Utils

Table of Contents

Fat

IMPORTS

ML_Core.Types |

DESCRIPTIONS

FUNCTION Fat

`DATASET(Types.NumericField)`	**Fat**
`(DATASET(Types.NumericField) d0, Types.t_FieldReal v=0)`	

Will take a potentially sparse file d and fill in the missing with value v for Numeric Field datasets

PARAMETER **d0** ||| TABLE (NumericField) — They myriad format Numeric Field dataset to be filled

PARAMETER **v** ||| REAL8 — The value to assign missing records

RETURN **TABLE ({ UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 number , REAL8 value })** — A full Numeric Field dataset with every field populated

FatD

IMPORTS

ML_Core.Types |

DESCRIPTIONS

FUNCTION FatD

`DATASET(Types.DiscreteField)`	**FatD**
`(DATASET(Types.DiscreteField) d0, Types.t_Discrete v=0)`	

Will take a potentially sparse file d and fill in the missing with value v for Discrete Field datasets

PARAMETER **d0** ||| TABLE (DiscreteField) — They myriad format Discrete Field dataset to be filled

PARAMETER **v** ||| INTEGER4 — The value to assign missing records

RETURN **TABLE ({ UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 number , INTEGER4 value })** — A full Discrete Field dataset with every field populated

Gini

DESCRIPTIONS

MACRO **Gini**

	Gini
`(infile, pivot, target, wi_name='wi')`	

Creates a file of pivot/target pairs with a Gini impurity value.

PARAMETER **infile** ||| INTEGER8 — the input file, any type with a work item field

PARAMETER **pivot** ||| INTEGER8 — the name of the pivot field

PARAMETER **target** ||| INTEGER8 — the name of the field used as the target

PARAMETER **wi_name** ||| INTEGER8 — the name of the work item field, default is "wi" return A table by Work Item and Pivot value giving count and Gini impurity value

RETURN **BOOLEAN** —

SequenceInField

DESCRIPTIONS

MACRO **SequenceInField**

SequenceInField
`(infile,infield,seq,wi_name='wi')`

Given a file which is sorted by the work item identifier and INFIELD (and possibly other values), add sequence numbers within the range of each infield. Slighly elaborate code is to avoid having to partition the data to one value of infield per node and to work with very large numbers of records where a global count project would be inappropriate. This is useful for assigning rank positions with the groupings.

PARAMETER **infield** ||| INTEGER8 — field name of grouping field

PARAMETER **infile** ||| INTEGER8 — the input file, any type

PARAMETER **seq** ||| INTEGER8 — name of the field to receive the sequence number

PARAMETER **wi_name** ||| INTEGER8 — work item field name, default is wi

RETURN **BOOLEAN** — a file of the same type with sequence numbers applied