

Примеры использования больших данных

Сейчас Big Data – это вполне рабочий набор технологий, способный принести пользу практически во всех сферах человеческой деятельности: от медицины и охраны общественного порядка до маркетинга и продаж. Сфера использования технологий больших данных обширна. К примеру, с помощью больших данных можно узнать о предпочтениях клиентов, об эффективности производства или провести анализ рисков. Big Data – это уже устоявшаяся сфера технологий, даже несмотря на относительно молодой ее возраст, получившая распространение во многих сферах бизнеса и играющая немаловажную роль в развитии компаний, предприятий, организаций.

Сложно выбрать отрасль, в которой можно обойтись без использования больших данных. Так, например, в настоящее время компании, стремящиеся к сокращению издержек, чтобы остаться конкурентоспособными на рынке, должны использовать технологии Big Data, это позволит им получить значительный экономический эффект в виде сокращения затрат на обработку информации в производстве, логистике и управлении.

Большие IT компании то место, где зародилась наука о данных. Кампания Google, родина парадигмы Map Reduce, создала внутри себя целое подразделение, единственной целью которого является обучение своих программистов технологиям машинного обучения. И в этом кроется их конкурентное преимущество: после получения новых знаний, сотрудники будут внедрять новые методы в тех проектах Google, где они постоянно работают. Представьте себе, насколько огромен список сфер, в которых кампания может совершить революцию. Один из примеров: нейронные сети используются для оптимизации затрат энергии в центрах обработки данных.

Корпорация Apple также не отстает от поискового гиганта и внедряет машинное обучение во все свои продукты. Ее преимущество наличие большой экосистемы, в которую входят все цифровые устройства, используемые в повседневной жизни. Это позволяет Apple достигать невозможного уровня: у

кампании есть столько данных о пользователях, сколько нет ни у какой-либо другой.

Большие данные позволяют решать большое многообразие различной сложности задач, например, прогноз изменения климата, функциональные проблемы криптографии, расчет безопасности атомных станций, моделирование газодинамического обтекания тел (самолеты, автомобилестроение, движение масс воздуха в городе, моделирование лекарств, анализ генома (предсказания медицинских рисков на основе данных об активности генов, больших космологических процессов), создание новых материалов, нефтедобыча и другие фундаментальные и прикладные задачи. Приведем несколько примеров использования больших данных.

Большие данные в автопроме. Современный автомобиль – накопитель информации: он аккумулирует все данные о водителе, окружающей среде, подключенных устройствах и о себе самом. В каждом автомобиле установлено огромное количество датчиков, до сих пор большинство информации от этих датчиков оставалось внутри автомобиля, но можно собрать эти сведения, чтобы лучше понять, как на самом деле функционирует автомобиль и как покупатели используют наши автомашины. Уже скоро одно транспортное средство, которое подключено к сети интернета, наподобие той, что объединяет электромобили Tesla Model S, будет генерировать до 25 Гб данных за час. Сейчас лоббируется более сложный метод сбора данных, который в полной мере задействует Big Data. Он позволит собирать данные о состоянии водителя, например, о проблеме осанки, о том, как человек сейчас сидит. Все сидят по-разному в зависимости от длины ног, формы спины и, если установить сенсоры в каждое сиденье можно вывести индивидуальный индекс для каждого из людей, что-то вроде отпечатков пальцев. Ученые в Токио это используют как потенциальную противоугонную систему для машин. Идея в том, что угонщик садится за руль, пытается угнать, но машина опознает, что за рулем водитель без доступа и блокирует двигатель до тех, пока не будет введен пароль на приборной панели. А если бы везде была такая технология,

что можно было бы сделать? Могли бы определить индикаторы приборной панели, лучше всего оповещающие о возможности аварии в следующие 5 секунд. И тогда можно задокументировать усталость водителя, на основании этой информации «машина» смогла бы опознать, когда водитель проваливается в позу «усталости», автоматически определить и послать внутренний сигнал: вибрация руля, салонный гудок, внимание следи за дорогой. Такого типа вещи можно делать при документировании все больших аспектов нашей жизни. Использование больших данных помогает сделать машину более безопасной и функциональной. Так, компания Toyota путем встраивания информационных коммуникационных модулей (DCM) подключает новые автомобили к Toyota Big Data Center. Этот инструмент, использующийся для больших данных, обрабатывает и анализирует данные, собранные DCM, чтобы в дальнейшем извлекать из них пользу.

Большие данные в розничной торговле. Понимание пользовательских запросов и таргетинг одна из самых больших и максимально освещенных широкой публике областей применения инструментов Big Data. Крупные ритейл-сервисы используют поведенческие данные своих клиентов для продвижения бизнеса (рис.7). Большие данные в ритейле могут состоять из различных сведений о потребителях, истории их покупок, детальной информации с каждого чека, привлечших внимание скидках, фактов посещения разных магазинов и т.д.

Использование больших данных в *маркетинговых исследованиях* позволили создавать наиболее точный портрет целевого потребителя, предсказание реакции потребителей на маркетинговые «сообщения» и предложения того или иного продукта, персонализацию рекламных сообщений, оптимизацию производства и стратегию распределения, создание цифрового маркетинга и рекламно-просветительских компаний, сохранение большего числа клиентов путём наименьших трат, получение лучшего представления о собственном продукте компании и т. д.

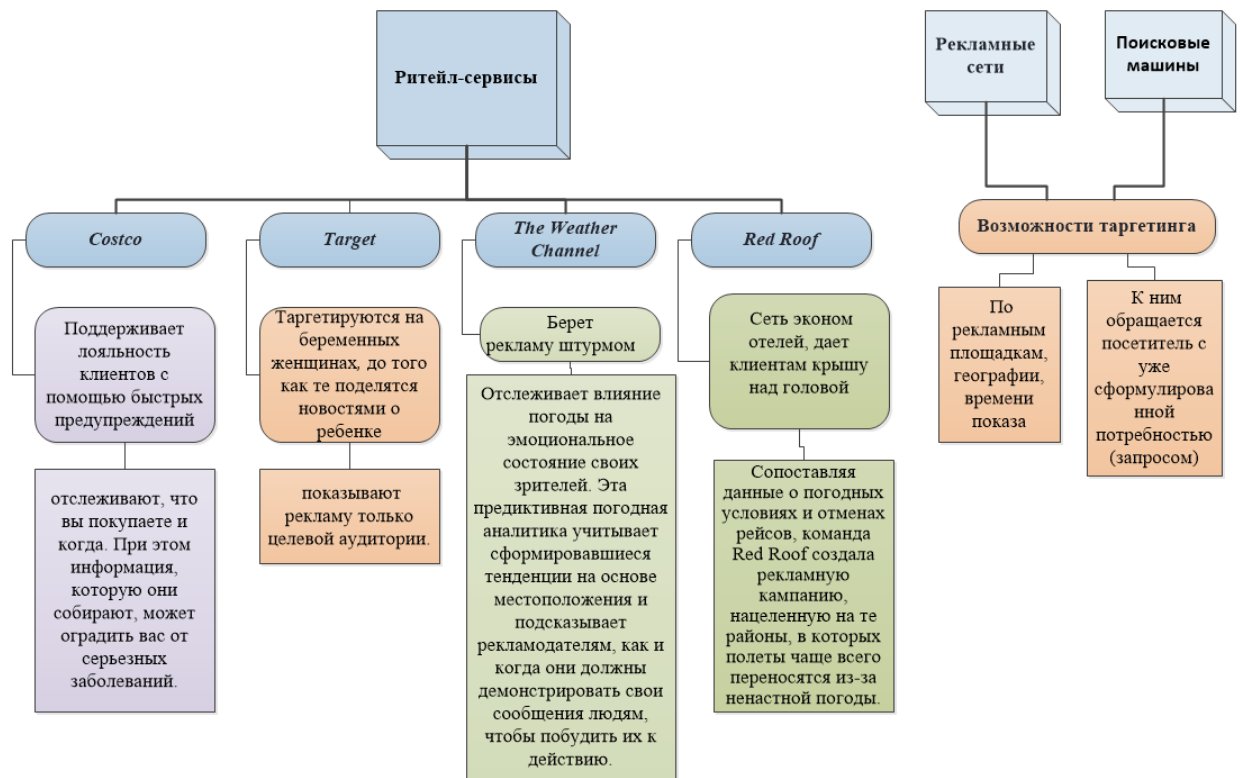


Рис. 7. Крупные ритейл-сервисы

Для использования аналитики в своей работе, бизнес должен научиться собирать и хранить данные из внутренних и внешних источников. Для этого необходимо определять способы и методы извлечения преимуществ и прибыли от использования больших данных, и разрабатывать стратегию применения больших данных. Топ-5 примеров того, как бизнес осуществляет развертывания для усиления своей конкурентоспособности представлен на рисунке 8.

Крупные компании, как Amazon, Uber и Netflix, используют «big data» для самых разнообразных аспектов бизнеса – от разработки новых продуктов до прогнозирования наиболее популярных фильмов. Большие данные определяются не только большим объемом, но также особой структурой, перед которой пасует традиционная математика. Для обработки информации такого вида нужны алгоритмы, которые считают в разы быстрее, а главное, могут приспосабливаться под задачу и обучаться самостоятельно, без вмешательства человека.



Рис.8. Топ-5 примеры того, как бизнес осуществляет развертывания для усиления своей конкурентоспособности

Большие данные в медицине. Реализация технологий Big Data в медицинской сфере позволяет врачам более тщательно изучить болезнь и выбрать эффективный курс лечения для конкретного случая. Благодаря анализу информации, медработникам становится легче предсказывать рецидивы и предпринимать превентивные меры. Как результат – более точная постановка диагноза и усовершенствованные методы лечения. Новая методика позволила взглянуть на проблемы пациентов с другой стороны, что привело к открытию ранее неизвестных источников проблемы. Сбор и анализ данных позволяет узнавать о больных намного больше: от предпочтений в еде и стиля жизни до генетической структуры ДНК и метаболитах клеток, тканей, органов. Так, Центр детской Геномной медицины в Канзас-Сити использует технологии анализа данных для быстрой расшифровки ДНК пациентов и анализа мутаций генетического кода, которые вызывают рак.

Индивидуальный подход к каждому пациенту с учетом его ДНК поднимет эффективность лечения на качественно иной уровень.

Исследователи недавно задались вопросом биопсии раковых заболеваний и задали компьютеру определить, анализируя данные и процент выживаемости, клетки раковые или нет. И конечно же, при наличии достаточного количества данных посредством алгоритма машинного обучения были выделены 12 признаков лучше всего прогнозирующих, что данный биоптат клеток рака груди действительно раковый, в медицинской литературе упоминалось лишь о 9 из них. Три признака не были известны ранее, но машина их нашла.

Большие данные на страже закона и порядка. За последние несколько лет правоохранительным структурам удалось выяснить, как и когда использовать большие данные. Общеизвестным фактом является, применяет технологии больших данных, для предотвращения террористических актов и более мелких преступлений.

Большие данные определяют развитие и коммунальной отрасли. Возможность собирать и анализировать информацию, поступающую со счётчиков учета воды, газа и электроэнергии – это первый и главный шаг на пути к умному потреблению ресурсов как на уровне домохозяйств, так и в масштабе ЖКХ-компаний. Так, например, применение больших данных позволило эстонской распределительной компании Elektrilevi, совместно с Ericsson реализующей запуск интеллектуальной системы учета электроэнергии, всего за первые два года проекта повысить эффективность на 20% и избежать дорогостоящих ошибок за счет оперативного обнаружения неисправностей.

