Comparative Analysis of Holt-Winters Algorithms on the Oracle Machine Learning Platform

Oleksandr Redych[1] and Ruslan Boichuk[2]
[1]*State Tax University, Universytetska Str. 31, 08205 Irpin, Ukraine*
[2]*Main Department of the State Tax Service in Ivano-Frankivsk region, Nezalezhnosti Str. 20, 76000 Ivano-Frankivsk, Ukraine*
redich2@gmail.com, if.dps@ukr.net

Keywords: Machine Learning, Oracle Cloud, APEX, Holt-Winters, Personal Income Tax.

Abstract: The Oracle Cloud Machine Learning toolkit was used in the work to study time series forecasting methods. The choice of tools was dictated by conducting research at the State Tax University (Irpin) for further use of the results in the work of regional and central divisions of the state tax service, whose information systems are developed on the Oracle platform. Holt-Winters exponential smoothing algorithms, currently represented by the functions of the application programming interface for machine learning models of the PL/SQL package DBMS_DATA_MINING, were investigated. The research used personal income tax data for 2015-2023, obtained from the tax office in the Ivano-Frankivsk region. Higher accuracy (MAE=3,57%) was shown by the algorithm of the Holt-Winters multiplicative exponential smoothing model with a fading multiplicative trend and multiplicative seasonality. Oracle Machine Learning for SQL implements exponential smoothing using a state of the art state space method that incorporates a single source of error (SSOE) assumption which provides theoretical and performance advantages. Access to research results is organized using the APEX web application creation tool. The considered toolkit will help in making decisions when assessing the base of the tax potential of the region and planning tax revenues. The results of the Holt-Winters exponential smoothing model algorithms research are presented and the losses of personal income tax in Ivano-Frankivsk region in 2022 are estimated.

1 INTRODUCTION

The information infrastructure of the Ministry of Finance of Ukraine, the State Tax Service, the National Bank and other central bodies of Ukraine is being developed on software products of the Oracle Corporation [1-3]. The core of the information systems of these institutions is the Oracle DBMS, which has proven itself to be a secure and reliable system. An important component of the operational activity of financial institutions is data analysis, detection of hidden patterns, forecasting, classification. This paper shows the results of the Holt-Winters exponential smoothing algorithm research on the Oracle Machine Learning platform for forecasting Personal Income Tax (PIT) revenues in the Ivano-Frankivsk region (Ukraine). The relevance of personal income tax research is determined by its high social and economic significance. Before the start of the war, during 2019-2021, the specific weight of personal income tax in the structure of tax revenues and revenues of the Consolidated Budget of Ukraine was more than 20%; in the structure of tax revenues of local budgets – more than 60% [4]. During the war in 2022-2023, the share of personal income tax in the total revenues of the Consolidated Budget decreased to 19,5% and 15,6%, respectively, which is largely explained by the mobilization of taxpayers and the absence of taxation of the financial support of conscripts.

Forecasting PIT revenues enables more accurate budgeting and expenditure planning for the government. This allows for the adaptation of budget priorities, rationalization of expenses, and ensuring financial stability. Forecasting tax revenue is a crucial tool for tax policy planning. Based on it, decisions can be made regarding the adjustment of tax rates, defining priorities in tax collection, and improving administrative procedures [5].

2 ORACLE MACHINE LEARNING PLATFORM

Oracle Cloud Machine Learning (OML) supports algorithms for data exploration, training and modeling with machine learning using SQL, R, Python, REST. OML includes more than 30 high-performance algorithms for working with databases that create models for use in applications [6]. It allows data analysts and other data professionals to quickly build models by automating key elements of the machine learning lifecycle. The technological advantage of using the OML platform is Oracle Corporation's focus on expanding cloud services, in particular, providing machine learning directly on the Oracle database. OML offers machine learning features such as classification, regression, clustering, feature extraction, anomaly detection, association (market basket analysis), time series, BigData, and more. Each machine learning function has several variants of calculation algorithms. OML uses built-in Oracle database features for maximum scalability and performance [7].

Using open source packages of R and Python, users can extend the set of methods and algorithms, as shown in the examples on the OML cloud environment (in the OML4Py and OML4R notebooks). In Oracle DB21c, Oracle Data Mining has been renamed to Oracle Machine Learning for SQL (OML4SQL), but the name of the PL/SQL package DBMS_DATA_MINING has not changed. The DBMS_DATA_MINING package exposes APIs that are leveraged by the Oracle Machine Learning for SQL. Users who wish to create machine learning models in their own schema require the CREATE MINING MODEL system privilege.

