



User Manual

a Bioada Product

MODELLING TEAM¹

July 3, 2020

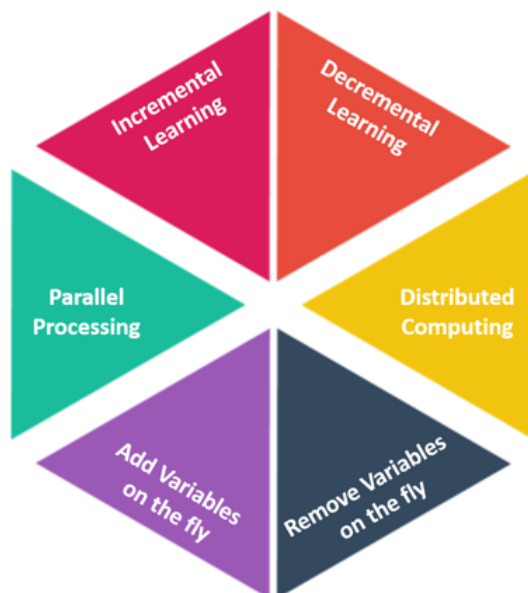
¹www.bioada.com

Purpose and Overview

Xarang - A Real Time Machine Learning Platform

Although machine learning algorithms are widely used in extremely diverse situations, in practice, one or more major limitations almost invariably appear and significantly constrain successful applications. Frequently, these problems are associated with large increases in the rate of generation of data, the quantity of data and the number of attributes (variables) to be processed. Increasingly, the data situation is now beyond the capabilities of conventional data mining methods.

The term “Real Time” is used to describe how well a machine learning algorithm can accommodate an ever increasing data load instantaneously. However, such real time problems are usually closely coupled with the fact that conventional algorithms operate in a batch mode where having all of the relevant data at once is a requirement. Xarang as a real time machine learning toolbox has the following characteristics, independent of the amount of data involved.



1. Incremental learning (Learn): immediately updating a model with each new observation without the necessity of pooling new data with old data.
2. Decremental learning (Forget): immediately updating a model by excluding observations identified as adversely affecting model performance without forming a new dataset omitting this data and returning to the model formulation step.
3. Variable addition (Grow): Adding a new attribute (variable) on the fly, without the necessity of pooling new data with old data.
4. Variable deletion (Shrink): immediately discontinuing use of an attribute identified as adversely affecting model performance.
5. Distributed processing: separately processing distributed data or segments of large data (that may be located in diverse geographic locations) and re-combining the results to obtain a single model.
6. Parallel processing: carrying out parallel processing extremely rapidly from multiple conventional processing units (multi-threads, multi-processors or a specialized chip).

Contents

| | | |
|----------|---------------------------------------|-----------|
| 1 | Introduction | 2 |
| 1.1 | What is Xarang | 2 |
| 1.2 | What can you do with Xarang | 2 |
| 2 | Project in Xarang | 3 |
| 2.1 | Creating a new project | 3 |
| 2.2 | Opening an existing Project | 4 |
| 2.3 | Saving a project | 4 |
| 2.4 | Merging two projects | 5 |
| 3 | Data in Xarang | 7 |
| 3.1 | File | 7 |
| 3.2 | Folder | 8 |
| 3.3 | Database | 9 |
| 3.4 | Cloud | 9 |
| 4 | Learner in Xarang | 11 |
| 5 | Explorer in Xarang | 13 |
| 5.1 | Univariate | 13 |
| 5.2 | Bi-variate | 14 |
| 5.2.1 | Correlation | 14 |
| 5.2.2 | Hypothesis Testing | 15 |
| 5.2.3 | ANOVA | 16 |
| 5.2.4 | Test of Independence | 17 |
| 5.3 | Multi-Variate | 18 |
| 5.3.1 | Extraction Methods | 19 |
| 5.3.2 | Number of Factors | 19 |
| 5.3.3 | Rotation Methods | 19 |
| 6 | Modeler in Xarang | 20 |
| 6.1 | Binary Classification | 20 |
| 6.2 | Regression | 22 |
| 7 | Predictor in Xarang | 25 |
| 7.1 | Select a dataset | 25 |
| 7.2 | Select a model | 26 |
| 7.3 | Predict using a model | 27 |
| 7.4 | Evaluate Prediction Model | 27 |

| | |
|------------------------------|-----------|
| <i>CONTENTS</i> | 1 |
| 8 Deploying in Xarang | 29 |
| 8.1 Scorecard | 30 |
| 8.2 A/B Test | 31 |

1

Introduction

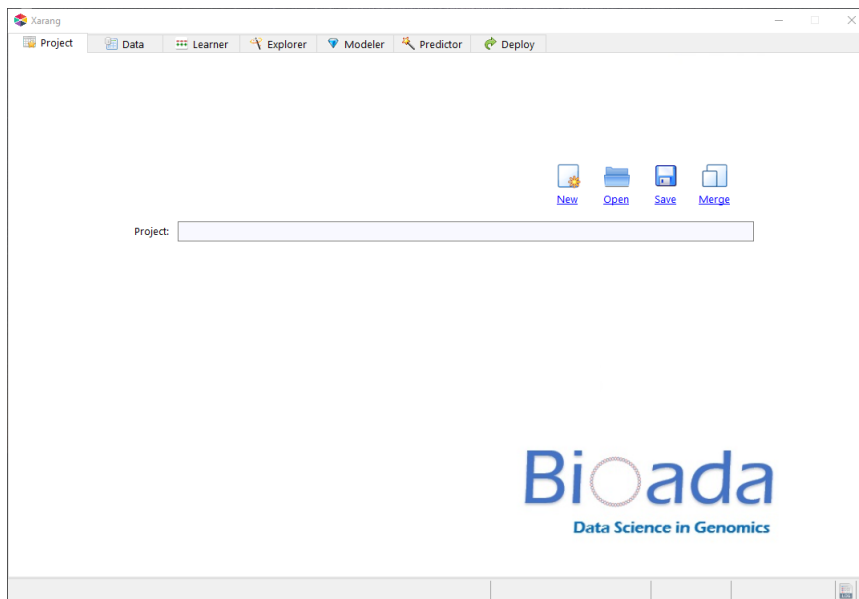
1.1 What is Xarang

1.2 What can you do with Xarang

2

Project in Xarang

The Project tab is the first screen of Xarang. We can create a new project, open or save an existing project, or merge multiple projects. Projects are stored as a set of files.

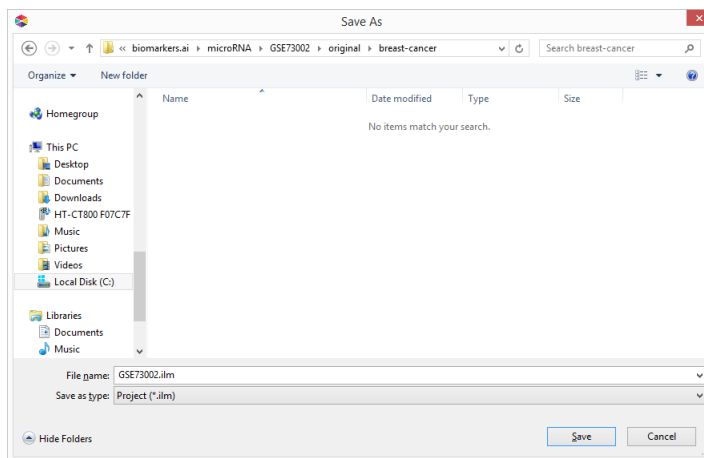


2.1 Creating a new project

1. To create a new project, click "New" project icon.

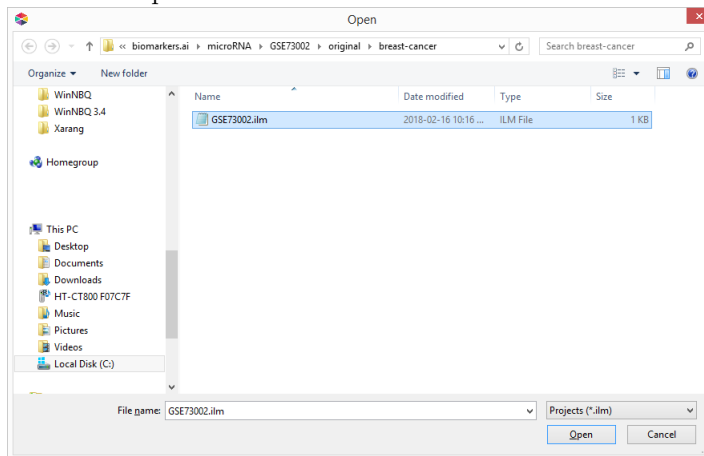


2. Select a folder.
3. Enter project name in the "File name" box.
4. Click "Save".



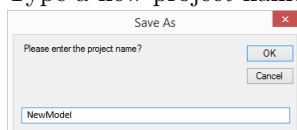
2.2 Opening an existing Project

1. To open an existing project, click on the Open icon.
2. Select the project file (e.g., GSE73002.ilm) click the Open button.
3. Click the "Open" button.



2.3 Saving a project


1. To save an existing project under a different name, click on the Save icon.
2. Type a new project name (e.g., NewModel) and click OK.

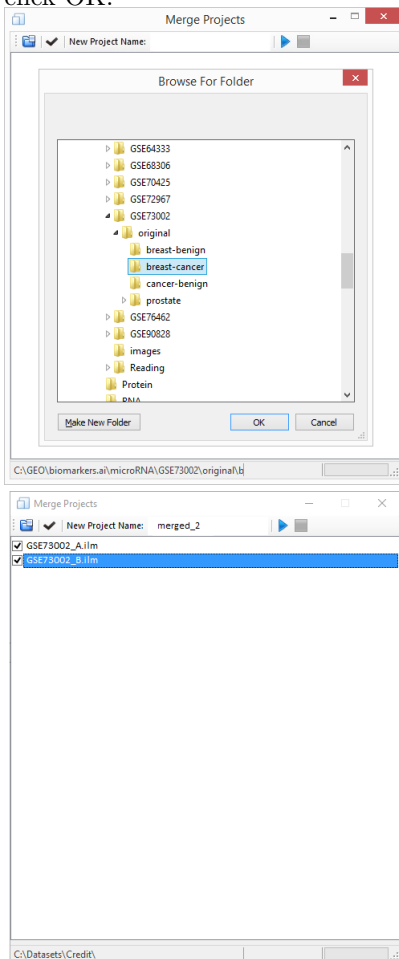


2.4 Merging two projects

1. To merge two or more existing projects, click on the Merge icon.



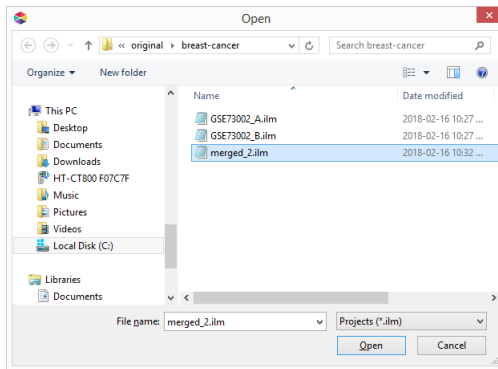
2. Click on the Open Folder icon to open "Browse For Folder"  dialogue box.
3. Select the folder that contains the projects you would like to merge and click OK.



