Production ML with Autonomous Data Warehouse

Harry Snart
harry.snart@oracle.com, linkedin.com/in/harrysnart
Future and past TechCast:

2019 TechCast Archive

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Speaker(s)</th>
<th>Video Replay</th>
<th>Slides</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 6, 2019</td>
<td>Extending Oracle Analytics Cloud’s Data Visualization Using Custom Plug-ins</td>
<td>Wayne Van Sluys</td>
<td>Video</td>
<td>Slides</td>
</tr>
<tr>
<td>July 11, 2019</td>
<td>Developing an RDF4J Adapter for Oracle Spatial and Graph – SPARQL focus</td>
<td>Timea Turdean, Software Engineer, Semantic Web Company</td>
<td>Video</td>
<td>Slides</td>
</tr>
</tbody>
</table>

Submit a topic to share at https://analyticsanddatasummit.org/techcasts/
Helpful Links

ORACLE AUTONOMOUS CLOUD
https://cloud.oracle.com/tryit

ORACLE AUTONOMOUS HANDS ON LAB FOR DEVELOPERS
https://go.oracle.com/e/f2?LP=82486

ORACLE MACHINE LEARNING ON OTN
https://www.oracle.com/technetwork/database/options/oml/overview/index.html

OML TUTORIALS
Credit_Scoring_100K Targeting Top Customers: https://oracle.github.io/learning-library/workshops/adwc4dev/?version=Self-Guided&page=SG-intro.md&elqTrackId=e57daae9db8d44bfac4a9e6614175e5a&elqaid=82487&elqat=2&source=%3Aow%3Alp%3Acpo%3A

ORACLE ANALYTICS CLOUD
Analytics and Data Summit

All Analytics. All Data. No Nonsense.

February 25-27, 2020

Oracle campus in Santa Clara, CA

Formerly called the BIWA Summit with the Spatial and Graph Summit
Same great technical content...new name!

www.AnalyticsandDataSummit.org
Production Machine Learning with the Autonomous Data Warehouse

Harry Snart
Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

The development, release, timing, and pricing of any features or functionality described for Oracle’s products may change and remains at the sole discretion of Oracle Corporation.
What is Machine Learning?
Any technique which enables computers to mimic human behavior.

Artificial Intelligence (AI)
- **1950’s**: Early conceptual work
- **1960’s**: Early applications
- **1970’s**: Further advancements
- **1980’s**: Expansion of research
- **1990’s**: Commercialization
- **2000’s**: Rapid growth
- **2010s**: Impact on society

**Machine Learning (ML)**
- AI techniques that give computers the ability to learn without being explicitly programmed to do so.

**Deep Learning**
- A subset of ML which make the computation of multi-layer neural networks feasible.
Types of Machine Learning Algorithms

Regression

Classification

Unsupervised

Supervised
About today’s demo

- Using binary classification models to classify labels in a dataset
- Data from UCI ML repository: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- Has been used in the past to run Kaggle competitions
- Dataset is cell attributes of breast tissue where a tumour has been found
- We use a historic dataset to predict whether a given tumour is benign or malignant
Goals of Demo

- Use Data Visualization to do Exploratory Data Analysis (EDA)
  - Profile data
  - Identify any inputs we might want to exclude if model is overfitting
  - Identify shape of training dataset
  - Rapid prototype a model and evaluate using DVML
- Use the Autonomous Data Warehouse to Productionise our Model
  - Load data into ADW
  - Use the Oracle Machine Learning Notebooks to build an in-database model
  - Expose model as a set of REST APIs
  - Document our APIs for Application Developers to use in their Apps
Why is it difficult to drop Machine Learning Models into Production?
61%

Organizations have Data Science initiatives
87%

Data Science Projects Fail
Data Scientist != Computer Scientist
Machine Learning Models

**Supervised Learning**
- Regression
- Classification
- Image Recognition
- Object Classification
- Deep Learning

**Unsupervised Learning**
- Sentiment Analysis
- Neural Networks
- Dimensionality Reduction
- Sound Recognition
- Language Processing
- Clustering
- Anomaly Detection
- Time Series
Example - Openstack

- Flask
- Flask
- Flask

NGINX
OpenSSL

Scikit Learn + Python
What Makes Oracle Data Science Solutions Different

**ALL Data Management**
- Integrated Data Management Platform with big data / data lake capabilities
- Self-driving, self-tuning, self-recovering data management and administration

