Department of Information Technology

# Department of Information Technology Data warehouse and data mining (BTIT-706-18)



**Submitted To: Submitted By:**

Dr. Amanpreet Kaur Sundram Kumar

1802644

IT Z

# Chandigarh Group of College Landran, Mohali, Punjab

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Name of Experiment** | **Page No.** | **Remarks** |
| **1.** | To Explore Rapid Miner Tool capabilities for  Data Science |  |  |
| **2.** | To Explore Oracle Data Miner for Data Science |  |  |

**Experiment -1**

**Aim -** To Explore Rapid Miner Tool capabilities for Data Science

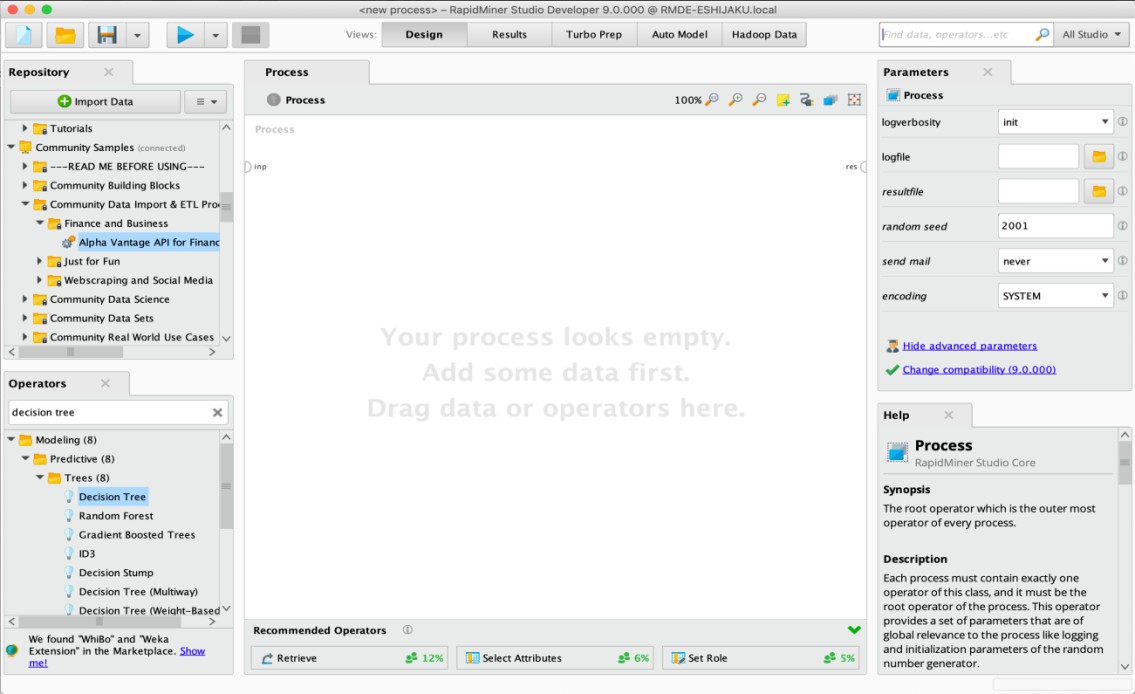
### Introduction-

RapidMiner Studio combines technology and applicability to serve a user-friendly integration of the latest as well as established data mining techniques. Defining analysis processes with RapidMiner Studio is done by drag and drop of operators, setting parameters and combining operators. The process structure is described internally by XML and developed by means of a graphical user interface. In the background, RapidMiner Studio constantly checks the process currently being developed for syntax conformity and automatically makes suggestions in case of problems. This is made possible by the so-called meta data transformation, which transforms the underlying meta data at the design stage in such a way that the form of the result can already be foreseen and solutions can be identified in case of unsuitable operator combinations (quick fixes). In addition, RapidMiner Studio offers the possibility of defining breakpoints and of therefore inspecting virtually every intermediate result. Successful operator combinations can be pooled into building blocks and are therefore available again in later processes. RapidMiner Studio contains more than 1500 operations altogether for all tasks of professional data analysis, from data partitioning, to market-based analysis, to attribute generation, it includes all the tools you need to make your data work for you. But also methods of text mining, web mining, the automatic sentiment analysis from Internet discussion forums (sentiment analysis, opinion mining) as well as the time series analysis and -prediction are available. RapidMiner. **RapidMiner** is a [data](https://en.wikipedia.org/wiki/Data_science) [science](https://en.wikipedia.org/wiki/Data_science) software platform developed by the company of the same name that provides an integrated environment for [data preparation,](https://en.wikipedia.org/wiki/Data_pre-processing) [machine learning](https://en.wikipedia.org/wiki/Machine_learning), [deep learning](https://en.wikipedia.org/wiki/Deep_learning), [text mining,](https://en.wikipedia.org/wiki/Text_mining) [and predictive](https://en.wikipedia.org/wiki/Predictive_analytics) [analytics.](https://en.wikipedia.org/wiki/Predictive_analytics) It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results [visualization,](https://en.wikipedia.org/wiki/Information_visualization) model validation and optimization.



### Feature of rapidMiner:

1. Application and Interfaces.
2. Data Access
3. Data Exploration
4. Data Prep
5. Modeling
6. Validation
7. Scoring
8. Code Control



### Elements of Rapid miner:

1. **Process**

A connected set of Operators that help you to transform and analyze your data.

### Ports

To build a process, you must connect the output from each [Operator](https://docs.rapidminer.com/9.9/studio/getting-started/design-view.html#operators) to the input of the next via a port*.* To connect two ports, click on them. Hover a port to see a tooltip with additional information. When connecting two Operators, you need to make sure that the output port of the first is compatible with the input port of the second, or you will get an error message.

### Repository

Repository is a central data storage entity. It holds connections, data, processes and results, either locally or remotely*.*

### Operators

The elements of a Process, each Operator takes input and creates output, depending on the choice of parameters

