cycle August 2018

## All about Tech Creating a New Reality



Plateau will be reached:

- less than 2 years
- 2 to 5 years
- 5 to 10 years
- ▲ more than 10 years
- ⊗ obsolete before plateau



# Gartner 2020 hype cycle

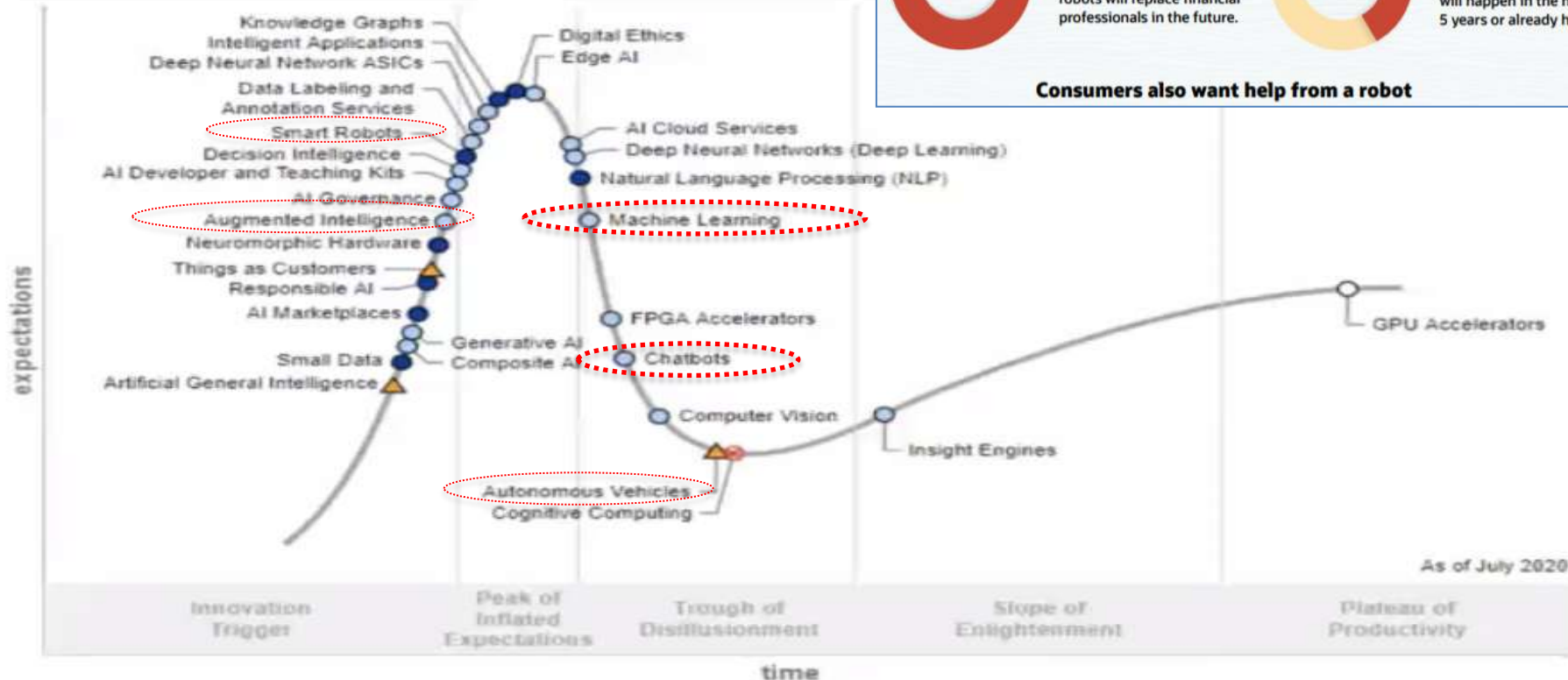


**82%**  
of consumers believe that robots will replace financial professionals in the future.



**42%**  
of consumers believe it will happen in the next 5 years or already has.

**Consumers also want help from a robot**



Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau

# What's comes *after* the Exadata Zone?

YOU will soon be in for more...