**Data Science Capabilities**
- Parallel, scalable Machine Learning algorithms
- In-database processing -> move the algorithm to the data
- Notebook and GUI development interfaces
- Collaborative development environments
- Integrated ML Capabilities (Graph, Text, Spatial...)
- Enriched with the best of Open Source

**High Performance Cloud Infrastructure**
- Integrated easy to use platform that minimizes the difficulties of development and adoption
- Focus on solving business problems by maximizing end-user efficiency
Oracle Machine Learning Algorithms and Analytics

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

Regression
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear Regression
- Neural Network
- Lasso

Clustering
- Hierarchical K-means
- Hierarchical O-cluster
- Expectation Maximization (EM)

Anomaly Detection
- One-class SVM

Time Series
- Forecasting - Exponential Smoothing
- Includes Popular Models E.G. Holt-winters With Trends, Seasonality, Irregularity, Missing Data

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
- 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 and Python Packages
- Third-party R And Python Packages Through Embedded Execution
- Spark Mllib Algorithm Integration
The Machine Learning Process

1. **Business Understanding**
   - Objectives
   - Success Criteria

2. **Data Understanding**
   - Capture Data
   - Describe Data
   - Explore Data

3. **Data Preparation**
   - Select Data
   - Clean Data
   - Construct Data
   - Combine Data

4. **Modeling**
   - Select Modeling Technique
   - Build Model
   - Assess Model

5. **Evaluation**
   - Evaluate Result
   - Decide Next Step

6. **Deployment**
The Process – Rapid Prototyping Using OAC

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

- Visualization
- Explain
- Data Prep
- Data Flow
- Model Details
- Visualization
Automated data preparations
Applied before training and applying a model

- **Fill NA**
  - Drop columns with NULL percentage higher than Maximum Null Value Percentage
  - Replace NA values
    - Categorical using most or least frequent value
    - Numeric using mean, max, min or median value
  - Same for all columns of the same type

- **Sanitize**
  - Remove categorical (text) columns with high number of unique values (> 75%)
  - Remove numerical columns that have a uniform distribution
  - Remove highly correlated attributes (above 0.75)

- **Feature Encoding**
  - Convert string values to numeric values
  - Two ways
    - OneHot
    - Indexer

- **Standardize and clean**
  - Optional for some algorithms
  - Standardization of Numeric columns
    - Remove mean and scale to unit variance
    - Purpose to normalize the range of values
Drop into Production

Apps

REST

Oracle Application Express
<table>
<thead>
<tr>
<th>Data Element</th>
<th>Data Type</th>
<th>Treat As</th>
<th>Aggregation</th>
<th>Sample Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Number</td>
<td>Attribute</td>
<td>None</td>
<td>917092, 90041602, 8735</td>
</tr>
<tr>
<td>DIAGNOSIS</td>
<td>Text</td>
<td>Attribute</td>
<td>None</td>
<td>B; M</td>
</tr>
<tr>
<td>RADIUS_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>12.34, 11.71; 12.46; 11.80</td>
</tr>
<tr>
<td>TEXTURE_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>19.83; 16.84; 20.52; 18.2</td>
</tr>
<tr>
<td>PERIMETER_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>87.76; 82.61; 134.70; 87</td>
</tr>
<tr>
<td>AREA_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>512.2; 1264.0; 477.3; 57</td>
</tr>
<tr>
<td>SMOOTHNESS_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.10070; 0.10750; 0.1150</td>
</tr>
<tr>
<td>COMPACTNESS_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.12060; 0.11470; 0.1289</td>
</tr>
<tr>
<td>CONCAVITY_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.00000; 0.12040; 0.2417</td>
</tr>
<tr>
<td>CONCAVE_POINTS_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.00000; 0.02864; 0.0161</td>
</tr>
<tr>
<td>SYMMETRY_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.1717; 0.1714; 0.1601; 0</td>
</tr>
<tr>
<td>FRACTAL_DIMENSION_NUCA</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.05667; 0.06113; 0.0591</td>
</tr>
<tr>
<td>RADIUS_NUCB</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>0.2860; 0.2204; 0.2315; 0.1</td>
</tr>
<tr>
<td>TEXTURE_NUCB</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>1.1500; 1.2800; 0.8561; 1</td>
</tr>
<tr>
<td>PERIMETER_NUCB</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>1.778; 2.225; 1.667; 1.59</td>
</tr>
<tr>
<td>AREA_NUCB</td>
<td>Number</td>
<td>Measure</td>
<td>Sum</td>
<td>46.87; 46.92; 47.43; 47.39</td>
</tr>
</tbody>
</table>
Basic facts about DIAGNOSIS

DIAGNOSIS is a Text Attribute with 2 unique values across 569 rows (100%). The most common DIAGNOSIS is B (63%) and the least common is M (37%).
Explain DIAGNOSIS

Basic Facts about DIAGNOSIS
What are the values of DIAGNOSIS and how do they relate to each other?