### Keyword Search

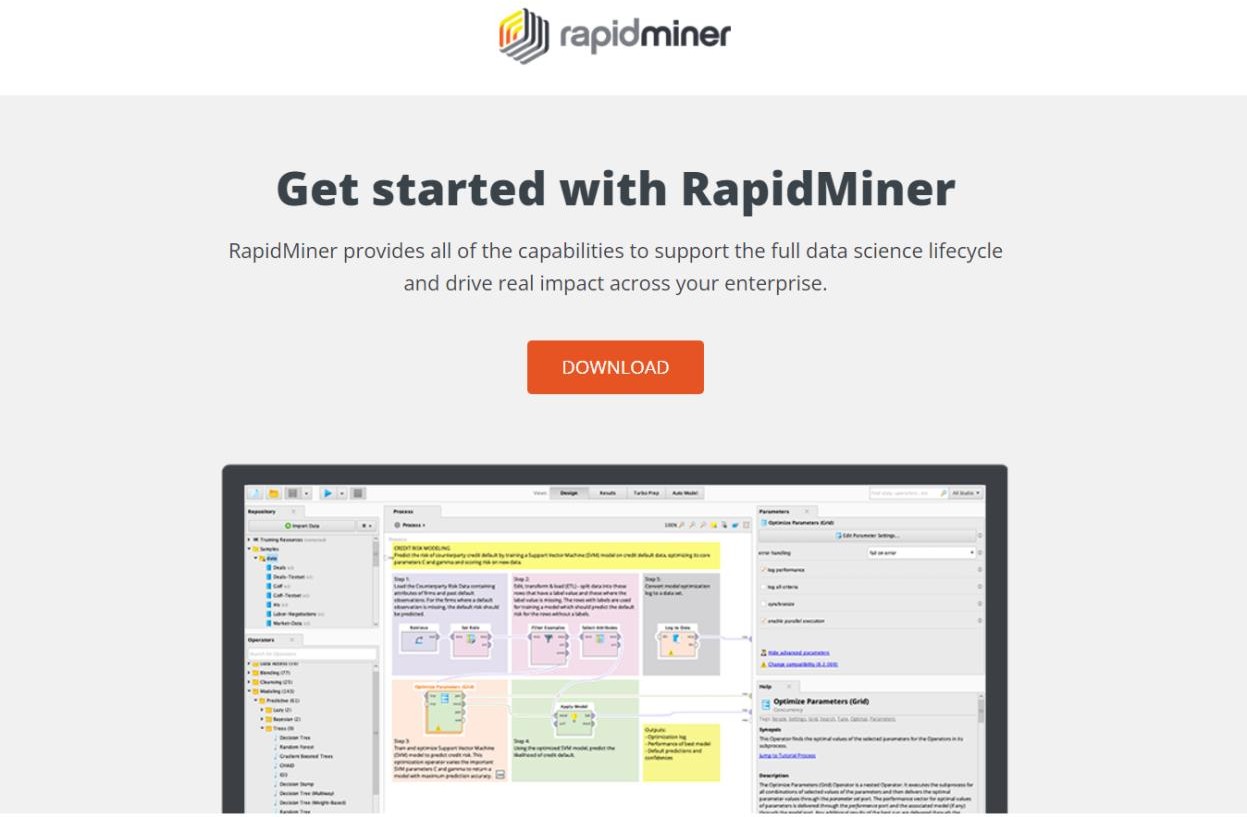
The alternative is keyword search. Although the Operators Panel includes a search field, the recommended procedure is to use the [global search](https://docs.rapidminer.com/9.9/studio/getting-started/global-search/), in the upper right corner of the user interface. The global search finds not just Operators, but data and processes from the Repository, extensions from the [Marketplace,](https://marketplace.rapidminer.com/) and even actions you can take from the menu.

### Parameter

The content of the Parameters Panel is context-dependent. Select any [Operator](https://docs.rapidminer.com/9.9/studio/getting-started/design-view.html#operators) that is displayed in the [Process