

5. Оценка потенциала роста производства и экспорта конкурентоспособной продукции агропромышленного комплекса: препринт / Н.И. Шагайда и др.; Институт прикладных экономических исследований РАНХиГС. Москва, 2020. – 55 с. DOI 10.2139/ssrn.3860802.

6. Шагайда Н.И., Терновский Д.С. О риске падения внутренних цен на зерно и необходимости подготовки превентивных мер по поддержке цен для производителей зерна // Экономическое развитие России. 2022. Т. 29, № 8. – С. 39–45. EDN ZSOVKU.

7. Применение математических методов в управлении АПК Беларуси и России / Под ред. Н.М. Светлова, В.И. Буця. Москва: ЦЭМИ РАН, 2020. – 177 с. DOI 10.33276/978-5-8211-0782-4.

8. Светлов Н.М., Шишкина Е.А. Пространственная модель частичного равновесия на оптовых рынках сельхозпродукции субъектов РФ // Искусственные общества. 2023. Спецвыпуск. DOI 10.18254/S207751800028484-1.

УДК 004:338.43

КЛАСТЕРНЫЙ АНАЛИЗ ДЛЯ ИССЛЕДОВАНИЯ ЭКОНОМИЧЕСКИХ УСЛОВИЙ ЭФФЕКТИВНОГО ФУНКЦИОНИРОВАНИЯ АПК РЕГИОНОВ

Алетдинова А.А., д.э.н., доцент

ФГБОУ ВО «Новосибирский государственный аграрный университет»

Ключевые слова: методы кластерного анализа, программное обеспечение, экономические условия функционирования, достоинства и недостатки

Key words: cluster analysis methods, software, economic conditions of operation, advantages and disadvantages

Аннотация: В статье рассмотрены возможности реализации методов кластерного анализа с помощью современного статистического программного обеспечения Loginom и Orange Data Maning. Выделены достоинства и недостатки их применения к показателям экономических условий эффективного функционирования агропромышленных комплексов (АПК) регионов. Ключевыми факторами в выборе методов такого анализа должны стать понимание необходимого количества кластеров и принятие наличия отдельных регионов или отраслей АПК с уникальными экономическими условиями. Выбор в любом случае остается за исследователем и объясняется отсутствием методов оценки качества полученных расчетов.

Summary: In the article, the author considered the possibilities of implementing cluster analysis methods using modern statistical software Loginom and Orange Data Maning. There are advantages and disadvantages of their application to

indicators of the economic conditions for the effective functioning of agro-industrial complexes in the regions. The key factors in choosing the methods of such analysis should be understanding the required number of clusters and accepting the presence of individual regions or branches of agriculture with unique economic conditions. In any case, the choice of the method belongs to the researcher, and we can explain this by the lack of methods for evaluating the quality of the calculations obtained.

В современных условиях необходимо не просто искать условия для эффективного функционирования АПК, но и проводить их сравнительный анализ в разрезе регионов, определяя лидеров и аутсайдеров, находя сформировавшиеся конкурентные преимущества первых, которые можно брать за пример к реализации. Для этих целей хорошо подходит кластерный анализ. При этом, понимая большое количество показателей экономических условий, влияющих на эффективность функционирования АПК, его реализация должна быть автоматизирована. Современные статистические пакеты хорошо справляются с этой задачей. Но в условиях введения санкций и импортозамещения необходим пересмотр используемого программного обеспечения. Еще одна из проблем для специалистов по кластерному анализу заключается в неполном охвате методов программным продуктом. Пока единственным решением остается комбинация такого инструментария.

Рассмотрим достоинства и недостатки реализации кластерного анализа с помощью пакетов Loginom и Orange Data Mining. Нужно исходить из понимания, что показатели экономических условий для функционирования АПК регионов числовые и текстовые. С помощью пакета Loginom для нашей задачи хорошо подойдут методы (рис. 1): k-means, EM-кластеризация, самоорганизующаяся сеть (сеть Коханена, нейронный газ).

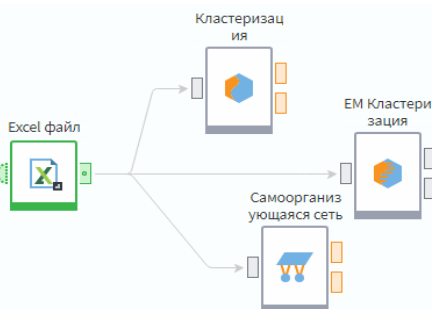


Рисунок 1. Реализация методов кластерного анализа в Loginom

Метод k-means реализуется в пакете Loginom узлом «Кластеризация», а в пакете Orange Data Maning – с одноименным названием (рис. 1). Его достоинство заключается в низком требовании к объему данных, он легко настраивается исследователем, готов предоставить необходимое заданное количество кластеров. Его недостатком считается невозможность делить данные, формирующие на координатной плоскости фигуры сложной формы [1], но на нашей практике показателям экономических условий функционирования АПК это не характерно. Вторым недостатком, который создает в исследованиях проблемы является отсутствие выбросов. Не всегда все объекты нужно группировать, т.к. из-за этого регионы/ отрасли АПК с сложившимися уникальными экономическими условия будут искусственно притянуты к определенным центрам.

