



## ANALYTICS AND DATA TechCasts

# From Data to Downpours: Oracle AI's Futuristic Adventure in Rainfall Forecasting

Craig Shallahamer – Applied AI Scientist, Viscosity North America [craig@viscosityna.com](mailto:craig@viscosityna.com)

# Helpful Links –

<https://www.youtube.com/@OracleAnalytics/videos>

<https://www.youtube.com/watch?v=cgnJeVu-plE&list=PL6gBNP-Fr8KWZkXpZnjr7lTMfDTj9-dfK&pp=iAQB>

<https://community.oracle.com/products/oracleanalytics>

<https://www.oracle.com/business-analytics/data-visualization/demos/>



July 13th

### Advancing Analytics at Rosendin

Presented by **Cathy Pendley**



Aug 10th

### Our FAV Features of OAC

Presented by **Dan Vlamis & Wayne Van Sluys**



Oct 5th

### What AnDOUC Learned at Cloud

Presented by **Cathy Pendley, Dan Vlamis, Tim Vlamis, Abi Giles-Haigh**

## TechCast Archive

[Click to see Live TechCast page](#)

2023	2022	2021	2020	2019
Date	Title	Presenter(s)	Replay	Download(s)
May 4	Oracle APEX: A Swiss Army Knife Story for Your Analytics	Lucas Hirscheegger & Simon Collins	<a href="#">Video</a>	<a href="#">Slides</a>
Apr 20	From Data to Insights with Oracle Analytics	Joel Acha	<a href="#">Video</a>	<a href="#">Slides</a>
Apr 6	Data Platform Migrations – Few Learnings	Sujata Balupala & Sanjay Sabnis	<a href="#">Video</a>	<a href="#">Slides</a>
Mar 14–16	AnDOUC Summit 2023	AnDOUC	--	--
Feb 16	Favorite New Features in Jan 2023 OAC	--	--	--
Jan 26	Summit Preview of Presentations	Various Presenters	<a href="#">Video</a>	<a href="#">Slides</a>
Jan 12	Oracle Analytics & Spatial Studio	Wayne Van Sluys and David Lapp	<a href="#">Video</a>	<a href="#">Slides</a>

Submit a topic to share at <https://andouc.org/techcasts/>

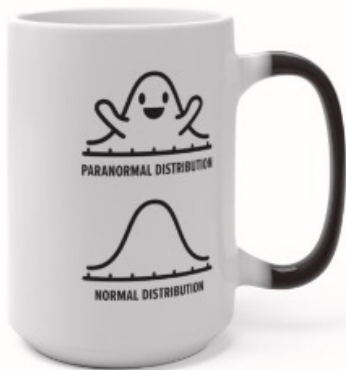


[www.andouc.org](http://www.andouc.org)





Visit the AnDOUC Store at [ANDOUC.ORG](https://ANDOUC.ORG)





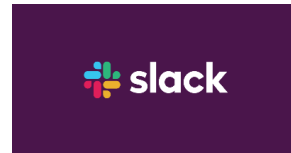


Let's Connect



**Website**

<http://andouc.org/>



**AnDOUC Techcast series**

**Chat with the Experts**

<https://bit.ly/Join-ANDOUC-Slack>



**Watch Previous TechCasts**

<https://bit.ly/3qmGgHN>



**@AnalyticAndData**



<https://www.facebook.com/AnDOracleUserCommunity>



<https://www.linkedin.com/company/analytics-and-data-oracle-user-community>



**Spatial + Graph SIG**  
[bit.ly/Spatial-Graph-LinkedIn](https://bit.ly/Spatial-Graph-LinkedIn)



*Call for Speakers  
now open!*



*Save the Date!*

## **Analytics and Data Summit 2024**

March 19-21, 2024  
Oracle Conference Center  
Redwood Shores, California

[www.andouc.org/andsummit2024](http://www.andouc.org/andsummit2024)



## From Data to Downpours: Oracle AI's Futuristic Adventure in Rainfall Forecasting!

AnDOUC TechCasts | Knowledge, Expertise, and Ideas for All  
September 7, 2023



ANALYTICS AND DATA  
**TechCasts**



Craig Shallahamer

Oracle ACE Director  
Applied AI Scientist | OraPub Founder

[@orapub](https://twitter.com/orapub)

[linkedin.com/in/craig-shallahamer-571a94a/](https://www.linkedin.com/in/craig-shallahamer-571a94a/)

[craig.shallahamer@viscosityna.com](mailto:craig.shallahamer@viscosityna.com)

[viscosityna.com](https://viscosityna.com)

[@ViscosityNA](https://twitter.com/ViscosityNA)



**Craig Shallahamer**

**Oracle ACE Director  
Applied AI Scientist | OraPub Founder**

 [@orapub](https://twitter.com/orapub)

 [linkedin.com/in/craig-shallahamer-571a94a/](https://www.linkedin.com/in/craig-shallahamer-571a94a/)

 [craig.shallahamer@viscosityna.com](mailto:craig.shallahamer@viscosityna.com)

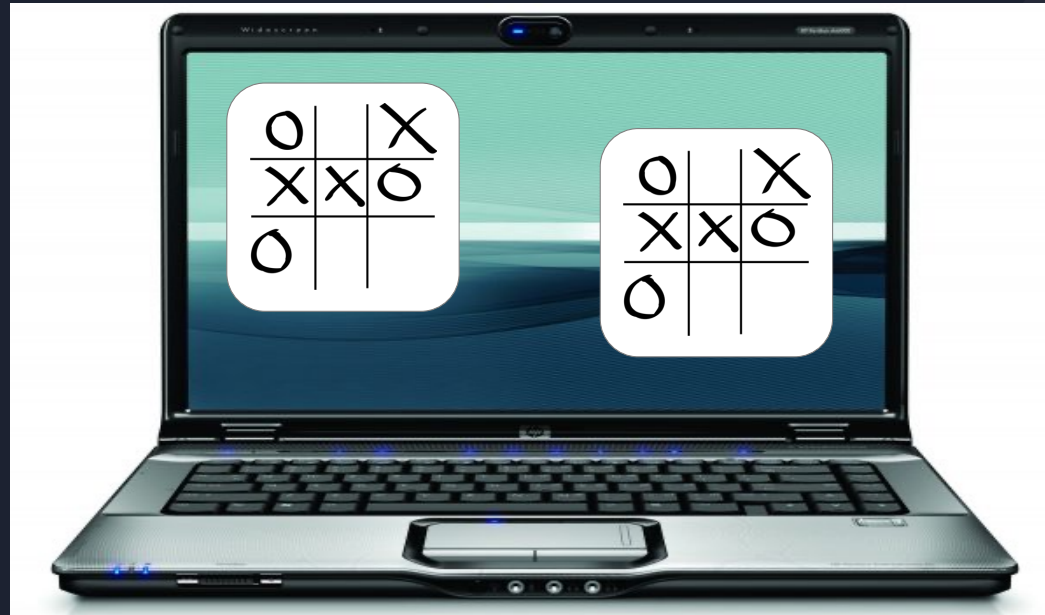
Craig Shallahamer is a respected figure in the field of machine learning and Oracle database performance tuning.

He has a wealth of experience in building and teaching predictive modeling techniques, including creating a Reinforcement Machine Learning bot in 1990. In addition to his work as an Applied AI Scientist at Viscosity and the founder of OraPub, Craig has written two highly regarded books, *Oracle Performance Firefighting* and *Forecasting Oracle Performance*.

He has received recognition as an Oracle ACE Director for his valuable contributions to the Oracle community through his technical expertise and leadership. Craig is an active participant in Oracle user groups, frequently presenting at conferences and serving as a board member and volunteer.

His extensive knowledge and experience in both machine learning and Oracle database performance tuning make him a leading authority in these fields.