4. Check the boxes next to the files you would like to merge. To select all files, you can click "Select All " button .
5. To merge the selected projects, click "Start ►" button . You can stop the merger by clicking "Stop ■" button at any time.
6. merged2.ilm is the newly merged project.

2. PROJECT IN XARANG

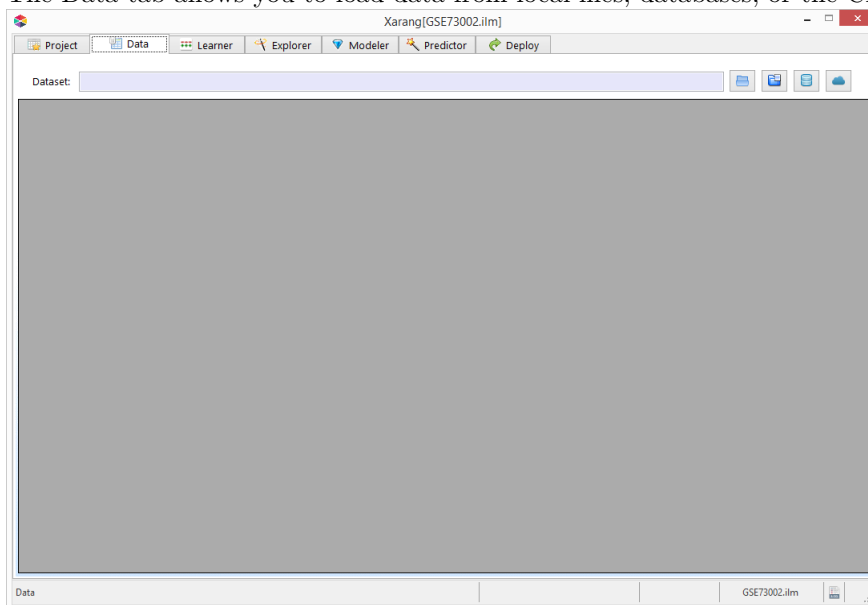
6




3

Data in Xarang

The Data tab allows you to load data from local files, databases, or the Cloud.





3.1 File

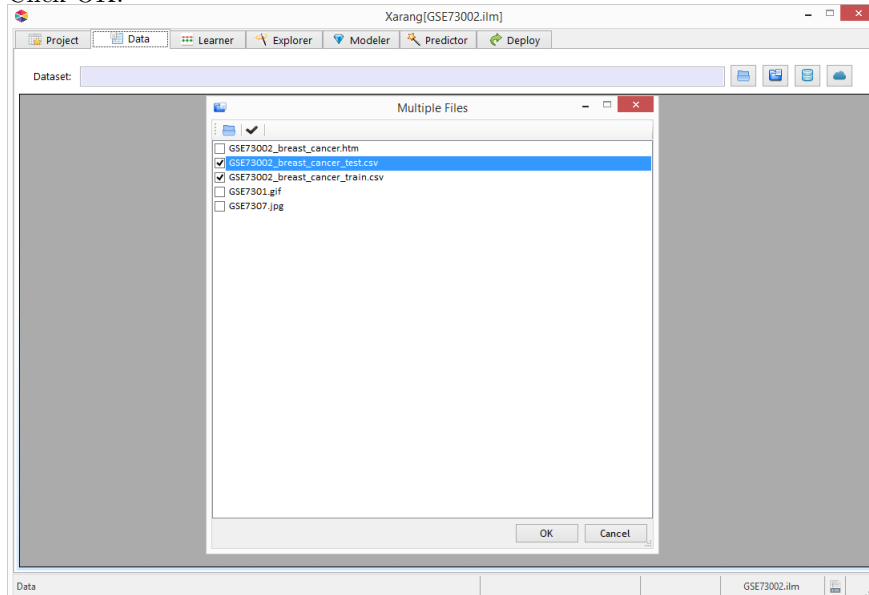
1. To load a local dataset, click the Open File  button.
2. Select the data file.
3. Click Open.

The screenshot shows the Xarang software interface with a data table. The table has 10 columns representing different miRNA accessions and rows of numerical data. The columns are: geo_accession, hsa-miR-1307-3p, hsa-miR-4783-3p, hsa-miR-8073, hsa-miR-4532, hsa-miR-6787-5p, hsa-miR-6861-5p, hsa-miR-1233-5p, hsa-miR-4675, and hsa-miR-92a-2-5p. The rows contain numerical values for each of these accessions.



| geo_accession | hsa-miR-1307-3p | hsa-miR-4783-3p | hsa-miR-8073 | hsa-miR-4532 | hsa-miR-6787-5p | hsa-miR-6861-5p | hsa-miR-1233-5p | hsa-miR-4675 | hsa-miR-92a-2-5p |
|---------------|-----------------|-----------------|--------------|--------------|-----------------|-----------------|-----------------|--------------|------------------|
| 1876422 | 7.826695 | 9.315201 | 6.307992 | 14.32252 | 9.210176 | 7.909871 | 11.83945 | 10.2466 | 5.91683 |
| 1876425 | 8.651167 | 9.051309 | 6.796376 | 13.71886 | 9.09567 | 8.22059 | 12.71636 | 9.68074 | 6.163979 |
| 1876427 | 9.11246 | 8.549548 | 7.190443 | 13.31739 | 9.162703 | 8.13832 | 12.60987 | 9.116076 | 7.157032 |
| 1876428 | 9.433894 | 9.556254 | 7.449193 | 14.16441 | 9.247772 | 8.091199 | 12.78285 | 9.899 | 6.950449 |
| 1876429 | 8.458979 | 8.897826 | 7.310657 | 13.47416 | 9.067359 | 7.822359 | 12.17858 | 9.646892 | 6.777309 |
| 1876430 | 8.62834 | 8.987145 | 6.677789 | 13.39477 | 9.119116 | 8.459085 | 13.03032 | 9.101669 | 6.84856 |
| 1876431 | 8.92551 | 8.95489 | 7.692006 | 13.68114 | 9.258071 | 8.388326 | 13.16057 | 9.884966 | 6.716098 |
| 1876432 | 8.828325 | 9.247807 | 7.138846 | 14.01279 | 9.262071 | 7.714129 | 11.55853 | 9.719329 | 6.097628 |
| 1876433 | 8.653539 | 8.289276 | 7.396927 | 12.97289 | 9.168592 | 8.27291 | 13.12131 | 9.431019 | 8.55323 |
| 1876437 | 7.976727 | 8.182042 | 6.681544 | 11.48957 | 8.81449 | 8.023083 | 14.3661 | 9.648237 | 7.370079 |
| 1876438 | 8.362527 | 8.715057 | 6.524507 | 13.0686 | 8.88969 | 7.994096 | 12.5294 | 8.705749 | 6.518305 |
| 1876439 | 9.110463 | 9.07073 | 7.105684 | 13.87262 | 9.545561 | 8.374212 | 12.52688 | 9.409822 | 7.25053 |
| 1876441 | 8.261886 | 9.020363 | 6.867811 | 13.22965 | 8.8639 | 7.799739 | 12.60262 | 9.505303 | 6.545929 |
| 1876442 | 10.05067 | 9.406485 | 7.757532 | 14.14914 | 9.295083 | 8.570689 | 12.36004 | 9.273005 | 6.644682 |
| 1876443 | 9.667789 | 9.42463 | 7.697715 | 13.60557 | 9.151859 | 8.507047 | 12.43994 | 9.619081 | 6.171601 |
| 1876446 | 9.411946 | 9.189084 | 7.568458 | 14.64995 | 9.329148 | 8.074501 | 13.05857 | 9.068426 | 7.382524 |
| 1876447 | 8.798788 | 8.742432 | 6.407903 | 13.44173 | 8.493637 | 7.33468 | 11.98914 | 9.190237 | 6.97578 |
| 1876449 | 9.989674 | 9.720611 | 8.328921 | 14.32419 | 9.259361 | 8.550528 | 12.63676 | 9.643961 | 6.955666 |
| 1876450 | 9.499095 | 9.755453 | 7.715518 | 14.13907 | 9.315679 | 8.373251 | 12.82136 | 9.970511 | 6.742391 |
| 1876451 | 9.026791 | 9.615182 | 7.413162 | 14.19006 | 9.1558 | 7.49616 | 11.33785 | 9.904745 | 6.584494 |
| 1876452 | 9.928274 | 9.594193 | 7.964289 | 15.1328 | 9.74538 | 8.932371 | 11.90529 | 9.65484 | 6.630105 |
| 1876454 | 9.472902 | 9.797751 | 7.769221 | 13.70424 | 9.922887 | 8.797158 | 13.30747 | 10.03412 | 6.716614 |
| 1876455 | 9.32481 | 9.692487 | 7.641259 | 13.6208 | 8.972921 | 8.669937 | 13.57547 | 8.919545 | 7.17508 |
| 1876457 | 9.208461 | 8.700366 | 7.715806 | 13.58448 | 9.678354 | 8.790678 | 12.18102 | 9.428689 | 7.698641 |
| 1876458 | 9.743966 | 9.688868 | 6.989293 | 14.67 | 8.747888 | 8.423214 | 12.66583 | 9.867255 | 6.433856 |

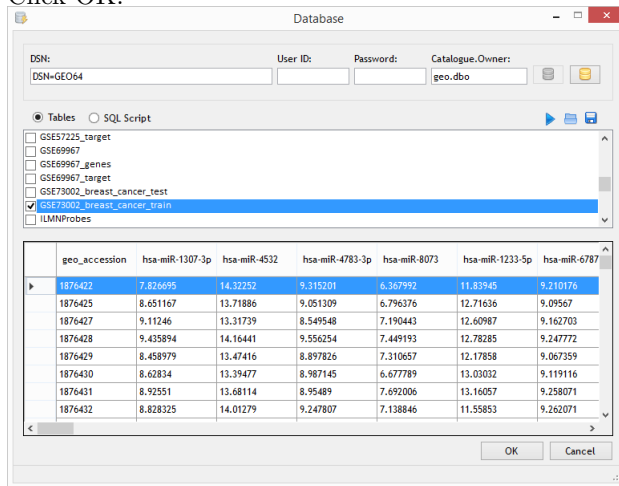
3.2 Folder

1. To load more than one dataset, click the Open Folder  button.
2. Click Open button  and select the folder that contains data files.
3. Select one or more files.
4. Click OK.