	<u>Directly Addressable</u>	<u>Indirect/Extended</u>
<u>4 Bit:</u>	16	(640)
<u>8 Bit:</u>	256	(65,536)
<u>16 Bit:</u>	65,536	(1,048,576)
<u>32 Bit:</u>	4,294,967,296	
<u>64 Bit:</u>	18,446,744,073,709,551,616	
<u>128 Bit:</u>	3.4 x 10e+38... quantum leap!	



- Qubits allow multiple states so that you can look at all of the possibilities/probabilities at one time (IBM Q is 20 qubits).
- The “**Quantum Zone**” next (Quantum Physics is incomplete – Einstein)
  - Just 512 qubits would store 512-bits of addressable memory or  $2^{512}$  (which is well over a googol or 1 with 100 zero's after it – **a googol is about  $2^{332}$** ).
  - Brush up on your Eigenvectors, Eigenvalues, Pauli Matrices & Grover's Algorithm
  - Create Singularity ... all atoms of a person by 2045 (I think earlier); *12-Monkeys*
  - Private universes – Is there one for each person? (Schroeder's cat – I think not)
  - Rearranging atoms to create new objects; Nanotech + Quantum Physics coming!

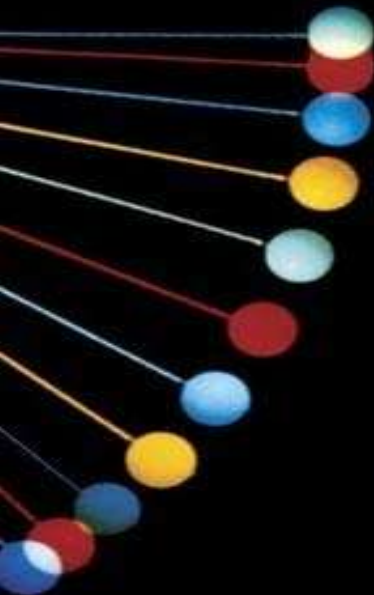
# The Age of Entanglement!



"A masterful account of the phenomenon Einstein thought so crazy it could not possibly be true. (Only it is!)"  
—DR. MARCUS CHOWN, author of *The Universe Next Door*

