

**Смирнова Светлана Михайловна**

Старший преподаватель, Московский финансовый промышленный университет «Синергия», г. Москва  
[smsmirnova@rambler.ru](mailto:smsmirnova@rambler.ru)

## **ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ МНОГОМЕРНОЙ КЛАССИФИКАЦИИ ДЛЯ ОБОСНОВАНИЯ ПРИМЕНЕНИЯ МЕР ПРОМЫШЛЕННОЙ ПОЛИТИКИ: НА ПРИМЕРЕ ПРОМЫШЛЕННОГО КЛАСТЕРА**

Моделирование стадии развития промышленного кластера значимо для формирования промышленной политики и создания институциональных условий экономического роста. В работе рассматриваются вопросы моделирования стадии жизненного цикла промышленного кластера, приводятся результаты апробации моделирования на примере нефтегазохимической промышленности Российской Федерации, а также освещаются вопросы апробации данного метода с помощью применения результатов моделирования в качестве обучающей выборки для построения дерева классификации и интерпретации полученных результатов для обоснования инвестиционных и инфраструктурных проектов.

**Ключевые слова:** отраслевая структура, национальная конкурентоспособность, кластер конкуренции, региональные рынки, математическое моделирование, многомерные методы, кластерный анализ, деревья классификации

**Svetlana Smirnova**

Lecturer, Moscow University for industry and finance Synergy, Moscow  
[smsmirnova@rambler.ru](mailto:smsmirnova@rambler.ru)

## **APPLICATION OF MULTIDIMENSIONAL CLASSIFICATION TOOLS FOR INDUSTRIAL POLICY MEASURES SUBSTANTIATION: CASE OF INDUSTRIAL CLUSTER**

The modeling stage of development of industrial cluster is significant for the formation of industrial policy and institutional conditions for economic growth. The

work deals with the modeling stage of the life cycle of industrial cluster, the results of testing applied to the petrochemical industry of the Russian Federation. Also, highlights issues approbation of this method by applying the results of simulation as a training set for building the classification tree and the interpretation of the results to justify investment and infrastructure projects.

**Keywords:** industrial structure, national competitiveness, cluster competition, regional markets, mathematical modeling, multivariate methods, cluster analysis, classification trees

Согласно значительному количеству эмпирических исследований, организация производства в структуре промышленного кластера позволяет региональной экономике достигать более высоких темпов экономического роста. Исследования в этой области также показали, что эффективность промышленного кластера существенным образом от мер промышленной политики по поддержанию его развития. При этом в ряде исследований указывается, что генезис промышленного кластера как социально-экономической системы обладает качествами жизненного цикла, то есть обладает последовательной сменой состояний, которые можно охарактеризовать как формирование, рост и развитие [5].

Изучение опыта формирования и развития кластеров указывает на то, что кроме страновой специфики, также весьма значимой является отраслевая специфика промышленного кластера. То есть кластеры, функционирующие в рамках одной национальной экономики и одной отрасли, будут иметь существенные сходства в развитии. Это также может выражаться в идентичном наборе факторов, влияющих на состояние промышленного кластера и идентичной оценке показателей, характеризующих стадии жизненного цикла кластера конкуренции. Таким образом, модель жизненного цикла должна

позволять оценить стадию жизненного цикла промышленного кластера на основе сравнения с другими кластерами исследуемой области.

Математическим методом, позволяющим решение поставленной задачи, являются алгоритмы многомерной классификации, в частности кластерный анализ, позволяющий разделять объекты в соответствии со схожими характеристиками. При этом признаковом пространством задачи многомерной классификации будет массив оценок факторов, характеризующих состояние кластера, например, таких, как единство рынка сбыта с соседними регионами/странами, наличие инфраструктуры для осуществления расширенного производства, и доступность сырьевых и трудовых ресурсов.

Стадия жизненного цикла кластера конкуренции — результат сложившейся системы поддерживающих связей, состояния спроса и предложения, а также управленческих решений и влияния предприятий-участников промышленного кластера. Так как экономическая среда неравномерна, то такие объекты объединятся по одной или нескольким классифицирующим характеристикам в одну группу в зависимости от стадии жизненного цикла.

На основе указанных предпосылок была сформулирована последовательность оценки стадии жизненного цикла промышленного кластера конкуренции.

Таблица.