Банки и страховые компании также обладают возможностью собирать информацию о клиентах, их действиях, финансовых транзакциях и даже перемещениях как по городу, так и по миру. Банк может определить даты важных событий в жизни своих клиентов – свадьба, рождение ребенка, смена работы, переезд и т.д. Эти сведения можно использовать для увеличения

продаж и работы над лояльностью клиентов. Так «Сбербанк» использует большие данные и машинное обучение во многих областях, в том числе в кредитном скоринге. Для решения этой задачи компания использует не только традиционные данные, такие как социально-демографические параметры, кредитная история, история транзакций, финансовая отчетность, но и ряд других. Для кредитного скоринга «Сбербанк» использует также графы связей клиентов, построенные на основе данных о денежных переводах и данных социальных сетей. Для кредитного скоринга компаний используются тексты новостей с их упоминанием, для которых проводится автоматический анализ тональности. С 2015 года компания добавила в модели данные сотовых операторов, что позволило улучшить качество классификатора по коэффициенту Джини. Большое число активных сим-карт и небольшое время их работы, мелкие и многочисленные пополнения счетов, подозрительная география звонков указывают на мошенничество и снижают вероятность одобрения кредитной заявки [13].

В телекоммуникационных компаниях большие данные – это вся служебная информация с подключенных устройств, история использования сервисов, геолокационные сведения и даже весь трафик, который может быть проанализирован, вплоть до текстов SMS. У операторов есть доступ к такого рода информации, но по закону «О персональных данных», они не могут использовать ее без согласия владельца устройства. Но могут производить, например, полнотекстовый анализ трафика, очищенного от персональных сведений. Главное, для чего используются большие данные в телекоме – более точная сегментация клиентов по типам, в соответствии с их потребительским поведением и предпочтениями. Маркетологи операторов связи, зная больше о клиентах, могут делать им более точные предложения, вовлекать их в использование дополнительных услуг, поддерживать лояльность, и тем самым больше на них зарабатывать.

Общемировая тенденция такова, что размеры этих клиентских сегментов, которые помогает выявлять Big Data, становятся все точнее, вплоть

до обращения с каждым отдельным клиентом, как с полноценным сегментом. Такой персонифицированный маркетинг – дело будущего, для операторов самых передовых государств. Но мышление маркетологов, подкрепляемое технологиями, движется именно в этом направлении, т.к. борьба за клиентов усиливается, и выигрывают те, кто умеет наладить личное взаимодействие с каждым потребителем.

Помимо использования Big Data в маркетинговых целях, телеком-операторы (впрочем, как и банки) могут применять такие технологии для обнаружения и предотвращения случаев фрода (мошеннические действия киберпреступников, направленные на воровство финансовых средств). Кроме того, операторы, как обладатели огромного количества информации о потребителях, потенциально могут стать центром экономической экосистемы, предоставляя партнерским компаниям из других сфер доступ к клиентам. В самом простом случае оператор может стать каналом маркетинговых коммуникаций, то есть попросту рассылать рекламу компании-партнера. Но делать это не массово, а адресно, точно направляя сообщение тем людям, которые могут быть в нем заинтересованы. К примеру, скидки на покупки в новом магазине косметики будут предлагаться только женщинам, живущим неподалеку. Тем, кто тратит часы, простаивая в пробках на дорогах, может прийти реклама аудиокниг.

Использование больших данных в геоинформационном пространстве для эффективного предоставления доступа к сети интернет

Одним из распространенных типов больших данных – это пространственно-временная информация о событиях. Например, это массивы данных содержащие информацию о транзакциях, совершаемых пользователями в сети Internet. Визуализация таких данных является серьезной проблемой. Чтобы найти смысл в гигантских массивах данных, сначала необходимо произвести их агрегирование. Для решения этой проблемы существует инструмент – пространственный анализ в геоинформационных системах (ГИС).

Практически 80 процентов данных в мире можно нанести на карту как слой данных в ГИС. Это относится к снимкам и информации дистанционного зондирования, огромному массиву векторных слоев, описательным и табличным базам данных, видео, коллекциям в режиме реального времени и историческим, данным с сенсоров и массивным облакам точек наблюдений. При современном уровне развития облачных вычислений, ГИС-информация в мире становится все более общедоступной. Веб-платформы дают возможность синтезировать и визуализировать данные из огромного количества источников данных в режиме реального времени. Это позволяет принимать обоснованные решения на основе актуальной, точной и своевременной информации, извлеченной из данных, в самых разных областях.

Использование больших данных не обходит стороной и сферу предоставления доступа к сети интернет в рамках поддержки и использования единого геоинформационного пространства. В рамках проведения НИР в данной области должны быть определены свойства, параметры и структуры накапливаемой информации, технологии распределенной обработки, управления данными, формирования и использования знаний о геопространстве, многомерного пространственного моделирования, представления в сети интернет. Создаваемое геоинформационное пространство становится составной частью общего информационного поля и на основе анализа больших данных предоставляет возможность отображать пространственные свойства объектов окружающего нас мира. Для эффективного предоставления доступа к сети интернет при работе с единым геоинформационным пространством требуется организовывать специализированные хранилища больших данных, получаемых на основе мониторинга. Определение состава информации позволит систематизировать данные предметной области, сформировать структуру их хранения и обеспечить возможность для поэтапного наращивания количественного и качественного состава данных в хранилище. Научное исследование в данном направлении проведено в Рыбницком филиале ПГУ им. Т.Г. Шевченко,

собраны и проанализированы данные в первом приближении, они не могут претендовать на «большие», но начало исследованию положено.

Для выполнения исследования выбрана система ArcGIS. В системе ArcGIS выполнили фильтрацию и подготовили данные, для отображения на карте и выполнения над ним пространственного анализа. Система ArcGIS позволяет объединять не только огромные объемы данных, но и данные с несопоставимыми типами. Источниками выступили как уже имеющиеся данные Государственной администрации г. Рыбница и Рыбницкого района, так и данные в обычных ГИС форматах. Пространственный анализ выполнили, как в браузере, так и в настольной геоинформационной системе.

Исследование проводилось по 23 селам Рыбницкого района: Андреевка, Пыколово, Шмалена, Колбасна, Большой Молокиш, Белочи, Броштяны, Бутучаны, Вадатурково, Воронково, Выхвантинцы, Гидирим, Ержово, Жура, Красненькое, Малый Молокиш, Михайловка, Мокра, Плоть, Попенки, Советское, Ульма, Ленино. Рассмотрены данные:

- о наличии фиксированного доступа к сети интернет;
- наличие мобильного доступа к сети интернет;
- наличие беспроводного доступа wi-fi интернет;
- наличие устройств, осуществляющих доступ к сети интернет, а именно персональный компьютер, мобильные компьютеры, такие как ноутбук, нетбук, мобильные устройства, игровая приставка, телевизор со специальным устройством Smart TV, иные средства.
- основные причины отсутствия доступа к сети интернет по селам.

Рассмотрены показатели – потребности:

- по установке точек доступа к сети интернет:
 - фиксированные точки доступа к сети интернет;
 - мобильные точки к сети интернет;
 - беспроводные точки доступа к сети интернет.
- по количеству устройств коммуникаций:

- персональные компьютеры;
- ноутбуки;
- планшеты;
- иные средства.

Анализ данных о том, какими полезными интернет-приложениями желают пользоваться жители показал, к сожалению, что основная доля – это общение и развлечение (табл. 4).

Интернет приложения

Таблица 4

Какими полезными Интернет-приложениями пользуются жители села	<ol style="list-style-type: none"> 1. Социальные сети, Skype, Odnoklassniki, доступ к различным сайтам. 2. Телевидение. 3. Портал государственных услуг. 4. Viber. 5. Просмотр фильмов.
---	--

Визуализация данных о наличии фиксированного доступа к сети интернет представлена на рисунке 9.



Рис. 9. Наличие фиксированного доступа к сети интернет

Наиболее обеспечены доступом к сети интернет села: Ерзово, Гидирим, Попенки, Бутучаны, Мокра, наименее обеспечены села: Белочи, Броштяны, Красненькое и др.

По наличию мобильного доступа к сети интернет лидируют следующие села: Ержово, Воронково, Выхватинцы, менее обеспеченные мобильным доступом: Андреевка, Шмалена, Пыкалово, Броштяны, Советское и другие (рис.10).



Рис. 10. Наличие мобильного доступа к сети интернет

Наибольшее количество Wi-fi точек наблюдается Бутучанах, в селе Плоть, а также в Ержово. В некоторых селах нет беспроводного доступа к сети (рис.11).



Рис. 11. Наличие беспроводного доступа (Wi-fi) к сети интернет

По наличию персональных компьютеров лидируют села Воронково, Выхватинцы, Ержово, Мокра. Около 50% сел имеют в наличии от 100 и выше персональных компьютеров (рис.12).



Рис. 12. Наличие персональных компьютеров в селах

По наличию мобильных компьютеров (ноутбук, нетбук) лидирует село Михайловка, в этом селе зарегистрировано наибольшее количество мобильных компьютеров, а наименее обеспеченны села – Вадатурково, Большой Молокиш, Попенки, Ульма (рис. 13).

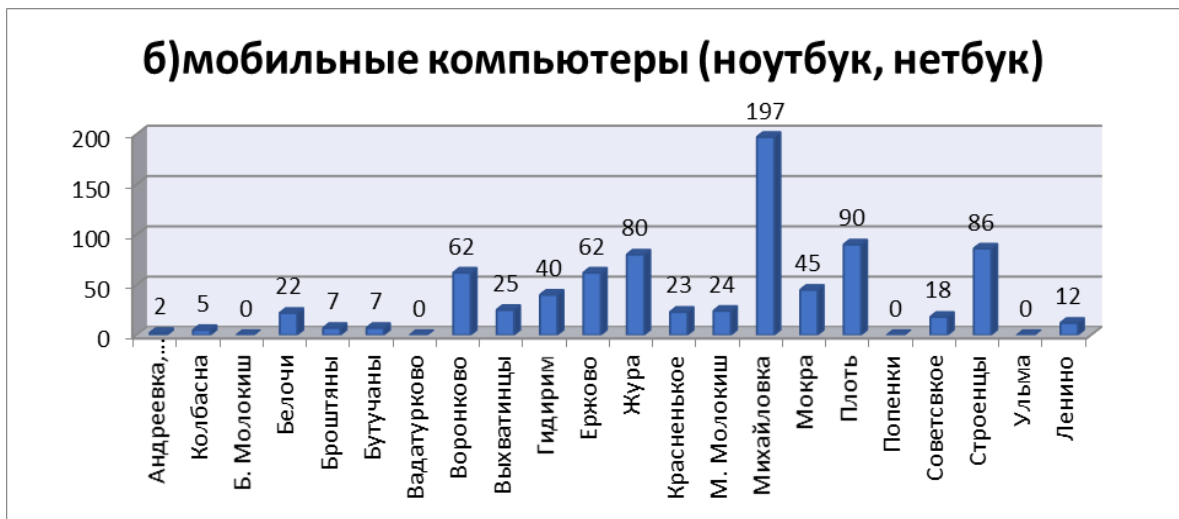


Рис. 13. Наличие мобильных компьютеров по селам

Одинаковое количество мобильных устройств, таких как: мобильный телефон или смартфон, устройство для чтения электронных книг находятся в селах Воронково и Ержово (1800 устройств), наименьшее количество мобильных устройств в селах Андреевка, Вадатурково, Михайловка, Попенки, Ульма. Игровыми приставками обеспечение в селах минимально. Например, в селе Белочи зарегистрировано 5 видеоприставок, в Красненьком – 5 и в Строенцах – 4 (рис. 14).



