CHAPTER 1

Oracle Data Miner
Understanding your data and gaining an insight into the behavior of your data (and hence your customers) have been goals that companies have been chasing for some time now. Myriad technology solutions are available to help you achieve this understanding and insight, but just as you accomplish those goals, some new challenge emerges to force you to look at new alternatives. One alternative that can help you achieve this deep insight into the behavior of your data is technology that makes use of some of the advanced machine learning algorithms. These are commonly referred to as data mining algorithms.

The data mining algorithms can be considered a process of searching data to discover patterns and trends that may have some competitive element to it. The following is one of the most commonly used definitions of data mining:

*Data mining is the non-trivial extraction of previously unknown and potentially useful information from data.*

Data mining can be used in addition to the traditional data analytics and statistical approaches to discover patterns in the data that could not be discovered using traditional approaches. The patterns that are discovered need to be evaluated to access their usefulness, as the results from data mining are not always useful.

One particular aspect of a data mining or a data science project that differentiates itself from other analytical or statistical approaches is that a very clear project problem definition is given. This allows the data mining or data science project to be directed by the problem definition to determine what needs to be achieved, and hence the desired outcome can be clearly defined. This approach can then be used to identify what tools and techniques are most appropriate to achieve these goals.

Data mining algorithms are implementations of advanced statistical and machine learning algorithms that can be used in certain problems on certain types of data to identify patterns that exist in the data. The data miner or data scientist can then examine these patterns to determine what they mean and relate them back to the business and the problem definition.

As the volume of data that companies capture and store in their database or the use of other data storage techniques increases, the need for these data mining techniques increases. This is particularly the case as companies begin to use data from their Big Data sources and as data increasingly becomes available from new data sources, such as the data that will become available with the Internet of Things (IOTs). Data mining will play a significant and central role with the management, integration, and delivery of decisions through the company.

Data mining is not a new technology. It has been around for many decades now, but started to become widespread in some industries, such as telecommunications, insurance, and finance, during the 1990s and 2000s. Data mining does not need to have a large volume of data before it can start delivering useful results. Many
companies have been able to achieve a competitive advantage by using a data set consisting of a few thousand records.

## In-Database Data Mining with Oracle Data Mining

In most typical data mining applications, the data in the database must be extracted from the database, transported to the data mining application server, loaded into the data mining application, and prepared for data mining; the data mining model is then produced and evaluated. This process occurs whenever you want to build a data mining model or to refresh an existing one. As you can imagine, the steps of extracting, transporting, and loading into the data mining application can take a considerable amount of time. As the data volumes in companies are growing at an ever-increasing rate, particularly when we consider the data that is available via Big Data sources, the time required to extract and move your data will become longer and longer. In addition to all of these steps, whenever you want to use the data mining models in a real-time scenario, such as giving a car insurance quote, the Risk Scoring data mining model will need to perform a set of steps similar to those just outlined. This will make the applications run seemingly at a snail’s pace. As the volume of data increases, the number of customers increases, and so on, the Risk Scoring data mining model will take longer and longer to run.

But what if you could eliminate most of these steps that involve the movement of data? What if you could do all of this in the database? After all, that is where the data will be!

With in-database data mining, the data mining algorithms are built into the database. You no longer need another application and another server to run the data mining applications. This is what Oracle has done with Oracle Data Mining (ODM). It has taken a suite of data mining algorithms and built them into the kernel of the Oracle Database. A common phrase associated with this is as follows:

…move the algorithms to the data, not the data to the algorithms.

With in-database data mining, there is no need to extract the data, move the data to the data mining application, and load the data into the data mining application. The data is already there in the database, so there is zero data movement. By eliminating the data movement, you can save a significant amount of computer processing time, as well as the human effort required to set it up and to manage the process. This can result in your data mining projects running in a fraction of the time that the alternative approach requires. Many companies are reporting such success stories. Some examples of these companies are listed in the section “Customer Success Stories” later in this chapter; several of these companies have actually reduced their data mining projects from weeks to minutes. Figure 1-1 illustrates the Oracle in-database advanced analytics proposition.
Predictive Analytics Using Oracle Data Miner

Data mining within Oracle Database offers many advantages:

- **No Data Movement** Most data mining products require that data be exported from a database, converted to a format required by the data mining tool, then loaded into the data mining tool before any database mining can be performed. With Oracle Data Mining, no data movement, data conversion, or loading of the data into another application is needed. This simplifies the entire data mining process, making it less complex and consuming significantly less time. In addition to giving the data scientist more time to work on the defined problem, it also allows the data scientists to work with a larger data set. As companies expand into the Big Data world, having very little or even no data movement can facilitate these projects being completed in a quicker time frame.