Key Drivers of DIAGNOSIS
What elements in this data best explain the values of DIAGNOSIS?

Segments that Explain DIAGNOSIS
What hidden groups in the data can predict outcomes for DIAGNOSIS?

Anomalies of DIAGNOSIS
What groups in the data exhibit unexpected results for DIAGNOSIS?

About Explain pages
Unbalanced dataset, suggest oversampling of class M

COMPACTNESS_NUCB by DIAGNOSIS

CONCAVITY_NUCB by DIAGNOSIS

CONCAVE_POINTS_NUCB by DIAGNOSIS

RADIUS_NUCC by DIAGNOSIS

PERIMETER_NUCC by DIAGNOSIS

Some inputs look like poor predictors, suggest dropping from model build.
RADIUS_NUCA, TEXTURE_NUCA by DIAGNOSIS, ID

Diagnosis: B vs. M
- **RADIUS_NUCA**
  - Range: [10.00, 40.00]
- **TEXTURE_NUCA**
  - Range: [10.00, 40.00]
<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Owner</th>
<th>Modified</th>
<th>Refreshed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WDBC_MODEL_TRAIN</td>
<td>harry.snart@ora</td>
<td>Just now</td>
<td>2 minutes ago</td>
</tr>
<tr>
<td></td>
<td>WDBC_MODEL_VALIDATE</td>
<td>harry.snart@ora</td>
<td>Just now</td>
<td>Just now</td>
</tr>
<tr>
<td></td>
<td>WDBC_ALL</td>
<td>harry.snart@ora</td>
<td>31 minutes ago</td>
<td>36 minutes ago</td>
</tr>
</tbody>
</table>
Select Train Two-Classification Model Script

- Neural Network for Classification
- Random Forest for model training
- Logistic Regression for model training
- Naive Bayes for Classification
- SVM for Classification
- CART for model training

[Search] [Grid] [Cancel] [OK]
Train Binary Classifier

Model Training Script: SVM for Classification

- Target: DIAGNOSIS
  - Target value to be learned/predicted
- Positive Class in Target: M
  - Positive class in the target value. Default is Yes.
- Categorical Column Imputation: Most Frequent
  - Method for categorical features with NA: Two options: mostFrequent and leastFrequent. Default is mostFrequent.
- Numerical Column Imputation: Mean

<table>
<thead>
<tr>
<th>ID</th>
<th>ab</th>
<th>DIAGNOSIS</th>
<th>RADIUS_NUCA</th>
<th>TEXTURE_NUCA</th>
<th>PERIMTER_NUCA</th>
<th>AREA_NUCA</th>
<th>SMOOTHNESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>842302</td>
<td>M</td>
<td>M</td>
<td>17.99</td>
<td>10.38</td>
<td>122.80</td>
<td>1,001</td>
<td>0.1184</td>
</tr>
<tr>
<td>842517</td>
<td>M</td>
<td>M</td>
<td>20.57</td>
<td>17.77</td>
<td>132.00</td>
<td>1,036</td>
<td>0.0847</td>
</tr>
</tbody>
</table>
### Select Columns

**No items to display.**

**Selected (32/32)**

- ID
- DIAGNOSIS
- RADIUS NUCA
- TEXTURE NUCA

<table>
<thead>
<tr>
<th>ID</th>
<th>DIAGNOSIS</th>
<th>RADIUS NUCA</th>
<th>TEXTURE NUCA</th>
<th>PERIMETER NUCA</th>
<th>AREA NUCA</th>
<th>SMOOTHN</th>
</tr>
</thead>
<tbody>
<tr>
<td>898690</td>
<td>B</td>
<td>11.47000000000000000000</td>
<td>16.03</td>
<td>73.02000000000000000000</td>
<td>402.70000000000000000000</td>
<td>0.09076</td>
</tr>
<tr>
<td>800147</td>
<td>B</td>
<td>11.95000000000000000000</td>
<td>11.96</td>
<td>77.23000000000000000000</td>
<td>426.70000000000000000000</td>
<td>0.11580</td>
</tr>
</tbody>
</table>
WDBC_CLASSIFICATION_MODEL
Binary Classification Model

