

User Manual a Bioada Product

Modelling $Team^1$

July 3, 2020

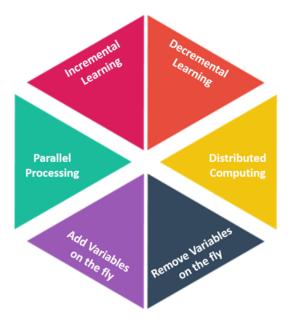
 1 www.bioada.com

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	Intr 1.1 1.2	voduction 2 What is Xarang 2 What can you do with Xarang 2
2	Pro	ject in Xarang 3
	2.1	Creating a new project
	2.2	Opening an existing Project
	2.3	Saving a project
	2.4	Merging two projects
3	Dat	a in Xarang 7
	3.1	File
	3.2	Folder
	3.3	Database
	3.4	Cloud
4	Lea	rner in Xarang 11
5	Exp	blorer in Xarang 13
	5.1	Univariate
	5.2	Bi-variate
		5.2.1 Correlation
		5.2.2 Hypothesis Testing
		5.2.3 ANOVA
		5.2.4 Test of Independence
	5.3	Multi-Variate
		5.3.1 Extraction Methods
		5.3.2 Number of Factors
		5.3.3 Rotation Methods
6	Mo	deler in Xarang 20
	6.1	Binary Classification
	6.2	Regression
7	Pre	dictor in Xarang 25
-	7.1	Select a dataset
	7.2	Select a model
	7.3	Predict using a model
	7.4	Evaluate Prediction Model 27

CONTENTS

8	Dep	oloying in	Xaı	ran	g														29
	8.1	Scorecard																	30
	8.2	A/B Test					•									•			31

1

1

Introduction

- 1.1 What is Xarang
- 1.2 What can you do with Xarang

Project in Xarang

 $\mathbf{2}$

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. PROJECT IN XARANG

)		Save As				
🖻 🌛 🔻 🕇 퉬 « biomarkers.ai 🕨 i	microRNA → GSE73002 → ori	ginal → breast-cancer	~ ¢	Search breast-c	ancer	J
Organize 👻 New folder						(
Mame Name This PC Desktop Desktop Desktop HT-C1800 F07C7F Wides Local Disk (C:) Ubraries	~	Date modified No items match your	Type search.	Size		
Documents Music						
File name: GSE73002.ilm						
Save as type: Project (*.ilm)						
Hide Folders				Save	Cancel	

2.2 Opening an existing Project

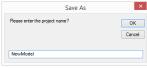
- 1. To open an existing project, click on the Open icon.
- 2. Select the project file (e.g., GSE73002.ilm) click the Open button.
- 3. Click the "Open" button.

-

۶			C	pen			
🗧 🕘 🕤 🚹 🚺	« biomark	ers.ai	৮ microRNA → GSE73002 → origin	al → breast-cancer	v C	Search breast-cancer	,
Organize 👻 Nev	v folder					H ·	
🐌 WinNBQ		^	Name	Date modified	Туре	Size	
鷆 WinNBQ 3.4 鷆 Xarang			GSE73002.ilm	2018-02-16 10:16	ILM File	1 KB	
🍓 Homegroup							
1 This PC							
) Desktop							
Downloads							
PHT-CT800 F07	C7F						
🚺 Music							
📔 Pictures							
Videos							
🚢 Local Disk (C:)							
Local Disk (C:)		~					
~	File <u>n</u> ame:	v GSE73	1002.ilm		~	Projects (*.ilm)	~

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.

Image Projects Image	TOP	10				A 14				
Browse For Folder					Merge Pro	ojects		-		
Image: Section of the section of th		New	Project N	lame:						
Image: Section of the section of th										
>				В	rowse For	Folder			×	
>										
>										
>			ь 📭	GSE6433	2				^	
> GST2025 > GST2057 > GST2057 > GST2002 > Brast-benign > Brast-benign > Brast-benign > GST2022 > Brast-benign										
Image: State Stat										
Image: Set										
Image: Cancel in the set of the set										
Werge Projects) Werge Projects) Werge Strate:) Constraints:) Constraints: <t< td=""><td></td><td></td><td>4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>			4							
Image: Second and Second										
Image:										
Images			Þ							
Images			Þ 鵳	GSE7646	2					
Image: Section and Sect					8					
AGEO/biomarkets.ai/microRNA/GSE73002/origina/bi										
VGEOLbiomarker: all/microBNA/GSE73002/originally Merge Projects										
Medic Hew Folder OK Cancel									~	
\GECVbiomarkers.al\microRNA\GSE73002\original\d										
Merge Projects Merge Projects Merge All GSE73002_All GSE73002_5.1im		Make	New Celd					Connel		
Merge Projects Merge Projects Merge All GSE73002_All GSE73002_5.1im		Make	New Fold			ОК	:	Cancel		
GSE73002_Ailm GSE73002_Bilm		Make	New Folds			ОК	:	Cancel		
G\$£73002_8.llm		biomark	ers.ai\mic	F	SE73002\ori					
] Mer	biomark rge Proje	ers.ai\mic cts	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
] Mer	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N ilm	er roRNA\G						×
	GSE7	biomark rge Proje / New 3002_A.i	ers.ai\mic cts Project N Im Im	er roRNA\G						· · · · · · · · · · · · · · · · · · ·

- 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.

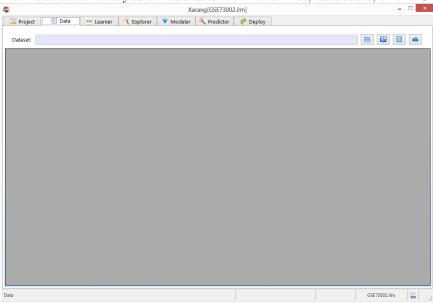
2. PROJECT IN XARANG

		Open				
🗧 🌛 🔹 🕇 📕 🐇 orig	inal ⊦ breas	t-cancer	~ Ċ	Search breast	-cancer	, P
Organize 👻 New folder					800 -	
	^ Na	ame	^		Date mod	dified
🖳 This PC		GSE73002 A.ilm			2018-02-1	16 10:27
E Desktop		GSE73002_B.ilm			2018-02-1	16 10:27
Documents Downloads	1	merged_2.ilm			2018-02-1	16 10:32
HT-CT800 F07C7F						
Music						
E Pictures						
崖 Videos						
📥 Local Disk (C:)						
词 Libraries						
Documents	~ <					
File par	ne: merged	2 ilm	Ý	Projects (*.ilr	20)	~
The Har	incigeo_				-	
				Open	0	ancel

Data in Xarang

3

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



3.1 File

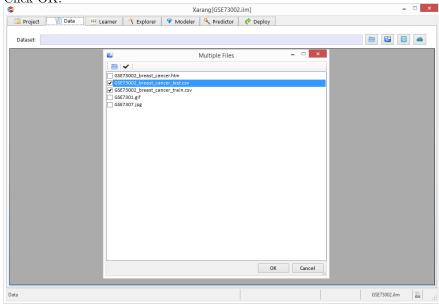
- 1. To load a local dataset, click the Open File \blacksquare button.
- 2. Select the data file.
- 3. Click Open.

3. DATA IN XARANG

į,	Project	Data 🚥	Learner 🛛 🜱 E	Explorer 🛛 💎 Mo	deler 🛛 🤻 Pred	ictor 🛛 🥏 Dep	loy			
at	a File: C:\GEO\b	iomarkers.ai\m	icroRNA\GSE730	02\original\breast-c	ancer\GSE73002_	breast_cancer_tra	ain.csv			
	geo_accession	hsa-miR-130	7-3p hsa-miR-478	3-3p hsa-miR-8073	hsa-miR-4532	hsa-miR-6787-5	ip hsa-miR-6861-5p	hsa-miR-1233-5p	hsa-miR-4675	hsa-miR-92a-2-
	1876422	7.826695	9.315201	6.367992	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	9.680874	6.163979
	1876427	9.11246	8.549548	7.190443	13.31739	9.162703	8.13832	12.60987	9.116076	7.157032
	1876428	9.435894	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	7.769321	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 ${\ensuremath{\ensuremath{\mathbb{G}}}}$ 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 \blacksquare button.
- 2. Enter the DSN, User ID/ Password (if needed) in the related fields and click $\blacksquare.$
- 3. Select one or more tables from the list. All tables must have the same schema.
- 4. Click OK. _ 🗆 🗙 Database DSN: DSN=GEO64 lser ID: Catalogue.Owner: 8 geo.dbo Tables O SQL Script Tables SQL Script
 GSE57225_target
 GSE69967
 GSE69967
 GSE69967_genes
 GSE50967_target
 GSE73002_breast_cancer_test
 GSE73002_breast_cancer_train ILMNProbes hsa-miR-1233-5p hsa-miR-6787 geo_accession hsa-miR-1307-3p hsa-miR-4532 hsa-miR-4783-3p hsa-miR-8073 1876425 8.651167 13,71886 9.051309 6.796376 12.71636 9.09567 9.11246 9.435894 8.549548 9.556254 7.190443 7.449193 9.162703 1876427 13.31739 12.60987 1876428 14.16441 12.78285 8.458979 8.62834 8.897826 8.987145 7.310657 6.677789 1876429 13.47416 12.17858 9.067359 1876430 13.39477 13.03032 9.119116 1876431 8.92551 13.68114 8.95489 7.692006 13.16057 9.258071 1876432 8.828325 14.01279 9.247807 7.138846 11.55853 9.262071 ОК Cancel
- 5. You can also enter, save or load SQL script for extracting data.