Рис. 14. Наличие мобильных устройств по селам

Наибольшее количество телевизоров со специальным устройством (Smart TV), зарегистрировано в селе Плоть и Советское – 50, в Гидириме – 15, и Малый Молокиш – 8. По иным устройствам лидирует Гидирим – 120 штук (рис. 15).

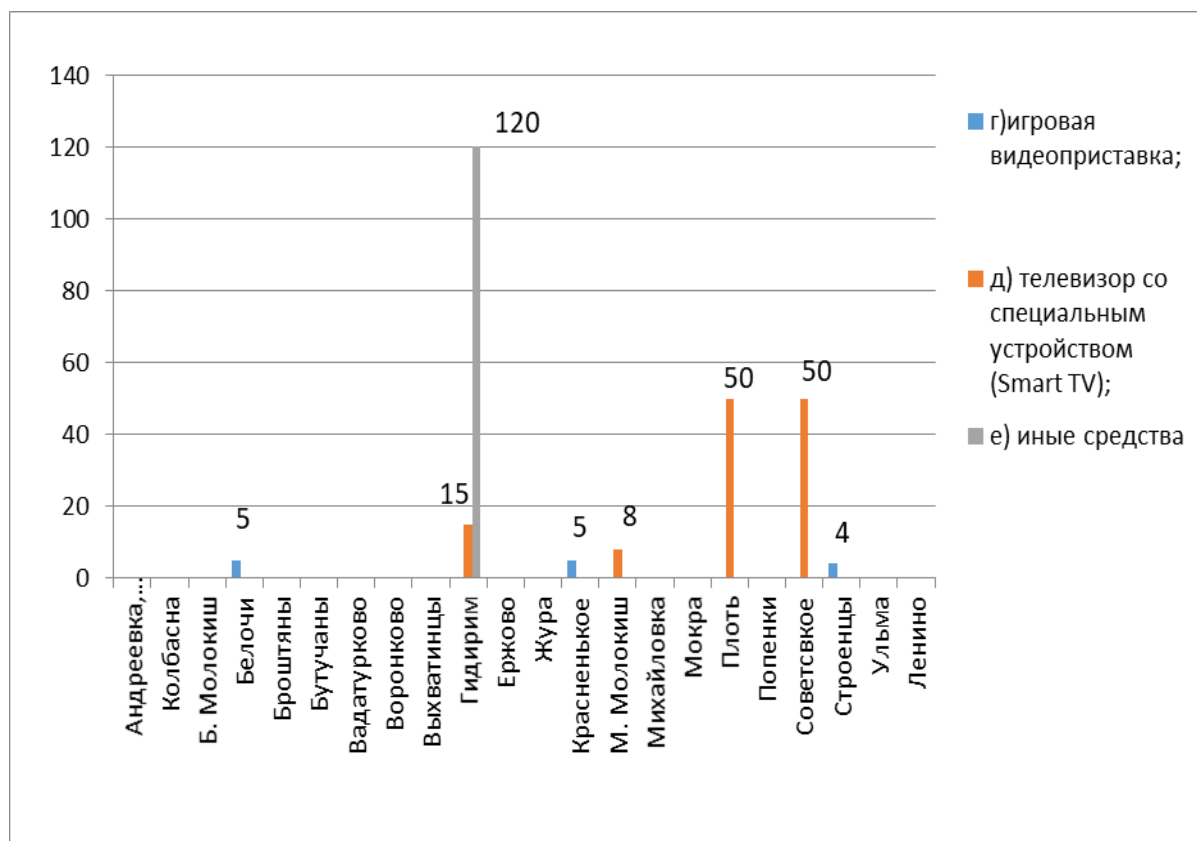


Рис. 15. Сводная диаграмма по видео приставкам, Smart TV и иным устройствам

Потребность сел Рыбницкого района по количеству устройств коммуникаций (персональные компьютеры, ноутбуки, планшеты и иные средства) представлена на рисунке 16. Например, анализ заявок в гос. Администрацию от глав сел на подключение оборудования выявил, что в Малом Молокише имеется востребованность в сельском клубе от 3 до 5 персональных компьютеров и от 3 до 5 ноутбуков. В селе Мокра необходимо по одному компьютеру в детский сад и дом культуры. В Попенках есть потребность в 3 персональных компьютерах или в 3 ноутбуках в дом культуры, а также в 1 персональном компьютере и 1 ноутбуке в детский сад. Необходим один персональный компьютер в дом культуры сел Советское и

Строенцы. В селе Ленина требуются устройства по коммуникации в администрацию села, в клуб и в библиотеку (три единицы). В селе Белочи и Красненькое есть потребность в 10 планшетах. В селе Броштяны есть потребность в: администрация села – 1 персональный компьютер, детский сад – 1 персональный компьютер, фельдшерско-акушерский пункт (ФАП) – 1 персональный компьютер, дом культуры – 1 персональный компьютер, библиотека – 1 персональный компьютер. В Воронково потребность в проведении проводной линии связи с подключением сети интернет. В селе Гершуновка, доступ к сети интернет.



Рис. 16. Потребность сел по количеству устройств коммуникаций

В селе Гидирим есть потребность по устройствам коммуникации: в

Гидиримской школе – 20 компьютеров, Гидиримский детский сад – 2 компьютера, в администрации села Гидирим – 3 компьютера, и в дом культуры села Гидирим – 2 компьютера. В селе Ержово имеется потребность в линии оптоволокна.

Графическое представление данных о потребностях сел по установке точек доступа к сети интернет, как фиксированных точек доступа к сети интернет, мобильных точек доступа к сети интернет так и беспроводных точек доступа к сети интернет представлено на рисунке 17.



Рис. 17. Потребность сел по установке точек доступа к сети интернет

Например, в Малом Молокише необходимо установить беспроводную сеть в клубе. В Попенках есть потребность в установке доступа к сети интернет в доме культуры (ДК), детском саду, ФАП и сельской врачебной

амбулатории (СВА). В селе Зозуляны необходимо установить точку доступа к сети Интернет в дом культуры. В селе Ержово есть потребность в установке точки доступа к сети интернет в библиотеке, детском саду и СВА. В селе Ленино потребность в установке точек доступа к сети интернет в администрации села, в клубе, в библиотеке, три точки беспроводного доступа к сети интернет (Wi-Fi).

Требуется точка доступа в центр села Шмалена, в детский сад и в парк села Андреевка. В селе Белочах существует потребность в установке 4 мобильных точек: администрация, школа, дом культуры, район «Дубрава». В селе Бутучаны потребность в установке точек доступа к сети интернет в администрацию села – выделенная линия, в детский сад – беспроводная точка доступа, в ФАП – беспроводная точка доступа, в дом культуры (актовый зал) – беспроводная точка доступа, в дом культуры (спорт зал) – беспроводная точка доступа, в дом культуры (летняя площадка) – беспроводная точка доступа, в библиотеку – беспроводная точка доступа. В селе Выхватинцы в ДК, музей и ФАП. В селе Гидирим требуется семь точек доступа к сети интернет: в МОУ «Гидиримский детский сад», ФАП, ДК, библиотеку, в музей. В село Строенцы: школа, детский сад, ДК, ФАП, туристическая база, магазин, администрация. Результаты исследования выявили основные проблемные причины при подключении сел к сети интернет (табл. 5).

Таблица 5

Название пункта	Основная причина отсутствия доступа к сети интернет
Андреевка, Пыкало, Шмалена	Высокие затраты на подключение. Отсутствие технической возможности подключения.
Колбасна	Высокие затраты на подключение. Отсутствие технической возможности подключения.
Б. Молокиш	Доступ с сети есть на работе, нет необходимости. Высокие затраты на подключение. Недостаток навыка для работы в сети. Отсутствие технической возможности подключения. По соображению конфиденциальности и др.
Белочи	Доступ с сети есть на работе, нет необходимости. Высокие затраты на подключение. Недостаток навыка для работы в сети.

	Отсутствие технической возможности подключения. По соображению конфиденциальности и др.
Бутучаны	Высокие затраты на подключение и на ежемесячное абонентское обслуживание.
Вадатурково	Нет необходимости, высокие затраты на подключение, отсутствие технической возможности подключения.
Воронково	Недостаток навыка работы в сети, отсутствие технической возможности подключения; по соображениям конфиденциальности, др. причины.
Выхватинцы	Высокие затраты на подключение. Отсутствие технической возможности подключения.
Гидирим	Доступ к сети есть на работе, нет необходимости. Высокие затраты на подключение. Недостаток навыка для работы. Отсутствие технической возможности подключения. По соображениям конфиденциальности, др. причины.
Ержово	Недостаток навыка работы в сети, высокие затраты на подключение
Жура	Недостаток навыка работы в сети.
Красенькое	Недостаток навыка работы в сети, высокие затраты на подключение
М. Молокиш	Нет необходимости, высокие затраты на подключение и дорогостоящее оборудование
Михайловка	Для некоторых нет необходимости, высокие затраты на подключение и дорогая плата за интернет.
Мокра	Для некоторых нет необходимости, высокие затраты на подключение, отсутствие технической возможности подключения.
Плоть	Для некоторых нет необходимости, высокие затраты на подключение, отсутствие технической возможности подключения.
Попенки	Дом культуры, детский сад отсутствие технической возможности подключения, а также отсутствие устройств, осуществляющих доступ к сети «Интернет», ФАП с. Попенки, СВА с. Зозуляны отсутствие устройств, осуществляющих доступ к сети «Интернет», отсутствие технической возможности подключения.
Советское	Нет необходимости.
Строенцы	Нет необходимости, высокие затраты на подключение, отсутствие денежных средств.
Ульма	Доступ к сети есть на работе, нет необходимости, высокие затраты на подключение, недостаток навыка для работы в сети, отсутствие технической возможности подключения, по соображениям конфиденциальности, др. причины.
Ленино	Для некоторых нет необходимости, высокие затраты на подключение и дальнейшего пользования, отсутствие технической возможности подключения.

Из данных представленных в таблице, можно выделить основные, часто встречаемые причины отсутствия сети интернет (в количестве сел) Рыбницкого района (рис. 18).



Рис. 18. Диаграмма часто встречаемых причин отсутствия сети интернет

Из представленной на сайте «Государственная администрация Рыбницкого района и г. Рыбница» – информации о количестве жителей сел Рыбницкого района были выделены такие группы населения как: школьник, пенсионеры, трудоспособное население. В результате исследования рассчитано количество жителей, пользующихся сетью интернет по каждому селу, рассчитан коэффициент обеспеченности жителей села интернетом (деления суммы всех исходных показателей на количество жителей, пользующихся интернетом) (табл. 6). Проведенное исследование позволяет сделать выводы о том, что около 30% пенсионеров пользуются сетью интернет.

Таблица 6

Обеспеченность жителей сел интернетом

Название села	Всего жителей	Школьники	Пенсионеры	Трудоспособное население	Количество пользующихся интернетом	Общее кол-во устройств	коэф. обес. точек доступа
Андреевка, Пыкало, Шмалена	590	43	136	350	434	17	0.04
Колбасна	1003	184	326	500	782	516	0.66

Б. Молокиш	889	63	264	462	604	735	1.22
Белочи	534	97	142	295	435	454	1.04
Броштяны	439	89	142	234	366	80	0.22
Бутучаны	1342	265	591	486	928	2166	2.33
Вадатурково	760	139	472	149	430	215	0.50
Воронково	2735	388	960	1387	2063	3654	1.77
Выхватинцы	1127	173	309	617	883	1938	2.20
Гидирим	1208	325	290	593	1005	1935	1.93
Ержово	2583	554	836	1193	1998	5639	2.82
Жура	1342	245	294	803	1136	581	0.51
Красненькое	1671	368	679	624	1196	1785	1.49
М. Молокиш	889	63	264	462	604	251	0.42
Михайловка	658	106	162	390	545	792	1.45
Мокра	1277	164	389	780	1061	1383	1.30
Плоть	1002	115	346	527	746	1475	1.98
Попенки	2258	293	519	1394	1843	597	0.32
Советское	500	83	344	73	259	207	0.80
Строенцы	565	35	129	313	387	465	1.20
Ульма	725	153	237	442	666	333	0.50

На основании данных о жителях сел построена общая диаграмма численности населения по селам (рис. 19). Анализ данных позволил выявить, какие села являются густонаселенными, что в селах Ержово, Воронково и Красненьком больше всего школьников и пенсионеров, а трудоспособного населения наибольшее количество в селах Ержово, Воронково и Попенках.

