

ОПРЕДЕЛЕНИЕ ПОТРЕБНОСТИ В ТРАНСПОРТНО-ТЕХНОЛОГИЧЕСКИХ МАШИНАХ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ ЦЕХОВ ПО ДОБЫЧЕ НЕФТИ И ГАЗА

Е. С. Козин

Тюменский индустриальный университет, Тюмень, Россия
e-mail: kozines@tyuiu.ru

Аннотация. В статье рассмотрен анализ производственных показателей цехов по добыче нефти и газа с целью их кластеризации для последующего определения потребности в автомобилях и технологических машинах. Цеха имеют разные размеры, мощность, находятся в различных условиях, характеризуются различными производственными показателями, но при этом комплектуются техникой по одинаковому нормативам. Это приводит к проблемам при обеспечении бесперебойного транспортно-технологического сервиса основного производства. В ряде цехов возникают ситуации, когда запланированного количества транспортно-технологических машин недостаточно для выполнения технологических операций по ремонту или обслуживанию скважин. В таком случае направляется техника из другого подразделения, ограничивая тем самым собственные возможности транспортного обслуживания. Планирование техники происходит зачастую с учетом исторически сложившихся условий работы цеха, что в целом применимо для старых цехов с устоявшимся фондом скважин, но практически не работает для новообразованных цехов с большими объемами вновь вводимых скважин и осложненными условиями добычи. Комплектацию техникой этих подразделений осуществляют применительно к существующим цехам со схожими показателями, что чаще всего приводит к недостаточному количеству машин и простоям основного производства по причине отсутствия техники. В связи с этим требуется поиск и обоснование тех производственных показателей цехов, которые определяют их дифференциацию. Целью исследования является повышение эффективности транспортно-технологического сервиса объектов нефтегазодобычи на основе определения закономерностей влияния производственных показателей цехов по добыче нефти и газа на потребность в транспортно-технологических машинах и разработки на этой основе дифференцированных нормативов оснащения подразделений техникой. С использованием методов машинного обучения проведена кластеризация производственных подразделений, а также выявлены факторы, определяющие распределение цехов на четыре группы. К основным факторам можно отнести фонд скважин в цехе и степень сложности этого фонда. Группы определяются по степени изменения указанных факторов. Представленный подход и полученное распределение можно использовать в качестве основы для более эффективного нормирования потребности цехов в автомобилях и технологических машинах, а также как часть систем поддержки принятия решений по управлению парком транспортных средств.

Ключевые слова: машинное обучение, кластеризация, метод главных компонент, потребность в технике, транспортно-технологические машины, добыча нефти и газа.

Для цитирования: Козин Е. С. Определение потребности в транспортно-технологических машинах на основе кластеризации цехов по добыче нефти и газа // Интеллект. Инновации. Инвестиции. – 2022. – № 4. – С. 140–150, <https://doi.org/10.25198/2077-7175-2022-4-140>.

DETERMINATION OF THE REQUIREMENT FOR TRANSPORTATION AND TECHNOLOGICAL MACHINES BY CLUSTERIZATION OF OIL AND GAS PRODUCTION DEPARTMENTS

E. S. Kozin

Industrial University of Tyumen, Tyumen, Russia
e-mail: kozines@tyuiu.ru

Abstract. The article considers the analysis of production indicators of oil and gas production departments with the aim of clustering them for the subsequent determination of the need for automobiles and technological machines. The departments have different sizes, power, are in different conditions, are characterized by different performance indicators, but at the same time they are equipped with vehicles according to the same standards. This leads to problems in ensuring the uninterrupted transport and technological service of the main production.

In a number of departments, situations arise when the planned number of transport and technological machines is not enough to perform technological operations for the repair or maintenance of wells. In this case, vehicles are sent from another sub-division, thereby limiting their own transport service capabilities. Fleet planning often takes place taking into account the historical conditions of the department, which is generally applicable for old departments with an established well stock, but practically does not work for newly formed departments with large volumes of newly commissioned wells and complicated production conditions. These subdivisions are equipped with vehicles in relation to existing workshops with similar indicators, which most often leads to an insufficient number of machines and downtime of the main production due to lack of machines. In this regard, it is necessary to search for and justify those production indicators of departments that determine their differentiation. The aim of the paper is to increase the efficiency of transport and technological service of oil and gas production facilities based on determining the patterns of influence of production indicators of production and gas shops on the need for transport and technological machines and developing, on this basis, differentiated standards for equipping units with vehicles. Using machine learning methods, the clustering of production units was carried out, and the factors that determine the distribution of departments into four groups were identified. The main factors include the stock of wells in the department and the degree of complexity of this stock. Groups are determined by the degree of change in these factors. The presented approach and the resulting distribution can be used as a basis for more efficient standardization of the needs of departments in automobiles and technological machines and also as part of decision support systems for vehicle fleet management.

Key words: machine learning, clustering, principal component analysis, the need for vehicles, transport and technological machines, oil and gas production.

Cite as: Kozin, E. S. [Determination of the requirement for transportation and technological machines by clusterization of oil and gas production departments]. *Intellekt. Innovacii. Investicii* [Intellect. Innovations. Investments]. Vol. 4, pp. 140–150, <https://doi.org/10.25198/2077-7175-2022-4-140>.

