

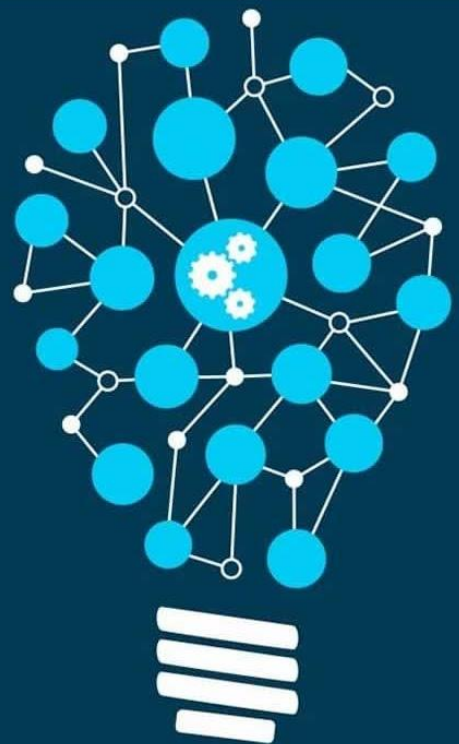
LonGrea

No Code

Community V1.0.0

ML.NET

MACHINE
LEARNING



 **DevExpress**[®]

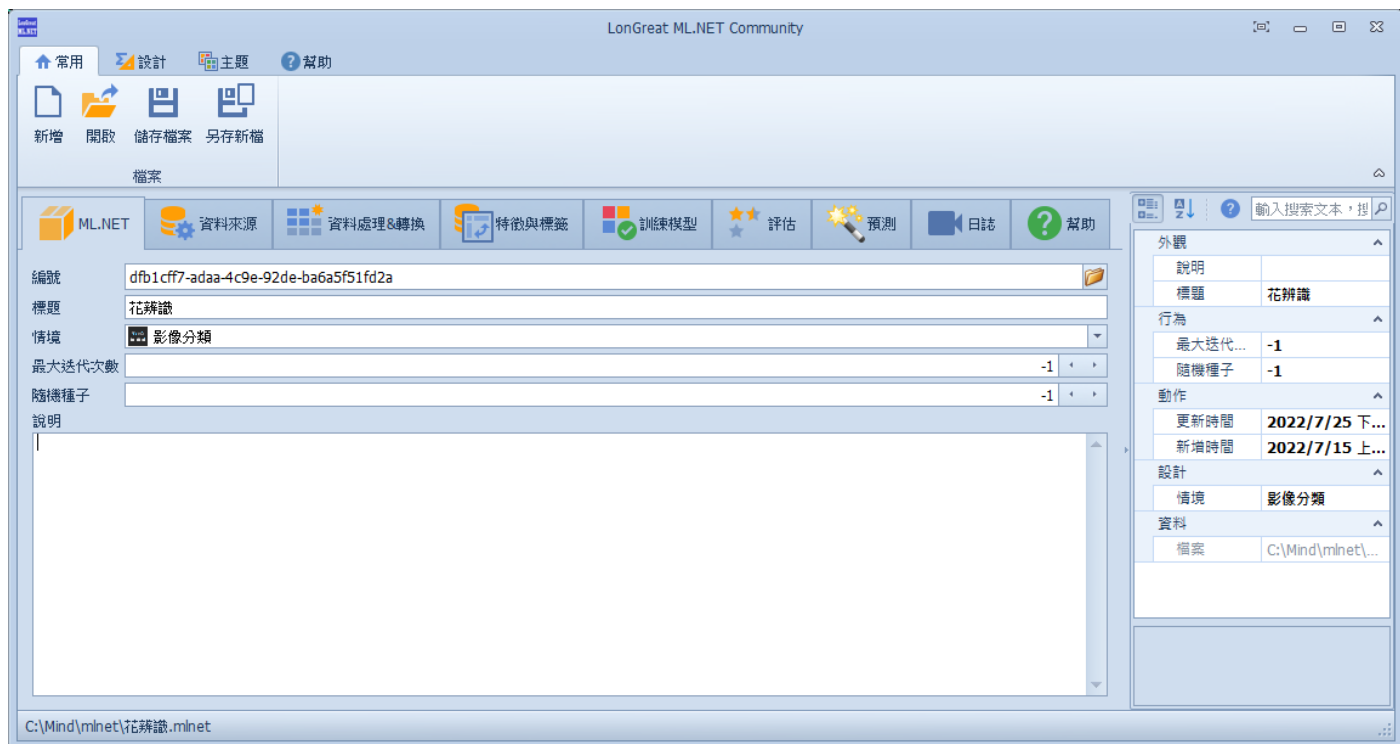
目 录

内容

简介.....	2
支持所有数据源，无须撰写程序代码.....	2
多种格式汇出与打印.....	3
强大图表.....	3
数据透视表与联动图表.....	4
资料网格检视.....	5
数据源.....	6
数据联合 Federation DataSource.....	7