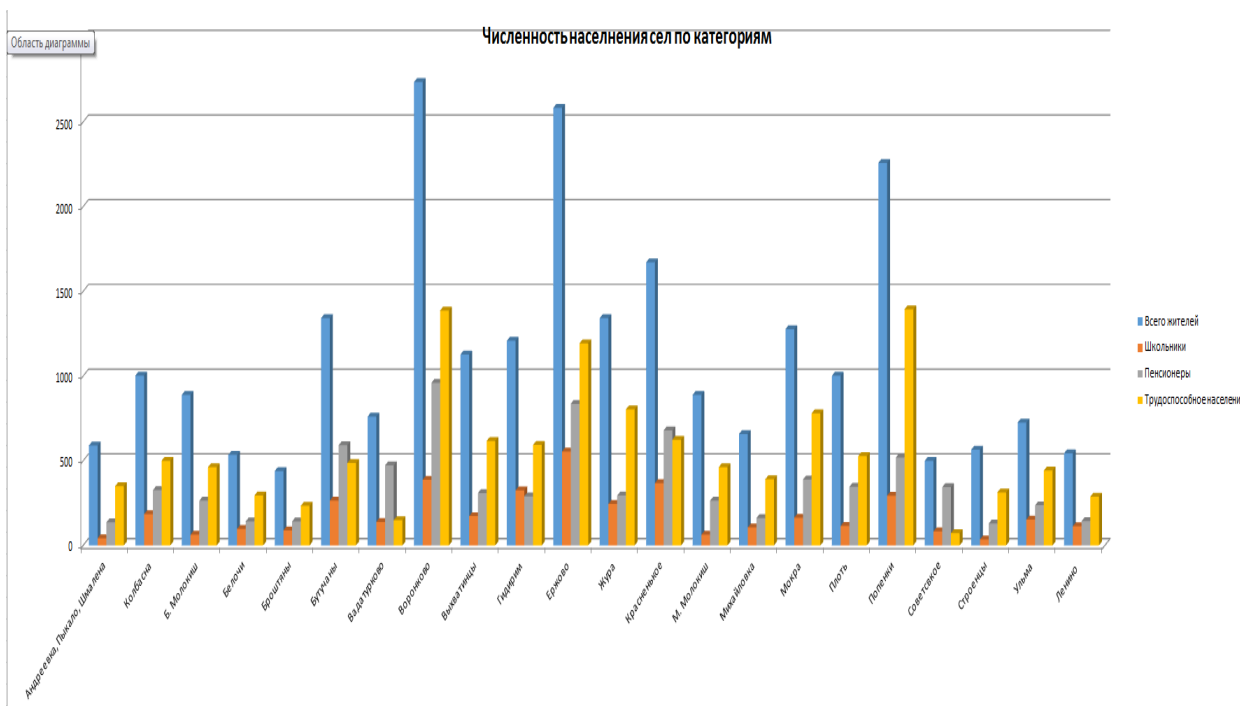


Рис. 19. Численность населения по категориям

Результаты проведенного исследования позволяют сформулировать следующие рекомендации. В селах Вадатурково, Советское, Броштяны, Строенцы, Белочи, Ленино – следует предоставить льготы на подключение к сети интернет.

В селах Ерзово, Гидирим, Попенки, Красненькое, Воронково, Бутучаны – наблюдается наибольшее количество школьников. Из этих сел в Попенках наблюдается недостаточное количество устройств коммуникаций.

На сайте «Государственная администрация Рыбницкого района и г. Рыбница» представлены данные по категориям жителей сел Рыбницкого района (табл. 7, рис. 20). Данные представлены по возрастным категориям:

- до 7 лет,
- от 7 до 15 лет,
- от 15 до 30 лет,
- старше 50 лет.

Данные по категориям жителей

Название села	Кол-во домохоз-в.	Возрастная категория				
		до 7 лет	от 7 до 15 лет	от 15 до 30 лет	от 30 до 50 лет	старше 50 лет
Андреевка, Пыкало, Шмалена	399	44	59	117	178	202
Колбасна	904	67	146	184	320	386
Б. Молокиш	541	77	87	149	269	327
Белочи	408	41	51	37	238	158
Броштяны	248	20	46	87	120	160
Бутучаны	589,606 кв.	69	149	234	409	440
Вадатурково	482	35	49	96	191	47
Воронково	1659	171	204	474	736	1032
Выхватинцы	519	78	101	147	672	559
Гидирим	598	79	96	262	349	425
Ержово	1748	193	235	520	100	1115
Жура	692	83	132	290	386	412
Красненькое	1084	66	155	461	463	687
М. Молокиш	241	31	27	65	104	126
Михайловка	371	39	62	123	206	218
Мокра	787	76	123	211	402	487
Плоть	751	51	82	141	232	370
Попенки	1034	139	177	578	660	659
Советское	390	26	40	83	140	177
Строенцы	423	50	60	102	185	195
Ульма	506	62	84	286	280	248

Анализ данных по возрастным категориям показал расхождение данных предоставленных главами сел от данных предоставленных Государственной администрацией Рыбницкого района и г. Рыбница. К примеру, по данным администрации села Андреевка школьников числится 43 человека, а по данным гос. администрации – 103 человека. По категории пенсионеры так же наблюдается разница данных. К примеру, по данным администрации села Ержово количество пенсионеров составляет 836 человек, в то время как по данным гос. администрации количество жителей старше 50 лет составляет 1115 человек.

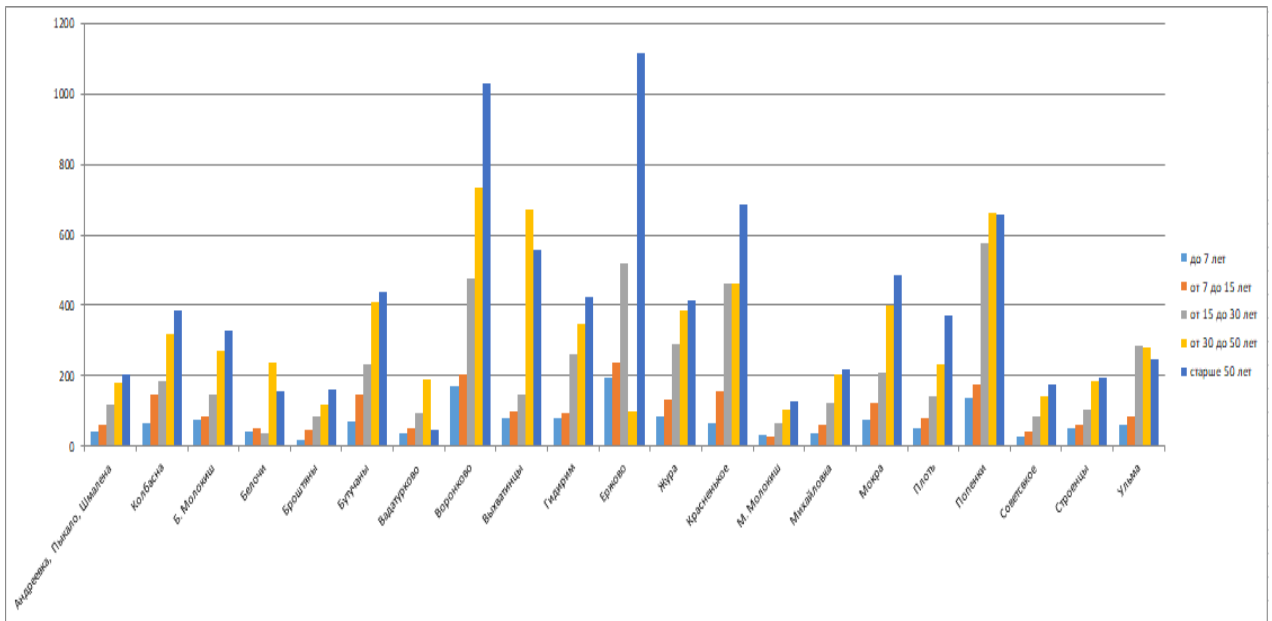


Рис. 20. Диаграмма жителей сел по возрастным категориям

На основании рассчитанного коэффициента обеспеченности жителей интернетом, построена таблица, определяющая качество обеспечение интернетом. За основу было принято, что если количество устройств, имеющихся в наличии у жителей села равно 50%, то село плохо обеспечено точками доступа к сети Интернет (коэффициенты 0,04-0,5). Если же количество жителей равно количеству устройств, то село обеспечено точками доступа к сети Интернет (коэффициенты 0,51-0,8). Если коэффициент выше 1, то село хорошо обеспечено (рис. 21).



Рис. 21. Шкала обеспеченности точками доступа к сети интернет

На основании шкалы обеспеченности точками доступа к сети интернет и расчета обеспеченности сел сетью интернет, построена диаграмма обеспеченности сел интернетом в процентном соотношении (рис.22).

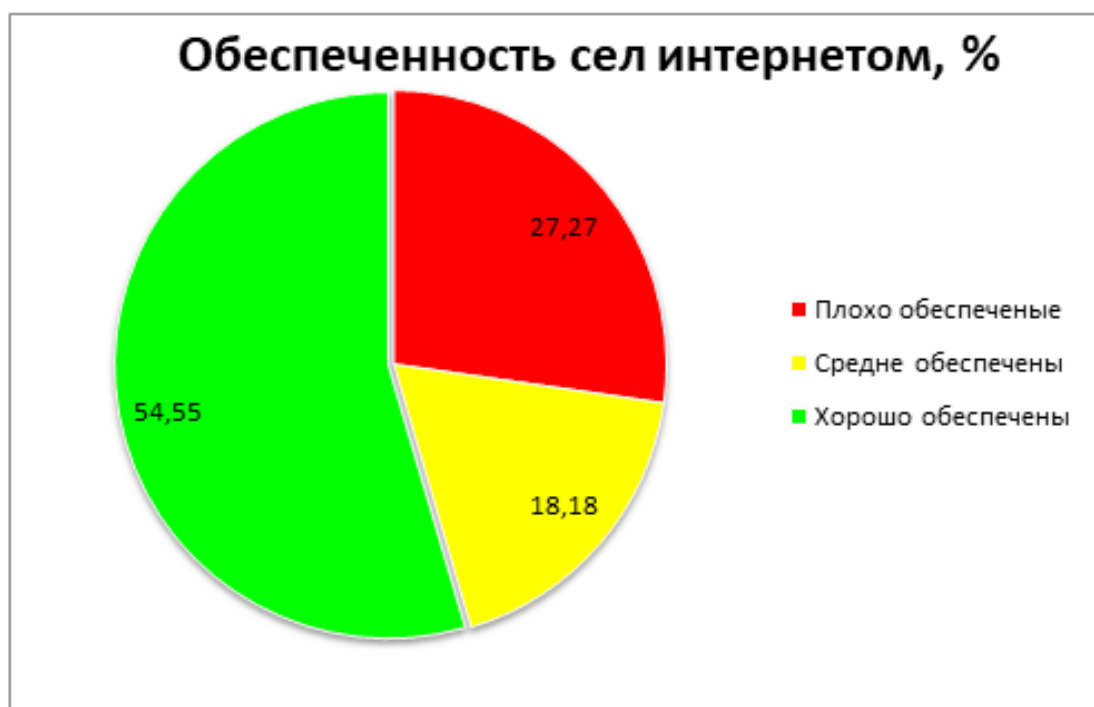






Рис. 22. Обеспеченность сел интернетом %

На диаграмме «Обеспеченность сел интернетом, %» цветами выделено состояние обеспеченности сел интернетом: красным цветом показано плохое обеспечение, желтым – средне обеспечение, и зеленым – хорошее обеспечение.