Введение

Автомобили и технологические машины участвуют в производстве работ по добыче нефти и газа. Производственные цеха имеют различные производственные показатели и для одного предприятия могут находиться в различных географических локациях с разными природно-климатическими условиями и удаленностью от основных инфраструктурных объектов и мест базирования техники. Однако комплектация цехов автомобилями и технологическим транспортом реализуется, как правило, для всех цехов по единым нормативам. При этом фактическая потребность цеха в технике может отличаться от нормативной. На практике это решается резервированием машин или временным участием техники одного цеха в работах другого. Каждая возникающая проблема решается индивидуально, что отнимает у транспортной службы много времени и ресурсов. Лежащей в основе этого научной проблемой является отсутствие учета производственно-технологических факторов при формировании потребности подразделений в технологическом транспорте.

В научной литературе представлено достаточно большое количество исследований по определению потребности в технике [2, 5, 20], по оценке эффективности использования транспорта [1, 6], а также по различным методам расчета и оптимизации количества машин для предприятий [3, 4], однако влияние производственно-технологических факторов на эффективность транспортно-технологического сервиса процессов добычи нефти и газа исследовано недостаточно.

В таких условиях рациональным является системное решение вопроса определения потребности в технике для каждого цеха или группы цехов [19].

Гипотеза исследования заключается в возможности объединения цехов по добыче нефти и газа в группы по ряду производственных показателей и формировании на этой основе нормативов обеспечения автомобилями и технологическими машинами.

Целью исследования является повышение эффективности транспортно-технологического сервиса объектов нефтегазодобычи на основе определения закономерностей влияния производственных показателей цехов по добыче и газа на потребность в транспортно-технологических машинах и разработки на этой основе дифференцированных нормативов оснащения подразделений техникой.

Методы исследования

В работе использованы методы машинного обучения и анализа данных, реализуемые в рамках фреймворка *sklearn* языка программирования *Python* [7, 8, 15]. Для обработки и представления информации был использован пакет работы с данными *pandas*, позволяющий обрабатывать большие массивы данных с использованием гибкой системы форматирования и запросов к ним. Для анализа данных и выявления взаимосвязи факторов между собой использовались методы корреляционного анализа и построения корреляционной матрицы [13]. Группирование данных осуществлялось с использованием группы методов обучения без учителя (*unsupervised learning*): кластерного анализа ме-

тодом иерархической древовидной кластеризации, методом k -средних (k -means) и метода главных компонент (*principal component analysis* – PCA) [10, 21, 25]. Метод кластерного анализа является довольно эффективным при наличии значительного объема данных и необходимости их разбиения на однородные группы (кластеры или классы). Метод главных компонент был использован также для определения перечня скрытых факторов, определяющих кластеризацию производственных подразделений, а также для визуализации полученных кластеров [9]. Для сравнения точности кластеризации и прогнозирования отнесения цехов к тому или иному кластеру был использован наивный классификатор Байеса (*naive_bayes* пакета *sklearn*) [12, 26]. Визуализация данных была получена методами библиотеки *matplotlib* и пакета *seaborn* (тепловая диаграмма корреляции переменных). Все вычисления проводились средствами *Python* и среды *Jupyter Notebook* [16]. Полученные результаты могут быть использованы в качестве основы для применения методов нормирования потребности в технике [17, 24].

**Основная часть
Исходные данные**

В качестве независимых переменных (исходных факторов) по результатам экспертного анализа был выбран ряд факторов, рассмотренный ниже. Планирование объемов работ транспортно-технологических машин на рассматриваемых предприятиях по добыче нефти и газа происходит на календарный год вперед с использованием общего объема машино-часов W , планируемого пробега L и наработки верхнего (навесного) оборудования N_{eq} . Эти показатели затем при необходимости пересчитываются в количество единиц техники с учетом продолжительности рабочего дня и количества смен.

К производственным показателям цехов относятся: общий фонд скважин в цехе T_f , который включает все скважины, в том числе недействующие; действующий фонд скважин A_f , включающий только действующие добывающие скважины; количество скважин на кустовой площадке N_{cl} ; количество рабочих N_e ; ввод новых скважин N_{nw} ; дебит скважин D ; текущие и капитальные ремонты скважин N_{rep} . Отдельная группа факторов – факторы, характеризующие осложненность фонда скважин (выраженные в количестве осложненных скважин), к которым можно отнести гидратообразование N_h , солеотложения N_s , взвешенные частицы N_{ss} , парафины N_p , обобщенный показатель осложненности фонда N_{com} , представляющий собой сумму скважин с указанными частными видами осложненности. В качестве зависимой переменной рассматривались три показателя: суммарное количество машино-часов работы техники за определенный период W ; суммарный пробег L ; суммарное количество машино-часов работы навесного оборудования N_{eq} .

По результатам сбора и предварительной обработки первичной информации была сформирована исходная таблица данных, содержащая численные значения указанных выше переменных за три года: с 2014 по 2016 для каждого цеха по добыче нефти и газа одной из крупнейших нефтегазодобывающих компаний в РФ.

Первые четыре строки сформированного для последующего анализа датасета численных значений зависимых и независимых переменных для последующего исследования по определению влияния производственных показателей цехов по добыче нефти и газа на потребность в транспортно-технологических машинах представлены в таблице 1.

Исходные данные обрабатывались с использованием библиотеки анализа данных *pandas* языка программирования *Python*.

Таблица 1. Исходная таблица данных для последующего анализа (фрагмент)

№ цеха	W, маш-ч	L, км.	N_{eq} , мото-ч	T_f , ед.	A_f , ед.	N_{cl} , шт.	N_e , чел.	N_h , ед.	N_s , ед.	N_{ss} , ед.	N_p , ед.	N_{com} , ед.	N_{nw} , ед.	D, м ³ /час	N_{rep} , ед.
1	166026	3183816	8393	2669	830	188	91	0	0	36	6	42	4	93	446
2	117809	1583133	5832	857	274	57	56	2	30	250	38	320	19	68	249
3	155694	2382768	9298	2672	1090	215	95	0	114	94	30	238	11	180	345

Источник: разработано автором

Корреляционная матрица

Предварительным этапом обработки данных является построение корреляционной матрицы взаимного влияния факторов друг на друга. Матрица корреляции была получена методом *corr()* для структуры данных *DataFrame* (Рисунок 1).