- **Security** The extensive security features of Oracle Database protect your data. This ensures that you can easily comply with all your data security and audit requirements at company, industry, and regulatory levels. Using and running the data mining algorithms within the Oracle Database require specific database privileges. This allows for full control over who is allowed to perform what activities on your data.

**FIGURE 1-1. The Oracle in-database advanced analytics proposition**
Data Preparation and Administration  Most data must be cleansed, filtered, normalized, sampled, and transformed in various ways before it can be mined. Up to 80 percent of the effort in a data mining project is often devoted to data preparation. Oracle developers have the key skills required to perform these tasks and can utilize the various in-database functionality to prepare the data for data mining, therefore eliminating the need for any external applications for data preparation. This also ensures that your company can easily comply with any data security and traceability requirements.

Ease of Data Refresh  The data mining processes within the Oracle Database already have ready access to data required and can easily produce a refreshed data set with minimum effort. This can allow the real-time scoring of data as it is being captured and can allow for highly efficient updating of your Oracle Data Mining models.

Oracle Database Analytics  The Oracle Database comes with many features for statistical analysis, including many specific data analytics methods that are common in business intelligence and data warehousing. As these features are all embedded in the Oracle Database, the Oracle Data Mining functions can easily utilize these features.

Oracle Technology Stack  You can take advantage of all aspects of Oracle's technology stack to integrate data mining within a larger framework for business intelligence, scientific inquiry, and Big Data analytics.

Application Programming Interfaces  The PL/SQL application programming interface (API) and SQL language operators provide direct access to Oracle Data Mining functionality in Oracle Database.

Oracle Advanced Analytics Option

The Oracle Advanced Analytics option comprises Oracle Data Mining and Oracle R Enterprise. The Oracle Advanced Analytics option is available as an extra license cost option with the Oracle Database Enterprise Edition. By combining the powerful in-database advanced data mining algorithms and the power and flexibility of R, Oracle has provided a set of tools that allows everyone, from the data scientist to the Oracle developer and the database administrator (DBA), to perform advanced analytics on their data to gain a deeper insight into their data as well as to achieve a competitive advantage over their competitors.

The following sections give an overview of the components of the Advanced Analytics option. In a later section, the architecture and the various elements of Oracle Data Mining will be explored.
NOTE

The Oracle Database Advanced Analytics option came into existence when Oracle released Oracle R Enterprise. Oracle Data Mining and Oracle R Enterprise comprise the Oracle Advanced Analytics option.

Oracle Data Mining

Oracle Data Mining contains a suite of advanced data mining algorithms that are embedded in the database that allows you to perform advanced analytics on your data. The data mining algorithms are integrated into the Oracle Database kernel and operate natively on data stored in the tables in the database. This integration removes the need for extraction or transfer of data into stand-alone mining/analytic servers, as is typical with most data mining applications. This can significantly reduce the time frame of data mining projects by requiring nearly no data movement.

In addition to the suite of data mining algorithms, which are listed in Table 1-1, Oracle has a variety of interfaces to enable you to use these algorithms. These interfaces include PL/SQL packages that you can use to build and apply models to new data,

<table>
<thead>
<tr>
<th>Data Mining Technique</th>
<th>Data Mining Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anomaly Detection</td>
<td>One Class Support Vector Machine</td>
</tr>
<tr>
<td>Association Rule Analysis</td>
<td>Apriori</td>
</tr>
<tr>
<td>Attribute Importance</td>
<td>Minimum Description Length</td>
</tr>
<tr>
<td>Classification</td>
<td>Decision Tree</td>
</tr>
<tr>
<td></td>
<td>Generalized Linear Model</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>Clustering</td>
<td>Expectation Maximization</td>
</tr>
<tr>
<td></td>
<td>K-Means</td>
</tr>
<tr>
<td></td>
<td>Orthogonal Partitioning Clustering</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>Non-Negative Matrix Factorization</td>
</tr>
<tr>
<td></td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td></td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>Regression</td>
<td>Generalized Linear Model</td>
</tr>
<tr>
<td></td>
<td>Support Vector Machine</td>
</tr>
</tbody>
</table>

