

Д. Сапонов

ПРОГНОСТИЧЕСКАЯ МОДЕЛЬ ПОВЕДЕНИЯ ПРИ ВЫБОРЕ СТРАТЕГИИ ПЕНСИОННЫХ НАКОПЛЕНИЙ: ВЫБОР АЛГОРИТМА И ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТА

ПРОГНОСТИЧЕСКАЯ МОДЕЛЬ ПОВЕДЕНИЯ ПРИ ВЫБОРЕ СТРАТЕГИИ ПЕНСИОННЫХ НАКОПЛЕНИЙ: выбор алгоритма и визуализация результата

PROGNOSTIC MODEL WHEN CHOOSING PENSION SAVING STRATEGIES: selection algorithm and visualization of the results

САПОНОВ Дмитрий Игоревич - научный сотрудник лаборатории методологии федеративных исследований ИНСАП РАНХиГС при Президенте РФ. E-mail: dsaponov1@yandex.ru

SAPONOV Dmitrii Igorevich - Researcher, Laboratory for Federative Research Methodology, Russian Presidential Academy for National Economy and Public Administration. E-mail: dsaponov1@yandex.ru

Аннотация. Статья посвящена разработке прогностической модели поведения на материале опроса людей трудоспособного возраста: (мужчины 18-60 лет, женщины 18-55 лет).

Показано, что для моделирования нелинейной зависимости доли тех, кто перевел накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный от возраста хорошо подходит алгоритм CHAID. Выделено пять возрастных групп, различающихся по однородности и внутренней структуре с точки зрения пенсионного поведения. Особое внимание уделено визуализации полученной модели с помощью анализа соответствий.

Abstract. The article is devoted to the elaboration of prognostic model of behavior; the study is based on the survey involving people of working age (men aged 18-60, women aged 18-55).

It is shown that the CHAID algorithm is good for modeling non-linear dependence on the age for those who transferred the funded part of the pension from the state fund to the non-state fund on their age. Five age groups showing different retirement behavior were singled out. A special attention was paid to the visualization of the model obtained by correspondence analysis.

Ключевые слова: пенсионное обеспечение, алгоритм CHAID, анализ соответствий, нелинейная зависимость, визуализация результатов классификационной модели

Keywords: pension provision, CHAID algorithm, correspondence analysis, non-linear dependence, visualization of results of classification model

Центр методологии федеративных исследований РАНХиГС при Президенте РФ в ноябре 2012 г. провел общероссийский телефонный опрос людей трудоспособного возраста: (мужчины 18–60 лет, женщины 18–55 лет)¹. В ходе исследования получена информация об

¹ Проект «Прогнозный анализ и сценарное моделирование организационно-экономического механизма повышения эффективности накопительной составляющей пенсионного обеспечения в РФ» выполнен по заказу Института Гайдара.

ожиданиях, связанных с пенсионным обеспечением и доходами после выхода на пенсию, и данные о фактическом поведении в части реализации пенсионных стратегий.

В списке показателей, которые составляют опросный инструмент, особую роль занимали признаки, явным образом указывающие на поведение респондента, связанное с будущей пенсией. Эти признаки, с одной стороны, не являются общими социально-демографическими характеристиками респондента, в отличие от таких показателей, как «Пользование интернетом» или «Наличие действующего загранпаспорта», которые указывают на тип поведения в повседневных практиках через соотнесение с определенными стандартами потребления. С другой стороны, они не являются выражением субъективного смысла, мнения и установок относительно будущей пенсии, потому что указывают не на мнение, а на конкретные решения, которые респондент принял или собирается принять.

В опросном инструменте исследования два таких признака. Первый, более слабый, реализован в вопросе: «А готовы ли вы выйти на пенсию на 5 лет позже, чтобы потом получать пенсию в 2 раза больше?». Второй, более сильный, измерялся вопросом: «Переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный (из ПФР в НПФ)?» Первый признак определен нами как более слабый по следующим причинам: 1) он проспективен (связан с еще не совершенным действием); 2) он моделирует вымышленную ситуацию, так как значения «на 5 лет» и «в 2 раза» являются условными; 3) значение этого показателя нельзя проверить, сверив с федеральной статистикой, за неимением таковой.

Второй признак, определенный нами как более сильный, отличается от первого по всем трем указанным выше параметрам: 1) он задается по уже совершенному действию, т.е. ретроспективен; 2) он связан с реальным действием; 3) значение показателя можно сравнить с данными Федеральной статистики. По данным Минтруда, большая часть населения остается «молчунами» — сейчас свои средства в НПФ перевели только 20 млн человек из 70 млн работающих, что составляет 28% (www.rosmintrud.ru). Данное исследование дает более низкую, но и более близкую оценку. Если считать только долю от валидных ответов (все респонденты, принявшие участие в обследовании, за вычетом затруднившихся ответить на данный вопрос, и те, кто не слышал о существовании накопительной части пенсии), то доля тех, кто перевел свои средства в НПФ, составляет 21% (Таблица 1). Эта оценка автоматически относит к «молчунам» тех, кто затруднился ответить или никогда не слышал о существовании накопительной части пенсии, т.е. является нижней границей для доли тех, кто перевел свои средства в НПФ.

Таблица 1 Переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный (из ПФР в НПФ)?

		Человек	%	Валидный %
Валидные	Да	357	17,9	21,0
	Нет	1345	67,3	79,0
	Итого	1703	85,1	100,0
Пропущенные	Затрудняюсь ответить	53	2,7	
	Вопрос не применим (Никогда не слышали о том, что в России существует обязательная накопительная часть пенсии)	244	12,2	
	Итого	297	14,9	
Итого		2000	100,0	

Оба показателя мы отнесли к группе поведенческих признаков. Традиционно подобные признаки, содержательно определяющие поведение в исследуемой сфере, становятся целевыми (зависимыми) переменными в прогностических моделях. Как правило, поведение (совершение действия) кодируется бинарной переменной (да/нет). С точки зрения анализа данных, бинарная целевая переменная определяет набор моделей, которые могут быть использованы: логистическая регрессия, дискриминантный анализ, деревья классификации и т.д. Особенно полезными и востребованными подобные прогностические модели поведения становятся в ситуации, когда: 1) речь идет о растущей (имеющей входящий поток) во времени совокупности объектов (людей), при этом входящие объекты требуют классификации на основе данных о поведении уже вошедших ранее объектов с последующим принятием решения по каждому объекту на основе результатов классификации (выдать кредит или отказать, положить входящее письмо в папку «спам» или нет и т.д.). В этом случае модель увеличивает долю правильных решений; 2) в стабильной совокупности требуется выявить группы с наибольшей (наименьшей) вероятностью определенного поведения с тем, чтобы сделать воздействие по стимулированию (предотвращению) определенного поведения более эффективным, адресным.