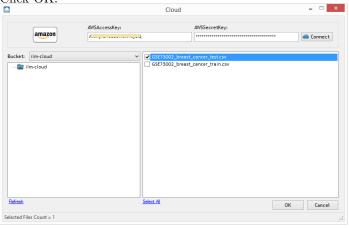
50/1		Usi	er ID: Passv		logue.C	Jwner:		
EU04				geo	.000			
bles 💿 SQL Sc	ript						Þ 🖴 6	
from GSE7300	2_breast_cancer_	train						^
								~
								^
geo_accession	hsa-miR-1307-3p	hsa-miR-4532	hsa-miR-4783-3p	hsa-miR-8073	hsa-	miR-1233-5p	hsa-miR-67	37
1876422	7.826695		9.315201					
1876422 1876425	7.826695 8.651167	14.32252 13.71886	9.315201 9.051309	6.367992 6.796376	11.83 12.71		9.210176 9.09567	
						636		
1876425	8.651167	13.71886	9.051309	6.796376	12.71	636 1987	9.09567	
1876425 1876427	8.651167 9.11246	13.71886 13.31739	9.051309 8.549548	6.796376 7.190443	12.71	636 1987 1285	9.09567 9.162703	
1876425 1876427 1876428	8.651167 9.11246 9.435894	13.71886 13.31739 14.16441	9.051309 8.549548 9.556254	6.796376 7.190443 7.449193	12.71 12.60 12.78	636 1987 1285 1858	9.09567 9.162703 9.247772	
1876425 1876427 1876428 1876429	8.651167 9.11246 9.435894 8.458979	13.71886 13.31739 14.16441 13.47416	9.051309 8.549548 9.556254 8.897826	6.796376 7.190443 7.449193 7.310657	12.71 12.60 12.78 12.17	636 1987 1285 1858 1032	9.09567 9.162703 9.247772 9.067359	
1876425 1876427 1876428 1876429 1876430	8.651167 9.11246 9.435894 8.458979 8.62834	13.71886 13.31739 14.16441 13.47416 13.39477	9.051309 8.549548 9.556254 8.897826 8.987145	6.796376 7.190443 7.449193 7.310657 6.677789	12.71 12.60 12.78 12.17 13.03	636 1987 1285 1858 1032 1057	9.09567 9.162703 9.247772 9.067359 9.119116	
1876425 1876427 1876428 1876429 1876430 1876431	8.651167 9.11246 9.435894 8.458979 8.62834 8.92551	13.71886 13.31739 14.16441 13.47416 13.39477 13.68114	9.051309 8.549548 9.556254 8.897826 8.987145 8.95489	6.796376 7.190443 7.449193 7.310657 6.677789 7.692006	12.71 12.60 12.78 12.17 13.03 13.16	636 1987 1285 1858 1032 1057	9.09567 9.162703 9.247772 9.067359 9.119116 9.258071 9.262071	
	from GSE7300	Nes	Nes © SQL Script from GSE73002_breast_cancer_train	from GSE73002_breast_cancer_train	les © SQL Script from GSE73002_breast_cancer_train	les SQL Script from GSE73002_breast_cancer_train	les © SQL Script from CSE73002_breast_cancer_train	Ites © SQL Script from CSE73002_breast_cancer_train

3.4 Cloud

- 1. To load data from the cloud, click the Open Cloud ${\color{red} {\bullet}}$ button.
- 2. Enter the AWS Access Key and AWS Secret Key.
- 3. Click the Connect button.

3. DATA IN XARANG

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



Learner in Xarang

1. Select "Target" variable.

4

- 2. Click \bigcirc 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 .

·····	Xarang[GSE73002.ilm]	
🐺 Project 🛛 🔡 Data 🛛 🎫 Le	earner 🌱 Explorer 🛛 💎 Modeler 🤻 Predictor 📌 Deploy	
Target:		
breast cancer	~	
Schema:	——————————————————————————————————————	Decremental
Variable	Туре	
geo_accession		
hsa-mir-1307-3p	123 Numeric	
hsa-mir-4783-3p	String	
hsa-mir-8073	E List	
hsa-mir-4532	🛞 Ignore	
hsa-mir-6787-5p	N	
hsa-mir-6861-5p	N	
hsa-mir-1233-5p	N	
hsa-mir-4675	N	
hsa-mir-92a-2-5p	N	
hsa-mir-6756-5p	N	
hsa-mir-1228-5p	N	
hsa-mir-6766-5p	N	
hsa-mir-4758-5p	N	
hsa-mir-6802-5p	N	
hsa-mir-6769a-5p	N	
hsa-mir-4419b	N	
hsa-mir-6784-5p	N	
hsa-mir-4428	N	
hsa-mir-342-5p	N	
hsa-mir-885-3p	N	
hsa-mir-8059	N	
hsa-mir-1238-5p	N V	

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

4. LEARNER IN XARANG

Project 🛛 🔛 Data	🎫 Learner 🛛 🐴	Explorer		Modeler 🔍	Predictor 🥏	Deploy				
- Hojett - Butu		- Explorer		Anoucles V	(Calcion	Depioy				
Target:				Count= 3180						
breast cancer			¥							
Schema:			2					Decreme	ntal	
Variable		Туре	^	x	Y	Count	SumX	SumY	SumXY	Su
				hsa-mir-3617						843
hsa-mir-1307-3p		N		hsa-mir-4299	hsa-mir-4730	3180	11073.047116	28569.55415	94162.009422	3264
hsa-mir-4783-3p		N		hsa-mir-6858	hsa-mir-4317	3180	26082.088826	8956.6315705	75669.433707	259
hsa-mir-8073		N		hsa-mir-197-5p	hsa-mir-4261	3180	23823.218595	9490.4312665	73751.278737	230
hsa-mir-4532		N		hsa-mir-6787	hsa-mir-185-5p	3180	26572.039734	10742.7873992	92295.285359	337
hsa-mir-6787-5p		N		hsa-mir-6717	hsa-mir-320c	3180	18207.95712	11811.1918142	73600.684353	240
hsa-mir-6861-5p		N		hsa-mir-1307	hsa-mir-1343	3180	20176.460186	19493.607815	127653.08907	6079
hsa-mir-1233-5p		N		hsa-mir-6858	hsa-mir-320b	3180	26082.088826	13458.430689	112314.27287	4659
hsa-mir-4675		N		hsa-mir-3652	hsa-mir-4652	3180	17651.779029	9531.6785507	55911.308720	1443
hsa-mir-92a-2-5p		N		hsa-mir-15b-5p	hsa-mir-4538	3180	9721.1870189	9852.3426383	36264.825674	7718
hsa-mir-6756-5p		N		hsa-mir-4532	hsa-mir-6756	3180	37819.793448	26387.835696	315909.50815	324
hsa-mir-1228-5p		N		hsa-mir-146a	hsa-mir-3605	3180	9539.4763515	9234.9866194	35117.338561	7649
hsa-mir-6766-5p		N		hsa-mir-6858	hsa-mir-320d	3180	26082.088826	10426.299679	87643.167455	326
hsa-mir-4758-5p		N		hsa-mir-6729	hsa-mir-6742	3180	39988.818160	15603.293363	195189.40462	122
hsa-mir-6802-5p		N		hsa-mir-1307	hsa-mir-4758	3180	20176.460186	27358.405775	177151.15526	113
hsa-mir-6769a-5p		N		hsa-mir-4515	hsa-mir-320c	3180	13162.855749	11811.1918142	56097.312238	1664
hsa-mir-4419b		N		hsa-mir-4515	hsa-mir-320b	3180	13162.855749	13458.430689	62949.687282	200
hsa-mir-6784-5p		N		hsa-mir-4515	hsa-mir-320e	3180	13162.855749	9779.2163157	49321.108774	145
hsa-mir-4428		N		hsa-mir-4515	hsa-mir-320d	3180	13162.855749	10426.299679	52056.409137	156
hsa-mir-342-5p		N		hsa-mir-4727	hsa-mir-4299	3180	13145.0730264	11073.047116	53139.308991	152
hsa-mir-885-3p		N		hsa-mir-8073	hsa-mir-1247	3180	18582.975499	19886.631499	118692.61353	492
hsa-mir-8059		N		hsa-mir-4783	hsa-mir-6742	3180	19816.954327	15603.293363	101006.78232	3864
hsa-mir-1238-5p		N	~	<						>

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

😨 Project 🛛 🔛 Data 🛛 🚥 Learr	ier 🌱 Explorer	5	Modeler 🤻	Predictor 🔗	Deploy			
	(capitorer		(inductor)	Treateror (bepioy			
larget:			Count= 3180					
breast cancer		~						
Schema:	B ()					✓ Decrementa	al
Variable	Туре	^	x	Y	Count	SumX	SumY	SumXY Su
geo_accession	Х		hsa-mir-4299	hsa-mir-4730	0	0	0	0
hsa-mir-1307-3p	N		hsa-mir-6858	hsa-mir-4317	0	0	0	0
hsa-mir-4783-3p	N		hsa-mir-197-5p	hsa-mir-4261	0	0	0	0
hsa-mir-8073	N		hsa-mir-6787	hsa-mir-185-5p	0	0	0	0
hsa-mir-4532	N		hsa-mir-320b	hsa-mir-1247	0	0	0	0
hsa-mir-6787-5p	N		hsa-mir-4259	hsa-mir-221-3p	0	0	0	0
hsa-mir-6861-5p	N		hsa-mir-1307	hsa-mir-1343	0	0	0	0
hsa-mir-1233-5p	N		hsa-mir-6858	hsa-mir-320b	0	0	0	0
hsa-mir-4675	N		hsa-mir-6756	hsa-mir-1343	0	0	0	0
hsa-mir-92a-2-5p	N		hsa-mir-4646	hsa-mir-4732	0	0	0	0
hsa-mir-6756-5p	N		hsa-mir-4532	hsa-mir-6756	0	0	0	0
hsa-mir-1228-5p	N		hsa-mir-146a	hsa-mir-3605	0	0	0	0
hsa-mir-6766-5p	N		hsa-mir-4317	hsa-mir-4652	0	0	0	0
hsa-mir-4758-5p	N		hsa-mir-614	hsa-mir-25-3p	0	0	0	0
hsa-mir-6802-5p	N		hsa-mir-4515	hsa-mir-320c	0	0	0	0
hsa-mir-6769a-5p	N		hsa-mir-4515	hsa-mir-320b	0	0	0	0
hsa-mir-4419b	N		hsa-mir-4515	hsa-mir-320e	0	0	0	0
hsa-mir-6784-5p	N		hsa-mir-4515	hsa-mir-320d	0	0	0	0
hsa-mir-4428	N		hsa-mir-4727	hsa-mir-4299	0	0	0	0
hsa-mir-342-5p	N		hsa-mir-8073	hsa-mir-1247	0	0	0	0
hsa-mir-885-3p	N		hsa-mir-4299	hsa-mir-6717	0	0	0	0
hsa-mir-8059	N		hsa-mir-130a	hsa-mir-6742	0	0	0	0
hsa-mir-1238-5p	N	~	<					>

$\mathbf{5}$

Explorer in Xarang

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

\$	Xa	arang[GSE73002.ilm]		- 🗆 🗙
📴 Project 🛛 🔠 Data 🛛 🎫 Learner		🤻 Predictor 🛛 🥐 Deploy		
Variable 1:	All	Variable 2:		All
breast cancer hsamir-1028-3p hsamir-1028-3p hsamir-1028-5p hsamir-1238-5p hsamir-1246-5p hsamir-136-3p hsamir-136-3p hsamir-136-3p hsamir-136-3p hsamir-136-3p hsamir-136-3p hsamir-136-3p hsamir-136-5p hsamir-136-5p		 ▲ breast cancer baa-mir-103m-3p baa-mir-103m-3p baa-mir-1228-5p baa-mir-1238-5p baa-mir-1238-5p baa-mir-1246 baa-mir-1246 baa-mir-130m-3p baa-mir-146a-5p baa-mir-146a-5p 		v
	ount Sum	Mean Variance	StDev CV	Skewness Kurtosis
hsa-mir-103a-3p	3180 9880.66	3.10713 4.0844	2.02099 65.04%	0.74 -0.02
Univariate			1	GSE73002.ilm

- 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.