По результатам анализа корреляционной матрицы наблюдается сильная корреляция факторов W , L и N_{eq} . Очевидно, для дальнейшей работы в качестве зависимой переменной следует выбрать один из них – «Машино-часы работы техники», что также соответствует экспертному мнению специалистов

организации. Факторы N_{cl} и N_e коррелируют с фондом скважин (T_f и A_f соответственно). Наблюдается корреляция между вновь вводимыми скважинами N_{nw} и гидратами N_h . Также имеется взаимосвязь суммарного количества скважин осложненного фонда

(N_{com}) с отдельными показателями осложненности: парафинами и количества взвешенных частиц (N_p и N_{ss}). Для дальнейших исследований исключим показатель суммарного количества осложненных скважин N_{com} в силу его субъективности.



Рисунок 1. Диаграмма корреляции факторов между собой
 Источник: разработано автором

Кластерный анализ

Исходная таблица данных по результатам анализа корреляции факторов была скорректирована и направлена на проведение кластерного анализа в целях определения наличия в структуре данных групп (кластеров) и объяснения их значения [22, 23]. Из таблицы 1 были исключены следующие столбцы: W, L, N_{eq}, N_{com} . После чего с помощью цикла методом k -средних был осуществлен перебор количества кластеров, описывающих дисперсию данных (от 2 до 8) [11, 18]. Визуализация критерия каменистой осыпи, описывающего зависимость суммы квадратов расстояний от точек до центра их ближайшего кластера от количества кластеров, представлена на рисунке 2.

Критерий каменистой осыпи позволяет графическим методом определить количество кластеров, которое будет соответствовать точке, где убывание

собственных значений замедляется наиболее сильно. Такая точка находится на уровне четырех кластеров. Таким образом, было определено, что в заданной структуре данных можно выделить четыре кластера, отличающихся между собой по каким-то (пока еще не установленным) факторам. Каждому цеху (строка в таблице 1) была присвоена метка класса, после чего проанализирована сходимость результатов по годам. Только один цех за три года был отнесен к разным классам, что говорит об устойчивости модели. Распределение на кластеры было визуализировано с помощью иерархического дерева кластеризации – дендрограммы (рисунок 3).

Дендрограмма позволяет определить расстояние между кластерами как наибольшее расстояние между любыми двумя объектами в различных кластерах (т.е. «наиболее удаленными соседями») и визуализирует степень сгруппированности точек

данных (цеха по добыче нефти и газа по оси y) по мере изменения Евклидова расстояния, отложенного по оси x, являющегося мерой близости точек

в кластере [14]. Можно заметить, что четыре кластера определены на уровне значения 0,018 Евклидова расстояния.

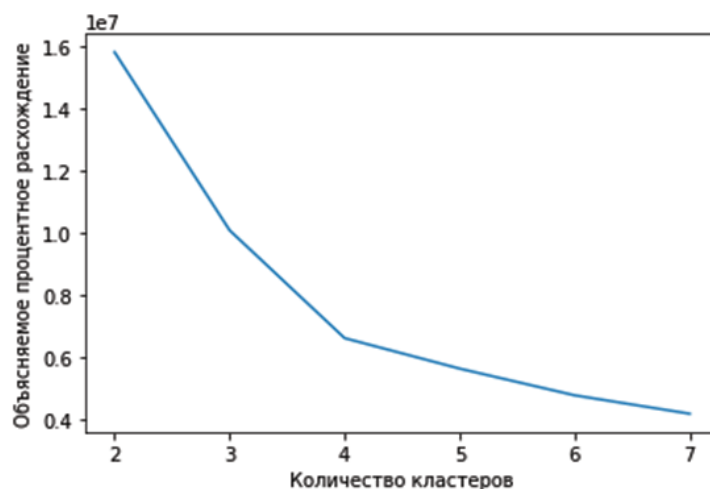


Рисунок 2. Определение количества кластеров методом каменной осыпи
 Источник: разработано автором

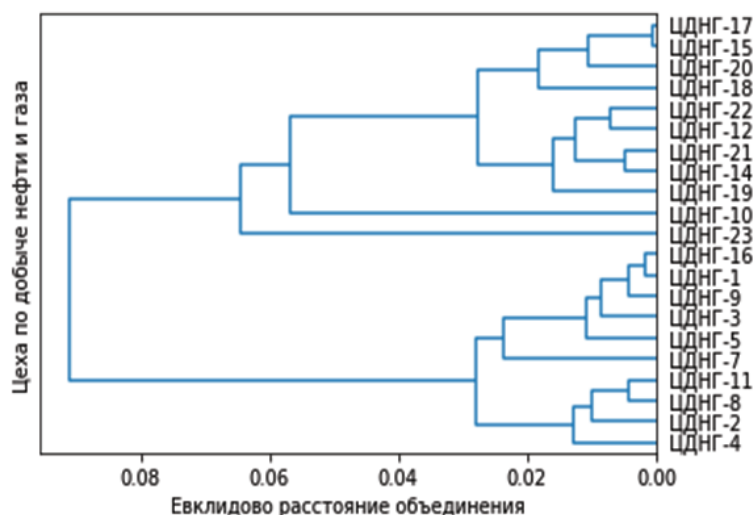


Рисунок 3. Горизонтальная дендрограмма результатов кластеризации цехов по добыче нефти и газа для набора данных 2016 года

Источник: разработано автором

Факторы, определяющие группирование на классы

Полученное распределение на кластеры говорит о наличии факторов, влияющих на структуру данных, и подтверждает гипотезу о возможности кластеризации цехов по добыче нефти и газа, однако при этом не отвечает на вопросы о сущности этих факторов. Для выделения влияющих на структуру данных факторов был использован метод главных компонент. Метод также позволяет понизить размерность многомерных данных и по-

лучить их визуальную структуру без существенной потери качества информации.