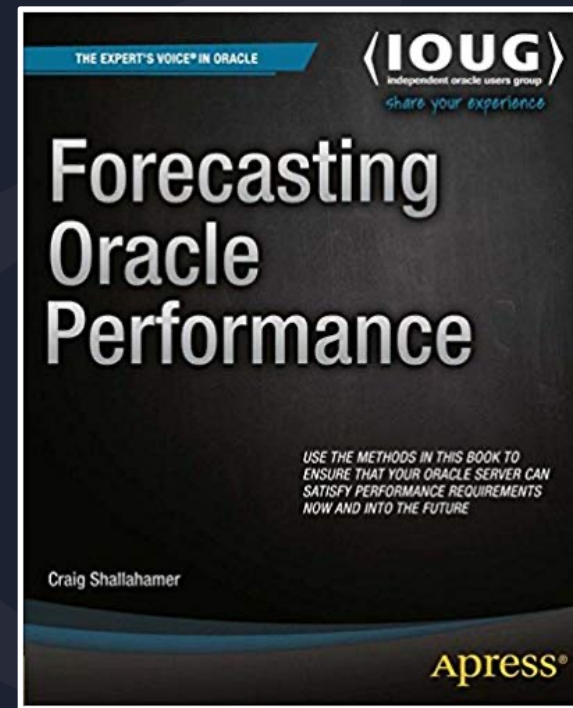
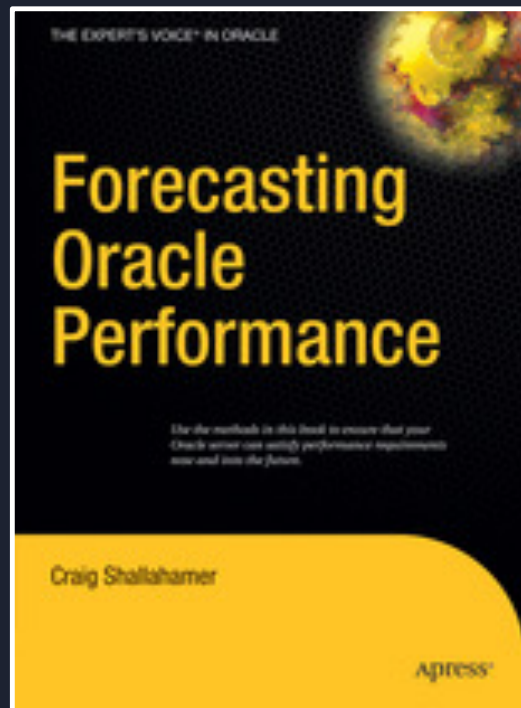
# Then again in 1990



I created a program where I could play the game tic-tac-toe with a computer, but also enabled the computer to play itself... and learning through that process.



# 2007



## MACHINE LEARNING 2

### Performance Prediction



ORAPUB

## MACHINE LEARNING 1

### Anomaly Detection

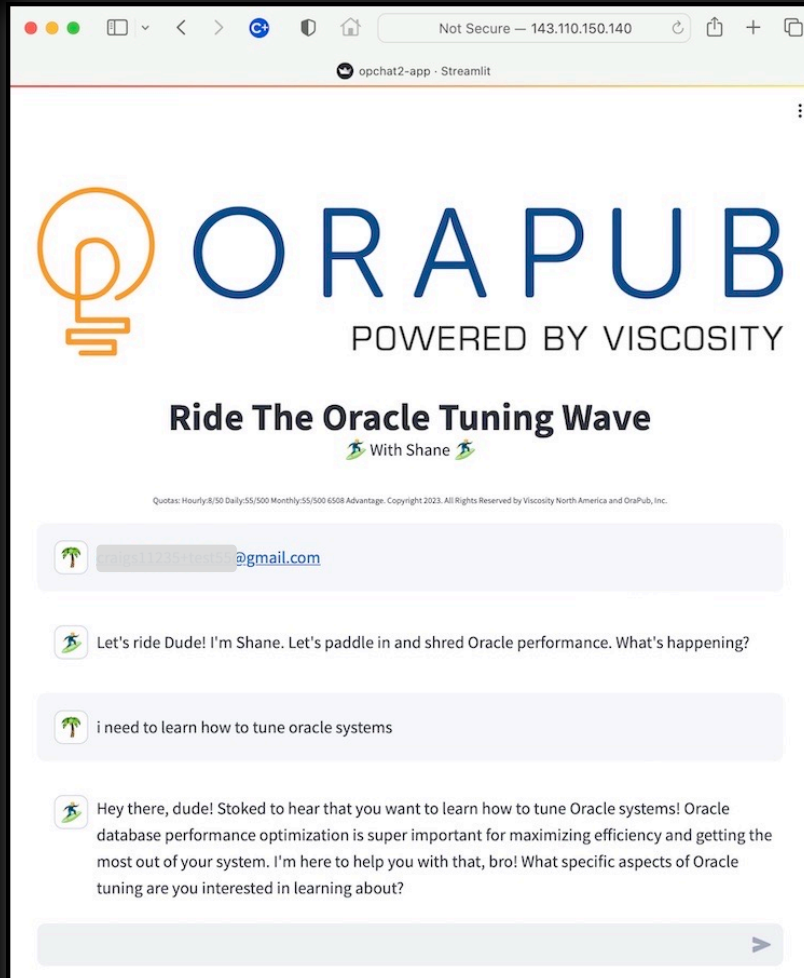


ORAPUB



# From a more business perspective...

- Internet marketing, focusing on identifying "hot leads" for very specific promotions.
- IT Operations when many Oracle systems must be monitored and Operations possibly alerted; both anomaly detection and performance prediction.
- Domain specific generative Q&A Bots and Conversational Assistants With Personality in multiple business domains. For example, Oracle Performance Tuning, Biblical Texts, Real estate investing platform.



# Want to try?

Focused on  
Oracle  
performance  
and internals.

# Viscosity Pillars and Delivery Models



## DATA

Oracle & SQL Server Postgres  
Performance Tuning  
Data Replication  
Data Warehousing Analytics  
Data Integration  
ERP Blue Prints  
Database Upgrades



## APPS

APEX  
EBS  
Web/Mobile Apps  
.Net and C#  
E-Business Suite  
SAAS/PAAS



## CLOUD

Azure Gold Partner  
Cloud Migrations  
Engineered Systems  
Oracle Cloud Partner  
Google Partner  
AWS Partner Hybrid Cloud

Workshops

Assessments

Proof of  
Concepts

Training

Turnkey  
Projects

Managed  
Services

# Viscosity's Oracle ACEs

## The Oracle ACE Program

The Oracle ACE Program recognizes and rewards individuals for their contributions to the Oracle community.



**Charles Kim**  
CEO | Co-Founder

 @racdba  
 ACE Director



**Rich Niemiec**  
Chief Innovation Officer

 @richniemiec  
 ACE Director



**Craig Shallahamer**  
Applied AI Scientist

 @orapub  
 ACE Director



**Sean Scott**  
Principal Consultant

 @oraclesean  
 ACE Director

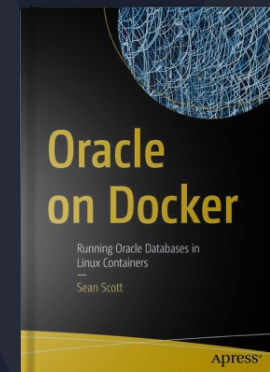
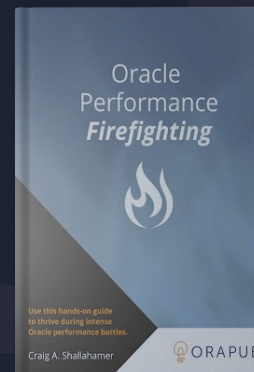
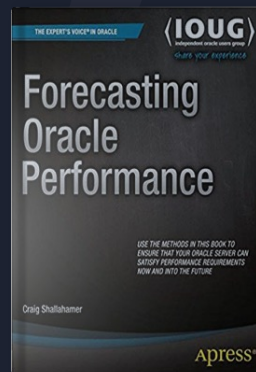
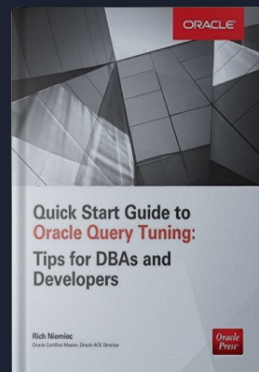
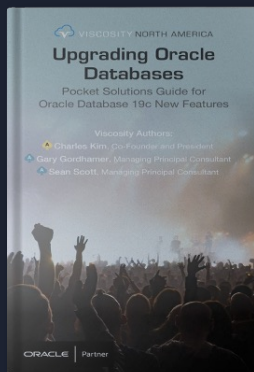
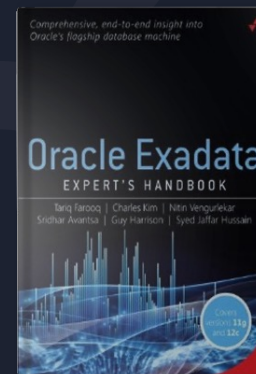
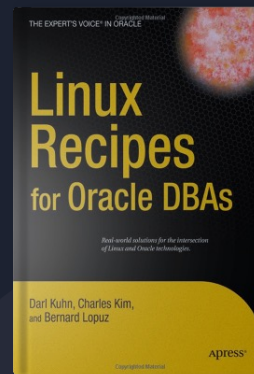
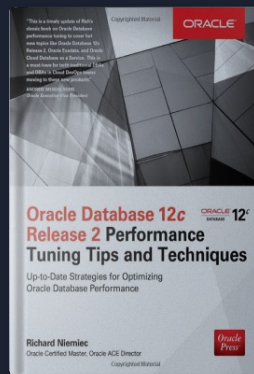
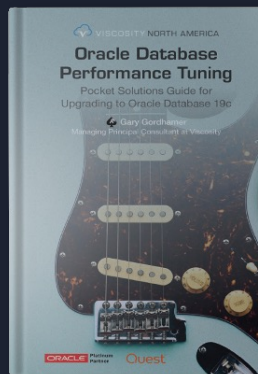
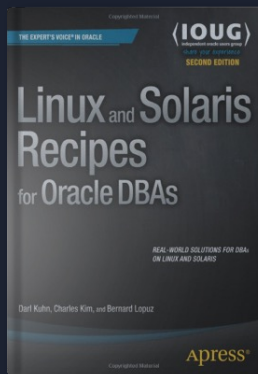