Если говорить, например, о компании по привлечению граждан в НПФ, мы имеем дело со второй ситуацией. Итак, у нас есть «слабый» («А готовы ли вы выйти на пенсию на 5 лет позже, чтобы потом получать пенсию в два раза больше?» — Да/Нет) и «сильный» показатели поведения в сфере пенсионирования («Переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный (из ПФР в НПФ)?» — Да/Нет). В настоящей работе мы проанализируем прогностические модели, построенные с этими двумя показателями, взятыми в качестве целевых (зависимых) переменных.

Блок независимых переменных, которые тестировались на предмет включения в модель, приведен в таблице 2. В первой колонке отмечены переменные, вошедшие в окончательную модель².

Таблица 2 Блок независимых переменных, которые тестировались на предмет включения в модель

	Показатель	Статистическая значимость связи с целевой переменной
*	Скажите, пожалуйста, сколько вам полных лет?	0,000
X	Федеральный округ (со слов респондента)	0,000
	Какая из оценок наиболее точно характеризует материальное положение вашей семьи?	0,019
	Приходилось ли вам одновременно работать больше, чем на одной работе за последние три года?	0,112
X	А помогают ли вам близкие родственники деньгами?	0,138

² Можно заметить, что показатель «Есть ли у вас действующий загранпаспорт?» включен в окончательную модель, хотя значимость связи с целевой переменной у него больше, чем у других переменных, которые не вошли в окончательную модель. Это связано с тем, что модель учитывает взаимодействия независимых переменных (влияние одной переменной в группах, образованных разбиением по другой переменной), в отличие от статистической значимости связи с целевой переменной.

	Скажите, пожалуйста, помогаете ли вы близким родственникам деньгами?	0,160
	Пол респондента	0,165
X	Пользуетесь ли вы интернетом? Если да, то каждый день, несколько раз в неделю или несколько раз в месяц и реже?	0,179
X	Есть ли у вас действующий загранпаспорт?	0,257
	Вы когда-нибудь задумывались о том, чтобы открыть собственное дело, стать предпринимателем?	0,259
	За последние три года открывали ли вы депозит, срочный вклад в банке?	0,268
	Скажите, пожалуйста, какое у вас образование?	0,406
	Вы живете в региональном центре, другом городе или в сельской местности?	0,533
	Приходилось ли вам когда-нибудь сдавать комнату, квартиру или дом?	0,612
	В настоящее время вы работаете, учитесь, временно не работаете или на пенсии?	0,845
* В первом столбце отмечены переменные, которые вошли в окончательную модель		

Основной результат обработки демографического блока предикторов (независимых переменных) вполне ожидаем. Вместе с тем он поддается внятной интерпретации и является исходной точкой для дальнейшего усложнения и уточнения моделей. Среди демографических переменных «возраст» лучше всего предсказывает поведение в сфере пенсионирования (это относится в равной мере к обеим целевым переменным). Фактически, он определяет дистанцию относительно события «выход на пенсию», которая и формирует в значительной степени поведение. В самом общем виде сильная связь целевых поведенческих признаков с возрастом означает, что существуют возрастные группы (кластеры), соотношенные с принципиально различными пенсионными стратегиями, и в то же время внутренне однородные по поведению в сфере пенсионирования. Однако характер зависимости для двух целевых переменных совершенно различен. Для более слабого показателя зависимость от возраста практически возрастающая (Рисунок 1), т.е. с его увеличением доля тех, кто не готов выйти на пенсию на 5 лет позже, возрастает, достигая в группе «49 и старше» 56%.

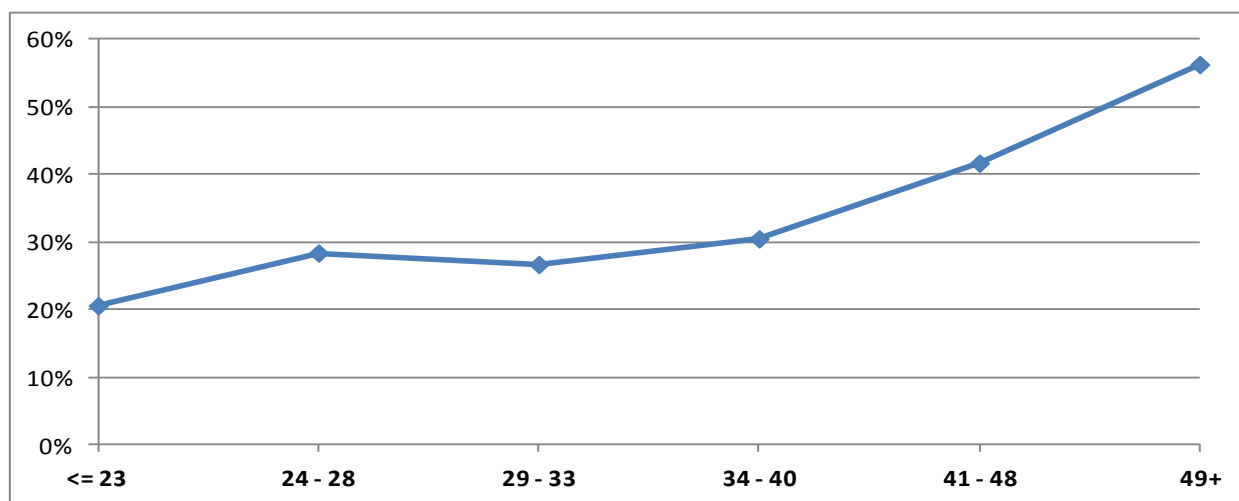


Рисунок 1 - Не готовы выйти на пенсию на 5 лет позже, чтобы потом получать пенсию в 2 раза больше.

Характер зависимости более сильного показателя от возраста совершенно другой, существенно нелинейный. Отчетливо выделяется группа среднего возраста (34–40 лет) с самым высоким значением показателя, где доля тех, кто перевел накопительную часть пенсии в НПФ доходит до 30% при средней доле 21%. В более младших и более старших группах наблюдается снижение показателя, до значений, существенно ниже средних (Рисунок 2).