Panel,](https://docs.rapidminer.com/9.9/studio/getting-started/design-view.html#process) and the Parameters Panel displays the options for configuring that Operator. Because RapidMiner Studio includes many Operators, each with its own unique functionality, the range of parameters is also quite diverse. By default, RapidMiner .

### Installation Steps of Rapid Miner:

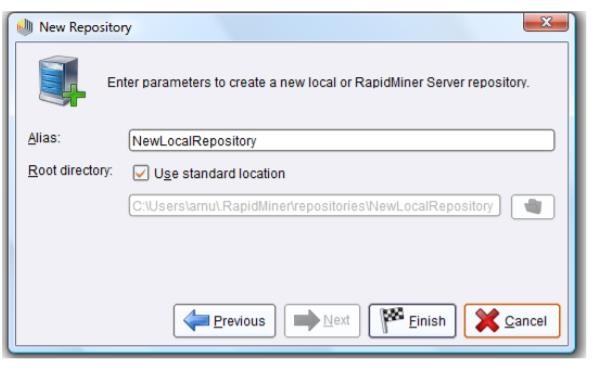
1. Download rapid miner from” <https://rapidminer.com/get-started/>” .



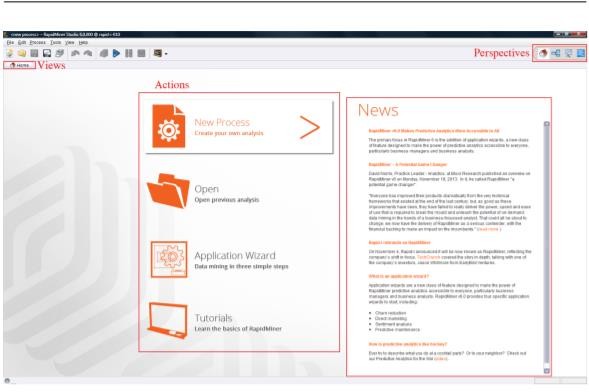
1. After downloading, while install the RapidMiner, windows prompt will open now click on run.
2. If you are downloading first time you will asked to create repository.Create a local repository on your computer to begin with the first use of RapidMiner Studio.



