

Week 07

Classification algorithms in Rapidminer

- **Dataset:** playtennis and risk
week 07 – datasets.xls
- **Algorithms:** Zero-R, One-R, NaiveBayes, K-NN

The image shows the RapidMiner software interface. The main workspace displays a workflow with two operators: 'Read Excel' and 'W.ZeroR'. The 'Read Excel' operator is connected to the 'W.ZeroR' operator. The 'W.ZeroR' operator has several parameters, and the 'D' parameter is checked. A red circle highlights the 'D' parameter with the text 'run in debug mode?'. The 'Parameters' panel on the right shows the 'W.ZeroR' operator with the 'D' parameter checked. The 'Overview' panel on the left shows a tree view of operators, with 'W.ZeroR' selected. The 'Problems' panel at the bottom shows 'No problems found'. The 'Log' panel is also visible.

Try with k=1

The screenshot displays the RapidMiner software interface. The main workspace shows a process flow starting with a 'Read Excel' operator, followed by a 'k-NN' operator. The 'k-NN' operator is highlighted with a red arrow from the left sidebar. The right sidebar shows the 'Parameters' for the 'k-NN' operator, with the 'k' parameter set to 1 and the 'weighted vote' checkbox checked. A red arrow points from the 'weighted vote' checkbox to a callout bubble. The bottom status bar shows 'No problems found'.

Indicates if the votes should be weighted by similarity.



Overview

Repositories

Operators

Apply

Modeling (3)

Association and Item Se

Model Application (2)

Thresholds (1)

Apply Threshold

R (4)

Bio (4)