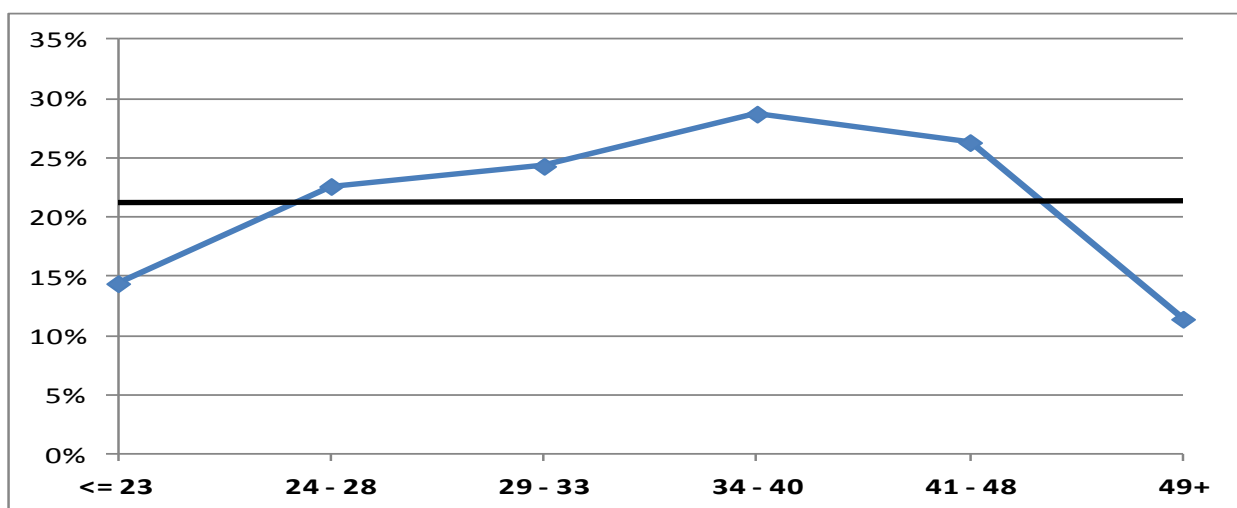


Рисунок 2 - Перевели накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный

Нелинейный характер зависимости определяет выбор метода для построения прогностической модели. Логистическая регрессия, не чувствительная к нелинейности, оценивает вклад возраста как незначимый ($p=0,056$) (Таблица 3).

Таблица 3 Оценка коэффициентов логистической регрессии с единственной независимой переменной «Возраст» и целевой переменной «Перевод накопительной части пенсии в НПФ»

	В	Стд. Ошибка	Вальд	ст.св.	Знч.	Exp(B)
Возраст	,010	,005	3,648	1	,056	1,010
Константа	,941	,208	20,373	1	,000	2,562

Таким образом, построение прогностической модели на основе логистической регрессии в нашей ситуации было бы ошибкой³. Далее, в части рассмотрения различных алгоритмов прогнозирования, мы ограничимся деревьями классификации, которые хорошо выявляют нелинейные зависимости и устойчивы к наличию выбросов и нетипичных значений в данных. Деревья классификации выделяют 5 возрастных групп таким образом, чтобы доли активных (тех, кто перевел накопительную часть в НПФ) максимально различались при переходе от одной группы к другой. Выделены следующие группы по возрасту: (18 лет – 22

³ Речь идет о включении возраста как независимой количественной переменной в логистическую регрессию. Если возраст категоризировать и включить в модель категории, то модель будет адекватной, хотя и более сложной.

года), (23 года — 33 года), (34 года — 41 год), (42 года — 49 лет), (более 49 лет). Такая группировка оптимальна с точки зрения соответствия различным стратегиям в сфере пенсионирования и выявляет полярные группы (рис. 3).

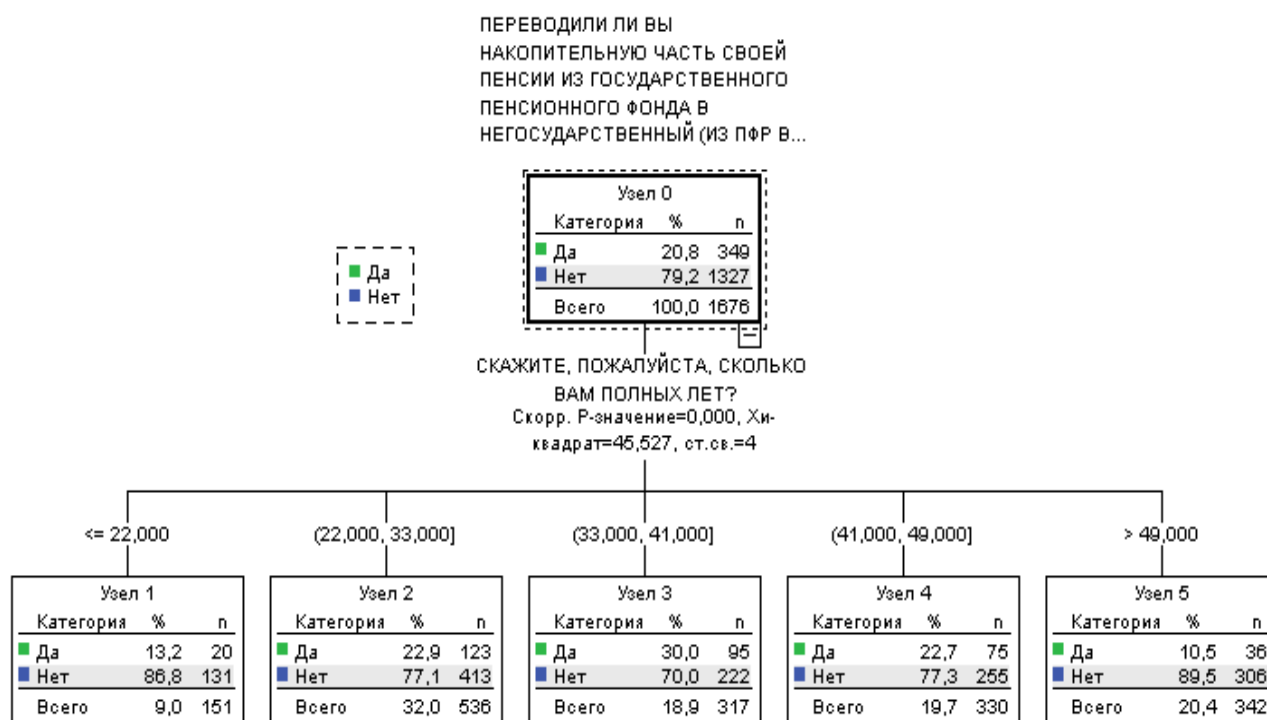


Рисунок 3 - Выделение возрастных групп, соответствующих разным пенсионным стратегиям на основе алгоритма CHAID

Посмотрим на полученную периодизацию с более общих позиций, поместив ее в контекст долгосрочного развития карьеры, начиная с обучения и заканчивая выходом на пенсию.