Последовательность оценки этапа жизненного цикла промышленного кластера

| № п.п  | Содержание этапа   |
|--------|--|
| Этап 1 | Выявление факторов внешней среды, влияющих на процесс формирования промышленного кластера.   |
| Этап 2 | Оценка существующих производств на основе выявленных факторов.   |
| Этап 3 | Эконометрический анализ признакового пространства задачи кластеризации с целью ограничения влияния мультиколлинеарности на результаты разбиения. |
| Этап 4 | Решение задачи кластерного анализа по алгоритму k-средних с целью выявления характеристик отдельных стадий промышленных кластеров.               |
| Этап 5 | Интерпретация результатов исследования.  |

Предложенный метод моделирования стадии жизненного цикла кластера конкуренции будет апробирован на основе данных кластеров производства

нефтегазохимии России. Так, в результате анализа структуры нефтехимического производства, было выявлено 50 связанных регионально обособленных предприятий, которые потенциально могут рассматриваться как региональные промышленные кластеры. При этом два из указанных кластеров — производство полиэтилена в Республике Татарстан, поэтому данные этих кластеров можно использовать в качестве проверки обоснованности разделения промышленных кластеров на группы в зависимости от стадии их жизненного цикла с помощью вычисления апостериорной вероятности принадлежности к классу промышленных кластеров, реализованной в алгоритме дискриминантного анализа.

На основе сформированных данных, за исключением двух контрольных значений потенциальных кластеров, была решена задача кластерного анализа по алгоритму k-средних, при этом получены следующие результаты (рис. 1).