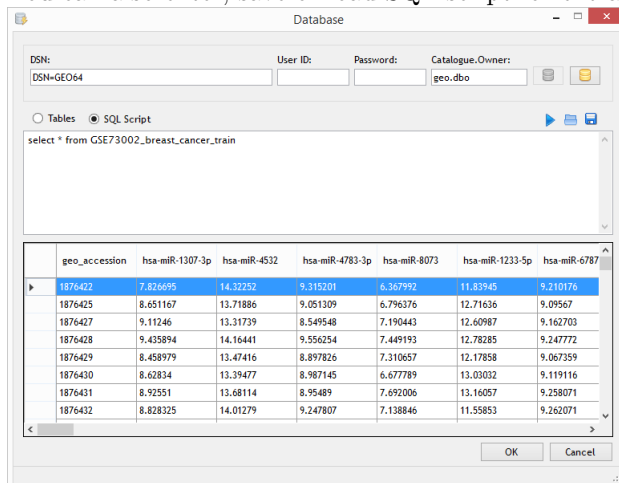


3.3 Database


1. To load data from a database, click the Open Database  button.
2. Enter the DSN, User ID/ Password (if needed) in the related fields and click .
3. Select one or more tables from the list. All tables must have the same schema.
4. Click OK.



5. You can also enter, save or load SQL script for extracting data.



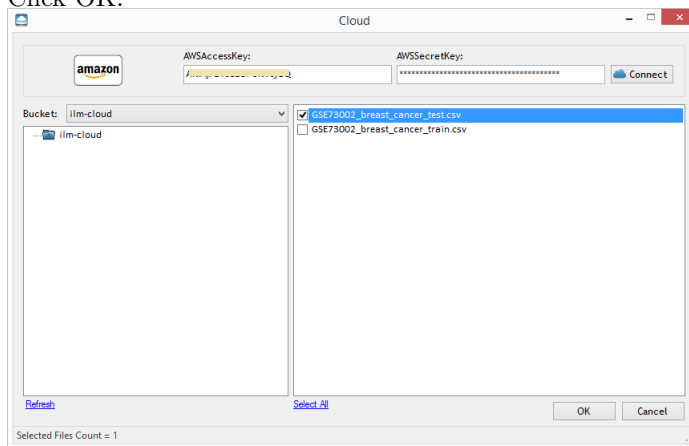
3.4 Cloud

1. To load data from the cloud, click the Open Cloud  button.
2. Enter the AWS Access Key and AWS Secret Key.
3. Click the Connect button.

3. DATA IN XARANG



10

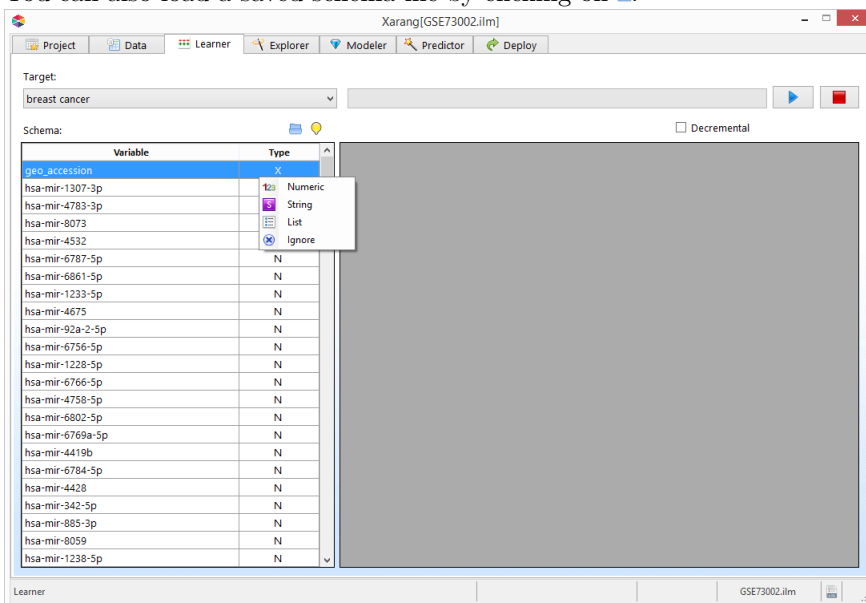
4. Select a Bucket.
5. Select a Folder.
6. Select one or more files. All files must have the same schema.
7. Click OK.





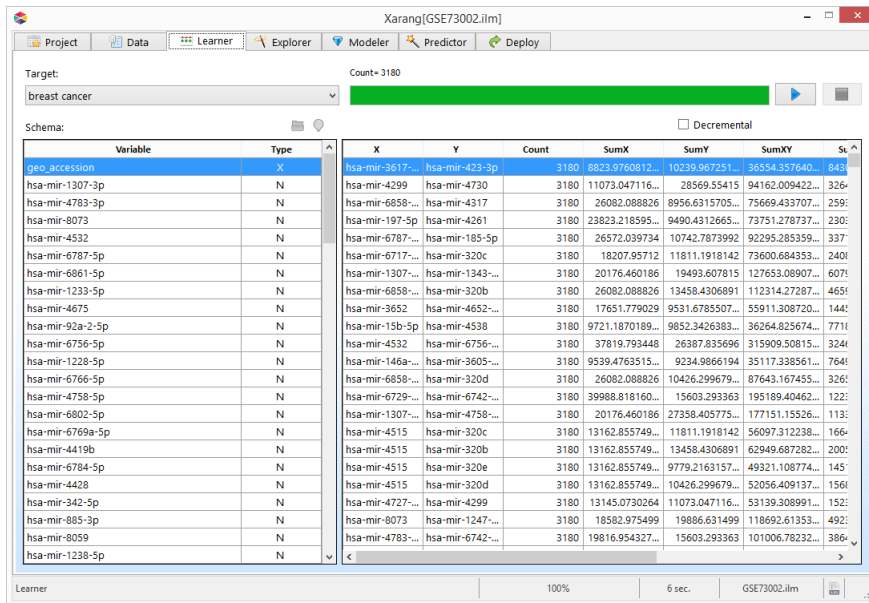
4

Learner in Xarang

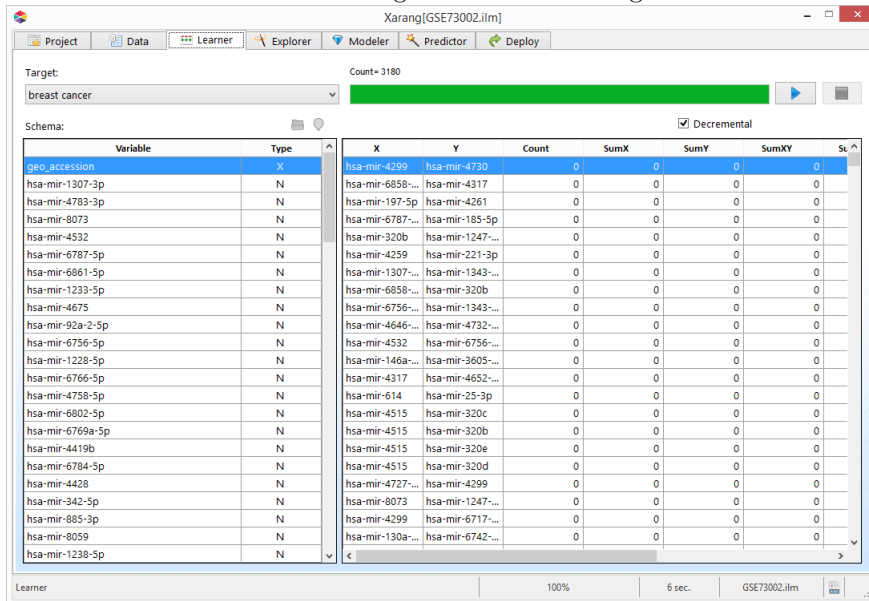
1. Select "Target" variable.
2. Click  to guess the type of each variable. You can change the type of variable by right clicking on that variable.
3. Select "Ignore" if you would like to exclude a variable from the learning process.
4. You can also load a saved schema file by clicking on .



5. Click the Start Learner button. 
6. Stop the process any time by clicking the Stop button 



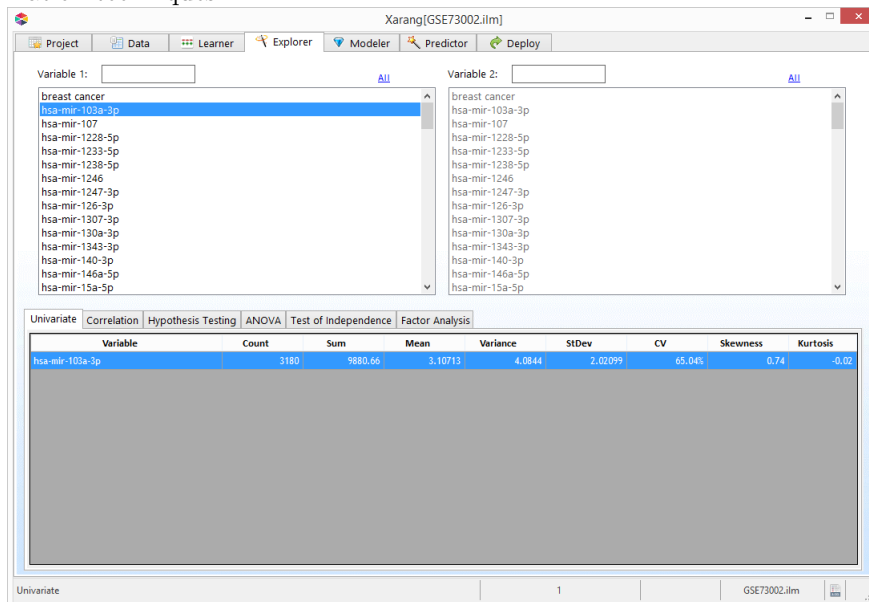
7. You can rollback the learning (called forgetting) by checking the Decremental checkbox and then clicking the Start button again.



5

Explorer in Xarang

The Explorer tab allows you to explore the data statistical analysis and visualization techniques.



- Univariate (descriptive statistics)
- Bivariate (inferential statistics)
- Multivariate (factor analysis)

5.1 Univariate

Univariate data analysis explores attributes (variables) one by one using statistical analysis. Attributes are either numerical or categorical (encoded to binary).