5. EXPLORER IN XARANG

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

🖉 Project 🛛 🔡 Data	🚥 Learner 🛛 😤 Explo	rer 🛛 💎 Mod	eler 🛛 💐 Predio	ctor 🛛 🥐 Deplo	у			
Variable 1:			<u>All</u> V	/ariable 2:				All
breast cancer				breast cancer				^
hsa-mir-103a-3p				hsa-mir-103a-3p				
hsa-mir-107				hsa-mir-107				
hsa-mir-1228-5p				hsa-mir-1228-5p				
hsa-mir-1233-5p				hsa-mir-1233-5p				
hsa-mir-1238-5p hsa-mir-1246				hsa-mir-1238-5p hsa-mir-1246				
hsa-mir-1247-3p				hsa-mir-1247-3p				
hsa-mir-126-3p				hsa-mir-126-3p				
hsa-mir-1307-3p				hsa-mir-1307-3p				
hsa-mir-130a-3p				hsa-mir-130a-3p				
hsa-mir-1343-3p				hsa-mir-1343-3p				
hsa-mir-140-3p				hsa-mir-140-3p				
				hsa-mir-146a-5p				
hsa-mir-146a-5p								
hsa-mir-146a-5p hsa-mir-15a-5p Inivariate Correlation Hypot	hesis Testing ANOVA T	est of Independe	v	hsa-mir-15a-5p				~
hsa-mir-146a-5p hsa-mir-15a-5p	hesis Testing ANOVA To Count	est of Independe Sum	v	hsa-mir-15a-5p	StDev	cv	Skewness	v Kurtosis
hsa-mir-146a-5p hsa-mir-15a-5p Inivariate Correlation Hypot Variable	- 1		ence Factor Ana	hsa-mir-15a-5p	StDev 0.46567	CV 146.47%	Skewness 0.78	
hsa-mir-146a-5p hsa-mir-15a-5p Inivariate Correlation Hypot Variable preast cancer	Count	Sum	ence Factor Anal	hsa-mir-15a-5p lysis Variance				Kurtosis
hsa-mir-146a-Sp hsa-mir-15a-Sp Inivariate Correlation Hypot Variable preast cancer nsa-mir-103a-3p	Count 3180	Sum 1011	rnce Factor Anal Mean 0.317925	hsa-mir-15a-5p lysis Variance 0.216849	0.46567	146.47%	0.78	Kurtosis -1.39
hsa-mir-146a-5p hsa-mir-15a-5p Inivariate Correlation Hypot Variable preast cancer nsa-mir-103a-3p nsa-mir-107	Count 3180 3180	Sum 1011 9880.66	ence Factor Anal Mean 0.317925 3.10713	hsa-mir-15a-5p lysis Variance 0.216849 4.0844	0.46567 2.02099	146.47% 65.04%	0.78 0.74	Kurtosis -1.39 -0.02
hsa-mir-146a-Sp hsa-mir-15a-Sp Inivariate Correlation Hypot Variable sa-mir-107 sa-mir-107 sa-mir-128-Sp	Count 3180 3180 3180 3180	Sum 1011 9880.66 10189.37	rnce Factor Ana Mean 0.317925 3.10713 3.20421	hsa-mir-15a-5p lysis Variance 0.216849 4.0844 4.03717	0.46567 2.02099 2.00927	146.47% 65.04% 62.71%	0.78 0.74 0.71	Kurtosis -1.39 -0.02 -0.02
hsa-mir-146a-5p hsa-mir-15a-5p nivariate Correlation Hypot Variable reast cancer sa-mir-103a-3p sa-mir-107 tsa-mir-1235-5p tsa-mir-1233-5p	Count 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02	rnce Factor Anal Mean 0.317925 3.10713 3.20421 11.7337	hsa-mir-15a-5p lysis Variance 0.216849 4.0844 4.03717 0.2564	0.46567 2.02099 2.00927 0.50636	146.47% 65.04% 62.71% 4.32%	0.78 0.74 0.71 -0.8	Kurtosis -1.39 -0.02 -0.02 0.21
hsa-mir-146a-5p hsa-mir-15a-5p nivariate <u>Correlation</u> Hypot <u>Variable</u> sta-mir-103-3p nsa-mir-107 sa-mir-1238-5p nsa-mir-1238-5p	Count 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15	rnce Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557	hsa-mir-15a-5p lysis Variance 0.216849 4.0844 4.03717 0.2564 1.14236	0.46567 2.02099 2.00927 0.50636 1.06881	146.47% 65.04% 62.71% 4.32% 9.50%	0.78 0.74 0.71 -0.8 0.94	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27
hsa-mir-146a-5p hsa-mir-15a-5p Variable variable stas-mir-103a-3p sas-mir-107 sas-mir-107 sas-mir-1228-5p sas-mir-1228-5p sas-mir-1228-5p sas-mir-1246	Count 3180 3180 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15 18638.24	rnce Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557 5.86108	hsa-mir-15a-5p lysis 0.216849 4.0844 4.03717 0.2564 1.14236 1.06906	0.46567 2.02099 2.00927 0.50636 1.06881 1.03395	146.47% 65.04% 62.71% 4.32% 9.50% 17.64%	0.78 0.74 0.71 -0.8 0.94 0.72	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27 0.54
Insa-mir-146a-Sp hsa-mir-15a-Sp inivariate Correlation Hypot variable vrasble variable vrasble variable sa-mir-103a-3p sa-mir-103 sa-mir-1238-5p sa-mir-1238-5p sa-mir-1238-5p sa-mir-1248-5p sa-mir-1246-5p sa-mir-1246 sa-mir-1246-3p sa-mir-1247-3p	Count 3180 3180 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15 18638.24 14978.86	rnce Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557 5.86108 4.71033	hsa-mir-15a-5p lysis Variance 0.216849 4.0844 4.03717 0.2564 1.14236 1.06906 8.51288	0.46567 2.02099 2.00927 0.50636 1.06881 1.03395 2.91768	146.47% 65.04% 62.71% 4.32% 9.50% 17.64% 61.94%	0.78 0.74 0.71 -0.8 0.94 0.72 0.59	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27 0.54 -0.74
hsa-mir-146a-5p hsa-mir-15a-5p Variable	Count 3180 3180 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15 18638.24 14978.86 19886.63	rnce Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557 5.86108 4.71033 6.25366	hsa-mir-15a-5p Variance 0.216849 4.0844 4.03717 0.2564 1.14236 1.06906 8.51288 0.569432	0.46567 2.02099 2.00927 0.50636 1.06881 1.03395 2.91768 0.754607	146.47% 65.04% 62.71% 4.32% 9.50% 17.64% 61.94% 12.07%	0.78 0.74 0.71 -0.8 0.94 0.72 0.59 0.8	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27 0.54 -0.74 0.1
Isa-mir-146a-5p Isa-mir-15a-5p Variable Variable costs costs sa-mir-103a-3p sa-mir-103a-3p sa-mir-1228-5p sa-mir-1228-5p sa-mir-1228-5p sa-mir-1228-5p sa-mir-1246 sa-mir-1247-3p sa-mir-1247-3p sa-mir-1307-3p	Count 3180 3180 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15 18638.24 14978.86 19886.63 9058.02	Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557 5.86108 4.71033 6.25366 2.84843	hsa-mir-15a-5p Variance 0.216849 4.0844 4.03717 0.2564 1.14236 1.06906 8.51288 0.569432 2.71116	0.46567 2.02099 2.00927 0.50636 1.06881 1.03395 2.91768 0.754607 1.64656	146.47% 65.04% 62.71% 4.32% 9.50% 17.64% 61.94% 12.07% 57.81%	0.78 0.74 0.71 -0.8 0.94 0.72 0.59 0.8 0.41	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27 0.54 -0.74 -0.74 0.1
hsa-mir-146a-5p hsa-mir-15a-5p Inivariate Correlation Hypot	Count 3180 3180 3180 3180 3180 3180 3180 3180	Sum 1011 9880.66 10189.37 37313.02 35793.15 18638.24 14978.86 19886.63 9058.02 20176.46	Factor Anal Mean 0.317925 3.10713 3.20421 11.7337 11.2557 5.86108 4.77033 6.25366 2.84843 6.3448	hsa-mir-15a-5p Variance 0.216849 4.0844 4.03717 0.2564 1.14236 1.06906 8.51288 0.569432 2.771116 3.70182	0.46567 2.02099 2.00927 0.50636 1.06681 1.03395 2.91768 0.754607 1.64655 1.92401	146.478 65.048 62.718 4.328 9.508 17.648 61.948 12.078 57.818 30.328	0.78 0.74 0.71 -0.8 0.94 0.72 0.59 0.8 0.41 0.86	Kurtosis -1.39 -0.02 -0.02 0.21 -0.27 0.54 -0.74 -0.74 0.1 -0.48 -0.95

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.