General

Positive Label for DIAGNOSIS: M

Quality

F1 Value: 0.93
Model Accuracy: 93%
Precision: 100%
Recall: 87%
False Positive Rate: 0%

Predicted Values

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>33</td>
<td>5</td>
<td>38 (50%)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>38</td>
<td>38 (50%)</td>
</tr>
<tr>
<td>Total</td>
<td>33 (43%)</td>
<td>43 (57%)</td>
<td>76 (100%)</td>
</tr>
</tbody>
</table>

Correct Prediction: 100%
Incorrect Prediction: 100%
Apply Model

Model: WDBC_CLASSIFICATION_MODEL

Outputs

Create | Output | Column Name

- PredictedValue: PREDICTED_DIAGNOSIS

Parameters

- Maximum Null Value Percent: 80
The Data Flow was successfully saved.

Save Data Set

Name: WDBC_MODEL_APPLY

Columns

- Name: ID
  - Treat As: Attribute
  - Default Aggregation: 
- Name: RADIUS_NUC
  - Treat As: Measure
  - Default Aggregation: Sum
- Name: TEXTURE_NUC
  - Treat As: Measure
  - Default Aggregation: Sum
- Name: PERIMETER_NUC
  - Treat As: Measure
  - Default Aggregation: Sum
- Name: AREA_NUC
  - Treat As: Measure
  - Default Aggregation: Sum
- Name: SMOOTHNESS_NUC
  - Treat As: Measure
  - Default Aggregation: 

Data Set Storage

- When Run: Prompt to specify Data Set

Data:

<table>
<thead>
<tr>
<th>ID</th>
<th>RADIUS_NUC</th>
<th>TEXTURE_NUC</th>
<th>PERIMETER_NUC</th>
<th>AREA_NUC</th>
<th>SMOOTHNESS_NUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>898690</td>
<td>11.470</td>
<td>16.03</td>
<td>73.02</td>
<td>402.7</td>
<td>0.09076</td>
</tr>
<tr>
<td>800147</td>
<td>11.050</td>
<td>14.06</td>
<td>77.23</td>
<td>428.7</td>
<td>0.11580</td>
</tr>
</tbody>
</table>
case when PREDICTED_DIAGNOSIS = DIAGNOSIS then 'Correct' else 'incorrect' end

Calculation validated
WDBC Prediction Results

# Validation Records

189

Results Confusion Matrix

<table>
<thead>
<tr>
<th>PREDICTED_DIAGNOSIS</th>
<th>B</th>
<th>M</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGNOSIS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>83.60%</td>
<td>4.76%</td>
<td>88.36%</td>
</tr>
<tr>
<td>M</td>
<td>1.06%</td>
<td>10.58%</td>
<td>11.64%</td>
</tr>
<tr>
<td>Total</td>
<td>84.66%</td>
<td>15.34%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Actual Diagnoses

- B: 88.36%
- M: 11.64%

Proportion of Correct Predictions

- Correct: 94.18%
- Incorrect: 5.82%

Predicted Diagnoses

- B: 84.66%
- M: 15.34%
Create Connection

Oracle Autonomous Data Warehouse Cloud

- **Connection Name**: HealthML
- **Description**: 
- **Client Credentials**: cwallet.sso
- **Username**: OML
- **Password**: ********
- **Service Name**: healthcareml_medium

[Create, Cancel, Save]
<table>
<thead>
<tr>
<th>Name</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>WDBC</td>
<td>SVM Classification Model for tumor data</td>
</tr>
<tr>
<td>Connection</td>
<td>Global</td>
</tr>
</tbody>
</table>
SELECT * FROM WDBC_MODEL_TRAIN;