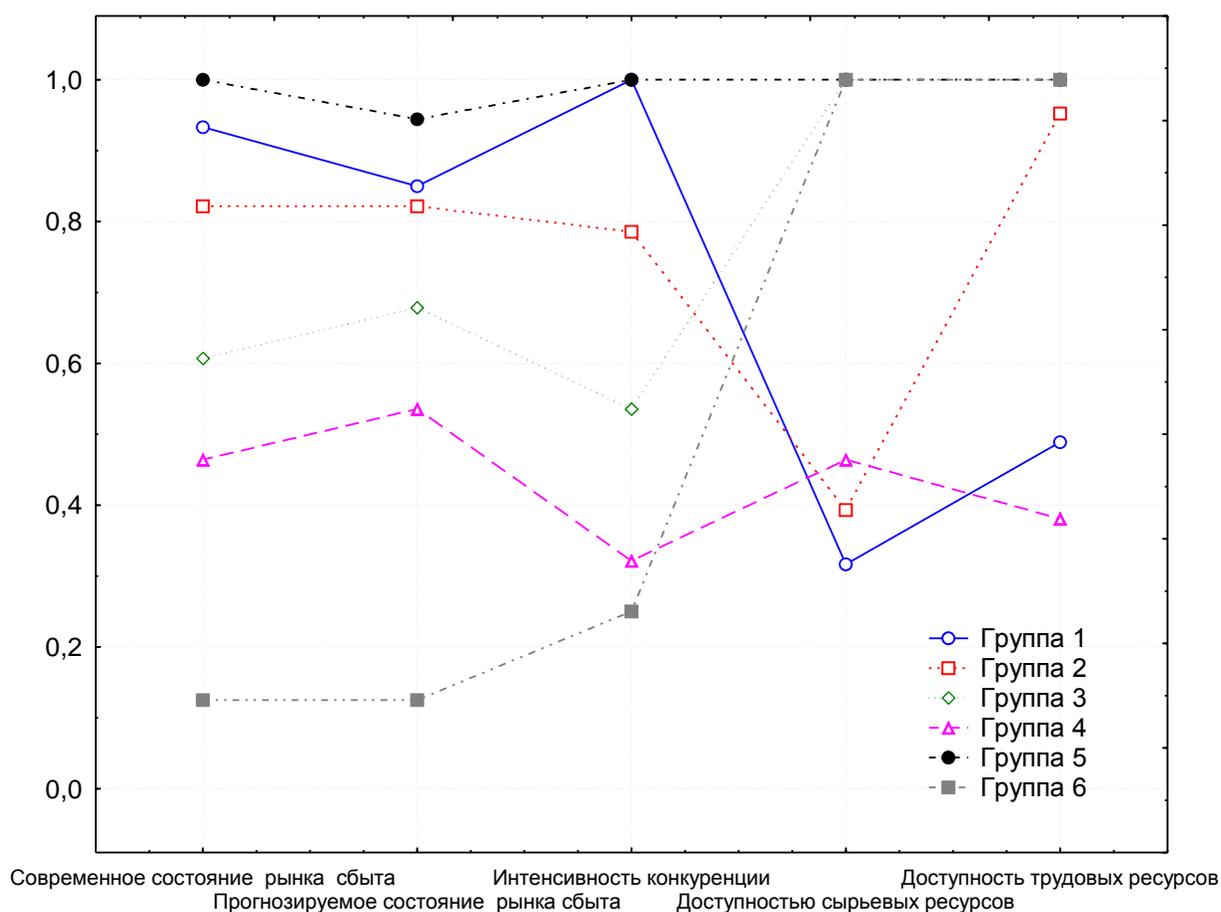


Рис.1  
Средние значения кластеров

Признаковое пространство задачи кластеризации было разделено на 5 групп. К кластерам в стадии формирования можно отнести участников двух групп 1 и 6, к кластерам в стадии роста — 2 и 3, к кластерам в стадии развития — 5. Структура разделения на группы представлена на рис. 2.



Рис .2

Структура разделения промышленных кластеров по стадиям

Таким образом, было получено следующее разделение, структура которого указана на рисунке 2.

В результате проверки полученных значений, разделения на основе двух объектов, которые являются промышленными кластерами на стадии зрелости, с помощью дискриминантного анализа было определено, что указанные объекты были распределены в 5 кластеров с апостериорной вероятностью более 70%, что говорит о том, что предложенный метод позволяет достоверно оценивать характеристики стадии жизненного цикла кластера конкуренции.

реализация инфраструктурных проектов и некоторых видов промышленной политики может существенным образом повлиять на

качественное изменение таких факторов, как доступность инфраструктуры и конкурентоспособность продукции — за счет сокращения совокупных затрат.

Таким образом, изменение факторов будет трансформировать и структуру распределения промышленных агломеративных объединений в разрезе стадии жизненного цикла, — то есть появятся различия в количестве кластеров, находящихся на различных этапах развития.

Выявление и оценка направления развития промышленного кластера на основе диагностики стадии его развития является значимым направлением развития управленческих инструментов стратегического управления и промышленной политики. Опишем подход к математическому инструментарию, позволяющему осуществлять оценку реализуемых и планируемых мероприятий по развитию различных функциональных пространств деятельности промышленного комплекса и их инфраструктурных составляющих. С этой целью может быть использован метод деревьев классификации. Так, указанный метод может прогнозировать принадлежность к определенному классу на основе одной или нескольких предикторных переменных.

Использованный метод позволяет оперировать с результатами кластеризации, или разбиения на группы объектов в зависимости от стадии жизненного цикла, а также позволяет использовать оценки в условных шкалах, и добавлять к ним новые случаи наблюдения.

Следует отметить, что указанные предикторные переменные могут иметь качественные характеристики, в нашем случае это принадлежность объекта к определенной стадии жизненного цикла.

Для получения дерева классификации необходимо:

- построить дерево классификации максимального размера;
- определить оптимальный размер дерева;
- использовать полученные результаты для классификации новых наблюдений.

Особенностью процесса классификации является то, что разделение на классы является заранее определенным, как определенным является и количество классов. При этом используемая, в основе метода классификация может быть отнесена к экзогенной переменной, позволяющей устанавливать принадлежность к указанным классам иных переменных. При этом, определяются ключевые правила отнесения объекта в виде неравенств значений определенных показателей. Например, изменение одного показателя позволяет изменить принадлежность к классу определенной стадии жизненного цикла промышленного кластера.

Опишем постановку построения деревьев классификации.

Результат решения задачи классификации представлен в виде иерархической структуры, состоящей из узлов принятия решений по оценке значений определенных переменных связанных между ребрами. Основой дерева является корневой узел, на основе которого идет построение дерева, в конце цепочки дерева находятся листовые узлы. Узлы содержат условия выбора, а ребра — взаимоисключающие результаты проверки соответствия этим критериям. Каждый листовой узел соответствует небольшой части исходного множества.

При каждой проверке условия происходит выборочек таким образом, что каждый элемент данных определяется как соответствующий только одному ребру.

Зададим матрицу переменных  $A'$  с  $M$  переменными  $X_j$  и  $N$  наблюдениями. Вектор  $Y$  состоит из  $N$  наблюдений, которые могут принадлежать к  $K$  классам.

Далее для построения дерева классификации можно применять алгоритм, основанный на построении бинарных деревьев решений, при котором происходит ветвление не более чем на две ветви (согласно принятой в практике эконометрического анализа такой алгоритм принято называть Classification and Regression Tree — CART) и алгоритм, позволяющий строить деревья с любым количеством ветвлений в узлах на основе проверки гипотезы о независимости

двух переменных по критерию  $\chi^2$ - (Chi-square Automatic Interaction Detector — CHAID). Для целей исследования в данной работе наиболее оптимальным является применение алгоритма бинарного ветвления, так как он позволяет интерпретировать полученные результаты с позиции принятия инвестиционных решений.