Перед помещением данных в модель PCA они были стандартизированы, после чего послужили основой для обучения модели с количеством главных компонент, равным двум. Следует отметить, что первая и вторая компоненты описывают около 63% всей дисперсии данных (3.93, 2.26 соответственно). Для улучшения интерпретации структуры данных они были визуализированы на рисунке 4.

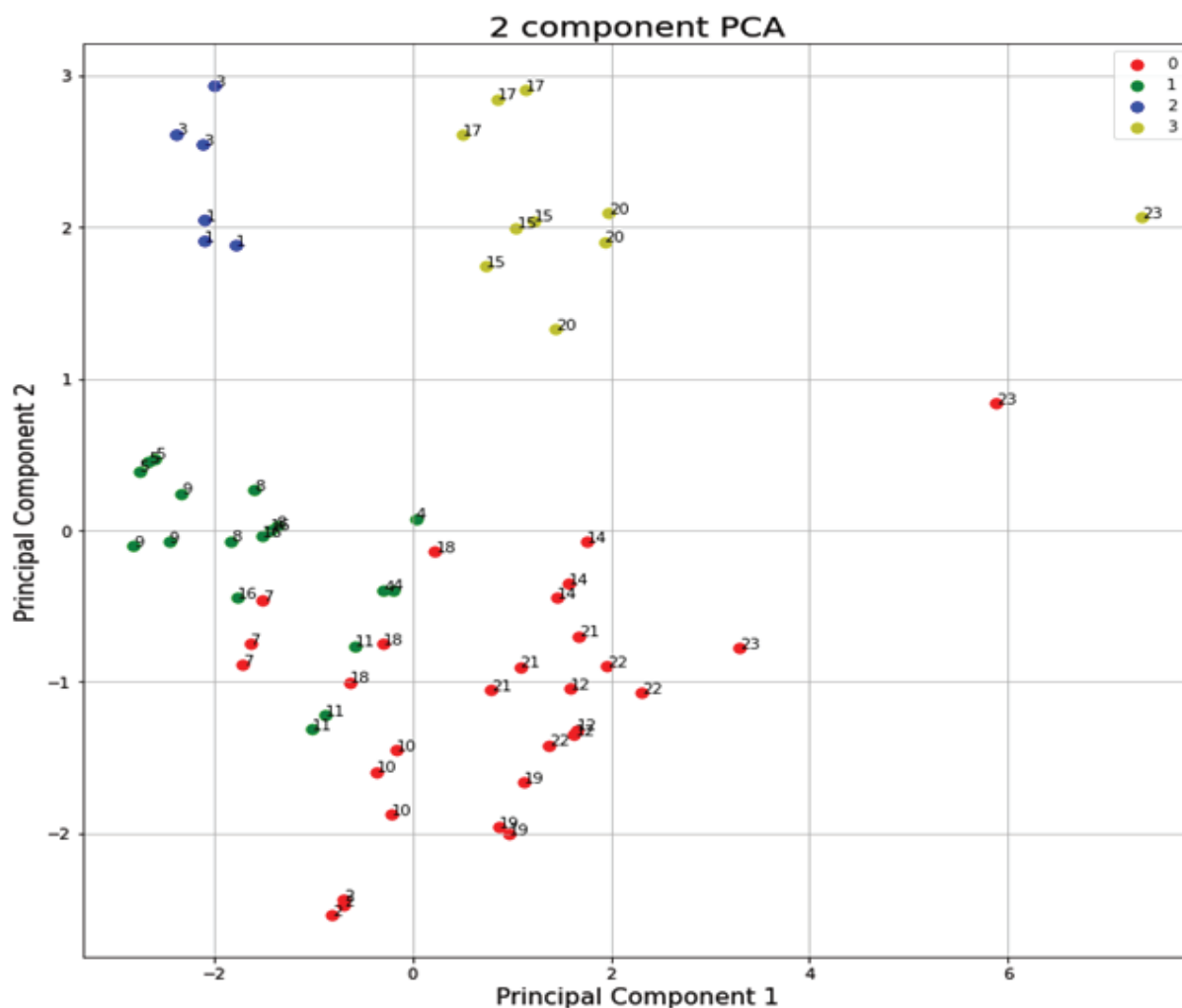


Рисунок 4. Визуализация кластеризации данных с использованием метода главных компонент
 Источник: разработано автором

В качестве осей координат использованы значения главных компонент (1 и 2 соответственно). В этих координатах были представлены цеха по добыче нефти и газа (каждый кластер от 0 до 3-го показан разным цветом) и метки цехов.

На графике наблюдается довольно четкое разделение на четыре кластера – синие точки слева вверху (самый малый по численности), желтые точки вверху по центру (второй по возрастанию кластер), зеленые точки слева по центру (средний по числен-

ности кластер), красные точки в центральной части графика (самый многочисленный кластер). Количество кластеров, визуально наблюдаемое на графике рисунка 4, совпадает с количеством полученных методом *k*-средних.