## THE GOD EFFECT

QUANTUM ENTANGLEMENT,  
SCIENCE'S STRANGEST PHENOMENON

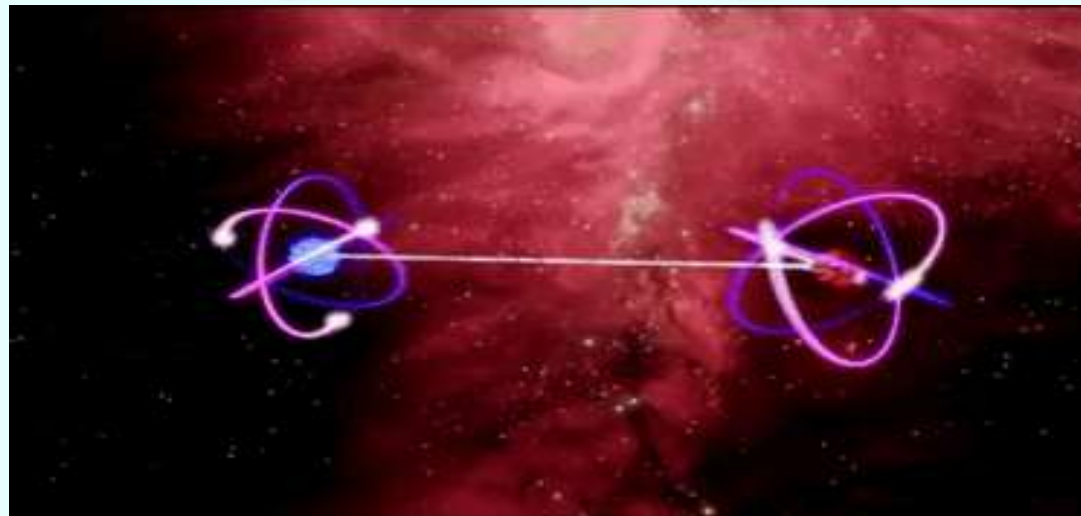


TIME TRAVEL,  
TELEPORTATION, AND  
THE ULTIMATE COMPUTER

BRIAN CLEGG

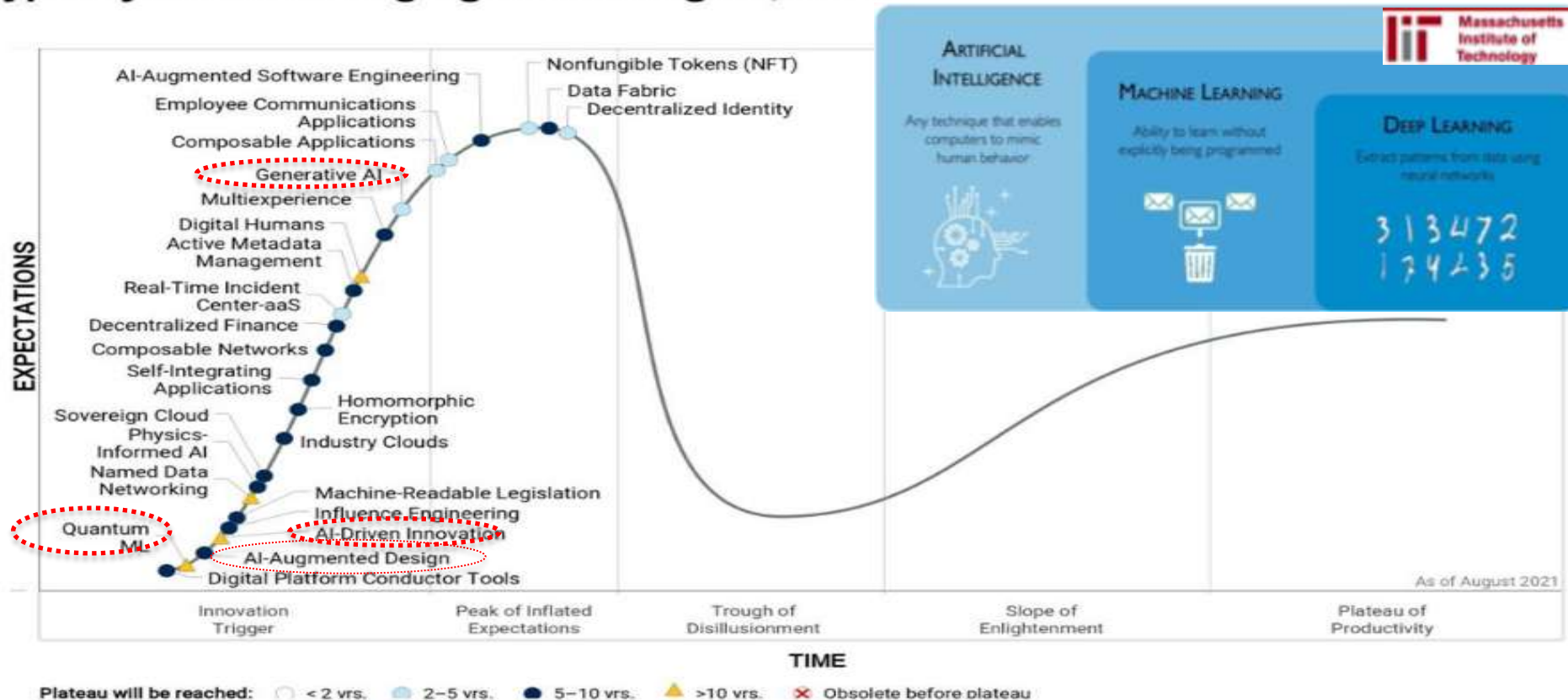
We are now shifting from  
the Information Age to:

**The Age of Entanglement**



# Gartner 2021 hype cycle for Emerging Tech

## Hype Cycle for Emerging Technologies, 2021





# 3 Types of Artificial Intelligence

## Artificial Narrow Intelligence (ANI)



Stage-1

### Machine Learning

- ▶ Specialises in one area and solves one problem



Siri



Alexa



Cortana

## Artificial General Intelligence (AGI)



Stage-2

### Machine Intelligence

- ▶ Refers to a computer that is as smart as a human across the board

## Artificial Super Intelligence (ASI)



Stage-3

### Machine Consciousness

- ▶ An intellect that is much smarter than the best human brains in practically every field

# Final Thoughts... world changing fast!

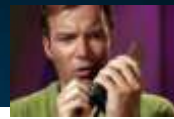
*“Those who use things of the world should not become attached to them. For the world in its present form is passing away.”*

1 Corinthians 7:31



# Star Trek

- Communicator - Motorola Flip Phone
- Phaser - EPM or Stun Gun / Taser
- Tablets (Medical) - Tablet Computers
- Tricorders - Many Medical Devices (below)
- Translators - Google Translate (others)
- Tractor Beam - MIT has it in concept
- Telepresence - Zoom Virtual Conferences
- Geordi's Visor - Robotic Eyes / Implants
- Communicator Badges - Many Security Badges
- Food Replicator - 3D Printer
- Holodeck - VR
- Teleportation - Quantum Entanglement
- Big Screen TV - Everyone has it
- Hands Free Phone - Bluetooth headset / Airpods



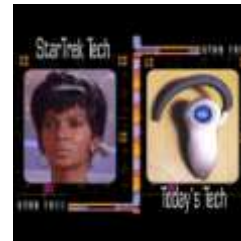
Hololens VR Apps Above



Microsoft working on Holodeck



MouthLab

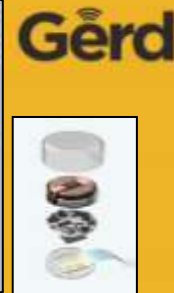
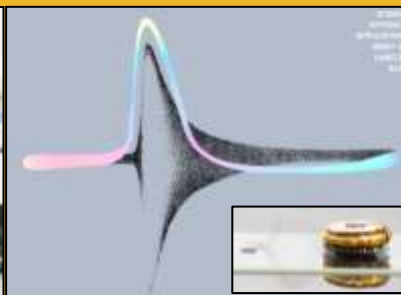




# The Digital Transformation Ahead

## Digital Transformation 2000 to 2050

A historically significant  
change in humanity...



**Disconnected**  
*The Dark Ages*

Using  
Digital

Wearing  
Digital

Implanting  
Digital

The Hive  
Mind



# Summary – Goal is Apply Technology & W.I.N.

- ❑ You've Just Entered the Twilight Zone (20+)
- ❑ Twilight Zone **We See Now**
- ❑ Big Data and IOT - Data Is Coming Fast
- ❑ **ML & Oracle - Overview & ADB**
- ❑ Applications of ML Algorithms & AutoML
- ❑ Machine Learning Future, Robots & VR
- ❑ Future Tech from the Twilight Zone+



Quantum Entanglement

"Spooky action-at-a-distance"  
(A. Einstein)



# Final Thoughts... Catch your Ride!



*“Things may come to those who wait, but only the things left by those who hustle.”*

— Abraham Lincoln



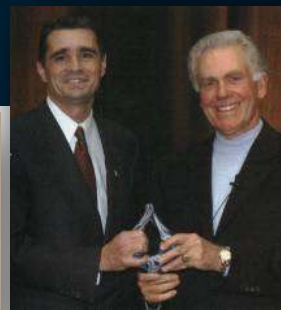


The  
TWILIGHT  
ZONE

# Oracle never caught from behind

## Oracle's 45<sup>th</sup> Anniversary in 2022

- Great Sales/Marketing
- Great Database
- Applications Leader
- BI Leader
- In the lead except Cloud
- **GAME OVER**
- Hardware/Software Engineering!
- Have Everything to Win in Cloud + AI!