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

5. EXPLORER IN XARANG

Project 🛛 🖓 Data 🛛 🚥 Learner	arr Explorer 🛛 😵 Modeler 🦂 Predictor 🛛 🥐 Deploy		
🚪 Project 🛛 🔡 Data 🛛 🎹 Learner	🌱 Explorer 🛛 🗣 Modeler 🛛 🤻 Predictor 🛛 🥐 Deploy		
Variable 1:	All Variable 2:		All
breast cancer has min-107 bas-min-1228-5p has-min-1228-5p has-min-1228-5p has-min-1248-6p has-min-1247-3p has-min-1247-3p has-min-1247-3p has-min-1307-3p has-min-1307-3p	 breast cancer hsa-mir-107 hsa-mir-1225-5p hsa-mir-1225-5p hsa-mir-1235-5p hsa-mir-1235-5p hsa-mir-1247-3p hsa-mir-1247-3p hsa-mir-1307-3p hsa-mir-1307-3p 		
hsa-mir-1343-3p hsa-mir-140-3p hsa-mir-146a-5p hsa-mir-15a-5p	hsa-mir-1343-3p hsa-mir-140-3p hsa-mir-146a-5p hsa-mir-146a-5p hsa-mir-146a-5p		ب <u>الله</u>
hsa-mir-1343-3p hsa-mir-140-3p hsa-mir-146a-5p hsa-mir-15a-5p	hsa-mir-140-3p hsa-mir-146a-5p	Covariance	
hsa-mir-1343-3p hsa-mir-140-3p hsa-mir-140-5p hsa-mir-15a-5p Jnivariate Correlation Hypothesis Testing A Variable 1	NOVA Test of Independence Factor Analysis	Covariance 1.09392	dl <u>Plot</u>
Isa-mir 1343-3p Isa-mir 146a-3p Isa-mir 146a-3p Isa-mir 15a-5p Isa-mir 15a-5p Isa-mir 15a-5p Isa-mir 15a-5p Isa-mir 146a-3p Isa-mir 15a-5p Isa-mir 15a-5p Isa-15a-5p	ANOVA Test of Independence Factor Analysis Variable 2		dll <u>Plot</u> Correlation
Isa-mir-1033-3p Isa-mir-1463-5p Isa-mir-1463-5p Isa-mir-15a-5p Variable 1 Variable 1 breast cancer Isa-mir-103a-3p	ANOVA Test of Independence Factor Analysis bas-mir-148-5p hsa-mir-15a-5p tsa-mir-15a-5p Variable 2 hsa-mir-1246	1.09392	طاً <u>Plot</u> Correlation 0.805136
Isamir 1343-3p Isamir 1463-5p Isamir 1463-5p Isamir 15a-5p Isamir 15a-5p Variable 1 breast concer Isamir 103a-3p Isamir 103a-3p	NOVA Test of Independence Factor Analysis Variable 2 hsa-mir-1246 hsa-mir-1246	1.09392 5.0029	dl Plot Correlation 0.805136 0.848436
Isa-mir 1343-3p Isa-mir 146a-3p Isa-mir 1-46a-3p Isa-mir 1-5a-5p Isa-mir 15a-5p Variable 1 breast cancer Isa-mir 103a-3p Isa-mir 107 Isa-mir 107	ANOVA Test of Independence Factor Analysis Variable 2 hsa-mir-1246 hsa-mir-1246 hsa-mir-1246	1.09392 5.0029 4.97258	Correlation 0.805136 0.848436 0.848213
Isa-mir-1233-3p Isa-mir-1463-5p Isa-mir-1463-5p Isa-mir-15a-5p Variable 1 Variable 1 breast cancer Isa-mir-103a-3p Isa-mir-107 Isa-mir-1233-5p	ANOVA Test of Independence Factor Analysis ANOVA Test of Independenc	1.09392 5.0029 4.97258 -1.09027	dl Plot Correlation 0.805136 0.848436 0.848213 -0.737964
Isa mir 1343-3p Isa mir 146a-3p Isa-mir 146a-3p Isa-mir 146a-3p Isa-mir 15a-5p Isa-mir 103a-3p Isa-mir 103a-3p Isa-mir 103a-3p Isa-mir 107 Isa-mir 1028-5p Isa-mir 1238-5p	ANOVA Test of Independence Factor Analysis Variable 2 hsa-mir-168-5p hsa-mir-15a-5p hsa-mir-1246 hsa-mir-1246 hsa-mir-1246	1.09392 5.0029 4.97258 -1.09027 2.34972	Image: constraints 0.805136 0.848436 0.848213 -0.737964 0.753489
Isa-mir 1343-3p Isa-mir 1-46a-3p Isa-mir 1-46a-3p Isa-mir 1-5a-5p Isa-mir 1-5a-5p Isa-mir 1-03a-3p Isa-mir 1-03a-3p Isa-mir 1-03a-3p Isa-mir 1-228-5p Isa-mir 1-233-5p Isa-mir 1-233-5p Isa-mir 1-233-5p	ANOVA Test of Independence Factor Analysis Variable 2 hsa-mir-15a-5p hsa-mir-15a-5p hsa-mir-1246 hsa-mir-1246 hsa-mir-1246 hsa-mir-1246 hsa-mir-1246 hsa-mir-1246 hsa-mir-1246	1,09392 5,0029 4,97258 -1.09027 2,34972 2,0316	Image: constraint of the second sec
hsa-mir-1343-3p hsa-mir-140-3p hsa-mir-146-5p hsa-mir-15a-5p Jnivariate Correlation Hypothesis Testing A	ANOVA Test of Independence Factor Analysis ANOVA Test of Independenc	1.09392 5.0029 4.97258 -1.09027 2.34972 2.34972 2.0316 8.51288	all Plot Correlation 0.805136 0.848436 0.848213 -0.737964 0.753489 0.67344 1
Isa-mir-1243-3p Isa-mir-1463-5p Isa-mir-1463-5p Isa-mir-1463-5p Isa-mir-1463-5p Isa-mir-103a-3p Isa-mir-103a-3p Isa-mir-1238-5p Isa-mir-1238-5p Isa-mir-1238-5p Isa-mir-1238-5p Isa-mir-1238-5p Isa-mir-1246 Isa-mir-1247-3p	ANOVA Test of Independence Factor Analysis NOVA Test of Independenc	1.09392 5.0029 4.97258 -1.09027 2.34972 2.0316 8.51288 1.55768	Image: control and plot 0.805136 0.848213 -0.737964 0.753489 0.67344 1 0.707487
Isa mir 1343-3p Isa mir 146a-3p Isa mir 1-46a-3p Isa mir 1-46a-3p Isa mir 1-5a-5p Isa mir 1-5a-5p Isa mir 103a-3p Isa mir 103a-3p Isa mir 103a-3p Isa mir 1238-5p Isa mir 1246 Isa mir 1246 Isa mir 1246-3p	ANOVA Test of Independence Factor Analysis ANOVA Test of Independence Factor Analysis Variable 2 Analysis Variable 2 Analysis Variable 2 Analysis Analysis Analysis Analysis Analysis Analysis Analysis Analysis Analysi	1,09392 5,0029 4,9728 -1,09027 2,34972 2,0316 8,51288 1,55768 3,78709	Image: Correlation 0.805136 0.848436 0.848436 0.7337864 0.6733449 0.673448 0.67347 1 0.707487 0.788297

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 *d* to visualize the result.

						>	arang[GS	E73002.ilm]			- 1	
🚪 Project	🔡 Dat	a 🛛	E Learner	Rex Expl	orer	Vodeler	🤻 Pre	dictor Deploy				
Variable 1						All		Variable 2:			All	
breast ca hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1 hsa-mir-1	03a-3p 07 228-5p 233-5p 238-5p 246 247-3p 26-3p 307-3p 30a-3p 343-3p							breast cancer hsa-mir-103a-3p hsa-mir-103a-3p hsa-mir-1228-5p hsa-mir-1228-5p hsa-mir-1228-5p hsa-mir-1228-5p hsa-mir-1247-3p hsa-mir-1247-3p hsa-mir-1247-3p hsa-mir-130a-3p hsa-mir-1343-3p hsa-mir-140-3p	_			
hsa-mir-1 hsa-mir-1	46a-5p 5a-5p	1.0					v	hsa-mir-146a-5p hsa-mir-15a-5p				~
hsa-mir-1 hsa-mir-1	46a-5p 5a-5p	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence	V Factor A	hsa-mir-15a-5p nalysis	6	breast cancer	(Complement)	*
hsa-mir-1 hsa-mir-1 Univariate	46a-5p 5a-5p Correlation	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence	V Factor A	hsa-mir-15a-5p nalysis hsa-mir-1246	6	breast cancer	[Complement]	
hsa-mir-1 hsa-mir-1 Jnivariate	46a-5p 5a-5p Correlation s Variable:	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence	Factor A	hsa-mir-15a-5p nalysis	6	breast cancer 1011 8.15116		216
hsa-mir-1 hsa-mir-1 Jnivariate	46a-5p 5a-5p Correlation s Variable:	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence		hsa-mir-15a-5p nalysis hsa-mir-1246 Count	6	1011	3.1	210 06
hsa-mir-1 hsa-mir-1 Jnivariate	46a-5p 5a-5p Correlation s Variable:	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence	Factor A	hsa-mir-15a-5p nalysis hsa-mir-1246 Count Mean	6	1011 8.15116 3.90171	3.1	216 065
hsa-mir-1 hsa-mir-1 Jnivariate	46a-5p 5a-5p Correlation s Variable: 246	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence		hsa-mir-15a-5p nalysis bsa-mir-1246 Count Mean Variance Z Value	6	1011 8.15116 3.90171 71.024	3.1	216 065
hsa-mir-1 hsa-mir-1 Univariate Continuou hsa-mir-1	46a-5p 5a-5p Correlation s Variable: 246 iable 1:	Hypothe	sis Testing	ANOVA 1	Fest of In	ndependence		hsa-mir-15a-5p nalysis bsa-mir-1246 Count Mean Variance	6	1011 8.15116 3.90171	3.1	216 065
hsa-mir-1 hsa-mir-1 Univariate Continuou hsa-mir-1 Binary Var	46a-5p 5a-5p Correlation s Variable: 246 iable 1: ncer	Hypothe	sis Testing	ANOVA 1	Fest of In	✓	Z Test	hsa-mir-15a-5p nalysis bsa-mir-1246 Count Mean Variance Z Value	δ	1011 8.15116 3.90171 71.024	3.1	216 065
hsa-mir-1 hsa-mir-1 Univariate Continuou hsa-mir-1 Binary Var breast car	46a-5p 5a-5p Correlation s Variable: 246 iable 1: ncer	Hypothe	sis Testing	ANOVA 1	Fest of In		Z Test T Test F Test	hsa-mir-15a-5p nalysis bsa-mir-1246 Count Mean Variance Z Value	5	1011 8.15116 3.90171 71.024	3.1	× 216 065
hsa-mir-1 hsa-mir-1 Jnivariate Continuou hsa-mir-1. Binary Var breast car	46a-5p 5a-5p Correlation s Variable: 246 iable 1: ncer	Hypothe	sis Testing	ANOVA 1	Fest of In	✓	Z Test T Test F Test	hsa-mir-15a-5p nalysis bsa-mir-1246 Count Mean Variance Z Value	6	1011 8.15116 3.90171 71.024	3.1	216 065

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.

5. EXPLORER IN XARANG

- 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 d to visualize the result.