档案目录转数据表.....	8
数据处理&转换.....	9
数据处理程序.....	9
数据转换函式.....	9
特征与卷标.....	11
训练模型.....	12
简单与专业属性.....	13
评估.....	14
预测.....	14
预测应用.....	14
使用相关技术.....	14
相关连结.....	15

简介

支持所有数据源，无须撰写程序代码



NO CODE

Step by Step

Wizard

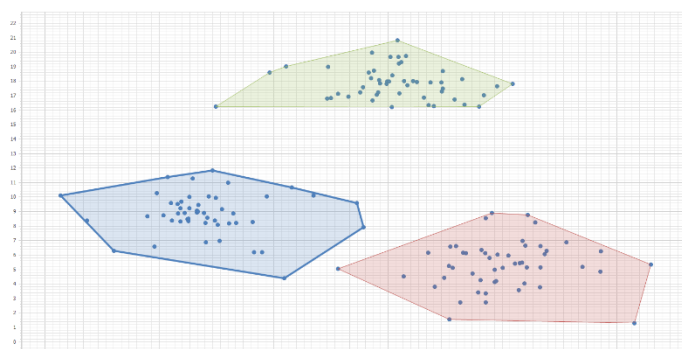
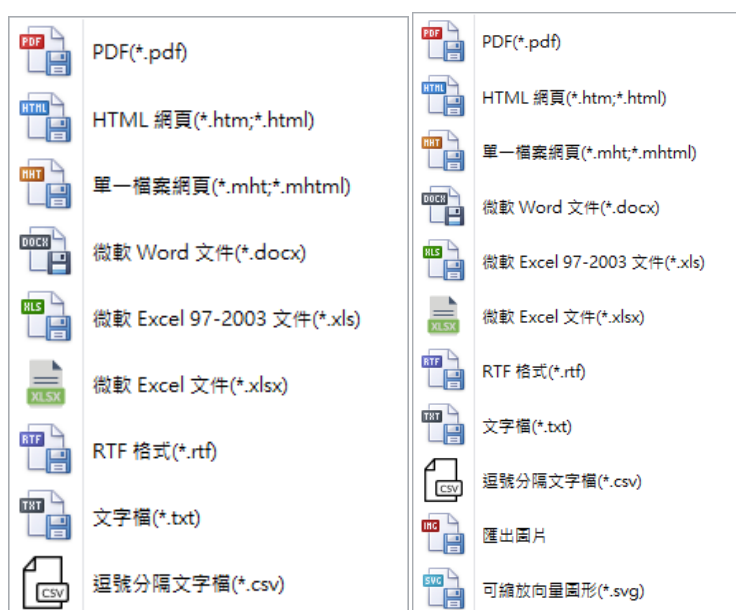
Easy & Fast & No Code