**Gary Gordhamer**  
Principal Consultant

 @ggordham  
 ACE Pro



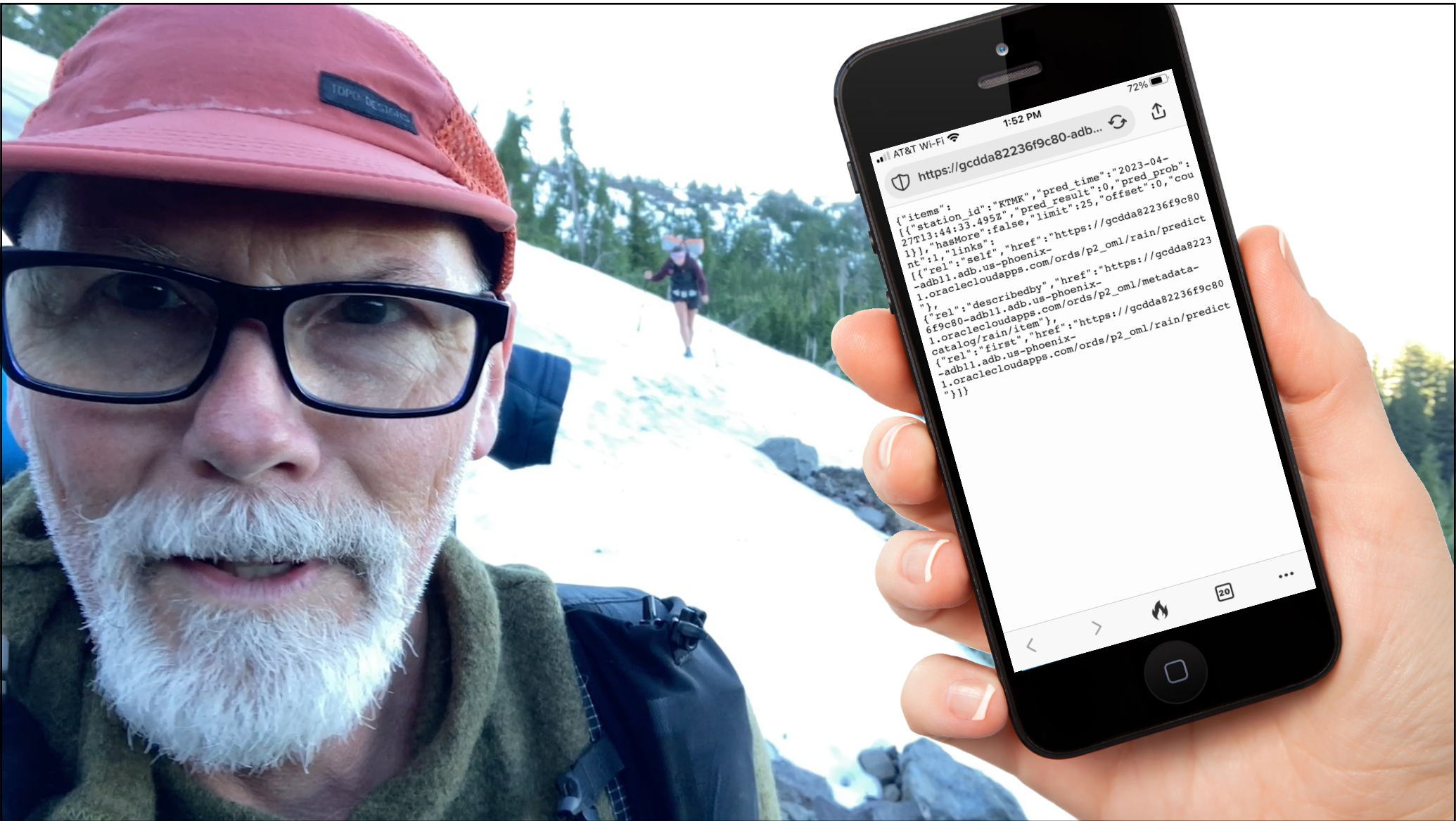
# We've written over 25 technical books!



READY?







AT&T Wi-Fi 1:52 PM 72%  
https://gcdda82236f9c80-adb...  
{  
 "items":  
 [{  
 "station\_id": "RTMK",  
 "pred\_time": "2023-04-27T13:44:33.495Z",  
 "pred\_result": 0,  
 "pred\_prob": 0,  
 "count": 1,  
 "links":  
 [{  
 "rel": "self",  
 "href": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2\_oml/rain/predict-1",  
 "describedby": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2\_oml/metadata-catalog/rain/item",  
 "first": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2\_oml/rain/predict-1",  
 "offset": 0,  
 "limit": 25,  
 "hasMore": false  
 }]  
 }  
 ]  
 }  
}





```
{
  "items": [
    {
      "station_id": "RTMK",
      "pred_time": "2023-04-27T13:44:33.495Z",
      "pred_result": 0,
      "pred_prob": 1,
      "hasMore": false,
      "limit": 25,
      "offset": 0,
      "count": 1,
      "links": [
        {
          "rel": "self",
          "href": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/rain/predict"
        },
        {
          "rel": "describedby",
          "href": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/metadata-catalog/rain/item"
        },
        {
          "rel": "first",
          "href": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/rain/predict"
        }
      ]
    }
  ]
}
```





<http://bit.ly/bp4d-rain>



# The Situation

It's helpful if I know it's going to rain.  
Most of my hikes are less than six hours.  
What I need is a six hour rain forecast.  
Ever hear of a 6hr rain forecast? No.  
And, local forecast aren't really so local.  
Perhaps AI can help?





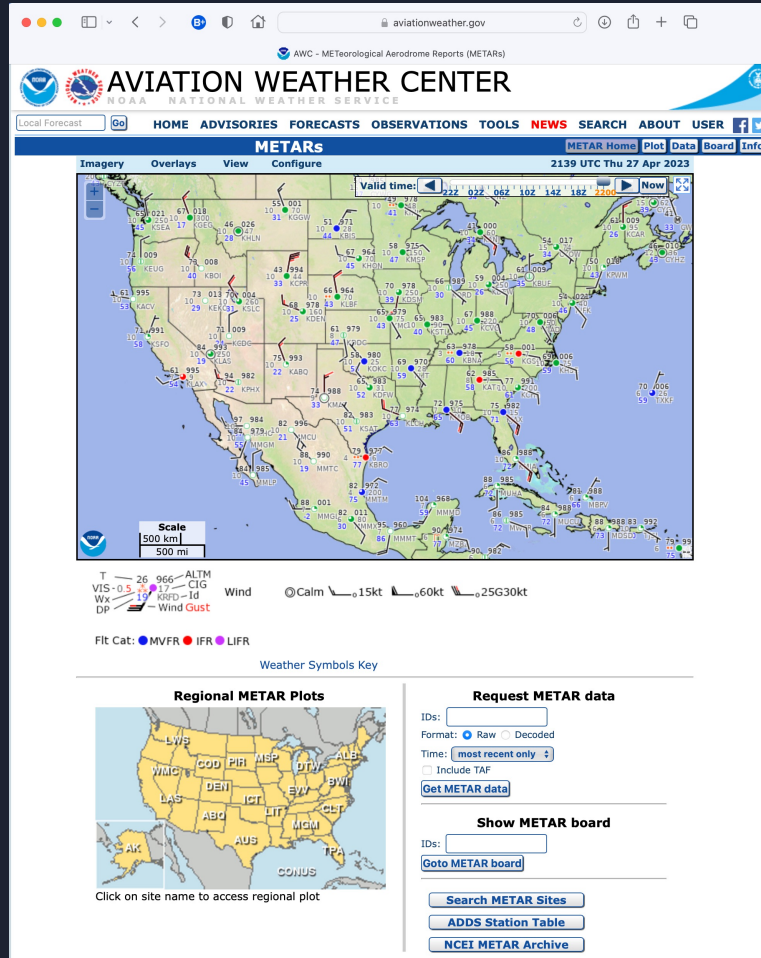
# Objective Summary

Build an AI,  
to forecast the likelihood of rain,  
from now out to six hours into the future.  
Needs to run from a URL on my phone.