2.1 Holt-Winters Models in Time Series Forecasting

In its simplest form, exponential smoothing is a one-parameter moving average method that models the exponential decrease in the influence of past levels on future values. Winters [8] developed Holt's [9] exponential smoothing model with a trend and added seasonality to it. The advantage of this method lies in its ability to make forecasts over a long period. However, in order to make a forecast, for example, for 1 year, data for at least 2 complete years are required, preferably for 3-5 complete years.

Let's consider the Holt-Winters Triple Exponential Smoothing model, which takes into account three smoothing parameters: level, trend, and seasonality . This model uses additive trend and multiplicative seasonality [10]:

1) An additive trend implies that the trend changes over time with a constant absolute change. For example, if the trend value is 10, then at each time step it will increase by 10. This is suitable for data where trend changes occur with constant absolute values.
2) Multiplicative seasonality means that seasonal fluctuations change as a percentage of the baseline. For example, if the baseline data has a value of 100 and the seasonal fluctuation is 10%, then during the season the data will fluctuate between 90 and 110. This approach is more suitable when relative changes in seasonality are more significant than absolute values.

The Holt-Winters method introduces two types of smoothing: exponential smoothing of the level and exponential smoothing of the trend. These smoothing parameters are denoted as α (alpha) for the level, β (beta) for the trend, and γ (gamma) for seasonality. The following formulas are used in this Holt Winters model:

Level:

$$L(t) = \alpha * (Y(t) / S(t-L)) + (1 - \alpha) * (L(t-1) + T(t-1)),$$

Trend:

$$T(t) = \beta * (L(t) - L(t-1)) + (1 - \beta) * T(t-1),$$

Seasonality:

$$S(t) = \gamma * (Y(t) / L(t)) + (1 - \gamma) * S(t-L),$$

Forecast:

$$F(t+m) = (L(t) + m * T(t)) * S(t+m)$$

where: $L(t)$ is the smoothed value for the current period; $Y(t)$ is the actual value for the current period; $S(t)$ is the seasonality coefficient for the current period; $T(t)$ is the trend value for the current period; $F(t+m)$ forecast for m periods ahead; α, β, γ smoothing coefficients for: series level, trend and seasonality, respectively:

- Level: This represents the average value of the time series. The model assumes that the series fluctuates around a certain level. The level indicates the baseline level of the time series without the influence of trend and seasonality.

Table 1: Algorithms of the Holt-Winters exponential smoothing model.

#	Holt-Winters model algorithm	Value of the model type parameter
1	Holt-Winters triple exponential smoothing model, additive trend, multiplicative seasonality. Offering a robust model for data with both linear trend and proportional seasonal variation.	EXSM_WINTERS
2	Holt-Winters multiplicative exponential smoothing model with damped multiplicative trend, multiplicative seasonality. Effectively moderating the exponential increase or decrease of both trend and seasonal components over time.	EXSM_WINTERS_MUL _TREND_DMP
3	Holt-Winters multiplicative exponential smoothing model with damped trend, additive trend, multiplicative seasonality. Moderating the linear trend over time while still capturing proportional seasonal changes.	EXSM_WINTERS_DAM PED

- Trend: This denotes the direction and speed of changes in the time series. It indicates whether the series is increasing, decreasing, or remaining constant over time. The trend can be linear or nonlinear.
- Seasonality: This refers to regular oscillations or patterns in the time series that repeat at fixed intervals. For example, annual seasonal effects or monthly fluctuations.

A total of 14 exponential smoothing algorithms are supported. OML implements exponential smoothing using a state-of-the-art state-space method that incorporates the Single Source of Error (SSOE) assumption, providing theoretical and operational advantages. An appealing feature of SSOE is that the estimates of the state variables converge in probability to their true values, thereby leading to a formal inferential structure for the ad-hoc exponential smoothing methods for forecasting [11].

The maximum accuracy of the forecast in OML is achieved by successive selection of smoothing coefficients of series, trend and seasonality α, β, γ in the range from 0 to 1. This type of model considers various combinations of additive and multiplicative trend, seasonality and error, with and without trend attenuation. With various extensions, exponential smoothing covers a wider class of models than the Box-Jenkins auto-regression integrated moving average (ARIMA) approach.