*In Memory of Ken Jacobs, Dr. DBA & Joel Kallman*



# 12c R2 Book – Available Now!

DBMS			Database Model				
Jan 2022	Dec 2021	Jan 2021			Jan 2022	Dec 2021	Jan 2021
1.	1.	1.	Oracle	Relational, Multi-model	1266.89	-14.85	-56.03
2.	2.	2.	MySQL	Relational, Multi-model	1206.05	+0.01	-46.01
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	944.81	-9.21	-96.42
4.	4.	4.	PostgreSQL	Relational, Multi-model	606.56	-1.86	+54.33
5.	5.	3.	MongoDB	Document, Multi-model	488.57	+3.89	+31.34
6.	6.	7.	Redis	Key-value, Multi-model	177.98	+4.44	+22.97
7.	7.	6.	IBM Db2	Relational, Multi-model	164.20	-2.98	+7.03
8.	8.	8.	Elasticsearch	Search engine, Multi-model	160.75	+3.03	+9.50
9.	10.	11.	Microsoft Access	Relational	128.95	+2.96	+13.61
10.	9.	9.	SQLite	Relational	127.43	-1.25	+5.54
11.	11.	10.	Cassandra	Wide column	123.55	+4.35	+5.47
12.	12.	12.	MariaDB	Relational, Multi-model	106.42	+2.06	+12.63
13.	13.	13.	Splunk	Search engine	90.45	-3.87	+2.79
14.	14.	15.	Microsoft Azure SQL Database	Relational, Multi-model	86.32	+3.07	+14.96
15.	15.	16.	Hive	Relational	83.45	+1.52	+13.02
16.	16.	17.	Amazon DynamoDB	Multi-model	79.85	+2.23	+10.72
17.	17.	37.	Snowflake	Relational	76.82	+5.79	+61.30
18.	18.	14.	Teradata	Relational, Multi-model	69.13	-1.17	-3.46
19.	20.	20.	Solr	Search engine, Multi-model	58.53	+0.80	+6.04
20.	19.	19.	Neo4J	Graph	58.03	0.00	+4.25
21.	21.	21.	SAP HANA	Relational, Multi-model	56.92	+2.34	+6.05
22.	22.	22.	FileMaker	Relational	55.86	+1.99	+8.47
23.	23.	18.	SAP Adaptive Server	Relational, Multi-model	51.05	-0.33	-3.56
24.	24.	24.	Google BigQuery	Relational	45.62	-0.18	+9.62
25.	25.	23.	HBase	Wide column	43.99	-1.55	-2.29
26.	26.	25.	Microsoft Azure Cosmos DB	Multi-model	40.04	+0.33	+7.07
27.	27.		PostGIS	Spatial DBMS, Multi-model	31.87	-0.57	
28.	29.	27.	InfluxDB	Time Series, Multi-model	30.09	+1.70	+3.77
29.	28.	26.	Couchbase	Document, Multi-model	28.86	+0.42	-2.77
30.	30.	30.	Firebird	Relational	27.28	-0.31	+4.52
31.	32.	29.	Amazon Redshift	Relational	25.85	+1.48	+2.93

## Top New Release

amazon  
Try Prime

All ▾ niemiec

Departments ▾ Your Amazon.co

Books Advanced Search New Releases

Books > Computers & Technology > Databases & Big

### Oracle Database 12c

by Richard Niemiec (Author)

**#1 New Release** in Oracle Databases

**Oracle Database 12c**  
**Release 2 Performance Tuning Tips and Techniques**  
Best Practices for Optimizing Database Performance

# Join Thousands Of Happy Customers And End The Frustration In Tuning Oracle Databases

Our services give you the training you need so your Oracle database runs faster and more efficiently. And, you get the credit!