# Data...

# Lots of it!



COUNT(*)	LOAD_ID	STATION_ID
910	K4S1-20230214	K4S1
910	K4S1-20230302	K4S1
910	K4S1-20230331	K4S1
910	K6S2-20230214	K6S2
914	K6S2-20230302	K6S2
910	K6S2-20230331	K6S2
553	KAST-20230214	KAST
580	KAST-20230302	KAST
515	KAST-20230331	KAST
478	KEUG-20230214	KEUG
430	KEUG-20230302	KEUG
425	KEUG-20230331	KEUG
501	KOTH-20230214	KOTH
557	KOTH-20230302	KOTH

10K rows

Autonomous Databases x Oracle Database Activi x Notebooks - OML x https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/oml/index.html?ojr=note...

UTOUG - Google... Cape Kiwanda RV... Surf Cam - Mome... Options Analysis -...

ORACLE Machine Learning P2\_OML Project P2\_OML Workspace P2\_OML

### Notebooks

Go to OML Notebooks EA

+ Create Edit Delete Duplicate Move Copy Save as Template Import

Export Version Copy to OML Notebooks EA Filter

<input type="checkbox"/>	Name ^	Comment	Created On ^	Created By ^	Last Update ^
<input type="checkbox"/>	P2 Build and Eval Model	2	4/27/23, 10:45 AM	P2_OML	4/27/23, 10:55 AM
<input type="checkbox"/>	P2 Deployment	3	4/27/23, 10:39 AM	P2_OML	4/27/23, 11:45 AM
<input type="checkbox"/>	P2 Load & Preprocess	1	4/27/23, 1:03 PM	P2_OML	4/27/23, 1:03 PM



OML Jobs automatically run the Deployment Notebook every twenty minutes, resulting in a forecast row inserted into a table.

The screenshot displays the Oracle Machine Learning (OML) Jobs interface. The page title is "ORACLE Machine Learning". The breadcrumb navigation shows "P2\_OML Project" and "P2\_OML Workspace". The "Jobs" section is active, showing a list of jobs. The table columns are: Name, Notebook, Owner, Last Start Date, Last Status, Next Run Date, Status, and Schedule. The first job listed is "P2 Predict Rain", which is a deployment notebook named "P2\_OML...." owned by "P2\_OML". It was last started on "4/27/23, 1:44 PM" and its status is "SUCCEEDED". The next run date is "4/27/23, 2:04 PM", and its status is "SCHEDULED". The schedule is "Every 20 min".

Name	Notebook	Owner	Last Start Date	Last Status	Next Run Date	Status	Schedule
P2 Predict Rain	P2_OML....	P2_OML	4/27/23, 1:44 PM	SUCCEEDED	4/27/23, 2:04 PM	SCHEDULED	Every 20 min

The REST API  
can be requested  
at any time from  
a smart phone to  
retrieve the rain  
forecast!

The screenshot displays the Oracle REST API interface for the 'predict' endpoint. The breadcrumb navigation shows 'REST > Modules > Predict Rain > predict > GET'. The endpoint name 'predict' is highlighted, with a green circle icon labeled 'GET'. It indicates 'Last Updated: 2 hours ago' and 'No comments available'. The source type is 'json/collection'. The URL is 'https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2\_oml/rain/predict'. A red arrow points to the URL. Below the endpoint details, the 'Source' section shows the SQL query: 

```
1 select station_id,
2 prediction_time_pacific pred_time,
3 prediction_results pred_result,
4 round(prediction_probability,3) pred_prob
5 from weather_prediction_results
```

 A green arrow points to the SQL query. On the right, the 'Handler Parameters' and 'Implicit Parameters' sections are visible, with parameters like ':body', ':body\_text', and ':content\_type'. The status bar at the bottom indicates '8:45:27 PM - REST call resolved successfully.' and 'Powered by ORDS'.

<http://bit.ly/bp4d-rain>



```
{
  "items": [
    {
      "station_id": "RTMK",
      "pred_time": "2023-04-27T13:44:33.495Z",
      "pred_result": 0,
      "pred_prob": 1,
      "hasMore": false,
      "limit": 25,
      "offset": 0,
      "count": 1,
      "links": {
        "self": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/rain/predict",
        "describedby": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/metadata-catalog/rain/item",
        "first": "https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/rain/predict"
      }
    }
  ]
}
```