**TABLE 1-1.** Oracle Data Mining algorithms available in Oracle 12c
a variety of SQL functions for real-time scoring of data, and the Oracle Data Miner tool that provides a graphical workflow interface for creating your data mining projects.

TIP
Oracle used to provide a set of Java APIs to the in-database data mining algorithms, but these are no longer supported.

TIP
Oracle Data Mining is the term used to describe the in-database data mining algorithms, functions, and procedures. Oracle Data Miner is the graphical workflow tool that comes as part of SQL Developer.

Oracle R Enterprise (ORE)
Oracle R Enterprise (ORE) was introduced in 2011 and enables the open source R statistical programming language and environment to be run on the database server and within the database (see Figure 1-2). Oracle R Enterprise integrates R with the Oracle Database. When Oracle R Enterprise was released, it, along with Oracle Data Mining, was repackaged to form the Oracle Advanced Analytics option.

While analysts interactively analyze data and develop R scripts, the data they will be using will reside in the database. With ORE data scientists can still write their R scripts and analyze the data, but now the data can remain within the Oracle Database. No data has to be downloaded to the data scientists’ computers, thus saving a significant amount of time and allowing data scientists to concentrate on solving the business problem at hand. By allowing the R scripts to run in the database, these scripts can utilize the ability of the database to manage the processing of many millions of records in an efficient manner and they can utilize

![Architecture of Oracle R Enterprise](image-url)
other database performance options, including the Parallel option. This capability overcomes many of the limitations of running R on the data scientists’ computers.

Oracle R Enterprise consists of the following components:

- **Transparency Layer**  The transparency layer is a collection of packages that map R data types to Oracle Database objects and generate SQL transparently in response to R expressions on mapped data types. The transparency layer allows R users to interact directly with database-resident data using R language constructs. This enables R users to work with data too large to fit into the memory of a user’s desktop system.

- **Statistics Engine**  This is a collection of statistical functions and procedures corresponding to commonly used statistical libraries. The statistics engine packages execute in Oracle Database.

- **Statistics Extensions**  This supports the R engine execution through the database on the database server. These SQL extensions enable R scripts to be run inside the database.

In addition to Oracle R Enterprise (ORE), Oracle also provides the following:

- **Oracle R Distribution**  This Oracle-supported distribution of open source R is provided as a free download from Oracle and also comes preinstalled on the Oracle Big Data Appliance. Oracle also provides some additional libraries to enhance the performance of certain functions. These libraries include Intel’s Math Kernel Library, AMD’s Core Math Library, and the Solaris Sun Performance Library.

- **ROracle**  This is an open source R package, maintained by Oracle and enhanced to use the Oracle Call Interface (OCI) libraries to handle database connections, that provides a high-performance, native C-language interface to the Oracle Database.

- **Oracle R Connector for Hadoop**  Oracle provides R users high-performance native access to the Hadoop Distributed File System (HDFS) and MapReduce programming framework. The Oracle R Connector for Hadoop is a component of the Oracle Big Data Connectors software suite.

### History of Data Mining in Oracle

Oracle Data Mining has been available in the Oracle Database since Oracle 9i R1, but it wasn’t until Oracle 9i R2 that Oracle Data Mining became an important feature. Originally Oracle Data Mining was called Darwin. Darwin was a data mining product developed by Thinking Machines back in the mid-1990s. Table 1-2
Chapter 1: Oracle Data Miner