(18 лет — 22 года) — обучение в вузе; (23 года — 33 года) — первая фаза карьеры; (34 года — 41 год) — развитие карьеры; (42 года — 49 лет) — стабилизация карьеры. В центре этой группы находится возраст 45 лет, соответствующий 1967 году рождения. Дело в том, что сейчас обязательная госпенсия формируется за счет взносов в Пенсионный фонд, ставка которых составляет 22%. У родившихся в 1967 году и позже эта цифра делится следующим образом: 6% направляется на накопительную часть (средства инвестируются для будущих выплат); из остальных 16% (идут на выплаты нынешним пенсионерам), 10% учитывается на индивидуальном счете работника и формирует объем его будущих пенсионных прав, а 6% направляется в солидарную часть и на объем пенсионных прав не влияет. У родившихся в 1966 году и раньше накопительной части нет, в индивидуальную направляется 16% из 22%, в солидарную — те же 6%. У людей старше 49 лет — предпенсионный возраст. Таким образом, полноценными участниками программы накопительной части пенсии следует считать россиян, которым в 2012 году было не более 45 лет. Тот факт, что в группе предпенсионного возраста встречаются ответы о переводе накопительной части пенсии в НПФ, может означать, что перевод был осуществлен ранее 2004 г., в котором россияне старшего возраста были исключены из системы с накопительной частью.

На рисунке 3 изображено дерево классификации с одним уровнем разбиения по единственному предиктору — возраст. На следующем шаге, для уточнения модели, мы добавим еще один уровень разбиения, по-прежнему работая только с социально-демографическим блоком предикторов. Основные результаты по уточненной модели таковы:

- 1 Выделенные нами возрастные группы обнаруживают различную внутреннюю структуру при попытке дальнейшего разбиения, отделяющего «молчунов» от «активных». Во-первых, некоторые группы обнаруживают высокую степень однородности, не расщепляясь на подгруппы. Во-вторых, предикторы, наиболее сильно отделяющие «молчунов» от «активных», различаются для разных групп, тем самым указывая на специфику группы.
- 2 Крайние возрастные категории (моложе 22 и старше 49) не дифференцируются ни одной из демографических переменных, т.е. имеют очень высокую степень однородности в смысле пенсионных стратегий (перевод накопительной части в НПФ) (Рисунок 4, Рисунок 6).
- 3 Основным признаком в группах (23 года — 33 года) и (34 года — 41 год), который расщепляет их по доле передавших накопительную часть в НПФ, является территориальный критерий (федеральный округ) (Рисунок 4, Рисунок 5).
- 4 Основным признаком для группы (42 года — 49 лет), дифференцирующей группу по доле передавших накопительную часть в НПФ, является «Пользование интернетом» (Рисунок 6).

Рассмотрим специфику каждой группы подробнее.

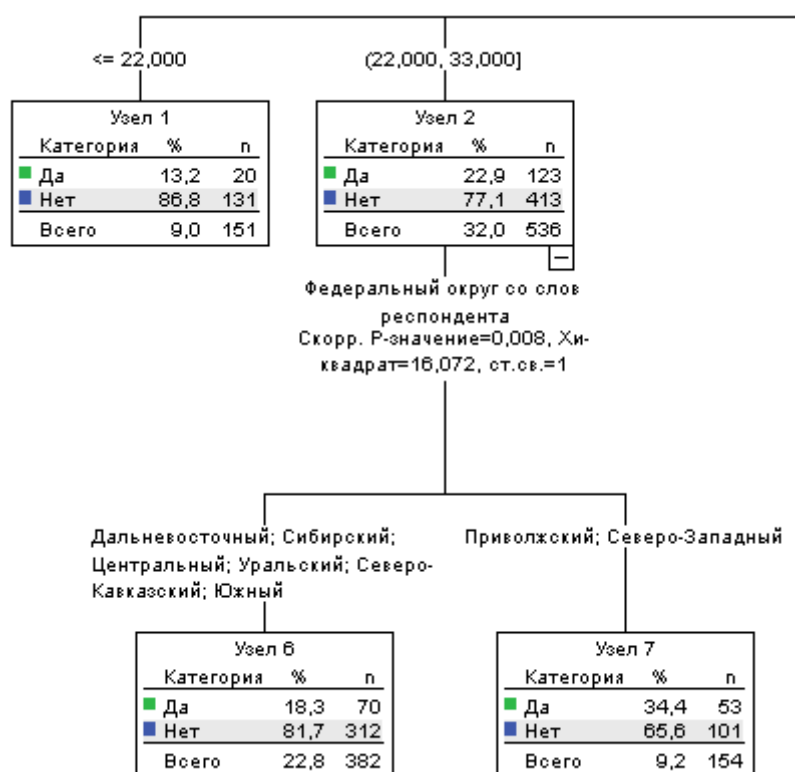


Рисунок 4 - Второй уровень разбиения для возрастных групп (18 лет – 22 года) и (23 года – 33 года)

Как уже было отмечено, группа (18 лет – 22 года) демонстрирует высокую степень однородности и не расщепляется демографическими переменными для дальнейшей поляризации (отделения «молчунов» от «активных»). В группе (23 года – 33 года), соответствующей началу карьеры, основной признак, выделяющий тех, кто перевел накопительную часть пенсии в НПФ – территориальный. В Приволжском и Северо-Западном округах доля активных в смысле пенсионного поведения россиян 34%. Отметим, что для данной возрастной группы Центральный и Северо-Западный федеральные округа входят в разные подгруппы.

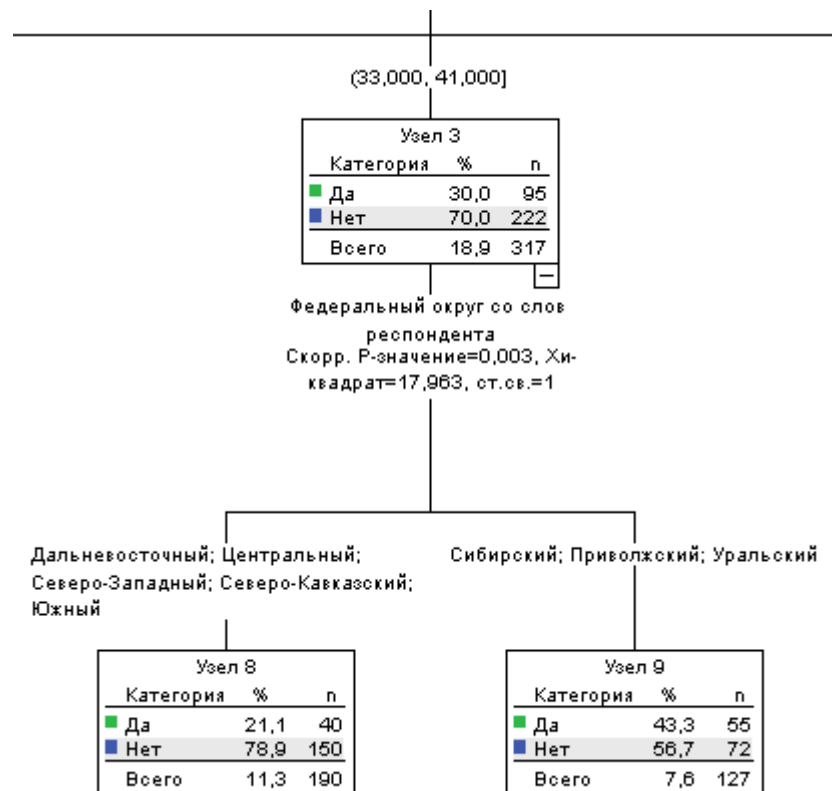


Рисунок 5 - Второй уровень разбиения для возрастной группы (34 года — 41 год)

Для возраста (34 года — 41 год) также характерна дифференциация пенсионного поведения по федеральным округам, однако комбинации округов, дающие наибольшую долю «активных» россиян, отличаются от группы (23 года — 33 года). На этот раз в нее входят не Приволжский и Северо-Западный федеральные округа, а Сибирский, Приволжский, Уральский. Северо-Западный и Центральный теперь входят в группу, дающую высокий процент «молчунов».