MEMBERSHIPS



TRAINING



FREE TOOLS



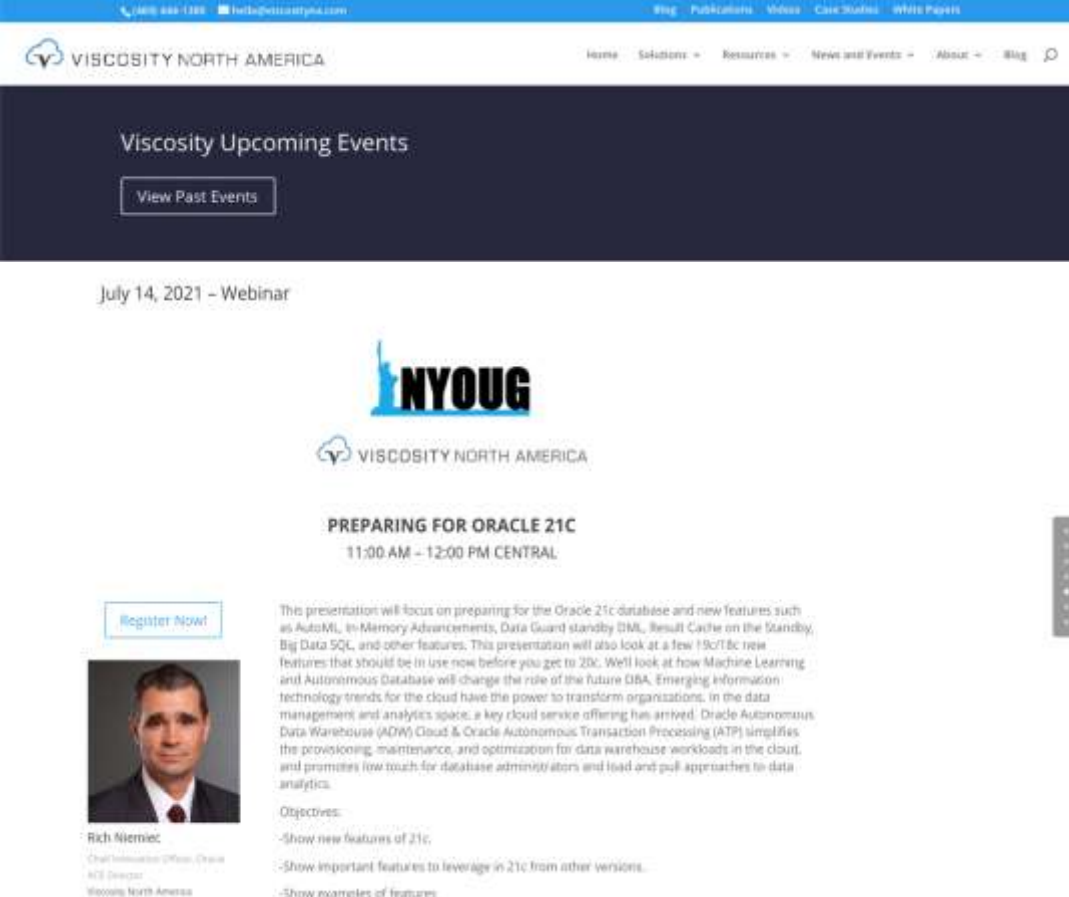
BOOKS

# Viscosity Speaking Events

Register for upcoming webinars:  
[viscosityna.com/event](https://viscosityna.com/event)

Follow us on  
[LinkedIn](#), [Facebook](#) or  
[Twitter](#): **@viscosityna**

Email me at  
[katie.barnes@viscosityna.com](mailto:katie.barnes@viscosityna.com)