Export Diversity

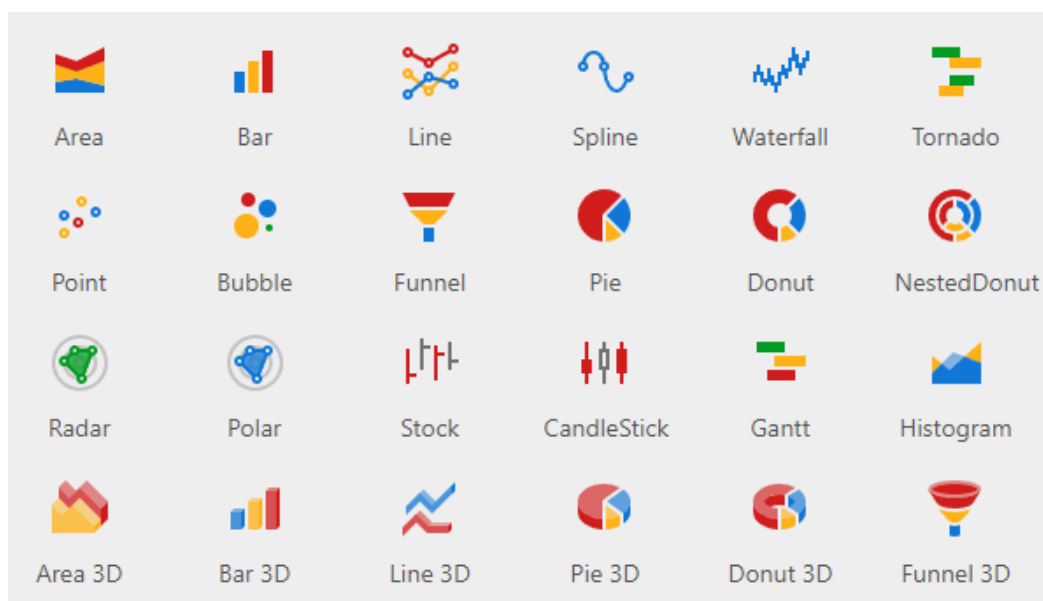
Visualize



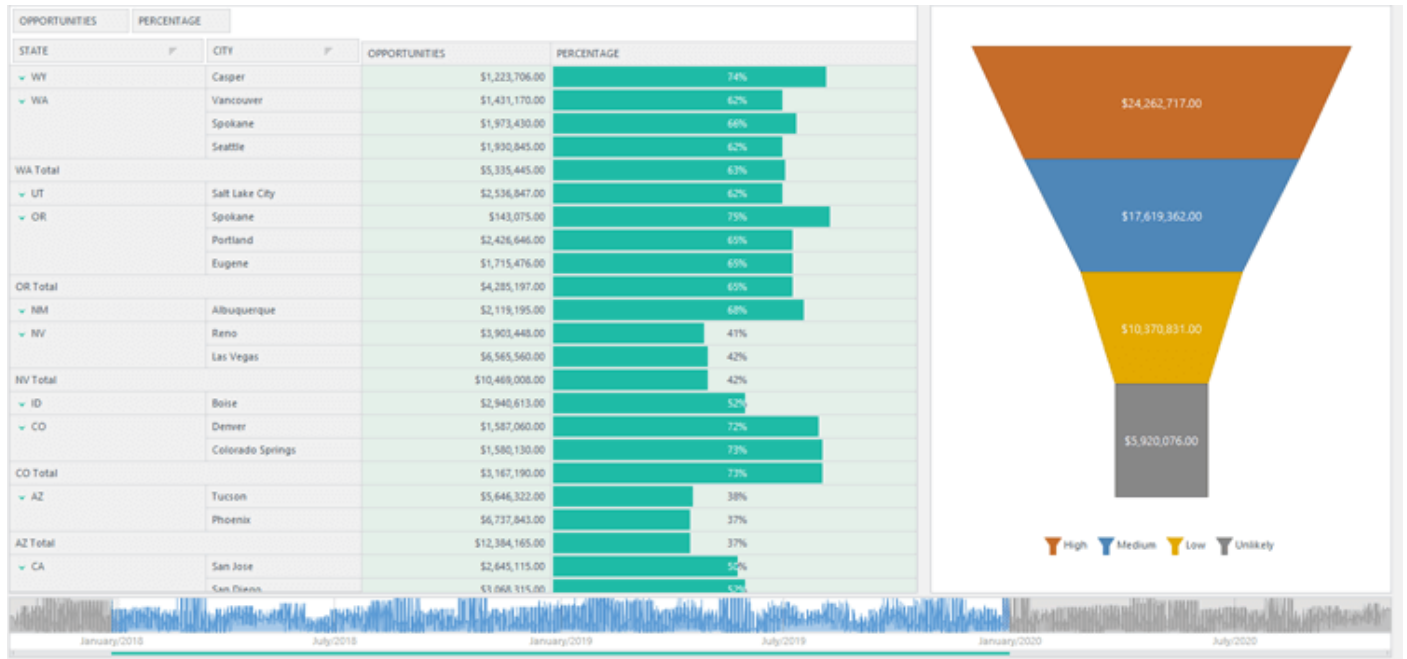
多种格式汇出与打印



强大图表



数据透视表与联动图表



资料网格检视

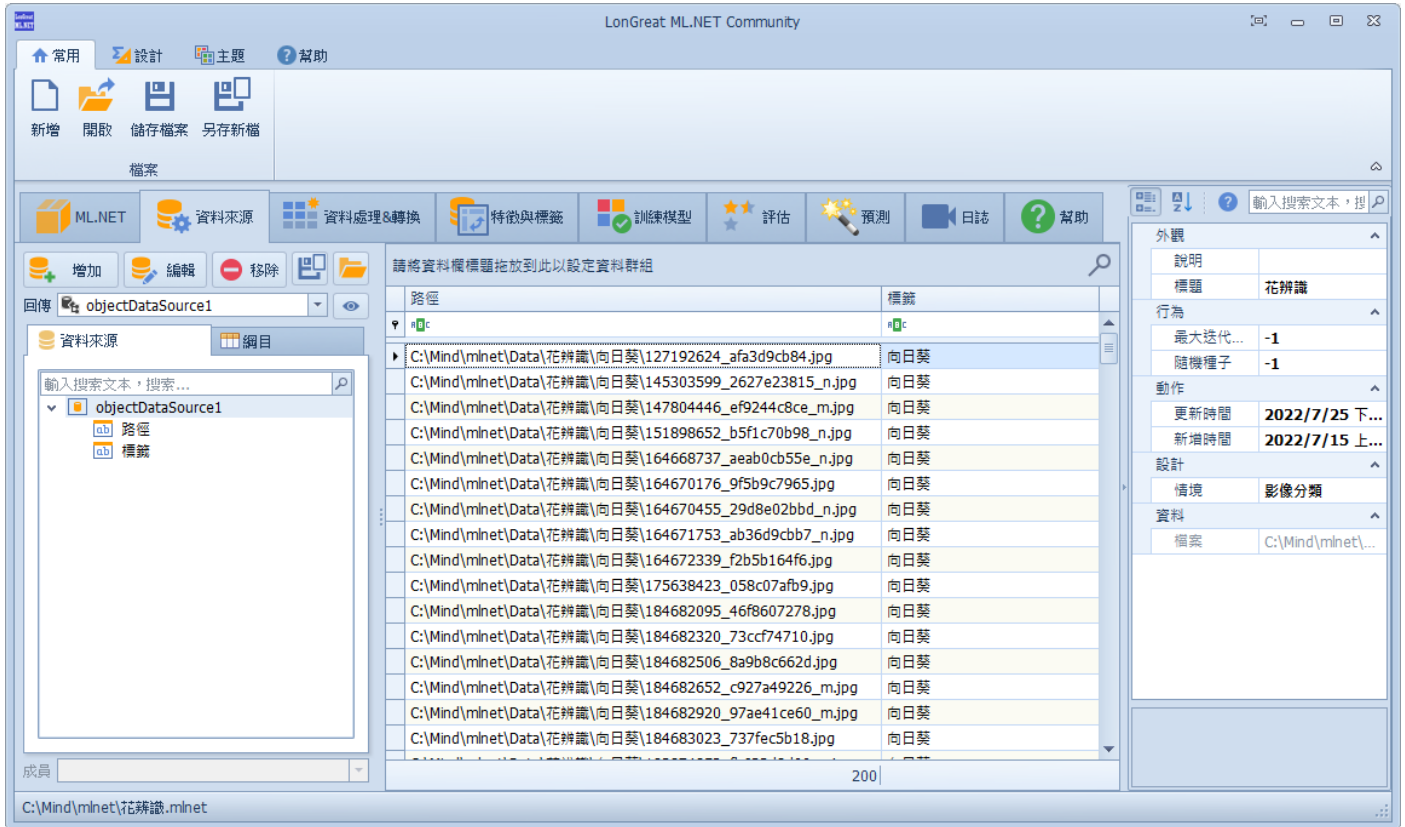
快速、群组、筛选、、、

The screenshot shows the LonGreat ML.NET Community application window. The interface is in Chinese and includes a menu bar with options like '常用' (Common), '設計' (Design), '主題' (Theme), and '幫助' (Help). Below the menu bar is a toolbar with icons for '新增' (Add), '開啟' (Open), '儲存檔案' (Save File), and '另存新檔' (Save As). The main area is divided into several sections. On the left, there is a list of data types: vendor_id (String), rate_code (Single), passenger_count (Single), trip_time_in_secs (Single), trip_distance (Single), payment_type (String), fare_amount (Single), VendorIdEncoded (String), RateCodeEncoded (String), PaymentTypeEncoded (String), PassengerCountMV (Single), TripTimeMV (Single), TripDistanceMV (Single), and Label (Single). The main area displays a data grid with columns: vendor_id, rate_code, passenger..., trip_time..., trip_dista..., payment..., fare_amo..., Vendor Id..., Rate Code..., Payment ..., Passenger..., Trip Time ..., Trip Dista..., and Label. The data is sorted by fare_amo... in descending order. The grid shows 10 rows of data. At the bottom, there is a status bar indicating '資料 4 / 100000' (Data 4 / 100000).