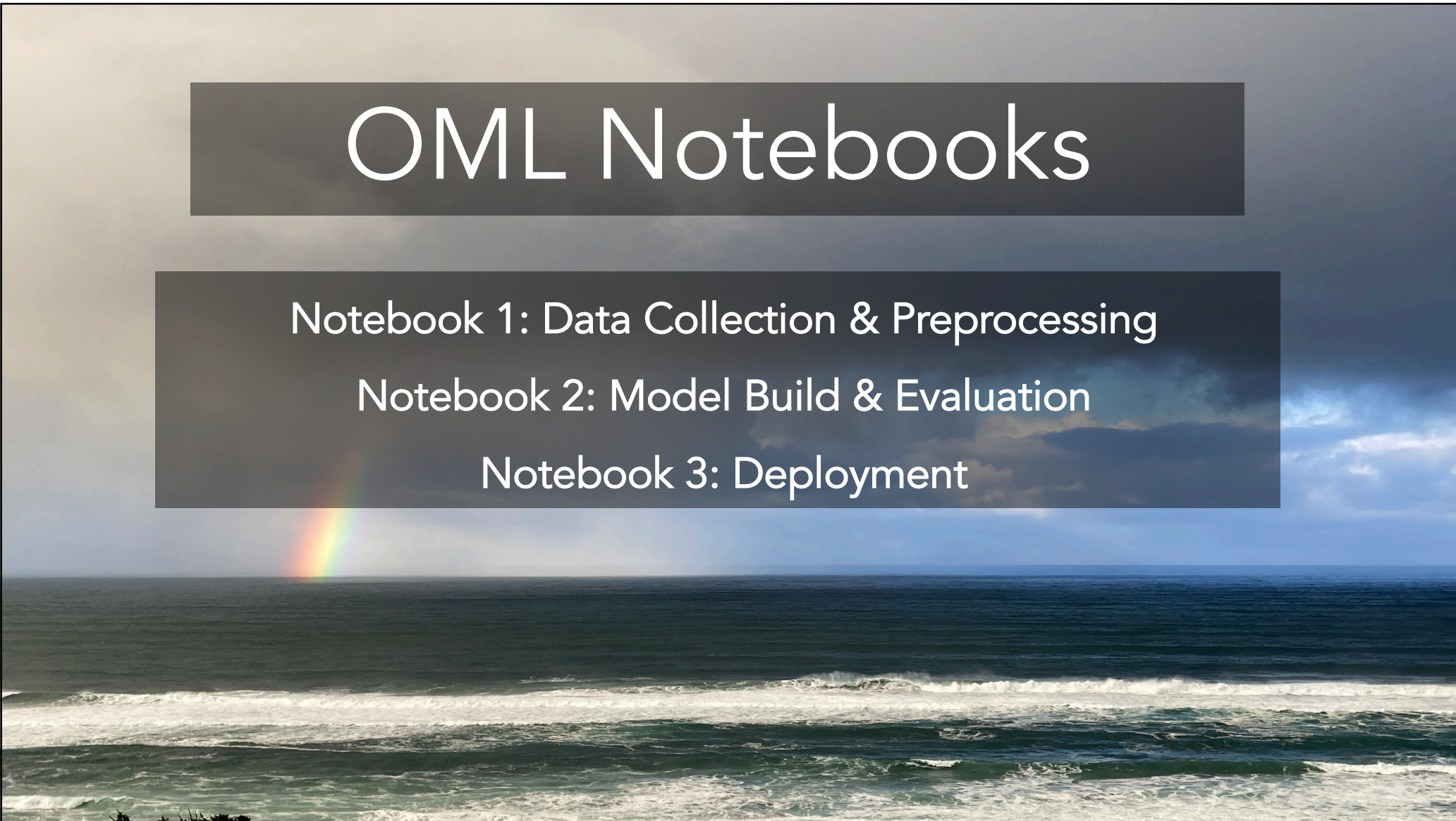


# OML Notebooks

Notebook 1: Data Collection & Preprocessing

Notebook 2: Model Build & Evaluation

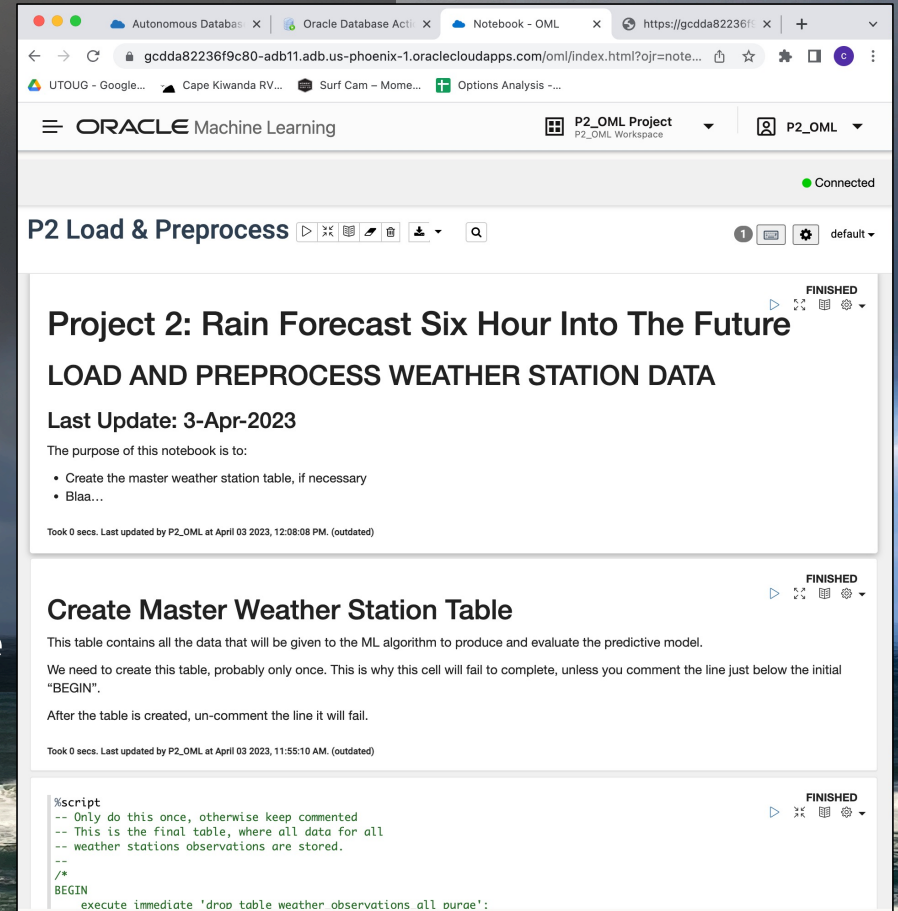
Notebook 3: Deployment





# Notebook 1: Data Collection & Preprocessing

- Create master weather station table
- Load raw METAR weather station data
- Explore working data
- Impute missing data values
- Feature engineering
- Rain label engineering
- Merge preprocessed data into master table



The screenshot displays the Oracle Machine Learning web interface. The browser address bar shows the URL: `https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/oml/index.html?ojr=note...`. The page header includes the Oracle Machine Learning logo and the project name 'P2\_OML Project'. The notebook title is 'P2 Load & Preprocess'. The content area shows the following text:

**Project 2: Rain Forecast Six Hour Into The Future**  
**LOAD AND PREPROCESS WEATHER STATION DATA**  
**Last Update: 3-Apr-2023**

The purpose of this notebook is to:

- Create the master weather station table, if necessary
- Blaa...

Below this, there is a section titled 'Create Master Weather Station Table' with the following text:

This table contains all the data that will be given to the ML algorithm to produce and evaluate the predictive model. We need to create this table, probably only once. This is why this cell will fail to complete, unless you comment the line just below the initial "BEGIN".

After the table is created, un-comment the line it will fail.

At the bottom, there is a code block with the following content:

```
%script
-- Only do this once, otherwise keep commented
-- This is the final table, where all data for all
-- weather stations observations are stored.
--
/*
BEGIN
execute immediate 'drop table weather observations all purge';
```



# Notebook 1: Data Collection & Preprocessing

"Zero down. Easy payments..."

The next time you hear how EASY it is to do AI/ML, remember all the preprocessing I needed to do.

Ask yourself or the speaker/teacher:

- Was the data already preprocessed?"
- Does "auto ML" do the preprocessing?

The screenshot displays the Oracle Machine Learning (OML) interface. The browser address bar shows a URL from oraclecloudapps.com. The notebook is titled "P2 Load & Preprocess" and is connected to a workspace named "P2\_OML Project". The main content area shows a project titled "Project 2: Rain Forecast Six Hour Into The Future" with the subtitle "LOAD AND PREPROCESS WEATHER STATION DATA". The last update is dated 3-Apr-2023. The notebook content includes a purpose statement, a list of tasks (e.g., "Create the master weather station table, if necessary"), and a code block for creating a master weather station table. The code block is marked as "FINISHED" and contains SQL commands to create a table and purge data.

ORACLE Machine Learning

P2\_OML Project  
P2\_OML Workspace

Connected

P2 Load & Preprocess

Project 2: Rain Forecast Six Hour Into The Future

LOAD AND PREPROCESS WEATHER STATION DATA

Last Update: 3-Apr-2023

The purpose of this notebook is to:

- Create the master weather station table, if necessary
- Blaa...

Took 0 secs. Last updated by P2\_OML at April 03 2023, 12:08:08 PM. (outdated)

Create Master Weather Station Table

This table contains all the data that will be given to the ML algorithm to produce and evaluate the predictive model.

We need to create this table, probably only once. This is why this cell will fail to complete, unless you comment the line just below the initial "BEGIN".

After the table is created, un-comment the line it will fail.

Took 0 secs. Last updated by P2\_OML at April 03 2023, 11:55:10 AM. (outdated)

```
%script
-- Only do this once, otherwise keep commented
-- This is the final table, where all data for all
-- weather stations observations are stored.
--
/*
BEGIN
  execute immediate 'drop table weather observations all purge';
```

# Explore Working Data

## Explore Working Data

Our working data has lots of problems and it can be improved before given to the chosen ML algorithm.

Took 0 secs. Last updated by P2\_OML at April 03 2023, 11:55:22 AM. (outdated)

FINISHED

```
%sql
-- Explore the data
select * from weather_preprocess order by load_id, station_id, observation_time;
```

FINISHED

       settings

LOAD_ID	STATION_ID	OBSERVATION_TIME	TEMP_C	DEWPOINT_C	ALTIM_IN_HG	WIND_SPEED_KT	WIND_GUST_KT	PRECIP_IN
KTKI-20230331	KTKI	2023-03-16 21:44:00	20.6	17.8	29.639763	10	17	
KTKI-20230331	KTKI	2023-03-16 21:53:00	20.6	18.3	29.639763	12		
KTKI-20230331	KTKI	2023-03-16 22:13:00	20.6	18.3	29.648623	12		
KTKI-20230331	KTKI	2023-03-16 22:28:00	20.6	18.3	29.681103	12		0.005
KTKI-20230331	KTKI	2023-03-16 22:36:00	20.6	18.3	29.66929	8	18	0.005
KTKI-20230331	KTKI	2023-03-16 22:51:00	20	18	29.619095	17		0.005
KTKI-20230331	KTKI	2023-03-16 22:53:00	20	18.3	29.630905	14		0.005
KTKI-20230331	KTKI	2023-03-16 23:53:00	14.4	11.7	29.740158	26	38	

Took 1 sec. Last updated by P2\_OML at April 03 2023, 11:55:23 AM. (outdated)



# Impute Missing Data Values

## Impute Missing Data Values

Machine learning algorithms expect numeric input data. While OML takes care of basic prep as we want.

In our dataset, it is appropriate and correct to “impute” (ie, set) a missing value to zero. OML

Took 0 secs. Last updated by P2\_OML at April 03 2023, 11:55:24 AM. (outdated)

```
%script
```

```
/*
```

```
Our data contains three columns that may contain a non-value.  
We will impute any non-value with a zero.
```

```
This cell can be re-run repeatedly because rows are updated only if the  
contain a missing value. Once the value is imputed, the row will long be  
updated.
```

```
*/
```

```
update weather_preprocess set wind_speed_kt = 0 where wind_speed_kt is null;  
update weather_preprocess set wind_gust_kt = 0 where wind_gust_kt is null;  
update weather_preprocess set precip_in = 0 where precip_in is null;
```

# Feature Engineering

## Feature Engineering

We will create, that is, engineer new features. Specifically, the engineered features

This allows a non-temporal algorithm to produce a temporally aware model.