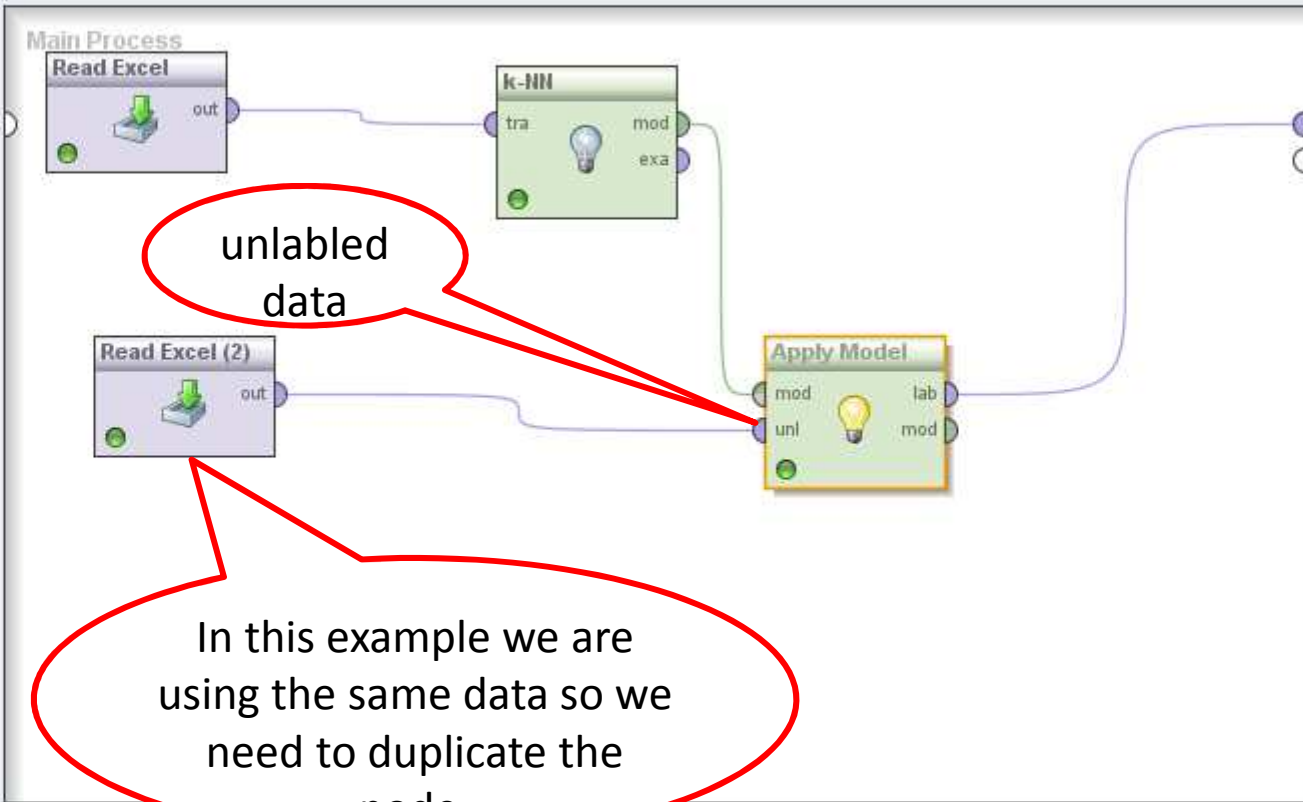
Apply Threshold

Apply Log

Apply Scaling

Apply Filter (Tamayo

Process XML Process



unlabeled data

In this example we are using the same data so we need to duplicate the node.

Parameters

Apply Mo

applicatio...

create view

Comment

Help

Apply Model

Synopsis

Problems Log

No problems found

Message Fixes Location

RapidMiner@feup-i023

File Edit Process Tools View Help

Result Overview ExampleSet (Read Excel (2))

Meta Data View Data View Plot View Annotations

Plotter
Scatter

x-Axis
prediction(PlayTennis)

y-Axis
PlayTennis

Color Column
None

Jitter

Export Image...

Repositories
Samples (none)
DB (null)
RapidRepository

Yes

No

No Yes

Log

System Monitor

Oct 26, 2011 12:04:13 PM INFO: Saving results.
Oct 26, 2011 12:04:13 PM INFO: Process finished successfully after 1 s
Oct 26, 2011 12:05:16 PM INFO: No filename given for result file, using stdout for logging results!
Oct 26, 2011 12:05:16 PM INFO: Loading initial data.
Oct 26, 2011 12:05:16 PM INFO: Process starts
Oct 26, 2011 12:05:18 PM INFO: Saving results.

Max:	1.1 GB
Total:	1.1 GB

risk dataset

The screenshot displays the RapidMiner software interface. The main workspace shows a workflow diagram with the following components:

- Read Excel**: The first process in the workflow.
- k-NN**: A classification model process, highlighted with a red box. Its parameters are visible in the right-hand pane.
- Read Excel (2)**: A second data source process.
- Apply Model**: A process that applies the k-NN model to the data.

The **Parameters** pane for the **k-NN** process is shown on the right, with the following settings:

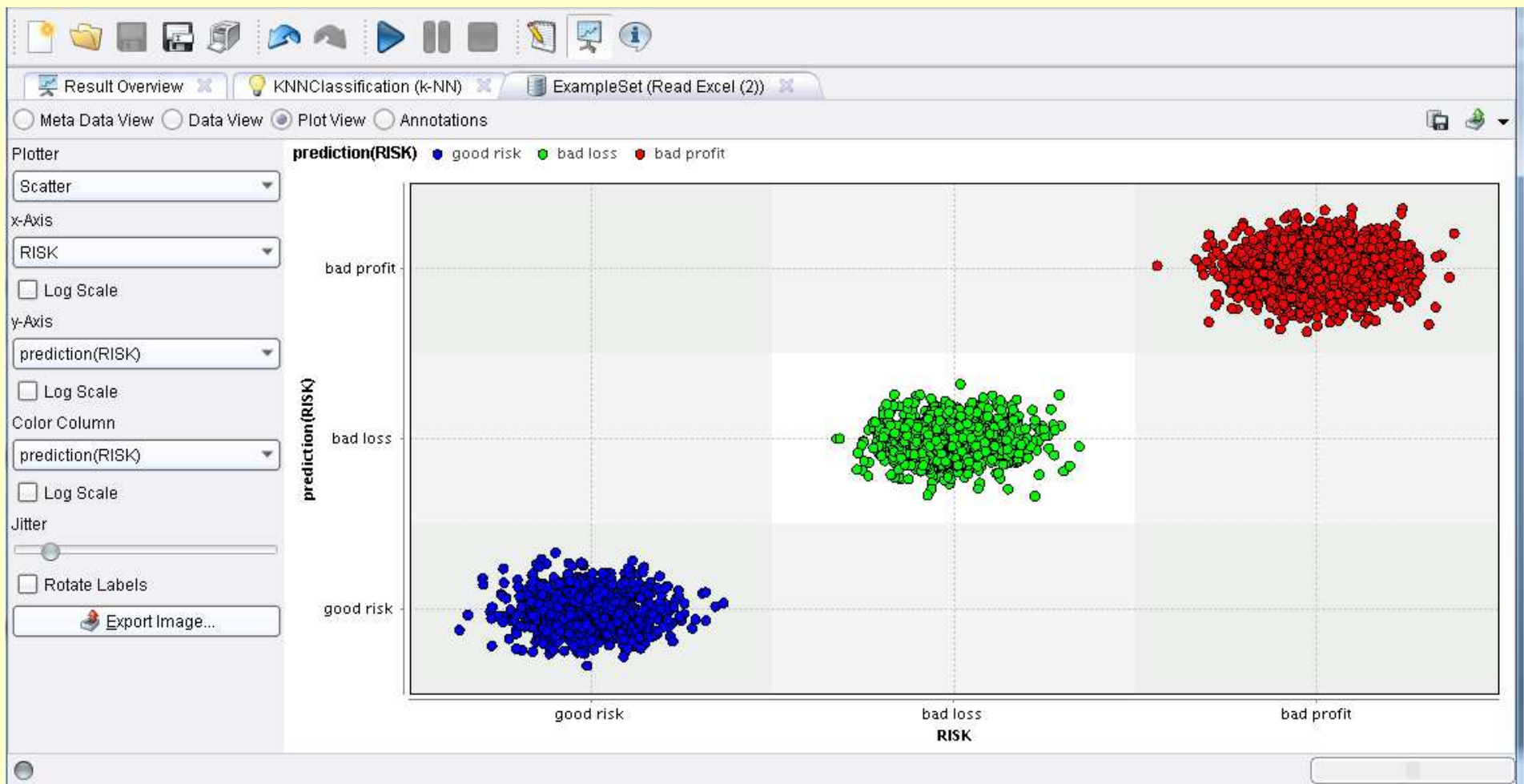
- k**: 1 (highlighted with a red circle)
- weighted vote**:
- measure types**: MixedMeas...
- mixed measure**: MixedEuclid...

The **Comment** pane at the bottom right contains the following text:

K-NN
Synopsis
Classification with k-NN based on

The **Problems** pane at the bottom shows "No problems found".

Why is this so accurate? What is wrong here?



The image shows the RapidMiner software interface. The main workspace displays a process diagram titled "Main Process". It contains two operators: "Read Excel" and "Validation". The "Read Excel" operator is connected to the "Validation" operator. A red circle highlights the "Validation" operator, with a callout bubble containing the text "double click".

The interface includes a menu bar (File, Edit, Process, Tools, View, Help), a toolbar with various icons, and several panels:

- Overview:** Shows a tree view of repositories and operators. The "Validation" operator is highlighted under the "Evaluation" category.
- Parameters:** Shows settings for the selected operator, including "logverbos..." set to "init" and a "logfile" field.
- Comment:** A text area for adding comments.
- Help:** A help icon.
- Process:** A section with "Synopsis" and "Description" for the selected operator.

The "Synopsis" section states: "The root operator which outer most operator of e process." The "Description" section states: "Each process must cont exactly one operator of the class, and it must be the".

At the bottom, there is a "Problems" panel showing "No problems found" and a "Log" panel with columns for "Message", "Fixes", and "Location".

RapidMiner@feup-i023

File Edit Process Tools View Help

Overview Process XML

Process Validation

Training Testing

k-NN Apply Model Performance

Parameters Performance

use example weights

Comment Help

Performance

Synopsis

This operator delivers performance values automatically determined in order to fit the learning type.

Description

In contrast to the other performance evaluation methods like for example [Performance \(Classification\)](#) or [Performance \(Binomial\)](#)

Problems Log

No problems found

Message	Fixes	Location

RapidMiner@feup-i023

File Edit Process Tools View Help

Result Overview PerformanceVector (Performance) KNNClassification (k-NN)

Table / Plot View Text View Annotations

Criterion Selector

- accuracy
- kappa

Multiclass Classification Performance Annotations

Table View Plot View

accuracy: 66.31% +/- 2.60% (mikro: 66.31%)

	true good risk	true bad loss	true bad profit	class precision
pred. good risk	417	158	233	51.61%
pred. bad loss	170	438	299	48.29%
pred. bad profit	217	310	1875	78.06%
class recall	51.87%	48.34%	77.90%	

Repositories

- Samples (none)
- DB (null)
- RapidRepository

Log

Oct 26, 2011 12:30:05 PM INFO: Loading initial data.
 Oct 26, 2011 12:30:05 PM INFO: Process starts
 Oct 26, 2011 12:30:10 PM INFO: Saving results.