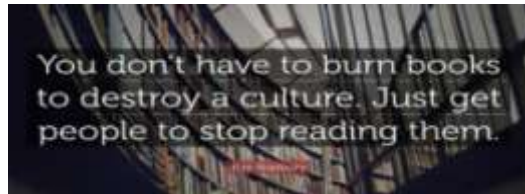


The screenshot shows the Viscosity North America website. At the top, there is a navigation bar with links for Home, Solutions, Resources, News and Events, About, and Blog. Below the navigation bar, the main heading reads "Viscosity Upcoming Events" with a "View Past Events" button. The featured event is titled "July 14, 2021 - Webinar" and is presented by NYOUG (New York Oracle User Group). The event is "PREPARING FOR ORACLE 21C" scheduled for "11:00 AM - 12:00 PM CENTRAL". A "Register Now" button is visible. To the left of the event description is a portrait of Rich Niemiec, identified as Chief Innovation Officer, Oracle ACE Director, and Viscosity North America. The event description states: "This presentation will focus on preparing for the Oracle 21c database and new features such as AutoML, In-Memory Advancements, Data Guard standby DML, Result Cache on the Standby, Big Data SQL, and other features. This presentation will also look at a few 19c/18c new features that should be in use now before you get to 20c. We'll look at how Machine Learning and Autonomous Database will change the role of the future DBA. Emerging information technology trends for the cloud have the power to transform organizations. In the data management and analytics space, a key cloud service offering has arrived. Oracle Autonomous Data Warehouse (ADW) Cloud & Oracle Autonomous Transaction Processing (ATP) simplifies the provisioning, maintenance, and optimization for data warehouse workloads in the cloud, and promises low touch for database administrators and load and pull approaches to data analytics." The objectives listed are: "-Show new features of 21c", "-Show important features to leverage in 21c from other versions.", and "-Show examples of features".

# References



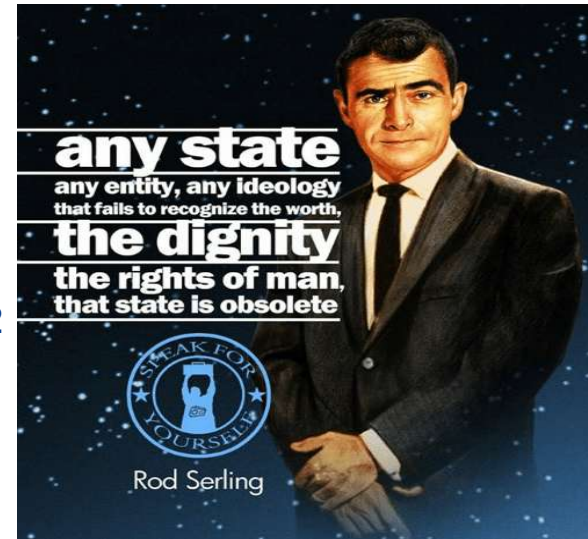
- The Emerging Technology Roadmap, Scott Klososky
- Futurist Gerd Leonhard The Futures Agency, Technology vs. Humanity, Gerd Leonhard, oracle.com & Juan Loaiza / Doug Hood presentations, amazon.com, smartcitiescouncil.com, youtube.com, business coach, libelium.com, monetate, en.wikipedia.org, Netflix, Black Mirror, hometoys.com, FPOV, huffingtonpost.com, theguardian.com, nationalgeographic.com, newscientist.com, enswmu.blogspot.com, dailymail.co.uk, FutureRobot, theguardian.com, thinkhealthwireless.blogspot.com, ge.com, cmswire.com, runningsupplement.co.uk, quickmeme.com, nike.com, thisiswhyimbroke.com, businessinsider.com, slideshare.com, forrester.com, spiceworks.com, mwaintel.com, humancapitalist.com, wired.com and Wired Magazine, shodanhq.com, developer.nokia.com, extremetech.com, Getty Images, Dr. Quantum, Leo Kouwenhoven, Telecom Tech News, Actimage & any other company products are the property of their respective companies.