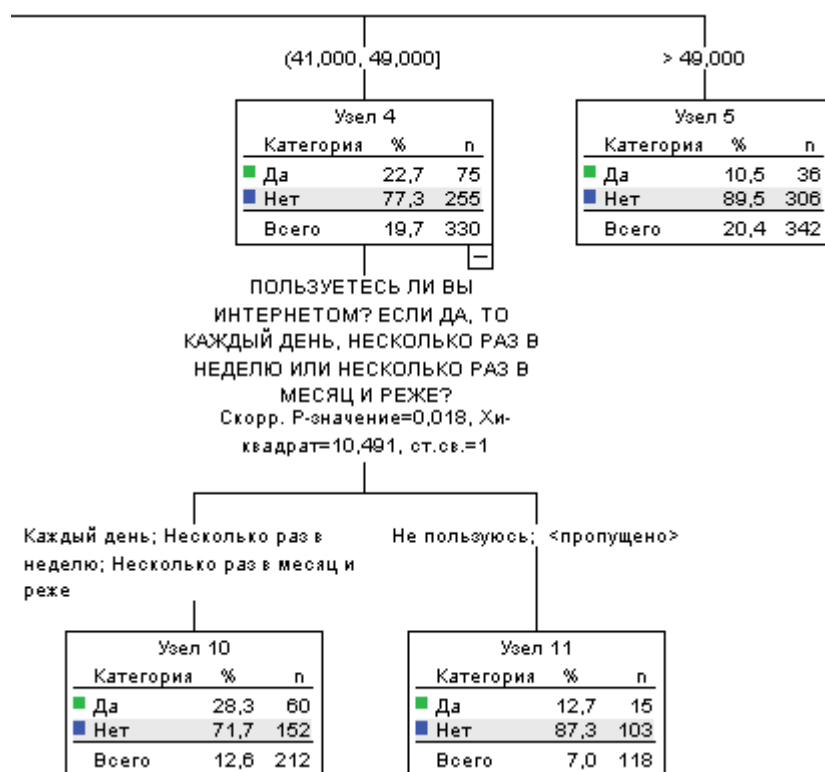


Рисунок 6 - Второй уровень разбиения для возрастных групп (42 года – 49 лет) и (более 49 лет)

Возрастную группу (42 года – 49 лет) расщепляет по доле переводящих накопительную часть пенсии в НПФ уже не территориальный признак, а пользование интернетом. Не пользующиеся им россияне в возрасте от 42 до 49 лет демонстрируют более высокую вероятность оказаться «молчуном» по сравнению с теми, кто пользуется интернетом несколько раз в месяц или чаще.

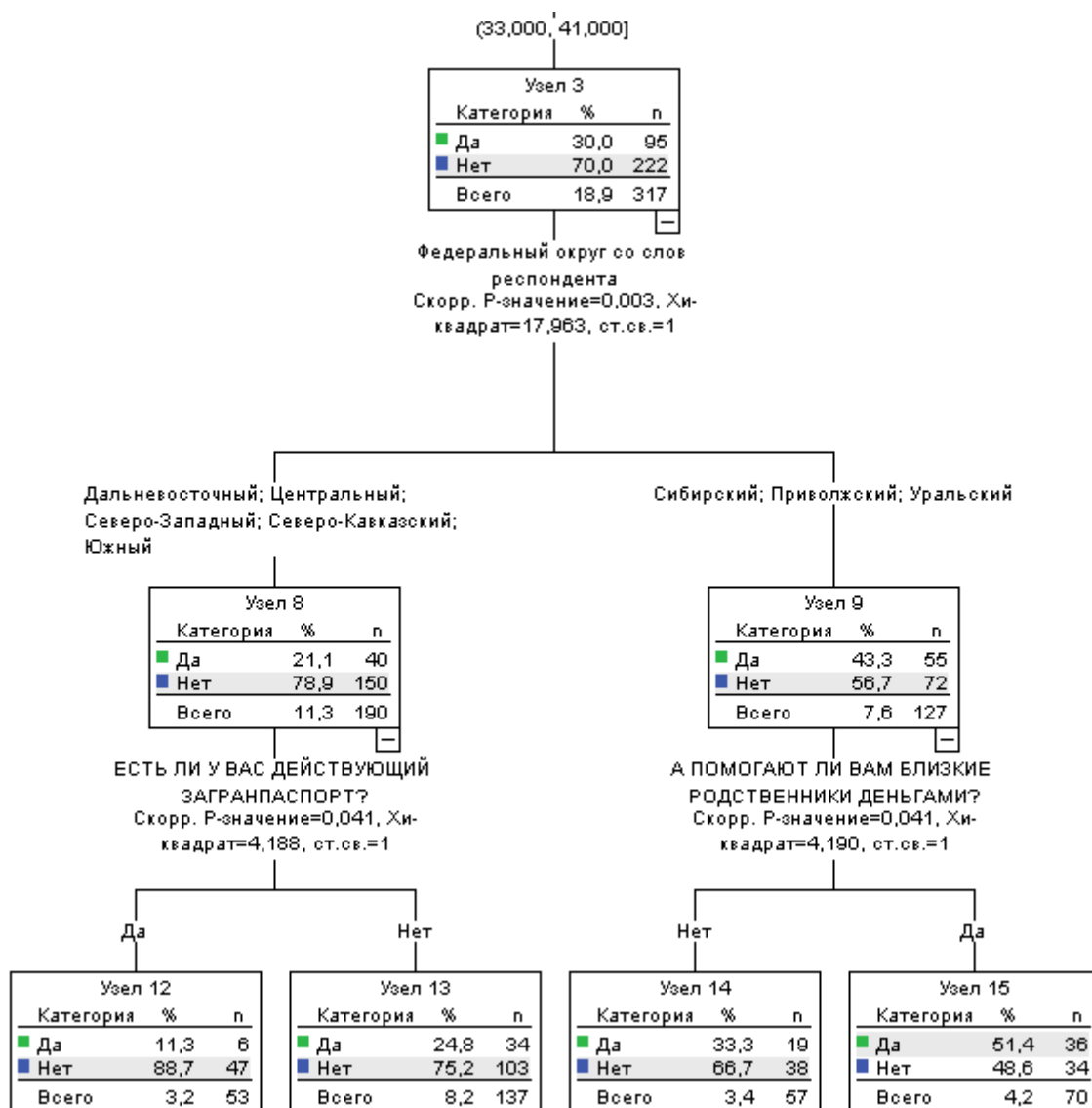


Рисунок 7 - Второй и третий уровень разбиения для возрастной группы (34 года – 41 год)