2.2 Setting Parameters of the Holt-Winters Model for Forecasting

In this work, OML exponential smoothing algorithms for SQL ("*OML4SQL Time Series ESM")* are investigated on PIT data for 2015-2023(Jan-May) and for its forecasting in 2023-2024. The "OML4SQL Time Series ESM" algorithm was selected because it was the sole time series forecasting algorithm available on the OML platform at the time of the study. Examples of OML algorithms of models on training datasets are provided in the form of notebooks (OML Notebooks), which can be taken as a basic version and used. This paper highlights the results of the three most accurate exponential smoothing algorithms studied (Table 1).

The studied time series is the personal income tax data from January 2015 to May 2023 as of the 1st of each month (PIT1523 time series table with 101 values). Date - column of dates of receipts and column PIT - actual receipts of personal income tax. The analysis of the series showed the presence of a linear trend and seasonality after 12 months, so in the Holt-Winters forecasting model, the value of the seasonality parameter can be set to 12 and the forecast window can also be specified as 12 months.

The main stage is setting the parameters of the selected model. The parameter ALGO_NAME defines the main algorithm of the model, which we have exponential smoothing

```
- ALGO_EXPONENTIAL_SMOOTHING.
```

The EXSM_MODEL parameter defines the algorithm type itself. In OML, the Triple Exponential Holt-Winters Smoothing Model algorithm is called "EXSM_WINTERS". So, we set the parameters of the EXSM_WINTERS model (Figure 1):

- Sets the interval of the data set or the interval size unit, such as day, week, month, and so on. This setting applies only to a time column with a date and time type. For example, we need to provide for monthly PIT, so we set this parameter to `EXSM_INTERVAL _MONTH`. The values of the parameters are taken from the OML documentation [12]:

```
v_setlst('EXSM_INTERVAL')
:= 'EXSM_INTERVAL_MONTH';
```

'EXSM_WINTERS' Create a Holt-Winters Model using monthly parameters FINISHED

```
%script

BEGIN DBMS_DATA_MINING.DROP_MODEL('PIT1523');      -- name of the model
EXCEPTION WHEN OTHERS THEN NULL; END;
/
DECLARE
    v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN

    v_setlst('ALGO_NAME')             := 'ALGO_EXPONENTIAL_SMOOTHING'; -- name of the algorithm
    v_setlst('EXSM_INTERVAL')         := 'EXSM_INTERVAL_MONTH';        -- accumulation int'l = month
    v_setlst('EXSM_PREDICTION_STEP')  := '12';                         -- prediction step = 12 months
    v_setlst('EXSM_MODEL')            := 'EXSM_WINTERS';               -- ESM model Holt-Winters
    v_setlst('EXSM_SEASONALITY')      := '12';                         -- seasonal cycle = 12 months

    DBMS_DATA_MINING.CREATE_MODEL2(
        MODEL_NAME          => 'PIT1523',                  -- name of the model
        MINING_FUNCTION     => 'TIME_SERIES',              -- Oracle Data Miining  function = TIME_SERIES
        DATA_QUERY          => 'select * from PIT1523',    -- SELECT from table PIT1523
        SET_LIST            => v_setlst,                   -- reference table
        CASE_ID_COLUMN_NAME => 'Date',                     -- Date colamn from  PIT1523
        TARGET_COLUMN_NAME  =>'PIT');                      -- PIT data from a time series table PIT1523
END;
```

Figure 1: The Holt-Winters algorithm EXSM_WINTERS with the set parameters in the OML notebook.

- Setting the forecast window parameter. If we want to display each value representing a quarter, then a value of 4 gives four values (quarters) of forecasting for the future. For forecasting monthly PIT, this parameter is set to "12":

```
v_setlst('EXSM_PREDICTION_STEP'):=
'12';
```

Setting the algorithm of the exponential smoothing model to be used.

```
v_setlst('EXSM_MODEL'):='EXSM_WINTE
RS';
```

- Set the seasonality parameter. The parameter defines a positive integer value as the duration of the seasonal cycle. The value it takes must be greater than 1. For example, 4 means that each group of four values forms a seasonal cycle (a quarter for monthly data), which makes sense if you use 4 quarters to represent a year. In our case, when forecasting PIT, we set this parameter to 12.

```
v_setlst('EXSM_SEASONALITY'):=
'12';
```

- Setting the parameter of the method for processing missing values. In time series, the special value EXSM_MISS_AUTO indicates that if the series contains missing values, it should be treated as an invalid time series.

```
v_setlst('EXSM_SETMISSING'):=
'EXSM_MISS_AUTO';
```

- Also sets the values of the main parameters that point to the target value column and date column, etc:

```
CASE_ID_COLUMN_NAME => 'Date' and
TARGET_COLUMN_NAME => 'PIT';
```

By default, the confidence level is set to 0.95 and the EXSM_NMSE parameter is set to 3, which determines the window used to calculate the average mean square error (AMSE) metric.