<table>
<thead>
<tr>
<th>ID</th>
<th>DIAGNOSIS</th>
<th>RADIUS_NUCA</th>
<th>TEXTURE_NUCA</th>
<th>PERIMETER_NUCA</th>
<th>AREA_NUCA</th>
<th>SMOOTHNESS_NUCA</th>
<th>COMPACTNESS_NUCA</th>
<th>CONCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>842302</td>
<td>M</td>
<td>17.98999999999998</td>
<td>10.38</td>
<td>122.8</td>
<td>1001</td>
<td>0.118400000000000000001</td>
<td>0.277600000000000000001</td>
<td>0.300090000000000000001</td>
</tr>
<tr>
<td>842517</td>
<td>M</td>
<td>20.57</td>
<td>17.77</td>
<td>132.9</td>
<td>1326</td>
<td>0.084739999999999999996</td>
<td>0.078640000000000000002</td>
<td>0.086900000000000000002</td>
</tr>
<tr>
<td>8430093</td>
<td>M</td>
<td>19.6900000000000000001</td>
<td>21.25</td>
<td>130</td>
<td>1203</td>
<td>0.1096</td>
<td>0.159899999999999999996</td>
<td>0.197390000000000000002</td>
</tr>
<tr>
<td>84348301</td>
<td>M</td>
<td>11.42</td>
<td>20.38</td>
<td>77.58</td>
<td>386.1</td>
<td>0.142499999999999999996</td>
<td>0.283899999999999999999</td>
<td>0.241400000000000000002</td>
</tr>
<tr>
<td>84358402</td>
<td>M</td>
<td>20.29</td>
<td>14.34</td>
<td>135.1</td>
<td>1297</td>
<td>0.1003</td>
<td>0.1328</td>
<td>0.198000000000000000002</td>
</tr>
<tr>
<td>843786</td>
<td>M</td>
<td>12.45</td>
<td>15.7</td>
<td>82.57</td>
<td>477.1</td>
<td>0.1278</td>
<td>0.17</td>
<td>0.157800000000000000002</td>
</tr>
<tr>
<td>844359</td>
<td>M</td>
<td>18.25</td>
<td>19.98</td>
<td>119.6</td>
<td>1040</td>
<td>0.094630000000000000006</td>
<td>0.109</td>
<td>0.112690000000000000002</td>
</tr>
</tbody>
</table>

 Took 4 sec. Last updated by OML at July 04 2019, 11:02:57 AM.
DECLARE
v_sql varchar2(100);
BEGIN
/* drop build settings */
BEGIN
  v_sql := 'DROP TABLE WDBC_SET';
  EXECUTE IMMEDIATE v_sql;
  DBMS_OUTPUT.PUT_LINE(v_sql || ': succeeded');
  EXCEPTION
  WHEN OTHERS THEN
    DBMS_OUTPUT.PUT_LINE(v_sql || ': drop unnecessary - no table exists');
END;
/* drop any previous model. */
BEGIN
  v_sql := 'CALL DBMS_DATA_MINING.DROP_MODEL(:WDBCMODEL);';
  EXECUTE IMMEDIATE v_sql;
  DBMS_OUTPUT.PUT_LINE(v_sql || ': succeeded');
  EXCEPTION
  WHEN OTHERS THEN
    DBMS_OUTPUT.PUT_LINE(v_sql || ': drop unnecessary - no model exists');
END;
/* Create a Build Setting table for Model Build */
EXECUTE IMMEDIATE 'CREATE TABLE WDBC SET (setting_name VARCHAR2(30), setting_value VARCHAR2(4000))';
EXECUTE IMMEDIATE 'INSERT INTO WDBC SET (setting_name, setting_value) VALUES (''ALGO_NAME'', ''ALGO_SUPPORT VECTOR MACHINES'')';
EXECUTE IMMEDIATE 'INSERT INTO WDBC SET (setting_name, setting_value) VALUES (''PREP_AUTO'', ''ON'')';
DBMS_OUTPUT.PUT_LINE ('Created model build settings table: WDBC SET');
/* Build the 1-Class SVM model. */
EXECUTE IMMEDIATE 'CALL DBMS_DATA_MINING.CREATE_MODEL(:WDBCMODEL, ''CLASSIFICATION'', :WDBC_MODEL TRAIN'', ''ID'', ''DIAGNOSIS'', :WDBC_SET);'
DBMS_OUTPUT.PUT_LINE ('Created model: WDBC MODEL');
END;
Use Model for Ad-Hoc Analyses

```sql
SELECT `prediction(WDBC) using '19.01' as RADIUS_NAUC,
      '17.82' as TEXTURE_NAUC,
      '12.1' as PERIMETER_NAUC,
      '1107' as AREA_NAUC,
      '0.1089' as SMOOTHNESS_NAUC,
      '0.3484' as COMPACTNESS_NAUC,
      '0.21559999999999999' as CONVEXITY_NAUC,
      '0.1104' as CONCAVE_POINTS_NAUC,
      '0.8229999999999999' as SYMMETRY_NAUC,
      '0.06116' as FRACAL_DIMENSION_NAUC,
      '0.5658999999999999' as RADIUS_NUCB,
      '1.4079999999999999' as TEXTURE_NUCB,
      '3.650000000000000' as PERIMETER_NUCB,
      '0.7399999999999999' as AREA_NUCB,
      '0.059288' as SMOOTHNESS_NUCB,
      '0.82813' as COMPACTNESS_NUCB,
      '0.64256' as CONVEXITY_NUCB,
      '0.61176' as CONCAVE_POINTS_NUCB,
      '0.971717' as SYMMETRY_NUCB,
      '0.0032289999999999' as FRACAL_DIMENSION_NUCB,
      '23.96' as RADIUS_NUCB,
      '38.39' as TEXTURE_NUCB,
      '153.0' as PERIMETER_NUCB,
      '1740' as AREA_NUCB,
      '0.1514' as SMOOTHNESS_NUCB,
      '0.2725' as COMPACTNESS_NUCB,
      '0.5936' as CONVEXITY_NUCB,
      '0.2059999999999999' as CONCAVE_POINTS_NUCB,
      '1.8266' as SYMMETRY_NUCB,
      '0.0000' as FRACAL_DIMENSION_NUCB
FROM PREDICTED_DIAGNOSIS; ```
Download Client Credentials (Wallet)
Connections to Autonomous Data Warehouse use a secure connection. Your existing tools and applications will need to use this wallet file to connect to your Autonomous Data Warehouse instance. If you are familiar with using an Oracle Database within your own data center, you may not have previously used these secure connections.

Set Resource Management Rules
Set resource management rules to allocate CPU/I0 shares to consumer groups and to cancel SQL statements based on their runtime and amount of I0.

Set Administrator Password
Set or reset your database administrator user's (ADMIN) password and when locked unlock your administrator user account on Autonomous Data Warehouse.

Manage Oracle ML Users
Create new Oracle Machine Learning user accounts and manage the credentials for existing Oracle Machine Learning users.

REST Data Services
Oracle REST Data Services (ORDS) on your Autonomous Data Warehouse instance provides HTTPS access to your data.

Download Oracle Instant Client
This is a free, light-weight set of tools, libraries and SDKs for building and connecting applications. These libraries underly the Oracle APIs of languages including Node.js, Python and PHP and provide access for OCI, OCI, JDBC, ODBC and Pro*C applications. Tools such as SQL*Plus and Oracle Data Pump are also included - Oracle recommends using this version of Data Pump for moving existing Oracle Database schemas to Autonomous Data Warehouse.

Send Feedback to Oracle
Use our CloudCustomerConnect forum to provide feedback about the service to Oracle, post questions, connect with experts, and share your thoughts and ideas. Click here to link to the forum.
Handler Updated

**ORDS Handler Definition**

- **RESTful Service Module**: wdbc
- **Module Base Path**: /wdbc/
- **URI Template**: addpatient/
- **Full URL**: https://<IP>/addpatient/
- **Method**: POST
- **Source Type**: PL/SQL
- **Mime Types Allowed**: application/json
- **Comments**: Add new patient and return prediction probability for cell results

**Source**

