

# Package ‘woeR’

January 23, 2019

**Title** Weight of Evidence Based Segmentation of a Variable

**Version** 0.2.1

**Author** Kashish Soien

**Maintainer** Kashish Soien <kashish.soiens19@gmail.com>

## Description

Segment a numeric variable based on a dichotomous dependent variable by using the weight of evidence (WOE) approach (Ref: Siddiqi, N. (2006) <doi:10.1002/9781119201731.biblio>).

The underlying algorithm adopts a recursive approach to create segments that are diverse in respect of their WOE values and meet the demands of user-defined parameters. The algorithm also aims to maintain a monotonic trend in WOE values of consecutive segments. As such, it can be particularly helpful in improving robustness of linear and logistic regression models.

**Imports** dplyr

**Suggests** smbinning

**Depends** R (>= 3.4.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.0.1

**BugReports** <https://github.com/kraken19/woeR/issues>

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2019-01-23 08:30:04 UTC

## R topics documented:

apply_woe . . . . .	2
woe_binning . . . . .	2

<b>Index</b>	<b>4</b>
--------------	----------

---

`apply_woe`*Weight of Evidence based segmentation of a variable*

---

**Description**

This function applies the binning generated from `woe_binning` to new data

**Usage**

```
apply_woe(df, woe_object)
```

**Arguments**

`df` A data frame. The variable names and types need to be identical to the ones passed to `woe_binning`

`woe_object` Output object from `woe_binning` function

**Value**

Input data frame is returned with two new columns - bin & woe

**Examples**

```
library(smbinning)
## Not run: woe_object <- woe_binning(smbsimdf1, "cbs1", "fgood", initial_bins = 10)
out <- apply_woe(smbsimdf1, woe_object)
#Above example to create and apply woe segmentation
## End(Not run)
```

---

`woe_binning`*Weight of Evidence based segmentation of a variable*

---

**Description**

Create heterogeneous segmentations of a numeric variable based on a dependent variable using Weight of Evidence approach

**Usage**

```
woe_binning(df, variable, dv, min_perc = 0.02, initial_bins = 50,
            woe_cutoff = 0.1)
```

**Arguments**

df	A data frame containing input arguments - variable & dv
variable	character string specifying the column name of the variable you want to bin. Currently, the code supports only numeric and integer classes
dv	character string specifying the column name of the binary dependent variable(0,1) (NAs are ignored). Dependent variable should be either of integer or numeric class
min_perc	Minimum percentage of records in each segment. If the percentage of records in a segment falls below this threshold it is merged with other segments. Acceptable values are in the range 0.01-0.2
initial_bins	No of segments of the variable to be created in the 1st iteration. Default value = 50(2 percent) for sample size > 1500. Acceptable values are in the range 5-100
woe_cutoff	Threshold of the absolute difference in woe values between consecutive segments. If the difference is less than this threshold segments are merged. Acceptable values are in the range 0-0.2

**Details**

Weight of Evidence represents the natural log of the ratio of percent of 0's in the segment to percent of 1's in the segment. It is a proxy for how far the dv rate for a segment is from the sample dv rate (# of 1s/# of observations).

**Value**

Output is a list containing the following elements :

- a) variable - value of the input argument 'variable'
- b) dv - value of the input argument 'dv'
- c) breaks - vector specifying cut-off values for each segment. Pass it to 'breaks' argument of cut function to create segments of the variable
- d) woe - woe table for the final iteration
- e) IV - Information Value for the final iteration

**Examples**

```
library(smbinning)
woe_binning(smbimdf1, "cbs1", "fgood", initial_bins = 10)
```

# Index

`apply_woe`, 2

`woe_binning`, 2