

SKEWED DATA CLASSIFICATION USING UNDER SAMPLING WITH DEEP LEARNING IN KERAS

S. Santhiya¹, G. Tamilarasi², S. Poornima³,
UG Scholar, Department of CSE,
P.A. College of Engineering and Technology, Pollachi.

P. Sangeetha⁴
Assistant Professor, Department of CSE,
P.A. College of Engineering and Technology, Pollachi

Classification technology can solve most problems and applications of some data characteristics with small amount, complete annotation and relative balance distribution. Class imbalance presents significant challenges to data mining and knowledge discovery applications. The fundamental issue with the imbalanced learning problem is the ability of imbalanced data to significantly compromise the performance of most standard learning algorithms. It is necessary to improve the performance of learners built from the imbalanced data. The imbalanced big datas are classified using the keras. Classification accuracy depends vitally on the quality of the training data. Quality of data was improved using under sampling method and accuracy of classifier was measured for the animal image dataset. Same is done for the safe driver prediction text dataset. Classifications for the imbalanced and balanced dataset performances were compared. Both the datasets were taken and implemented using python in keras using tensorflow backend.

Thereby, this imbalanced big datas are balanced using the undersampling and it increases the performance of the classifiers.