Project 🔡 Data	🖽 Learner 🛛 😤 Explorer	Vodeler	🔍 Predicto	or 🛛 🥐 Deploy				
						-		
Variable 1:		All	Va	riable 2:				All
breast cancer			^ bi	reast cancer				
hsa-mir-103a-3p			h	sa-mir-103a-3p				
hsa-mir-107			h	sa-mir-107				
hsa-mir-1228-5p			hs	sa-mir-1228-5p				
hsa-mir-1233-5p			hs	sa-mir-1233-5p				
hsa-mir-1238-5p			h	sa-mir-1238-5p				
hsa-mir-1246				sa-mir-1246				
hsa-mir-1247-3p				sa-mir-1247-3p				
hsa-mir-126-3p				sa-mir-126-3p				
hsa-mir-1307-3p				sa-mir-1307-3p				
hsa-mir-130a-3p				sa-mir-130a-3p				
hsa-mir-1343-3p				sa-mir-1343-3p				
hsa-mir-140-3p				sa-mir-140-3p				
			h	sa-mir-146a-5p				
hsa-mir-146a-5p								
hsa-mir-15a-5p	esis Testing ANOVA Test of	Independence	✓ hs	sa-mir-15a-5p				
hsa-mir-15a-5p	esis Testing ANOVA Test of	Binary \	✓ hs	sa-mir-15a-5p				
hsa-mir-15a-5p nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246	NOVA	Binary \	Factor Analy: Variables:	sa-mir-15a-5p				
hsa-mir-15a-5p nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246	NOVA	Binary \	Variables: ast cancer	sa-mir-15a-5p	Ires	F Value		Probability
hsa-mir-15a-5p nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246 ANOVA O Stats	NOVA	Binary \ Øbrea	Variables: ast cancer	sa-mir-15a-5p	ires 17548.61	F Value	5856.69	Probability 0.00E+
hsa-mir-15a-5p nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246 ANOVA O Stats Source of Variace	NOVA [Binary \ Øbrea	Variables: ast cancer	sa-mir-15a-5p		F Value	5856.69	
hsa-mir-15a-5p [°] nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246 ANOVA Stats Source of Variace Between Groups	NOVA Sum of Squares 17548.61	Binary \ Øbrea	Variables: ast cancer reedom 1	Sa-mir-15a-5p	17548.61	F Value	5856.69	
hsa-mir-15a-5p nivariate Correlation Hypoth Continuous Variable: hsa-mir-1246 ANOVA O Stats Source of Variace Setween Groups Within Groups	NOVA [Sum of Squares 17548.61 9522.35	Binary \ Øbrea	v hs Factor Analys Variables: ast cancer reedom 1 3178	Sa-mir-15a-5p	17548.61	F Value	5856.69	

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.

- **D** X

٥	Xarang[GSE7	3002.ilm]				- 🗆 🗙
🐺 Project 🛛 🔠 Data 🛛 🎫 Learner 🛛 🜱 Explorer	💎 Modeler 🛛 💐 Predic	tor 🛛 🥐 Deplo	y			
Variable 1:	<u>All</u> V	ariable 2:				All
breast cancer hsamir-103a-3p hsamir-107 hsamir-128-5p hsamir-128-5p hsamir-128-5p hsamir-128-5p hsamir-1246 hsamir-1247-3p hsamir-1243-3p hsamir-130a-3p hsamir-143-3p hsamir-143-3p hsamir-146a-3p hsamir-15a-5p Univariate Correlation Hypothesis Testing ANOVA Test of I	×	reast cancer isa-mir-103a-3p isa-mir-107 isa-mir-1228-5p isa-mir-1238-5p isa-mir-1238-5p isa-mir-1247-3p isa-mir-1247-3p isa-mir-126-3p isa-mir-1307-3p isa-mir-1303-3p isa-mir-1303-3p isa-mir-140-3p isa-mir-140-5p isa-mir-15a-5p				~
Binary Variables in Rows:		Chi2	DF	Probability	Correlation	Entropy
breast cancer		3180.00	1	0.00E+00	1.00	0.0000
		Observed	breast cancer	Complement]		
		breast canc	1011	0		
Binary Variables in Columns:	Chi2	[Compleme	0	2169		
Binary Variables in Columns:	Chi2	[Compleme	0	2169		
	Chi2	[Compleme	0	2169		
	Chi2	[Compleme	0	2169		
	Chi2	[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".

			Xa	rang[GSE73002	2.ilm]				
🐺 Project 🛛 🔡 Dat	ta 🎫 Learner	Replore	er 💎 Modeler	Redictor	🥐 Deploy				
Variable 1:			All	Variab	le 2:				All
breast cancer hsarmin-103a-pp hsarmin-107 hsarmin-1238-5p hsarmin-1238-5p hsarmin-1238-5p hsarmin-1246 hsarmin-1246-3p hsarmin-1307-3p hsarmin-130a-3p hsarmin-1343-3p hsarmin-146a-5p hsarmin-146a-5p hsarmin-15a-5p				hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann hsann	t cancer nir-103a-3p nir-107 nir-1228-5p nir-1238-5p nir-1238-5p nir-1246 nir-1246 nir-126-3p nir-126-3p nir-126-3p nir-130a-3p nir-1343-3p nir-1343-3p nir-146a-5p nir-15a-5p				~
Univariate Correlation	Hypothesis Testing								
	Hypothesis resulty	ANOVA Tes	t of Independence	Factor Analysis					
	hypothesis resulty	ANOVA Tes	t of Independence Variable	Factor Analysis 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 ^
Extraction Method:	PrincipalComponen			Factor 1					
Extraction Method:			Variable	Factor 1 9.65 0.80	0.82 0.07	0.31 0.03	0.30 0.02	0.25 0.02	0.16 0.01
Extraction Method: Rotation Method:			Variable hsa-mir-103a-3p	Factor 1 9.65 0.80 0.93710332	0.82 0.07 0.25466849	0.31 0.03 -0.06526470	0.30 0.02 -0.01970297	0.25 0.02 -0.01309129	0.16 0.01 0.04064
	PrincipalComponen	ts v	Variable hsa-mir-103a-3p hsa-mir-107	Factor 1 9.65 0.80 0.93710332 0.93532493	0.82 0.07 0.25466849 0.25267532	0.31 0.03 -0.06526470 -0.05994179	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846	0.25 0.02 -0.01309129 0.00219564	0.16 0.01 0.04064 0.03049
	PrincipalComponen	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p	Factor 1 9.65 0.80 0.93710332 0.93532493 -0.84758667	0.82 0.07 0.25466849 0.25267532 0.31602695	0.31 0.03 -0.05526470 -0.05994179 0.22767903 -0.07266150 0.06566737	0.30 0.02 -0.01970297 -0.03026325 0.26815033	0.25 0.02 -0.01309129 0.00219564 0.14828440	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521
Rotation Method:	PrincipalComponen	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p hsa-mir-1233-5p	Factor 1 9.65 0.80 0.93710332 0.93532493 -0.84758667 0.87496760 0.82602537 0.88780471	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216	0.25 0.02 -0.01309129 0.00219564 0.03548619 -0.06937733 0.34378944	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443
Rotation Method:	PrincipalComponen	ts v	Variable 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	Factor 1 9.65 0.80 0.93710332 0.93532493 -0.84758667 0.87496760 0.82602537 0.88780471 0.85191361	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218 -0.19200869	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203 0.36641437	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216 -0.27393294	0.25 0.02 -0.01309129 0.00219564 0.14828440 0.03548619 -0.06937733 0.34378944 -0.01248447	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443 0.12731
Rotation Method:	PrincipalComponen	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p hsa-mir-1238-5p hsa-mir-1238-5p hsa-mir-1238-5p hsa-mir-1246 hsa-mir-1247-3p hsa-mir-126-3p	Factor 1 9.65 0.80 0.93710332 0.93532493 -0.84758667 0.87496760 0.82602537 0.88780471 0.85191361 0.92089004	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218 -0.19200869 0.28702517	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203 0.36641437 0.02909705	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216 -0.27393294 0.04856661	0.25 0.02 -0.01309129 0.00219564 0.14828440 0.03548619 -0.06937733 0.34378944 -0.01248447 -0.13112700	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443 0.12731 0.00751
Rotation Method:	PrincipalComponen None	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p hsa-mir-1238-5p hsa-mir-1246 hsa-mir-1247-3p hsa-mir-126-3p hsa-mir-1307-3p	Factor 1 9.65 0.80 0.93710332 0.93532493 -0.84758667 0.82602537 0.88780471 0.85181361 0.92089004 0.91573785	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218 -0.19200869 0.28702517 -0.25439354	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203 0.36641437 0.02909705 -0.06462981	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216 -0.27393294 0.04856661 -0.05656080	0.25 0.02 -0.01309129 0.00219564 0.014828440 0.03548619 -0.06937733 0.34378944 -0.01248447 -0.13112700 0.04067981	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443 0.12731 0.00751 -0.16136
Rotation Method:	PrincipalComponen None	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p hsa-mir-1238-5p hsa-mir-1238-5p hsa-mir-1247-3p hsa-mir-1247-3p hsa-mir-126-3p hsa-mir-1307-3p	Factor 1 9.65 0.80 0.93710332 0.93532493 0.837496760 0.82602537 0.88780471 0.82191361 0.92089004 0.92089004 0.91573785 0.93823038	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218 -0.19200869 0.28702517 -0.25439354 0.22699525	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203 0.36641437 0.02909705 -0.06462981 0.00427244	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216 -0.27393294 0.04856661 -0.05656080 0.03443057	0.25 0.02 -0.01309129 0.00219564 0.14828440 0.03548619 -0.06937733 0.34378944 -0.01248447 -0.13112700 0.04067981 -0.12930297	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443 0.12731 0.00751 -0.16136 -0.02591
Rotation Method:	PrincipalComponen None	ts v	Variable hsa-mir-103a-3p hsa-mir-107 hsa-mir-1228-5p hsa-mir-1238-5p hsa-mir-1246 hsa-mir-1247-3p hsa-mir-126-3p hsa-mir-1307-3p	Factor 1 9.65 0.80 0.93710332 0.93532493 0.837496760 0.82602537 0.88780471 0.82191361 0.92089004 0.92089004 0.91573785 0.93823038	0.82 0.07 0.25466849 0.25267532 0.31602695 -0.30678830 -0.44288548 0.06100218 -0.19200869 0.28702517 -0.25439354	0.31 0.03 -0.06526470 -0.05994179 0.22767903 -0.07266150 0.06566737 -0.21074203 0.36641437 0.02909705 -0.06462981	0.30 0.02 -0.01970297 -0.03026325 0.26815033 0.27660846 0.21601137 -0.08826216 -0.27393294 0.04856661 -0.05656080	0.25 0.02 -0.01309129 0.00219564 0.014828440 0.03548619 -0.06937733 0.34378944 -0.01248447 -0.13112700 0.04067981	0.16 0.01 0.04064 0.03049 0.10998 -0.02427 0.15521 0.16443 0.12731 0.00751 -0.16136

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

)		Xaı	ang[GSE73002	2.ilm]			-	
🌆 Project 🛛 📳 Data 🛛 🗄	🗄 Learner 🛛 🜱 Explore	💎 Modeler	Redictor	🥐 Deploy				
Classification Target:	🗌 OnTheFly	Iteration	: 5	Delta: 0.001	Show	🔍 Reducer	🔀 Model 📘	Script
		~						
input: []	-							
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 hsa-mir-211-3p hsa-mir-22-3p								
nsa-mir-22-3p hsa-mir-221-3p hsa-mir-222-3p		~						
Classification Regression								
odeler				1			GSE73002.ilm	110 8 10

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.