При добавлении в модель третьего уровня разбиения расщепляется на более мелкие подгруппы только возрастная группа (34 года – 41 год) (Рисунок 7). Мы можем сделать еще один важный вывод относительно структуры пенсионного поведения внутри возрастных групп: чем ближе группа находится к краю, тем она однороднее по доле тех, кто перевел накопительную часть пенсии в НПФ. Действительно, крайние группы (18 лет – 22 года) и (более 49 лет) не расщепляются вовсе, вторые с краю (23 года – 33 года) и (42 года – 49 лет) расщепляются только на втором уровне и не расщепляются на третьем. И только средняя группа (34 года – 41 год), в которой самый высокий процент «активных», расщепляется на третьем уровне по наличию загранпаспорта и финансовой помощи от близких родственников. Причем отсутствие загранпаспорта и принятие финансовой помощи повышают шансы на перевод накопительной части в НПФ. Обратим внимание, что оба признака (загранпаспорт и финансовая помощь) образуют группы на уровне значимости $P=0,041$, что уже приближается к стандартному порогу в 0,05, но все еще является значимым различием.

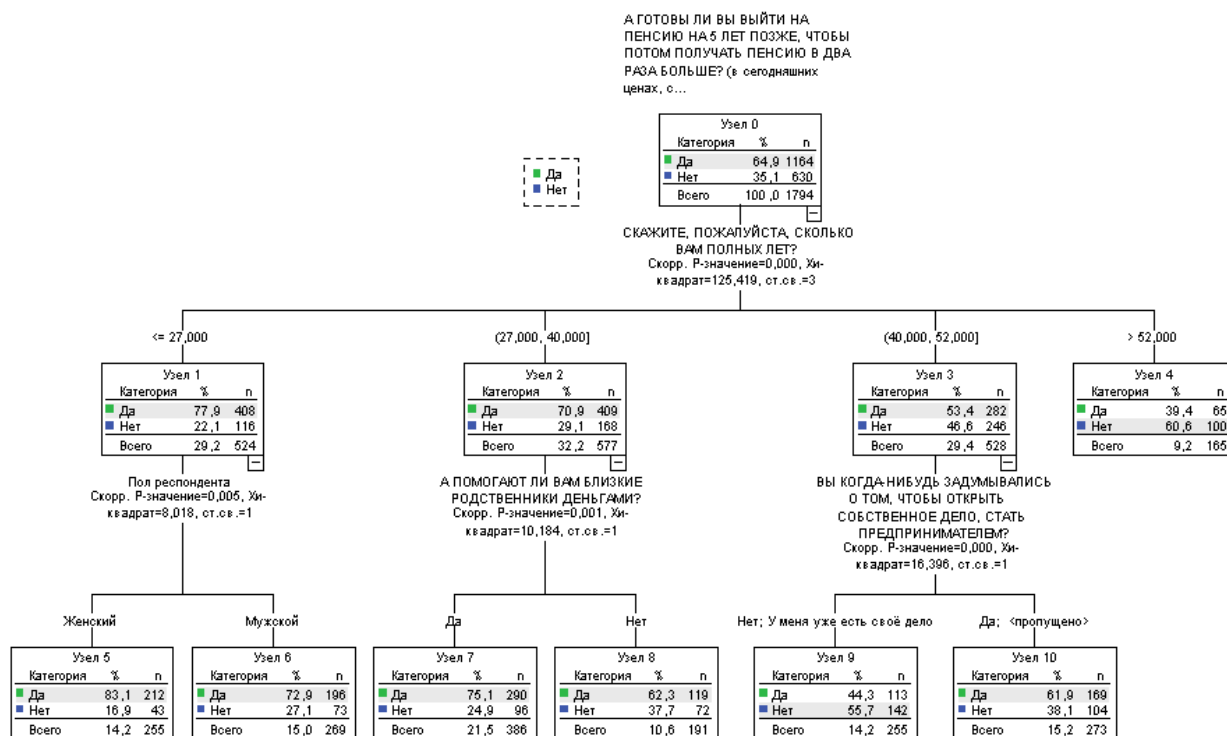


Рисунок 8 - Классификационная модель для целевой переменной «Готовы ли вы выйти на пенсию на 5 лет позже, чтобы получать в 2 раза больше».

Если говорить о классификационной модели для более слабого поведенческого признака, при ее построении самым сильным фактором также оказывается возраст. Для молодых (от 18 до 27 лет) значимый фактор - пол. Женщины в этом возрасте демонстрируют большую готовность выйти на пенсию на 5 лет позже. Для респондентов среднего возраста (от 28 до 40 лет) готовность выйти на пенсию на 5 лет позже увеличивается, если они принимают финансовую помощь от близких родственников. Для возрастной группы (от 41 года до 52 лет) специфическим признаком, влияющим на готовность к более позднему выходу на пенсию, оказывается обдумывание возможности открыть собственное дело.

Чтобы перейти от перечисления групп, выделенных по различным признакам, и указания доли целевой категории (определенного поведения) в этих группах к прогностической модели, необходимо сформулировать правило отнесения каждого респондента, попавшего в терминальный узел (узел, который далее не расщепляется) к одной из категорий целевой переменной. По умолчанию, порог отнесения составляет 50%. Например, если в 15-м узле (Рисунок 7) доля тех, кто перевел накопительную часть пенсии в НПФ, составляет 51,4% (более 50%), то все россияне, попадающие в этот узел (в возрасте 34 года – 41 год, из Сибирского, Приволжского, Уральского федеральных округов, принимающие финансовую помощь от близких родственников) будут отнесены моделью к группе «Перевел накопительную часть в НПФ». Представление о качестве классификационной модели дает таблица классификации (таблица сопряженности по целевой переменной и ее же прогнозом) (Таблица 4).

Таблица 4 Таблица классификации для модели с порогом отнесения 50%

Наблюденное	Предсказанное		
	Да	Нет	Процент корректных
Да	36	313	10,3
Нет	34	1293	97,4
Общий процент	4,2%	95,8%	79,3
Метод построения: CHAID Зависимая переменная: переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный?			