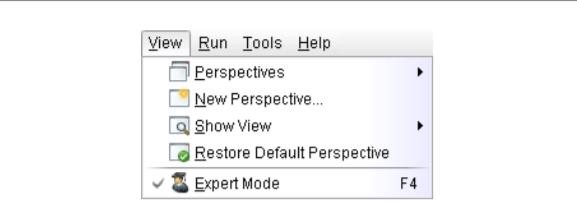
1. After clicking on next you will Define a new local repository for storing your data and analysis processes. It is advisable to create a new directory as a basis.



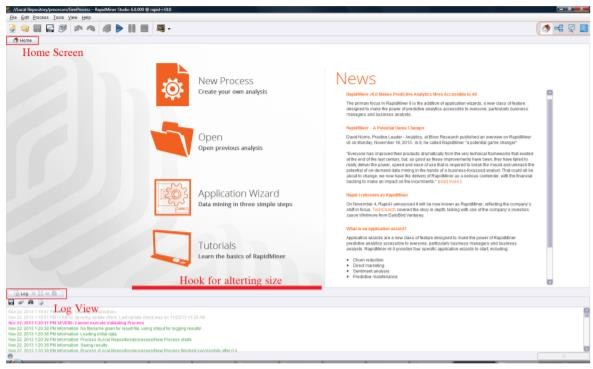
1. Click on finish button.
2. After clicking finish button a home perspective view of rapid miner is open.

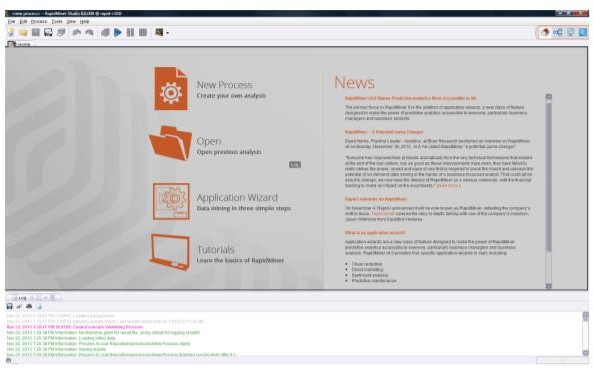


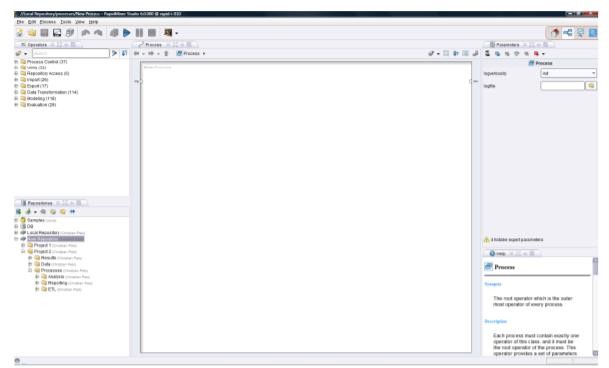
1. In the Home Perspective there is only one view, one preset at least, namely the home screen, which you are looking at now. You can activate further views by accessing the “View” menu.



1. In the subitem “Show View” you will find all available views of RapidMiner Studio. Views, which are now visible in the current perspective, are marked with a tick. Activate a further view by making a selection, for example the view with the name “Log”.You will now see that a second view with this name has been added in the Home Perspective.



1. Dragging the lower Log View to the middle and highlighting the new position.
2. Now switch to the Design Perspective by clicking in the toolbar.The Design Perspective is the central working environment of RapidMiner Studio. At last main window will open where you can perform your query or task.



