

# Обучение на несбалансированных выборках ансамбля классификаторов при анализе состояния сетевых сегментов

*И. С. Лебедев*<sup>1\*</sup>

<sup>1</sup> Санкт-Петербургский Федеральный исследовательский центр Российской академии наук,  
Санкт-Петербург, Россия

\* *isl\_box@mail.ru*

**Аннотация.** Актуальность рассматриваемой в статье темы состоит в решении проблемных вопросов идентификации редких событий в условиях дисбаланса обучающих множеств. Цель исследования – анализ возможностей ансамбля классификаторов, обученных на разных подмножествах разбалансированных данных. Обозначены предпосылки возникновения несбалансированных данных при формировании обучающих выборок. Предложено решение, основанное на применении ансамбля классификаторов, обученных на различных обучающих выборках, в которых имеется дисбаланс классифицируемых событий. Приведен анализ возможности применения несбалансированных тренировочных множеств для ансамбля классификаторов, где усреднение ошибок происходит за счет процедуры коллективного голосования. Осуществлена оценка распределений значений признаков в тестовых и тренировочных подмножествах. Исследована разбалансировка, заключающаяся в нарушении соотношений количества событий определенного вида внутри одного класса в обучающих подмножествах данных. Отсутствие данных в обучающей выборке приводит к увеличению эффекта разброса ответов, который усредняется увеличением сложности модели, включением в ее состав различных классифицирующих алгоритмов. В дальнейшем, в ходе функционирования возможно вносить изменения в состав классифицирующих алгоритмов, что позволяет повышать показатели точности идентификации потенциального деструктивного воздействия.

**Ключевые слова:** классификация, выявление аномалий, паразитный трафик, информационная безопасность

**Для цитирования:** *Лебедев И. С.* Обучение на несбалансированных выборках ансамбля классификаторов при анализе состояния сетевых сегментов // Прикладная информатика. 2021. Т. 16. № 3. С. 109–119. DOI: 10.37791/2687-0649-2021-16-3-109-119