Так как объем группы тех, кто перевел накопительную часть в НПФ, значительно меньше (21%), чем объем второй категории, доля правильно предсказанных наблюдений в этой категории — всего 10,3%. Мы можем сбалансировать количество правильно предсказанных наблюдений в двух категориях (Таблица 5). Однако при этом уменьшается общий процент правильно предсказанных наблюдений (с 79% до 65%).

Таблица 5 Таблица классификации для сбалансированной модели

Наблюденное	Предсказанное		
	Да	Нет	Процент корректных
Да	202	147	57,9
Нет	428	899	67,7
Общий процент	37,6%	62,4%	65,7
Метод построения: CHAID Зависимая переменная: Переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный?			

Интегральное описание, включающее обе целевые переменные

Если говорить об интегральной модели, основанной на двух поведенческих переменных: «А готовы ли вы выйти на пенсию на 5 лет позже, чтобы потом получать пенсию в 2 раза больше?» (Да/Нет) и «Переводили ли вы накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный (из ПФР в НПФ)?» (Да/Нет), логично рассмотреть 4 группы, на которые разбиваются респонденты по этим переменным (Таблица 6).

Таблица 6 Распределение по четырем категориям, образованным двумя целевыми переменными

Поведенческая группа	Доля от ответивших на оба вопроса
Перевели в НПФ, согласны работать дольше	13%
Молчуны, согласны работать дольше	49%
Перевели в НПФ, НЕ согласны работать дольше	8%
Молчуны, НЕ согласны работать дольше	30%

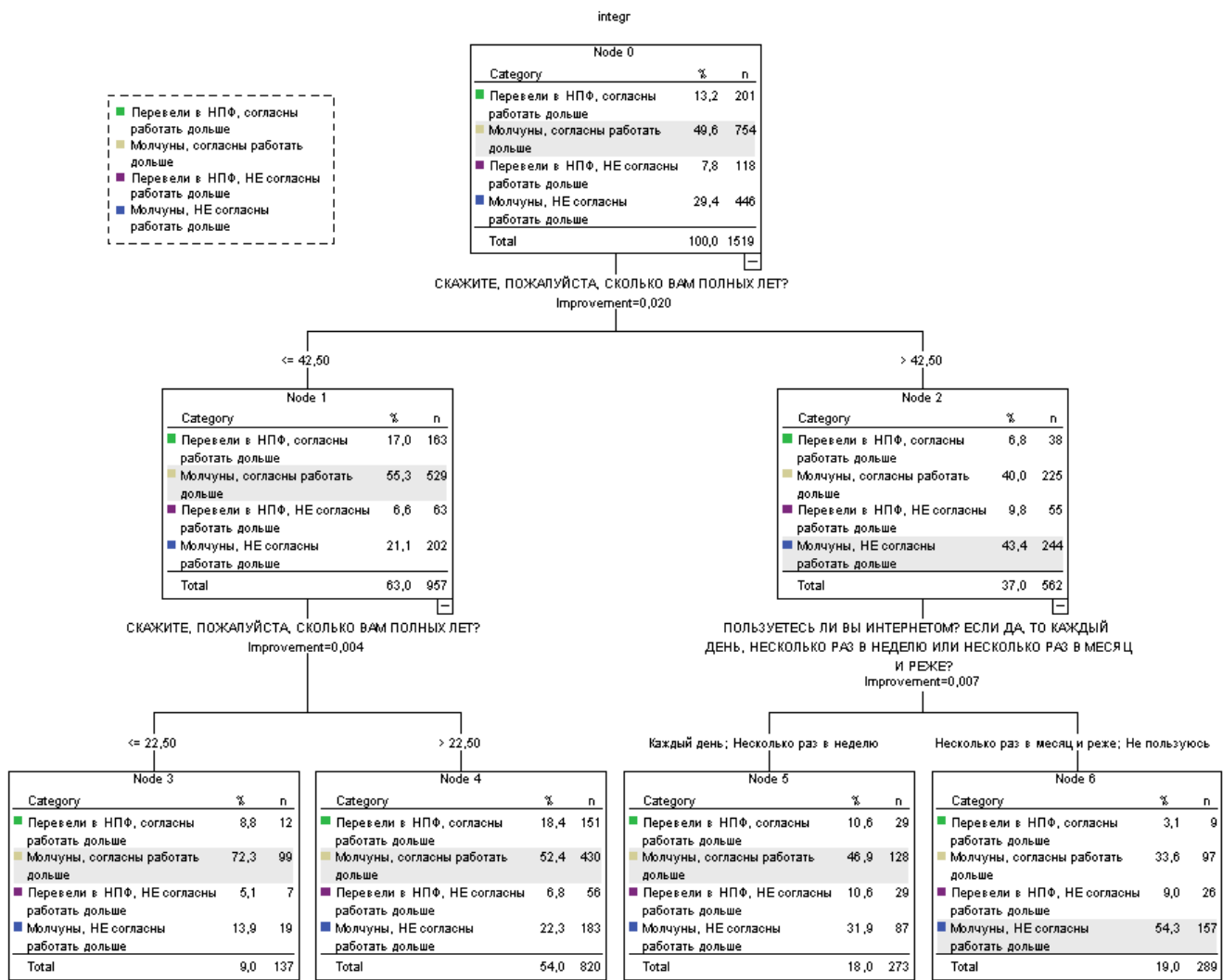


Рисунок 8 - Классификационная модель для обобщенной целевой переменной с четырьмя категориями

Анализ соответствий как способ визуализации результатов

Классификационная модель, построенная для обобщенной целевой переменной с четырьмя категориями, ставит в центр рассмотрения возраст и пользование интернетом. В результате мы имеем 4 группы, причем в каждой из них обобщенная целевая переменная имеет разные распределения. Для оценки результата воспользуемся анализом соответствий, применяющимся для визуализации таблиц сопряженности, позволяющим видеть столбцы и строки в виде точек, нанесенных на карту восприятия. В нашем случае входной таблицей для анализа соответствий будет таблица частот по категориям обобщенной целевой переменной и четырем выходным группам (Таблица 7). Точки строк и точки столбцов наносятся на одну и ту же карту и анализируются одновременно.

Таблица 7 Таблица частот по категориям обобщенной целевой переменной и четырем выходным группам

Виды пенсионных стратегий	18–22	23–42	Пользователи интернет	Не пользуются интернет
			43–59	43–59
Перевели в НПФ, согласны работать дольше	12	151	29	9
Молчуны, согласны работать дольше	99	430	128	97
Перевели в НПФ, НЕ согласны работать дольше	7	56	29	26
Молчуны, НЕ согласны работать дольше	19	183	87	157

Метод имеет ряд особенностей, которые выгодно отличают его от других. Обычная процедура при анализе таблиц сопряженности — выявление меры общей связи (при помощи параметра сравнения χ^2), однако этот метод ничего не говорит о связях отдельных строк и столбцов. Анализ соответствий решает эту проблему. А. Куталиев и А. Попов определяют анализ соответствий как особый случай метода анализа главных компонент строк и столбцов матрицы (таблицы сопряженности) [Куталиев, Попов, 2006: 302]. Тем не менее анализ соответствий и анализ главных компонент используются при разных обстоятельствах. Метод главных компонент — для анализа непрерывных величин; анализ соответствий — для анализа категориальных переменных. Основная цель анализа соответствий — графическое представление каждой строки и каждого столбца таблицы как точки на плоскости.