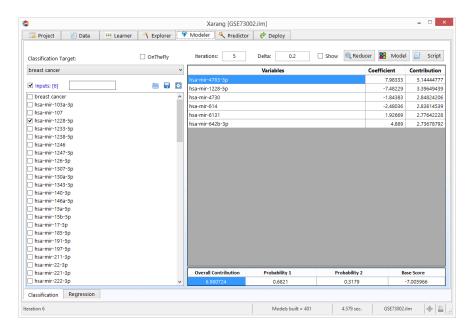
1. Select the Classification tab on the bottom left of the window.

6. MODELER IN XARANG

- 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 \equiv to open the selected variable list.
- 6. Click Model button to build the model.

\$				Xa	rang [GSE7300)2.ilm]					-	×	
📴 Project 🛛 🔠 Data	😐 Learner	🕆 Explorer	V	Modeler	Redictor	🧼 Deploy	(
Classification Target:		OnTheFly		Iteration	ıs: 5	Delta:	0.2	Show	🔍 Reducer	🔀 м	odel 📘	Script	
breast cancer			~										
✓ Inputs: [100]		🖴 🖬	*										
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													
✓ Isa-mir-130a-3p ✓ hsa-mir-1343-3p													
✓ hsa-mir-1343-3p													
✓ hsa-mir-140-sp ✓ 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			~										
Classification Regression													
Modeler							100%			GSE73	002.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.



- 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).

Zarang - Model Script	-		×
🗟 Save Model			
Ivaring_bcbreast cancer; 7005966.05 Ivas-mir(-78)=7,98330.6231747.N Ivas-mir(-78)=7,98330.6231747.N Ivas-mir(-730,-7.48330.8430.8137.N Ivas-mir(-730,-184330.8439.8137.N Ivas-mir(-740,-184330.8137.N Ivas-mir(-740,-184330.8137.N Ivas-mir(-740,-184330.8137.N Ivas-mir(-614,-182609.06380.08,N) Ivas-mir(-614,-182609.06380.08,N) Ivas-mir(-614,-182609.06380.08,N) Ivas-mir(-614,-184000.890017,N)			^
SQL Function CREATE FUNCTION dboxarang_bc { {sa_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 7.983307@bhss_mir.4783.3p -7.4822907@bhss_mir.1228_5p -1.8438070@bhss_mir.4730 -2.480507@bhss_mir.614 +1326507@bhss_mir.6131 +4890007@htss_mir.6131 +4890007@htss_mir.614b_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 F	ROM TABLE		
			Ŷ

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. 3. Select the input variables from the Inputs list.
- 4. Click 🖬 to save the selected variables.
- 5. Click \equiv to open the selected variable list.
- 6. Click Model button to build the model.

🐺 Project 🛛 🔠 Data 🛛 🎫 Learner	🕂 Explorer	💎 Modeler 🛛 🦄 Predict	or 🛛 🥐 Deploy			
tegression Target:	OnTheFly	Iterations:	Delta: 0.0	1 Show	🔍 Reducer 🛛 🔀 Mo	del 📔 Scrip
hsa-mir-103a-3p	~		hsa-mir-103a-3p		Coefficient	Correlation
		Intercept			0.853035	
Inputs: [99]		hsa-mir-107			0.286559	0.9541
breast cancer	^	hsa-mir-15a-5p			0.167128	0.9313
hsa-mir-103a-3p		hsa-mir-130a-3p			0.0614609	0.9235
Ansa-mir-107		hsa-mir-126-3p			0.113962	0.923
Asa-mir-1228-5p		hsa-mir-4317			0.120592	0.9181
Asa-mir-1233-5p		hsa-mir-25-3p			0.0837937	0.9152
Asa-mir-1238-5p		hsa-mir-191-5p			0.0247072	0.911
Asa-mir-1246		hsa-mir-15b-5p			0.0781068	0.9022
Asa-mir-1247-3p		hsa-mir-140-3p			-0.000675287	0.8995
✔ hsa-mir-126-3p ✔ hsa-mir-1307-3p		hsa-mir-4514			0.104445	0.8986
v nsa-mir-1307-3p V hsa-mir-130a-3p		hsa-mir-221-3p			0.0216432	0.8952
A hsa-mir-1343-3p		hsa-mir-23a-3p			0.0553365	0.8932
A hsa-mir-140-3p					0.0883064	
A hsa-mir-146a-5p		hsa-mir-23b-3p				0.8864
hsa-mir-15a-5p		hsa-mir-146a-5p			0.0080349	0.8853
hsa-mir-15b-5p		hsa-mir-6073			0.0906108	0.8845
A hsa-mir-17-3p		hsa-mir-17-3p			0.0287429	0.8844
Ansa-mir-185-5p		hsa-mir-22-3p			0.0287011	0.8836
Ansa-mir-191-5p		hsa-mir-222-3p			-0.0281572	0.8832
Ansa-mir-197-5p		hsa-mir-3617-5p			0.0238864	0.8773
A hsa-mir-211-3p		hsa-mir-4652-5p			-0.0295973	0.8683
✔ hsa-mir-22-3p		hsa-mir-320e			-0.0376097	0.8648
✓ hsa-mir-221-3p		PC1	PC2	PC3	PC4	PC5
✓ hsa-mir-222-3p	~	79.09%	3.50%	3,29%	1.07%	0.88%

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.

🐺 Project 🛛 🔡 Data	🚟 Learner	🕆 Explorer		Modeler 🤻 Predictor 🥏 Depl	ру			
Regression Target:		OnTheFly		Iterations: Delta:	0.01 Show	🔍 Reducer	Model	Script
hsa-mir-103a-3p			~	hsa-mir-10	3a-3p	Coefficie	nt (Correlation
		~ •		Intercept		-0.38	4258	
Inputs: [12]		-	*	hsa-mir-107		0.86	6163	0.954
breast cancer			^	hsa-mir-4734		0.021	5231	0.709
hsa-mir-103a-3p				hsa-mir-642b-3p		0.071	9442	0.674
Isa-mir-107				hsa-mir-1238-5p		0.056	5726	0.668
hsa-mir-1228-5p				hsa-mir-197-5p		0.022	8075	0.656
hsa-mir-1233-5p				hsa-mir-3151-5p		0.029	5397	0.653
✔ hsa-mir-1238-5p				hsa-mir-885-3p		0.022	6028	0.640
hsa-mir-1246				hsa-mir-4690-5p		-0.056		0.59
hsa-mir-1247-3p				hsa-mir-6742-5p			6705	0.589
hsa-mir-126-3p hsa-mir-1307-3p				hsa-mir-6729-5p			5252	-0.512
hsa-mir-1307-3p				hsa-mir-6816-5p			7567	-0.573
hsa-mir-1343-3p				hsa-mir-1228-5p			5106	-0.373
hsa-mir-140-3p				nsa-mir-1228-5p		-0.29	5106	-0.731
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				PC1 PC2	PC3	PC4		PC5
hsa-mir-222-3p			~	72.69% 10.28%	3.95%	3.28%		2.45%
Classification Regression								

- 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 - Model Script – 🗖	×
🔜 Save Model		
xarang_rr,hsa-mir-103a-3p,1,1		
intercept,-0.384258,0,N		
hsa-mir-107,0.866163,3.20420		
hsa-mir-4734,0.021523,12.509		
hsa-mir-642b-3p,0.071944,8.90		
hsa-mir-1238-5p,0.056573,5.86		
hsa-mir-197-5p,0.022808,7.491		
hsa-mir-3151-5p,0.029540,4.99		
hsa-mir-885-3p,0.022603,4.995		
hsa-mir-4690-5p,-0.056644,5.9		
hsa-mir-6742-5p,0.021671,4.90		
hsa-mir-6729-5p,0.205252,12.5		
hsa-mir-6816-5p,0.014757,10.5		
hsa-mir-1228-5p,-0.295106,11.	733655,N	
SQL Function		
CREATE FUNCTION dbo.xarang_r	r	
(
	37 FLOAT, @hsa_mir_4734 FLOAT, @hsa_mir_642b_3p FLOAT, @hsa_mir_1238_5p FLOAT, @hsa_mir_197_5p FLOAT,	
ELOAT, @hsa mir 1228 5p FLOAT	a_mir_885_3p FLOAT, @hsa_mir_4690_5p FLOAT, @hsa_mir_6742_5p FLOAT, @hsa_mir_6729_5p FLOAT, @hsa_mir_6816_5 	p
FLUAI, @nsa_mir_1228_5p FLU	41	
1		
RETURNS FLOAT		
AS		
BEGIN		
RETURN		
-0.384258*@Intercep		
+0.866163*@hsa_mi		
+0.021523*@hsa mi		
+0.071944*@hsa_mi		

Predictor in Xarang

7

 Interference of a field line of a field of a f

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.

7. PREDICTOR IN XARANG

	roject 🛛 🔢 E					C Depio				
ata	set: C:\GEO\bio	markers.ai\micro	RNA\GSE73002\	original\breast-ca	ncer\GSE73002_br	reast_cancer_test.	CSV			
	geo_accession	hsa-miR-1307-3	hsa-miR-4783-5	3p hsa-miR-8073	hsa-miR-4532	-	hsa-miR-6861-5p		hsa-miR-4675	hsa-miR-92a-2-5
	1876423	9.227366	8.695286	7.007205	13.76236	8.732082	8.332469		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	7.361416
	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	0.812455	0 /37155	7 782043	14 26802	0 500872	8 038166	12 44267	8 571600	7 632700
C	ata Mo	odel Pred	dict Eval	uate						

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.