1. Select the Univariate tab.
2. Select one or more variables from the Variable 1 list.

- To view only Binary or Numeric variables, click All, then Binary or Numeric.

The screenshot shows the Xarang software interface with the Explorer tab selected. Two variable lists are visible, both with 'All' selected. Below them is a summary table for the selected variables.


| Variable | Count | Sum | Mean | Variance | STDev | CV | Skewness | Kurtosis |
|-----------------|-------|----------|----------|----------|----------|---------|----------|----------|
| breast cancer | 3180 | 1011 | 0.317925 | 0.216849 | 0.46567 | 146.47% | 0.78 | -1.39 |
| hsa-mir-103a-3p | 3180 | 9880.66 | 3.10713 | 4.0844 | 2.02099 | 65.04% | 0.74 | -0.02 |
| hsa-mir-107 | 3180 | 10189.37 | 3.20421 | 4.03717 | 2.00927 | 62.71% | 0.71 | -0.02 |
| hsa-mir-1228-5p | 3180 | 37313.02 | 11.7337 | 0.2564 | 0.50636 | 4.32% | -0.8 | 0.21 |
| hsa-mir-1233-5p | 3180 | 25793.15 | 11.2557 | 1.14236 | 1.06881 | 9.50% | 0.94 | -0.27 |
| hsa-mir-1238-5p | 3180 | 18638.24 | 5.86108 | 1.06906 | 1.03395 | 17.64% | 0.72 | 0.54 |
| hsa-mir-1246 | 3180 | 14978.86 | 4.71033 | 8.51288 | 2.91768 | 61.94% | 0.59 | -0.74 |
| hsa-mir-1247-3p | 3180 | 19886.63 | 6.25366 | 0.569432 | 0.754607 | 12.07% | 0.8 | 0.1 |
| hsa-mir-126-3p | 3180 | 9058.02 | 2.84843 | 2.71116 | 1.64656 | 57.81% | 0.41 | -0.48 |
| hsa-mir-1307-3p | 3180 | 20176.46 | 6.3448 | 3.70182 | 1.92401 | 30.32% | 0.86 | -0.95 |
| hsa-mir-130a-3p | 3180 | 9371.36 | 2.94697 | 3.19721 | 1.78807 | 60.68% | 0.47 | -0.52 |
| hsa-mir-1343-3p | 3180 | 19493.61 | 6.13007 | 0.667911 | 0.817258 | 13.33% | 0.75 | 0 |

5.2 Bi-variate

Bi-variate analysis is the simultaneous analysis of two variables (attributes). It explores the concept of relationship between two variables, whether there exists an association and the strength of this association, or whether there are differences between two variables and the significance of these differences. There are four types of bi-variate analysis.

5.2.1 Correlation

Linear correlation quantifies the strength of a linear relationship between two numerical variables. When there is no correlation between two variables, there is no tendency for the values of one quantity to increase or decrease with the values of the second quantity.

- Select the Correlation tab.
- Select one or more variables from the Variable 1 list.
- Select the second variable from the Variable 2 list.
- Click  to visualize the result.

The screenshot shows the Xarang software interface with the Hypothesis Testing tab selected. The interface displays two variable lists (Variable 1 and Variable 2) and a correlation matrix table.


Variable 1: breast cancer, hsa-mir-103a-3p, hsa-mir-107, hsa-mir-1228-5p, hsa-mir-1233-5p, hsa-mir-1238-5p, hsa-mir-1246, hsa-mir-1247-3p, hsa-mir-126-3p, hsa-mir-1307-3p, hsa-mir-130a-3p, hsa-mir-1343-3p, hsa-mir-140-3p, hsa-mir-146a-5p, hsa-mir-15a-5p

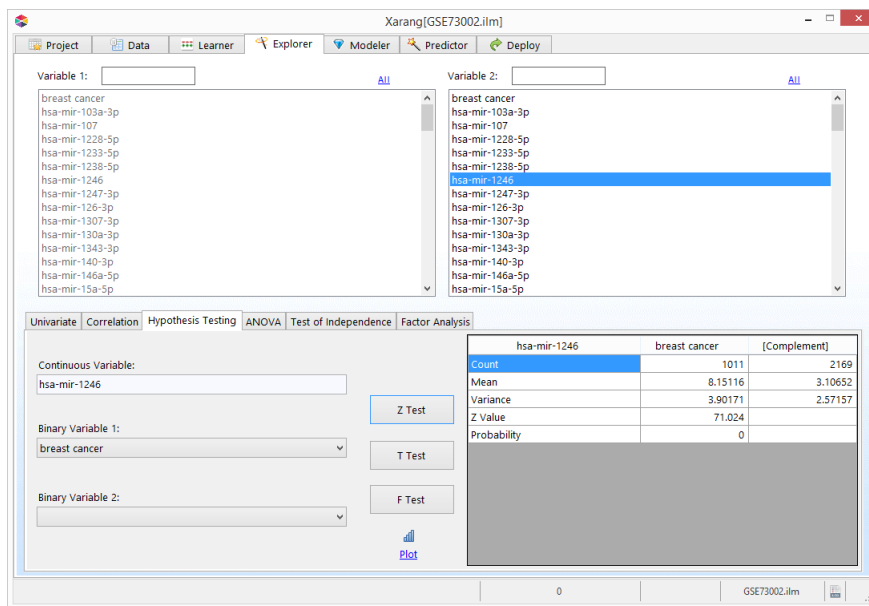
Variable 2: breast cancer, hsa-mir-103a-3p, hsa-mir-107, hsa-mir-1228-5p, hsa-mir-1233-5p, hsa-mir-1238-5p, hsa-mir-1246, hsa-mir-1247-3p, hsa-mir-126-3p, hsa-mir-1307-3p, hsa-mir-130a-3p, hsa-mir-1343-3p, hsa-mir-140-3p, hsa-mir-146a-5p, hsa-mir-15a-5p

Correlation Matrix:

| Variable 1 | Variable 2 | Covariance | Correlation |
|-----------------|--------------|------------|-------------|
| breast cancer | hsa-mir-1246 | 1.09392 | 0.805136 |
| hsa-mir-103a-3p | hsa-mir-1246 | 5.0029 | 0.848436 |
| hsa-mir-107 | hsa-mir-1246 | 4.97258 | 0.848213 |
| hsa-mir-1228-5p | hsa-mir-1246 | -1.09027 | -0.737964 |
| hsa-mir-1233-5p | hsa-mir-1246 | 2.34972 | 0.753489 |
| hsa-mir-1238-5p | hsa-mir-1246 | 2.0316 | 0.67344 |
| hsa-mir-1246 | hsa-mir-1246 | 8.51288 | 1 |
| hsa-mir-1247-3p | hsa-mir-1246 | 1.55768 | 0.707487 |
| hsa-mir-126-3p | hsa-mir-1246 | 3.78709 | 0.788297 |
| hsa-mir-1307-3p | hsa-mir-1246 | 4.50919 | 0.803254 |
| hsa-mir-130a-3p | hsa-mir-1246 | 4.1819 | 0.801585 |
| hsa-mir-1343-3p | hsa-mir-1246 | 1.87308 | 0.785521 |

5.2.2 Hypothesis Testing

1. Select the Hypothesis Testing tab.
2. Select a numerical variable from the Variable 2 list. Note: To view only Binary or Numeric variables, click All, then Binary or Numeric.
3. Select Binary Variable 1 and if necessary Binary Variable 2 from the drop down lists.
4. Click Z Test , T Test , or F Test button. The related result will be displayed.
5. Click  to visualize the result.




The details of the test performed are as follows:

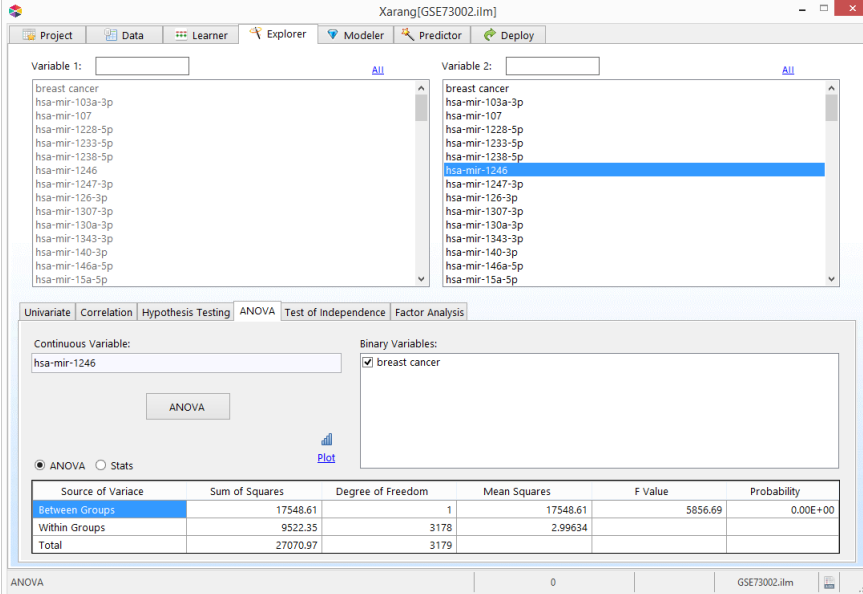
1. Z-Test : The Z test assesses whether the difference between averages of two attributes are statistically significant. This analysis is appropriate for comparing the average of a numerical attribute with a known average or two conditional averages of a numerical attribute given two binary attributes (two categories of the same categorical attribute).
2. T Test : The T test like Z test assesses whether the averages of two numerical attributes are statistically different from each other when the number of data points is less than 30. T test is appropriate for comparing the average of a numerical attribute with a known average or two conditional averages of a numerical attribute given two binary attributes (two categories of the same categorical attribute).
3. F Test : The F-test is used to compare the variances of two attributes. F test can be used for comparing the variance of a numerical attribute with a known variance or two conditional variances of a numerical attribute given two binary attributes (two categories of the same categorical attribute).

5.2.3 ANOVA

ANOVA (Analysis of Variance) assesses whether the averages of more than two groups are statistically different from each other, under the assumption that the corresponding populations are normally distributed. ANOVA is useful for comparing averages of two or more numerical attributes or two or more conditional averages of a numerical attribute given two or more binary attributes (two or more categories of the same categorical attribute).

1. Select the ANOVA tab.
2. Select a numerical variable from the Variable 2 list. Note: To view only Binary or Numeric variables, click All, then Binary or Numeric.

3. Select the Binary Variables from the Binary Variables list.
4. Click the ANOVA button. The ANOVA table will be displayed.
5. Click the Stats radio button to view the Count, Mean, and Variance.
6. Click  to visualize the result.



