

А. А. Крюкова, канд. экон. наук, ФГБОУ ВО ПГУТИ, г. Самара, kaasamara@mail.ru

С. В. Пальмов, доцент, канд. техн. наук, ФГБОУ ВО ПГУТИ, г. Самара,
ФГБОУ ВО СамГТУ, г. Самара, psv@psuti.ru

Исследование применимости методов технологии Data Mining для анализа клиентской базы телекоммуникационной компании

Сегодня коммерческой организации приходится действовать в жестких условиях рынка. Чтобы оставаться конкурентноспособной, необходимо активно использовать современные информационные технологии, в частности, Data Mining. Авторы демонстрируют эффективность применения упомянутой технологии в разрезе анализа клиентской информации на примере телекоммуникационной компании, а также выявляют метод, который наилучшим образом подойдет для обработки выбранного набора данных. С этой целью средствами аналитического пакета Orange была проведена серия экспериментов, заключавшаяся в оценке качества прогностических моделей, построенных на основе методов Data Mining. Все рассмотренные методы продемонстрировали высокие результаты, однако наиболее предпочтительными по совокупности характеристик являются «Случайный лес» и «Метод опорных векторов».

Ключевые слова: искусственный интеллект, интеллектуальный анализ данных, Orange, нейронная сеть, дерево решений, кластеризация, случайный лес, метод опорных векторов.

Введение

В настоящее время, чтобы оставаться конкурентноспособной, любая компания должна непрерывно совершенствовать свои процессы и качество предоставляемых услуг, поддерживать лояльность клиентов, стараться привлечь новых и т. д. Непременным условием реализации вышеперечисленного является использование корпоративных информационных систем (КИС). Современная КИС — это сложный аппаратно-программный комплекс, предназначенный для решения широкого круга задач и состоящий из большого числа различных модулей [3]. Однако отечественные организации, включая крупные, зачастую не используют совсем или используют крайне ограниченно

возможности КИС в разрезе технологий искусственного интеллекта, например, Data Mining (интеллектуальный анализ данных, ИАД), который может значительным образом способствовать решению упомянутых задач [9]. Они предпочитают применять более дешевые решения, в частности, *MS Excel*, который по определению не может быть использован для проведения серьезных расчетов в указанной сфере. Тем не менее методы искусственного интеллекта не являются «волшебной палочкой», гарантирующей получение требуемого результата. Следовательно, будет правильным утверждать, что существует проблема выбора метода ИАД, в наибольшей степени подходящего для определенной предметной области.