عنوان مقاله:

A Python-based Data mining to Address Class Imbalance Problem

محل انتشار:

سومین کنفرانس سراسری نوآوری های اخیر در مهندسی برق و کامپیوتر (سال: 1395)

تعداد صفحات اصل مقاله: 7

نویسندگان:

Seyyedali Fattahi - *Data Mining and Optimization Research Group, Centre for Artificial Intelligence, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, UKM Bangi, 43600, Selangor, Malaysia*

Zalinda Othman - *Data Mining and Optimization Research Group, Centre for Artificial Intelligence, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, UKM Bangi, 43600, Selangor, Malaysia*

خلاصه مقاله:

Orange canvas is an open source data mining tool that is based on Python scripting, visual programming and scientific computing. We developed analytical frameworks, which have advanced theoretical studies of practical learning methods, to address the class imbalance problem. In a two-class classification task, when the number of one class (majority) is greater than another (minority), this class is called imbalanced. The classification of this imbalanced class causes imbalanced distribution, poor predictive classification accuracy and a Class Imbalance Problem (CIP). We will focus on clarifying and writing a simple or clear Python script and visualize the frameworks of existing learning methods that address the CIP with well-known Synthetic over-sampling technique (SMOTE) based ensemble methods. The introduced orange workflows, Python scripting, and experimental results, will assist researchers and students to address the CIP simply. This study's aim is to design innovative methods to address CIP.

کلمات کلیدی:

Python, class imbalance problem, SMOTE, Orange Toolbox, data mining

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/576536