### Application of Rapid Miner:

1. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization.
2. Create data driven port folios.
3. Easy to use visual environment for building analytics processes.
4. Every analysis is a process, each transformation or analysis step is an operator, making design fast, easy to understand, and fully reusable
5. Convenient set of data exploration tools and intuitive visualizations
6. Support for scripting environments like R, or Groovy for ultimate extensibility
7. Transparent integration with RapidMiner Server to automate processes for data transformation, model building, scoring and integration with other applications
8. Extensible through open platform APIs and a Marketplace with additional functionality
9. More than 1500 operators for all tasks of data transformation and analysis.
10. Powerful Global Search sifts through repositories to quickly retrieve anything, including processes, models, operators, extensions and even UI actions.

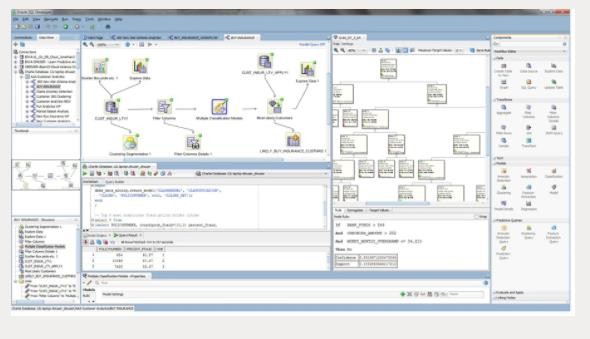
### References:

1. **Wikipedia**
2. **Rapidminer.com**
3. **https://academy.rapidminer.com/learning-paths/get-started-with- rapidminer-and-machine-learning.**
4. **https://rapidminer.com/products/studio/feature-list/**

**Experiment-2 Aim**- To explore Oracle Data Miner for Data Science

**Procedure**- It is a representative of the Oracle’s Advanced Analytics Database. Market leading companies use it to maximize the potential of their data to make accurate predictions. The system works with a powerful data algorithm to target best customers. Also, it identifies both anomalies and cross-selling opportunities and enables users to apply a different predictive model based on their need.

Further, it customizes customer profiles in the desired way. Oracle Data Miner enables data scientists, “citizen data scientists,” and business and data analysts to work directly with data inside the database using a graphical “drag and drop” workflow editor. Oracle Data Miner (ODMr), an extension to Oracle SQL Developer, captures and documents in graphical analytical workflows the steps users take while exploring data and developing machine learning methodologies. ODMr workflows are useful for re-executing analytical methodologies and for sharing insights with team members. ODMr generates SQL and PL/SQL scripts and offers a workflow API for accelerating model deployment throughout the enterprise.



### Features-

* Interactive workflow tool for creating, evaluating, modifying, sharing, and deploying machine learning methodologies
* Nodes from the ODMr tool palette
* Explore and Graph nodes for visualizing data - histograms, summary statistics, scatterplots, boxplots
* Transform node - supports popular and custom data transforms including binning and recoding variables, missing values treatment, and creating new “engineered features” based on user domain expertise to override Oracle Machine Learning automatic data preparation
* Column Filter node - uses an attribute importance / feature selection algorithm to identify the most influential attributes in supervised learning, and Kulback-Leibler divergence for unsupervised learning - identifying the strength of each attribute's correlation with other attributes
* Model Build node - automates common steps including creating a random sample for train and test datasets, automatic model testing and evaluation, computing a confusion matrix, lift chart, receiver operating characteristic (ROC) curve, and model statistics, with model visualizers including decision trees, cluster trees, and model attribute coefficients
* Ingest and process structured data in tables and views (numeric and varchar datatypes), unstructured data (CLOBs), transactional data, aggregations, and spatial and graph data
* Where multiple algorithms exist for given machine learning technique, the Model Build node automatically builds multiple machine learning models for comparison
* Integration with open source R for the execution of user-defined R functions at the database server, including data-parallel and task-parallel execution (see [Oracle Machine Learning for R](https://www.oracle.com/database/technologies/datawarehouse-bigdata/oml4r.html))
* Works with Big Data SQL to access data across the broad range of big data sources, including Oracle Database, Spark, Hadoop and other data sources.