Кроме того, отнесение цехов по добыче нефти и газа к кластерам, полученных методом *PCA*, совпадает с кластерами, определенными методом *k-means* (Таблица 2).

Таблица 2. Распределение цехов на кластеры, полученное методом *k-means* и методом *PCA*

Кластеры	Цеха, отнесенные к кластерам	
	<i>k-means</i>	<i>PCA</i>
1	1, 3	1, 3
2	15, 17, 20, 23	15, 17, 20
3	4, 5, 8, 9, 11, 16	4, 5, 8, 9, 11, 16
4	2, 7, 10, 12, 14, 18, 19, 21, 22	2, 7, 10, 14, 18, 19, 21, 22, 23

Источник: разработано автором

Для интерпретации значений главных компонент и выявления их взаимосвязи со значениями

факторов была построена таблица математической корреляции двух компонент и факторов (Таблица 3).

Таблица 3. Вычисление корреляции 11 исходных переменных с двумя скрытыми переменными, полученными методом PCA

№	T_f ед.	A_f ед.	N_{cl} шт.	N_e чел.	N_h ед.	N_s ед.	N_{ss} ед.	N_p ед.	N_{nw} ед.	D, м ³ /час	N_{rep} ед.
1	-0.31	0.14	-0.37	0.39	0.37	0.03	0.14	0.36	0.40	-0.24	0.28
2	0.45	0.55	0.30	0.27	0.07	0.08	-0.30	-0.23	0.20	0.09	0.37

Источник: разработано автором

Таблица корреляции позволяет определить, можно ли интерпретировать и назвать скрытые переменные на основании их отношений с оригиналами. Скрытые переменные (компоненты) можно выявить по наличию корреляционной взаимосвязи с действительными переменными. Первая главная компонента, отражающая примерно 40% общей информации в наборе, связана с осложненностью фонда скважин (корреляция с гидратами $N_h = 0,37$, парафинами $N_p = 0,36$). Вторая главная компонента, отражающая 23% информации в данных, связана с фондом скважин (корреляция с действующим $A_f = 0,55$ и общим фондом скважин $T_f = 0,45$).

Новый набор данных, закодированный двумя скрытыми переменными, был проверен на эффективность прогнозирования распределения цехов на группы с помощью наивного байесовского классификатора после перекодирования исходного набора данных. Модель показала высокую точность прогнозирования и допустимость использования для кластеризации цехов по добыче нефти и газа.

Результаты и выводы

Таким образом, с помощью метода кластерного анализа исходный набор данных может быть сгруппирован по четырем кластерам. Визуализация исходных данных методом главных компонент подтверждает наличие четырех групп точек данных. С учетом полученных главных компонент были определены границы производственных показателей кластеров. На дисперсию данных влияют две главные компоненты, первая из которых связана с осложненностью фонда скважин, т.е. со сложностью извлечения углеводородов и потребностью в дополнительных технологических операциях. Вторая компонента связана с фондом скважин в цехе, что также характеризует увеличение объема работ по добыче углеводородов. Первую группу цехов можно охарактеризовать как цеха с большим общим фондом скважин (здесь и далее – средние значения переменных, относящихся к цехам в классе: 2698),

большим количеством неиспользованных скважин (975 активных скважин) и низкой осложненностью фонда скважин (0 скважин с гидратами; 14 скважин с парафинами). К ним относятся 1 и 3 цеха. Это, как правило, старые цеха, находящиеся недалеко от инфраструктурных центров и значительно выработавшие свой ресурс. Среди других групп транспортное обеспечение этих цехов наиболее прогнозируемо и носит характер, близкий к стационарному.

Второй кластер, к которому относятся 15, 17 и 20 цеха, можно назвать как кластер с малым общим фондом скважин (1385), малым количеством неиспользованных скважин (1147 активных скважин) и низкой осложненностью фонда скважин (8 гидратов; 10 парафинов). Это относительно молодые цеха с большим остаточным ресурсом, в которых практически все скважины являются активными и добывающими. В них идет активное бурение, но при этом фонд довольно беспроblemен. Они требуют большое количество техники, но вследствие низкой осложненности не имеют резких и непрогнозируемых скачков заявок на транспорт.

Третий кластер, к которому относятся 4, 5, 8, 9, 11, 16 цеха, можно назвать кластером с малым общим фондом скважин (1385), большим количеством неиспользованных скважин (593 активных скважины) и низкой осложненностью фонда скважин (0 гидратов; 15 парафинов).

Четвертый кластер характеризуется малым фондом скважин (881), почти все из которых являются добывающими (602 акт. скважины), активной добычей, однако фонд скважин значительно осложнен (31 гидратов; 65 парафинов), что затрудняет добычу и увеличивает трудоемкость и объемы транспортно-технологического сервиса. К нему можно отнести 2, 7, 10, 14, 18, 19, 21, 22, 23 цеха. Такие цеха, как правило, имеют нестационарный поток заявок на технику, характеризуются большим количеством внеплановых операций с привлечением специальных автомобилей и наиболее сложно прогнозируемы с точки зрения нормирования транспорта.

Таким образом, установленные с помощью метода главных компонент производственно-технологические факторы позволят установить закономерности влияния производственных показателей цехов по добыче нефти и газа на потребность в транспортно-технологических машинах и осуществлять на этой основе формирование нормативов технологического транспорта правильно осуществлять в привязке к классификации цехов.