vendor_id	rate_code	passenger...	trip_time...	trip_dista...	payment...	fare_amo...	Vendor Id...	Rate Code...	Payment ...	Passenger...	Trip Time ...	Trip Dista...	Label
CMT	1	1	1271	3.8	CRD	17.5	[1.0]	[1.0]	[1.0]	0.42661828	1.514717	0.87563246	17.5
CMT	1	1	474	1.5	CRD	8	[1.0]	[1.0]	[1.0]	0.42661828	0.5648905	0.3456444	8
CMT	1	1	637	1.4	CRD	8.5	[1.0]	[1.0]	[1.0]	0.42661828	0.7591461	0.32260144	8.5
CMT	1	1	181	0.6	CSH	4.5	[1.0]	[1.0]	[1.0]	0.42661828	0.21570714	0.13825777	4.5
CMT	1	1	661	1.1	CRD	8.5	[1.0]	[1.0]	[1.0]	0.42661828	0.78774816	0.25347257	8.5
CMT	1	1	935	9.6	CSH	27.5	[1.0]	[1.0]	[1.0]	0.42661828	1.1142882	2.2121243	27.5
CMT	1	1	869	2.3	CRD	11.5	[1.0]	[1.0]	[1.0]	0.42661828	1.0356326	0.52998805	11.5
CMT	1	1	454	1.4	CRD	7.5	[1.0]	[1.0]	[1.0]	0.42661828	0.5410555	0.32260144	7.5
CMT	1	1	366	1.5	CSH	7.5	[1.0]	[1.0]	[1.0]	0.42661828	0.43618128	0.3456444	7.5
CMT	1	1	252	0.6	CSH	5	[1.0]	[1.0]	[1.0]	0.42661828	0.30032155	0.13825777	5
CMT	1	1	314	1.2	CRD	6	[1.0]	[1.0]	[1.0]	0.42661828	0.37421018	0.27651554	6
CMT	1	1	480	0.7	CRD	7	[1.0]	[1.0]	[1.0]	0.42661828	0.57204103	0.16130072	7
CMT	1	1	386	1.3	CRD	7	[1.0]	[1.0]	[1.0]	0.42661828	0.46001634	0.29955846	7
CMT	1	2	351	0.8	CSH	5.5	[1.0]	[1.0]	[1.0]	0.85323656	0.418305	0.18434368	5.5
CMT	1	1	407	1.1	CSH	7	[1.0]	[1.0]	[1.0]	0.42661828	0.48504314	0.25347257	7
CMT	1	2	370	0.6	CSH	10	[1.0]	[1.0]	[1.0]	0.85323656	1.1550005	1.3004058	10