На основании данных, проведенного анализа, построена «Карта обеспеченности сел сетью интернет». На карте указаны границы сел Рыбницкого района, по которым была представлена информация Государственной администрацией Рыбницкого района и г. Рыбница и данные обычных ГИС. Исходными показателями являются «Наличие фиксированного доступа к сети интернет» (), «Наличие мобильного (гаджеты, ноутбуки, планшеты, мобильные телефоны, спутниковые точки) доступа к сети интернет» (), «Наличие беспроводного доступа (Wi-Fi) интернет» (), «Наличие устройств, осуществляющих доступ к сети интернет» (). Указанные выше элементы размещены на территориях сел, по которым проводились исследования. На рисунке 23 представлен общий вид геоинформационной системы.

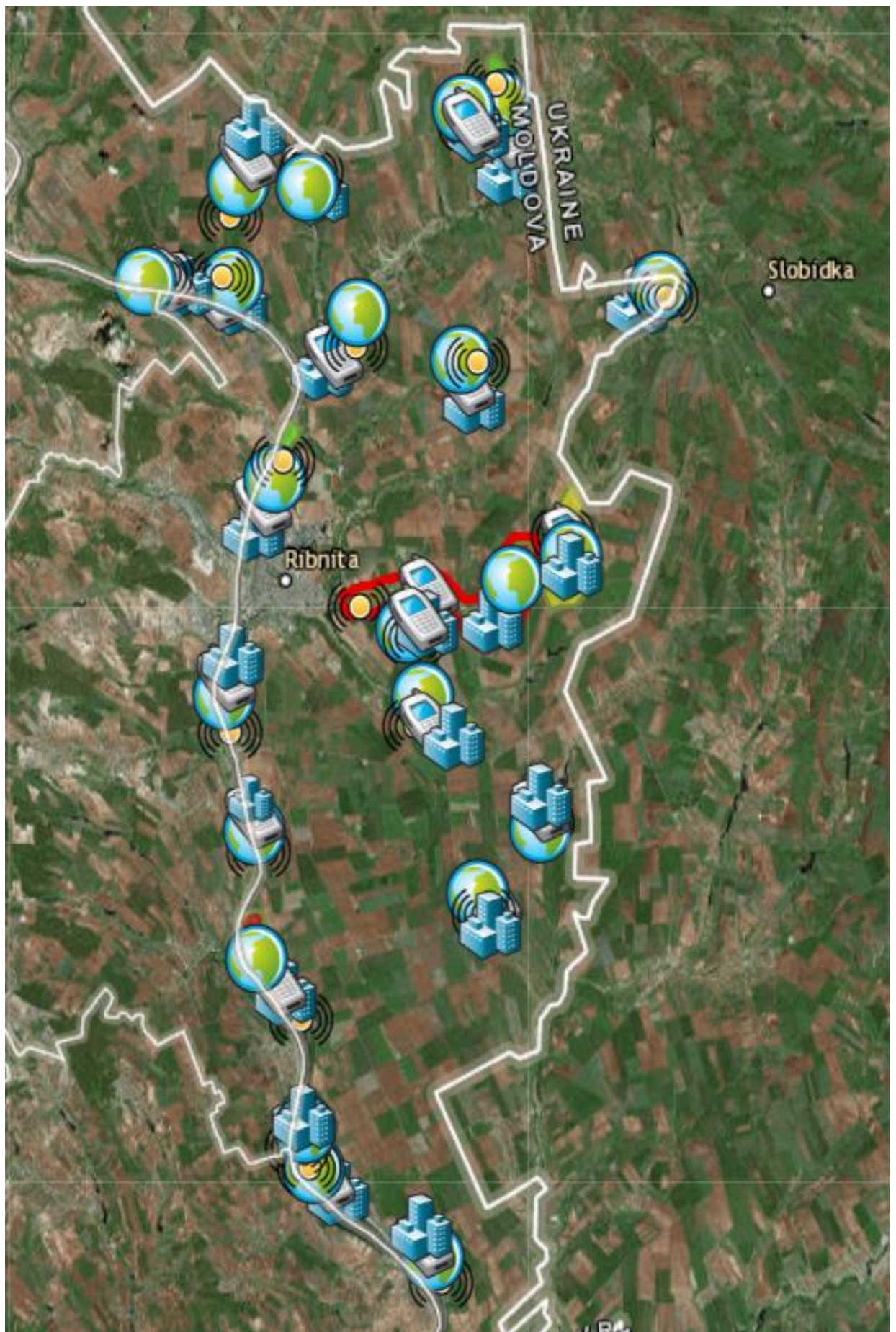


Рис. 23. Общий вид геоинформационной системы

Для каждого села Рыбницкого района отмечено на карте обеспеченности сел сетью интернетом введена общая информация. При выборе любого объекта этой карты, входящую в отмеченные границы села, всплывает окно с информацией (рис. 24).

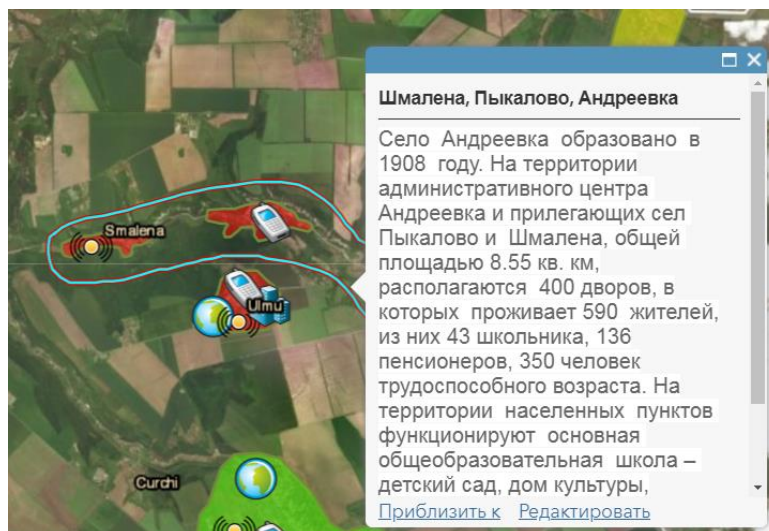


Рис. 24. Общая информация о селе

При выборе определенного элемента, всплывает соответствующая информация. На рисунке 25 показан пример выбора метки «Наличие мобильного доступа к сети интернет».

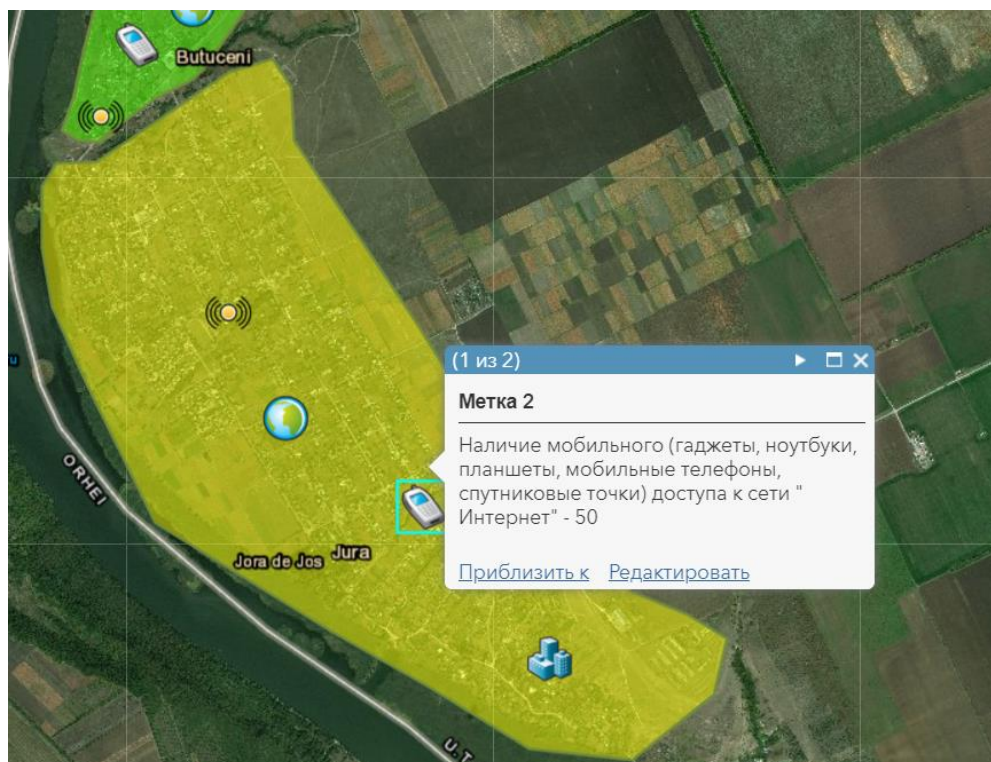


Рис. 25. Метка «Наличие мобильного доступа к сети интернет»

Исходя из рассчитанных показателей обеспеченности сел доступом к сети интернет, на карте села отмечены соответствующим цветом. Цветовое представление позволяет наглядно увидеть, какие села нуждаются в дополнительных подключениях к сети интернет (рис.26).

