Package ‘EpiSemble’

January 20, 2025

Type Package

Title Ensemble Based Machine Learning Approach for Predicting Methylation States

Version 0.1.1

Author Dipro Sinha [aut, cre],
Sunil Archak [aut],
Dwijesh Chandra Mishra [aut],
Tanwy Dasmandal [aut],
Md Yeasin [aut]

Maintainer Dipro Sinha `<diprosinha@gmail.com>`

Description DNA methylation (6mA) is a major epigenetic process by which alteration in gene expression took place without changing the DNA sequence. Predicting these sites in-vitro is laborious, time consuming as well as costly. This 'EpiSemble' package is an in-silico pipeline for predicting DNA sequences containing the 6mA sites. It uses an ensemble-based machine learning approach by combining Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting approach to predict the sequences with 6mA sites in it. This package has been developed by using the concept of Chen et al. (2019) <doi:10.1093/bioinformatics/btz015>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.0

Imports stats, devtools, tidyverse, seqinr, Biostrings, splitstackshape, entropy, party, stringr, tibble, doParallel, parallel, e1071, caret, randomForest, gbm, foreach, ftrCOOL, iterators

NeedsCompilation no

Repository CRAN

Date/Publication 2023-06-04 14:20:02 UTC

Contents

`epiPred` *Epigenetic Modification Prediction*

Description

Predicting sequences with 6mA sites.

Usage

```
epiPred(FastaData,Species)
```

Arguments

`FastaData`	Sequence file (.fasta format)
`Species`	Model organism

Value

MethStatus: Sequences with their methylation state (methylated or non-methylated)

References

Chen, W., Lv, H., Nie, F., & Lin, H. (2019). i6mA-Pred: identifying DNA N6-methyladenine sites in the rice genome. Bioinformatics, 35(16), 2796-2800.

Examples

```
library(EpiSemble)
data<-system.file("exdata/test.fasta", package = "EpiSemble")
pred<-epiPred(FastaData=data, Species="Rice")
```

`ImpFeatures` *Important Features*

Description

Find out the most suitable features for predicting sequences with 6mA sites.

Usage

```
ImpFeatures(Fastafile,Species)
```

Arguments

`Fastafile`	Sequence file in .fasta format
`Species`	Model organism

Value

test_data_input: A matrix containing important features for prediction

References

Chen, W., Lv, H., Nie, F., & Lin, H. (2019). i6mA-Pred: identifying DNA N6-methyladenine sites in the rice genome. Bioinformatics, 35(16), 2796-2800.

Examples

```
library(EpiSemble)
data<-system.file("exdata/test.fasta", package = "EpiSemble")
imp<-ImpFeatures(Fastafile=data, Species="Rice")
```

Index