

Random Forest Algorithm in Imbalance Genomics Classification

Sudhansu Shekhar Patra ✉, Om Praksah Jena, Gaurav Kumar, Sreyashi Pramanik, Chinmaya Misra, Kamakhya Narain Singh

Book Editor(s): Rabinarayan Satpathy, Tanupriya Choudhury, Suneeta Satpathy, Sachi Nandan Mohanty, Xiaobo Zhang

First published: 18 January 2021

<https://doi.org/10.1002/9781119785620.ch7>

Summary

Random Forest (RF) is a well-known ensemble technique based on decision tree designed to improve the accuracy of CART and can be applied for classification as well as regression problems. It can be applied to problems which are classified as “large p and small n” problems with highly data adaptive. It accounts the interaction among the features and correlation among features. It works by constructing many CART trees on the training dataset making the model less interpretable. To predict for a new data, each tree votes on the outcome, and we aggregate the outcome by majority voting. Genomics data are usually high dimensional multi-class imbalance data. This chapter describes and shows the usefulness of RF algorithm for high dimensional genomic data analysis and shows the accuracy of the result.

About Wiley Online Library

[Privacy Policy](#)

[Terms of Use](#)

[Cookies](#)

[Accessibility](#)

[Help & Support](#)

[Contact Us](#)

[DMCA & Reporting Piracy](#)

Opportunities

**Subscription Agents
Advertisers & Corporate Partners**

Connect with Wiley

**The Wiley Network
Wiley Press Room**

Copyright © 1999-2021 John Wiley & Sons, Inc. All rights reserved