**Scenario:** I am taking the state wise startup company’s expenditure (R&D Spend, Administration Spend, and Marketing Spend) and profit data.

The goal is to predict the Profit for the given set of expenditure values. I am not explaining details about the ML Algorithm and the parameter tuning here. I would like to show the end to end process of Data extraction from SAP HANA DB, analyzing, cleaning, feature selection, and applying machine learning model and finally write back the results and ML algorithm performance metrics to the HANA tables.

The linear regression is the most commonly used model in research and business and is the simplest to understand, so using the random forest regression method we will predict the Profit.

The below diagram shows the ML Prediction life cycle and steps followed in the use case.

The basic steps involved in this process are:

1. Check the HANA Table data and analyze it using SQL in HANA Studio/WEB IDE.

(Make sure you have required privileges to do DML operations on the tables in SAP HANA Database.)

2. Import Pyodbc, Pandas, Sklearn, Matplotlib, seaborn libraries in python.

3. Create a connection to the HANA database and execute the required SQL.

4. Extract all the historical data into the data frame object and start analyzing it in Python using pandas.

5. Do the feature engineering, data cleaning and then feed the final set of independent variables to Machine learning algorithm (Random Forest) to predict the dependent variable (Profit).

6. Analyze the Machine learning algorithm metrics and fine-tune for better accuracy by repeating step 5. Store the Machine learning algorithm metrics in log table and also update the predicted value of historical data into the HANA Table.

7. For the new data set, create the python program which reads the new data using the Pyodbc connection and predicts the dependent variable (Profit) and updates the actual transactional table for reporting.

8. Schedule this program and keep monitoring the model metrics and predicted value.

## **1. Check the HANA Table data and analyze it using SQL**

I have created two tables, one contains the actual company wise data which is used to store the transactional data. The second one to store the metrics of the Machine learning algorithm (Ex: MAE, R Squared, MAPE, RMSE, Accuracy, etc.)

Please find the structures of both tables below.

Main transactional data table structure.

# 

# 

# **References**

<https://blogs.sap.com/2019/04/05/new-r-and-enhanced-python-api-for-sap-hana-machine-learning-released/>

<https://blogs.sap.com/2018/04/06/whats-new-in-sap-hana-2.0-sps-03-by-the-sap-hana-academy/>

<https://datatofish.com/python-script-windows-scheduler/>

<https://github.com/dbader/schedule>

<https://blogs.sap.com/2018/10/29/python-client-api-for-machine-learning-in-sap-hana-2.0-express-edition-sps-03-revision-33/>

<https://blogs.sap.com/2018/04/20/enhancements-to-external-machine-learning-in-sap-hana-2.0-sps-03/>

<https://help.sap.com/viewer/42668af650f84f9384a3337bcd373692/2.0.02/en-US/b2a37c7ecec2416bbf1889b2f2883ade.html>

<https://towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706>