Обозначим  $t_p$  как исходный («родительский») узел, при этом  $t_L, t_R$  — последовательно левый и правый узел по отношению к исходному («родительскому») углу. Каждый узел (структура или класс) должен иметь ссылки на двух потомков Left и Right — аналогичные структуры. Также узел должен содержать идентификатор правила, а также содержать информацию о количестве или отношении примеров каждого класса обучающей выборки «прошедшей» через узел и иметь признак терминального узла — листа.

Дерево классификации, построенное в соответствии с правилом разделения, согласно которому разделяются на две группы максимально гомогенные наблюдения.

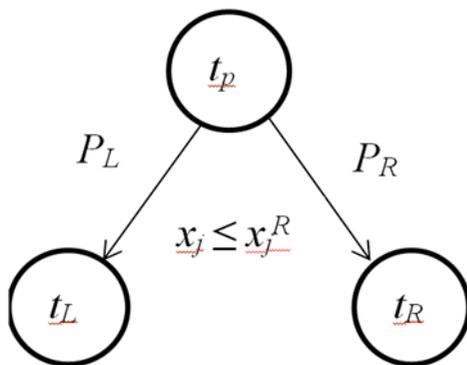


Рис. 3

где  $t_p$  — исходная точка,  $t_L, t_R$  — левая и правая ветви;  
 $X_j$  — переменная  $j$ ;  
 $x$  — лучшее разделение переменных

Максимальная однородность разделяемых узлов определяется с помощью функции «отсеивания» (impurity function)  $i(t)$ .

Для исходного узла  $t_p$  данная функция является константой для всех возможных значений для любых значений разделения переменной

$$X_j < x, j = 1, \dots, M,$$

максимальная разность однородной левой и правой узловой точки будет равнозначна максимизации изменений функции «отсеивания»  $\Delta i(t)$ :

$$\Delta i(t) = i(t_p) - E[i(t_c)] \quad (1)$$

где  $t_c$  — левые и правые исходного узла  $t_p$ .

Если предположить, что  $P_L, P_R$  — вероятности левого и правого узла, получаем:

$$\Delta i(t) = i(t_p) - P_L i(t_L) - P_R i(t_R) \quad (2)$$

Поэтому, для каждого угла классификации происходит максимизация согласно следующей формуле:

$$\arg \max_{x_j \leq x_j^R, j=1, \dots, M} [i(t_p) - P_L i(t_L) - P_R i(t_R)] \quad (3)$$

Данное выражение означает, что при классификации определяются все возможные значения всех переменных в матрице  $X$ , для принятия наилучшего решения,  $X_j < x^{\wedge}$  которое будет максимизировать изменения в оценке «отсеивания»  $\Delta i(t)$ .

Для определения функции «отсеивания» используются следующие механизмы: «отсеивание» Джини и правило «отсеивания» Тоувинга.

Правило «отсеивание» Джини (или индекс Джини) широко применяется различных направлениях теории статистики. По отношению к функции «отсеивания» он может быть определен следующим образом:

$$i(t) = \sum_{k \neq 1} p(k|t)p(1|t) \quad (4)$$

где  $k$  — индекс класса;

$p(k|t)$  — условная вероятность того, что класс  $K$  станет причиной разделения в узле  $t$ .

Применение индекса Джини к оценке функции «отсеивания» для максимизации однородности распределения определяется по следующей формуле:

$$\Delta i(t) = - \sum_{k=1}^K p^2(k|t_p) + P_L \sum_{k=1}^K p^2(k|t_L) + P_R \sum_{k=1}^K p^2(k|t_R) \quad (5)$$

Таким образом, алгоритм оценки функции «отсеивания» с помощью алгоритма Джини решает следующую проблему оценки:

$$\underbrace{\arg \max}_{x_j \leq x_j^R, j=1, \dots, M} \left[ - \sum_{k=1}^K p^2(k|t_p) + P_L \sum_{k=1}^K p^2(k|t_L) + P_R \sum_{k=1}^K p^2(k|t_R) \right] \quad (6)$$

Этот алгоритм позволяет находить обобщенные классы в значительно «зашумленных» массивах данных.