Заключение

В работе рассмотрен подход к нормированию потребности в технике для цехов по добыче нефти и газа с учетом распределения этих цехов на группы, связанные с разностью производственных их показателей. Были определены факторы, определяющие группирование цехов на четыре кластера и являющиеся объектом научной новизны настоящего исследования. К этим факторам можно отнести фонд скважин в цехе и осложненность этого фонда скважин, т. е. наличие примесей, усложняю-

щих процедуру извлечения углеводородов и увеличивающих объем сервисных работ над кустовой площадкой. Потребность в технике предлагается определять для каждой группы цехов по отдельности, что позволит «привязать» количество транспортно-технологических машин не к исторически сложившимся нормам в цехе, а к реальным показателям каждого. Такой подход позволит повысить точность прогнозирования потребности в технике и сократить простои основного производства по причине нехватки техники. Проведенные исследования позволили подтвердить гипотезу о возможности кластеризации цехов на группы, определяемые рядом производственно-технологических факторов, а также установить эти факторы. Целью дальнейших исследований является разработка норматива транспортно-технологических машин на производственный показатель цехов по добыче нефти и газа, а также формирование системы поддержки принятия решений по управлению парком техники в рамках предприятия.

Литература

1. Бражникова С. В. Методические подходы к оценке эффективности производственных процессов при реализации грузовых перевозок // Корпоративное управление экономической и финансовой деятельностью на железнодорожном транспорте: сборник научных трудов. – М.: МИИТ. – 2010. – № 8. – С. 45–51.
2. Кузнецов В. В., Гайворонская Н. Ф. Нормативное прогнозирование потребности в сельхозтехнике для растениеводства // Научное обозрение: теория и практика. – 2016. – № 12. – С. 50–60.
3. Макуев В. А. Частные случаи задачи по оптимальному расчету парка лесосечных машин // Вестник Московского государственного университета леса – Лесной вестник. – 2005. – № 2. – С. 125.
4. Милюткин В. А., Соловьёв С. А., Макаровская З. В. Оптимизация машинно-тракторного парка агропредприятия при выборе сельхозмашин (сеялок) по основным технико-технологическим показателям // Известия Оренбургского государственного аграрного университета. – 2017. – № 4 (66). – С. 122–124.
5. Определение потребности в транспортных средствах и специальной технике для транспортных подразделений нефтепроводной отрасли / В. И. Бауэр [и др.] // Инженерный вестник Дона. – 2015. – № 1–1(33). – С. 46.
6. Репетов А. Н. Основные направления повышения эффективности использования техники // Вестник Курской Государственной Сельскохозяйственной Академии. – 2008. – № 3. – С. 35–37.
7. Aguilar-Dominguez D. et al. (2021) Machine learning approach for electric vehicle availability forecast to provide vehicle-to-home. *5th Annual Conference of the Centre-for-Doctoral-Training (CDT) of the UK-Engineering-and-Physical-Sciences-Research-Council (EPSRC) on Energy Storage and its Applications (ESA)*. Vol. 7, pp. 71–80. DOI: 10.1016/j.egy.2021.02.053 (In Eng.).
8. Al-Tarawneh M., Huang Y. (2019) Road vehicle classification using machine learning techniques. *Sensors And Smart Structures Technologies For Civil, Mechanical, And Aerospace Systems 2019*. Vol. 10970. DOI: 10.1117/12.2514320 (In Eng.).
9. Arröspide J., Salgado L. (2012) Region-dependent vehicle classification using PCA features. *Proceedings – International Conference on Image Processing, ICIP. Lake Buena Vista, FL, SEP 30-OCT 03, 2012*. Pp. 453–456. DOI: 10.1109/ICIP.2012.6466894 (In Eng.).
10. Esenturk E. et al. (2022) Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Development for Autonomous Vehicles. *IEEE Access*. Vol. 10, pp. 6660–6675. DOI: 10.1109/ACCESS.2021.3140052 (In Eng.).
11. Esteves R., Hacker T., Rong C. (2013) Competitive K-means: A new accurate and distributed K-means algorithm for large datasets. *2013 IEEE Fifth International Conference On Cloud Computing Technology And Science (Cloudcom)*. Vol. 1, pp. 17–24. DOI: 10.1109/CloudCom.2013.89 (In Eng.).
12. Fang X. (2013) Inference-based naïve bayes: Turning naïve bayes cost-sensitive. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 25. No 10, pp. 2302–2313. DOI: 10.1109/TKDE.2012.196 (In Eng.).