### Installation Steps-

**Step One: Install the Database**

Perform a basic installation of Oracle Database Enterprise Edition and create a starter database.

1. In Windows Control Panel, Administrative Tools, choose Services. Stop any Oracle services that may be running on your computer.
2. From the Oracle Database installation directory, run SETUP.EXE to start Oracle Universal Installer. Follow the instructions as specified on the initial pages, then:
   * On the Select Installation Options page, choose Create and Configure a Database.
   * On the System Class page, choose Desktop Class.
   * On the Typical Install Configuration page, specify the installation directories, the database name, and the database system passwords.
3. The Installer performs prerequisite checks, displays summary information about the installation, and copies the Oracle executables and the starter database files to your computer.
4. The Installer configures Oracle Net to enable client connections.
5. The Installer starts Oracle Database Configuration Assistant to create the starter database.
6. Database Configuration Assistant displays summary information about the starter database. If you wish to use Oracle sample data or the Data Mining sample programs, unlock the SH account.
7. Exit the Installer. The starter database is now running, and Oracle Net is ready to accept client connections.
8. If necessary, restart the Oracle services that you stopped in step 1.

### Step Two: Verify the Data Mining Installation

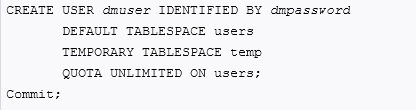
Once you have installed Oracle Database and created a starter database, you may wish to verify that Oracle Data Mining is functioning properly before proceeding further.

In Appendix A, you will find a series of SQL and PL/SQL commands that perform basic data mining operations. You can type these commands into SQL\*Plus or SQL Developer to make sure that the database is enabled for data mining.

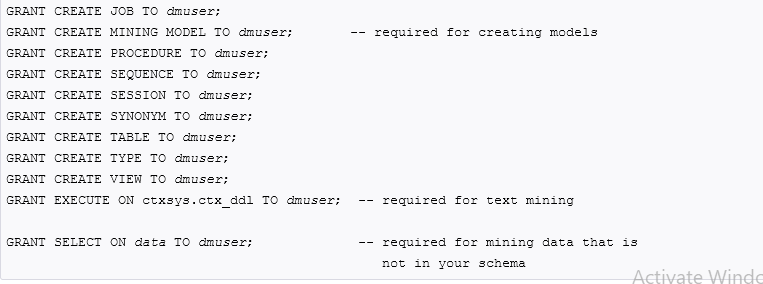
### Step Three: Create a Data Mining User

Create a database user with data mining privileges.

1. Log in to SQL\*Plus or SQL Developer as the system user.
2. Execute the following CREATE USER statement. Specify a user name for dmuser and password for dmpassword.



1. Execute these statements to grant data mining privileges to the user.



1. If the user will need to import or export data mining models, grant this additional privilege.



For import/export operations, the user must create a directory object. The user may also require additional privileges

### Step Four: If You Want to Use Oracle Data Miner

Oracle Data Miner, an extension to Oracle SQL Developer 3.0, is a graphical user interface to Oracle Data Mining. Oracle Data Miner uses a workflow paradigm to perform data mining tasks.

You can use Oracle Data Miner to explore data, build and evaluate multiple mining models, and apply the models to new data. By building workflows, you can capture and document the methodology you use to perform a range of mining tasks. You can save and share workflows.