Для чтения карты восприятия удобно пользоваться следующими практическими правилами:

- Точки вблизи начала координат имеют малое влияние.
- Точки одного набора, расположенные далеко от центра, но близко друг от друга, имеют схожие профили.
- Геометрически отдельный профиль строки притягивается к такому положению на графике, где расположены столбцы, имеющие бросающиеся в глаза особенности в данной строке.

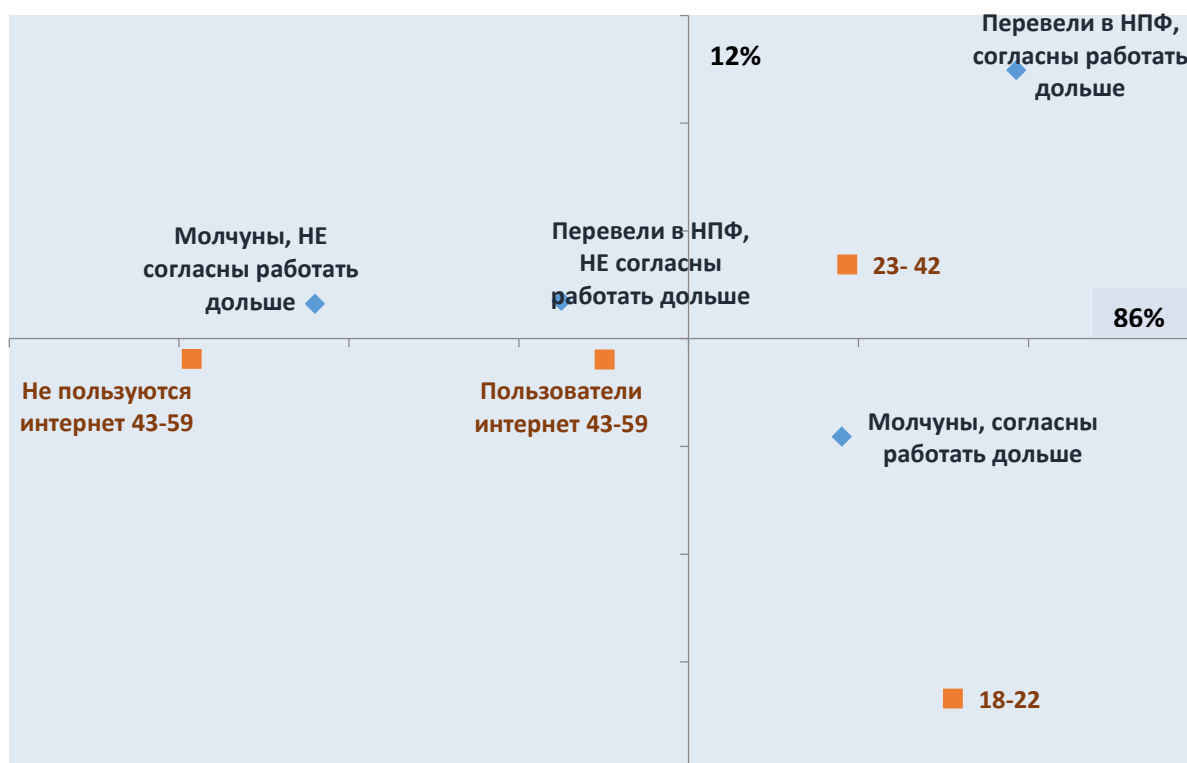


Рисунок 9 - Пример карты восприятия

Подписи на горизонтальной оси (86%) и на вертикальной оси (12%) указывают на вес осей в представлении информации. В нашем случае основная нагрузка ложится на горизонтальную ось, а дополнительное вертикальное измерение добавляет 12% информации.

Карту восприятия можно интерпретировать следующим образом.

- Пользование интернетом дополняет возраст в части предсказания поведения, образуя ось на одном конце которой молодые (18–22 года), а на другом — люди в возрасте 43–59 лет, не пользующиеся интернетом.
- Группа «Молчуны, НЕ согласны работать дольше», составляющая 30% от ответивших на оба вопроса, притягивается к группе 43–59 лет, не пользующиеся интернетом.
- Группа «Молчуны, согласны работать дольше» (49%) притягивается к противоположному концу оси «возраст–пользование интернетом», к группе 18–22 года.
- Группа «Перевели в НПФ, согласны работать дольше» притягивается (доля возрастает с 13% до 18%) к широкой возрастной группе 23–42, составляющей 54%.

Таким образом, в ходе построения прогностической модели поведения нами были сделаны следующие шаги:

- выбор целевых поведенческих переменных (зависимых переменных);

- выбор предикторов (независимых переменных);
- выбор алгоритма с учетом нелинейного характера зависимости целевой переменной от возраста;
- расчет и обоснование выбора группировки по возрасту, наилучшим образом соответствующей различным пенсионным стратегиям;
- пошаговое уточнение модели путем добавления уровней разбиения;
- оценка качества модели и балансировка долей правильно предсказанных наблюдений в категориях целевой переменной;
- построение интегральной модели с четырьмя категориями поведения;
- визуализация интегральной модели с помощью анализа соответствий.

Заключение

В ходе построения и уточнения модели была установлена нелинейность зависимости доли тех, кто перевел накопительную часть своей пенсии из государственного пенсионного фонда в негосударственный, от возраста. Максимум приходится на группу 34 года — 40 лет, в которой доля перевода накопительной части в НПФ составляет 30%.

Возрастные группы, полученные в ходе анализа, сильно различаются с точки зрения формирования пенсионных стратегий. На краях возрастной оси (до 22 лет и старше 49 лет) выделяются однородные группы. И в той, и в другой группе доля тех, кто перевел накопительную часть пенсии в НПФ, невысока (13 и 10% соответственно), а сами группы однородны, т.е. внутри них не обнаруживаются подгрупп с более высокой или низкой долей целевой категории. В средней возрастной группе (23 года — 41 год) различия по доле перевода накопительной части в НПФ привязаны к федеральным округам. Для более старшей возрастной группы (42 года — 49 лет) важным фактором, объясняющим долю перевода накопительной части в НПФ, является пользование интернетом. Среди респондентов в возрасте 42–49 лет, использующих интернет, доля перевода накопительной части в НПФ составляет 28% против 13%, среди не использующих интернет.