13. Forrester P., Zhang J. (2020) Parametrising correlation matrices. *Journal of Multivariate Analysis*. Vol. 178, 104619. DOI: 10.1016/j.jmva.2020.104619 (In Eng.).
14. Huang H. et al. (2016) Mining arbitrary shaped clusters and outputting a high quality dendrogram. *Database And Expert Systems Applications, DEXA 2016, PT I*. Vol. 9827, pp. 153–168. DOI: 10.1007/978-3-319-44403-1_10 (In Eng.).
15. Jaffar F. et al. (2020) Prediction of Drag Force on Vehicles in a Platoon Configuration Using Machine Learning. *IEEE Access*. Vol. 8, pp. 201823–201834. DOI: 10.1109/ACCESS.2020.3035318 (In Eng.).
16. Kadiyala A., Kumar A. (2017). Applications of Python to evaluate environmental data science problems. *Environmental Progress and Sustainable Energy*. Vol.36. No 6, pp. 1580–1586. DOI: 10.1002/ep.12786 (In Eng.)
17. Li Z. C., Wu Q. Y., Yang H. (2019) A theory of auto ownership rationing. *Transportation Research Part B-Methodological*. Vol. 127, pp. 125–146. DOI: 10.1016/j.trb.2019.07.008 (In Eng.).
18. Liu Y., Li B. (2020) Bayesian hierarchical K-means clustering. *Intelligent Data Analysis*. Vol. 24. Is. 5, pp. 977–992. DOI: 10.3233/IDA-194807 (In Eng.).
19. Lu C. C. et al. (2021) Optimal fleet deployment for electric vehicle sharing systems with the consideration of demand uncertainty. *Computers & Operations Research*. Vol. 135., 105437. DOI: 10.1016/j.cor.2021.105437 (In Eng.).
20. Melnikov A. N., Lyubimova I. I., Manayev K. I. (2016) Improvement of the vehicles fleet structure of a specialized motor transport enterprise. *Procedia Engineering*. Vol. 150. – P. 1200–1208. DOI:10.1016/j.proeng.2016.07.236.
21. Peng Y. et al. (2012) Vehicle type classification using PCA with self-clustering. *IEEE International Conference on Multimedia and Expo Workshops, Melbourne, Australia, JUL 09-13, 2012*. Pp. 384–389. DOI: 10.1109/ICMEW.2012.73 (In Eng.).
22. Rabbouch H., Saâdaoui F., Mraïhi R. (2017) Unsupervised video summarization using cluster analysis for automatic vehicles counting and recognizing. *Neurocomputing*. Vol. 260. pp. 157–173. DOI: 10.1016/j.neucom.2017.04.026 (In Eng.).
23. Ren M. et al. (2018) A Stochastic Model for Vehicle Clustering Performance Analysis. *IEEE International Conference on Communications, Proceedings Paper: Kansas City, MO MAY 20–24, 2018*. DOI: 10.1109/ICC.2018.8422606 (In Eng.).
24. Tan L. H. (2001) Rationing rules and outcomes: The experience of Singapore’s vehicle quota system. *IMF Working Papers*. Vol. 136, A001. DOI: 10.5089/9781451855814.001 (In Eng.).
25. Zhang D. et al. (2012) Study on military vehicle equipment grading method based on clustering analysis and fisher discriminant analysis. *IEEE International Conference on Automation and Logistics, ICAL*. Pp. 127–130. DOI: 10.1109/ICAL.2012.6308183 (In Eng.).
26. Zhang H., Su J. (2008) Naive Bayes for optimal ranking. *Journal of Experimental and Theoretical Artificial Intelligence*. Vol.20. No 2, pp. 79–93. DOI: 10.1080/09528130701476391 (In Eng.).

References

1. Brazhnikova, S. V. (2010) [Methodological approaches to assessing the effectiveness of production processes in the implementation of freight transportation]. *Korporativnoe upravlenie ekonomicheskoy i finansovoy deyatel'nosti na zheleznodorozhnom transporte: sbornik nauchnykh trudov* [Corporate management of economic and financial activities in railway transport: a collection of scientific papers]. Vol. 8, pp. 45–51. (In Russ.).
2. Kuznecov, V. V., Gajvoronskaya, N. F. (2016) [Normative forecasting of the need for agricultural machinery for crop production]. *Nauchnoe obozrenie: teoriya i praktika* [Scientific Review: Theory and Practice]. Vol. 12, pp. 50–60. (In Russ.).
3. Makuev, V. A. (2005) [Special cases of the problem of optimal calculation of the fleet of logging machines]. *Vestnik Moskovskogo gosudarstvennogo universiteta lesa – lesnoj vestnik* [Bulletin of the Moscow State Forest University – Forest Bulletin]. Vol. 2, pp. 125. (In Russ.).
4. Milyutkin, V. A., Solovyov, S. A., Makarovskaya, Z. V. (2017) [Optimization of the machine and tractor fleet of an agricultural enterprise when choosing agricultural machines (seeders) according to the main technical and technological indicators]. *Izvestiya Orenburgskogo gosudarstvennogo agrarnogo universiteta* [Proceedings of the Orenburg State Agrarian University]. Vol. 66, No. 4, pp. 122–124. (In Russ.).
5. Bauer, V. I. (2015) [Determining the requirement of vehicles and special techniques for transport units in the pipeline industry]. *Inzheneriy vestnik Dona* [Engineering journal of Don]. Vol. 33, No 1–1, pp. 46–52. (In Russ.).
6. Repetov, A. N. (2008) [The main directions of increasing the efficiency of use of technology]. *Vestnik Kurskoj Gosudarstvennoj Selskohozyajstvennoj Akademii* [Bulletin of the Kursk State Agricultural Academy]. Vol. 3, pp. 35–37. (In Russ.).