To install and launch Oracle Data Miner:

1. Go to the following page on the Oracle Technology Network:

[http://www.oracle.com/technetwork/database/options/odm/dataminerworkflow-](http://www.oracle.com/technetwork/database/options/odm/dataminerworkflow-168677.html) [168677.html](http://www.oracle.com/technetwork/database/options/odm/dataminerworkflow-168677.html)

1. Follow the installation instructions provided on the page. The basic steps can be summarized as:
   1. Install Oracle Database.
   2. Install Oracle SQL Developer 3.0, which includes Oracle Data Miner 11gR2.
   3. Install the Oracle Data Miner repository.
2. For further instructions, use the Online Help provided by SQL Developer.
3. For additional assistance, use the Oracle By Example tutorials. Links to the tutorials are available with the installation instructions at the URL provided in Step [1.](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/quickstart.htm#BABEFBJC)

### Step Five: If You Want to Use the Spreadsheet Add-In

The Oracle Spreadsheet Add-In for Predictive Analytics lets you perform certain limited data mining operations in an Excel spreadsheet.

To install and launch the Spreadsheet Add-In:

1. Install Oracle Client and create a Net Service Name as described in ["Oracle Client and Oracle Net".](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/connecting.htm#BCGIAEFJ)
2. Download the Spreadsheet Add-In from the Oracle Technology Network. [http://www.oracle.com/technetwork/database/options/odm/odm-pred-analytics-addin- 092973.html](http://www.oracle.com/technetwork/database/options/odm/odm-pred-analytics-addin-092973.html)
3. Open the zip file and extract the contents to the Microsoft Office Library, typically: C:\Program Files\Microsoft Office\Office\Library
4. Open Excel and click Tools > Add–Ins. Select Oracle Predictive Analytics from the Add– Ins dialog box. The OraclePA menu is added to the Excel toolbar.
5. From the Add-In menu in Excel, choose Connect.
6. In the Connect (Oracle Database) dialog:
   * Select the Net Service Name that you created.
   * Provide the user name and password of the data mining user. Click Connect to launch the Spreadsheet Add-In.

### Step Six: If You Want to Use the Sample Programs

Follow these steps to install and configure the sample Data Mining programs:

1. Install Oracle Database Examples as described in ["Optionally Install Oracle Database Examples"](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/install_odm.htm#CACCBDHG).
2. The sample programs are copied to the RDBMS\demo subdirectory of the Oracle home directory. You can find the PL/SQL programs by searching for dm\*.sql.
3. Start SQL\*Plus and connect to the database as the system user. Run the dmshgrants script. Specify the full path to the Oracle home directory and the name of the data mining user.

### @ ORACLE\_HOME\RDBMS\demo\dmshgrants dmuser

1. Connect to the database as the Data Mining user. Run the dmsh script.

### CONNECT dmuser

**@ ORACLE\_HOME\RDBMS\demo\dmsh COMMIT;**

1. Once you have completed steps [1](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/quickstart.htm#BABJGJEG) through [4,](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/quickstart.htm#BABDDABF) you can run the PL/SQL programs.

### Applications-

Eliminate data movement, achieve big data scalability, preserve security, and accelerate time from model development to model deployment

 Move easily between Oracle Database environments to support development staging and production deployment scenarios for Oracle Machine Learning models and associated data assembly, transformation, and preparation scripts

Empower employees with a diverse skillset with in-database machine learning algorithms, enabling data-driven projects

Easy to use "drag and drop" user face accelerates knowledge discovery and model building for "citizen data scientists"

 Workflows document the machine learning methodologies developed for sharing and automation

 Generates SQL and PL/SQL scripts from workflows to automate and accelerate model deployment throughout the enterprise

Workflow API enables programmatic workflows invocation

### References-

* 1. [**https://docs.oracle.com/cd/E11882\_01/datamine.112/e16807/quickstart.htm# DMADM203**](https://docs.oracle.com/cd/E11882_01/datamine.112/e16807/quickstart.htm#DMADM203)
  2. [**https://www.oracle.com/database/technologies/datawarehouse- bigdata/dataminer.html**](https://www.oracle.com/database/technologies/datawarehouse-bigdata/dataminer.html)