# Copyright Information

- Neither the Viscosity nor the author guarantee this document to be error-free. Please provide comments/questions to [rich.niemiec@viscosityna.com](mailto:rich.niemiec@viscosityna.com) & [richniemiec@gmail.com](mailto:richniemiec@gmail.com). I'm always looking to improve!
- Rich Niemiec ©2022. This document cannot be reproduced without expressed written consent from Rich Niemiec, but may be reproduced or copied for presentation and conference use.
- References include Rich Niemiec's Exadata Presentation & Oracle 12cR2 Database Performance Tuning Tips & Techniques book, [www.oracle.com](http://www.oracle.com), [en.wikipedia.org](http://en.wikipedia.org), [slashgear.com](http://slashgear.com), [gifsoup.com](http://gifsoup.com), [www.amazon.com](http://www.amazon.com), Tech Crunch, [www.rolta.com](http://www.rolta.com), The Twilight Zone, Information Week, Gartner, Computerworld, Quest, Data and Technology Today, Forbes, Quest, Orbit, Computer Weekly, Redmond, Database Trends & Applications, dsp, Dataversity, zdnet, DBVisit, Steve Jones, Kerry Osborne, Julian Dontcheff, Accenture, Quora, Brent Ozar, & Oracle OpenWorld



**THERE IS A FIFTH DIMENSION  
BEYOND THAT WHICH IS KNOWN TO  
MAN. IT IS A DIMENSION AS VAST  
AS SPACE AND AS TIMELESS AS INFINITY.**

# Getting Started—Oracle ML/AI Resources

ORACLE

## Oracle Advanced Analytics Overview Information

- [Oracle Machine Learning Newest Features and Road Map.pptx](#) presentation
- Blog post: [Simple Guide to Oracle's Machine Learning and Advanced Analytics](#)
- Oracle Advanced Analytics Public [Customer References](#)
- [Oracle's Machine Learning and Advanced Analytics Data Management Platforms](#) white paper on OTN



## YouTube recorded Oracle Advanced Analytics Presentations and Demos, White Papers

- [Oracle's Machine Learning & Advanced Analytics 12.2 & Oracle Data Miner 4.2 New Features YouTube video](#)
- [Library of YouTube Movies on Oracle Advanced Analytics, Data Mining, Machine Learning \(7+ "live" Demos e.g. Oracle Data Miner 4.0 New Features, Retail, Fraud, Loyalty, Overview, etc.\) Overview YouTube video of Oracle's Advanced Analytics and Machine Learning](#)

ORACLE  
UNIVERSITY

## Getting Started/Training/Tutorials

- [OAA/Oracle Data Miner Workflow GUI Online \(free\) Tutorial Series](#) on OTN
- [OAA/Oracle R Enterprise \(free\) Tutorial Series](#) on OTN
- [Try the Oracle Cloud Now!](#)
- [Getting Started w/ ODM blog entry](#)
- [New OAA/Oracle Data Mining 2-Day Instructor Led Oracle University course](#)
- [Oracle Data Mining Sample Code Examples](#)

ORACLE Help Center

## Additional Resources, Documentation & OTN Discussion Forum

- [Oracle Advanced Analytics Option on OTN page](#)
- [OAA/Oracle Data Mining on OTN page, ODM Documentation](#)
- [OAA/Oracle R Enterprise page on OTN page, ORE Documentation](#)
- [Oracle SQL based Basic Statistical functions on OTN](#)
- [Oracle R Advanced Analytics for Hadoop \(ORAAH\) on OTN](#)



**[Analytics and Data Summit](#)** , All Analytics, All Data, No Nonsense.

March 12-14, 2019, Redwood Shores, CA

# Quick FREE notes

Send email to (for slides):  
[hello@viscosityna.com](mailto:hello@viscosityna.com)

(richniemiec@gmail.com)

@richniemiec - twitter



*richniemiec@gmail.com*

*viscosityna.com*

*@richniemiec*

163