7. Aguilar-Dominguez, D. et al. (2021) Machine learning approach for electric vehicle availability forecast to provide vehicle-to-home. *5th Annual Conference of the Centre-for-Doctoral-Training (CDT) of the UK-Engineering-and-Physical-Sciences-Research-Council (EPSRC) on Energy Storage-and-its Applications (ESA)*. Vol. 7, pp. 71–80. DOI: 10.1016/j.egy.2021.02.053 (In Eng.).
8. Al-Tarawneh, M., Huang, Y. (2019) Road vehicle classification using machine learning techniques. *Sensors And Smart Structures Technologies For Civil, Mechanical, And Aerospace Systems 2019*. Vol. 10970. DOI: 10.1117/12.2514320 (In Eng.).
9. Arröspide, J., Salgado, L. (2012) Region-dependent vehicle classification using PCA features. *Proceedings – International Conference on Image Processing, ICIP*. Lake Buena Vista, FL, SEP 30–OCT 03, 2012. Pp. 453–456. DOI: 10.1109/ICIP.2012.6466894 (In Eng.).
10. Esenturk, E. et al. (2022) Identification of Traffic Accident Patterns via Cluster Analysis and Test Scenario Development for Autonomous Vehicles. *IEEE Access*. Vol. 10, pp. 6660–6675. DOI: 10.1109/ACCESS.2021.3140052 (In Eng.).
11. Esteves, R., Hacker, T., Rong, C. (2013) Competitive K-means: A new accurate and distributed K-means algorithm for large datasets. *2013 IEEE Fifth International Conference On Cloud Computing Technology And Science (Cloudcom)*. Vol. 1, pp. 17–24. DOI: 10.1109/CloudCom.2013.89 (In Eng.).
12. Fang, X. (2013) Inference-based naïve bayes: Turning naïve bayes cost-sensitive. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 25, No 10, pp. 2302–2313. DOI: 10.1109/TKDE.2012.196 (In Eng.).
13. Forrester, P., Zhang, J. (2020) Parametrising correlation matrices. *Journal of Multivariate Analysis*. Vol. 178, pp. 104619. DOI: 10.1016/j.jmva.2020.104619 (In Eng.).
14. Huang, H. et al. (2016) Mining arbitrary shaped clusters and outputting a high quality dendrogram. *Database And Expert Systems Applications, DEXA 2016, PT I*. Vol. 9827, pp. 153–168. DOI: 10.1007/978-3-319-44403-1_10 (In Eng.).
15. Jaffar, F. et al. (2020) Prediction of Drag Force on Vehicles in a Platoon Configuration Using Machine Learning. *IEEE Access*. Vol. 8, pp. 201823–201834. DOI: 10.1109/ACCESS.2020.3035318 (In Eng.).
16. Kadiyala, A., Kumar, A. (2017). Applications of Python to evaluate environmental data science problems. *Environmental Progress and Sustainable Energy*. Vol.36, No 6, pp. 1580–1586. DOI: 10.1002/ep.12786 (In Eng.).
17. Li, Z. C., Wu, Q. Y., Yang, H. (2019) A theory of auto ownership rationing. *Transportation Research Part B-Methodological*. Vol. 127, pp. 125–146. DOI: 10.1016/j.trb.2019.07.008 (In Eng.).
18. Liu, Y., Li, B. (2020) Bayesian hierarchical K-means clustering. *Intelligent Data Analysis*. Vol. 24. Is. 5, pp. 977–992. DOI: 10.3233/IDA-194807 (In Eng.).
19. Lu, C. C. et al. (2021) Optimal fleet deployment for electric vehicle sharing systems with the consideration of demand uncertainty. *Computers & Operations Research*. Vol. 135., 105437. DOI: 10.1016/j.cor.2021.105437 (In Eng.).
20. Melnikov, A. N., Lyubimova, I. I., Manayev, K. I. (2016) Improvement of the vehicles fleet structure of a specialized motor transport enterprise. *Procedia Engineering*. Vol. 150, pp. 1200–1208. DOI:10.1016/j.proeng.2016.07.236 (In Eng.).
21. Peng, Y. et al. (2012) Vehicle type classification using PCA with self-clustering. *IEEE International Conference on Multimedia and Expo Workshops*, Melbourne, Australia, JUL 09–13, pp. 384–389. DOI: 10.1109/ICMEW.2012.73 (In Eng.).
22. Rabbouch, H., Saâdaoui, F., Mraïhi, R. (2017) Unsupervised video summarization using cluster analysis for automatic vehicles counting and recognizing. *Neurocomputing*. Vol. 260, pp. 157–173. DOI: 10.1016/j.neucom.2017.04.026 (In Eng.).
23. Ren, M. et al. (2018) A Stochastic Model for Vehicle Clustering Performance Analysis. *IEEE International Conference on Communications, Proceedings Paper*. Kansas City, MO MAY 20–24, 2018. DOI: 10.1109/ICC.2018.8422606 (In Eng.).
24. Tan, L. H. (2001) Rationing rules and outcomes: The experience of Singapore’s vehicle quota system. *IMF Working Papers*. Vol. 136, A001. DOI: 10.5089/9781451855814.001 (In Eng.).
25. Zhang, D. et al. (2012) Study on military vehicle equipment grading method based on clustering analysis and fisher discriminant analysis. *IEEE International Conference on Automation and Logistics, ICAL*, pp. 127–130. DOI: 10.1109/ICAL.2012.6308183 (In Eng.).
26. Zhang, H., Su, J. (2008) Naive Bayes for optimal ranking. *Journal of Experimental and Theoretical Artificial Intelligence*. Vol.20. No 2, pp. 79–93. DOI: 10.1080/09528130701476391 (In Eng.).

Информация об авторе:

Евгений Сергеевич Козин, кандидат технических наук, доцент, доцент кафедры сервиса автомобилей и технологических машин, Тюменский индустриальный университет, Тюмень, Россия

ORCID ID: 0000-0002-6774-3285, **Researcher ID:** D-8474-2019-id, **Scopus Author ID:** 57052768700
e-mail: kozines@tyuiu.ru

Статья поступила в редакцию: 28.03.2022; принята в печать: 15.06.2022.
Автор прочитал и одобрил окончательный вариант рукописи.

Information about the author:

Evgeny Sergeevich Kozin, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Car Service and Technological Machines, Industrial University of Tyumen, Tyumen, Russia
ORCID ID: 0000-0002-6774-3285, **Researcher ID:** D-8474-2019-id, **Scopus Author ID:** 57052768700
e-mail: kozines@tyuiu.ru

The paper was submitted: 28.03.2022.
Accepted for publication: 15.06.2022.
The author has read and approved the final manuscript.