Правило «отсеивания» Тоувинга, в отличие от индекса Джини, осуществляет поиск двух классов, совокупная доля которых может быть выше 50% выборки наблюдений.

Максимизация однородности в данном случае определяется по следующей формуле:

$$\Delta i(t) = \frac{P_L P_R}{4} \left( \sum_{k=1}^K |p(k|t_L) - p(k|t_R)| \right)^2 \quad (7)$$

которая определяет решение следующей задачи:

$$\underbrace{\arg \max}_{x_j \leq x_j^R, j=1, \dots, M} \left( \frac{P_L P_R}{4} \left[ \sum_{k=1}^K |p(k|t_L) - p(k|t_R)| \right]^2 \right) \quad (8)$$

С целью решения поставленных во введении задач исследования, в данной работе был использован подход, основанный на оценке индекса Джини, так как он позволяет в большей степени определять выбор в тех случаях, когда объекты исследования равномерно распределены по классам.

Для получения итогового дерева классификации применяются различные алгоритмы остановки ветвления, в зависимости от необходимости получения целиком правильно классифицированных наблюдений или с допущением определенного уровня ошибки классификации.

Правило остановки ветвления может быть основано на оценке таких показателей, как «чистота» классификации — ошибка отнесения объекта к классу. Так как применение деревьев классификации необходимо для обоснования инвестиционных решений и мер промышленной политики, необходимо, чтобы дерево решений позволяло без ошибочно определять правила отнесения промышленных кластеров к стадиям жизненного цикла. Иные методы остановки ветвления основаны на алгоритмах, позволяющих ограничивать размер дерева в случаях, когда количество объектов наблюдения значительно. Однако по отношению к предмету исследования применение данных методов необоснованно, так как количество объектов исследования ограничено.

Апробация предложенного метода проведена на основе данных разделения на кластеры нефтегазохимического производства Российской Федерации в зависимости от стадии развития их жизненного цикла.

В качестве новых переменных использованы данные о планируемой реализации инвестиционных и инфраструктурных проектов участников отрасли нефтегазохимии, принадлежащих в различным агломеративным производственным объединениям. Схема дерева классификации и ее основные показатели представлены на рис. 4.

Данная схема была построена на основе предположения, что наиболее значимым для реализации мер промышленной политики, и инвестиционных и инфраструктурных проектов является выявление кластеров на стадии формирования и роста. Поэтому с целью построения функции использовались следующие весовые коэффициенты: для группы «не является кластером» — 2, для группы «формирование» — 2, для группы «рост» — 1, для группы «развитие» — 1.