数据源

几乎涵盖市面所有数据类型，数据联合 Federation DataSource 提供不同数据源合并





数据联合 Federation DataSource

是一个联合数据源，它集成了不同的数据源并通过联合查询提供统一的数据访问。

建立以下类型的联合查询：

Join

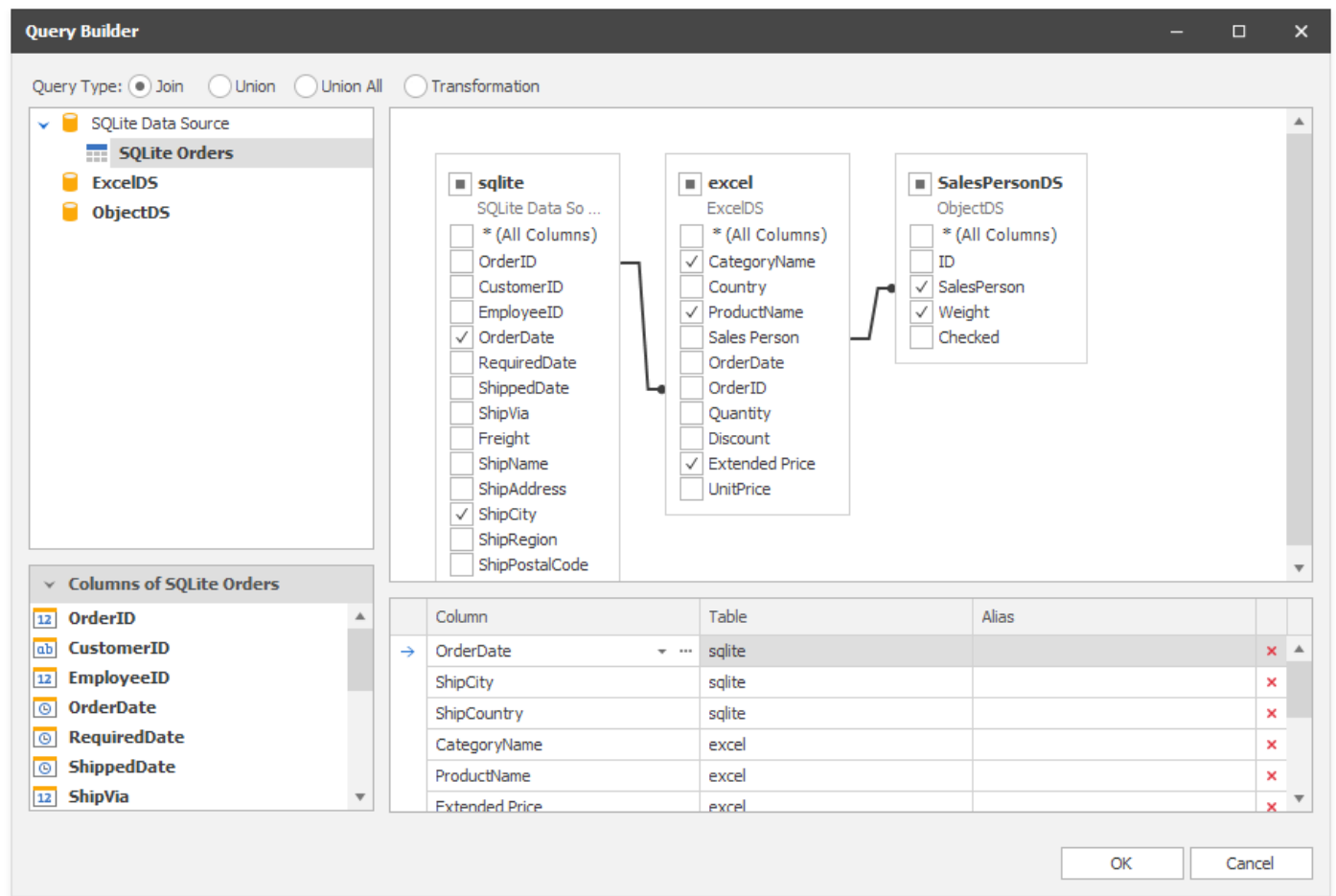
根据共享的列组合来自两个或多个表的行。连接类型指定在两个表中具有匹配值的记录。

Union / UnionAll

Union 查询将两个或多个表中的行组合成一个数据集，并删除合并表中的重复行。UnionAll 查询的操作类似于 Union，但是当它们包含相同的数据时，它们会复制来自不同表的行。您只能为包含同名列的数据源创建联合查询。此类列的数据类型应进行隐式转换。

Transformation

如果数据源包含复杂的列（对象），您可以转换其属性以在平面视图中将它们显示为单独的列。如果其中一个数据列是一个数组，您可以展开它的值并为数组的每个元素显示一个新的数据行。展开柱时，您可以将其展平并创建展平视图。



档案目录转数据表



数据处理&转换

数据处理程序

标题	形别
引导程序	BootstrapSample
快取	Cache
字段筛选	FilterRowsByColumn
按关键词段分数筛选	FilterRowsByKeyColumnFraction
缺失值栏筛选	FilterRowsByMissingValues
随机资料列	ShuffleRows
跳过资料列	SkipRows
撷取资料列	TakeRows