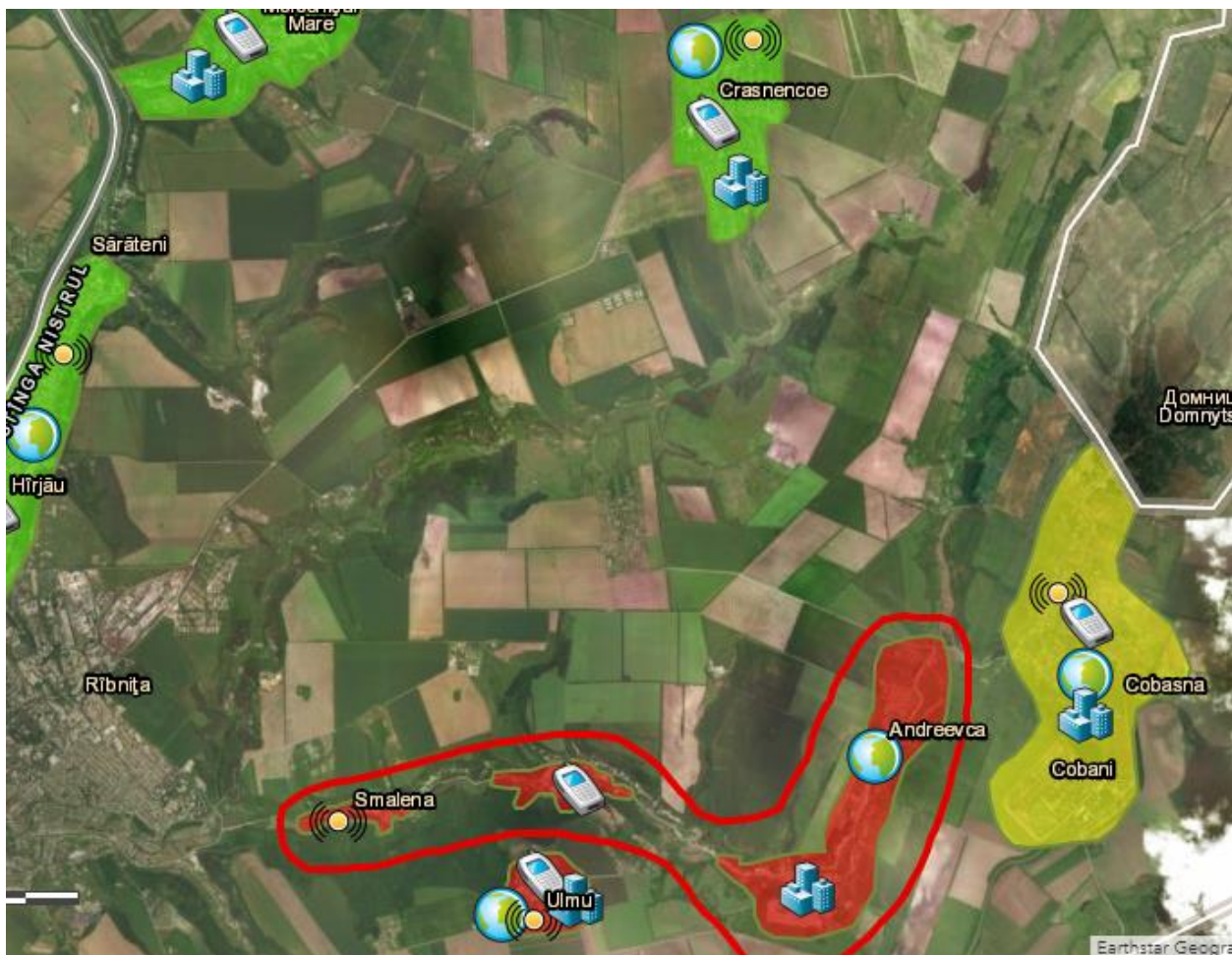


Рис. 26. Общая диаграмма обеспеченности сел в цветовом отображении

Выполненный анализ на основе информации: о существующих подключениях к сети Интернет; о количестве устройств коммуникаций, находящихся в селах; о причинах и проблемах отсутствия доступа к сети интернет позволил определить коэффициент обеспеченности жителей села интернетом и сформулировать рекомендации.

Селам Вадатурково, Советское, Броштяны, Строенцы, Белочи, Ленино – рекомендуется предоставить льготы на подключение к сети Интернет.

В селах Ержово, Гидирим, Попенки, Красненькое, Воронково, Бутучаны – наблюдается наибольшее количество школьников, во всех этих селах и

обеспеченность устройствами коммуникации сел хорошее кроме Попенки. В Попенках наблюдается недостаточное количество устройств коммуникаций.

На карте, созданной в ГИС ArcGis представлена информация по 23-м селам Рыбницкого района. Исходя из данных обеспеченности сел, каждое село наглядно выделено своим определенным цветом. В результате проделанной работы выявлено, что 6 сел (Андреевка, Пыкало, Шмалена, Броштяны, Попенки, Малый Молокиш, Вадатурково, Ульма) плохо обеспечены доступом к сети Интернет, 4 села (Жура, Колбасна, Советское, Ленино) обеспечены доступом на среднем уровне, и 12 сел (Белочи, Строенцы, Большой Молокиш, Мокра, Михайловка, Красненькое, Воронково, Гидирим, Плоть, Выхватинцы, Бутучаны, Ержово) хорошо обеспечены доступом к сети интернет.

Ссылка на ГИС: <https://arcg.is/1ObKKb>



Пример использования больших данных в ГИС для анализа покрытия сети связи на территории ПМР. Для выполнения этой задач была выбрана система Google Maps. Исходные данные являются официальными и были взяты с сайта, действующего на территории ПМР провайдера IDC (рис 27):

- уровень сигнала;
- тип сигнала;
- слои, нанесённые на карту ПМР.

Представлены следующие типы покрытий:

- CDMA 450 (рис.28).
- CDMA 800/ EDVO (рис.29, рис. 30).
- LTE/4G.

CDMA 450 МГц (IMT-MS-450) – технологический комплекс нового поколения, соединяющий в себе наивысшее качество передачи речи и исключительно высокие скорости передачи данных. CDMA (англ. Code Division Multiple Access – множественный доступ с кодовым разделением) – технология мобильной связи с каналами одной полосы частот, но отличными друг от друга кодовой модуляцией.



Рис. 27. Покрытие сети CDMA 450

Способность передавать большие объемы информации закладывалась разработчиками американской компании Qualcomm «по умолчанию», в отличие от стандарта GSM, где передача данных является так называемой «надстройкой» в телефонной системе. Свыше 250 операторов в 99 странах мира используют технологии CDMA в различных частотных диапазонах (450/800/900/2000 МГц). В России технология CDMA развивается в частотном диапазоне 450 МГц, поскольку он признан наиболее подходящим стандартом для поэтапной «цифровизации» сетей NMT-450i.

Преимущества стандарта CDMA 450 МГц.

1. Отсутствие посторонних шумов и высокая степень конфиденциальности разговоров – изначально стандарт разрабатывался для вооруженных сил США. Расшифровать сигнал без знания кода практически невозможно.

2. Повышенная емкость сети, предусматривающая реализацию безлимитных подключений как в голосовом трафике, так и в трафике передачи данных

3. Нет строгого количества каналов. С увеличением нагрузки увеличиваются потери в пакетах передачи данных, но при этом не происходит обрыва соединений.

4. Высокая скорость передачи данных, позволяющая реализовать полноценную работу в интернете и использовать все возможности бизнес-решений класса «удаленный офис» как с помощью мобильного телефона, так и подключив его к компьютеру.

5. Значительно более высокий уровень безопасности сотовых телефонов в сравнении с аппаратами других стандартов (пиковая мощность современных телефонов GSM составляет порядка 1,0 Вт, в то время как данный показатель CDMA-телефонов не превышает 0,2 Вт, что, в том числе, благотворно сказывается и на длительности работы без подзарядки).

6. Высокое качество голосовой связи, сравнимое с качеством передачи голоса в цифровых проводных линиях связи. Голосовая связь в радиоканале CDMA превосходит по качеству другие виды сотовой связи.

7. Повышенная конфиденциальность связи, позволяющая вести телефонные переговоры и передавать данные, не опасаясь прослушивания и перехвата со стороны посторонних лиц.

8. Наибольшая из всех систем сотовой связи пропускная способность базовых станций, позволяющая обеспечить 70-90 одновременных разговоров в соте одной базовой станции. По этому показателю CDMA превосходит GSM в 5-7 раз, D-AMPS в 8-10 раз.

9. Возможность получения значительной зоны радиопокрытия от одной базовой станции (радиус до 50 км). Это позволяет получить сплошное радиопокрытие при значительно меньшем числе базовых станций (в 5-7 раз по сравнению с GSM-900).

10. Наименьшая удельная стоимость канала связи в системе, что позволяет на этапе внедрения и развертывания сети получить экономию значительных финансовых ресурсов за счет меньших капиталовложений в инфраструктуру.

11. Повышенная экологическая безопасность абонентских терминалов CDMA. По этому показателю радиотелефон CDMA называют в мире «зеленым телефоном».

12. Абонентский терминал CDMA представляет собой многофункциональное устройство, которое позволяет передавать данные между компьютерами, цифровые и аналоговые факсы, а также обеспечивать прямой доступ в Интернет. Скорость передачи данных в сети CDMA уже сегодня составляет 14.4 Кбит/с, что выше аналогичного показателя в других сотовых сетях и этой услугой уже пользуются свыше 30% абонентов. Все это позволяет предоставлять в сетях CDMA услуги более высокого качества практически по тарифам.

13. Но все же главное преимущество системы CDMA заключается в ее принципиальной возможности осуществить плавный переход к сетям связи третьего поколения. При этом значительная часть уже установленного оборудования будет использоваться в новых сетях, существенно снижая капитальные затраты при внедрении услуг 3G.

На первом этапе частично модернизируется оборудование уже работающих базовых станций и устанавливается новое программное обеспечение. Этот этап развития сетей IS-95A называют первой фазой развития сети сотовой связи третьего поколения – 3G-1X. За счет организации дополнительного канала (Supplemental Channel) в среде модернизированных базовых станций появляется возможность передачи данных со скоростью 144 Кбит/с. При этом в комплекс мобильного центра коммутации (MSC) вводится дополнительное оборудование пакетной передачи данных, которое обеспечивает прямой высокоскоростной доступ в Интернет и в другие сети передачи данных.

На втором этапе формируется полнофункциональная сеть сотовой связи третьего поколения. Она позволяет передавать данные со скоростью до 2.45 Мбит/с с сохранением всех свойств сотовой сети предыдущего этапа 3G-1X и абонентской базы первоначальной сети IS-95.

Такая поэтапная эволюция сетей стандарта CDMA One, в направлении сотовых сетей третьего поколения CDMA-2000, может происходить в том же частотном диапазоне 800 МГц и в той же рабочей полосе частот 1.25 МГц (на первом этапе), которая была официально выделена операторам CDMA для развертывания сетей сотовой связи стандарта IS-95A.

Отсюда следует, что американский стандарт CDMA учитывает финансовые интересы не только операторов связи и производителей оборудования, но самое главное – интересы абонентов, которые могут свободно выбирать тот уровень услуг, который им действительно необходим в определенное время. Если для общения нужна качественная голосовая связь и низкоскоростная передача данных – можно выбрать услуги IS-95A. Если

возникла необходимость скоростного доступа в Интернет, можно модифицировать свой терминал и запросить услуги CDMA-2000-1X.

В отличие от IS-95, ни один из европейских стандартов сотовой связи не может предложить своим абонентам постепенный переход к новым услугам. Широкое внедрение GSM привело к практически полному вытеснению NMT-450, а грядущее внедрение UMTS – приведет к упадку сетей GSM в Западной Европе.

На этом фоне развитие сетей CDMA второго поколения (CDMA One) в направлении 3G имеет более благоприятную как техническую, так и экономическую перспективу, благодаря гибкости технологии и возможности ее поэтапного развития без кардинального изменения уже работающей инфраструктуры.

Операторы сетей CDMA идут именно этим путем, повышая уровень программного обеспечения, наращивая скорость передачи данных, повышая качество связи и предоставляя все новые виды услуг в своих сетях.