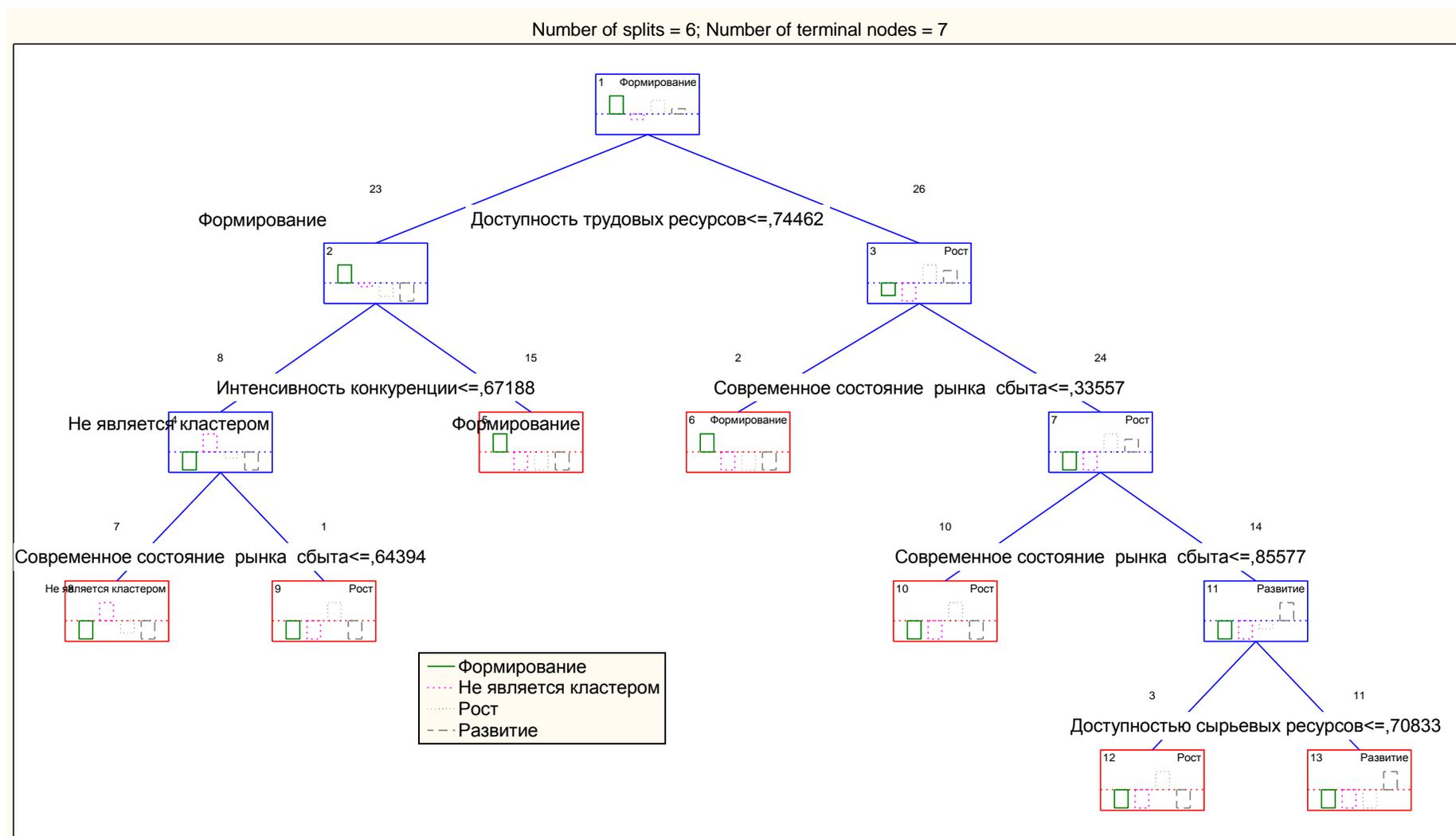


Рис. 4 Дерево классификации

В ходе анализа была оценена значимость показателей с позиции классификации и оценки вероятности принадлежности к классу. Согласно рис. 5 наиболее значимым фактором является интенсивность конкуренции, а наименее значимым — доступность трудовых ресурсов.