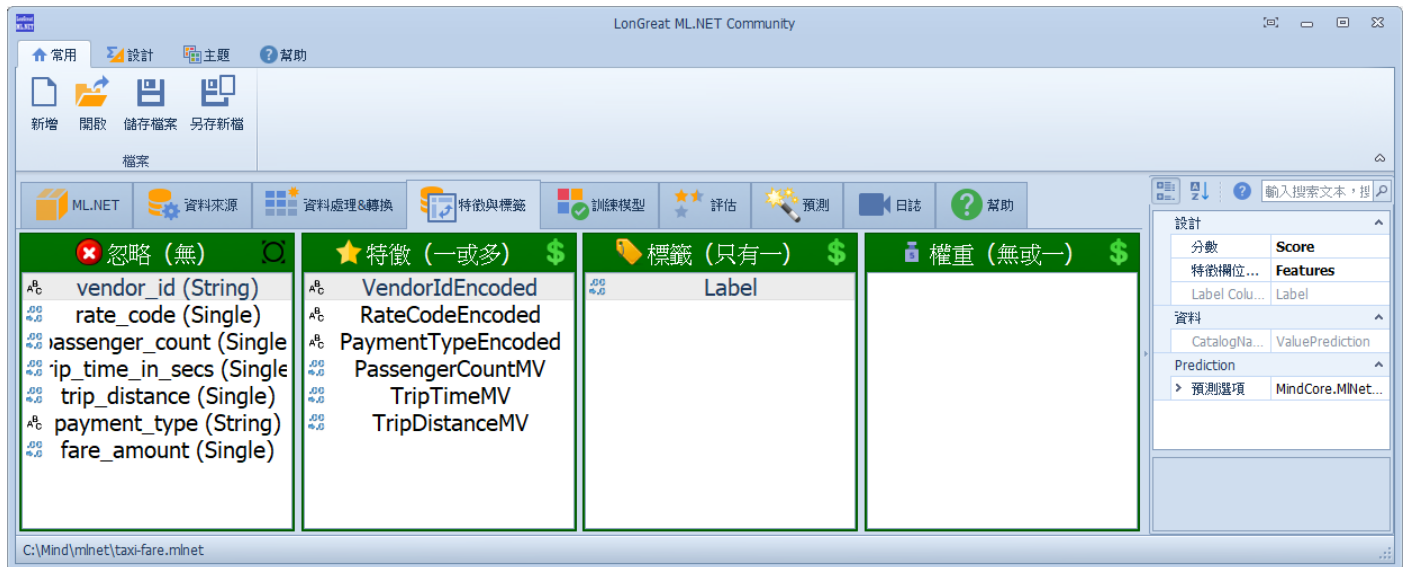
数据转换函式

转换类别	标题	形别
文字转换	文字内嵌	ApplyWordEmbedding
文字转换	特征化文字	FeaturizeText
文字转换	文字内嵌	LatentDirichletAllocation
文字转换	标准化文字	NormalizeText
文字转换	产生哈希 Ngrams	ProduceHashedNgrams
文字转换	产生哈希词袋	ProduceHashedWordBags
文字转换	产生 Ngrams	ProduceNgrams
文字转换	产生词袋	ProduceWordBags
文字转换	删除默认停用词	RemoveDefaultStopWords
文字转换	删除停用词	RemoveStopWords
文字转换	将字符标记为键	TokenizeIntoCharactersAsKeys
文字转换	分词成词	TokenizeIntoWords
文字转换	SRCNN 侦测时间异常	DetectAnomalyBySrCnn
文字转换	通过 Ssa 侦测变化点	DetectChangePointBySsa
文字转换	侦测变化点	DetectIidChangePoint
文字转换	侦测 i. i. d 尖峰	DetectIidSpike
文字转换	Ssa 侦测尖峰	DetectSpikeBySsa

文字转换	Ssa 预测	ForecastBySsa
自定义	表达式	DataExpression
核心扩充	近似内核映射	ApproximatedKernelMap
核心扩充	投影到主成分	ProjectToPrincipalComponents
特征选取	计数选择特征	SelectFeaturesBasedOnCount
特征选取	相互信息选择特征	SelectFeaturesBasedOnMutualInformation
深度学习	简单 Onnx 模型	SimpleApplyOnnxModel
资料行	连成数据域	Concatenate
资料行	复制数据域	CopyColumns
资料行	移除资料行	DropColumns
资料行	加载原始影像	LoadRawImageBytes
数据形别转换	转换形别	ConvertType
数据形别转换	哈希	Hash
数据形别转换	键转换二进制	MapKeyToBinaryVector
数据形别转换	键转换原始值	MapKeyToValue
数据形别转换	键转换浮点	MapKeyToVector
数据形别转换	类别转换键	MapValueToKey
标准化	标准 Bin 界线	NormalizeBinning
标准化	标准化全局对比度	NormalizeGlobalContrast
标准化	对数均值方差标准化	NormalizeLogMeanVariance
标准化	标准化调整单位度量	NormalizeLpNorm
标准化	均值方差标准化	NormalizeMeanVariance
标准化	最小值最大值标准化	NormalizeMinMax
标准化	分量范围调整标准化	NormalizeRobustScaling
标准化	标准化监督分量	NormalizeSupervisedBinning
遗漏值	指示遗漏值	IndicateMissingValues
遗漏值	取代遗漏值	ReplaceMissingValues
类别转换	最常编码向量	CategoricalOneHotEncoding
类别转换	哈希单一最大编码向量	CategoricalOneHotHashEncoding