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darwin</td>
<td>Mid-1990s</td>
<td>This product was first developed by Thinking Machines.</td>
</tr>
<tr>
<td>Darwin (Oracle)</td>
<td>1999</td>
<td>Oracle purchased Darwin from Thinking Machines in 1999 and continued to distribute Darwin. Work began to integrate the data mining algorithms into the next release of the database. At the same time, the Darwin team was renamed Oracle Data Mining. Also during this time, the Oracle Data Mining team developed Oracle Personalization, which was a web-based recommendation product.</td>
</tr>
<tr>
<td>Oracle 9i R2</td>
<td>May 2002</td>
<td>Although some data mining functionality was released with Oracle 9i R1, the main ODM release was with the R2 release of the database, with more algorithms, support for PMML, and automated binning. The API for using ODM consisted of a set of Java APIs and required the installation of an ODM server in the database.</td>
</tr>
<tr>
<td>Oracle 10.1g</td>
<td>February 2004</td>
<td>This release enhanced the existing algorithms and improved data processing. The DBMS_DATA_MINING PL/SQL package was introduced as an interface to the algorithms. A new Oracle Data Mining Client (DM4J) based on Oracle JDeveloper components enabled graphical specification of ODM objects and a graphic user interface for interacting with key Java objects and processes in the ODM server.</td>
</tr>
<tr>
<td>Oracle 10.2g</td>
<td>July 2005</td>
<td>The Decision Tree and One-Class SVM algorithms are added. A revised Java Data Mining (JDM) JSR-73–compliant Java API is provided. The DBMS_PREDICTIVE_ANALYTICS PL/SQL package is added. Oracle Spreadsheet Add-In for Predictive Analytics enabled Microsoft Excel users to mine their Oracle Database or Excel data using the automated methodologies provided by DBMS_PREDICTIVE_ANALYTICS; the Add-In is distributed on Oracle Technology Network. SQL functions are added that include PREDICTION and PREDICTION_PROBABILITY, among others. The first ODM graphical user interface (GUI) is provided for preparing data, building models, and applying to new data. This was a wizard type of application.</td>
</tr>
</tbody>
</table>

TABLE 1-2. How Oracle Data Mining Has Evolved Over Time (continued)
## Table 1-2. How Oracle Data Mining Has Evolved Over Time

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle 11.1g</td>
<td>September 2007</td>
<td>This release provided additional improvements in the algorithms with better automated and embedded data transformations. It also added multivariate linear regression and multivariate logistic regression. The Profile procedure was added to the <code>DBMS_PREDICTIVE_ANALYTICS</code> PL/SQL package. Additional SQL functions were added, including <code>PREDICTION_BOUNDS</code>, cost matrix in <code>PREDICTION</code>, and more details in <code>GET_MODEL_DETAILS</code> procedures.</td>
</tr>
<tr>
<td>Oracle 11.2g</td>
<td>September 2009</td>
<td>Support was added for the importing of external data mining models (linear and binary logistic regression) using PMML. The Java API was deprecated in this release. Support for native transactional data was added for Association Rules. The release also provided some additional small enhancements to the algorithms and the <code>DBMS_DATA_MINING</code> package.</td>
</tr>
<tr>
<td>SQL Developer 3</td>
<td>March 2011</td>
<td>SQL Developer 3 came with the first release of the Oracle Data Miner GUI that had the workflow functionality. This gave the user a graphical interface for querying and exploring the data, applying data transformations, building models, and applying models. The graphical workflow environment provided a simple interface for the SQL and PL/SQL data mining functions and procedures.</td>
</tr>
<tr>
<td>Oracle 12c</td>
<td>June 2013</td>
<td>This release added new and improved Clustering algorithms, new Principal Component Analysis algorithm, GLM models that now supported feature selection and creation, simplified text mining, and more SQL functions that enabled users to view model prediction details. Also added was Predictive Queries (also known as Dynamic Scoring), which allows scoring data dynamically without a predefined model.</td>
</tr>
</tbody>
</table>
summarizes the main developments and releases of Darwin and subsequently Oracle Data Mining. With each subsequent release of the Oracle Database and SQL Developer, Oracle has added greater Oracle Data Mining functionality. Check out the “new features” documentation for each release to see what new features have been added.