Our focus will be on engineering moving average features.

Took 0 secs. Last updated by P2\_OML at April 03 2023, 11:55:30 AM. (outdated)

```
%script
```

```
-- Engineer moving average features.
```

```
--
```

```
-- This cell can be re-run repeatedly.
```

```
BEGIN
```

```
execute immediate 'alter table weather_preprocess add (  
    precip_in_24hr_sum number default 0,  
    precip_in_12hr_sum number default 0,  
    precip_in_6hr_sum number default 0,  
    precip_in_3hr_sum number default 0,  
    precip_in_1hr_sum number default 0,  
    altim_in_hg_24hr_avg number default 0,  
    altim_in_hg_12hr_avg number default 0,  
    altim_in_hg_6hr_avg number default 0,  
    altim_in_hg_3hr_avg number default 0,  
    altim_in_hg_1hr_avg number default 0,  
    label number default -99  
);
```

```
EXCEPTION WHEN OTHERS THEN NULL;
```

```
END;
```

```
/
```



# Merge Preprocessed Data Into Master Table

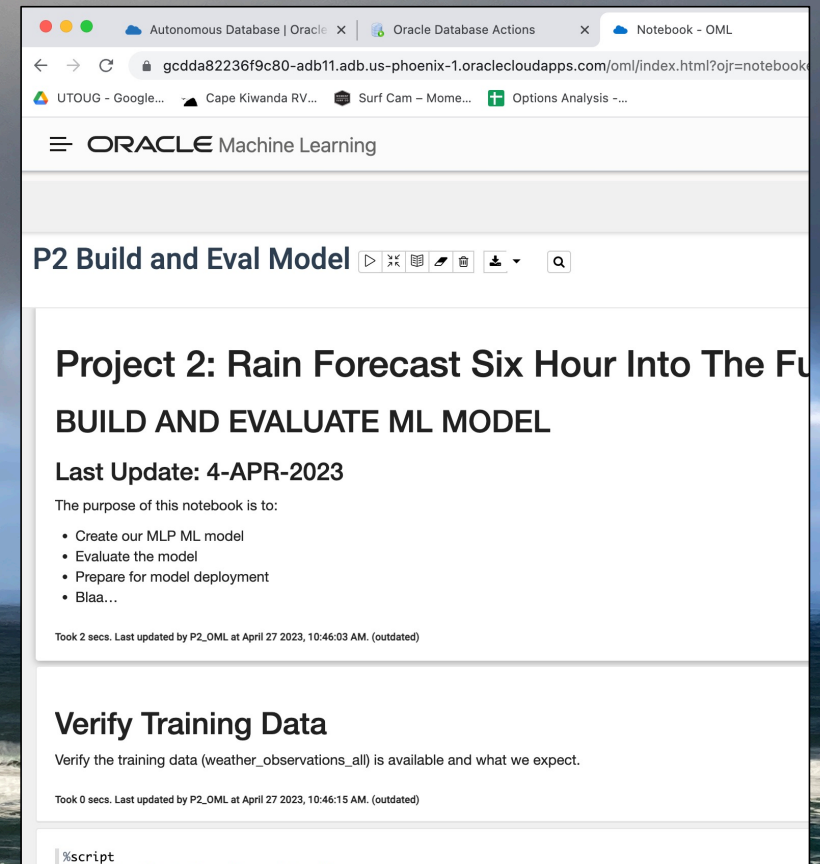
COUNT(*)	LOAD_ID	STATION_ID
910	K4S1-20230214	K4S1
910	K4S1-20230302	K4S1
910	K4S1-20230331	K4S1
910	K6S2-20230214	K6S2
914	K6S2-20230302	K6S2
910	K6S2-20230331	K6S2
553	KAST-20230214	KAST
580	KAST-20230302	KAST
515	KAST-20230331	KAST
478	KEUG-20230214	KEUG
430	KEUG-20230302	KEUG
425	KEUG-20230331	KEUG
501	KOTH-20230214	KOTH
557	KOTH-20230302	KOTH

COUNT(*)	LOAD_ID	STATION_ID
----------	---------	------------



# Notebook 2: Model Build and Evaluation

- Verify the preprocessed training data
- Create training data view
- Determine best features
- Create "training" and "testing" views; 70/30
- Build Multi-layer Perceptron (MLP) Model
- Evaluate model



The screenshot displays the Oracle Machine Learning (OML) interface in a web browser. The browser's address bar shows the URL: `gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/oml/index.html?ojr=notebook`. The page title is "ORACLE Machine Learning". The notebook title is "P2 Build and Eval Model". The notebook content includes a section titled "Project 2: Rain Forecast Six Hour Into The Future" and a sub-section "BUILD AND EVALUATE ML MODEL". The "Last Update" is "4-APR-2023". The purpose of the notebook is to: "Create our MLP ML model", "Evaluate the model", "Prepare for model deployment", and "Blaa...". The notebook was last updated by "P2\_OML" at "April 27 2023, 10:46:03 AM. (outdated)". Below this, there is a section titled "Verify Training Data" with the text "Verify the training data (weather\_observations\_all) is available and what we expect." and a note "Took 0 secs. Last updated by P2\_OML at April 27 2023, 10:46:15 AM. (outdated)". At the bottom, there is a code editor with the text "%script".



# Create Training Data View

## Create Model Building View And Verify

A view gives us flexibility on creating new columns (eg, PRIMARY\_KEY) and allows flexibility when using the data. All these benefits occur without physically duplicating or changing our preprocessed data.

Took 0 secs. Last updated by P2\_OML at April 27 2023, 10:47:05 AM. (outdated)

```
%script
/*
Create the weather_observations_all_v view and verify.

The PRIMARY_KEY column is created to allow an individual sample to be uniquely
identified. Some of the OML algorithms expect us to have a primary key.
*/

create or replace view weather_observations_all_v
as
select
  station_id||'-'||observation_time PRIMARY_KEY,
  TEMP_C,
  DEWPOINT_C,
  ALTIM_IN_HG,
  WIND_SPEED_KT,
  WIND_GUST_KT,
  PRECIP_IN,
  PRECIP_IN_24HR_SUM,
  PRECIP_IN_12HR_SUM,
  PRECIP_IN_6HR_SUM,
  PRECIP_IN_3HR_SUM,
  PRECIP_IN_1HR_SUM,
  ALTIM_IN_HG_24HR_AVG,
  ALTIM_IN_HG_12HR_AVG,
  ALTIM_IN_HG_6HR_AVG,
  ALTIM_IN_HG_3HR_AVG,
  ALTIM_IN_HG_1HR_AVG,
  LABEL
from weather_observations_all;

-- Verification counts
select count(*) from weather_observations_all;
```

Using a view reducing data volume, DDL and DML.

Faster and less expensive.

Very flexible and fast.



# Determine The Best Features

*The best feature is no feature!*

Fewer features mean:

- Less data (really?)
  - Less memory, IO and CPU
- Better performing model
- Faster training
- Faster predictions

## Feature Selection

The best feature is no feature. OML4SQL provides a simple and powerful algorithm.

Took 0 secs. Last updated by P2\_OML at April 27 2023, 10:47:31 AM. (outdated)

```
%script
-- Determine the importance of each feature.

begin dbms_data_mining.drop_model ('FEATURE_IMPORTANCE_WORKING');
exception when others then null; end;

declare
  v_setlist dbms_data_mining.setting_list;
begin
  v_setlist('ALGO_NAME') := 'ALGO_AI_MDL';
  v_setlist('PREP_AUTO') := 'ON';

  dbms_data_mining.create_model2(
    model_name => 'FEATURE_IMPORTANCE_WORKING',
    mining_function => 'ATTRIBUTE_IMPORTANCE',
    data_query => 'select * from weather_observations_all_v',
    set_list => v_setlist,
    case_id_column_name => 'PRIMARY_KEY',
    target_column_name => 'LABEL');

end;
```

PL/SQL procedure successfully completed.