特征与卷标

拖拉字段设定



训练模型

标题	情境	形别
二元平均线性分类模型	二元分类	AveragedPerceptron
二元快速树系模型	二元分类	FastForestBinary
二元判定树分类模型	二元分类	FastTreeBinary
字段感知分解机器模型	二元分类	FieldAwareFactorizationMachine
二元一般化加法模型	二元分类	GamBinary
二元逻辑回归模型	二元分类	LbfgsLogisticRegressionBinary
二元促进判定树模型	二元分类	LightGbmBinary
二元线性 SVM 分类模型	二元分类	LinearSvm
二元分类模型	二元分类	Prior
二元逻辑回归分类模型	二元分类	SdcaLogisticRegressionBinary
二元布尔回归分类模型	二元分类	SdcaNonCalibratedBinary
二元符号随机梯度下降	二元分类	SymbolicSgdLogisticRegressionBinary
单变量时间序列模型	预测	ForecastBySsa
DNN 影像分类模型	影像分类	ImageClassification
矩阵分解模型	建议	MatrixFactorization
随机 SVD 模型	异常侦测	RandomizedPca
多元 L BFGS 最大熵模型	多类分类	LbfgsMaximumEntropy
多元 LightGBM 促进式判定树模型	多类分类	LightGbm
多元贝氏机率分类模型	多类分类	NaiveBayes
最大熵多元分类模型	多类分类	SdcaMaximumEntropy
线性多元分类模型	多类分类	SdcaNonCalibrated
快速判定树回归模型	值预测	FastForestRegression
判定树回归模型	值预测	FastTreeRegression
复合波氏判定树回归模型	值预测	FastTreeTweedie
一般化加法回归模型	值预测	GamRegression
波氏回归模型	值预测	LbfgsPoissonRegression
促进判定树回归模型	值预测	LightGbmRegression
最小平方回归模型	值预测	OlsRegression
梯度下降回归模型	值预测	OnlineGradientDescent
随机双重坐标回归模型	值预测	SdcaRegression
K-means 分群	群集	KMeans

简单与专业属性

训练器提供简单与专业两种选择

允許空值繼續	True
分位數樣本數	100
分葉數目	20
分類拆分	False
分類拆分最小示例分數	0.001
平滑	0
每次分割特徵比例	1
每個特徵最大計數	255
每個節點的最大分類...	64
決策樹數	100
直方圖池大小	-1
捆綁	None
特徵比例	1
特徵重用懲罰	0
特徵群	True
特徵選擇種子	123
特徵優先使用懲罰	0
偏差	0
執行時間	False
袋子大小	0
袋子示例分數	0.7
最大分類分割點數	64
最小示例數	10
最小類範例計數	100
測試頻率	2147483647
稀疏閾值	0.7
群組欄位	
磁盤轉置	
熵係數	0
線程數	
隨機種子	123
隨機標籤	False
壓縮集合	False
獲得置信度水平	0
Softmax溫度	0

分葉數目	20
決策樹數	100
最小示例數	10

评估

评估纪录与相关分数纪录

編號	資料來源	筆數	開始時間	結束時間	時間	指標
0	DataSo...	0	2022-07-21-09-08-22 ...	2022年7月21日 上...	2022年7月21日 ...	00:00:02...

資料

平均絕對誤差 (MAE)	0.8756125531148911
判定係數	0.6380906389490668
均方誤差 (MSE)	42.38244664073625
根均方誤差 (RMSE)	6.5101802310486185
損失函數	42.38244700233495

判定係數
R-Squared 接近 1 表示模型擬合度更好。

预测

批次预测与输入预测

预测应用

训练模型产生 zip 文件，程序加载发布应用

使用相关技术

[NET6](#)

[ML.NET 1.7.1](#)

[Devexpress](#)

相关连结

 GitHub	<u>github.longreat.net</u>
 gitee	<u>gitee.longreat.net</u>