The screenshot shows the Xarangi software interface with the ANOVA configuration and results. The 'ANOVA' tab is active, and the 'Stats' radio button is selected. The ANOVA table is displayed below the configuration area.

| Source of Variance | Sum of Squares | Degree of Freedom | Mean Squares | F Value | Probability |
|--------------------|----------------|-------------------|--------------|---------|-------------|
| Between Groups | 17548.61 | 1 | 17548.61 | 5856.69 | 0.00E+00 |
| Within Groups | 9522.35 | 3178 | 2.99634 | | |
| Total | 27070.97 | 3179 | | | |

5.2.4 Test of Independence

The Chi2 test can be used to determine the association between categorical (binary) attributes. It is based on the difference between the expected frequencies and the observed frequencies in one or more categories in the frequency table. The Chi2 distribution returns a probability for the computed Chi2 and the degree of freedom. A probability of zero shows complete dependency between two categorical attributes and a probability of one means that two categorical attributes are completely independent.

1. Select the Test of Independence tab.
2. Select Binary variablea in Rows and a Binary variables in Columns.
3. Click the Chi2 button.

The screenshot shows the Xarang software interface with the 'Test of Independence' tab selected. The 'Binary Variables in Rows' and 'Binary Variables in Columns' both contain 'breast cancer'. The Chi2 test results are displayed in a table:

| Chi2 | DF | Probability | Correlation | Entropy |
|---------|----|-------------|-------------|---------|
| 3180.00 | 1 | 0.00E+00 | 1.00 | 0.0000 |

Below the table is a 2x2 contingency table:

| Observed | breast cancer | Complement |
|---------------|---------------|------------|
| breast cancer | 1011 | 0 |
| [Compleme... | 0 | 2169 |

5.3 Multi-Variate

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. For example, people may respond similarly to questions about income, education, and occupation, which are all associated with the latent variable socioeconomic status. The relationship of each variable to the underlying factor is expressed by the so-called factor loading. Here is an example of the output of a simple factor analysis. The first number underneath of every factor are "eigenvalue" and "percentage of variance explained".

The screenshot shows the Xarang software interface with the 'Factor Analysis' tab selected. The 'Extraction Method' is set to 'PrincipalComponents' and the 'Rotation Method' is set to 'None'. The output table shows the following factor loadings:

| Variable | Factor 1 9.65 0.80 | Factor 2 0.82 0.07 | Factor 3 0.31 0.03 | Factor 4 0.30 0.02 | Factor 5 0.25 0.02 | Factor 6 0.16 0.01 |
|-----------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| hsa-mir-103a-3p | 0.93710332 | 0.25466849 | -0.06526470 | -0.01970297 | -0.01309129 | 0.04064 |
| hsa-mir-107 | 0.93532493 | 0.25267532 | -0.05994179 | -0.03026325 | 0.00219564 | 0.03049 |
| hsa-mir-1228-5p | -0.84758667 | 0.31602695 | 0.22767903 | 0.26815033 | 0.14828440 | 0.10998 |
| hsa-mir-1233-5p | 0.87496760 | -0.30678830 | -0.07266150 | 0.27660846 | 0.03548619 | -0.02427 |
| hsa-mir-1238-5p | 0.82602537 | -0.44288548 | 0.06566737 | 0.21601137 | -0.06937733 | 0.15521 |
| hsa-mir-1246 | 0.88780471 | 0.06100218 | -0.21074203 | -0.08826216 | 0.34378944 | 0.16443 |
| hsa-mir-1247-3p | 0.85191361 | -0.19200869 | 0.36641437 | -0.27393294 | -0.01248447 | 0.12731 |
| hsa-mir-126-3p | 0.92089004 | 0.28702517 | 0.02909705 | 0.04856661 | -0.13112700 | 0.00751 |
| hsa-mir-1307-3p | 0.91573785 | -0.25439354 | -0.06462981 | -0.05656080 | 0.04067981 | -0.16136 |
| hsa-mir-130a-3p | 0.93823038 | 0.22699525 | 0.00427244 | 0.03443057 | -0.12930297 | -0.02591 |
| hsa-mir-1343-3p | 0.89554088 | 0.09654542 | 0.24863103 | 0.07360621 | 0.22806453 | -0.22143 |

5.3.1 Extraction Methods

Xarang supports seven extraction methods:

1. Alpha Factoring
2. Generalized Least Squares
3. Image Factoring
4. Iterative Principal Axis
5. Maximum Likelihood
6. Principal Components Analysis (PCA)
7. Unweighted Least Squares

PCA is the most popular extraction method. However, information on the relative strengths and weaknesses of these techniques is not well known. In general, Maximum Likelihood or Iterative Principal Axis will give you the best results, depending on whether your data are generally normally-distributed or significantly non-normal, respectively.

5.3.2 Number of Factors

After extraction you must decide how many factors to retain for rotation. Both over-extraction and under-extraction of factors retained for rotation can have damaging effects on the results. The default in most statistical software packages is to retain all factors with eigenvalues greater than 1.0. Alternate tests for factor retention include the scree test. The scree test involves examining the graph of the eigenvalues and looking for the natural bend or break point in the data where the curve flattens out. The number of datapoints above the “break” (i.e., not including the point at which the break occurs) is usually the number of factors to retain.

5.3.3 Rotation Methods

An important feature of factor analysis is that the axes of the factors can be rotated within the multidimensional variable space. Rotations that allow for correlation are called oblique rotations; rotations that assume the factors are not correlated are called orthogonal rotations.

Varimax is the most popular orthogonal rotation and Promax is the only oblique rotation method supported by Xarang.

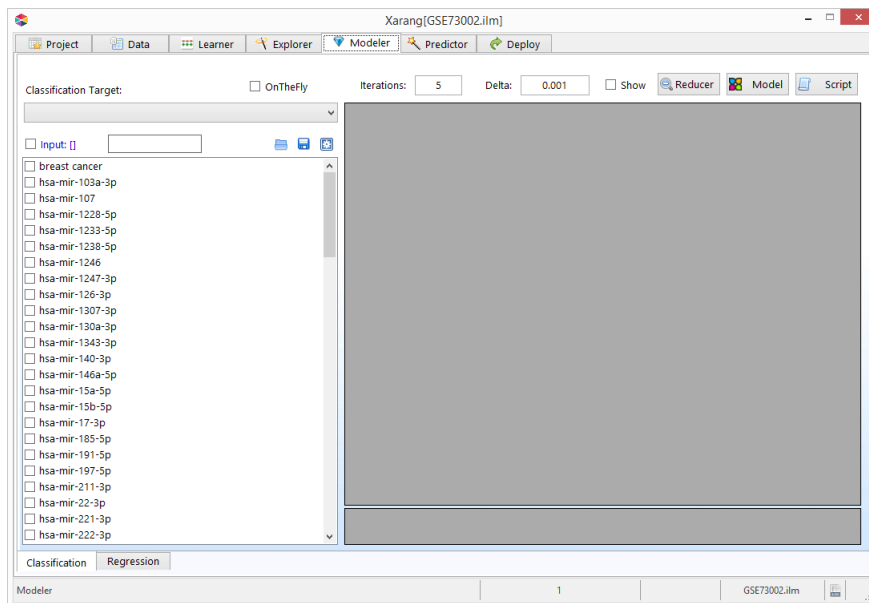
1. Equamax
2. Promax
3. Quartimax
4. Varimax

6

Modeler in Xarang

The Modeler constructs two types of predictive models:



- Classification
- Regression

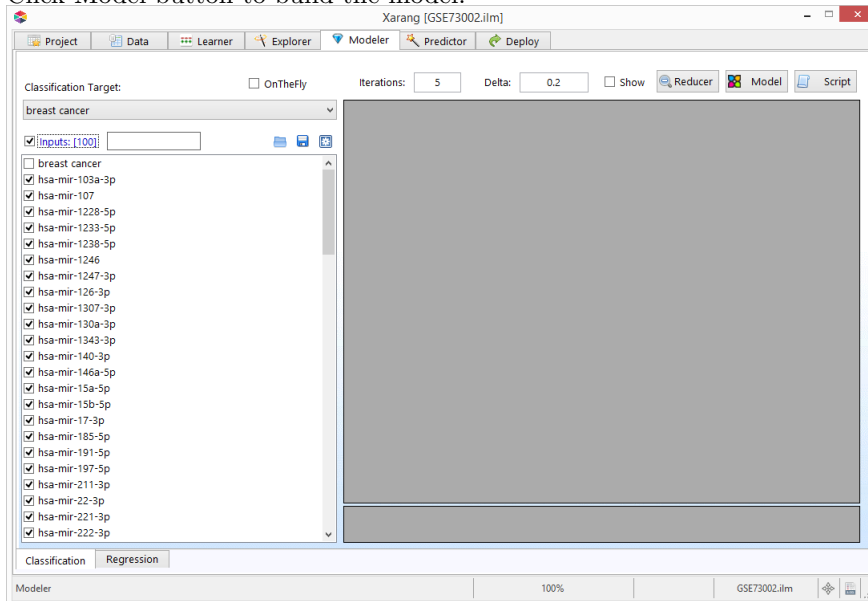


6.1 Binary Classification

Classification refers to the data mining task of attempting to build a predictive model when the target is categorical. If the number of unique values are just two (0,1) it is called Binary Classification. The main goal of classification is to divide a dataset into mutually exclusive groups such that the members of each group are as close as possible to one another, and different groups are as far as possible from one another.

1. Select the Classification tab on the bottom left of the window.

2. Select the Classification target from the dropdown list. Only binary variables will be displayed.
3. Select the input variables from the Inputs list.
4. Click  to save the selected variables.
5. Click  to open the selected variable list.
6. Click Model button to build the model.