**Oracle Data Mining Components**

Oracle Data Mining is an option in the database. It is fully integrated into the kernel of the database. This means that all the data mining algorithms are built into the database engine, and this allows these algorithms to be easily used in applications by using the SQL and PL/SQL functions and procedures. Any data mining objects created for data mining become first-class objects in the database and can be treated similarly to any other SQL function and other database options. Before you can use Oracle Data Mining, a repository needs to be built that manages all Oracle Data Mining objects and their usage.

**Oracle Data Mining Architecture**

Although Oracle Data Mining and all its algorithms come prebuilt into the database, you will need to create the Oracle Data Mining Repository before you can start using Oracle Data Miner. SQL scripts come as part of SQL Developer, which allows your database administrator (DBA) to create the ODM Repository in the database. An alternative approach is to allow functionality embedded in the SQL Developer tool to manage the creation of the ODM Repository. The architecture of Oracle Data Mining is illustrated in Figure 1-3.

The enterprise edition of the Oracle Database includes the following database features that support Oracle Data Miner:

- **Oracle Data Mining**  A component of the Oracle Advanced Analytics option to Oracle Database Enterprise Edition; Oracle Data Mining provides the model building, testing, and scoring capabilities for Data Miner.

- **Oracle XML DB**  This database provides services to manage the Data Miner Repository metadata, such as the details of the workflow specifications.

- **Oracle Scheduler**  This component provides the engine for scheduling the Data Miner workflows.

- **Oracle Text**  This feature provides services necessary to support Text Mining.
The Oracle Data Miner tool is a component of SQL Developer. The Oracle Data Miner tool is a GUI workflow–based tool that allows everyone from data scientists, data analysts, developers, and DBAs to build quickly and simply data mining workflows for their data and data mining business problem. The Oracle Data Miner workflow tool was first introduced in SQL Developer 3, and with all subsequent releases additional functionality has been added.

The Oracle Data Miner tool (see Figure 1-4) allows you to build workflows by defined nodes that enable you to accomplish the following:

- Explore your data using statistics and various graphical methods
- Build various data transformations that include sampling, various data reduction techniques, create new features, apply complex filtering techniques, and create custom transformations on your data
- Build data mining models using the variety of in-database data mining algorithms
- Apply your data mining models to new data to produce a scored data set that can be acted upon by your business users
- Create and use transient data mining models using Predictive Queries
- Create and apply complex Text Analytics models on your semistructured and unstructured data
When the time comes to productionize your workflows, full support is provided by the generation of all the required SQL scripts that are needed to run the workflow. These can be easily scheduled in the database to run on a regular basis.

**TIP**
The second part of this book contains chapters illustrating how you can build workflows for each of the main data mining techniques.

**Oracle Data Mining Using SQL and PL/SQL**

Data Mining models are database schema objects that perform data mining by applying a model created using one of the in-database data mining algorithms. A number of PL/SQL packages and SQL functions are provided to allow you to define, create, apply, and assess data mining models in the database.

The main PL/SQL package for Oracle Data Mining is called `DBMS_DATA_MINING`. This package contains all the procedures you will need to create your data mining
models, assess and explore the internals of these models, and apply these models to new data.

The `DBMS_DATA_MINING_TRANSFORM` PL/SQL package contains various functions that allow you to transform your data to prepare it for input into your Oracle Data Mining models. These procedures can be used in addition to any additional programming that is necessary to prepare the data. When using the `DBMS_DATA_MINING_TRANSFORM` procedures, you can embed their transformations into your Data Mining models.

The `DBMS_PREDICTIVE_ANALYTICS` PL/SQL package allows you to Profile, Explain, and Predict using the in-database Data Mining functions, but when using this package you do not have any input into the predictive analytics process. All the processing is done internally inside the Oracle Data Mining algorithms to determine the appropriate methods to use.

The Oracle data dictionary comes with a number of database views that allow you to see the Data Mining models you have and the attributes and settings that those models use. These data dictionary views include the following:

- `*_MINING_MODELS`
- `*_MINING_MODEL_ATTRIBUTES`
- `*_MINING_MODEL_PROPERTIES`

where `*` can be replaced by

- `ALL_` This view contains the Oracle Data Mining information that is accessible to the user.
- `DBA_` This view displays the Oracle Data Mining information that is accessible to DBA users.
- `USER_` This view contains the Oracle Data Mining information that is accessible to the current user.