Orange Data Maning позволяет выполнять методы кластерного анализа (рис. 2): k-means, Louvain, DBSCAN. EM-кластеризация основывается на оценке вероятности принадлежности объекта к какому-либо кластеру. Достоинством является возможность задать число групп. При этом есть одна особенность, метод стоит применять только в случае нормального распределения значений выбранных исследователем переменных.

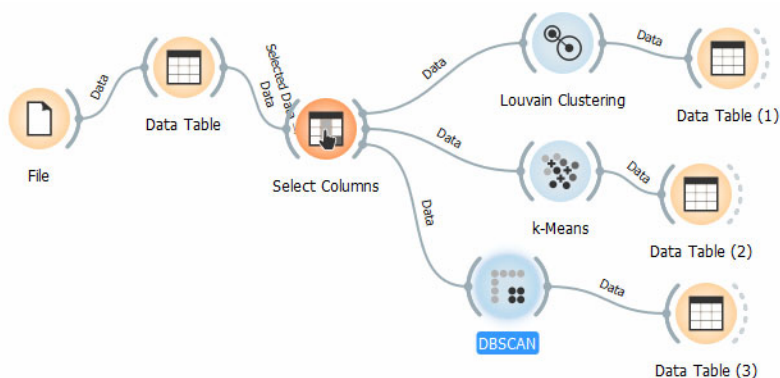


Рисунок 2. Реализация методов кластерного анализа в Orange Data Maning

Самоорганизующаяся сеть включает в себя реализацию сети Коханена и нейронная газа. Как и в случае метода k-means не отбрасывает выбросы, объединяя всё в кластеры. При анализе экономических условий функционирования АПК достаточно часто нейронный газ предлагает все объекты сгруппировать в один кластер, что тоже можно рассматривать как его недостаток. Неоднозначное мнение оставил о себе метод Лувена (Louvain). Количество кластеров определяется расчетами (рис. 3), а не желанием исследователя.

Но при этом полученные результаты на примере данных показателей экономических условий функционирования АПК регионов часто дают неплохой итог.

Info		Регионы	Cluster	Картофель 2020	Картофель 2021
82 instances (no missing data) 6 features No target variable. 2 meta attributes		1	Белгородская ...	337.734	344.416
Variables		2	Брянская обла...	1152.01	1119.68
<input checked="" type="checkbox"/> Show variable labels (if present)		3	Владимирская ...	180.571	168.701
<input type="checkbox"/> Visualize numeric values		4	Воронежская ...	717.916	636.203
<input checked="" type="checkbox"/> Color by instance classes		5	Ивановская об...	74.5462	66.7266
Selection		6	Калужская обл...	147.631	140.902
<input checked="" type="checkbox"/> Select full rows		7	Костромская о...	93.4071	80.3682
		8	Курская область	403.289	317.553
		9	Липецкая обла...	393.921	330.197

Рисунок 3. Пример реализации кластеризации методом Лувена

Метод DBSCAN хорошо отделяет выбросы. Применительно к нашим данным чаще всего это отдельные объекты со своими особенными, выделяющимися характеристиками. Кроме того, можно на основе варьирования количества соседей в группе и метрик расстояний менять количества кластеров. Но метод требует больших объемов данных, а значит, не всегда может быть реализован. Это можно отнести к его недостатку.

В заключении нужно отметить, что все методы кластерного анализа имеют свои достоинства и недостатки. Решать только исследователю. Выбрать какой-то метод достаточно сложно, т.к. в зависимости от значений переменных один может быть предпочтительнее другого. Рассмотренные методы кластерного анализа на этом не исчерпываются, их больше. К сожалению исследователям приходится комбинировать программное обеспечение для их реализации. Для исследований экономических условий эффективного функционирования АПК регионов можно порекомендовать методы k-means, Лувена, DBSCAN и сети Коханена.

Список использованной литературы

1. Ahmed, M., Seraj, R., Islam, S. M. S. The k-means algorithm: A comprehensive survey and performance evaluation // Electronics. – 2020. – Т. 9. – №. 8. – Р. 1295.
2. Алетдинова, А.А, Гааг, А.В., Лауферман, О.В. Проведение кластерного анализа с помощью алгоритма k-means и самоорганизующихся сетей // Роль аграрной науки в устойчивом развитии сельских территорий: сб. 8 Всерос. (национальной) науч. конф. с междунар. участием. – Новосибирск: Золотой колос, 2023. – С. 803–806.