7. To avoid attributes that do not contribute significantly to model prediction you can use the Reducer function. You can also adjust the Delta value and number of Iterations to influence the outcome of the Reducer. The Delta is the contribution threshold that a certain variable must provide to the model in order to be selected by the Reducer.

| Variables | Coefficient | Contribution |
|-----------------|-------------|--------------|
| hsa-mir-4783-3p | 7.98333 | 5.14444777 |
| hsa-mir-1228-5p | -7.48229 | 3.39649439 |
| hsa-mir-4730 | -1.84383 | 2.84824206 |
| hsa-mir-614 | -2.48036 | 2.83814539 |
| hsa-mir-6131 | 1.92669 | 2.77642228 |
| hsa-mir-642b-3p | 4.889 | 2.73678792 |

| Overall Contribution | Probability 1 | Probability 2 | Base Score |
|----------------------|---------------|---------------|------------|
| 6.980724 | 0.6821 | 0.3179 | -7.005966 |

8. If you did not use the Reducer, you will need to select one or more Input variables and build the model by clicking the Model button.
9. To change input variables in real time, check the OnTheFly checkbox. You can now select or unselect variables to instantly change and build the model.
10. To generate a script for the model, click the Script button. The Model Script window will open with the required scripts (Equation, SQL Script, VB Code and Java Code).

```

xarang_bc.breast cancer:-7.005966,0.5
hsa-mir-4783-3p,7.983330,6.231747,N
hsa-mir-1228-5p,-7.482290,11.733655,N
hsa-mir-4730,-1.843830,8.984137,N
hsa-mir-614,-2.480360,8.197558,N
hsa-mir-6131,1.926690,8.063808,N
hsa-mir-642b-3p,4.889000,8.903171,N

=====

-- SQL Function --
CREATE FUNCTION dbo.xarang_bc
(
  @hsa_mir_4783_3p FLOAT, @hsa_mir_1228_5p FLOAT, @hsa_mir_4730 FLOAT, @hsa_mir_614 FLOAT, @hsa_mir_6131 FLOAT, @hsa_mir_642b_3p FLOAT
)
RETURNS FLOAT
AS
BEGIN
  RETURN
    7.983330*@hsa_mir_4783_3p
    -7.482290*@hsa_mir_1228_5p
    -1.843830*@hsa_mir_4730
    -2.480360*@hsa_mir_614
    +1.926690*@hsa_mir_6131
    +4.889000*@hsa_mir_642b_3p
END



SELECT dbo.xarang_bc(hsa_mir_4783_3p, hsa_mir_1228_5p, hsa_mir_4730, hsa_mir_614, hsa_mir_6131, hsa_mir_642b_3p) as Score FROM TABLE
=====

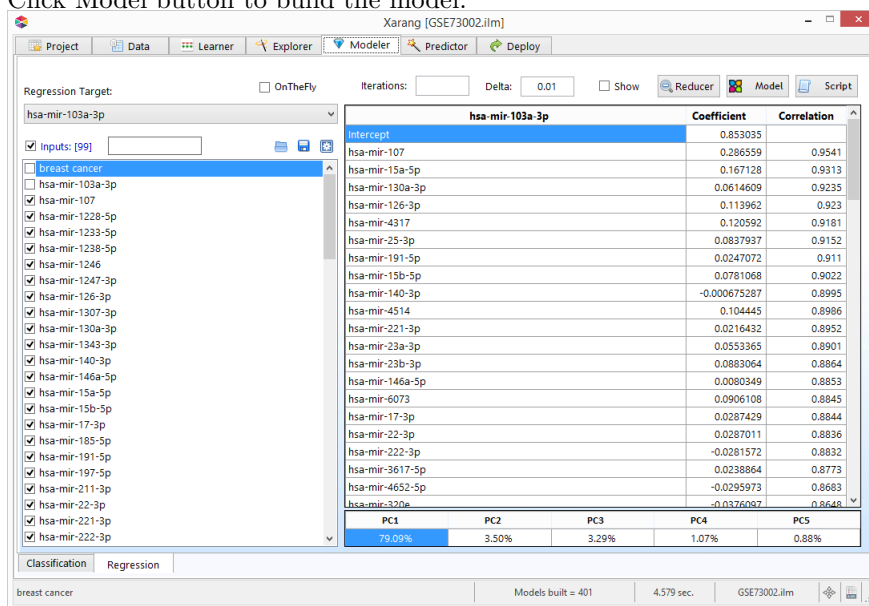
```

6.2 Regression

Regression refers to the data mining problem of attempting to build a predictive model when the target is numerical. The simplest form of regression, simple

linear regression, fits a line to a set of data.

1. Select the Regression tab on the bottom left of the window.
2. Select the Regression target from the dropdown list. You can select either a numeric or binary variable.
3. Select the input variables from the Inputs list.
4. Click  to save the selected variables.
5. Click  to open the selected variable list.
6. Click Model button to build the model.



Regression Target: OnTheFly Iterations: Delta: 0.01 Show

Regression Target: hsa-mir-103a-3p

Inputs: [99]

Inputs list:

- breast cancer
- hsa-mir-103a-3p
- hsa-mir-107
- hsa-mir-1238-5p
- hsa-mir-1238-5p
- hsa-mir-1246
- hsa-mir-1247-3p
- hsa-mir-126-3p
- hsa-mir-1307-3p
- hsa-mir-130a-3p
- hsa-mir-1343-3p
- hsa-mir-140-3p
- hsa-mir-146a-5p
- hsa-mir-15a-5p
- hsa-mir-15b-5p
- hsa-mir-17-3p
- hsa-mir-185-5p
- hsa-mir-191-5p
- hsa-mir-197-5p
- hsa-mir-211-3p
- hsa-mir-22-3p
- hsa-mir-221-3p
- hsa-mir-222-3p

| hsa-mir-103a-3p | Coefficient | Correlation |
|-----------------|--------------|-------------|
| Intercept | 0.853035 | |
| hsa-mir-107 | 0.286559 | 0.9541 |
| hsa-mir-15a-5p | 0.167128 | 0.9313 |
| hsa-mir-130a-3p | 0.0614609 | 0.9235 |
| hsa-mir-126-3p | 0.113962 | 0.923 |
| hsa-mir-4317 | 0.120592 | 0.9181 |
| hsa-mir-25-3p | 0.0837937 | 0.9152 |
| hsa-mir-191-5p | 0.0247072 | 0.911 |
| hsa-mir-15b-5p | 0.0781068 | 0.9022 |
| hsa-mir-140-3p | -0.000675287 | 0.8995 |
| hsa-mir-4514 | 0.104445 | 0.8986 |
| hsa-mir-221-3p | 0.0216432 | 0.8952 |
| hsa-mir-23a-3p | 0.0553365 | 0.8901 |
| hsa-mir-23b-3p | 0.0883064 | 0.8864 |
| hsa-mir-146a-5p | 0.0080349 | 0.8853 |
| hsa-mir-6073 | 0.0906108 | 0.8845 |
| hsa-mir-17-3p | 0.0287429 | 0.8844 |
| hsa-mir-22-3p | 0.0287011 | 0.8836 |
| hsa-mir-222-3p | -0.0281572 | 0.8832 |
| hsa-mir-3617-5p | 0.0238864 | 0.8773 |
| hsa-mir-4652-5p | -0.0295973 | 0.8683 |
| hsa-mir-320a | -0.0376097 | 0.8648 |

| PC1 | PC2 | PC3 | PC4 | PC5 |
|--------|-------|-------|-------|-------|
| 79.09% | 3.50% | 3.29% | 1.07% | 0.88% |

Classification Regression

breast cancer Models built = 401 4.579 sec. GSE73002.ilm

7. To avoid attributes that do not contribute significantly to model prediction you can use the Reducer function. You can also adjust the Delta value and number of Iterations to influence the outcome of the Reducer. The Delta is the contribution threshold that a certain variable must provide to the model in order to be selected by the Reducer.

Regression Target: OnTheFly Iterations: Delta: 0.01 Show

Inputs: [12]

breast cancer

- hsa-mir-103a-3p
- hsa-mir-107
- hsa-mir-1238-5p
- hsa-mir-1238-5p
- hsa-mir-1238-5p
- hsa-mir-1246
- hsa-mir-1247-3p
- hsa-mir-126-3p
- hsa-mir-1307-3p
- hsa-mir-130a-3p
- hsa-mir-1343-3p
- hsa-mir-140-3p
- hsa-mir-146a-5p
- hsa-mir-15a-5p
- hsa-mir-15b-5p
- hsa-mir-17-3p
- hsa-mir-185-5p
- hsa-mir-191-5p
- hsa-mir-197-5p
- hsa-mir-211-3p
- hsa-mir-22-3p
- hsa-mir-221-3p
- hsa-mir-222-3p

| hsa-mir-103a-3p | | Coefficient | Correlation |
|-----------------|--|-------------|-------------|
| Intercept | | -0.384258 | |
| hsa-mir-107 | | 0.866163 | 0.9541 |
| hsa-mir-4734 | | 0.0215231 | 0.7091 |
| hsa-mir-642b-3p | | 0.0719442 | 0.6742 |
| hsa-mir-1238-5p | | 0.0565726 | 0.6686 |
| hsa-mir-197-5p | | 0.0228075 | 0.6561 |
| hsa-mir-3151-5p | | 0.0295397 | 0.6537 |
| hsa-mir-885-3p | | 0.0226028 | 0.6403 |
| hsa-mir-4690-5p | | -0.0566438 | 0.599 |
| hsa-mir-6742-5p | | 0.0216705 | 0.5896 |
| hsa-mir-6729-5p | | 0.205252 | -0.5127 |
| hsa-mir-6816-5p | | 0.0147567 | -0.5733 |
| hsa-mir-1228-5p | | -0.295106 | -0.7311 |

| PC1 | PC2 | PC3 | PC4 | PC5 |
|--------|--------|-------|-------|-------|
| 72.69% | 10.28% | 3.95% | 3.28% | 2.45% |

Classification Regression

Models built = 1002 10.58 sec. GSE73002.ilm

8. If you did not use the Reducer, you will need to select one or more Input variables and build the model by clicking the Model button.
9. To change input variables in real time, check the OnTheFly checkbox. You can now select or unselect variables to instantly change and build the model.
10. To generate a script for the model, click the Script button. The Model Script window will open with the required scripts (Equation, SQL Script, VB Code and Java Code).