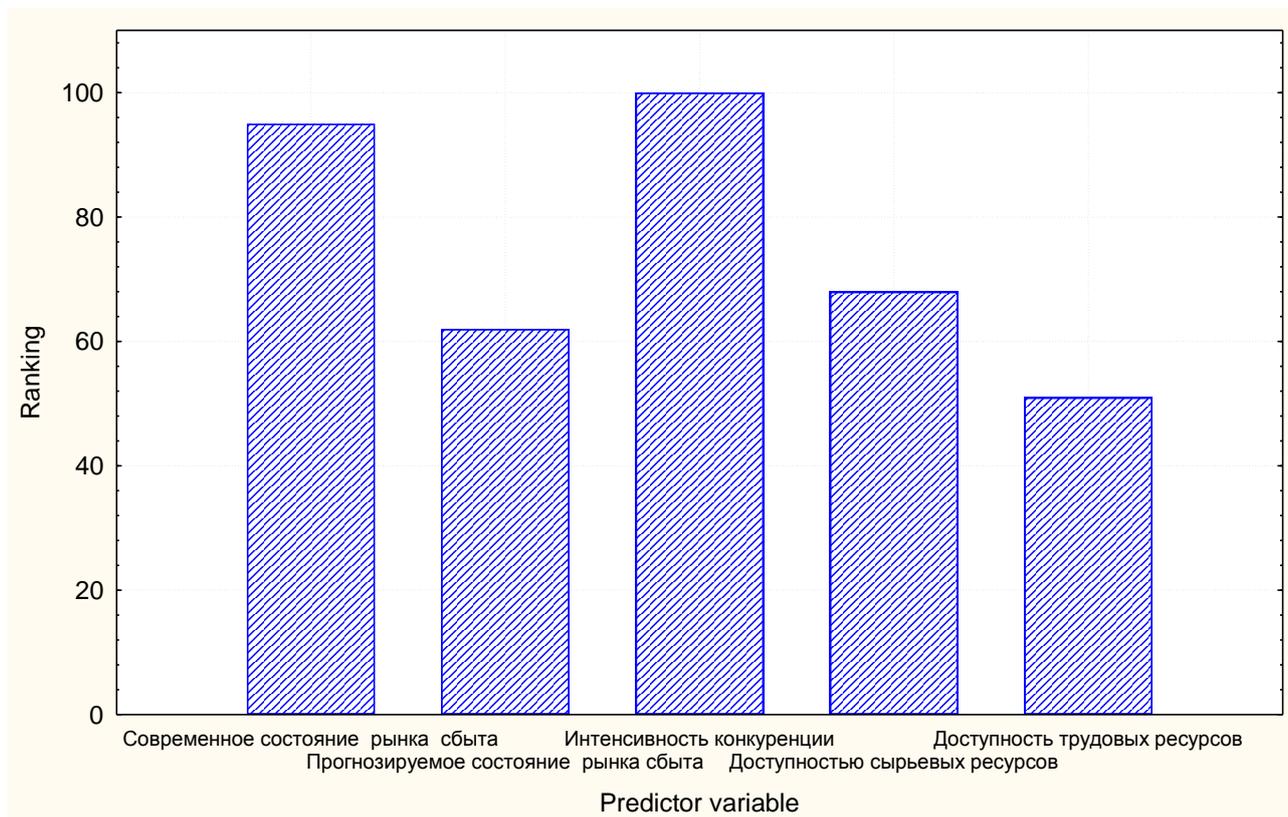


Рис. 5  
Значимость фактора для разделения на группы

В качестве контроля правила остановки, использовался алгоритм оценки индекса Джини. С целью оценки данного показателя был рассмотрен график накопления ошибки, в котором по оси X — последовательность уровней узловых точек, а по оси Y — «цена» ошибки. В частности, на район 4 уровня приходится замедление скорости возрастания значений, и это говорит о том, что классификация, включающая в себя 4 уровня, является наиболее оптимальной для целей настоящей работы.