2.3 Evaluation of Results

The tables of views listed in Table 2 contain information about the input data, model quality assessment, and global information. The names of the views include the model name "PIT1523".

Table 2: Tables views of the Holt-Winters model results.

#	Views with the model Details	Description
1	DM$VGPIT1523	Global Name-Value Pairs
2	DM$VPPIT1523	Exponential Smoothing Forecast
3	DM$VSPIT1523	Computed Settings
4	DM$VWPIT1523	Model Build Alerts
5	DM$VRPIT1523	Time Series Regression Build
6	DM$VGPIT1523	Global Name-Value Pairs

Table 3: Estimating the parameters of ESM algorithms.

#	Description	Name	Algorithm (see Table 1)		
			1	2	3
1	Smoothing constant, α	ALPHA	3860,9018	3849,8462	3855,8791
2	Trend smoothing constant, β	BETA	0,0001	0,0001	0,015
3	Seasonal smoothing constant, γ	GAMMA	0,0001	0,0001	0,0001
4	Akaike information criterion	AIC	3867,3779	3857,2197	3863,2526
5	Bayesian information criterion	BIC	3902,7437	3894,3033	3900,3361
6	Model mean absolute error	MAE	0,0455	0,0412	0,0433
7	Model standard error	STD	0,0565	0,0531	0,0551

These views are generated in the user schema and can be used to display modeling results in an APEX-based application. More details about Exponential Smoothing model parameters here [12].

Table 3 shows the estimated parameters of the model from the views DM$VGPIT1523 for all three algorithms when forecasting data for the period from June 2023 to June 2024. Some explanations for the estimates:

- *Akaike's Information Criterion* (AIC) is a measure of the trade-off between model accuracy and complexity. A lower AIC value indicates a better model.
- *Bayesian Information Criterion* (BIC). BIC is an alternative information superiority criterion that penalizes more complex models. A lower BIC value indicates a better model.

As can be seen from Table 3, by most estimates, the best algorithm for our time series is the Holt-Winters Multiple Exponential Smoothing Model with Damped Multiple Trend, Multiple Seasonality (EXSM_WINTERS_MUL_ TREND_DMP).

Given sufficient accuracy, this algorithm was used to estimate the decrease in PIT in 2022 due to the armed aggression of the Russian Federation, since the model was trained on pre-war time series and the predicted values could correspond to the peaceful state of the economy. The forecasting results for 2021-2023 and the absolute errors are shown in Table 4.

As can be seen from the table, the absolute error of the simulated PIT revenues for 2021 is 0,65%, and the forecast error for 2022 is 11,26%, which is explained by the fact that the model could not predict the war and its impact and the model error, in fact, determines the amount of PIT losses. Probable PIT losses in 2022 may reach an average of 674 ± 1% million UAH. An estimate of the simulated PIT values for 5 months of 2023 gives a 6.4% absolute forecast error. It should be noted that the mean absolute error (MAE) shows that the mean absolute difference between the observed values and the predicted values for all predictions is < 4%, which is a good result.

Figure 5 shows a graph of actual and simulated PIT data for 2015-2024 with confidence intervals (confidence level 0,95), obtained in the OML notebook (Holt-Winters model algorithm #2, Table 1). You can see significant differences between the forecast values and the actual values at the beginning of the pandemic (2020) and during the armed aggression of the Russian Federation. However, the following values are already taken into account by the model and the quality of the forecast improves.

As already mentioned, the simulation results can be used when creating analytical applications using Oracle APEX (Application Express). Application Express is a supported feature of the Oracle Database and is included, at no additional cost, with every Oracle Database, both on-premises and in the cloud. There are no additional licensing costs based on the number of developers, applications or end-users. Application Express is also included with every Oracle Database Cloud Service, from the low-priced Oracle Database Exadata Express Cloud

Table 4: Estimating actual and forecast revenues PIT.