Save Model

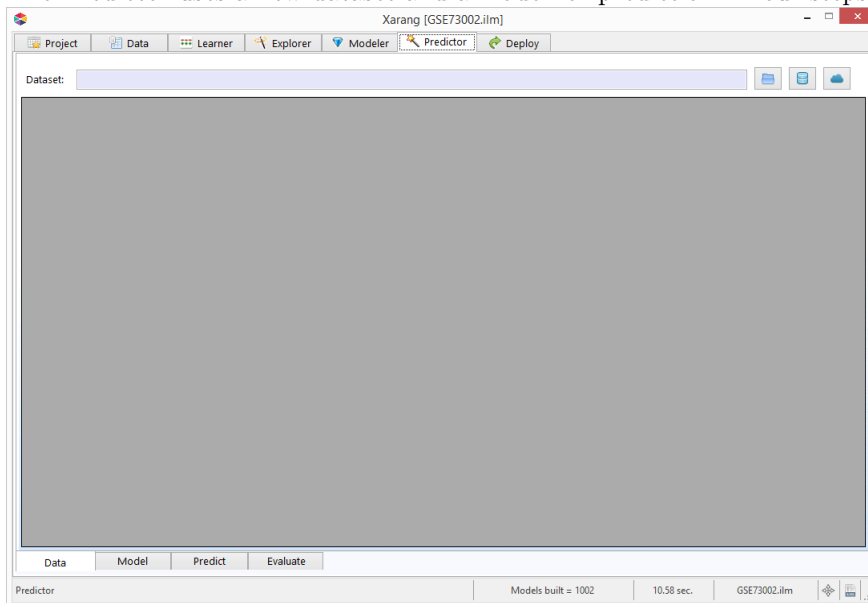
```
xarang_rr.hsa-mir-103a-3p.1.1
_intercept_-0.384258.0.N
hsa-mir-107.0.866163.3.204205.N
hsa-mir-4734.0.021523.12.509392.N
hsa-mir-642b-3p.0.071944.8.905171.N
hsa-mir-1238-5p.0.056573.5.861083.N
hsa-mir-197-5p.0.022808.7.491578.N
hsa-mir-3151-5p.0.029540.4.997937.N
hsa-mir-885-3p.0.022605.4.995746.N
hsa-mir-4690-5p.0.056644.5.960939.N
hsa-mir-6742-5p.0.021671.4.906696.N
hsa-mir-6729-5p.0.205252.12.575100.N
hsa-mir-6816-5p.0.014757.10.583252.N
hsa-mir-1228-5p.0.295106.11.733655.N
```

```
-- SQL Function --
CREATE FUNCTION dbo.xarang_rr
(
  @Intercept FLOAT, @hsa_mir_107 FLOAT, @hsa_mir_4734 FLOAT, @hsa_mir_642b_3p FLOAT, @hsa_mir_1238_5p FLOAT, @hsa_mir_197_5p FLOAT,
  @hsa_mir_3151_5p FLOAT, @hsa_mir_885_3p FLOAT, @hsa_mir_4690_5p FLOAT, @hsa_mir_6742_5p FLOAT, @hsa_mir_6729_5p FLOAT, @hsa_mir_6816_5p
  FLOAT, @hsa_mir_1228_5p FLOAT
)
RETURNS FLOAT
AS
BEGIN
  RETURN
    -0.384258*@Intercept
    +0.866163*@hsa_mir_107
    +0.021523*@hsa_mir_4734
    +0.071944*@hsa_mir_642b_3p
```

7

Predictor in Xarang

The Predictor uses a new dataset and a model for prediction in four steps:



7.1 Select a dataset

On the Predictor tab opens to the Data tab. Load the dataset from a local drive, a database or a Cloud service that you would like to use to make predictions.

The screenshot shows the Xarang software interface with a data table. The table has the following columns: geo_accession, hsa-miR-1307-3p, hsa-miR-4783-3p, hsa-miR-8073, hsa-miR-4532, hsa-miR-6787-5p, hsa-miR-6861-5p, hsa-miR-1233-5p, hsa-miR-4675, and hsa-miR-92a-2-5'. The rows contain numerical values for each of these identifiers.

| geo_accession | hsa-miR-1307-3p | hsa-miR-4783-3p | hsa-miR-8073 | hsa-miR-4532 | hsa-miR-6787-5p | hsa-miR-6861-5p | hsa-miR-1233-5p | hsa-miR-4675 | hsa-miR-92a-2-5' |
|---------------|-----------------|-----------------|--------------|--------------|-----------------|-----------------|-----------------|--------------|------------------|
| 1876423 | 9.227366 | 8.695286 | 7.007205 | 13.76236 | 8.732082 | 8.332469 | 12.43351 | 9.479579 | 7.434257 |
| 1876424 | 10.03153 | 9.47176 | 7.735587 | 14.29435 | 9.35634 | 8.492598 | 12.34068 | 9.364533 | 7.269479 |
| 1876426 | 9.092936 | 9.01008 | 7.308988 | 13.42158 | 9.182487 | 7.849389 | 13.20808 | 9.580367 | 6.773605 |
| 1876434 | 9.444392 | 9.140817 | 7.238101 | 14.04505 | 9.553611 | 8.252753 | 13.23759 | 9.347461 | 6.511993 |
| 1876435 | 8.959761 | 7.989191 | 6.783135 | 14.07617 | 8.888717 | 8.431629 | 11.56258 | 9.578112 | 7.178381 |
| 1876436 | 9.54439 | 8.625057 | 7.423399 | 14.30406 | 9.709041 | 8.619472 | 12.09327 | 9.735268 | 6.696023 |
| 1876440 | 9.867797 | 9.727315 | 7.621283 | 14.28897 | 8.67508 | 7.549451 | 12.09875 | 9.245347 | 4.940167 |
| 1876444 | 9.747062 | 8.994434 | 7.290945 | 14.81394 | 9.423681 | 8.285672 | 12.5677 | 9.579801 | 6.88196 |
| 1876445 | 8.842734 | 8.694386 | 6.837693 | 13.54887 | 8.568683 | 7.777652 | 13.88951 | 8.7743 | 7.261416 |
| 1876448 | 9.713699 | 9.598712 | 8.043567 | 14.41205 | 9.883377 | 8.775574 | 13.12187 | 9.651178 | 7.942512 |
| 1876453 | 10.24082 | 9.271383 | 8.310658 | 14.6707 | 9.496679 | 9.037232 | 12.01757 | 9.734557 | 7.237737 |
| 1876456 | 9.300765 | 8.84165 | 7.676995 | 13.96945 | 9.081041 | 8.423436 | 12.51888 | 9.437701 | 7.765482 |
| 1876460 | 8.766821 | 8.100648 | 7.16234 | 13.63278 | 9.084591 | 8.624661 | 11.08857 | 9.525038 | 8.156337 |
| 1876467 | 9.613644 | 9.368008 | 7.806978 | 13.96186 | 9.115763 | 8.272773 | 12.71169 | 9.021611 | 7.758393 |
| 1876470 | 9.282357 | 9.239434 | 7.306853 | 14.21551 | 9.323076 | 8.483621 | 12.45238 | 9.808333 | 7.448363 |
| 1876472 | 8.584151 | 8.591717 | 7.260065 | 13.4385 | 8.801219 | 8.168343 | 15.00761 | 9.622752 | 7.366673 |
| 1876476 | 8.264532 | 9.225657 | 7.773982 | 14.23271 | 9.703371 | 8.876334 | 13.67966 | 9.902658 | 6.412145 |
| 1876478 | 9.252275 | 9.268107 | 7.125161 | 14.42026 | 9.062788 | 8.134373 | 13.23411 | 9.309254 | 7.235056 |
| 1876481 | 10.26484 | 9.64172 | 8.017974 | 14.90087 | 9.528085 | 8.481521 | 13.33325 | 9.55059 | 6.436598 |
| 1876482 | 9.718472 | 9.692038 | 8.04897 | 14.578 | 9.568075 | 8.390263 | 13.31764 | 9.612351 | 6.640935 |
| 1876483 | 9.18927 | 9.120891 | 7.422479 | 14.36621 | 9.234879 | 8.608214 | 12.43797 | 9.674432 | 6.91387 |
| 1876484 | 8.551674 | 9.506077 | 7.040375 | 14.15666 | 9.295652 | 8.312333 | 12.69751 | 9.803865 | 6.680276 |
| 1876485 | 9.230846 | 9.096972 | 7.256461 | 14.19552 | 9.255583 | 8.105115 | 12.54229 | 8.861532 | 6.731208 |
| 1876487 | 8.812455 | 8.437155 | 7.782042 | 14.26802 | 8.608072 | 8.038166 | 12.44247 | 8.671606 | 7.632706 |

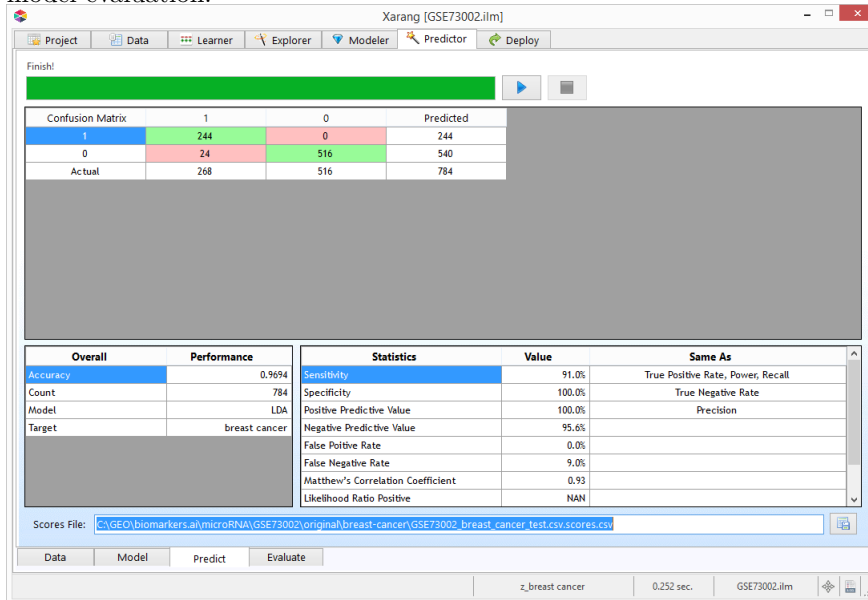
7.2 Select a model

1. Click the Model tab.
2. Select either Classification, Regression or MultiClass from the Model drop-down list.
3. Select one or more Input variables and a Target variable.
4. You can also append other variables to the output file by selecting them from the Key list.

The screenshot shows the Xarang software interface with the Model configuration screen. The Model is set to Classification. The Inputs list includes breast_cancer, hsa-mir-103a-3p, hsa-mir-107, hsa-mir-1228-5p, hsa-mir-1233-5p, hsa-mir-1238-5p, hsa-mir-1246, hsa-mir-1247-3p, hsa-mir-126-3p, hsa-mir-1307-3p, hsa-mir-130a-3p, hsa-mir-1343-3p, hsa-mir-140-3p, hsa-mir-146a-5p, hsa-mir-15a-5p, hsa-mir-15b-5p, hsa-mir-17-3p, hsa-mir-185-5p, hsa-mir-191-5p, hsa-mir-197-5p, and hsa-mir-211-3p. The Target is set to breast_cancer. The Key list includes breast_cancer, geo_accession, hsa-mir-103a-3p, hsa-mir-107, hsa-mir-1228-5p, hsa-mir-1233-5p, hsa-mir-1238-5p, hsa-mir-1246, hsa-mir-1247-3p, hsa-mir-126-3p, hsa-mir-1307-3p, hsa-mir-130a-3p, and hsa-mir-1343-3p.