Oracle has at least 15 SQL functions that allow you to score data using a data mining model. There are two primary SQL functions that you will use most commonly. These are `PREDICTION` and `PREDICTION_PROBABILITY`. The other SQL functions allow you to determine the outcomes of applying a Clustering data mining model; establish predicted features; and define various prediction values such as costs, value bounds, and prediction rules/details.
Part III of this book will cover the details of how you can use these data dictionary views, PL/SQL packages, and SQL functions. Examples will be given to illustrate how these can be used for different data mining techniques.

Oracle Statistical Functions
All versions of the Oracle Database come with a comprehensive collection of statistical functions built into the database. These statistical functions come standard with the Oracle Database and do not require any additional licences. The database features more than 110 SQL and PL/SQL statistical functions that can be grouped under a variety of headings, as illustrated in Table 1-3.

### Ranking Functions
- rank, dense_rank, cume_dist, percent_rank, ntile

### Window Aggregate Functions
(moving & cumulative)
- Avg, sum, min, max, count, variance, stddev, first_value, last_value

### LAG/LEAD Functions
Direct interrow reference using offsets

### Reporting Aggregate Functions
- Sum, avg, min, max, variance, stddev, count, ratio_to_report

### Statistical Aggregates
- Correlation, linear regression family, covariance

### Linear Regression
Fitting of an ordinary-least-squares regression line to a set of number pairs
Frequently combined with the COVAR_POP, COVAR_SAMP, and CORR functions

### Descriptive Statistics
- DBMS_STAT_FUNCS: summarizes numerical columns of a table and returns count, min, max, range, mean, median, stats_mode, variance, standard deviation, quantile values, +/- n sigma values, top/bottom 5 values

### Correlations
- Pearson’s correlation coefficients, Spearman’s and Kendall’s (both nonparametric)

### Cross Tabs
Enhanced with % statistics: chi squared, phi coefficient, Cramer’s V, contingency coefficient, Cohen’s kappa

### Hypothesis Testing
- Student t-test, F-test, Binomial test, Wilcoxon Signed Ranks test, Chi-square, Mann Whitney test, Kolmogorov-Smirnov test, one-way ANOVA

### Distribution Fitting
- Kolmogorov-Smirnov test, Anderson-Darling test, Chi-squared test, Normal, Uniform, Weibull, Exponential

---

**TABLE 1-3. Summary of Free Statistical Functions in Oracle**
Some advanced statistical functions were introduced in more recent versions of the database. These functions enable you to perform various statistical analysis using windowing or moving window calculations; these advanced functions include Pivot, Rankings, Lead/Lag, and more. These are particularly useful in data warehousing and some advanced analytics projects.

**TIP**
If the statistical function you want to use is not included in the Oracle Database but it exists in R, then you can use the R statistical function on your data in the database.

**Applications Powered by Oracle Data Mining**
Over the past few years, Oracle has been integrating into some of its applications the functionality available in the Oracle Advanced Analytics option. For example, Oracle has integrated Oracle Data Mining into the applications by building data mining models and making these available as features of these applications. The Oracle Data Mining functionality allows the applications to predict turnover, perform what-if analysis, real-time fraud and security analytics, customer churn and segmentation, anomaly detection, customer loyalty analysis, among many other tasks. The following are some of the Oracle applications that have Oracle Advanced Analytics and Oracle Data Mining built into them:

- Oracle Fusion HCM Workforce Predictions
- Oracle Fusion CRM Sales Prediction Engine
- Oracle Spend Classification
- Oracle Sales Prospector
- Oracle Adaptive Access Manager
- Oracle Airline Data Model
- Oracle Communications Data Model
- Oracle Retail Data Model
- Oracle Security Governor for Healthcare
- Oracle Social Network Analysis
Watch out for more applications that will have Oracle Advanced Analytics added to their functionality in the future.

