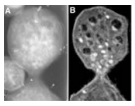


# Yeast

Donated on 8/31/1996

Predicting the Cellular Localization Sites of Proteins

## Dataset Characteristics
Multivariate

## Subject Area
Biology

## Associated Tasks
Classification

## Feature Type
Real

## # Instances
1484

## # Features
8

## Dataset Information

### Additional Information
Predicted Attribute: Localization site of protein. ( non-numeric ).

The references below describe a predecessor to this dataset and its development. They also ...

SHOW MORE

### Has Missing Values?
No

## Variables Table

| Variable Name | Role | Type | Description | Units | Missing Values |
|---|---|---|---|---|---|
| Sequence_Name | ID | Categorical | Accession number for the SWISS-PROT database | | no |

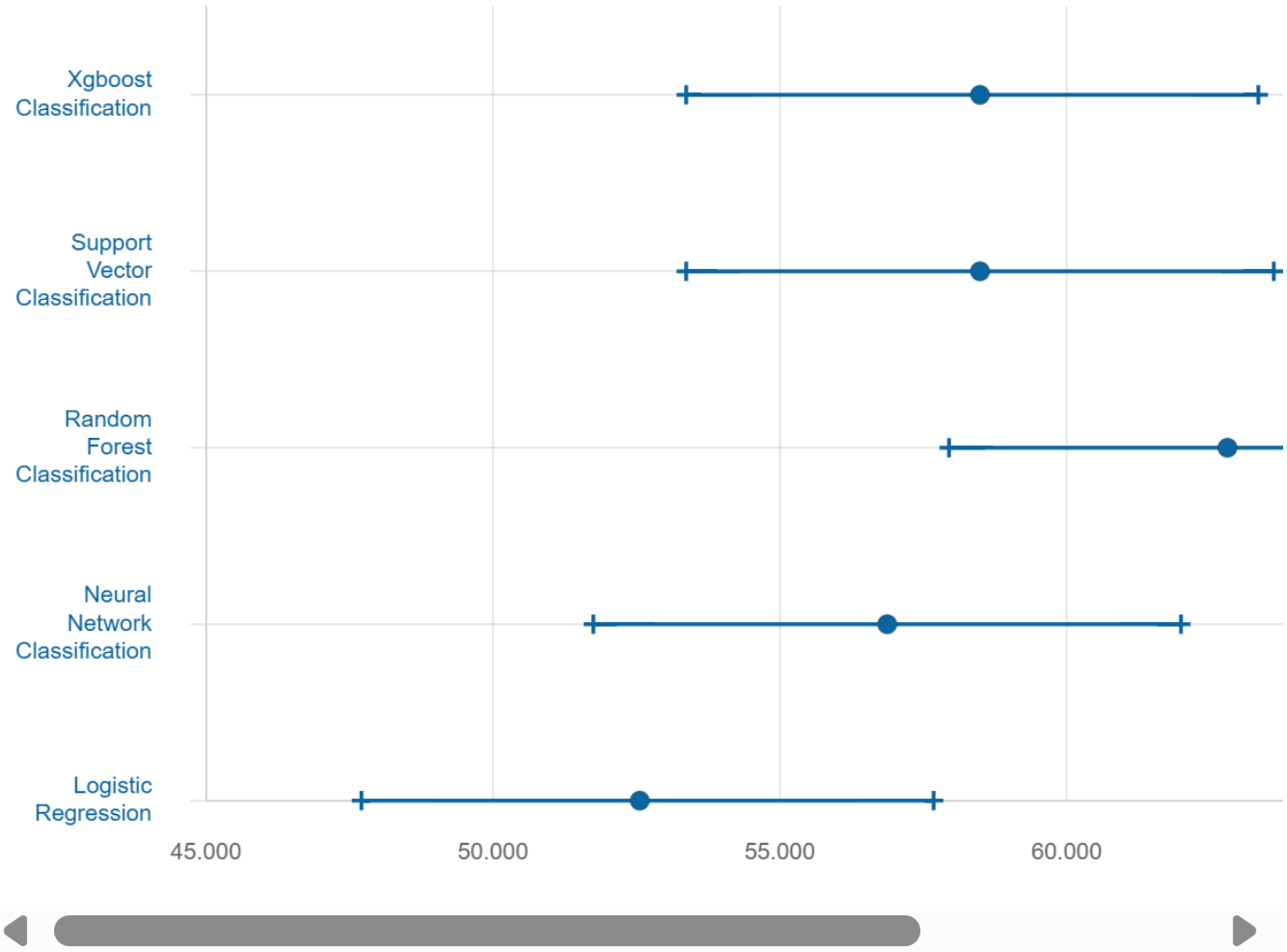
| Variable Name | Role | Type | Description | Units | Missing Values |
|---|---|---|---|---|---|
| mcg | Feature | Continuous | McGeoch's method for signal sequence recognition. | | no |
| gvh | Feature | Continuous | von Heijne's method for signal sequence recognition. | | no |
| alm | Feature | Continuous | Score of the ALOM membrane spanning region prediction program. | | no |
| mit | Feature | Continuous | Score of discriminant analysis of the amino acid content of the N-terminal region (20 residues long) of mitochondrial and non-mitochondrial proteins. | | no |
| erl | Feature | Continuous | Presence of HDEL substring (thought to act as a signal for retention in the endoplasmic reticulum lumen). Binary attribute. | | no |
| pox | Feature | Continuous | Peroxisomal targeting signal in the C-terminus. | | no |
| vac | Feature | Continuous | Score of discriminant analysis of the amino acid content of vacuolar and extracellular proteins. | | no |
| nuc | Feature | Continuous | Score of discriminant analysis of nuclear localization signals of nuclear and non-nuclear proteins. | | no |

| Variable Name | Role | Type | Description | Units | Missing Values |
|---|---|---|---|---|---|
| localization_site | Target | Categorical | | | no |

Rows per page 10 0 to 10 of 10

## Baseline Model Performance

Accuracy Precision




## Dataset Files

| File | Size |
| --- | --- |
| yeast.data | 92.8 KB |
| yeast.names | 3.2 KB |

## Papers Citing this Dataset

SORT BY YEAR, DESC

**On Possibility and Impossibility of Multiclass Classification with Rejection**
By Chenri Ni, Nontawat Charoenphakdee, Junya Honda, Masashi Sugiyama. 2019
Published in ArXiv.

**Incremental kernel PCA and the Nystr"om method**
By Fredrik Hallgren, Paul Northrop. 2018
Published in ArXiv.

**Degrees of Freedom and Model Selection for k-means Clustering**
By David Hofmeyr. 2018
Published in ArXiv.

**Multi-Resolution Dual-Tree Wavelet Scattering Network for Signal Classification**
By Amarjot Singh, Nick Kingsbury. 2017
Published in ArXiv.

**A Siamese Deep Forest**
By Lev Utkin, Mikhail Ryabinin. 2017
Published in ArXiv.

Rows per page 5 0 to 5 of 19

## Reviews

There are no reviews for this dataset yet.

WRITE A REVIEW

DOWNLOAD (18.5 KB)

IMPORT IN PYTHON

CITE

19 citations

22385 views

## Creators

Kenta Nakai

## DOI

10.24432/C5KG68

## License

View Datasets

Donate a Dataset

**LOGISTICS**

Contact

Privacy Notice

Feature Request or Bug Report