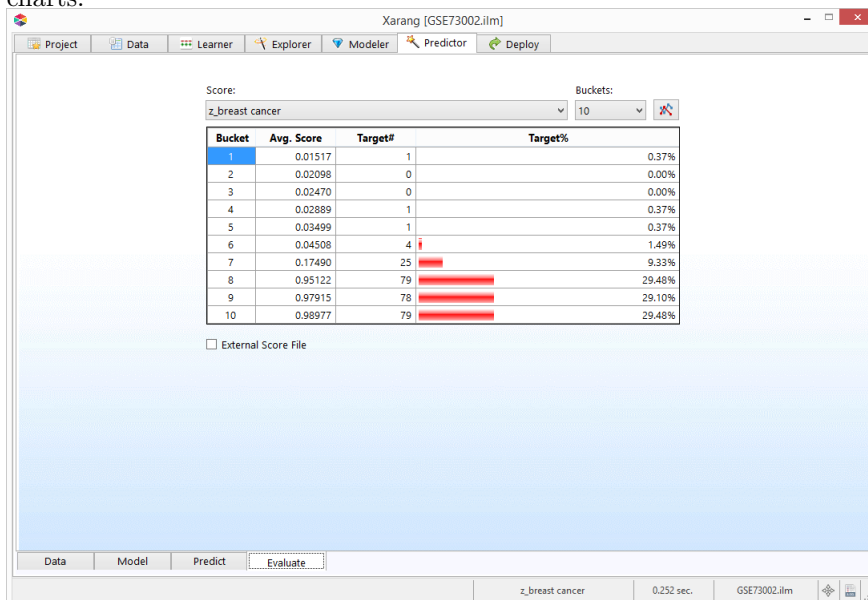
7.3 Predict using a model

To begin the Predictor, click the Start Predictor button. The results will be displayed and an output file will be created. Learn more about LDA, MLR and model evaluation.

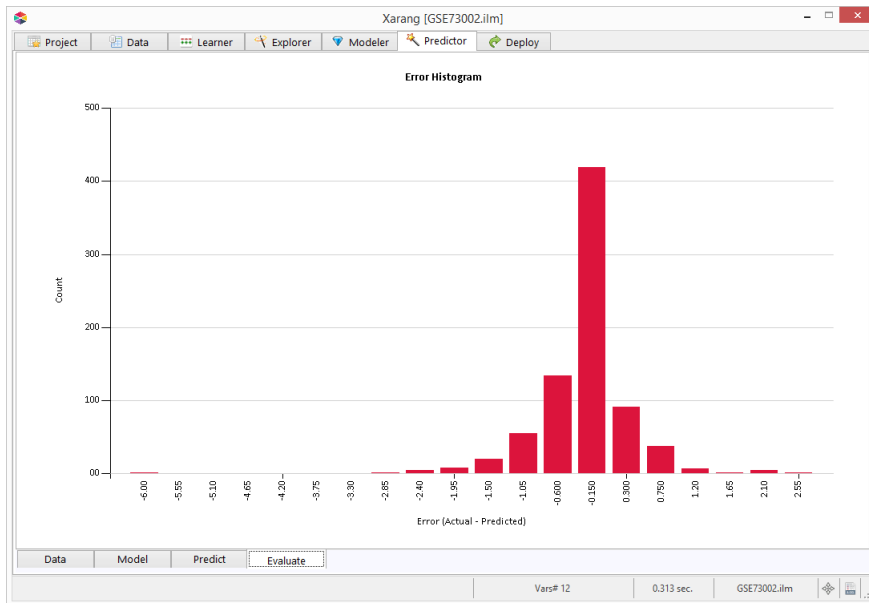


7.4 Evaluate Prediction Model

If you have used a Classification Model, click on  to view more evaluation charts.



This is Error Histogram for a regression model.

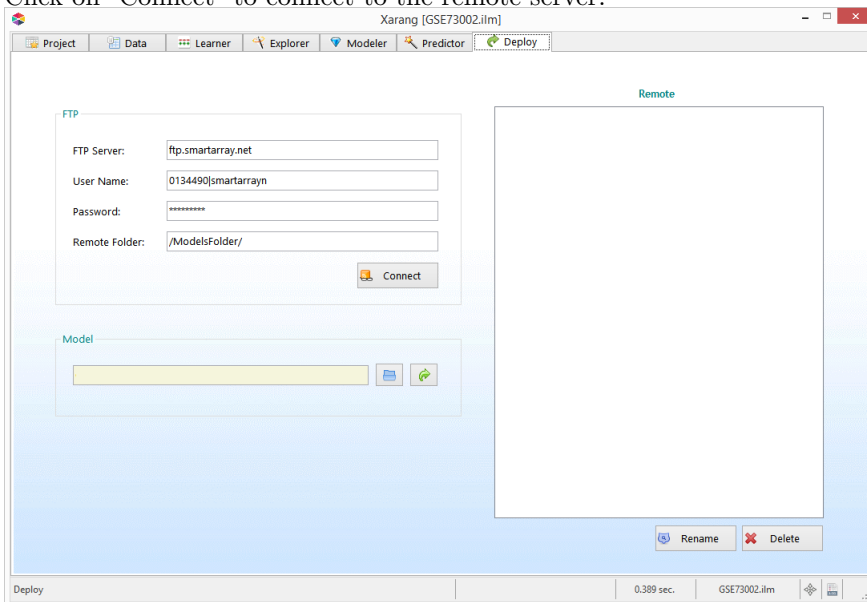


8

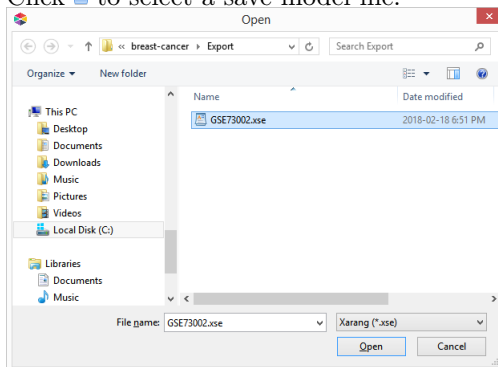
Deploying in Xarang


Models can be deployed to a remote server using FTP.

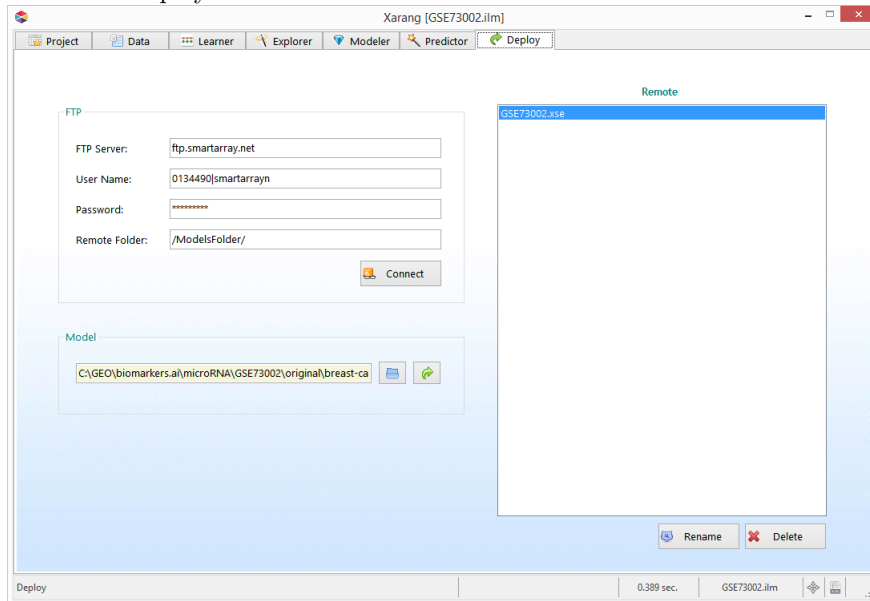
1. Click on "Connect" to connect to the remote server.



2. Click  to select a save model file.







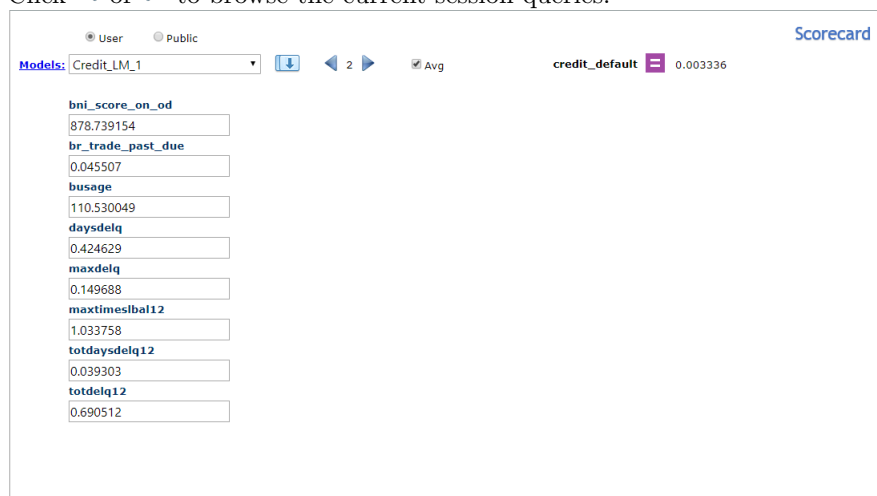
- Click  to deploy the model to the remote server.



- The models on the remote server can be renamed or deleted.



8.1 Scorecard

- Click "Models" to refresh the list.
- Select a model from the list and click  .
- Fill the scorecard or check "Avg" to fill the scorecard with the related average values.
- Click  to predict using the model.
- Click  or  to browse the current session queries.



8.2 A/B Test


A/B testing (also known Multivariate testing) is a method of comparing two versions of a model against each other to determine which one performs better.


1. Click  to refresh the list of models.
2. You can compare up to 5 models by selecting them from the corresponding list.
3. Assign a percentage between 1-100 to each model. The total percentage should equal to 100.
4. Select a dataset from the list.
5. Click .

A/B Test

User Public

| Group | Model | Traffic% | Count | Accuracy | Probability |
|----------|---|---------------------------------|-------|----------|-------------|
| A | <input type="text" value="Credit_LM_1"/> | <input type="text" value="50"/> | 1,774 | 87.66% | |
| B | <input type="text" value="Credit_NLM_1"/> | <input type="text" value="50"/> | 1,743 | 93.34% | 0.0000 |
| C | <input type="text"/> | <input type="text"/> | | | |
| D | <input type="text"/> | <input type="text"/> | | | |
| E | <input type="text"/> | <input type="text"/> | | | |



Datasets: 

Time Elapsed = 0.4 sec.