Year	Actual revenues PIT, UAH	Forecast of revenues PIT, UAH	Absolute error,per year, UAH	Relative error of revenue PIT per year, %	MAE
2021	5 836 364 394,25	5 798 285 924,96	-38078469,29	0,65%	0,0367
2022	5 986 078 790,44	6 659 949 560,32	673870769,88	-11,26%	0,0357
2023 5 months	2 470 910 417,81	2 628 752 677,00	157842259,19	6,40%	0,0397

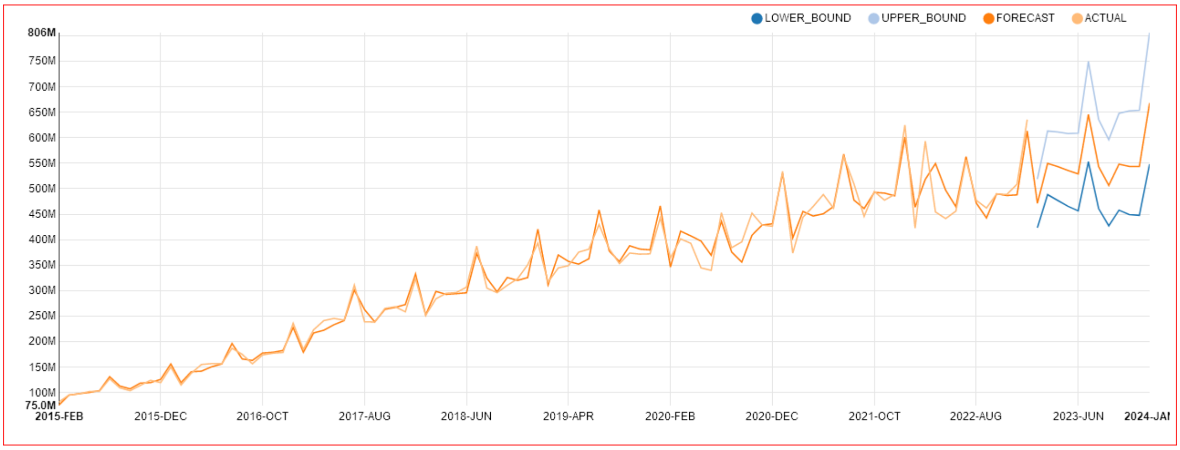

Figure 5: Chart of actual and forecast values of PIT in Ivano-Frankivsk region 2015-2024 (confidence level of 0.95).

Service all the way up to the Oracle Database Exadata Cloud Service [13].

Features of creating an Apex schema user to access OML:

1) Create the Workspace and the user schema on APEX.
2) When creating a Workspace, a user schema with the same name is automatically created in Cloud DB (ATP or ADW).
3) Register as an Oracle Cloud Database Administrator and go to the user administration tab.
4) Select the created user, activate the access to OML button in the edit menu.

For example, in the view DM$VPPIT1523, the model forms columns of actual, forecast and calculated values for the lines of the confidence level specified in the parameters (lower_bound and upper_bound). All calculated values of the model coefficients and detailed information can also be displayed on the APEX pages.

3 CONCLUSIONS AND FURTHER RESEARCH

The following results were obtained in the study:

1) Among the 14 Holt-Winters algorithms implemented on the OML platform and used in the forecasting of PIT receipts (data from the Ivano-Frankivsk region for 2015-2023), the best forecasting accuracy was shown by the algorithm of Holt-Winters multiplicative exponential smoothing model with damped multiplicative trend and multiplicative seasonality (MAE = 3,57%).
2) The effectiveness of the algorithms can be explained by the use of an extended state space method for exponential smoothing, including the assumption of a single source of error (SSOE).
3) The results of modeling tax revenues in 2022 indicate that estimated losses of PIT in Ivano-Frankivsk region. due to the aggression of the Russian Federation, they may amount to UAH 654± 1% million UAH.
4) APEX, which is a supported and free feature of the Oracle database, allows fast, low-code creation of applications using simulation results. This enables convenient access to the OML toolkit and simulation results for a wide range of employees of corporations, enterprises, regulatory agencies, in particular, divisions of the State Tax Service, whose information systems are being developed on the Oracle platform.

Further research will be focused on exploring the algorithms of Oracle Machine Learning for Python (OML4Py), which enable the utilization of specialized libraries for data processing, machine learning, and graphical analysis available through the Oracle Autonomous Database service. OML4Py is a module that allows users to manipulate data in Oracle database tables using Python syntax. OML4Py functions and methods transparently transform selected Python function sets into SQL commands for execution directly on the Oracle database.