How Are Customers Using Oracle Advanced Analytics

The Oracle Advanced Analytics option has a considerable amount of advanced functionality built into it. This functionality includes the in-database Oracle Data Mining features and the advanced features of Oracle R Enterprise. The typical data mining functionality available in these products can be grouped together, and the following are some examples of the typical application areas for data mining:

- Identify most important factor (Attribute Importance)
- Predict customer behavior (Classification)
- Predict or estimate a value (Regression)
- Find profiles of targeted people or items (Decision Trees)
- Segment a population (Clustering)
- Find fraudulent or “rare events” (Anomaly Detection)
- Determine co-occurring items in a “basket” (Associations)

These typical data mining areas and problems can be applied in a number of industries to gain a deeper insight into their customers, their environments, their processes, and so on. The following are some typical use cases:

- Targeting the right customer with the right offer
- Discovering hidden customer segments
- Finding the most profitable selling opportunities
- Anticipating and preventing customer churn
- Exploiting the full 360 degree customer opportunity
- Providing security and detecting suspicious activity
- Understanding sentiments in customer conversations
- Reducing medical errors and improving quality of health
- Understanding influencers in social networks
Customer Success Stories

Being able to see and understand how other companies have used Oracle Advanced Analytics can help you determine how you can use this option in your company. Oracle Advanced Analytics has a wide customer base spanning many industries and types of projects. Some Oracle Advanced Analytics customers have shared the details of their projects at various conferences and in various publications. The following table lists some of these customers and how they have used Oracle Advanced Analytics.

<table>
<thead>
<tr>
<th>Customer</th>
<th>How They Used Oracle Advanced Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkcell</td>
<td>Turkcell is the largest telecommunications (telco) company in Turkey. It is using Oracle Advanced Analytics and Oracle Data Mining to analyze customer call records with the aim of identifying potential fraudulent calls to their prepaid customers.</td>
</tr>
<tr>
<td>Dunnhumby</td>
<td>Dunnhumby is one of the world’s largest customer analytics companies. It provides customer insight for some of the largest retail companies in the world. Dunnhumby is using Oracle Data Miner to build many customer loyalty models. The company has succeeded in reducing the total time required on a monthly basis to process data, build models using Oracle Data Miner, and score its customers from weeks to minutes.</td>
</tr>
<tr>
<td>Stubhub</td>
<td>Stubhub is a fan-to-fan ticket marketplace. It is using Oracle Advanced Analytics to gain a better insight into the company’s customers and their online behavior. By using Oracle Advanced Analytics and the in-database capabilities, Stubhub can produce its advanced analytics in a much shorter time frame.</td>
</tr>
<tr>
<td>Oracle Racing</td>
<td>How did Oracle win the America’s Cup? The winning team used Oracle Advanced Analytics and Oracle Data Mining to analyze the racing performance of the team’s yacht using over 2,500 variables. During each training session and race, data was constantly being collected from sensors located all over the yacht. The data was constantly being collected and fed back to the team’s analytics database. The data was analyzed and the outputs were used to make adjustments and improvements to the configuration of the yacht.</td>
</tr>
</tbody>
</table>
## Customer

### Argonne National Laboratory

Argonne National Laboratory uses Oracle Data Mining to model and predict protein crystallization propensity from protein sequences. The lab’s scientists were able to use Oracle Data Mining to identify the set of attributes that correlated with the protein’s propensity to crystallize and used the SVM algorithm in Oracle Data Mining to build the model.

### StuartMaue

StuartMaue receives over 4,000 submissions per month for payment. This equates to over $200 million worth of legal invoices per month. StuartMaue wanted to automate and improve the review, categorization, and investigation of possibly noncompliant legal submissions. This was a very labor-intensive process that involved trying to spot potential fraudulent or erroneous submissions. Oracle Data Mining was used to mine the structured and unstructured data in the firm’s Oracle Database. By automating the process, StuartMaue was able to scale to large volumes, saving time and money.

### Xerox

Xerox is using Oracle Data Mining to analyze its customer data. The company has been able to save a significant amount of processing time by using Oracle Data Miner in its Oracle Databases, as it no longer has to move its data out of the database to another data mining application. Using the Oracle Data Mining in-database capabilities has saved Xerox both time and money.

### NOTE

Many more companies have used Oracle Advanced Analytics, but the details of their projects are not publicly available. Watch out for more customer case studies at conferences such as Oracle OpenWorld, in publications such as Oracle Magazine, and on the Oracle website.