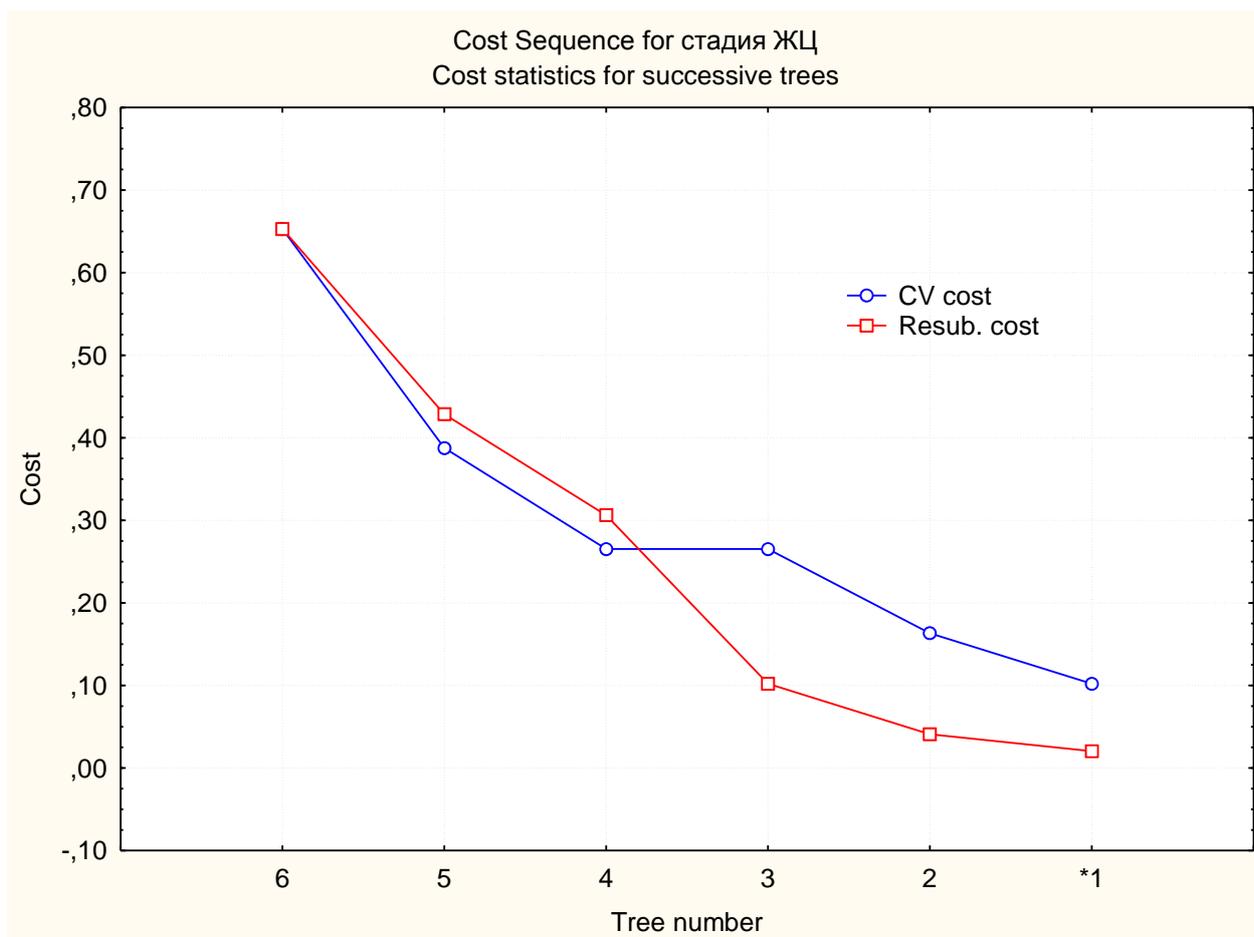


Рис.6

С целью апробации результатов классификации были проанализированы следующие инвестиционные и инфраструктурные проекты, реализуемые нефтегазохимическими производствами Российской Федерации:

1. Строительство Приморского ГКУ в Приморском Крае с целью расширения производства этилена — сырья для производства полиэтилена (совместный проект ОАО «Газпром» и ЗАО «Сибур Холдинг»).
2. Строительство ОАО «Роснефть» в Приморском Крае Восточной НХК для расширения доступности сырья, в частности, такого как смесевая нефтя и СУГ.
3. Модернизация ОАО «Роснефть» ЭП-300 в ОАО «Ангарский ЗП» с целью расширения производства полиэтиленов в Иркутской области.

4. Инвестпроект ОАО «Каустик» в строительство нового производства линейного полиэтилена в Республике Башкортостан.

5. Строительство ЗАО «Сибур Холдинг» продуктопровода для ШФЛУ на участке Западной Сибири и Ямало-Ненецкого автономного округа.

Влияние указанных проектов было отмечено и на показатели, используемые в качестве признакового пространства кластеризации. Кроме того, на построение дерева классификации оказывало влияние структурное состояние кластера. В частности, было определено, что из 10 кластеров и агломеративных объединений, структурно повлиять на изменение стадии жизненного цикла можно только в 4 случаях из 6.

Таким образом, можно сказать, что указанный метод позволяет оценить потенциальные изменения не только с позиции отдельного производственного комплекса, но и промышленного агломеративного объединения в целом.

Следует также отметить, что у указанного метода были выявлены следующие ограничения применения:

— количество анализируемых потенциальных кластеров должно быть не менее 20-ти;

— количество показателей, оцениваемых с целью определения разбиения, должно быть не менее 4-х;

— отраслевые и страновые особенности ведения деятельности анализируемыми предприятиями должны быть более значимы, чем региональные.

При этом, разработанный метод позволяет также выделять несколько типов структурного состояния промышленных кластеров на различных этапах жизненного цикла, и как следствие, более объективно оценивать состав и структуру выявляемых кластеров при оценке агломеративных объединений отрасли.

### **Литература**

1. Claus Steinle, Holder Schiele. When do industries cluster? A proposal on how to assess an industry's propensity to concentrate at a single region or nation//Research Policy. 2002. Vol. 31, iss. 6. P. 849-858.
2. Ganne B. Yveline L. Asian industrial cluster, global competitiveness and new policy initiatives: World Scientific, 2009, 601 p.
3. Tsuji M., Giovannetti E., Kagam M. Industrial agglomeration and new technologies. A global perspective: Edward Elgar, 2007, 394 p.
4. Смирнова С.М. Оценка и управление развитием кластера конкуренции национальной экономики.// Журнал ВЕСТНИК ИНЖЭКОНА, серия: Экономика, выпуск 2(29) 2009, с. 392-394
5. Смирнова С.М. Государственное регулирование и развитие кластера конкуренции регионального рынка // Научное обозрение: Экономика и управление № 4 , 2011 г. с. 54-60
6. Смирнова С.М. Оценка стадии развития промышленного кластера: подходы к моделированию// Материалы научно-практической конференции (7 апреля 2014 года): сб. ст./под ред. В.И. Хабарова.-М.: Московский финансово-промышленный университет «Синергия», 2014. — 272 с. — стр. 214-219
7. Смирнова С.М. Государственное регулирование развитием кластера конкуренции//Экономика: вчера, сегодня, завтра, № 1, 2011 г. С. 45-53