			Xarang [G	SE73002.ilm]		
Project 🔡 Da	ta 🎫 Learner	🕆 Explorer	💎 Modeler 🛛 💐 Pr	edictor Deploy		
М	odel: Classification	n v				
Inputs: [6]			≡ ∎	Target:		
breast cancer			~	breast cancer		
hsa-mir-103a-	p					
hsa-mir-107						
✓ hsa-mir-1228-						
hsa-mir-1233-						
hsa-mir-1238-5	^p p					
hsa-mir-1240				🗌 Key: []		
hsa-mir-1247-				breast cancer		^
hsa-mir-1307-				geo_accession		<u>^</u>
hsa-mir-130a-3				hsa-mir-103a-3p		
hsa-mir-1343-				hsa-mir-107		
hsa-mir-140-3				hsa-mir-1228-5p		
hsa-mir-146a-				hsa-mir-1233-5p		
hsa-mir-15a-5				hsa-mir-1238-5p		
hsa-mir-15b-5	0			hsa-mir-1246		
hsa-mir-17-3p				hsa-mir-1247-3p		
hsa-mir-185-5	,			hsa-mir-126-3p		
hsa-mir-191-5	, ,			hsa-mir-1307-3p		
hsa-mir-197-5	,			hsa-mir-130a-3p		
hsa-mir-211-3)		~	hsa-mir-1343-3p		*
Data Mod	el Predict	Evaluate				
st cancer				Models built = 1002	10.58 sec.	GSE73002.ilm

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.

Project Image: Statistic intervence Modeler Predictor Deploy nish Image: Statistic intervence Image: Statistintervence	📮 Project 🛛 🔡 Data)	Xarang [GSE73002.	lm]	-	
Overall Performance contresion 244 0 26 0 784 specificity 100.5 100.6 True Positive Rate, Power, Recall 0001 True Negative Rate 1024 Specificity 1025 Precision 1026 True Negative Rate 1026 Precision 1027 Specificity 1028 Precision 1029 Specificity 100.05 Precision		🚥 Learner 🛛 🜱 Expl	lorer 🛛 💎 Modele	r 🤻 Predictor	E Deploy		
Overall Performance contresion 244 0 26 0 784 specificity 100.5 100.6 True Positive Rate, Power, Recall 0001 True Negative Rate 1024 Specificity 1025 Precision 1026 True Negative Rate 1026 Precision 1027 Specificity 1028 Precision 1029 Specificity 100.05 Precision	inish						
I 244 0 244 0 24 56 540 Actual 268 516 784 Overall Performance ccuracy 0.9944 ount 784 Specificity 91.05 True Positive Fate, Power, Recall ount 784 Performance Specificity Out 784 Specificity 100.05 Pesitive Predictive Value 90.05 Arget breast cancer Fate Negative Rate 0.05 Scores File: CrCECD.biomark.ers.allymicroRNAA(GSET2002.toringenabl/breast-cancer.cet	111316						
I 244 0 244 0 24 56 540 Actual 268 516 784 Overall Performance ccuracy 0.9944 ount 784 Specificity 91.05 True Positive Fate, Power, Recall ount 784 Performance Specificity Out 784 Specificity 100.05 Pesitive Predictive Value 90.05 Arget breast cancer Fate Negative Rate 0.05 Scores File: CrCECD.biomark.ers.allymicroRNAA(GSET2002.toringenabl/breast-cancer.cet							
0 24 516 Actual 268 516 Overall Performance rccuracy 0.964 ount 784 Securacy 0.964 lodel LDA Predictive Predictive Value 51.05 rarget breast cancer Hegative Predictive Value 55.65 False Negative Rate 0.005 False Negative Rate 0.05 Scores File: CxCEChbiomarkers allymicroRHMA(GSET2002)toring/mhb/breast-cancer/GSET2002)_breast-cancer (Lest.cvy.scores.cv)	Confusion Matrix	1	0	Predicted			
Actual 268 516 784 Overall Performance Statistics Value Same As coursey 0.9594 Securitary 91.05 True Positive Rate, Power, Recall ount 784 Godel LDA Perfective Value 100.05 True Negative Rate 0.005 False Predictive Value 95.65 False Predictive Value 95.65 False Positive Rate 0.005 False Negative Rate 0.05 False Negative Rate 9.05 Matthew's Correlation Coefficient 0.93 UkeHood Ratio Positive NAM Scores File: Cr(ECO.biomarkers.ab/microRMA/GSET2002.breast_cancer/GSET2002.breast_cancers.csv)	1	244	0	244			
Overall Performance ccuracy 0.9954 sount 784 specificity 91.05 True Positive Rate, Power, Recall totel LDA Positive Predictive Value 95.65 False Negative Rate 0.05 Matthew's Correlation Coefficient 0.93 LikeHood Ratio Positive NAH Scores File: CrCECholomarkers.allymicroRNA(SSET2002).torigenal.blareast.cancer_CSET2002.breast.cancers.csv					_		
Current Performance Statistics Value Same As current 0.9564 Sentifiely 91.05 True Positive Rate ount 774 Specificity 100.05 True Positive Rate lodel LDA Perestancer Precision False Precisive Rate 0.05 rget breast cancer False Precisive Rate 0.05 Precision False Positive Rate 0.05 False Positive Rate 0.05 False Positive Rate 0.05 Matthew's Correlation Coefficient 0.93 Scores File: CxCEChbiomarkers.ab,microRNAA(SSET2002).toriginahbreast-cancer/GSE72002.breast_cancer/csEr2/so2 Weight Precision	Actual	268	516	784	_		
Curreary Outroit Statistics Value Same As unit 0.954 Secretifyity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate get breast cancer Predictive Value 95.65 False Positive Rate False Negative Rate 0.05 Precision False Positive Rate 0.05 False Negative Rate 0.05 Matthew's Correlation Coefficient 0.93 Correlation Scores File: ExisteDiomanitiers.ahmicroRNAK.055272002.breast_cancer.(dSE72002.breast_cancer.et.cs.cs.cs.cs.cs.cs.cs.cs.cs.cs.cs.cs.cs.							
Currary Output Statistics Value Same As currary 0.0964 Sentificity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate odel LDA Predictive Value 100.05 Precision Faste Negative Rate 0.06 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Scores File: CrCGD/biomarkers.ahmicroRNA/GSEF2002/original/bireast-cancer/GSEF2002/bireast_cancer/GSEF2002/bi							
Currary Output Statistics Value Same As currary 0.0964 Sentificity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate odel LDA Predictive Value 100.05 Precision Faste Negative Rate 0.06 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Scores File: CrCGD/biomarkers.ahmicroRNA/GSEF2002/original/bireast-cancer/GSEF2002/bireast_cancer/GSEF2002/bi							
Currary Output Statistics Value Same As currary 0.0964 Sentificity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate odel LDA Predictive Value 100.05 Precision Faste Negative Rate 0.06 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Scores File: CrCGD/biomarkers.ahmicroRNA/GSEF2002/original/bireast-cancer/GSEF2002/bireast_cancer/GSEF2002/bi							
Currey Output Statistics Value Same As currey 0.0934 Semithivity 91.05 True Positive Rate unit 784 Specificity 100.05 True Negative Rate odel LDA Predictive Value 95.65 False Pedictive Value 95.65 False Pedictive Rate 0.05 Freedoctive Rate 0.05 Freedoctive Value Matthew's Correlation Coefficient 0.93 Libelhood Rato Positive NAM Corres File: CAGEOL biomanities: submicroRNAX/cSGF28002, breast_cancer_(cSE728002, breast_cancer_rest.cvy:scores.cs) Emittive Scores.cs							
Currery Outroit Performance Statistics Value Same As sentificity 91.05 True Positive Rate 91.05 True Positive Rate unit 764 Specificity 100.05 True Negative Rate odel LDA Persisticancer Negative Predictive Value 95.65 False Portive Rate 0.05 Frection False Portive Rate Value 93.05 Matthew's Correlation Coefficient 0.93 Corres File: Existion (Cost27002, breast, cancer, (cs5273002, breast,							
Current Performance Statistics Value Same As Semithivity 91.05 True Positive Rate Incomparing and the predictive Value 100.05 True Negative Rate odel LDA Breast cancer Breast cancer 100.05 Precision Face Negative Rate watherwise Deress cancer Correlation Coefficient 0.03 Face Negative Rate 0.05 Face Negative Rate 0.05 True Dostive Rate 0.05 Face Negative Rate 0.05 Korres File: Ex(GEC) biomarkers.at/microRNAX/GS672002.breast_cancer.(GSF23002.breast_cancer.(GSF23002.breast_cancer.(GSF23002.breast_cancer.cs.cs) Face Face							
Current Performance Statistics Value Same As Semithivity 91.05 True Positive Rate Incomparing and the predictive Value 100.05 True Negative Rate odel LDA Breast cancer Breast cancer 100.05 Precision Face Negative Rate watherwise Deress cancer Correlation Coefficient 0.03 Face Negative Rate 0.05 Face Negative Rate 0.05 True Dostive Rate 0.05 Face Negative Rate 0.05 Korres File: Ex(GEC) biomarkers.at/microRNAX/GS672002.breast_cancer.(GSF23002.breast_cancer.(GSF23002.breast_cancer.(GSF23002.breast_cancer.cs.cs) Face Face							
Currary Output Statistics Value Same As currary 0.0964 Sentificity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate odel LDA Predictive Value 100.05 Precision Faste Negative Rate 0.06 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Scores File: CrCGD/biomarkers.ahmicroRNA/GSEF2002/original/bireast-cancer/GSEF2002/bireast_cancer/GSEF2002/bi							
Currary Output Statistics Value Same As currary 0.0964 Sentificity 91.05 True Positive Rate odel LDA Specificity 100.05 True Negative Rate odel LDA Predictive Value 100.05 Precision Faste Negative Rate 0.06 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Faste Negative Rate 0.05 Scores File: CrCGD/blormafkers.ahmicroRNA/GSEF2002/original/bireast-cancer/GSEF2002/breast_cancer/GS							
Current Performance Statistics Value Same #s							
Current Performance Statistics Value Same #s							
Current Performance Statistics Value Same #s							
ount 784 Specificity 100.05 True Negative Rate lodel LDA Positive Predictive Value 100.05 Precision arget breast cancer Negative Rate 9.05 False Policive Rate 0.05 Precision Justice Rate 9.05 Policive Rate Scores File: CGEO/biomarkers.al/microRNA/GSE72002/breast.cancer/GSE73002/breast.cancer/Est.cvs.cores.cov Policive Rate			1				
LDA Positive Predictive Value 100.05 Precision rget breast cancer Positive Predictive Value 95.65 False Protice Rate 0.05 False Protice Rate 0.05 False Protice Rate 9.05 Scores File: CGECholomarkars.abmicroRNA(SEST2000_toriginal breast-cancer/GSE73002_breast_cancer_test.csv NAH	Overall			atistics			^
Image: Second File Dreast cancer Megative Predictive Value 95,65 False Popular Rate 0,05 False Popular Rate 0,05 False Popular Rate 9,05 Matthew's Correlation Coefficient 0,93 Likelhood Ratio Positive NAH NAH Scores File: CGE0\blommarkers.ahmicroRHAA(SSE72002).toraint/breast-cancer_lest.csyscores.csy False	couracy	0.9694	Sensitivity	atistics	91.0	True Positive Rate, Power, Recall	
False Politive Rate 0.05 False Negative Rate 9.05 Matthew's Correlation Coefficient 0.93 Likelihood Ratio Positive NAN Scores File: C\GEO\biomarkers.alymicroRNA\GSE73002\original.breast-cancer\GSE73002_breast_cancer_test_cov_scores.cs/	ccuracy ount	0.9694 784	Sensitivity Specificity		91.0	True Positive Rate, Power, Recall	
False Negative Rate 9.0% Matthew's Correlation Coefficient 0.93 Likelihood Ratio Positive NAN Scores File: CGEC\biomarkers.abmicroRNA(SSE72002\original_breast-cancer\GSE73002\breast_cancer\test.csy	ccuracy punt Iodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive	Value	91.0 100.0 100.0	True Positive Rate, Power, Recall	
Matthew's Correlation Coefficient 0.93 Likethood Rato Positive NAN Scores File: EXGEC/biomarkersahmicroRNAKGSET3002toriginal/breast-cancert/GSET3002breast-cancert_est.csyscores.csy	ccuracy ount lodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive Negative Predictive	Value	91.07 100.07 100.07 95.65	True Positive Rate, Power, Recall True Negative Rate Precision	
Likelihood Ratio Positive NAN Scores File: [C_GEO(biomarkers.al)microRNA(GSE73002).original)breast-cancer/GSE73002,breast_cancer_test_csy.scores.csy	ccuracy punt Iodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate	Value e Value	91.07 100.07 100.07 95.65	True Positive Rate, Power, Recall True Negative Rate Precision	
Scores File: C\GEO\blomarkers.alymicroRNA\GSE73002\original.breast-cancer\GSE73002_breast_cancer_text_csv.scores.csv	ccuracy punt Iodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate False Negative Rat	Value e Value e	91.05 100.07 100.05 95.65 0.05 9.05	True Positive Rate, Power, Recall True Negative Rate Precision	
	ccuracy punt Iodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate False Negative Rat Matthew's Correla	Value e Value e ttion Coefficient	91.05 100.05 100.05 95.65 0.05 9.05 0.95	True Positive Rate, Power, Recall True Negative Rate Precision	
	ccuracy ount lodel	0.9694 784 LDA	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate False Negative Rat Matthew's Correla	Value e Value e ttion Coefficient	91.05 100.05 100.05 95.65 0.05 9.05 0.95	True Positive Rate, Power, Recall True Negative Rate Precision	
Data Model Predict Evaluate	couracy ount lodel arget	0.9694 784 LDA breast cancer	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate False Negative Rat Matthew's Correla Likelihood Ratio Po	Value e Value e tion Coefficient positive	91.0 100.0 100.0 95.6 0.0 9,0 0.9 0.9	True Positive Rate, Power, Recall True Negative Rate Precision	
	couracy ount lodel arget	0.9694 784 LDA breast cancer	Sensitivity Specificity Positive Predictive Negative Predictive False Poitive Rate False Negative Rat Matthew's Correla Likelihood Ratio Po	Value e Value e tion Coefficient positive	91.0 100.0 100.0 95.6 0.0 9,0 0.9 0.9	True Positive Rate, Power, Recall True Negative Rate Precision	
z_breast cancer 0.252 sec. GSE73002.iim 🚸 🔚	ccuracy odel rrget Scores File: C(CEO)bioma	0.9694 784 LDA breast cancer kers_alymicroRNA\GSE7300	Sensitivity Specificity Positive Predictive Negative Predictive False Potive Rate False Negative Rat Matthew's Correla Likelihood Ratio Pr 22/original/breast-co	Value e Value e tion Coefficient positive	91.0 100.0 100.0 95.6 0.0 9,0 0.9 0.9	True Positive Rate, Power, Recall True Negative Rate Precision	