```sql
declare
    PRED_LABEL VARCHAR2(1);
```

66 of 4000
## ORDS generated API for wdbc

<table>
<thead>
<tr>
<th>Method</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST</td>
<td>/addpatient/</td>
</tr>
<tr>
<td>PUT</td>
<td>/diagnosepatient</td>
</tr>
<tr>
<td>GET</td>
<td>/viewpatient</td>
</tr>
<tr>
<td>GET</td>
<td>/viewpredictions</td>
</tr>
</tbody>
</table>

[ BASE URL: /ords/oml/wdbc , API VERSION: 1.0.0 ]
ORS generated API for wdbc

default

POST /addpatient/

PUT /diagnosepatient

Implementation Notes
use this API to PUT a final diagnosis against a patient ID

Response Class (Status 200)
output of the endpoint

Model | Example Value
inline_model_0 {}

Response Content Type | application/json

Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Parameter Type</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>(required)</td>
<td></td>
<td>header</td>
<td>string</td>
</tr>
<tr>
<td>id</td>
<td>(required)</td>
<td></td>
<td>header</td>
<td>double</td>
</tr>
</tbody>
</table>

Try it out!
GET https://localhost:8081/rdbc/viewpatient?

ID=999007

```
```
import requests
import json

response = requests.get('https://wdbc/viewpredictions')
data = json.loads(response.content)
print(data['items'][0])
Summary

Using OAC we can explore our data
We can build a prototype predictive model
with OML we can build a scalable, production ready model inside the database
With ORDS we can expose our model as a REST API
Using REST we can deploy our model into custom applications