System Monitor

RapidMiner@feup-i023

File Edit Process Tools View Help

Overview Process XML

Main Process

inp

Read Excel

Split Data

k-NN

Apply Model

res res

Parameters

Split Data

partitions Edit Enu...

sampling type shuffled s...

use local random seed

Comment

Help

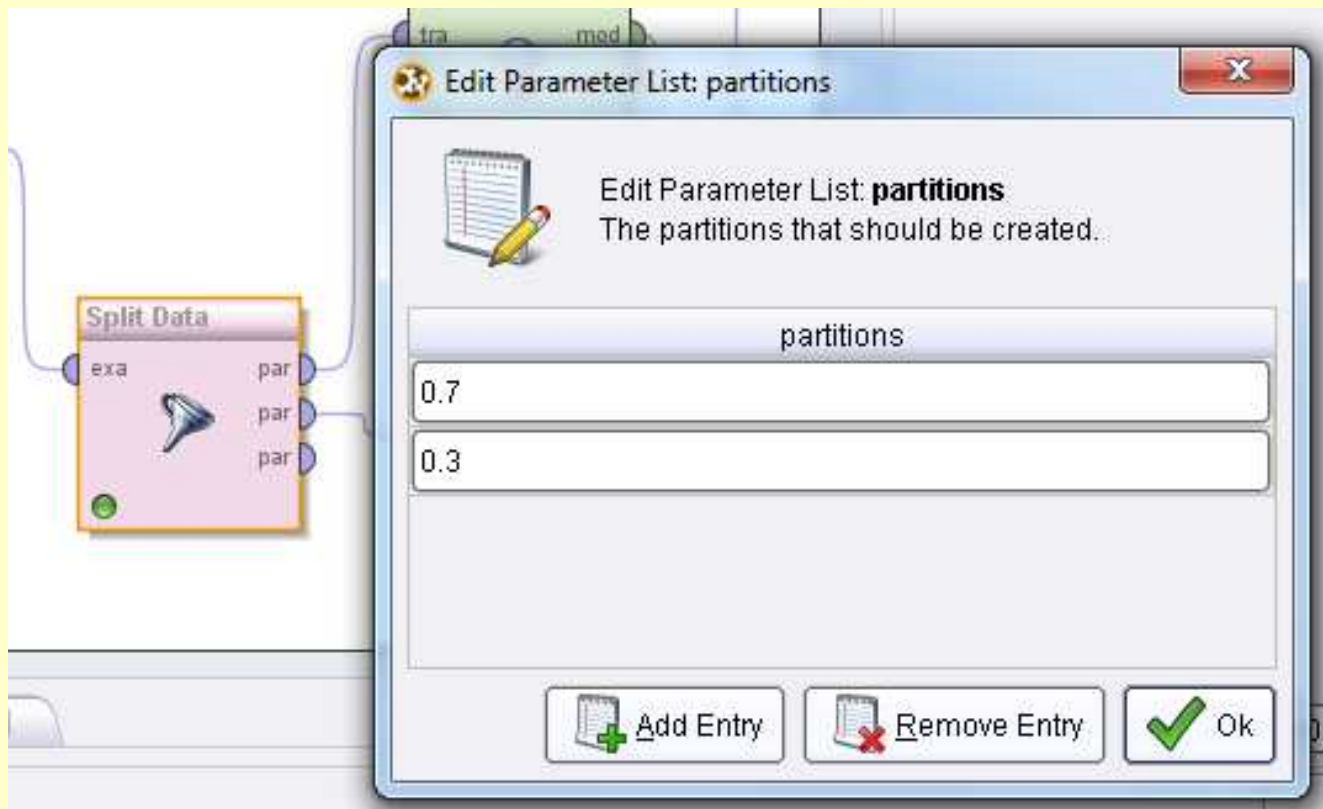
Parameters

- **partitions:** The partitions that should be created. *Range:* enumeration
- **sampling type:** Defines the sampling type of this operator. *Range:* linear sampling, shuffled sampling, stratified sampling; default: shuffled sampling
- **use local random seed:** Indicates if a local random seed should be used