# The Best Feature is NO feature.

```
%sql
-- Notice the DM$VA then the name of the attribute importance model.
select * from DM$VAFEATURE_IMPORTANCE_WORKING where rownum < 200;
```

FINISHED



settings

## Which features to keep?

Available Fields

PARTITION\_NAME ATTRIBUTE\_NAME ATTRIBUTE\_SUBNAME ATTRIBUTE\_IMPORTANCE\_VALUE ATTRIBUTE\_RANK

xAxis

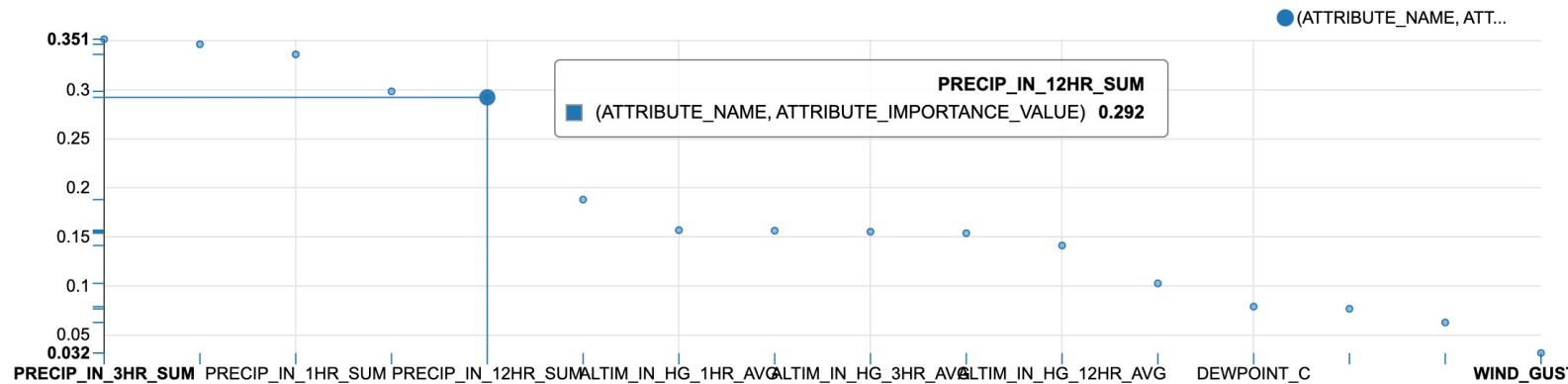
ATTRIBUTE\_NAME x

yAxis

ATTRIBUTE\_IMPORTANCE\_VALUE x

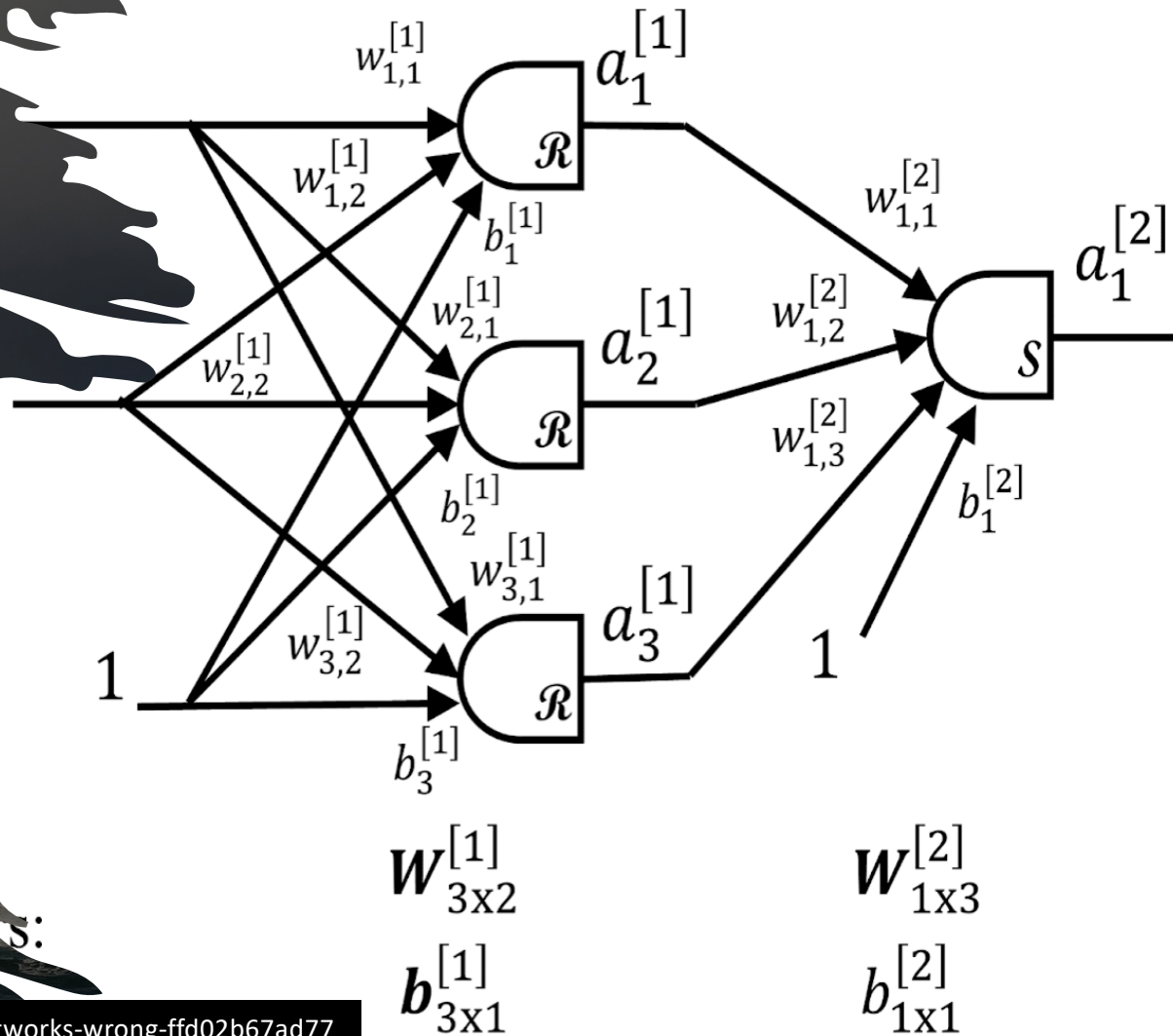
group

size





# Build Multi-layer Perceptron (MLP) Model



<https://towardsdatascience.com/please-stop-drawing-neural-networks-wrong-ffd02b67ad77>



# Build Multi-layer Perceptron Models (MLP)

Took 0 secs. Last updated by P2\_OML at April 27 2023, 10:50:03 AM. (outdated)

FINISHED  
▶ 🔍 📖 ⚙️ ▼

```
%script
-- ALL FEATURE VIEW
-- Build "ALL DATA" MLP model for weather_obs_features_all_v

begin
  dbms_data_mining.drop_model('MODEL_MLP_ALL');
  exception when others then null;
end;
/

DECLARE
  v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN
  v_setlst('PREP_AUTO')           := 'ON';
  --V_SETLST('NNET_SOLVER')       := 'NNET_SOLVER_ADAM';
  V_SETLST('NNET_HIDDEN_LAYERS') := '3';
  V_SETLST('NNET_NODES_PER_LAYER') := '16,32,16';
  V_SETLST('NNET_ITERATIONS')    := '1000';

  DBMS_DATA_MINING.CREATE_MODEL2(
    'MODEL_MLP_ALL',           -- Model name
    'CLASSIFICATION',
    'SELECT * FROM data_all_train_v', -- TRAIN_DATA
    V_SETLST,
    'PRIMARY_KEY',
    'LABEL');

END;
```

FINISHED  
▶ 🔍 📖 ⚙️ ▼

PL/SQL procedure successfully completed.

OML4SQL  
Build Multi-layer  
Perceptron Model  
(MLP)

# Evaluate MLP Classification Model: Confusion Matrix

## ALL Feature MLP Model

\*\*\*\* MODEL ACCURACY \*\*\*\*: .8433

ACTUAL_TARGET_VALUE	PREDICTED_TARGET_VALUE	VALUE
0	1	195
0	0	1404
1	0	308
1	1	1303

## SIX Feature MLP Model

\*\*\*\* MODEL ACCURACY \*\*\*\*: .8237

ACTUAL_TARGET_VALUE	PREDICTED_TARGET_VALUE	VALUE
0	1	203
0	0	1431
1	0	368
1	1	1236

What I really  
hate is believing  
it will NOT rain  
but then it rains!

Which model  
should I use?



# Notebook 3: Deployment

- Create model using all available data, not just training data.
- Do not evaluate mode, but could.
- Pull some new current weather data for specific station.
- Preprocess the new current data
- Forecast using new current data (one row) and re-created model.
- Store results, so REST API can access the forecast.