Subsequent investigations are planned to consider the integration of OML, REST, and Oracle Big Data services for a wide range of data collection and utilization tasks, validation of the obtained results on tax revenue data from other regions, utilization of the discussed toolkit in studies of tax policy, and the development of methodological recommendations for the implementation of machine learning technologies in the operational activities of tax authorities [14, 15].

REFERENCES

[1] O. Puhachenko and T. Fomina, "Information support for expert studies and tax audits of the procedure for calculating and paying taxes and fees," Economics and society, no. 52, 2023, [Online]. Available: https://doi.org/10.32782/2524-0072/2023-52-72.

[2] T.V. Donchenko and A.V. Oliynyk, "Stages of formation of the tax system of Ukraine in the context of the development of the digital economy," Collection of scientific works of the State Fiscal Service University of Ukraine, no. 1, pp. 45-62, 2019, [Online]. Available: https://core.ac.uk/download/pdf/287302574.pdf.

[3] S.P. Kanzyuba, R.M. Matviychuk, Y.M. Sydorovych, and P.M. Musienko, "Electronic document flow. Reengineering of administrative processes in public authorities," Electronic governance and electronic democracy. Education manual, vol. 9, 64 p., 2017, [Online]. Available: https://books.google.de/books?id=paW4DwAAQBAJ&printsec=frontcover&hl=uk&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

[4] O. Totska and I. Dmytruk, "Individual income tax in Ukraine: national and regional dimension," FCS, vol. 1, no. 8, pp. 30-39, Mar. 2023, [Online]. Available: https://doi.org/10.26565/2786-4995-2023-1-04.

[5] E.I. Maslennikov and A.O. Gusev, "Income tax of individuals in Ukraine: current state and ways of reform," Black Sea Economic Studies, no. 57, pp. 47-50, 2020, [Online]. Available: https://doi.org/10.32843/bses.57-8.

[6] Oracle, "Oracle Machine Learning for SQL Use Cases," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/21/mlsql/machine-learning-functions.html#GUID-E5D374FB-F7D5-49ED-BBD4-272B8B8CADC2.

[7] A. Png and H. Helskyaho, "Oracle Machine Learning in Autonomous Database. In: Extending Oracle Application Express with Oracle Cloud Features," Apress, Berkeley, CA, pp.139-191, 2022, [Online]. Available: https://doi.org/10.1007/978-1-4842-8170-3_5.

[8] P.R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, vol. 6, no. 3, pp.324-342, 1960.

[9] C.E. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Averages," ONR Memorandum, no. 52, Carnegie Institute of Technology, Pittsburgh, USA, 1957. Available from the Engineering Library, University of Texas at Austin, 2003. Published in International Journal of Forecasting, vol. 20, pp. 1-13, 2004, with additional commentaries.

[10] C. Chatfield, "The Holt-Winters Forecasting Procedure," Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 27, no. 3, pp. 264-279, 1978, [Online]. Available: https://doi.org/10.2307/2347162.

[11] J. K. Ord, et al., "Time Series Forecasting: The Case for the Single Source of Error State Space Approach," Working Paper, Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia, April 2005, [Online]. Available: https://robjhyndman.com/papers/SSOE.pdf.

[12] Oracle, "DBMS_DATA_MINING - Algorithm Settings: Exponential Smoothing," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/oracle-database/21/arpls/DBMS_DATA_MINING.html#GUID-A95A0A38-8A5A-4470-B49F-80D81C588BFC.

[13] Oracle, "Oracle® APEX," 2023, [Online]. Available: https://docs.oracle.com/en/database/oracle/apex/23.1/htmrn/index.html#Oracle%C2%AE-APEX.

[14] L. Canali, Z. Baranowski, and P. Kothuri, "Integration of Oracle and Hadoop: Hybrid Databases Affordable at Scale," Journal of Physics: Conference Series, vol. 898(2), 2017, doi: 10.1088/1742-6596/898/4/042055.

[15] A. Png and H. Helskyaho, "Oracle Machine Learning REST Services. In: Extending Oracle Application Express with Oracle Cloud Features," Apress, Berkeley, CA, pp.193-232, 2022, [Online]. Available: https://doi.org/10.1007/978-1-4842-8170-3_6.