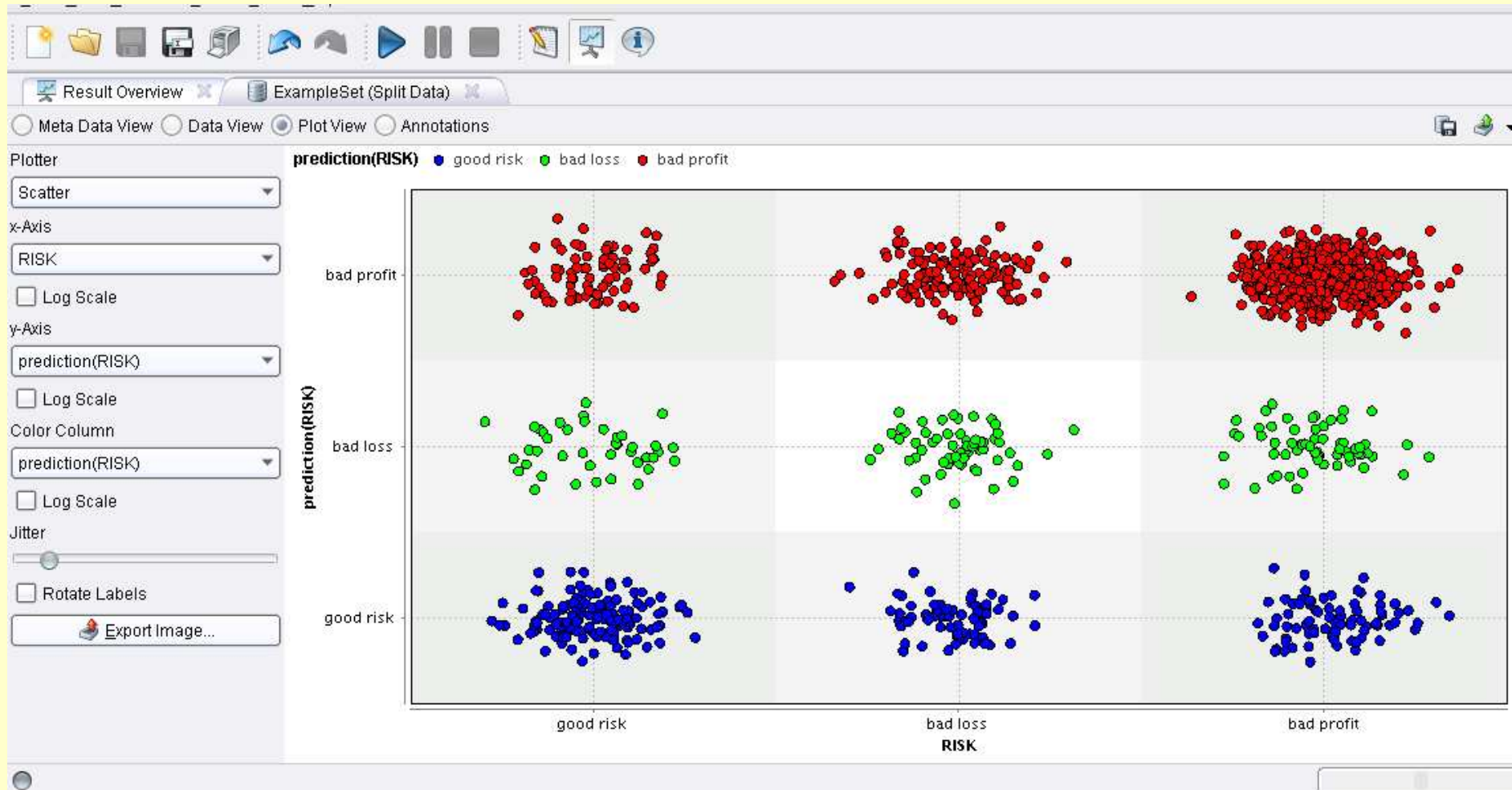
Problems Log

No problems found

Message	Fixes	Location



results with k-nn



results with the Naive Bayes model

