

Л. Ф. Вьюненко, Л. В. Гадасина (Санкт-Петербург, СПбГУ).
Анализ экономических данных с использованием методов машинного обучения.

Развитие страны в целом определяется развитием ее субъектов как территориальных единиц верхнего уровня. Анализ показателей развития субъектов позволяет исследовать возможности и проблемы государства. В настоящей работе используются методы машинного обучения для анализа данных об экономическом и инновационном развитии регионов Российской Федерации, полученных из открытых источников. Предлагаемый подход состоит в последовательном использовании нескольких методов машинного обучения. На первом этапе субъекты РФ разбиваются на отличающиеся друг от друга группы со схожими признаками. Затем номера групп регионов используются в качестве выходных переменных, для предсказания которых строятся классификаторы. На заключительном этапе выявляются признаки, наиболее важные с точки зрения влияния на формирование групп регионов. В качестве исходных данных рассматривались следующие показатели развития регионов, публикуемые Федеральной службой государственной статистики:

- экономические показатели регионов — оплата труда, приобретение оборудования предприятиями, текущие затраты, количество предприятий, средняя численность работников, оборот предприятий (объем товаров собственного производства и объем выполненных работ и услуг собственными силами), инвестиции в основной капитал, доходы от предпринимательской деятельности;

- показатели инновационного развития регионов — затраты на технологические инновации, исследование и разработка новых продуктов, услуг и методов для их продукции (денежные затраты), покупка машин и оборудования, связанных с технологическими инновациями (СТИ), приобретение новых технологий, приобретение прав на патенты и лицензии на использование изобретений, приобретение программного обеспечения, обучение и подготовка персонала, связанного с инновациями, маркетинговые исследования, используемые передовые производственные технологии, дизайн (деятельность по изменению формы, внешнего вида или удобства использования продуктов или услуг), инжиниринг производственных процессов.

Для выделения групп регионов использовался иерархический кластерный анализ с евклидовой метрикой вычисления расстояния между объектами и методом Варда для формирования кластеров. Классификация проводилась методом «случайный лес», позволяющим не только получить хорошие значения метрик качества классификации, но и выявлять показатели, определяющие различие классов (в данном случае групп регионов). На рисунках 1 и 2 представлены диаграммы относительной значимости анализируемых показателей экономического и инновационного развития для формирования групп регионов.

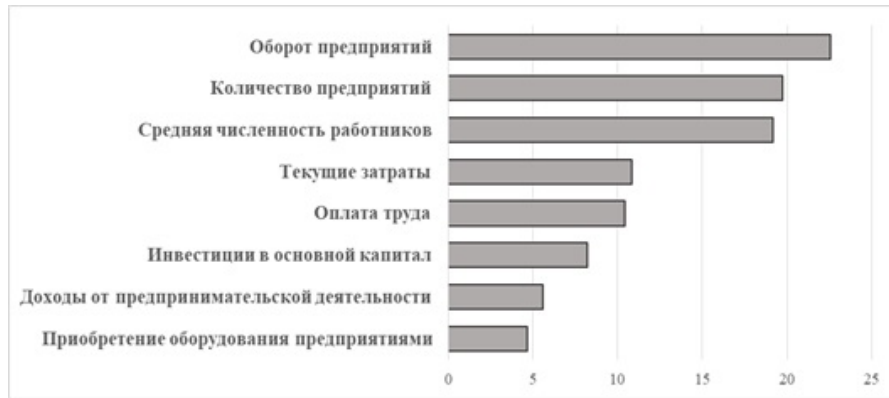


Рис. 1. Относительная значимость показателей экономического развития для формирования групп регионов.

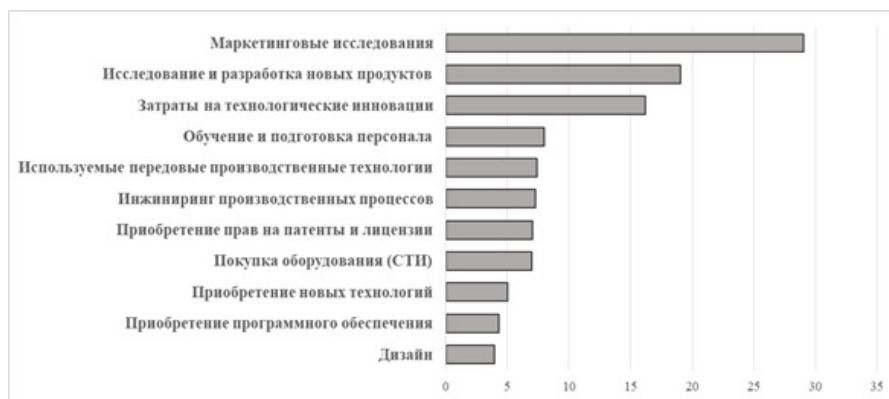


Рис. 2. Относительная значимость показателей инновационного развития для формирования групп регионов.

Предлагаемый подход позволяет выявить общие и специфические характеристики регионов. С точки зрения экономических показателей кластеры образуются главным образом на основе оборота предприятий, их количества и средней численности работников. При этом остальные факторы также оказывают влияние на формирование кластеров. Для инновационных показателей, например, такой фактор как дизайн не является кластерообразующим в настоящее время, что, по-видимому, объясняется недостаточным вниманием к нему со стороны многих предприятий.