The screenshot displays the Oracle Machine Learning (OML) workspace. The browser address bar shows the URL: `https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/oml/index.html?ojr=notebookedit%3Bbid%3D481%...`. The page header includes the Oracle Machine Learning logo and the project name "P2\_OML Project". The main content area is titled "P2 Deployment" and shows the status "FINISHED". The deployment notes are as follows:

**Project 2: Rain Forecast Six Hours Into The Future**  
**Deployment using OML Jobs and REST API**  
**Last Update: 27-Apr-2023**

**Key Deployment Notes:**

- Set the unique METARS weather station ID in the second code cell below.
- This entire notebook will be deployed, using the OML Jobs facility
- The job should be run no more often than once every 20 minutes.
- The job will reference and run this ENTIRE notebook.
- The notebook will INSERT a single row into the prediction table.
- The REST API will simply SELECT on the prediction table, returning the results.
- Therefore, this notebook must be able to be run top-to-bottom.

Here's how this deployment process works:

- The deployment model must already exist.
- The weather\_observations\_all table must already exist.
- The existing training data has already been preprocessed, so minimal processing is needed.
- New METAR is loaded for the station we want to predict the rain.
- New data must be preprocessed, just like the training data.
- Make the prediction
- Store the results in the results table (remove existing result for this station\_id)



# Make The Prediction

- **SELECT** makes the forecast
- **INSERT** into results table
- **SELECT** returns the prediction

## P2 Deployment

```
-- Get the station_id from the preprocess table
declare
v_station_id varchar2(20);
begin
select station_id
into v_station_id
from weather_preprocess
FETCH FIRST 1 ROWS ONLY;

-- Delete any existing previous prediction
delete
from weather_prediction_results
where station_id = v_station_id;

-- Make the prediction using the existing "ALL features" model
insert into weather_prediction_results
select v_station_id station_id,
(systimestamp at time zone 'US/Pacific') prediction_time_pacific,
prediction(MODEL_MLP_ALL_DEPLOY USING *) prediction_results,
prediction_probability(MODEL_MLP_ALL_DEPLOY USING *) prediction_probability
from WEATHER_PREPROCESS_PREDICT_v
order by primary_key desc
FETCH FIRST 1 ROWS ONLY;
end;
/

select station_id,
prediction_time_pacific
prediction_results,
prediction_probability
from weather_prediction_results;
```

# Create an OML Job to make forecast every 20 minutes.

The screenshot displays the Oracle Machine Learning (OML) interface. The top navigation bar includes the 'ORACLE Machine Learning' logo and the 'P2\_OML Project' workspace. The 'Jobs' section is active, showing a table of OML jobs. The table has columns for Name, Notebook, Owner, Last Start Date, Last Status, Next Run Date, Status, and Schedule. A job named 'P2 Predict Rain' is listed with a status of 'SCHEDULED' and a schedule of 'Every 20 min'.

	Name	Notebook	Owner	Last Start Date	Last Status	Next Run Date	Status	Schedule
<input type="checkbox"/>	P2 Predict Rain	P2_OML....	P2_OML	4/27/23, 1:44 PM	SUCCEEDED	4/27/23, 2:04 PM	SCHEDULED	Every 20 min



# Create A REST API enabling simple URL for forecast

Are we done?

The screenshot displays the Oracle REST API interface. The breadcrumb navigation shows the path: REST > Modules > Predict Rain > predict > GET. The main content area for the 'predict' endpoint includes a green circle with 'GET', the text 'Last Updated: 2 hours ago', 'No comments available', and 'Source Type: json/collection'. A green arrow points to the URL: `https://gcdda82236f9c80-adb11.adb.us-phoenix-1.oraclecloudapps.com/ords/p2_oml/rain/predict`. Below this, the 'Source' section shows an SQL query in a text editor:

```
1 select station_id,  
2 prediction_time_pacific pred_time,  
3 prediction_results pred_result,  
4 round(prediction_probability,3) pred_prob  
5 from weather_prediction_results
```

A red arrow points to the `round(prediction_probability,3)` expression in the query. To the right of the query editor, the 'Handler Parameters' and 'Implicit Parameters' sections are visible, with the latter showing parameters like `:body`, `:body_text`, and `:content_type`. At the bottom, a status bar indicates '8:45:27 PM - REST call resolved successfully.' and 'Powered by ORDS'.





# Reality Check : Challenges

- Use SQL\*Developer as much as possible: fast, stable and works.
- Essentially no OML online community.
- Documentation is very light on examples and difficult to understand.
- OML Notebooks are unstable, especially when switching between notebooks and other ADB Database Actions.
- Free Tier is slow, OML Notebooks are very unstable and can destroy an ADB environment.
- Paid ADB OML4SQL Jobs can be incredibly expensive... watch out.



Now it's your turn!

[www.orapub.com](http://www.orapub.com) for  
presentation, step-step  
instructions and notebooks.

[https://github.com/cshallahamer/  
o4sql-book/raw/main/weather-  
datasets](https://github.com/cshallahamer/o4sql-book/raw/main/weather-datasets)





# Get Trained!

## MACHINE LEARNING 1 Anomaly Detection



ORAPUB

## MACHINE LEARNING 2 Performance Prediction



ORAPUB

# Viscosity's Oracle ACEs

## The Oracle ACE Program

The Oracle ACE Program recognizes and rewards individuals for their contributions to the Oracle community.



**Charles Kim**  
CEO | Co-Founder

 @racdba

 ACE Director



**Rich Niemiec**  
Chief Innovation Officer

 @richniemiec

 ACE Director



**Craig Shallahamer**  
Applied AI Scientist

 @orapub

 ACE Director



**Sean Scott**  
Principal Consultant

 @oraclesean

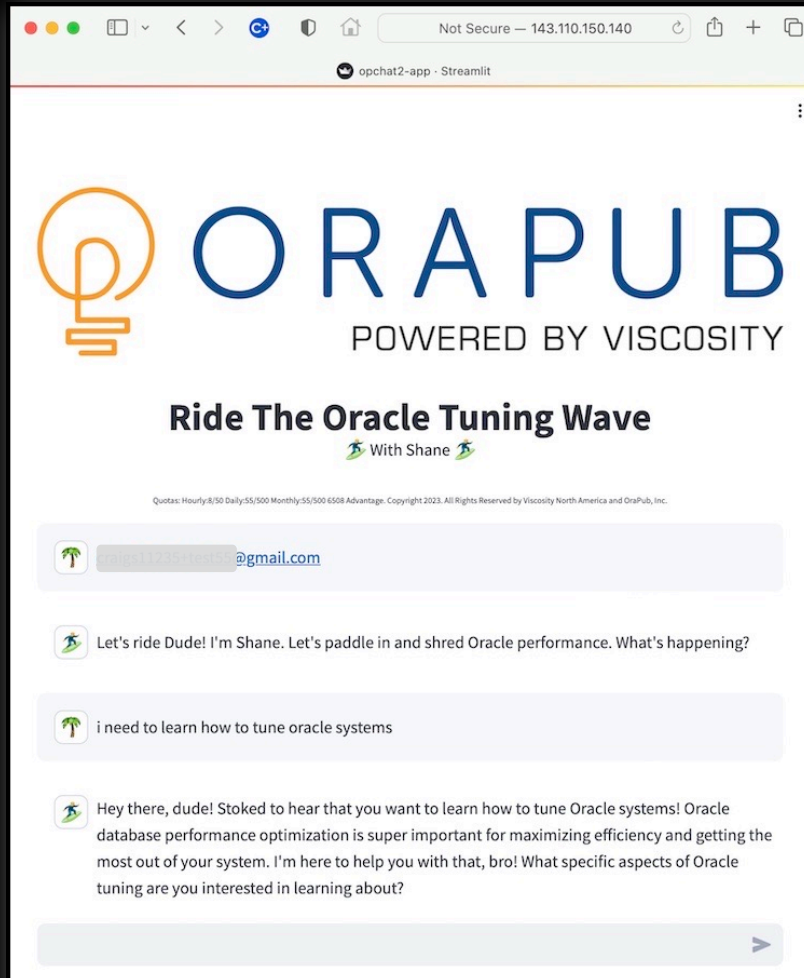
 ACE Director



**Gary Gordhamer**  
Principal Consultant

 @ggordham

 ACE Pro



# Want to try?

Focused on  
Oracle  
performance  
and internals.



# *Thank You!*



## From Data to Downpours: Oracle AI's Futuristic Adventure in Rainfall Forecasting!

AnDOUC TechCasts | Knowledge, Expertise, and Ideas for All  
September 7, 2023



ANALYTICS AND DATA  
**TechCasts**



Craig Shallahamer

Oracle ACE Director  
Applied AI Scientist | OraPub Founder

[@orapub](https://twitter.com/orapub)

[linkedin.com/in/craig-shallahamer-571a94a/](https://www.linkedin.com/in/craig-shallahamer-571a94a/)

[craig.shallahamer@viscosityna.com](mailto:craig.shallahamer@viscosityna.com)

[viscosityna.com](https://viscosityna.com)

[@ViscosityNA](https://twitter.com/ViscosityNA)