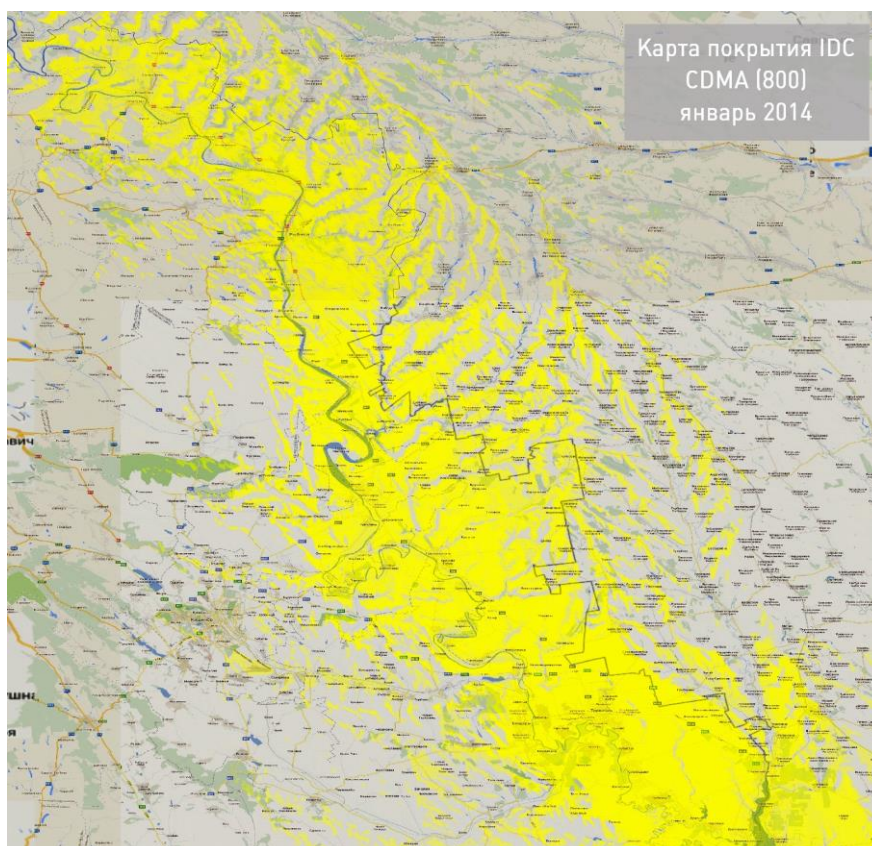


Рис. 28. Покрытие сети CDMA 800

LTE (буквально с англ. Long-Term Evolution – долговременное развитие, часто обозначается как 4G LTE) – стандарт беспроводной высокоскоростной передачи данных для мобильных телефонов и других терминалов, работающих с данными. Он основан на сетевых технологиях GSM/EDGE и UMTS/HSPA, увеличивая пропускную способность и скорость за счёт использования другого радиointерфейса вместе с улучшением ядра сети. Стандарт был разработан 3GPP (консорциум, разрабатывающий спецификации для мобильной телефонии) и определён в серии документов Release 8, с незначительными улучшениями, описанными в Release 9.

LTE является естественным обновлением как для операторов с сетью GSM/UMTS, так и для операторов с сетью CDMA2000. В разных странах используются различные частоты и полосы для LTE, что делает возможным подключать к LTE-сетям по всему миру только многодиапазонные телефоны.

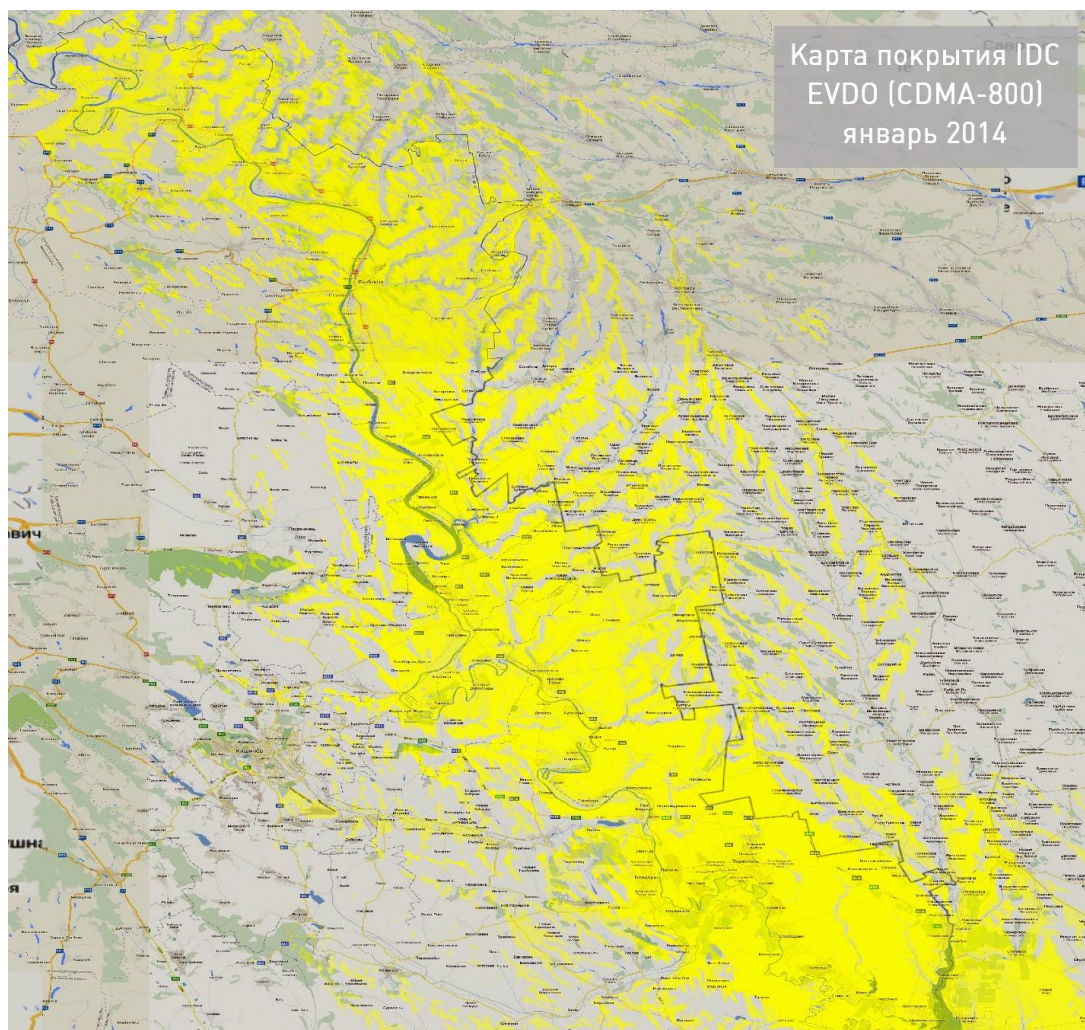


Рис. 29. Карта покрытия EV-DO сети

Создание карты зоны соприкосновения покрытия мобильной связи провайдеров ПМР и Республики Молдова

Google Earth набор приложений, построенных на основе бесплатного картографического сервиса и технологии, предоставляемых компанией Google. Сервис представляет собой карту и спутниковые снимки всего мира (а также Луны и Марса). С сервисом интегрирован бизнес-справочник и карта автомобильных дорог, с поиском маршрутов. Основу новинки составляет технология, полученная в результате приобретения фирмы Keyhole. Воспользоваться новым приложением совсем просто. Достаточно скачать бесплатную базовую версию (объем 10 Мб) и можно отправляться в «путешествие» по миру. Для отображения объемной картины земной поверхности Google Earth использует технологию широкополосной потоковой передачи и трехмерную графику. На основе данного программного продукта построена карта покрытия мобильной связи Приднестровской Молдавской Республики (рис. 30).

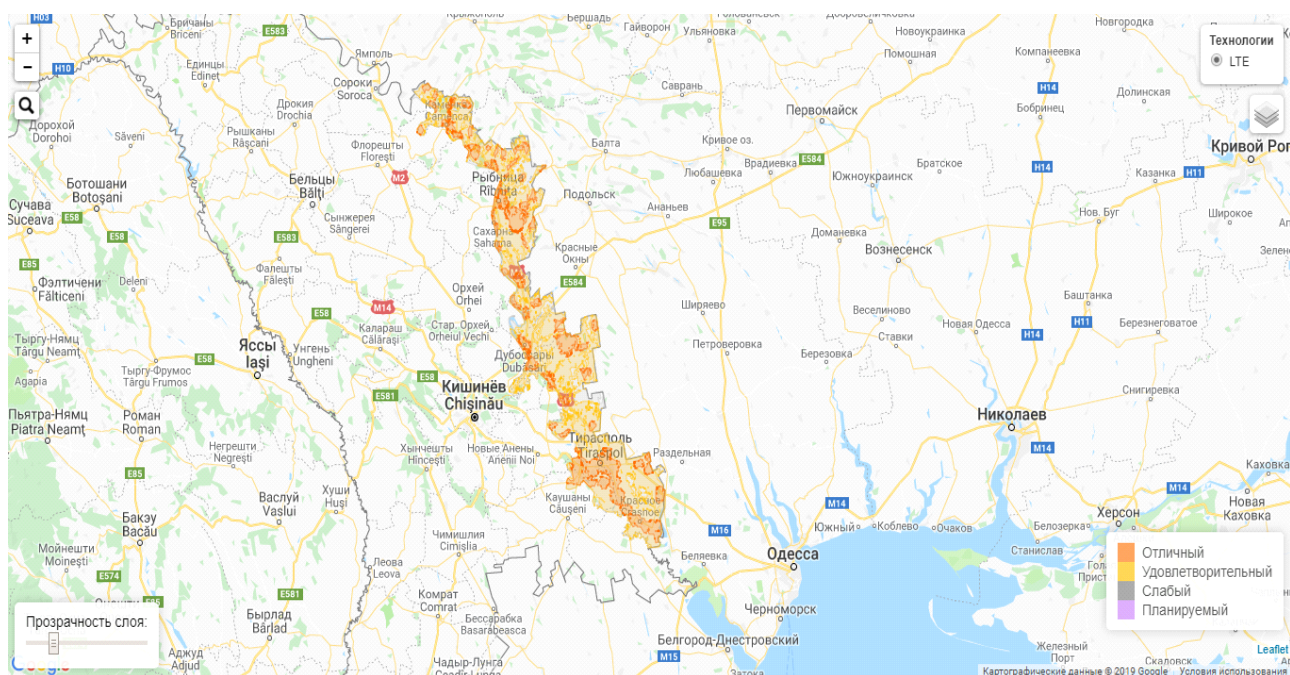


Рис. 30. Общий вид карты покрытия

Для определения зоны соприкосновения покрытия с провайдерами Республики Молдова был выбран провайдер MOLDCELL и карты покрытия с сайта провайдера (рис. 31).

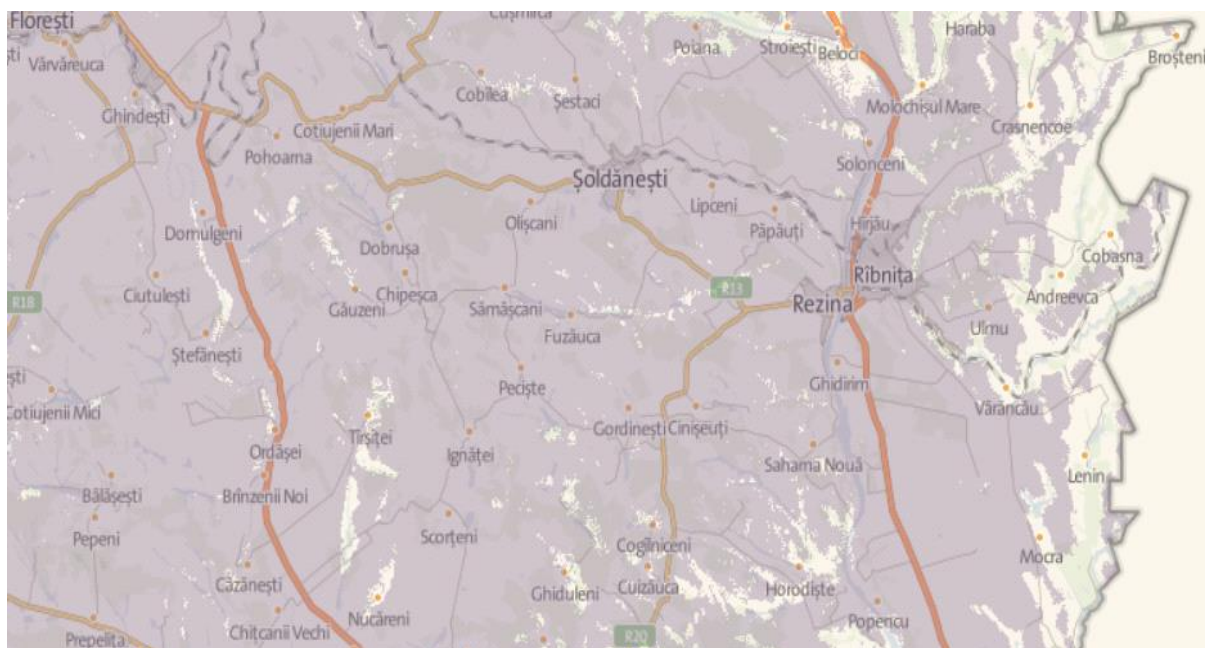


Рис. 31. Карта покрытия Рыбницкого района провайдера Moldcell

На карте фиолетовым цветом помечена зона, на которой осуществляется приём сигнала. Белым цветом отмечены районы, где сигнал отсутствует.