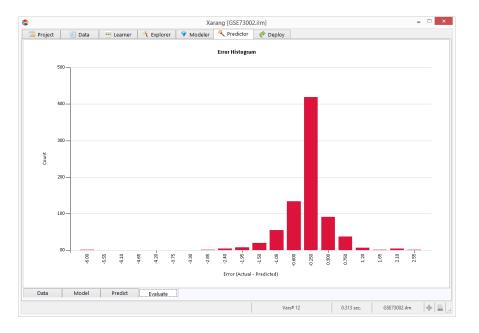
7.4 Evaluate Prediction Model

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

				-	SE73002.ilm]			
Project	🔡 Data	🖽 Learner	🌱 Explorer 🛛 🛛	Modeler 🕺 Pr	edictor 🧼 Deploy			
		Score:				Buckets:		
		z_breast c	ancer		~	10 👻 🕅		
		Bucket	Avg. Score	Target#	Target%			
		1	0.01517	1		0.37%		
		2	0.02098	0		0.00%		
		3	0.02470	0		0.00%		
		4	0.02889	1		0.37%		
		5	0.03499	1		0.37%		
		6	0.04508	4		1.49%		
		7	0.17490	25 📻	•	9.33%		
		8	0.95122	79 📻		29.48%		
		9	0.97915	78		29.10%		
		10	0.98977	79 📻		29.48%		
		🗌 Externa	l Score File					
Data	Model	Predict	Evaluate					
Data	Model	Predict	Evaluate		z, breast car	rcer 0.252 s	sc. GSE73002.ilm	*

This is Error Histogram for a regression model.

7. PREDICTOR IN XARANG



8

Deploying in Xarang

Models can be deployed to a remote server using FTP.

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

							Remote		
FTP									
ETO	Server:	ftp.smartarray.r	et						
Use	r Name:	0134490 smarta	irrayn						
Pas	sword:	*******							
Ren	note Folder:	/ModelsFolder/							
				_					
				🚨 Co	innect				
Mode									
				2	6				
							Rer	iame 🛛 🗱	Delete

2. Click \equiv to select a save model file.

\$					Open						×	1
🔄 🏵 – 🕇]	« breast-o	cano	er	▹ Export		~ ¢		Search Export			ρ,	
Organize 🔻 Ne	w folder								8== •	•		
_		^		Name		^			Date	modified	ł	
🌉 This PC			[🔠 GSE73	002.xse				2018-	02-18 6:	51 PM	
Documents												
鷆 Downloads												
🔰 Music												
╞ Pictures												
Videos												
📥 Local Disk (C:)											
词 Libraries												
Documents		п										
🁌 Music		~	<									>
	File <u>n</u> ame:	GS	E73	002.xse			¥	Xarang (*.xse)	1		~	
								<u>O</u> pen		Cano		

8. DEPLOYING IN XARANG

3. Click otinto deploy the model to the remote server.

Projet Data III Learner Poporer Modeler Predictor Deploy Remote FIP User Name: 0134490[smartarray.net User Name: 0134490[smartarray.net Password:	\$				Xa	rang [GSE7300	2.ilm]				-	□ ×
FTP FTP Server: thp.smartarray.net User Name: O134490[smartarrayn Password: @ Connet Model Cr\GEO\biomarkers.a\microRNA\GSE73002\originah\breast-ca @ Rename @ Rename	🕛 Project	🔡 Data	🚥 Learner	🜱 Explorer	V Modeler	Redictor	🥐 Deploy]				
FIP Server: tp.smartarray.net User Name: 0134490[smartarrayn Password: ******** Remote Folder: (ModelsFolder/ Image: Ima									Remote			
User Name: 0134490jsmartarrayn Password: Remote Folder: //ModelsFolder/ C<\GEO\biomarkers.ahmicroRNA\GSE73002\originah\breast-ca	FTP						GSE73002	.xse				
Password: ******* Remote Folder: /ModelsFolder/ C <geo\biomarkers.ahmicrorna\gse73002\originah\breast-ca< td=""><td>FT</td><td>Server:</td><td>ftp.smartarray.r</td><td>net</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></geo\biomarkers.ahmicrorna\gse73002\originah\breast-ca<>	FT	Server:	ftp.smartarray.r	net								
Remote Folder: /ModelsFolder/ C<\GEO\blomarkers.ahmicroRNA\GSE73002\originah\breast-ca	Us	er Name:	0134490 smarta	arrayn								
Model C <geo\biomarkers.ahmicrorna\gse73002\originahbreast.ca< td=""><td>Pa</td><td>ssword:</td><td>*******</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></geo\biomarkers.ahmicrorna\gse73002\originahbreast.ca<>	Pa	ssword:	*******									
Model Ct\GEO\biomarkers.al\microRNA\GSE73002\original\breast-ca	Re	mote Folder:	/ModelsFolder,	<i>(</i>								
C\GEO\blomarkers.ahmicroRNA\GSE73002\originahbreast-ca					🚨 Co	nnect						
C\GEO\blomarkers.ahmicroRNA\GSE73002\originahbreast-ca												
Rename Selete	Mod	el										
Rename Selete	C:	\GEO\biomarke	rs.ai\microRNA\G	SE73002\original	\breast-ca	ê						
1 200 sec (557300) 3m & 10									🕓 Re	name 🗱	Delete	
	Deploy								0.389 sec.	GSE73002	.ilm 🔶	

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

8.1 Scorecard

- 1. Click "Models" to refresh the list.
- 2. Select a model from the list and click.
- 3. Fill the scorecard or check "Avg" to fill the scorecard with the related average values.
- 4. Click \blacksquare to predict using the model.
- 5. Click \triangleleft or \triangleright to browse the current session queries.

🖲 User 🔍 Public					Scorecard
Models: Credit_LM_1	•	2 🕨	🗹 Avg	credit_default 😑 0.00333	6
bni_score_on_od					
878.739154					
br_trade_past_due					
0.045507					
busage					
110.530049					
daysdelq					
0.424629					
maxdelq					
0.149688					
maxtimeslbal12					
1.033758					
totdaysdelq12					
0.039303					
totdelq12					
0.690512					

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 🤯.

Group	Model		Traffic%	Count	Accuracy	Probabil
Α	Credit_LM_1	•	50	1,774	87.66%	
В	Credit_NLM_1	•	50	1,743	93.34%	0.0000
С		۲]		
D		٠]		
E		۲]		
			a			
<u>Datasets:</u>	File - Credit test	•	٢			
	Time Elapsed = 0.4 sec.					