Карта покрытия Рыбницкого района провайдера IDC представлена на рисунке 32. На карте нанесены слои, отображающие уровень покрытия сети и зону покрытия.

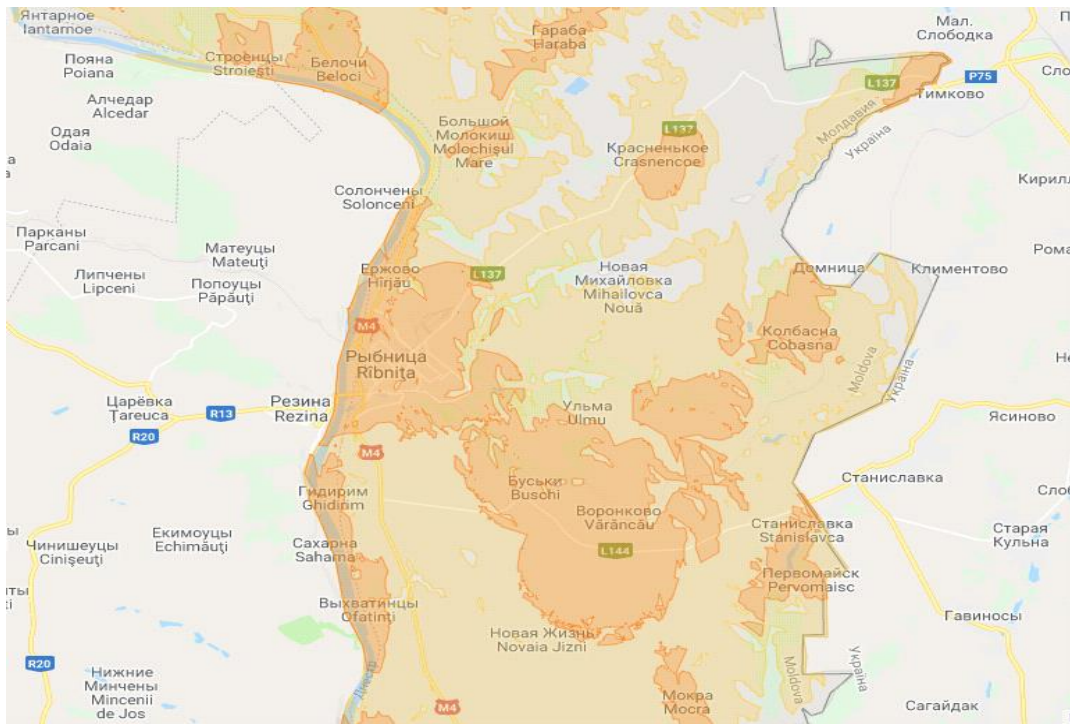


Рис. 32. Карта покрытия Рыбницкого района провайдера IDC

Как видно исходя из сравнения провайдер IDC охватывает все сёла Рыбницкого района (рис. 33 а)), а провайдер Moldcell не может обеспечить комфортную работу на территории Рыбницкого района. Такая же ситуация обстоит в центральной части республики (рис. 33 в)).

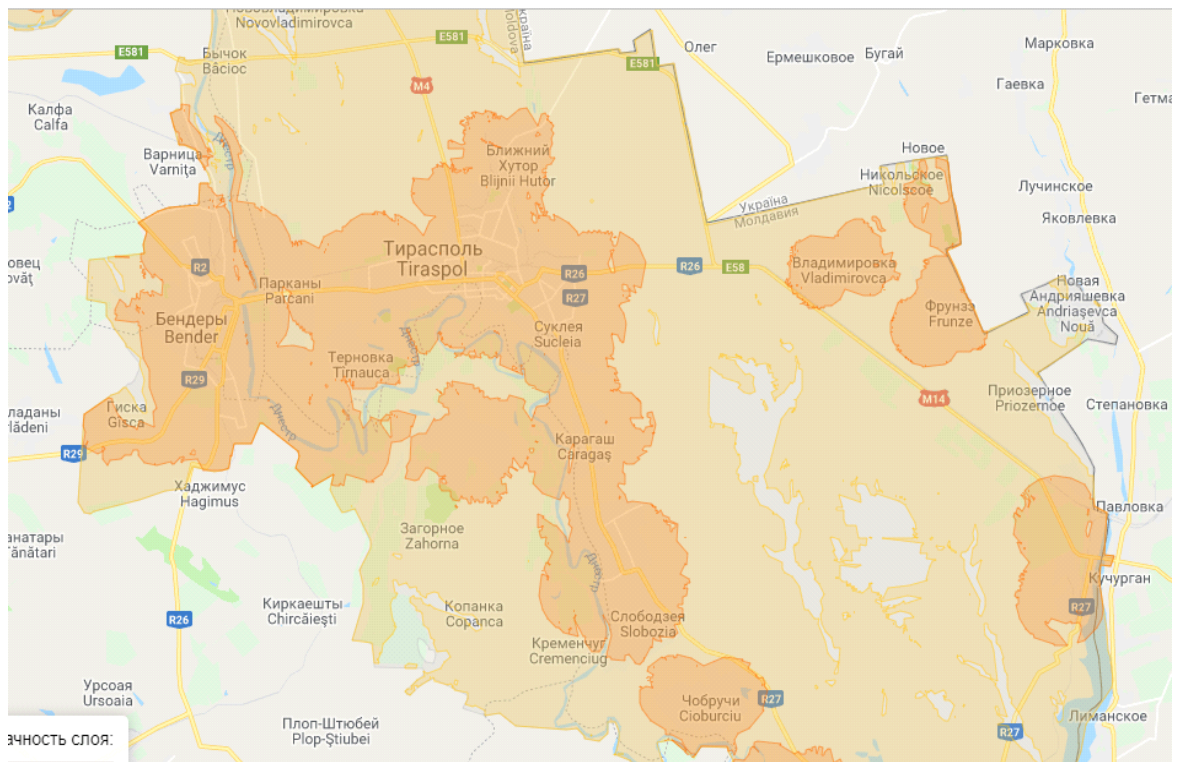


Рис. 33. а) Карта покрытия провайдера IDC



Рис. 33. в) Карта покрытия провайдера Moldcell

Исходя из анализируемых данных (информация по типу покрытия сети, о существующих типах подключениях к сети интернет и мобильной связи) выявлено, что на данный момент карта покрытия сети охватывает всю территорию ПМР, но из-за перепадов в рельефе, высоковольтных линий, и территориальной удаленности некоторых населённых пунктах республики отсутствует возможность комфортного использования услуг мобильной связи и интернета. Для решения данных проблем нужно расширять зону покрытия сети в районах, удалённых от ретранслятора сигнала. Так же для нахождения приоритетных территорий, на которых наиболее большое количество абонентов нуждается в данных услугах, а также выявления проблемных зон.

Операторы мобильной связи, осваивающие технологии больших данных, могут использовать большие данные для повышения качества обслуживания, оптимизации каналов коммуникации с клиентами, аналитики и отчетности, анализа данных для развития сети, анализа M2M-данных, борьбы с мошенничеством и спамом, персонализации услуг. Применение технологий Big Data позволяет решать задачи, в том числе управлять и

измерять качество оказания услуг на уровне каждого абонента, бороться со спамом и мобильным мошенничеством, формировать индивидуальные предложения продуктов и услуг, планировать развитие инфраструктуры связи, а также развивать розничную сеть и многое другое.

В будущем телекоммуникационным компаниям придется иметь дело с большими данными все чаще – распространение технологий M2M приведет к тому, что к 2020 году на планете будет гораздо больше подключенных устройств, чем людей. Согласно видению компании, Ericsson к 2020 году в мире будет насчитываться более 50 млрд подключенных устройств. Каждое из таких устройств будет генерировать данные, и ежемесячный трафик только лишь мобильных данных превысит 25 ЭБ. В итоге объемы информации, созданной машинами и людьми, достигнет к 2020 году, по прогнозам IDC, 44 зеттабайт.

Выводы

Направление больших данных получило широкое распространение в западных странах и России. Неоспоримо, что за большими данными большое будущее и на сегодняшний момент, они являются одним из ключевых двигателей развития информационных технологий. Но есть и темные стороны больших данных, о которых не стоит забывать и первая из них – это идея того, что мы можем быть наказаны за прогнозы, например, полиция может использовать информацию в своих интересах. Есть термин «предиктивная безопасность» или «алгоритмическая криминология», и идея о том, что, если взять много данных, например, места совершенствования преступлений можно прогнозировать, куда посылать патруль. Это логично, но проблема, конечно же, в том, что дело не закончится лишь данными местоположения. Это пойдет до уровня частного лица. Устройства, измеряющие биологические данные, покажут присутствие агрессивных мыслей. И тогда прогнозирующий алгоритм, может показать, что мы собираемся совершить преступление, и нас могут привлечь к ответственности, еще до момента действия.

Конфиденциальность была главной проблемой во время малых данных. Во время больших данных, она многократно усиливается – это проблема охрана свободы воли, свободы выбора, свободы желаний, свободы действий. Человечество лишь на подступах к эре больших данных, и еще не научилось управлять всей получаемой информацией. Это проблема не только для служб безопасности. Коммерция собирает много информации и ею так же злоупотребляет. Необходимо стать более умными в этом и это займет какое-то время. Это, примерно, как проблема огня для первобытного человека. Это орудие, но это орудие, которое, если мы не будем осторожны, нас обожжет.

Еще одна проблема – большие данные сократят количество рабочих мест. Большие данные и алгоритмы изменят работу белых воротников, профессиональные знания в 21 веке так же, как автоматизация фабрик и конвейеризация бросили вызов синим воротничкам в 20 веке. Работа профессионалов, столкнется с радикальными изменениями в их работе или даже полным ее устранением. Нам нравится думать, что технологии создают рабочие места в долгосрочной перспективе после короткого временного периода дезорганизации и это имеет место быть в реалиях, в которых мы живем. Но есть некоторые категории деятельности, которые просто исчезают и никогда не возвращаются. Индустриальная революция вам не подействовала, если бы вы были лошадью. Нам нужно быть хозяевами этой технологии, а не ее прислугой.

Большие данные преобразуют нашу жизнь, нашу работу и наше мышление. Они помогут развивать нашу карьеру и вести жизнь, полную радости и надежд, счастья и благополучия, но в прошлом мы часто, смотря на ИТ, видели только «Т» – технологии, технику потому что это вещественно. Сейчас нам нужно устремить взор на «И» – информацию, которая менее видна, но в какой-то мере более важна. Человечество, наконец-то может извлечь знания из собираемой информации, как часть наших вечных поисков в понимании мира и нашего места в нем, и поэтому большие данные – дело большое.